

РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ

АКАДЕМИЧЕСКИЙ УЧЕБНИК

Фумио Хайяши

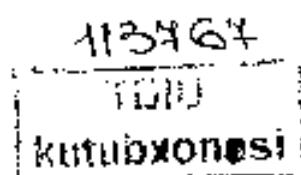
ЭКОНОМЕТРИКА



530.115(07)

X-151

Фумио Хайяши
ЭКОНОМЕТРИКА



ОРТУ

1977

*Перевод с английского
под научной редакцией В. П. Носко*

Рекомендуется Российской академией народного хозяйства и государственной службы при Президенте Российской Федерации в качестве учебника для студентов, обучающихся по экономическим направлениям и специальностям, а также для студентов бакалавриата, магистратуры, аспирантов, преподавателей экономических факультетов вузов. (Основание – приказ Министерства образования и науки №130 от 22 февраля 2017 г.)



Москва • 2017

УДК 330.4
ББК 65.05
Х 12

Перевод с английского:

А.А. Скрябин (гл. 2, 5, 6, 9, 10), В.Д. Петренко (гл. 1, 8), В.Е. Зямалов (гл. 3),
Ю.К. Ачкасов (гл. 7), А.В. Зубарев, В.Д. Петренко (гл. 4)

Хайяши, Фумио

Х 12 Эконометрика / Фумио Хайяши; пер. с англ. под науч. ред. В.П. Носко. — М.: Издательский дом «Дело» РАНХиГС, 2017. — 728 с. — (Академический учебник).

ISBN 978-5-7749-1197-4

Эконометрика Хайяши обещает стать средним большим резюме современной эконометрики. Она входит аспирантов первого года в стандартный материал аспирантских курсов эконометрики, рассматриваемый с современной точки зрения. Исследование охватывает весь стандартный материал, необходимый для понимания основных методов эконометрики, от обычного метода наименьших квадратов до коинтеграции. Книга совершенно самобытна в разработке и анализа временных рядов, и кросс-секционного анализа, предоставляя читателю единую структуру для понимания и интерпретирования результатов.

Эконометрика имеет много полезных особенностей и в сжатой форме охватывает все важные темы в эконометрике. Восемь из десяти глав включают серьезное эмпирическое применение, взятое из экономики труда, экономики отраслевых рынков, внутренних и международных финансов и макроэконометрики. Все результаты формулируются как утверждения, так что студенты понимают обсуждаемую тему, а также условия, при которых соответствующие результаты имеют силу. Большинство утверждений доказывается в тексте.

Для тех, кто собирается писать диссертацию на прикладной тематике, эмпирические приложения, собранные в книге, являются хорошим способом обучиться тому, как следует проводить эмпирическое исследование. Для скланных к теории бескомпромиссное рассмотрение основных методов является хорошей подготовкой к более продвинутым теоретическим курсам.

УДК 330.4
ББК 65 05

ISBN 978-5-7749-1197-4

Copyright © 2000 by Princeton University Press

Все права сохранены. Никакая часть данной книги не может быть воспроизведена в какой бы то ни было форме с помощью каких-либо электронных или механических средств, включая изготовление фотоконий, запись, копирование и хранение информации, без письменного разрешения издателя.

© ЧР БОУ ВО «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации», 2017

Содержание

Предисловие	17
Обязательные предварительные требования	17
Структура книги	18
Построение курса на базе этой книги	19
Контрольные вопросы и аналитические упражнения	19
Экспериментальное управление	20
Математические обозначения	21
Благодарности	21
Предисловие к русскому изданию	22
Глава 1. Свойства OLS в конечных выборках	24
1.1. Классическая линейная модель регрессии	25
Предположение линейности	25
Матричные обозначения	27
Предположение о строгой экзогенности	28
Следствия строгой экзогенности	29
Строгая экзогенность в моделях переменных ридов	31
Остальные предположения модели	32
Классическая регрессионная модель для случайных выборок	34
«Фиксированные» регрессоры	34
1.2. Алгебра метода наименьших квадратов	36
OLS минимизирует сумму квадратов остатков	36
Нормальные уравнения	37
Два выражения для OLS-оценок	39
Еще несколько терминов из алгебры	40
Анализ влиятельных наблюдений (дополнительно)	43
Замечание относительно вычисления OLS-оценок	46
1.3. Свойства OLS в конечных выборках	48
Распределение $\hat{\beta}$ в конечных выборках	48
Смещение χ^2 в конечных выборках	52

Оценка $\text{Var}(b X)$	53
1.4. Тестирование гипотез и предположения нормальности	55
Нормальное распределение ошибки	55
Тестирование гипотез об отдельных коэффициентах регрессии	57
Правило принятия решений на основании t -теста	59
Дополнительный интерес	60
и значение	61
Данные гипотезы	61
F -тест	63
Более удобное выражение для F	65
Сравнение t и F	66
Пример тестовой статистики, распределённой которой зависит от X	68
1.5. Связь с методом максимального правдоподобия	69
Метод максимального правдоподобия	70
Условная и безусловная функции правдоподобия	70
Логарифмическое правдоподобие для регрессионной модели	71
Максимизация правдоподобия с помощью нестандартизованной функции правдоподобия	71
Правила Крамера — Рао для классической регрессионной модели	73
F -тест как тест отношения правдоподобия	76
Метод квазimaxимального правдоподобия	76
1.6. Обобщённый метод наименьших квадратов (GLS)	78
Последствия ослабления предположения 1.4	79
Эффективное оценивание при известной матрице V	79
Частный случай взвешенный метод наименьших квадратов (WLS)	81
Сравнение оценок с помощью GLS	82
1.7. Приложения: отдачи от масштаба и производные в распределении электроэнергии	84
Отрасль, производство и распределение электроэнергии	84
Данные	84
Для чего нам нужна эконометрика	86
Технология Кобба — Дугласа	86
Откуда мы знаем, что действительно имеем дело с технологией Кобба — Дугласа?	88
Выполнены ли предположения OLS?	88
Метод наименьших квадратов с ограничениями	89
Тестирование однородности функций издержек	90

Лирическое отступление: об осторожности с интерпретацией K^2	91
Тестирование постоянной отдачи от масштаба	92
Важность графического анализа остатков	93
Дальнейшая разработка темы	94
Набор задач для главы 1	96
Ответы на избранные вопросы	110

Глава 2. Теория больших выборок **114**

2.1. Обзор предельных теорем для последовательностей случайных величин	115
Различные виды сходимости	115
Три полезных результата	118
Взгляд на оценки как на последовательности случайных величин	121
Законы больших чисел и центральные предельные теоремы	122
2.2. Фундаментальные понятия в анализе временных рядов	124
Необходимость эргодической стационарности	124
Различные классы случайных процессов	125
Другая формулировка отсутствия сериальной зависимости	133
CLT для эргодических стационарных последовательностей мартингал-разностей	134
2.3. Распределение OLS-оценки на больших выборках	136
Модель	137
Асимптотическое распределение OLS-оценки	141
χ^2 состоятельность	143
2.4. Тестирование гипотез	145
Тестирование линейных гипотез	145
Тест является состоятельным	147
Асимптотическая мощность	148
Тестирование нелинейных гипотез	150
2.5. Состоятельное оценивание $E(x_i^2 x_i)$	151
Использование остатков вместо ошибок	151
Преобразование S в терминах матриц данных	153
Рассмотрение конечных выборок	154
2.6. Последствия условной гомоскедастичности	154
Противопоставление условной и безусловной гомоскедастичности	155
Редукция к формулам для конечных выборок	155
Распределение на больших выборках t - и F -статистик	156

Варианты асимптотических тестов при условной гомоскедастичности	157
2.7. Тестирование условной гомоскедастичности	159
2.8. Оценивание параметризованной условной гетероскедастичности (дополнительно)	161
Функциональная форма	162
WLS с известным σ^2	163
Регрессия σ^2 на $\hat{u}$, дает состоятельную оценку σ^2	164
WLS с оцененным σ^2	164
Противопоставление OLS и WLS	165
2.9. Проекция методом наименьших квадратов	166
Оптимальное представление z в линейной комбинации переменных	166
Наилучший линейный предиктор	167
OLS действительно оценивает коэффициенты проекции	169
2.10. Тестирование серийной корреляции	170
Статистики Бокса — Пирса и Льюнга — Боков	171
Выборочные автокорреляции, вычисленные с помощью остатков	175
Тестирование с предположениями о строгой экзогенности регрессоров	175
Тест, основанный на остатках регрессии	176
2.11. Приложение: экзогенность рациональных ожиданий	179
Типичная функциональная форма	179
Тестирование следствия	181
Тестирование на серийную корреляцию	183
Является ли комбинация процентная ставка оптимальным предиктором?	185
H_t не является строго экзогенной	187
Проглашающее развитие	188
2.12. Регрессия на время	190
Асимптотическое распределение OLS-оценки	190
Тестирование гипотез в регрессиях на время	192
Приложение 2.А. Асимптотика с фиксированными регрессорами	194
Приложение 2.В. Доказательство утверждения 2.10	194
Набор задач для главы 2	197
Ответы на набранные вопросы	213

Глава 3. Одномерный GMM

3.1. Оценка из-за избыточности: пример Уордмана	218
Система одновременных уравнений рыночного равновесия	218

Смещение из-за эндогенности	219
Наблюдаемые сдвиги предложения	220
3.2. Другие примеры	223
Простая макроэкономическая модель	224
Ошибки в переменных	224
Производственная функция	226
3.3. Общая постановка задачи	228
Регрессоры и инструмент	228
Идентификация	230
Переданные условия идентификаруемости	232
Предельное асимптотическое нормальности	232
3.4. Обобщенный метод моментов (GMM)	234
Метод моментов	236
Обобщенный метод моментов	236
Оценки ошибок	237
3.5. Асимптотические свойства GMM	238
Асимптотическое распределение GMM-оценок	238
Оценки дисперсии ошибок	239
Тестирование гипотез	240
Оценки B	241
Эффективная GMM-оценка	242
Асимптотическая мощность	243
Свойства в малых выборках	244
3.6. Тестирование сверхидентифицирующих ограничений	246
Тестирование подмножеств условий ортогональности	248
3.7. Тестирование гипотез по принципу отклонения	
градоподобия	251
LM-статистика для модели регрессии	253
Тест добавления переменной (информативный	
материал)	253
3.8. Приложения условной homoskedasticity	255
Эффективный GMM становится 2SLS	256
J становится статистикой Саргана	257
Свойства 2SLS на конечных выборках	258
Альтернативные выводы 2SLS	259
Когда регрессоры предопределены	261
Тестирование подмножества условий ортогональности	262
Тестирование условий homoskedasticity	264
Тестирование на наличие эндогенности	264
3.9. Примеры: отдача от образования	265
Данные NLS-Y	266
Целулогарифмическое уравнение заработной платы	267
Смещение из-за исключения переменных	267

<i>IQ</i> как мера способностей	268
Ошибки с переменных	270
Коррекция смещения с помощью 2SLS	271
Дальнейшее развитие	272
Набор задач для главы 3	273
Ответы на избранные вопросы	293

Глава 4. Обобщенный метод моментов для систем уравнений..... 287

4.1. Модель с несколькими уравнениями	288
Линейность.....	288
Стационарность и эргодичность	289
Условия ортогональности.....	291
Идентифицируемость.....	291
Предположение для асимптотической нормальности	293
Связь с «одной» системой однострочных уравнений	294
4.2. GMM для системы уравнений	294
4.3. Теория больших выборок	296
4.4. Сравнение оценивания одного уравнения и системы уравнений	299
Когда они «эквивалентны»?.....	300
Совместное оценивание может быть эффективным	302
4.5. Частные случаи GMM оценивания системы уравнений:	
FIVE, 3SLS и SUR	303
Условная гомоскедастичность.....	303
Инструментальная оценка с полной информацией (FIVE).....	304
Трехшаговый метод наименьших квадратов (3SLS)	305
Кажущиеся несвязанными регрессии (SUR)	308
SUR и OLS.....	310
4.6. Общие коэффициенты	314
Модель с общими коэффициентами	315
GMM-оценка	316
OLS в модели дула.....	319
Приведение формул к более красивому виду.....	320
Ограничение, которое не является ограничивающим	322
4.7. Приложение. Спрос на взаимосвязанные факторы	325
Транслогарифмическая функция издержек.....	325
Доли секторов	326
Эластичности замещения	327
Свойства функции издержек	329
Стохастические спецификации	329
Природа ограничений	330

Модель множественной регрессии с перекрестными ограничениями	331
Какое уравнение исключать	333
Результаты	334
Набор задач для главы 4	336
Ответы на избранные вопросы	349

Глава 5. Панельные данные **352**

5.1. Модель компонент ошибки	353
Компоненты ошибки	354
Групповое среднее	356
Регрессионная ошибка	357
5.2. Оценка фиксированных эффектов	359
Формула	359
Свойства на больших выборках	360
Отсутствует когда η_i является сферической	362
Сравнение случайных эффектов с фиксированными	363
Остаточные условия homoskedasticity	365
5.3. Несбалансированные панели (необязательны параллели)	367
Обнуление пропущенных наблюдений	367
Обнуление против сжатия	368
Отсутствие смещения выборочного отбора	369
5.4. Подвохи: Междугрупповые различия и темп роста	371
Выделение внешнего уравнения	371
Добавление ошибок	373
Трактовка σ^2	373
Состоятельное оценивание скорости конвергенции	374
Приложение 5 А. Распределение статистики Хаусмана	376
Набор задач для главы 5	378
Ответы на избранные вопросы	392

Глава 6. Серийная корреляция **394**

6.1. Моделирование серийной корреляции: линейные процессы	394
$MA(q)$	395
$MA(\infty)$ как предел в среднем квадратическом	395
Фильтры	399
Обращение лаговых полиномов	401
6.2. Процессы ARMA	405
$AR(1)$ и его $MA(\infty)$ -представление	405
Автокорреляция $AR(1)$	408
$AR(p)$ и его $MA(\infty)$ -представление	408
$ARMA(p, q)$	410

ARMA(p, q) с общими корнями	412
Обратимость	412
Производная функция автокорреляции и спектр	413
6.3. Векторные процессы	416
6.4. Оценивание авторегрессий	421
Оценивание AR(i)	421
Оценивание AR(p)	423
Выбор глубины запаздываний	424
Оценивание VAR	427
Оценивание ARMA(p, q)	428
6.5. Асимптотика для выборочных средних сериально коррелированных процессов	430
LLN для корреляции и стационарных процессов	430
Две центральные предельные теоремы	431
Множественные обобщения	434
6.6. Включение сериальной корреляции в GMM	436
Модель и асимптотические результаты	436
Оценивание S , когда автокорреляция обращается в нуль после конечного количества лагов	437
Использование ядер для оценивания S	438
VARIAC	440
6.7. Оценивание при условной homoskedастичности (одномерно)	443
Ядрное оценивание S при условной homoskedастичности	443
Представление оцененной долговременной дисперсии через матрицу данных	444
Связь с GLS	445
6.8. Приложение: Форвардные обменные курсы как оптимальные предикторы	448
Индикатор эффективности рынка	449
Тестирование равенства нулю безусловного среднего	450
Асимптотические тесты	453
Численные задачи для главы 6	458
Открытые и нерешенные вопросы	471

Глава 7. Экстремальная оценка

7.1. Экстремальные оценки	476
«Измеримость» $\hat{\theta}$	476
Два класса экстремальных оценок	477
Метод максимального правдоподобия (ML)	478
Метод условного максимального правдоподобия	481

Инициализация оценки ML	483
Нелинейный метод наименьших квадратов (NLS)	484
Линейный и нелинейный GMM	485
7.2 Состоятельность	487
Две теоремы о состоятельности асимптотических оценок	487
Состоятельность M-оценок	490
Возвратность после репараметризации	492
Идентификация в NLS и ML	493
NLS	493
ML	494
Состоятельность GMM	499
7.3 Асимптотическая нормальность	501
Асимптотическая нормальность M-оценок	502
Состоятельность оценивания асимптотической ковариационной матрицы	505
Асимптотическая нормальность условного ML	506
Два примера	509
Асимптотическая нормальность GMM	511
Составление GMM с ML	514
Выражение для ошибки оценки в едином формате	515
7.4 Проверка гипотез	520
Нулевая гипотеза	520
Рабочие предположения	521
Статистика Вальда	522
Статистика множителей Лагранжа (LM-статистика)	524
Статистика отношения правдоподобий (LR-статистика)	526
Итоговые выводы для трех статистик	527
7.5 Численная оптимизация	529
Алгоритм Ньютона — Рафсона	530
Алгоритм Гаусса — Ньютон	530
Запись алгоритмов Ньютона — Рафсона и Гаусса — Ньютона в общем формате	531
Уравнения, нелинейные только по параметрам	531
Набор задач для главы 7	533
Опять на избранные вопросы	537

8.1. Модели с качественными откликами	539
Вывод и тесты для таблонизации T	540
Состоятельность	541
Асимптотическая нормальность	542
8.2. Многочленные регрессионные модели	543
Многочлен	543
Многочленные распределения	544
Функция правдоподобия	546
Репараметризация функции правдоподобия	547
Проверка состоятельности и асимптотической нормальности	547
Выявление некорректных предположений	549
8.3. Цензурированные регрессионные (тобит) модели	550
Функция правдоподобия для тобит-модели	551
Репараметризация	552
8.4. Многомерная регрессия	554
Переформулирование модели многомерной регрессии	554
Функция правдоподобия	555
Максимизация функции правдоподобия	557
Состоятельность и асимптотическая нормальность	558
8.5. FIML	559
Система одновременных уравнений с ошибками инструментальными переменными	559
Полная система одновременных уравнений	562
Взаимосвязь между (T_0, V_0) и Δ	563
Функция правдоподобия для FIML	564
Концентрированная функция правдоподобия для FIML	566
Тестирование сверхидентифицирующих ограничений	566
Свойства FIML-оценок	567
ML-оценивание SLR-моделей	569
8.6. FIML	571
Определение FIML	571
Выявление FIML	573
Сравнение FIML и 2SLS	574
8.7. Совместно коррелированные уравнения	576
Две переменные	577
Безусловный ML для связанных наблюдений	578
ML-оценка процесса AR(1)	579
Метод ML-оценивания AR(1) процессов	580
Метод ML-оценивания процессов AR(p) и VAR(p)	583
Набор задач главы 8	585

Глава 9. Эконометрика единичного корня	501
9.1. Моделирование трендов	591
Ускоренное представление процесса	593
Почему важно знать, является ли рассматриваемый процесс $I(1)$ -процессом	595
Что следует брать в качестве нулевой гипотезы, $I(0)$ или $I(1)$?	597
Другие подходы к моделированию трендов	598
9.2. Инструментарий для проверки единичных корней	598
Дифференцирование $I(0)$ -процессов	599
Аппроксимация $I(1)$ -процессов случайными блужданиями	600
Связь с моделями АРМА	601
Винеровский процесс	602
Полезная лекция	605
9.3. Тесты Дики — Фуллера	608
AR(1)-модель	608
Получение предельного распределения при нулевой гипотезе $I(1)$	609
Включение постоянной составляющей	613
Включение временного тренда	616
9.4. Расширенные тесты Дики — Фуллера	621
Расширенная авторегрессия	621
Предельное распределение OLS-оценок	622
Получение тестовых статистик	625
Тестирование гипотезы о ρ	627
Что делать, когда ρ неизвестно	627
Предложение для выбора $\rho_{max}(T)$	629
Включение в регрессию постоянной составляющей	630
Включение линейного тренда	633
Сводный обзор DF- и ADF-тестов и других тестов на единичный корень	635
9.5. Какой тест на единичный корень использовать	637
Асимптотика доказательств к единичности	637
Свойства и качества выборок	638
9.6. Предложения: Паритет покупательной способности	639
Смущающая зависимость от модели случайного блуждания?	640
Набор задач для главы 9	641
Ответы на избранные вопросы	655

Глава 10. Кointеграция	660
10.1. Кointегрируемые системы	661
Латентные векторные I(0)- и I(1)-процессы	661
Разложение Бевриджа—Польсона	663
Определение кointеграции	663
10.2. Асимптотические представления кointегрированных систем ..	671
Трех-элементная система Филлипса	671
VAR и кointеграция	673
Векторная модель коррекции ошибок (VECM) ..	675
ML-процедура Йохансена	677
10.3. Тестирование нулевой гипотезы об отсутствии	
кointеграции	680
Ложные регрессии	680
Тест на кointегтацию, основанный на остатках ..	681
Тестирование нулевой гипотезы о кointегтации ..	687
10.4. Статистические выводы и кointегрирующихся векторов	688
Оценка SOLS	688
Двухмерный пример	690
Предсказание двумерного процесса	691
Допущение сарвавной корреляции	692
Общий случай	695
Другие свойства и свойства на конечных выборках ..	697
10.5. Приложение: Справочник учителя SMLA	697
Данные	698
$(\alpha_1 - \mu, \mu, B)$ как кointегрированная система	698
DOLS	700
Нестабильный сплит на деньги	701
Набор задач для главы 10	704
 Приложение А. Блочные (клеточные) матрицы	
и произведении Кронекера	708
Блочные (клеточные) матрицы	708
Сложение и умножение блочных (клеточных) матриц	708
Обращение блочных (клеточных) матриц	710
Произведение Кронекера	711
 Предметный указатель	712

Предисловие

Эта книга призвана служить в качестве учебника по курсу эконометрики для студентов первого и второго курсов. Она имеет две отличительные черты. Во-первых, охватывает целый спектр методов, используя в качестве организующего принципа так называемый обобщенный метод моментов. Я считаю, что такой унифицированный подход является наиболее эффективным способом изложить материал первого курса в доступной, но еще строгой манере. Во-вторых, в большинстве глав имеется параграф, в котором детально рассматриваются оригинальные прикладные статьи из таких различных областей экономики, как экономика отраслевых рынков, труд, финансы, международная экономика и макроэкономика. Таким образом, читатель будет знать, как использовать методы, рассматриваемые в соответствующей главе, и при каких условиях они применимы.

Конспекты лекций, на основе которых базируется эта книга, использовались на протяжении нескольких последних лет в Университете Пенсильвании, Колумбийском университете, Принстонском университете, Университете Токио, Бостонском колледже, Гарвардском университете и Университете штата Огайо. Студентам, похоже, книга весьма нравится. Мой собственный опыт преподавания по ней таков, что, по мнению студентов, эта книга лучше, чем преподаватель.

Обязательные предварительные требования

Предполагается, что читатель владеет основами математического анализа, теории вероятностей и линейной алгебры. Считается, что собой представляют понимание следующих понятий: функции нескольких переменных, частные производные, интегралы, случайные события, совместные распределения, независимость, безусловные и условные математические ожидания, нормальные распределения, распределения хи-квадрат, умножение матриц, обратные матрицы, собственные значения и собственные векторы, определитель матрицы. Близкие аналогии с этими будут вводиться по ходу обсуждения. Результаты, касающиеся блочных матриц и кроссверсионных произведений, собраны в приложения. Предварительного изучения бакалаврского курса эконометрики не требуется.

1979-1984

© 1984 by John Wiley & Sons, Inc.

Структура книги

Чтобы понять организацию материала в этой книге, полезно различать *модели* и *процедуры оценивания*. Основная предпосылка эконометрики состоит в том, что экономико-социальные данные (такие как панельный REIT США) являются случайными величинами. *Модель* представляет семейство распределений вероятностей, которое могло породить эти данные. *Процедура оценивания* представляет основанный на данных протокол выбора из модели некоторого конкретного распределения, которое, скорее всего, породило эти данные. Большинство процедур оценивания, используемых в эконометрике, является тем или иным частным случаем принципа GMM-оценивания. Например, когда GMM применяется к модели, известной как классическая линейная модель регрессии, результирующая процедура есть просто обычный метод наименьших квадратов (*Ordinary Least Squares* — OLS) — основная процедура оценивания в эконометрике. Такая точка зрения является организующим принципом в первых шести главах книги, где представлено большинство стандартных процедур оценивания.

GMM можно было бы изложить сразу в главе 1, но это лишило бы читателя возможности рассматривать ряд вопросов, специфических для OLS, не отталкиваясь на GMM. По этой причине в книге используются две первые главы для представления теории OLS для конечных и бесконечных выборок. Сам GMM находится в главе 3 как обобщение OLS.

Главная новизна представления материала в книге — это трактовка процедур оценивания нескольких уравнений, таких как SUR (*Seemingly Unrelated Regression* — кажущиеся несвязанными регрессии), 3SLS (*Three-Stage Least Squares* — трехэтапный метод наименьших квадратов), метод случайных эффектов, рассматриваемый в главе 4, и метод фиксированных эффектов, рассматриваемый в главе 5, — как частных случаев GMM для одного уравнения из главы 3. В главе 6 книги обсуждение GMM завершается указанием на то, каким образом можно инкорпорировать в GMM автокоррелированность ошибок.

Для некоторых эконометрических моделей более естественным, чем GMM, принципом оценивания является принцип максимального правдоподобия (*Maximum Likelihood* — ML), который рассматривается в главах 7 и 8. Чтобы сделать связь между GMM и ML более ясной, обсуждение ML начинается в книге в главе 7 с принципа экстремальных оценок (*Extremum Estimators*), который включает ML и GMM в качестве частных случаев. В главе 8 рассматривается применение ML к различным моделям.

Книга содержит также всестороннее рассмотрение анализа временных рядов. Основной материал, изложенный в параграфе 2.2 и в первой части главы 6, дает возможность перейти к изучению важных недав-

них достижений в анализе нестационарных временных рядов, которые рассматриваются в двух последних главах книги.

Построение курса на базе этой книги

На базе этой книги можно построить несколько различных курсов.

- Если курс читается два раза в неделю по полтора часа, то восемь недель будет достаточно для покрытия ядра теории, изложенного в главах 1-4 и 6 (исключая параграфы, посвященные конкретным экономическим приложениям), главе 7 (с пропуском доказательств и примеров) и главе 8.
- Двенадцатинедельный семестровый курс может покрывать, в дополнение к ядру теории, главу 5 и экономические приложения, содержащиеся в главах 1-6.
- Краткий (скажем, шестинедельный) курс, посвященный GMM-оценке и анализу кросс-секционных и панельных данных, мог бы покрывать главы 1-5 (исключая параграфы 2.10-2.12, но включая параграф 2.2), главы 7 и 8 (исключая параграф 8.7) могли бы быть дополнительными компонентами курса, посвященной ML-оценке.
- Краткий курс временных рядов, покрывающий недавние достижения с экономическими приложениями, мог бы включать главу 1 (исключая параграф 1.7), главу 2, части глав 6 и 7 (параграфы 6.1-6.5, 6.8 и 7.1), а также главы 9 и 10. Если курс фокусируется только на теории, прикладные параграфы в главах 2, 6, 9, и 10 можно опустить.

Контрольные вопросы и аналитические упражнения

В конце каждой главы книги помещено несколько коротких контрольных вопросов с большим количеством указаний (и даже ответов). Эти вопросы можно использовать для проверки того, что читатель действительно разбирался в материале параграфа. Если не при первом, то при повторном прочтении читатели должны пытаться ответить на указанные вопросы, прежде чем двигаться дальше. Ответы на некоторые контрольные вопросы доступны на сайте книги: <http://rpr.princeton.edu/files/6946.html>

В конце каждой главы имеется несколько аналитических упражнений, предлагающих читателю доказать результаты, оставленные без доказательства в тексте, или дополнительные результаты, полезные сами по себе. Если не указано противное, аналитические упражнения можно пропустить без потери непрерывности.

Эмпирические упражнения

Каждая глава, как правило, содержит один большой эмпирический упражнение. В нем предлагаются воспроизвести эмпирические результаты оригинальной статьи, обсуждавшейся в тринадцатом параграфе этой главы, и произвести оценивание различных обобщений рассмотренной в статье модели с использованием процедур оценивания, представленных в главе. Набор данных для оценивания, который обычно тот же, что и использованный в оригинальной статье, можно скачать с упомянутого выше сайта книги.

Для применения процедуры оценивания к множеству данных тестовых должны прогнать на компьютере пакет статистических программ. В эконометрике используется довольно много статистических пакетов, включая GAUSS (www.artech.com), MATLAB (www.mathworks.com), Eviews (www.eviews.com), LIMDEP (www.limdep.com), RATS (www.estima.com), SAS (www.sas.com), STATA (www.stata.com) и TSP (www.tsp.com).

GAUSS и MATLAB отличаются от остальных пакетов тем, что они являются матричными языками, а не набором процедур. Рассмотрим, например, реализацию метода OLS посредством пакетов GAUSS или MATLAB. После загрузки данных в рабочее пространство компьютера в нескольких строках отражаются матричные операции OLS для получения OLS-оценки и сопутствующих статистик (например, R^2). В других пакетах, которые процедурно-ориентированы и иногда называются «canned packages» («стандартные, готовые»), эти несколько строк могут заменяться односторонней командой, вызывающей OLS-процедуру. Например, в пакете TSP такой командой является OLSQ.

Стандартные пакеты имеют свои преимущества и недостатки. Ясно, что для выполнения того же самого в них требуется меньше строк, что ведет к меньшим затратам времени на программирование. С другой стороны, в этих пакетах процедуры, принимающие данные и выдающие точечные оценки и сопутствующие статистики, являются по существу, черным ящиком. Из документации пакета иногда неясно, как вычисляются те или иные статистики. Хотя новые достижения в эконометрике регулярно инкорпорируются в такие пакеты, желаемые процедуры все же могут отсутствовать в текущий момент времени, и тогда возникает необходимость в написании своих собственных процедур в GAUSS или MATLAB. Но это может быть и благом, реальное выписывание соответствующих матричных операций предоставляет нам отличную возможность понять процедуру оценивания.

За несколькими исключениями, все вычисления, необходимые для эмпирических упражнений, предлагаемых в книге, можно выполнить, используя любой из вышеперечисленных стандартных пакетов. Поэтому я рекомендую использовать GAUSS или MATLAB тем аспирантам-

экономистам, которые планируют написание практической диссертации, используя современные процедуры оценивания, или написание теоретической диссертации, представляющей новую процедуру. В традиционном случае студентам следует использовать алфавит из стандартных шрифтов, упомянутых выше.

Математические обозначения

В эконометрике нет единых математических обозначений. Система обозначений в этой книге следует наиболее общепринятой, если не универсальной, практике. Векторы рассматриваются как *матрицы-столбцы*, и записываются полужирными строчными буквами. Для обозначения *матрицы* используются полужирные прописные буквы. Транспонированная матрица для A обозначается как A' . Для скалярных переменных используются (по большей части строчные) курсивные буквы.

Благодарности

Я весьма признателен за помощь, полученную от следующих лиц и учреждений: Марк Уотсон (Mark Watson), Дэйл Норберсон (Dale Norbergson), Бо Стопе (Bo Stope), Серена Нг (Serena Ng), Масан Обаса (Masan Ogasaki) и Юанг Бао (Jungtae Ba) были добротными донорами и спонсорами, чтобы использовать ранние варианты этого труда в качестве учебников для их курсов эконометрики. Комментарии, сделанные ими и их студентами, учителя и данной финальной версии, Юю Хонда (Yuzo Honda) прочитал рукопись и сделал ряд полезных предложений. Наото Кунтомо (Naoto Kuntomo), Уитни Ньюэй (Whitney Newey), Серена Нг (Serena Ng), Пьер Перрон (Pierre Perron), Джим Сток (Jim Stock), Кадзута Танака (Katsuto Tanaka), Марк Уотсон (Mark Watson), Хэл Уайт (Hal White) и Йошихиро Ядзима (Yoshihiro Yajima) выжили время для ответа на мои вопросы и запросы. Два студента Токийского университета — Мари Сасаки (Mari Sasaki) и Наоки Шимои (Naoki Shimoi) — вычитали всю рукопись с целью устранения опечаток. Их работа была поддержана грантом Zengin Foundation for Studies on Economics and Finance. Питер Дугласс (Peter Dougherty), главный редактор в Princeton University Press, обеспечил публикацию и надлежащее издание. Стефан и Хелен (Stephanie Hugel) были достаточно разносторонним экспертом по L^AT_EX, чтобы приспособиться к моим пристрастиям в разметке набора. Элли Фус (Ellen Foss) контролировала производство книги. Наконец, что не менее важно, Джессика Хелфорд (Jessica Helford) согласилась, вероятно по дружбе, выполнить дизайн обложки.

В течение более пяти лет вся моя свободное время уходило на подготовку этой книги. Теперь, когда книга известна себе как человек, только что победивший свой страх перед книгой. Моя последовательная деятельность, истощавшая, но надеюсь, профессия этого не заметила.

Предисловие к русскому изданию

Я весьма польщен тем, что мой учебник «Эконометрика» предлагается российским читателям.

Прошло уже 15 лет с момента появления этой книги на английском языке. За эти годы у меня была возможность общаться со студентами из России, слушавшими мой курс или комментировавшими содержание этой книги. Я был впечатлен их технической компетентностью, причина которой, как я полагаю, кроется в математическом образовании в России. Ведь Россия — это страна, которая дала Ляпунова, Слуцкого, Колмогорова, Гордину и многие другие апалитические умы. И поэтому я со скаринностью и в то же время с удовольствием воспринимаю то, что моя книга, которая не пытается идти на компромиссы в отношении строгости, наконец достигает широкого круга студентов с сильной базой знаний.

В 1980-е и 1990-е годы произошли два важных изменения в стиле и содержании эмпирических исследований. Одним из них является широкое использование GMM (обобщенного метода моментов), а другим является изменение анализа временных рядов на процессы с единичным корнем (описываемые двумя последними главами этой книги). Конечно, авторитетная программа аспирантуры хотела бы предложить для второго года курсы, посвященные этим новым разработкам. Однако я пришел из опыта преподавания профильного курса эконометрики для студентов первого года аспирантуры, что вместо того, чтобы обобщенный метод моментов относить на второй год обучения, им можно — и нужно — заменить традиционное изложение базового материала на первом году обучения. На сегодняшний день, моя книга остается единственным базовым учебником эконометрики, использующим GMM в качестве организующего принципа.

Выглядела бы моя книга иначе, если бы я начал писать книгу? Наверное. Разумеется, многое произошло с тех пор, как эта книга была опубликована на английском языке в 2000 году. Два наиболее важных продвижения произошли в использовании контролируемых экспериментов и в байесовской эконометрике. Первое только подчеркивает важность методологии, уже хорошо освещенной в книге. Как вы увидите из главы 3, вам достаточно принять лишь один предположение, которое сделает пригодным метод инструментальных переменных. Как вы

калечиваемых экспериментов — сделать доступными наборы данных, для которых это идентифицирующее предположение гарантировано не выполняется. В отличие от этого возрастающая популярность байесовской эконометрики, несомненно обязанная постоянному снижению стоимости вычислений, может в один прекрасный день заставить серьезно пересмотреть материал первого года. Это потому, что байесовская точка зрения радикально отличается от классического, или частотного, подхода, который принят в большинстве профильных учебников эконометрики, в том числе и в моем. Однако смена парадигмы, если она вообще состоится, дело далекого будущего.

Поэтому я уверен, что моя книга предлагает наиболее эффективный и по сей день путь обучения эконометрике на первом году аспирантуры.

Фрэнк Хайкин
Июль 2015

Глава 1. Свойства OLS в конечных выборках

Оценивание методом наименьших квадратов (*Ordinary Least Squares OLS*) является основной процедурой оценивания в эконометрике. В этой главе описываются свойства OLS-оценок¹ в малых, или **конечных, выборках**, то есть статистические свойства OLS-оценок, которые действительны для выборки любого заданного размера. Материалы, представленные в этой главе, совершенно стандартны. Изложение здесь отличается от большинства других книг поддержкой роли, которую играет предположение о «строгой экзотенности» регрессоров.

В заключительном параграфе мы применяем теорию для конечных выборок для оценивания функций издержек, используя кросс-секционные данные по отдельным фирмам. Важнее практическое знание имеет вопрос, поставленный в работе [Nerlove, 1963]: имеет ли место возрастающая отдача от масштаба в отрасли электроснабжения? Если да, то микроэкономика говорит вам, что отрасль должна регулироваться государством. Кроме предоставления вам практических навыков использования методов для проверки интересующих гипотез в работе Нерлова имеется тщательное обсуждение вопроса о том, почему OLS является подходящим методом оценивания в конкретном прикладном исследовании.

¹В оригинале здесь используется термин *estimator* («оценщик»). Обычно при переводе этого термина на русский язык используют термин «оценка» и для метода (функции), и для результата применения этого метода (формулы) в конкретной выборке данных, то есть для числового значения, которое при этом получается. Поскольку сам автор, скорее всего (см. абзац, следующий за формулой (1.2.5)) о том, что в этой книге оба термина *estimator* («оценщик») и *estimate* («оценка») используются практически как синонимы, мы будем придерживаться указанной выше практики перевода. — *Прим. перев.*

1.1. Классическая линейная модель регрессии

В данном параграфе мы приводим предположения, на которых базируется классическая линейная модель регрессии.

В этой модели рассматриваемая переменная (которая называется **зависимой переменной** (*dependent variable*), **регрессируемой переменной** (*regressand*), или, по происхождению, **переменной в левой части** (*left-hand / side variable*), связана с несколькими другими переменными (которые называются **регрессорами** (*regressors*), **объясняющими переменными** (*explanatory variables*), или **переменными в правой части** (*right-hand / side variables*)). Допустим, мы наблюдаем n значений этих переменных. Пусть y_i обозначает i -е наблюдение зависимой переменной, а $(x_{i1}, x_{i2}, \dots, x_{ik})$ — i -е наблюдение k регрессоров. **Выборка** (*sample*) или **данные** (*data*) — это набор этих n наблюдений.

Данные и коэффициенты не могут быть получены в результате экспериментов (исключения экспериментальную экономику), так что и зависимая, и объясняющая¹ переменные рассматриваются как случайные величины — величины, значения которых подвержены случайности. **Модель** (*model*) — это набор предположений на совместное распределение зависимой и объясняющих переменных. Иначе говоря, модель представляет собой совокупность совместных распределений, удовлетворяющих некоторому набору предположений. Классическая линейная модель регрессии — это совокупность совместных распределений, удовлетворяющих предположениям 1.1–1.4, сформулированным ниже.

Предположение линейности

Первое предположение заключается в том, что связь между зависимой переменной и регрессором линейна.

Предположение 1.1 (линейность):

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \quad (i = 1, 2, \dots, n). \quad (1.1.1)$$

где $\beta_1, \dots, \beta_k$ — неизвестные параметры, подлежащие оценке; ε_i — наблюдаемая величина ошибки с определенными свойствами, указанными ниже.

Выражение $\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$ в правой части уравнения называется **регрессией** (*regression*), или **функцией регрессии** (*regression function*), а коэффициенты β_i — **коэффициентами регрессии** (*regression coefficients*). Они представляют собой предельные эффекты отдельных регрессоров. Например, β_1 представляет изменение зависимой

¹ Это переменная в правой части. — *Прим. науч. ред. перевод.*

переменной, когда второй регрессор увеличивается на единицу, в то время как остальные остаются неизменными. Математически это может быть записано как $\partial y_i / \partial x_{i2} = \beta_2$. Линейность означает, что предельные эффекты не зависят от значений регрессоров. Ошибка представляет собой часть зависимой переменной, которая остается не объясненной регрессором.

Пример 1.1 (функция потребления): Простая функция потребления, знакомая из вводного курса экономики, — это

$$CON_i = \beta_1 + \beta_2 YD_i + \varepsilon_i, \quad (1.1.2)$$

где CON — потребление, YD — располагаемый доход. Если данные представляют собой годовые макроэкономические временные ряды, то CON_i и YD_i — агрегированное потребление и доход в году i . Если же данные получены из обследования домашних хозяйств, то CON_i — это потребление i -го домашнего хозяйства в кросс-секционной выборке из n домашних хозяйств. Функцию потребления можно записать в виде (1.1.1), обозначив $y_i = CON_i$, $x_{i1} = 1$ (константа) и $x_{i2} = YD_i$. Ошибка ε_i представляет другие переменные помимо располагаемого дохода, которые влияют на потребление. Они включают как переменные — такие как финансовые активы, — которые могут быть наблюдаемы, но исследователь решил их не включать в качестве регрессоров, так и другие переменные — такие как «настроение» потребителей, — которые трудно измерить. Когда уравнение включает только один регрессор, отличный от константы, как в этом примере, это называется **простой моделью регрессии**¹ (*simple regression model*).

Предположение линейности не является ограничительным, как может показаться на первый взгляд, так как зависимая переменная и регрессоры могут быть следствием преобразований рассматриваемых переменных. Рассмотрим

Пример 1.2 (уравнение заработной платы): Упрощенная версия уравнения заработной платы, обычно оцениваемая в экономиче-

$$\log(WAGE_i) = \beta_1 + \beta_2 S_i + \beta_3 TENURE_i + \beta_4 EXP_i + \varepsilon_i, \quad (1.1.3)$$

где $WAGE$ — уровень заработной платы человека, S — количество лет обучения, $TENURE$ — количество лет, отработанных на текущем рабочем месте, EXP — количество лет трудового стажа (то есть количество лет, проработанных человеком на текущем и предыдущих рабочих местах). Уравнение заработной платы пишется в общий формат (1.1.1), если обозначить $y_i = \log(WAGE_i)$. Такую классификацию уравнения называют **полулוגарифмической** (*semi-log*), поскольку логарифмическому

¹ В русскоязычной литературе в этом случае говорят также о простой линейной регрессии.

преобразованию подвергается только зависимая переменная. Указанные уравнения выводятся из следующей нелинейной связи между уровнем заработной платы и регрессорами:

$$\ln WAGE_i = \exp(\beta_1) \exp(\beta_2 S_i) \times \\ \times \exp(\beta_3 T F E I' B E_i) \exp(\beta_4 E X P R_i) + \varepsilon_i. \quad (1.1.4)$$

Взяв логарифм от обеих частей (1.1.4) и учитывая, что $\log[\exp(x)] = x$, получаем (1.1.3). Коэффициенты в полулогарифмической форме имеют интерпретацию *процентного изменения*, а не *абсолютный уровень*. Например, значение 0,05 коэффициента β_2 означает, что дополнительный год обучения приводит к пятипроцентному росту уровня заработной платы. Разница в интерпретации получается из-за того, что зависимая переменная — это логарифм заработной платы, а не сама заработная плата, т. е. изменение в значении логарифма равно² процентному изменению уровня.

К виду (1.1.1) могут быть сведены и некоторые другие разновидности нелинейного влияния. Предположим, например, что предельный эффект образования увеличивается по мере роста уровня образования. Это может быть достигнуто путем подбора в уравнении заработной платы функции от регрессора — квадрата количества лет обучения (S^2). Если обозначить коэффициент перед словом регрессором как β_2 , то предельный эффект от образования будет рассчитываться по формуле:

$$\beta_2 (1 + 2S_i) = d \log(WAGE_i) / dS_i.$$

Если коэффициент β_2 отрицателен, предельный эффект от образования снижается при росте уровня образования.

Конечно, существуют нелинейные спецификации, не сводимые к линейным. Например, уравнение регрессии (1.1.4) было бы невозможно свести к линейному, если бы эффект отклика не мультипликативен, а аддитивен:

$$\ln WAGE_i = \exp(\beta_1) \exp(\beta_2 S_i) \exp(\beta_3 T F E I' B E_i) \exp(\beta_4 E X P R_i) + \varepsilon_i.$$

Методы оценивания нелинейных уравнений регрессии (таблицы 10 и 11) будут обсуждаться в главе 7.

Матричные обозначения

Прежде чем обсуждать в главе 2 предположения классической линейной регрессионной модели, напомним некоторые матричные обозначения. Эти обозначения будут полезны как для формулировки прочих предположений, так и для вывода OLS-оценки β . Обозначим K -мерные векторы

¹ К-размерными векторами x_i — K -мерному уровню заработной платы. Прим. автора ред. перевода.

² Прологарифмическое равно. Прим. автора ред. перевода.

столбцы β_1 и β как

$$\underset{(n \times 1)}{x_1} = \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{bmatrix}, \quad \underset{(K \times 1)}{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}. \quad (1.1.5)$$

По определению скалярного произведения векторов, $x_1'\beta = \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_K x_{1n}$. Поэтому уравнения в предположении 1.1 могут быть переписаны следующим образом:

$$y_i = x_1'\beta + \varepsilon_i \quad (i = 1, 2, \dots, n). \quad (1.1.6)$$

Также определим

$$\underset{(n \times 1)}{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \underset{(n \times 1)}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \underset{(n \times K)}{X} = \begin{bmatrix} x_{11} \\ \vdots \\ x_{n1} \end{bmatrix} = \begin{bmatrix} x_{11} & \dots & x_{1K} \\ \vdots & \dots & \vdots \\ x_{n1} & \dots & x_{nK} \end{bmatrix}. \quad (1.1.6)$$

В (1.1.5) количество строк в векторах и матрицах равно числу наблюдений. По этой причине y и X иногда называют **вектором данных** и **матрицей данных** (*data vector*, *data matrix*). Поскольку числа столбцов в X равняются числу строк в β , X и β являются согласованными, и $X\beta$ — вектор размера $n \times 1$. i -й элемент этого вектора равен $x_1'\beta$. Поэтому предположение 1.1 можно переписать более компактно:

$$\underset{(n \times 1)}{y} = \underbrace{\underset{(n \times K)}{X} \underset{(K \times 1)}{\beta}}_{(n \times 1)} + \underset{(n \times 1)}{\varepsilon}.$$

Предположение о строгой экзогенности

Сформулируем следующее предположение классической регрессионной модели

Предположение 1.2 (строгая экзогенность):

$$E(\varepsilon_i | X) = 0 \quad (i = 1, 2, \dots, n). \quad (1.1.7)$$

Здесь математическое ожидание (среднее) рассчитывается как условное относительно значений регрессоров для этих наблюдений. Это замечание можно сделать более наглядным, если переписать данное предположение без асимптотической единицы *mathematical*:

$$E(\varepsilon_i | x_1, \dots, x_K) = 0 \quad (i = 1, 2, \dots, n).$$

Это предположение можно сформулировать иначе. Для любого наблюдения i возьмем совместное распределение $nK + 1$ случайных величин, $f(\varepsilon_i, x_1, \dots, x_n)$ и рассмотрим условное распределение $f(\varepsilon_i | x_1, \dots, x_n)$. Условное математическое ожидание $E(\varepsilon_i | x_1, \dots, x_n)$ является, в общем случае, целой функцией от $(x_1, \dots, x_n)$. Предположение о строгой экзогенности говорит, что эта функция равна константе, а точнее нулю¹.

Предположение о том, что эта константа равна нулю, не является ограничительным, если регрессоры включают константу, поскольку тогда уравнение может быть переписано так, что условное математическое ожидание ошибки равно нулю. Чтобы продемонстрировать это, предположим, что $E(\varepsilon_i | X)$ равно μ и что $z_{i1} = 1$. Тогда уравнение можно записать в виде:

$$\begin{aligned} y_i &= \beta_1 + \beta_2 x_{i2} + \dots + \beta_K x_{iK} + \varepsilon_i = \\ &= (\beta_1 + \mu) + \beta_2 x_{i2} + \dots + \beta_K x_{iK} + (\varepsilon_i - \mu). \end{aligned}$$

Если переобозначить β_1 как $\beta_1 + \mu$ и ε_i как $\varepsilon_i - \mu$, то условное математическое ожидание новой ошибки равно нулю. Почти во всех приложениях регрессоры включают постоянную составляющую.

Пример 1.3 (продолжение примера 1.1): Для простой регрессионной модели из примера 1.1 предположение строгой экзогенности можно записать как

$$E(\varepsilon_i | YD_1, YD_2, \dots, YD_n) = 0.$$

Поскольку $x_i = (1, YD_i)'$, вы могли бы захотеть переписать предположение о строгой экзогенности в виде:

$$E(\varepsilon_i | 1, YD_1, 1, YD_2, \dots, 1, YD_n) = 0.$$

Однако, поскольку константа не несет в себе информации, условное математическое ожидание при условии

$$(1, YD_1, 1, YD_2, \dots, 1, YD_n)$$

сходится с математическим ожиданием при условии

$$(YD_1, YD_2, \dots, YD_n).$$

Следствия строгой экзогенности

Предположение о строгой экзогенности имеет ряд важных следствий.

- Безусловное математическое ожидание ошибки равно нулю, то есть

$$E(\varepsilon_i) = 0 \quad (i = 1, 2, \dots, n). \quad (1.1.8)$$

¹ Некоторые авторы определяют понятие «строгой экзогенности» немного иначе. Например, в [Koopmans and Plosser, 1953] и [Engle, Hendry and Richard, 1983], те регрессоры называются строго экзогенными, если ε_i независим от ε_j для всех i, j . Это определение является более сильным, чем определение строгой экзогенности, используемое здесь, но не противоречит ему.

Это так, поскольку, по закону полных математических ожиданий¹, $E[E(\varepsilon_i | \mathbf{X})] = E(\varepsilon_i)$.

- Если смешанный момент $E(xy)$ двух случайных величин x и y равен нулю, мы будем говорить, что x **ортогональна** y (или y ортогональна x). При выполнении строгой экзогенности регрессоры ортогональны ошибке для *всех* наблюдений, то есть

$$E(x_{jk}\varepsilon_i) = 0 \quad (i, j = 1, \dots, n; k = 1, \dots, K)$$

или

$$E(x_{j\cdot}\varepsilon_i) = \begin{bmatrix} E(x_{j1}\varepsilon_i) \\ E(x_{j2}\varepsilon_i) \\ \vdots \\ E(x_{jK}\varepsilon_i) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(K \times 1)} \quad (\text{для всех } i, j). \quad (1.1.9)$$

Доказательство этого утверждения, приведенное ниже, представляет собой прекрасный пример применения свойств условных математических ожиданий.

Доказательство

Поскольку x_{jk} является элементом $\mathbf{X}$, строгая экзогенность влечет

$$E(\varepsilon_i | x_{jk}) = E[E(\varepsilon_i | \mathbf{X}) | x_{jk}] = 0 \quad (1.1.10)$$

по формуле повторного математического ожидания². Из этого следует, что

$$\begin{aligned} E(x_{jk}\varepsilon_i) &= E[E(x_{jk}\varepsilon_i | x_{jk})] \\ &\quad (\text{по формуле повторного математического ожидания}) \\ &= E[x_{jk} E(\varepsilon_i | x_{jk})] \\ &\quad (\text{из линейности условных математических ожиданий}^3) \\ &= 0. \end{aligned}$$

Следует подчеркнуть, что строгая экзогенность требует ортогональности регрессоров не только ошибке в том же наблюдении (то есть $E(x_{jk}\varepsilon_i) = 0$ для всех k), но также ошибкам в других наблюдениях (то есть $E(x_{jk}\varepsilon_i) = 0$ для всех k и для $j \neq i$).

¹Закон полных математических ожиданий имеет вид: $E[E(y, \mathbf{x})] = E(y)$.

²Формула повторного математического ожидания имеет вид: $E[E(y | \mathbf{x}, z) | \mathbf{x}] = E(y | \mathbf{x})$.

³Линейность условных математических ожиданий утверждает, что $E[f(\mathbf{x})g(\mathbf{x})] = f(\mathbf{x}) E(g | \mathbf{x})$.

- Так как математическое ожидание ошибки равно нулю, условия ортогональности (1.1.9) эквивалентны условиям нулевой корреляции. Это верно, так как

$$\begin{aligned} \text{Cov}(x_i, \varepsilon_i) &= E(x_i \varepsilon_i) = E(x_i) E(\varepsilon_i) \\ &\quad (\text{по определению ковариации}) \\ &= E(x_i) E(\varepsilon_i) \quad (\text{так как } E(\varepsilon_i) = 0, \text{ см. (1.1.8)}) = \\ &= 0 \quad (\text{по условию ортогональности (1.1.9)}). \end{aligned}$$

В частности, для $i = j$ $\text{Cov}(x_i, \varepsilon_i) = 0$. Поэтому строгая экзогенность подразумевает (иногда тем, кто ранее изучал эконометрику) требование того, что регрессоры не коррелированы с ошибкой в том же наблюдении (*contemporaneously uncorrelated*).

Строгая экзогенность в моделях временных рядов

В моделях временных рядов, где i обозначает время, следствие строгой экзогенности (1.1.9) может быть переформулировано следующим образом: регрессоры ортогональны прошлым, текущим и будущим ошибкам (это эквивалентно ортогональности ошибок прошлым, текущим и будущим значениям регрессоров). Но для большинства моделей временных рядов это условие (и тем более строгая экзогенность) не выполняется. Нынешняя классическая теория конечных выборок, базирующаяся на понятии строгой экзогенности, датируя будущее изначальным и прошлым, редко можно применить к случаю временных рядов. Однако в следующей главе будет показано, что для больших выборок предложенная оценка обладает рядом хороших свойств и без предположения о строгой экзогенности.

Самый очевидный пример невыполнения свойства строгой экзогенности — это модель, в которой в качестве регрессора выступает **лаг** зависимой переменной. Вот простейшая из подобных моделей:

$$y_i = \beta_0 y_{i-1} + \varepsilon_i \quad (i = 1, 2, \dots, n). \quad (1.1.11)$$

Она называется **моделью авторегрессии первого порядка** (*first-order autoregressive model* — **AR(1)**). Мы рассмотрим эту модель более подробно в главе 6.1. Предположим, что (в соответствии с предположением о строгой экзогенности) регрессор для наблюдения i , y_{i-1} , ортогонален ошибке для того же наблюдения i , то есть $E(y_{i-1} \varepsilon_i) = 0$. Тогда

$$\begin{aligned} E(y_i \varepsilon_i) &= E(\beta_0 y_{i-1} + \varepsilon_i) \varepsilon_i = \quad (\text{из (1.1.11)}) \\ &= \beta_0 E(y_{i-1} \varepsilon_i) + E(\varepsilon_i^2) = \\ &= E(\varepsilon_i^2) \quad (\text{так как } E(y_{i-1} \varepsilon_i) = 0 \text{ по предположению}). \end{aligned}$$

Пытаться (если исключить случай, когда ошибки всегда равны нулю) $E(y_i, \varepsilon_i)$ не равно нулю. Однако x_i является регрессором для наблюдения $i + 1$. Таким образом, регрессор неортогонален прошлой ошибке, что нарушает строгую экзогенность.

Остальные предположения модели

Ниже приведены остальные предположения классической регрессионной модели.

Предположение 1.3 (отсутствие мультиколлинеарности): Ранг матрицы данных X размера $n \times K$ равен K с вероятностью 1.

Предположение 1.4 (сферичность ковариационной матрицы ошибок):

(гомоскедастичность) $E(\varepsilon_i^2 | X) = \sigma^2 > 0 \quad (i = 1, 2, \dots, n)$. (1.1.12)
(отсутствие корреляции между наблюдениями)

$$E(\varepsilon_i \varepsilon_j | X) = 0 \quad (i, j = 1, 2, \dots, n; i \neq j). \quad (1.1.13)$$

Чтобы понять предположение 1.3, необходимо вспомнить из линейной алгебры, что ранг матрицы равен числу линейно независимых столбцов этой матрицы. Это предположение гласит, что ни один из K столбцов матрицы данных X не может быть выражен как линейная комбинация остальных столбцов матрицы X . Это означает, что матрица X имеет полный столбцовый ранг (*full column rank*). Так как K столбцов не могут быть линейно независимыми, если их размерность меньше K , из данного предположения вытекает, что $n \geq K$, то есть наблюдений должно быть не меньше, чем столько же, сколько регрессоров. Говорят, что регрессоры (строго) мультиколлинеарны (*strictly multicollinear*), если это предположение не выполняется. В некоторых случаях несложно увидеть, что регрессоры мультиколлинеарны в каких-то случаях возникают вследствие этого.

Пример 1.4 (продолжение примера 1.2): Если в выборке ни один из индивидов никогда не менял место работы, то $TENURE_i = EXP_i$ для всех i , что противоречит нашему предположению об отсутствии мультиколлинеарности. Очевидно, что в этом случае невозможно разделить влияние на заработную плату количества лет, проработанных

¹ Если вычислить, что момент распределения ε_i данным случаем второй момент, $E(\varepsilon_i^2 | X)$ равен какому-либо значению (здесь σ^2), то это означает, что этот момент существует и конечен. Далее мы будем говорить об этом сокращенно.

на текущем рабочем месте, и влияние общего трудового стажа. Если подставить это разложение в уравнение заработной платы, чтобы исключить переменную $TENURE$, уравнение заработной платы примет вид:

$$\log(WAGE_i) = \beta_1 + \beta_2 S_i + (\beta_3 + \beta_4) T \cdot XIPR_i + \varepsilon_i,$$

из чего следует, что можно оценить только сумму коэффициентов $\beta_3 + \beta_4$, но не коэффициенты β_3 и β_4 по отдельности.

Предположение гомоскедастичности (1.1.12) означает, что условный момент второго порядка (который в общем случае является непарной функцией от X) — постоянная величина. Иногда требование строгой экзогенности, это предположение можно переформулировать в более привычных терминах. Рассмотрим условную дисперсию $\text{Var}(\varepsilon_i | X)$. Она равна этой же постоянной, так как

$$\begin{aligned} \text{Var}(\varepsilon_i | X) &= E(\varepsilon_i^2 | X) - E(\varepsilon_i | X)^2 \\ &= E(\varepsilon_i^2 | X) \quad (\text{по определению условной дисперсии}) \\ &= E(\varepsilon_i^2 | X) \quad (\text{так как } E(\varepsilon_i | X) = 0 \text{ вследствие строгой экзогенности}). \end{aligned}$$

Аналогично (1.1.13) эквивалентно следующему требованию:

$$\text{Cov}(\varepsilon_i, \varepsilon_j | X) = 0 \quad (i, j = 1, 2, \dots, n; i \neq j).$$

То есть в совместном распределении $(\varepsilon_i, \varepsilon_j)$ при условии X ковариация равна нулю. Приравнительно к моделям временных рядов (1.1.13) эквивалентно требованию отсутствия **серийной корреляции** у ошибок.

Так как элемент (i, j) матрицы $\varepsilon \varepsilon'$ размера $n \times n$ равен $\varepsilon_i \varepsilon_j$, предположение 1.4 можно переписать в более компактной форме

$$E(\varepsilon \varepsilon' | X) = \sigma^2 I_n. \quad (1.1.14)$$

Дискуссия в предыдущем параграфе демонстрирует, что это предположение можно представить и в следующем виде:

$$\text{Var}(\varepsilon | X) = \sigma^2 I_n.$$

Однако представление (1.1.14) предпочтительнее, поскольку более удобной мерой разброса являются вторые моменты (такие как $E(\varepsilon_i^2 | X)$), а не дисперсии. Смысл этих замечаний станет понятен после изучения следующей главы, в которой речь пойдет о теории бинарных выборов. Предположение 1.4 иногда называют предположением **сферичности** ковариационной матрицы, поскольку матрица вторых моментов размера

и x_i (элементами которых являются дисперсии и ковариации) пропорциональна единичной матрице I_n . В дальнейшем это предположение будет ослаблено.

Классическая регрессионная модель для случайных выборок

Выборка (y, X) называется **случайной**, если последовательность $\{y_i, x_i\}$ является i.i.d. (независимой и одинаково распределенной) по наблюдениям. Так как по предположению 1.1 ε_i — это функция от (y_i, x_i) и поскольку $\{y_i, x_i\}$ независимы от (y_j, x_j) для $j \neq i$, (ε_i, x_i) независимы от x_j для $j \neq i$. Поэтому

$$\begin{aligned} E(\varepsilon_i | X) &= E(\varepsilon_i | x_i), \\ E(\varepsilon_i^2 | X) &= E(\varepsilon_i^2 | x_i), \end{aligned} \quad (1.1.15)$$

$$\text{и } E(\varepsilon_i \varepsilon_j | X) = E(\varepsilon_i | x_i) E(\varepsilon_j | x_j) \quad (\text{для } i \neq j).$$

(Доказательство последнего равенства в (1.1.15) вынесено в контрольные вопросы.) Следовательно, предположения 1.2 и 1.4 сводятся к таким:

$$\text{предположение 1.2: } E(\varepsilon_i | x_i) = 0 \quad (i = 1, 2, \dots, n), \quad (1.1.16)$$

$$\text{предположение 1.4: } E(\varepsilon_i^2 | x_i) = \sigma^2 > 0 \quad (i = 1, 2, \dots, n) \quad (1.1.17)$$

Следствием одинакового распределения случайной выборки является то, что совместное распределение (ε_i, x_i) одинаково для всех i . То есть безусловный второй момент $E(\varepsilon_i^2)$ не зависит от i (это называют **безусловной гомоскедастичностью** (unconditional homoskedasticity)) и функциональная форма условного второго момента $E(\varepsilon_i^2 | x_i)$ одинакова для всех i . Однако из предположения 1.4 не следует, что значения условного второго момента одинаковы для всех i . Поэтому предположение 1.4 остается ограничительным в ситуации случайной выборки: без этого предположение условный второй момент $E(\varepsilon_i^2 | x_i)$ мог бы быть различным для разных i вследствие его потенциальной зависимости от x_i . Чтобы подчеркнуть разницу, мы будем называть ограничения на условные моменты второго порядка, (1.1.12) и (1.1.17), **условной гомоскедастичностью**.

«Фиксированные» регрессоры

Классическая линейная регрессионная модель была изложена в предположении, что регрессоры являются стохастическими. Такое изложение противоречит большинству случаев, в которых регрессоры предопределяются детерминированными, или «фиксированными» (*deterministic*, «fixed»). Если регрессоры X фиксированы, то нет необходимости делать

различия между условным, $f(\varepsilon_i | x_1, \dots, x_k)$, и безусловным, $f(\varepsilon_i)$, распределением ошибки. Тогда предположения 1.2 и 1.4 можно переписать в виде:

$$\text{предположение 1.2} \quad E(\varepsilon_i) = 0 \quad (i = 1, \dots, n), \quad (1.1.18)$$

$$\begin{aligned} \text{предположение 1.4} \quad E(\varepsilon_i^2) &= \sigma^2 \quad (i = 1, \dots, n); \\ \Gamma(\varepsilon_i \varepsilon_j) &= 0 \quad (i, j = 1, \dots, n; i \neq j). \end{aligned} \quad (1.1.19)$$

Очевидно, что такое переформулирование является формалистическим в неэкспериментальной науке (в частности, в эконометрике). Однако предположения о детерминированности регрессоров остаются популярными, так как регрессионную модель с фиксированными X можно интерпретировать как набор утверждений, условных относительно X , что позволяет исключить « X » из рассуждений, в частности из предположений 1.2 и 1.4.

Отсюда за экономичность в обозначениях необходимо платить свою цену. При таких обозначениях очень легко упустить из виду предположение, что ошибка не коррелирована с текущими, прошлыми и будущими значениями регрессоров. Кроме того, в случае детерминированных регрессоров исчезает разница между условной и безусловной homoskedастичностью. Поэтому в этой книге регрессоры предполагаются стохастическими, и, если не указано иное, утверждения, условные относительно X , явно записываются с помощью вставки « X » (если не сказано иного).

Контрольные вопросы

1. (Смена единиц измерения в полуспараметрической модели.) Как изменится условное значение заработной платы (1.1.3) в примере 1.2, если WAGE измерять не в долларах, а в центях? Указание: $\log(xy) = \log(x) + \log(y)$.
2. Докажите, что левая часть в (1.1.5):
Указание: $E(\varepsilon_i | x_1, \dots, x_k) = E(\varepsilon_i | x_1, \dots, x_k | X)$, (ε_i, X) независимы от переменных $\{\varepsilon_j, x_1, \dots, x_k, -1 \leq x_1, \dots, x_k\}$ для $i \neq j$.
3. (Одновременное выполнение линейности и строгой экзогенности.) Покажите, что из предположений 1.1 и 1.2 следует, что

$$E(y_i | X) = x_i'\beta \quad (i = 1, 2, \dots, n). \quad (1.1.20)$$

Докажите в обратную сторону: из выполнения условия (1.1.20) следует, что существуют ошибки, которые удовлетворяют предположениям 1.1 и 1.2.

4. (Нормально распределенная случайная выборка.) Рассмотрим случайную выборку из потребителев с доходом $(C/N, Y/D)_i$ ($i = 1, 2, \dots, n$). Предположим, что совместное распределение $(C/N, Y/D)_i$ (оно одинаково для всех i в соответствии с предположением о случайности выборки) нормально. Очевидно, что предположение 1.3 является истинным, если X может быть меньше K только из-за случайности. Покажите, что прочие предположения (1.1, 1.2 и 1.4)

также выполнены. **Указание:** Если две случайные величины y и x , совместно нормально распределены, тогда условное математическое ожидание линейно по x , то есть

$$E(y|x) = \beta_1 + \beta_2 x,$$

и условная дисперсия, $\text{Var}(y|x)$, не зависит от x . Ключевым является тот факт, что это распределение одинаково для всех i , так как в противном случае β_1 и β_2 могут отличаться для разных i .

5. (Мультиколлинеарность в модели парной регрессии.) Покажите, что предположение 1.3 в случае парной регрессии равносильно утверждению, что регрессор, отличный от постоянной, (x_{j2}) не является постоянной (то есть x_{i2} / x_{j2} для некоторых пар $\{i, j\}$, $i \neq j$ с вероятностью, равной единице).
6. (Упражнение на условные и безусловные математические ожидания.) Покажите, что из предположений 1.1 и 1.4 следует, что

$$\text{Var}(e_i) = \sigma^2 \quad (i = 1, \dots, n)$$

$$\text{и } \text{Cov}(e_i, e_j) = 0 \quad (i \neq j; i, j = 1, \dots, n). \quad (*)$$

Указание: Из строгой экзогенности вытекает $E(e_i) = 0$. Поэтому (*) эквивалентно

$$E(e_i^2) = \sigma^2 \quad (i = 1, \dots, n)$$

$$\text{и } E(e_i e_j) = 0 \quad (i \neq j; i, j = 1, \dots, n).$$

1.2. Алгебра метода наименьших квадратов

Этот параграф посвящен вычислительной процедуре получения OLS-оценки $\hat{b}$ неизвестного вектора коэффициентов β . Также вводятся некоторые понятия, связанные с вектором b .

OLS минимизирует сумму квадратов остатков

Хотя мы не наблюдаем ошибку, мы можем при некотором заданном гипотетическом значении $\tilde{\beta}$ вектора коэффициентов β вычислить значение

$$y_i - x_i' \tilde{\beta}.$$

Это называется **остатком** (*residual*) для наблюдения i . Тогда **сумма квадратов остатков** (*sum of squared residuals — SSR*) представляет собой

$$SSR(\tilde{\beta}) \equiv \sum_{i=1}^n (y_i - x_i' \tilde{\beta})^2 = (y - X\tilde{\beta})'(y - X\tilde{\beta}).$$

Эту сумму называют также **остаточной суммой квадратов** (*residual sum of squares — RSS*), или **суммой квадратов ошибок** (*error sum of squares — ESS*). Сумма квадратов остатков является функцией от $\tilde{\beta}$, так

как от него зависит остаток. **OLS-оценка**, $\hat{b}$, вектора параметров β — это такой вектор $\tilde{\beta}$, который минимизирует эту функцию

$$\hat{b} = \underset{\tilde{\beta}}{\operatorname{argmin}} SSR(\tilde{\beta}). \quad (1.2.1)$$

Взаимосвязь между неизвестным вектором коэффициентов β , его OLS-оценкой $\hat{b}$ и гипотетическим значением $\tilde{\beta}(\tilde{\beta})$ продемонстрирована на рис. 1.1 для $K = 1$. Так как $SSR(\tilde{\beta})$ — квадратичная функция от $\tilde{\beta}$, ее график имеет U-образную форму. Значение $\tilde{\beta}$, соответствующее минимуму, — это OLS-оценка $\hat{b}$. Поскольку она зависит от выборки (y, X) , OLS-оценка $\hat{b}$ в общем случае отличается от «истинного» значения β . Равенство может достигаться только по чистой случайности.

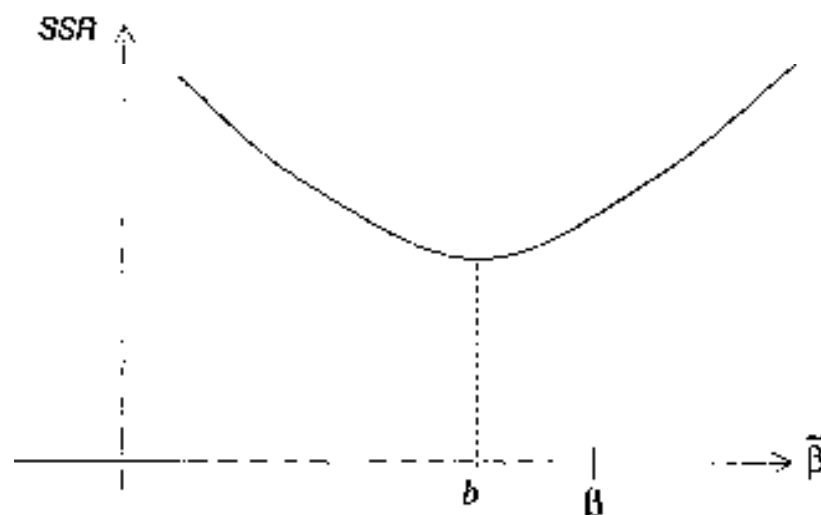


Рис. 1.1. Гипотетическое, истинное и оцененное значения

Поскольку в целевой функции используются квадраты остатков, этот метод приписывает больший «штраф» большим значениям остатков. OLS-оценка используется, чтобы предотвратить большие значения остатков для некоторых наблюдений, допуская при этом небольшие остатки для большого числа наблюдений. В следующем параграфе мы увидим, что применение именно такого критерия приводит к некоторым хорошим свойствам оценки.

Нормальные уравнения

Классический способ решения задачи минимизации состоит в получении условий первого порядка приравниванием частных производных нулю. Для этого необходимо найти K -мерный вектор частных производных

$\partial SSR(\tilde{\beta})/\partial \tilde{\beta}^1$. Задача упрощается, если записать $SSR(\tilde{\beta})$ иначе:

$$SSR(\tilde{\beta}) = (y - X\tilde{\beta})'(y - X\tilde{\beta})$$

(так как i -й элемент $y - X\tilde{\beta}$ — это $y_i - x_i'\tilde{\beta}$)

$= (y' - \tilde{\beta}'X')y - X\tilde{\beta}$ (поскольку $(X\tilde{\beta})' = \tilde{\beta}'X') =$

$$= y'y - \tilde{\beta}'X'y - y'X\tilde{\beta} - \tilde{\beta}'X'X\tilde{\beta} =$$

$$= y'y - 2y'X\tilde{\beta} - \tilde{\beta}'X'X\tilde{\beta}$$

($\tilde{\beta}'X'y = y'X\tilde{\beta}$, так как значение скаляра

при транспонировании не изменится)

$$= y'y - 2\alpha'\tilde{\beta} + \tilde{\beta}'A\tilde{\beta}, \quad \text{где } \alpha \equiv X'y \text{ и } A \equiv X'X. \quad (1.2.2)$$

Первое слагаемое $y'y$ не зависит от $\tilde{\beta}$, поэтому им можно пренебречь при дифференцировании $SSR(\tilde{\beta})$. Напомним из линейной алгебры, что

$$\frac{\partial(\alpha'\tilde{\beta})}{\partial \tilde{\beta}} = \alpha \quad \text{и} \quad \frac{\partial(\tilde{\beta}'A\tilde{\beta})}{\partial \tilde{\beta}} = 2A\tilde{\beta} \quad \text{для симметричной матрицы } A,$$

так что K -мерный вектор частных производных равен

$$\frac{\partial SSR(\tilde{\beta})}{\partial \tilde{\beta}} = 2\alpha + 2A\tilde{\beta}.$$

Условия первого порядка получаются приравнованием этого выражения нулю. Применяя (1.2.2) к нашей задаче, заменяя вектор α на $X'y$, A — на $X'X$ и переупорядочивая, получаем условия первого порядка:

$$\underbrace{X'X}_{(K \times K)(K \times K)} b = \underbrace{X'y}_{(K \times 1)(1 \times 1)}. \quad (1.2.3)$$

Мы захотели $\tilde{\beta}$ на b , поскольку OLS-оценка b — это значение $\tilde{\beta}$, удовлетворяющее условиям первого порядка. Эти K уравнений называются **нормальными уравнениями**.

Вектор остатков, вычисленный в точке $\tilde{\beta} = b$,

$$e = y - Xb. \quad (1.2.4)$$

¹ Если $b \in \mathbb{R}^K$ — скалярная функция от K -мерного вектора x , то b и x представляют собой K -мерный вектор, k -й элемент которого равен $\partial b(x)/\partial x_k$, где x_k — k -й элемент x (Этот K -мерный вектор называется **градиентом**). В нашем случае $x = \tilde{\beta}$, а функция $b(x) = SSR(\tilde{\beta})$.

называется вектором **OLS остатков** (*OLS residuals*). Если i -й элемент $e_i \equiv y_i - x_i'b$. Перегруппировывая (1.2.3), получаем:

$$X'(y - Xb) = 0, \text{ или } X'e = 0, \text{ или } \frac{1}{n} \sum_{i=1}^n x_i \cdot e_i = 0, \text{ или } \frac{1}{n} \sum_{i=1}^n x_i \cdot (y_i - x_i'b) = 0. \quad (1.2.3')$$

это уравнение показывает, что нормальные уравнения могут быть при интерпретированы как выборочные аналоги условий ортогональности $E(x_i \cdot e_i) = 0$. Эта идея будет более подробно обсуждаться в последующих главах.

Однако условия первого порядка являются лишь необходимыми условиями экстремума. Следовательно, чтобы убедиться, что в b достигается минимум, а не максимум, нам необходимо проверить условия второго порядка. Читатели, которые знакомы с теорией гессовых¹ функции нескольких переменных², могут сразу понять, что условия второго порядка выполнены, поскольку (как отмечено выше) матрица $X'X$ положительно определена. Однако то, что в b достигается минимум, возможно доказать и непосредственно. Для этого необходимо вычесть и прибавить к левой функции одну и ту же величину (такая стратегия эффективна в случае квадратичной целевой функции, как здесь). Применение этой стратегии к алгебре метода наименьших квадратов оставлено читателям в качестве аналитического упражнения.

Два выражения для OLS-оценки

Итак, мы получили систему из K линейных одновременных уравнений с K неизвестными, составляющими вектор b . По предположению 1.3 (об отсутствии мультиколлинеарности), матрица коэффициентов $X'X$ положительно определена (по поводу доказательства этого утверждения см. контрольный вопрос 1 текущего параграфа) и, следовательно, невырождена. То есть система нормальных уравнений имеет единственное решение b , которое можно найти, умножив обе части (1.2.3) слева на $(X'X)^{-1}$:

$$b = (X'X)^{-1} X'y. \quad (1.2.5)$$

Рассматривая (1.2.5) как функцию от выборки (y, X) , это выражение иногда называют **OLS-оценщиком** (*OLS estimator*). Для любой конкретной выборки (y, X) значение этой функции называется **OLS-оцен-**

¹ В оригинале *Hessian*. Этот термин автор употребляет, имея в виду именно матрицу вторых частных производных, которую в русскоязычной литературе часто именуют матрицей Гессе (франц. *matr. Hess.*, русск. перевод).

² Гессовы функции $b(x)$ представляют собой квадратичные матрицы, в которой элемент (k, l) равен $\partial^2 b(x) / \partial x_k \partial x_l$.

кой (*OLS-estimate*). В этой книге (как и в большинстве других) эти термины будут использоваться почти как синонимы¹.

Поскольку $(X'X)^{-1}X'y = (X'X/n)^{-1}X'y/n$, OLS-оценка может быть также записана как

$$b = S_{xx}^{-1}s_{xy}, \quad \text{где} \quad (1.2.5')$$

$$S_{xx} = \frac{1}{n}X'X = \frac{1}{n}\sum_{i=1}^n x_i x_i' \quad (\text{выборочное среднее } x_i x_i'), \quad (1.2.6a)$$

$$s_{xy} = \frac{1}{n}X'y = \frac{1}{n}\sum_{i=1}^n x_i \cdot y_i \quad (\text{выборочное среднее } x_i \cdot y_i). \quad (1.2.6b)$$

Интерпретовать матрицу данных из (1.2.5) более удобно для вывода свойств оценки в малых выборках, а использование выборочных средних из предположений (1.2.5') — для теории больших выборок.

Еще несколько терминов из алгебры

После того как мы вывели OLS-оценку, следует определить еще несколько терминов.

- **Подобранным значением** (*fitted value*) объясняющей переменной для наблюдения i называется $\hat{y}_i = x_i' b$. Вектор подобранных значений $\hat{y}$ равен Xb . Таким образом, вектор OLS-остатков можно записать как $e = y - \hat{y}$.
- Определим **проекционную матрицу** (*projection matrix*) P и **аннулятор** (*annihilator*) M :

$$P = X(X'X)^{-1}X'. \quad (1.2.7)$$

$$M = I_n - P. \quad (1.2.8)$$

Они обладают рядом привычных свойств (доказательство которых оставлено читателям в качестве одного из контрольных вопросов).

$$\text{Матрицы } P \text{ и } M \text{ симметричны и идемпотентны}^2. \quad (1.2.9)$$

$$PX = X \quad (\text{отсюда термин «проекционная матрица»}). \quad (1.2.10)$$

$$MX = 0 \quad (\text{отсюда термин «аннулятор»}). \quad (1.2.11)$$

¹Поскольку автор указывает на то, что упомянутые два термина используются в тексте почти как синонимы, мы употребляем при переводе термин «оценка» как для *OLS-estimate*, так и для *value*, что соответствует традиции, принятой в русскоязычной литературе. — *Прим. пер. ред. перевода.*

²Квадратная матрица A называется **идемпотентной**, если $A = A^2$.

Поскольку e — вектор остатков при $\tilde{\beta} = b$, то сумма квадратов остатков OLS (SSR) равна $e'e$. Ее можно также переписать как:

$$SSR = e'e = e'Me. \quad (1.2.12)$$

(Доказательство этого поставлено в качестве контрольного вопроса.) Это выражение, связывающее SSR и истинные ошибки ε , будет использовано далее.

- OLS-оценка σ^2 (дисперсии ошибки), которую обозначим s^2 , равна сумме квадратов остатков, деленной на $n - K$:

$$s^2 \equiv \frac{SSR}{n - K} = \frac{e'e}{n - K}. \quad (1.2.13)$$

(В этом определении предполагается, что $n > K$. В противном случае s^2 не определена.) Как будет показано в утверждении 1.2, приведенном ниже, при делении суммы квадратов остатков на $n - K$ (число степеней свободы), а не на n (размер выборки), оценка σ^2 оказывается несмещенной. Интуитивное объяснение состоит в том, что сначала необходимо оценить K параметров (β) и лишь потом получить вектор остатков e , необходимый для вычисления s^2 . С формальной точки зрения e должен удовлетворять системе K нормальных уравнений (1.2.9), которая налагает ограничения на возможные значения остатков.

- Квадратный корень из s^2 (s) называется **стандартной ошибкой регрессии** (*standard error of regression — SER*), или **стандартной ошибкой уравнения** (*standard error of equation — SEE*). SER представляет собой оценку стандартного отклонения ошибок.
- **Ошибка оценки** (*sampling error*)¹ определяется как $b - \beta$. Ее также можно связать с ε :

$$\begin{aligned} b - \beta &= (X'X)^{-1}X'y - \beta = \quad (\text{из (1.2.5)}) \\ &= (X'X)^{-1}X'(X\beta + \varepsilon) - \beta \\ &(\text{так как } y = X\beta + \varepsilon \text{ по предположению 1.1}) \\ &= (X'X)^{-1}(X'X)\beta + (X'X)^{-1}X'\varepsilon - \beta = \\ &= \beta - (X'X)^{-1}X'\varepsilon + \beta \approx (X'X)^{-1}X'\varepsilon. \end{aligned} \quad (1.2.14)$$

¹В оригинале *sampling error* (бухгалтерская ошибка выборки). Мы используем при переводе термин «ошибка оценки», поскольку в англоязычной литературе термин *sampling error* и сопоставляемый ему в русскоязычной литературе термин «ошибка выборки» обычно означают нечто иное (например, систематическую ошибку выборки). — Прим. науч. ред. переводы.

- **Нецентрированный R^2 .** Одной из мер изменчивости зависимой переменной является сумма квадратов $\sum u_i^2 = \mathbf{y}'\mathbf{y}$. Поскольку OLS остатки удовлетворяют системе нормальных уравнений, мы получаем следующую декомпозицию $\mathbf{y}'\mathbf{y}$:

$$\begin{aligned}\mathbf{y}'\mathbf{y} &= (\hat{\mathbf{y}} - \mathbf{e})'(\hat{\mathbf{y}} - \mathbf{e}) \quad (\text{так как } \mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}) = \\ &= \hat{\mathbf{y}}'\hat{\mathbf{y}} - 2\hat{\mathbf{y}}'\mathbf{e} + \mathbf{e}'\mathbf{e} \\ &= \hat{\mathbf{y}}'\hat{\mathbf{y}} + 2\mathbf{b}'\mathbf{X}'\mathbf{e} + \mathbf{e}'\mathbf{e} \quad (\text{поскольку } \hat{\mathbf{y}} \equiv \mathbf{X}\mathbf{b}) \\ &= \hat{\mathbf{y}}'\hat{\mathbf{y}} + \mathbf{e}'\mathbf{e} \\ & \quad (\text{так как } \mathbf{X}'\mathbf{e} = \mathbf{0} \text{ из системы нормальных} \\ & \quad \text{уравнений; см. (1.2.3)}).\end{aligned}\tag{1.2.15}$$

Нецентрированный R^2 (uncentered R^2) определяется как

$$R_{uc}^2 \equiv 1 - \frac{\mathbf{e}'\mathbf{e}}{\mathbf{y}'\mathbf{y}}.\tag{1.2.16}$$

Применяя декомпозицию (1.2.15), получаем, что R_{uc}^2 равен:

$$\frac{\hat{\mathbf{y}}'\hat{\mathbf{y}}}{\mathbf{y}'\mathbf{y}}.$$

Поскольку $\hat{\mathbf{y}}'\hat{\mathbf{y}}$ и $\mathbf{e}'\mathbf{e}$ неотрицательны, $0 \leq R_{uc}^2 \leq 1$. Таким образом, нецентрированный R^2 может быть интерпретирован как доля изменчивости зависимой переменной, объясненная независимыми переменными. Чем ближе подобранные значения к значениям истинной зависимой, тем ближе нецентрированный R^2 к единице.

- **(Центрированный) R^2 , коэффициент детерминации.** Если единичный регрессор в модели — константа (так что $K = 1$ и $\mathbf{x}_{i1} = 1$), то тогда из (1.2.5) вытекают, что $b_1 = \bar{y}$ ($\bar{y}$ — выборочное среднее зависимой переменной). Из этого следует, что $\hat{y}_i = \bar{y}$ для всех i . Из (1.2.15) $\hat{\mathbf{y}}'\hat{\mathbf{y}} = n\bar{y}^2$, а $\mathbf{e}'\mathbf{e} = \sum_{i=1}^n (y_i - \bar{y})^2$. Если же в модели, кроме константы, присутствуют и другие регрессоры, то можно доказать (доказательство оставлено в качестве аналитического упражнения), что $\sum_{i=1}^n (y_i - \bar{y})^2$ может быть разложена как

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n e_i^2, \quad \text{где } \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.\tag{1.2.17}$$

Коэффициент детерминации R^2 определяется следующим образом:

$$R^2 \equiv 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}.\tag{1.2.18}$$

В силу декомпозиции (1.2.17), R^2 равен:

$$\frac{\sum_{j=1}^n (\bar{y}_j - \bar{y})^2}{\sum_{j=1}^n (y_j - \bar{y})^2}.$$

Поэтому, если регрессоры включают константу (так что разложение (1.2.17) выполнено), $0 \leq R^2 \leq 1$. Таким образом, R^2 , определенный в (1.2.18), является мерой, объясняющей способности регрессоров, не идящих констант.

Если среди регрессоров нет константы, то при вычислении R^2 по формуле (1.2.18) (как делают некоторые программные пакеты) R^2 может получиться отрицательным. Это происходит потому, что регрессия без константы может аппроксимировать объясняемую переменную хуже, чем выборочное среднее. С другой стороны, некоторые эконометрические пакеты (например, STATA) при исключении константы из модели автоматически используют формулу (1.2.16) для расчета коэффициента детерминации во избежание получения отрицательных значений R^2 . Это папка с двух концов. Предположим, что среди регрессоров нет константы, однако линейная комбинация регрессоров равна константе. Такое может произойти, например, когда константу в модели заменяют набором сезонных дамми-переменных¹. Такая регрессия по сути не изменится, если один из регрессоров в линейной комбинации заменить на константу. Более того, не изменится вектор подобранных значений. Однако, если R^2 рассчитывается по формуле (1.2.16) для регрессии без константы или по формуле (1.2.18) для регрессии с константой, R^2 будет меньше при добавлении константы (см. контрольный вопрос 7 ниже).

Анализ влиятельных наблюдений (дополнительно)

Поскольку метод наименьших квадратов стремится предотвратить появление нескольких больших остатков, допуская большое количество сравнительно малых остатков, только несколько наблюдений могут быть чрезвычайно влиятельными, в том смысле, что при исключении таких наблюдений из выборки некоторые элементы вектора параметров $\mathbf{b}$ могут значительно измениться. Существует систематический подход к поиску таких наблюдений, называемых **влиятельными наблюдениями** (*influential observations*)². Обозначим через $\hat{\mathbf{b}}^{(i)}$ OLS-оценку вектора коэффициентов β , полученную с использованием выборки, из которой было исключено i -е наблюдение. Ключевым для дальнейшего анализа является

¹Понятие дамми-переменной будет введено в эмпирических приложениях в этой главе.

²Для более подробного ознакомления с темой см. [Kassam, Kohn, and Welsh, 1983].

следующее уравнение:

$$y^{(i)} - \hat{y} = \left(\begin{array}{c} 1 \\ \vdots \\ 1 - \mu_i \end{array} \right) (X'X)^{-1} x_i' e_i, \quad (1.2.19)$$

где x_i' , как и прежде, обозначает i -ю строку матрицы данных X , e_i — OLS остаток для наблюдения i , а μ_i вычисляется по формуле:

$$\mu_i = x_i' (X'X)^{-1} x_i. \quad (1.2.20)$$

то есть μ_i является i -м диагональным элементом проекционной матрицы P . (Доказательство (1.2.19) было бы хорошим упражнением в матричной алгебре, но мы не будем приводить его здесь.) Нетрудно показать (см. контрольный вопрос 7 к параграфу 1.3), что

$$0 \leq \mu_i \leq 1 \quad \text{и} \quad \sum_{i=1}^n \mu_i = K. \quad (1.2.21)$$

Поэтому μ_i в среднем равно K/n .

Для иллюстрации при применении (1.2.19) на конкретном примере рассмотрим взаимосвязь между y и x инвестициями в основной капитал и экономическим ростом в беднейших странах мира в период с 1960 по 1985 г. На рис. 1.2 изображена диаграмма рассеяния для 13 стран, ВВП которых на одного работающего в 1985 году составлял менее 10% аналогичного показателя в США¹. По горизонтальной оси откладывается среднегодовое отношение инвестиций в основной капитал к ВВП за период с 1960 по 1985 г. По вертикальной — среднегодовой рост ВВП на работающего за тот же период. На основе визуального анализа рисунка нетрудно заметить, что положение оцененной линии регрессии в значительной мере будет зависеть от единственного выделяющегося наблюдения (Ботсваны). Действительно, при исключении Ботсваны из выборки, оцененный коэффициент наклона падает с 0,37 до 0,058. В случае парной регрессии (как здесь) выделяющиеся наблюдения легко обнаружить с помощью визуального анализа диаграммы рассеяния, подобной рис. 1.2. Однако эта стратегия не может быть использована в случае, когда имеется более одного регрессора, отличного от постоянной. Анализ, основанный на формуле (1.2.19), не ограничен парной регрессией. В табл. 1.1 для каждого наблюдения приведены OLS остатки, значения μ_i и значения (1.2.19). Значение μ_i для Ботсваны, равное 0,7196, намного превосходит среднее, равное 0,154 (= K/n = 2/13). То есть это наблюдение является весьма **влиятельным**, о чем свидетельствуют два последних столбца таблицы. Обратите внимание, что мы бы не смогли обнаружить это влиятельное

¹ Данные были взяты из Penn World Table, переименованной в [DeLong and Summers, 1991]. Однако в их анализе использовалась вся выборка, включающая 61 страну.

наблюдения, рассматривая остатки. Это неудивительно, поскольку метод наименьших квадратов алгебраически устроен так, чтобы не допускать больших значений остатков (за счет большого количества малых остатков для остальных наблюдений).

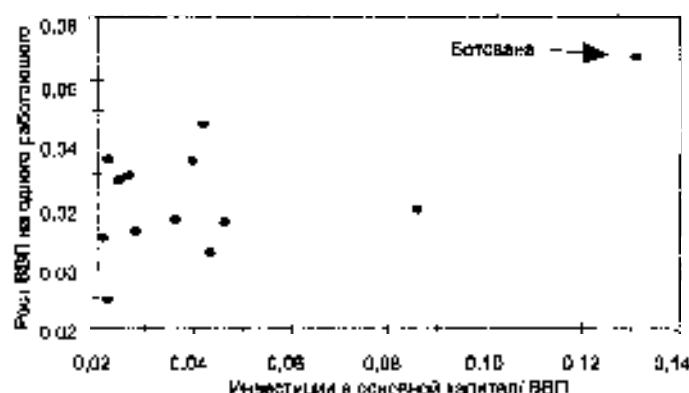


Рис. 1.2. Инвестиции в основной капитал и рост

Таблица 1.1

Анализ влияния

Страна	Рост ВВП на одного работающего	Инвестиции в основной капитал/ВВП	Остаток	r_i	(1.2.19) для β_1	(1.2.19) для β_2
Ботсвана	0,0676	0,1310	0,0119	0,7196	0,0104	-0,3129
Камерун	0,0458	0,0413	0,0233	0,0773	-0,0021	0,0045
Эфиопия	0,0054	0,0912	0,0056	0,1493	0,0010	-0,0119
Индия	0,0115	0,0276	-0,0008	0,0980	0,0009	-0,0027
Индонезия	0,0345	0,0221	0,0192	0,1160	0,0034	0,0094
Кот-д'Ивуар	0,0076	0,0243	0,0117	0,1084	-0,0019	0,0213
Кения	0,0136	0,0452	0,0096	0,0775	0,0007	0,0023
Малайзия	0,0102	0,0219	-0,0254	0,1107	0,0045	-0,0027
Мальта	0,0133	0,0381	0,0032	0,0817	0,0006	0,0036
Мали	0,0044	0,0433	-0,0183	0,0760	0,0016	-0,0006
Пакистан	0,0297	0,0283	0,0126	0,1422	0,0020	0,0205
Танзания	0,0184	0,0656	-0,0209	0,2361	-0,0021	0,0052
Танзани	0,0341	0,0395	0,0123	0,0784	0,0012	0,0057

Что делать с выделяющимися наблюдениями? Смотря по обстоятельствам. Если выделяющиеся наблюдения удовлетворяют регрессионной модели, они содержат в себе ценную информацию о выборке, которая не содержится в других наблюдениях. В таком случае их можно исключить. Однако более вероятным кажется то, что выделяющиеся наблюдения — это наблюдения, чуждые для выборки, из-за того, что они

не удовлетворяют модели. Тогда они однозначно должны быть исключены из выборки. В рассматриваемом примере наблюдался рост мирового спроса на алмазы (главную статью экспорта Ботсваны), а для их добычи необходима дорогостоящая буровая техника. Если мы предполагаем, что за экономическую моделью стоит следующая логическая цепочка: инвестиции означают внедрение новых технологий, которые, в свою очередь, ведут к росту производительности труда и, следовательно, к экономическому росту, — тогда Ботсвана, высокие темпы роста ВВП которой были обусловлены повышенным спросом, должна быть исключена из выборки.

Замечание относительно вычисления OLS-оценок¹

До сих пор основное внимание уделялось концептуальным аспектам выборки OLS. Но прикладным исследователям, которые на практике вычисляют OLS-оценки с использованием компьютеров, важно знать некоторые особенности компьютерных вычислений, чтобы избежать возможности получения недостоверных результатов. Потенциальным источником проблем может быть то, что компьютеры аппроксимируют натуральные числа так называемыми **числами с плавающей запятой**. Когда в арифметической операции участвуют и очень большие, и очень маленькие числа, использование чисел с плавающей запятой может привести к неточным результатам. При вычислении OLS-оценки такая ситуация может возникнуть, когда регрессоры значительно отличаются по величине. Например, одним из регрессоров может быть процентная ставка, выраженная дробью, а другим — ВВП США в долларах. Матрица $X'X$ будет тогда содержать как очень маленькие, так и очень большие числа, и операция вычисления обратной матрицы может дать недостоверный результат.

Эту проблему можно решить путем выбора единиц измерения таким образом, чтобы регрессоры были близки по величине. Например, указывать ставку процента в процентах, а ВВП США — в триллионах долларов. Такая мера предосторожности в большинстве случаев помогает избежать проблем. Более систематический подход к преобразованию матрицы X состоит в вычислении выборочных средних из всех регрессоров и делении на выборочные стандартные ошибки. Стандартизация переменных проводится для формирования $X'X$. Полученные OLS-оценки надо скорректировать, переводя их в масштаб исходных данных. Большинство компьютерных пакетов, вычисляющих OLS-оценки (например, TSP), используют более сложное преобразование матрицы X (которое называется **QR-декомпозицией**).

¹ Более подробно эта тема изложена в Section 1.5 [Davidson and MacKinnon, 1993].

Контрольные вопросы

1. Докажите, что матрица $X'X$ положительно определена, если матрица X имеет полный столбцовый ранг. Указание: Необходимо показать, что $e'X'Xe > 0$ для $e \neq 0$. Определим $z \equiv Xe$. Тогда $e'X'Xe = z'z = \sum_{i=1}^n z_i^2$. Если X — матрица полного столбцового ранга, то $z \neq 0$ для всех $e \neq 0$.
2. Проверьте, что $X'X/n = \frac{1}{n} \sum_{i=1}^n x_i x_i'$ и $X'y/n = \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i$ как в (1.2.6). Указание: Элемент (k, l) матрицы $X'X$ равен $\sum_{i=1}^n x_{ik} x_{il}$.
3. (OLS-оценка для простой регрессии одной переменной.) В простой регрессионной модели $K = 2$ и $x_{i1} = 1$. Покажите, что

$$S_{XX} = \begin{bmatrix} 1 & \bar{x}_2 \\ \bar{x}_2 & \frac{1}{n} \sum_{i=1}^n x_{i2}^2 \end{bmatrix}, \quad S_{XY} = \begin{bmatrix} \bar{y} \\ \frac{1}{n} \sum_{i=1}^n x_{i2} y_i \end{bmatrix},$$

где

$$\bar{y} \equiv \frac{1}{n} \sum_{i=1}^n y_i \quad \text{и} \quad \bar{x}_2 \equiv \frac{1}{n} \sum_{i=1}^n x_{i2}.$$

Покажите, что

$$b_0 = \frac{\frac{1}{n} \sum_{i=1}^n (x_{i2} - \bar{x}_2)(y_i - \bar{y})}{\frac{1}{n} \sum_{i=1}^n (x_{i2} - \bar{x}_2)^2} \quad \text{и} \quad b_1 = \bar{y} - \bar{x}_2 b_0.$$

(Можно заметить, что в числителе выражения для b_2 стоит выборочная дисперсия регрессора, считанного от константы. А в знаменателе находится выборочная ковариация зависимой переменной и регрессора, отличного от константы.) Указание:

$$\frac{1}{n} \sum_{i=1}^n x_{i2}^2 - (\bar{x}_2)^2 = \frac{1}{n} \sum_{i=1}^n (x_{i2} - \bar{x}_2)^2$$

и

$$\frac{1}{n} \sum_{i=1}^n x_{i2} y_i - \bar{x}_2 \bar{y} = \frac{1}{n} \sum_{i=1}^n (x_{i2} - \bar{x}_2)(y_i - \bar{y}).$$

Вы можете взять уравнение (1.2.5) и просто произвести обращение матрицы. Однако есть и другой путь: выписать систему из двух нормальных уравнений. Первое уравнение системы будет иметь следующий вид: $b_1 = \bar{y} - \bar{x}_2 b_0$. Подставьте это выражение для b_1 во второе уравнение и найдите b_0 .

4. Докажите (1.2.9)–(1.2.11). Указание: Эти свойства вытекают непосредственно из определений P и M .
5. (Матричная алгебра для подобранных значений и остатков.) Покажите следующее:

(a) $\hat{y} = Py$, $e = My = My - Mx$. Указание: Используйте (1.2.5).

(b) (1.2.12), а именно $SSR = e'Mx$.

6. (Изменение единиц измерения и R^2 .) Влияет ли изменение единиц измерения зависимой переменной на R^2 ? А независимые единицы измерения регрессоров? **Указание:** Проверьте, влияют ли вышеуказанные изменения на числитель и знаменатель в определении R^2 .
7. (Взаимосвязь между R^2_{OLS} и R^2 .) Покажите, что

$$1 - R^2 = \left(1 - \frac{n \cdot \bar{y}^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right) (1 - R^2_{OLS}).$$

Указание: Используйте (1.2.16), (1.2.18) и тождество $\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n \cdot \bar{y}^2$.

8. Покажите, что

$$R^2_{OLS} = \frac{y'Py}{y'y}.$$

9. (Вычисляемые статистики.) Проверьте, что b , SSR , s^2 и R^2 можно вычислить, используя следующие выборочные средние: $\bar{y}$, $\bar{x}_1x_1$, $\bar{x}_1x_2$, $\bar{y}^2/y/n$ и $\bar{y}$. (Если константа исключена в модель, то $\bar{y}$ является аргументом $\bar{x}_1x_1$, соответствующим константе.) Поэтому для получения коэффициентов регрессии и связанных с ними статистик эти выборочные средние достаточно вычислять только один раз.

1.3. Свойства OLS в конечных выборках

Мы уже вывели OLS-оценку, теперь нам предстоит исследовать ее свойства в конечных выборках, а именно свойства распределения оценки, которые имеют место для любого объема выборки n .

Распределение b в конечных выборках

Утверждение 1.1 (свойства OLS-оценки β в конечных выборках):

- (a) (Несмещенность.) При выполнении предположений 1.1–1.3,
 $E(b|X) = \beta$.
- (b) (Выражение для дисперсий¹.) При выполнении предположений 1.1–1.4,
 $\text{Var}(b|X) = \sigma^2 \cdot (X'X)^{-1}$.
- (c) (Теорема Гаусса — Маркова.) При выполнении предположений 1.1–1.4,
 OLS-оценка является «эффективной» в классе линейных несмещен-

¹Здесь и далее у автора «variance» — «дисперсия», хотя в действительности имеется в виду ковариационная матрица случайного вектора (которую также называют дисперсионной или дисперсионно-ковариационной). — *Прим. науч. ред. перевод*

ных оценок. То есть для любой несмещенной оценки $\tilde{\beta}$, линейной по y , $\text{Var}(\tilde{\beta}|X) \geq \text{Var}(b|X)$ в матричном смысле¹

- (d) При выполнении предположений 1.1–1.4, $\text{Cov}(b, e|X) = 0^2$, где $e \equiv y - Xb$.

Прежде чем перейти к доказательству, остановимся подробнее на смысле этого утверждения.

- Матричное неравенство в пункте (с) говорит, что матрица $\text{Var}(\tilde{\beta}|X) - \text{Var}(b|X)$ размерности $K \times K$ положительна полуопределена. Значит,

$$\alpha' \text{Var}(\tilde{\beta}|X) - \text{Var}(b|X) \alpha \geq 0$$

или

$$\alpha' \text{Var}(\tilde{\beta}|X) \alpha \geq \alpha' \text{Var}(b|X) \alpha$$

для любого K -мерного вектора α . В частности, и для вектора, состоящего из нулей, кроме k -го элемента, равного единице. Для такого вектора α квадратичная форма $\alpha' A \alpha$ выбирает элемент (k, k) из матрицы A . Элемент (k, k) матрицы $\text{Var}(\tilde{\beta}|X)$, например, равен $\text{Var}(\tilde{\beta}_k|X)$, где $\tilde{\beta}_k$ — k -й элемент вектора $\tilde{\beta}$. То есть из матричного неравенства (с) следует, что

$$\text{Var}(\tilde{\beta}_k|X) \geq \text{Var}(b_k|X) \quad (k = 1, 2, \dots, K). \quad (1.3.1)$$

То есть для любого коэффициента регрессии дисперсия его OLS-оценки не превышает дисперсии любой альтернативной линейной несмещенной оценки.

- Из выражения (1.2.5) легко видеть, что OLS-оценка линейна по y . Вообще существует много других линейных и несмещенных оценок $\tilde{\beta}$ (в качестве одного из контрольных вопросов вам будет предложено построить одну из таких оценок). Теорема Гаусса–Маркова утверждает, что OLS-оценка является **эффективной** в том смысле, что ее условная дисперсия $\text{Var}(b|X)$ является наименьшей среди дисперсий всех линейных несмещенных оценок. По этой причине OLS-оценку называют **наилучшей линейной несмещенной оценкой** (*Best Linear Unbiased Estimator* — BLUE).

¹ Пусть A и B — две квадратные матрицы одинакового размера. Говорят, что $A \geq B$, если разность $A - B$ положительно полуопределена. Матрица C размерности $K \times K$ называется положительно полуопределенной (или неотрицательно определенной), если $\alpha' C \alpha \geq 0$ для любого K -мерного вектора α .

²Здесь и далее $\text{Cov}(e, e)$ — матрица взаимных ковариаций случайных векторов e и e — *Вект. матр. ред. переписки*.

- OLS-оценка b является функцией от выборки (y, X) . Поскольку выборка (y, X) случайна, это же можно сказать и о b . Теперь предположим, что мы зафиксировали конкретную матрицу данных X , вычислили b для всех выборок, соответствующих всем возможным реализациям y , а затем взяли среднее значение b (вам дисперсия сделать это в выражении на метод Монте Карло в этой главе). Это среднее представляет собой (теоретическое) условное математическое ожидание $E(b|X)$. Пункт (а) (наша цель) говорит о том, что это математическое ожидание равно истинному значению β .
- Существует еще одно определение несмещенности, которое является более слабым, чем указанное в пункте (а). По формуле полного математического ожидания, $E[E(b|X)] = E(b)$. То есть из (а) следует

$$E(b) = \beta. \quad (1.3.2)$$

Это уравнение утверждает, что если мы вычислим b для всех возможных выборок, различающихся не только зависимой переменной y , но и X , то его среднее будет равно истинному значению. Это безусловное утверждение, скорее всего, является более применимым к экономике, так как на практике выборки отличаются как y , так и X . То есть важность условного утверждения (а) состоит в том, что из него следует безусловный аналог (1.3.2), который, в свою очередь, ближе к практике.

- Те же рассуждения применимы к утверждению (с) относительно дисперсии. В контрольном вопросе, приведенном далее, вам будет предложено показать, что из утверждений (а) и (б) следует, что

$$\text{Var}(\hat{\beta}) \geq \text{Var}(b), \quad (1.3.3)$$

где $\hat{\beta}$ — любая линейная несмещенная оценка (то есть $E(\hat{\beta}|X) = \beta$).

Теперь настало время перейти к доказательству этого важного результата. Может показаться, что оно слишком длинное, однако эта видимость создается лишь из-за его подробности. На самом деле это доказательство весьма простое. Тем не менее при первом чтении вам, возможно, захочется пропустить доказательство пункта (с). Доказательство части (d) оставлено читателям в качестве контрольных вопросов.

Доказательство

- (а) (Доказательство того, что $E(b|X) = \beta$) $E(b - \beta|X) = 0$ тогда и только тогда, когда $E(b|X) = \beta$. Итак, мы доказали первое из вышеупомянутых утверждений. Из выражения для ошибки оценки (1.2.14), $b - \beta = A\varepsilon$, где $A = (X'X)^{-1}X'$. То есть

$$E(b - \beta|X) = E(A\varepsilon|X) = A E(\varepsilon|X).$$

Второе равенство следует из линейности условного математического ожидания: A — функция от X и, следовательно, может рассматриваться как детерминированная матрица. Так как $E(\varepsilon|X) = 0$, последнее выражение равно нулю.

(3) (Доказательство того, что $\text{Var}(\hat{\beta}|X) = \sigma^2 \cdot (X'X)^{-1}$)

$$\begin{aligned}\text{Var}(\hat{\beta}|X) &= \text{Var}(\hat{\beta} - \beta|X) \quad (\text{так как } \beta \text{ детерминирован}) \\ &= \text{Var}(A\varepsilon|X) \quad (\text{из (1.2.14) и } A \equiv (X'X)^{-1}X') = \\ &= A \text{Var}(\varepsilon|X)A' \quad (\text{поскольку } A \text{ — функция от } X) \\ &= A E(\varepsilon\varepsilon'|X)A' \quad (\text{по предположению 1.2}) = \\ &= A(\sigma^2 I_n)A' \quad (\text{по предположению 1.4, см. (1.1.14)}) = \\ &= \sigma^2 AA' = \\ &= \sigma^2 \cdot (X'X)^{-1} \\ &(\text{так как } AA' = (X'X)^{-1}X'X(X'X)^{-1} = (X'X)^{-1}).\end{aligned}$$

4е) (Теорема Гаусса — Марков.) Поскольку $\hat{\beta}$ линейна по y , его можно записать как $\hat{\beta} = Cy$ для некоторой матрицы C , которая, помимо того, является функцией от X . Пусть $D \equiv C - A$ или $C = D + A$, где $A \equiv (X'X)^{-1}X'$. Тогда

$$\begin{aligned}\hat{\beta} &= (D + A)y = \\ &= Dy + Ay = \\ &= D(X\beta + \varepsilon) + b \\ &(\text{так как } y = X\beta + \varepsilon \text{ и } Ay = (X'X)^{-1}X'y = b) \\ &= DX\beta + D\varepsilon + b.\end{aligned}$$

После взятия условного математического ожидания от обеих частей получаем:

$$E(\hat{\beta}|X) = DX\beta + E(D\varepsilon|X) + E(b|X)$$

Поскольку обе скаляры, b и $\hat{\beta}$, несмещенные и поскольку $E(D\varepsilon|X) = DE(\varepsilon|X) = 0$, откуда следует, что $DX\beta = 0$. Чтобы это было справедливым для любого вектора β , необходимо, чтобы $DX = 0$. Поэтому $\hat{\beta} = D\varepsilon + b$ и

$$\begin{aligned}\hat{\beta} - \beta &= D\varepsilon + (b - \beta) = \\ &= (D - A)\varepsilon \quad (\text{из (1.2.14)}).\end{aligned}$$

Следовательно,

$$\begin{aligned}\text{Var}(\hat{\beta}|X) &= \text{Var}(\hat{\beta} - \beta|X) = \\ &= \text{Var}[(D + A)e|X] = \\ &= (D + A) \text{Var}(e|X)(D' + A') \\ &\quad (\text{так как } D \text{ и } A \text{ — функции от } X) \\ &= \sigma^2 \cdot (D + A)(D' + A') \quad (\text{поскольку } \text{Var}(e|X) = \sigma^2 I_n) = \\ &= \sigma^2 \cdot (DU' + AD' + DA' + AA').\end{aligned}$$

Однако $DA' = DX(X'X)^{-1} = 0$, поскольку $DX = 0$. Кроме того, $AA' = (X'X)^{-1}$, как было показано в (b). Значит,

$$\begin{aligned}\text{Var}(\hat{\beta}|X) &= \sigma^2 \cdot [DD' + (X'X)^{-1}] \geq \\ &\geq \sigma^2 \cdot (X'X)^{-1} \\ &\quad (\text{так как матрица } DD' \text{ положительно полуопределена}) \\ &= \text{Var}(b|X) \quad (\text{из (b)}).\end{aligned}$$

Следует подчеркнуть, что предположение о строгой экзогенности (предположение 1.2) является ключевым для доказательства свойства несмещенности. Без строгой экзогенности нельзя было бы заключить этого. Например, недостаточно предположить, что $E(\varepsilon_i|x_i) = 0$ для всех i или что $E(\varepsilon_i \cdot x_j) = 0$ для всех i, j . Как уже отмечалось в параграфе 1.1, большинство моделей временных рядов не удовлетворяют условию строгой экзогенности, даже если выполнены более слабые требования (таким как условие ортогональности $E(\varepsilon_i \cdot x_j) = 0$). Следовательно, для подобных моделей OLS-оценка не обладает свойством несмещенности.

Свойства s^2 в конечных выборках

Мы уже определили OLS-оценку s^2 в (1.2.13). Она также является несмещенной.

Утверждение 1.2 (несмещенность s^2): При выполнении предположений 1.2–1.4, $E(s^2|X) = \sigma^2$ (и, следовательно, $E(s^2) = \sigma^2$), при $n > K$ (то есть когда s^2 определена).

Эти утверждения можно легко доказать, используя свойства матрицы следа¹.

Доказательство Поскольку $s^2 = e'e/(n - K)$, то доказательство сводится к тому, чтобы показать, что $E(e'e|X) = (n - K)\sigma^2$. Как было

¹Следом (или квадратной матрицы A) называют сумму диагональных элементов A (или $\text{tr}(A) = \sum_i a_{ii}$).

показано в контрольном вопросе 5 (а) задания 1.2, $e'e = e'Me$, где M — аннулятор. Таким образом, необходимо доказать два свойства:
1) $E(e'Me|X) = \sigma^2 \cdot \text{trac}(M)$ и 2) $\text{trac}(M) = n - K$.

(1) Доказательство того, что $E(e'Me|X) = \sigma^2 \cdot \text{trac}(M)$. Поскольку $e'Me = \sum_{i=1}^n \sum_{j=1}^K m_{ij} \varepsilon_i \varepsilon_j$ (это видно, если расписать квадратичную форму $e'Me$), получаем:

$$\begin{aligned} E(e'Me|X) &= \sum_{i=1}^n \sum_{j=1}^K m_{ij} E(\varepsilon_i \varepsilon_j | X) \quad (\text{так как } m_{ij} = \text{функция от } X, \\ &\quad E(m_{ij} \varepsilon_i \varepsilon_j | X) = m_{ij} E(\varepsilon_i \varepsilon_j | X)) \\ &= \sum_{i=1}^n m_{ii} \sigma^2 \\ &\quad (\text{поскольку } E(\varepsilon_i \varepsilon_j | X) = 0 \text{ для } i \neq j \text{ по предположению 1.4}) \\ &= \sigma^2 \sum_{i=1}^n m_{ii} = \\ &= \sigma^2 \cdot \text{trac}(M). \end{aligned}$$

(2) Доказательство того, что $\text{trac}(M) = n - K$:

$$\begin{aligned} \text{trac}(M) &= \text{trac}(I_n - P); \\ &\quad (\text{так как } M \equiv I_n - P, \text{ см. (1.2.8)}) \\ &= \text{trac}(I_n) - \text{trac}(P) \\ &\quad (\text{факт: оператор следа является линейным}) \\ &= n - \text{trac}(P); \\ \text{trac}(P) &= \text{trac}[X(X'X)^{-1}X'] \\ &\quad (\text{так как } P \equiv X(X'X)^{-1}X', \text{ см. (1.2.7)}) \\ &= \text{trac}[(X'X)^{-1}X'X] \quad (\text{факт: } \text{trac}(AB) = \text{trac}(BA)) \\ &= \text{trac}(I_K) = K. \end{aligned}$$

Итак, $\text{trac}(M) = n - K$. ■

Оценка $\text{Var}(b|X)$

Если $\hat{\sigma}^2$ — оценка σ^2 , то естественной оценкой $\text{Var}(b|X) = \sigma^2 \cdot (X'X)^{-1}$ является

$$\widehat{\text{Var}}(b|X) \equiv \hat{\sigma}^2 (X'X)^{-1}. \quad (1.3.4)$$

Это одна из стрелок, которые входят в комплектную выдачу после OLS-оценки в любом контрольном пакете.

Контрольные вопросы

1. (Роль предположения об отсутствии мультиколлинеарности.) Где мы использовали в утверждениях 1.1 и 1.2 предположение 1.3 о том, что $\text{rank}(X) = K$? **Указание:** Условие отсутствия мультиколлинеарности необходимо нам, чтобы обеспечить обратимость матрицы $X'X$.
2. (Пример линейной оценки.) Для примера 1.1 с функцией потребления предложите линейную и несмещенную оценку β_2 , отличную от OLS-оценки. **Указание:** Как насчет $\hat{\beta}_2 = (CON_2 - CON_1)/(YD_2 - YD_1)$? Является ли она линейной по $(CON_1, \dots, CON_n)$? Является ли она несмещенной? $E(\hat{\beta}_2 | YD_1, \dots, YD_n) = \beta_2$?
3. (Чего не утверждает теорема Гаусса — Маркова.) Существует ли при предположениях 1.2-1.4 линейная, но необязательно несмещенная оценка β , дисперсия которой меньше, чем у OLS-оценки? Если да, то насколько мал может быть дисперсия? **Указание:** Если оценка $\hat{\beta}$ — константа, то она тривиальным образом линейна по y .

4. (Теорема Гаусса — Маркова и безусловная дисперсия.)

(а) Докажите: $\text{Var}(\hat{\beta}) = E[\text{Var}(\hat{\beta} | X)] + \text{Var}[E(\hat{\beta} | X)]$.

Указание: По определению,

$$\text{Var}(\hat{\beta} | X) \equiv E[(\hat{\beta} - E(\hat{\beta} | X))(\hat{\beta} - E(\hat{\beta} | X))' | X]$$

и

$$\text{Var}[E(\hat{\beta} | X)] = E\left\{[E(\hat{\beta} | X) - E(\hat{\beta})][E(\hat{\beta} | X) - E(\hat{\beta})]'\right\}.$$

Используя свойства приближения линейными функциями, пишите $\hat{\beta} = E(\hat{\beta} | X)$ и приближайте и вычитайте $E(\hat{\beta})$.

- (б) Докажите (1.3.3). **Указание:** Если $\text{Var}(\hat{\beta} | X) > \text{Var}(\hat{\beta} | X)$, то

$$E[\text{Var}(\hat{\beta} | X)] > E[\text{Var}(\hat{\beta} | X)].$$

5. Предложите точечную оценку α^2 , считая, что α вам известен. **Указание:** Как насчет α^2/α ? Является ли она несмещенной?

6. Докажите часть (б) утверждения 1.1. **Указание:** По определению,

$$\text{Cov}(b, c | X) \equiv E[(b - E(b | X))(c - E(c | X))' | X].$$

Так как $E(b | X) = \beta$, получаем $b - E(b | X) = A\epsilon$, где $A = (X'X)^{-1}X'$. Искомое $M\epsilon = e$ (см. контрольный вопрос 5 параграфа 1.2), чтобы показать, что $e = E(e | X) = M\epsilon$, $L(A\epsilon\epsilon' A' | X) = A L(\epsilon\epsilon' | X) A'$, поскольку A и M являются функциями от X . Наконец, используйте $MX = 0$ (см. (1.2.11)).

7. Докажите (1.2.2). **Указание:** Поскольку матрица P симметрична по построению, ее диагональные элементы неотрицательны. Заметьте, что $\sum_{i=1}^n p_{ii} = \text{tr}(P)$.

1.4. Тестирование гипотез в предположении нормальности

Очень часто экономическая теория, мотивирующая спецификацию регрессионного уравнения, также накладывает ограничения на коэффициенты модели. Предположим, что экономическая теория накладывает ограничение, что коэффициент β_2 равен 1. Хотя предположение 1.1 гарантирует, что в среднем b_2 (OLS-оценка β_2) равна 1, если ограничение верно, b_2 может не быть в точности равной единице на конкретной выборке, которой вы располагаете. Очевидно, мы не можем заключить, что ограничение не выполнено, только на том основании, что оценка b_2 отличается от 1. Чтобы решить, является ли ошибка оценки $b_2 - 1$ «слишком большой», чтобы принять истинность ограничения, необходимо на основании ошибки оценки построить тестовую статистику, распределение которой было бы известным при условии истинности нулевой гипотезы. Может показаться, что для этого необходимо определить совместное распределение (X, e) , поскольку (как следует из (1.2–14)) ошибка оценки — это функция от (X, e) . Однако в этом параграфе будет изложен замечательный факт состоящий в том, что при нормальном условном распределении X распределение e не нужно специфицировать распределение X , чтобы найти распределение тестовой статистики.

На языке проверки гипотез тестируемое ограничение (такое, например, как « $\beta_2 = 1$ ») называется **нулевой гипотезой** (*null hypothesis*, *null*). Нулевая гипотеза является ограничением на **поддерживаемую гипотезу** (*maintained hypothesis*) — набор предположений, которые совместно с нулевой гипотезой позволяют получить тестовую статистику с известным распределением. Для рассматриваемого случая тестирования коэффициентов регрессии, для формирования такой поддерживаемой гипотезы, к классической линейной регрессионной модели (предположения 1.2–1.4) необходимо добавить только предположение о нормальности условного распределения e . (Как только что отмечалось, специфицировать совместное распределение (X, e) нет необходимости.) Иногда поддерживаемую гипотезу достаточно просто называют «моделью». Говорят, что модель **правильно специфицирована**, если поддерживаемая гипотеза верна. Хотя слишком большое значение тестовой статистики интерпретируется как невыполнение нулевой гипотезы, такие выводы возможно сделать лишь в случае правильной спецификации модели. Распределение тестовой статистики может отличаться от предполагаемого, если нулевая гипотеза верна, но модель неправильно специфицирована.

Нормально распределенные ошибки

Во многих приложениях ошибка включает большое количество разнообразных факторов, не охваченных регрессорами. Центральная предельная теорема наводит на мысль о том, что такая ошибка имеет нормаль-

но распределена. В других приложениях ошибка возникает вследствие ошибок измерения зависимой переменной. Важно, что ошибки в измерениях часто распределены нормально (из-за чего даже нормальное распределение изначально выводится для ошибок измерения). Поэтому заслуживает внимания следующее предположение о нормальности.

Предположение 1.5 (нормальность ошибки): *Условное распределение ε при условии X является многомерным нормальным распределением.*

Вспомогательным средством проверки, что нормальное распределение объясняет рядов данных, является:

- Это распределение зависит только от математического ожидания и дисперсии. Таким образом, при известных математическом ожидании и дисперсии можно написать функцию плотности. Если распределение условного отклика y при X нормальное, то его математическое ожидание и дисперсия могут зависеть от X . Следовательно, если распределение условного отклика y при X нормально, то математическое ожидание и дисперсия не зависят от X , то есть маргинальное (то есть безусловное) распределение является тем же самым нормальным распределением.
- В общем случае из независимости двух случайных величин не следует их некоррелированность, то есть наоборот. Однако если две случайные величины совместно нормально распределены, то из их некоррелированности следует их независимость. Иными словами, для совместно нормально распределенных случайных величин некоррелированность и независимость эквивалентны. Это верно и для условного распределения: если две случайные величины совместно нормальны и некоррелированы при условии X , то они независимы при условии X .
- Лinéйная функция от случайных величин с совместным нормальным распределением также распределена нормально. То же верно и для условных распределений. Если условное распределение y при условии X нормально, то Ay (где элементы матрицы A — функции от X) имеет условное нормальное распределение при условии X .

Именно благодаря этим особенностям нормального распределения предположение 1.5 обеспечивает следующие свойства, которые используются для доказательства теорем статистики.

- Математическое ожидание и дисперсия распределения y условного относительно X , уже были указаны в предположениях 1.2 и 1.4. Добавление к предположениям 1.2 и 1.4 предположения 1.5

приводит к тому, что распределение ε , условное относительно X , является нормальным распределением $N(0, \sigma^2 I_n)$.

$$\varepsilon|X \sim N(0, \sigma^2 I_n). \quad (1.4.1)$$

Таким образом, распределение ε , условное относительно X , не зависит от X . Из этого следует, что ε и X независимы. Поэтому, в частности, маргинальное, или безусловное, распределение ε нормально $N(0, \sigma^2 I_n)$.

- Из (1.2.14) нам известно, что ошибка оценки $\hat{\beta} - \beta$ линейна по ε при условии X . Так как ε распределен нормально при условии X , то это же справедливо для ошибки оценки. Его математическое ожидание и дисперсия заданы частями (а) и (б) утверждения 1.1. Таким образом, при предположениях 1.1–1.5

$$(\hat{\beta} - \beta)|X \sim N(0, \sigma^2 (X'X)^{-1}). \quad (1.4.2)$$

Тестирование гипотез об отдельных коэффициентах регрессии

Сначала рассмотрим гипотезу о k -м коэффициенте:

$$H_0: \beta_k = \bar{\beta}_k.$$

Здесь $\bar{\beta}_k$ — некоторое известное значение, определяемое нулевой гипотезой. Мы хотим протестировать эту гипотезу против альтернативы $H_1: \beta_k \neq \bar{\beta}_k$ на уровне значимости α . Берем k -ю компоненту (1.4.2), накладываем ограничения, заданные нулевой гипотезой, и получаем:

$$(\beta_k - \bar{\beta}_k) | X \sim N(0, \sigma^2 \cdot ((X'X)^{-1})_{kk}),$$

где $((X'X)^{-1})_{kk}$ — элемент (k, k) матрицы $(X'X)^{-1}$. Если мы определим статистику t_k как результат от деления разности $\hat{\beta}_k - \bar{\beta}_k$ на ее стандартное отклонение,

$$t_k = \frac{\hat{\beta}_k - \bar{\beta}_k}{\sqrt{\sigma^2 \cdot ((X'X)^{-1})_{kk}}}, \quad (1.4.3)$$

то t_k будет иметь стандартное нормальное распределение $N(0, 1)$.

Времененно предположим, что σ^2 известно. Тогда статистика t_k обладает рядом желательных свойств как тестовая статистика. Во-первых, ее значение может быть вычислено по выборке. Во-вторых, ее распределение, условное относительно X , не зависит от X (не путать с фактом, что значение t_k зависит от X). Третье, t_k и X независимо распределены,

и вне зависимости от значения $\mathbf{X}$ распределение z_k совпадает с безусловным распределением. Это удобно, так как разные выборки отличаются между собой не только значениями y , но и значениями $\mathbf{X}$. В-третьих, распределение указанной статистики известно. В частности, оно не зависит от неизвестных параметров (таких как β). (Если распределение статистики зависит от неизвестных параметров, эти параметры называются **мешающими** (*nuisance parameters*).) Используя эту статистику, мы можем определить, является ли ошибка оценки слишком большой: она является слишком большой, если тестовая статистика принимает значение, которое нетипично для реализации из указанного распределения.

Если мы не знаем значение σ^2 , то естественно заменить мешающий параметр σ^2 на его OLS-оценку s^2 . Получаемая подстановкой s^2 вместо σ^2 -статистика называется **t-отношением**, или **t-значением** (*t-ratio*, *t-value*). Знаменатель этой статистики называется **стандартной ошибкой** (*standard error*) OLS-оценки коэффициента β_k и иногда обозначается $SE(b_k)$:

$$SE(b_k) \equiv \sqrt{s^2 \cdot ((\mathbf{X}'\mathbf{X})^{-1})_{kk}} = \sqrt{(k, k)\text{-й элемент } \widehat{\text{Var}}(\mathbf{b}|\mathbf{X}) \text{ в (1.3.4)}}. \quad (1.4.4)$$

Поскольку s^2 — случайная величина (так как она является функцией от выборки), эта подстановка изменяет распределение статистики, но, к счастью, новое распределение также известно и не зависит ни от мешающих параметров, ни от $\mathbf{X}$.

Утверждение 1.3 (распределение t-статистики):

Пусть предположения 1.1–1.5 выполнены. Тогда при нулевой гипотезе $H_0: \beta_k = \bar{\beta}_k$, t-статистика, определенная как

$$t_k \equiv \frac{b_k - \bar{\beta}_k}{SE(b_k)} \equiv \frac{b_k - \bar{\beta}_k}{\sqrt{s^2 \cdot ((\mathbf{X}'\mathbf{X})^{-1})_{kk}}}. \quad (1.4.5)$$

распределена как $t(n-K)$ (t-распределение Стьюдента с $n-K$ степенями свободы).

Доказательство

Чтобы отразить замену s^2 на σ^2 , мы можем переписать t-статистику в виде:

$$\begin{aligned} t_k &= \frac{b_k - \bar{\beta}_k}{\sqrt{s^2 \cdot ((\mathbf{X}'\mathbf{X})^{-1})_{kk}}} \cdot \sqrt{\frac{\sigma^2}{s^2}} = \frac{z_k}{\sqrt{s^2/\sigma^2}} = \\ &= \frac{z_k}{\sqrt{\frac{e'e/(n-K)}{\sigma^2}}} = \frac{z_k}{\sqrt{\frac{q}{n-K}}}, \end{aligned}$$

где $q \equiv e'e/\sigma^2$. Мы уже показали, что z_k имеет распределение $N(0, 1)$. Теперь мы покажем, что

$$(1) \quad q|X \sim \chi^2(n - K),$$

(2) две случайные величины z_k и q независимы условно относительно X .

Тогда, по определению t -распределения, отношение z_k к $\sqrt{q/(n - K)}$ распределено как t -распределение с $n - K$ степенями свободы¹, и на этом доказательство будет завершено.

(1) Поскольку $e'e = e'Me$ из (1.2.12), мы имеем:

$$q = \frac{e'e}{\sigma^2} = \frac{e'}{\sigma} M \frac{e}{\sigma}.$$

Матрица M , стоящая в середине, идемпотентна (так как она является аннулятором). Кроме того, $e/\sigma|X \sim N(0, I_n)$ из (1.4.1). Поэтому эта квадратичная форма распределена как χ^2 с количеством степеней свободы, равным $\text{rank}(M)^2$. Но $\text{rank}(M) := \text{trase}(M)$, так как матрица M идемпотентна². В процессе доказательства утверждения 1.2 мы уже показали, что $\text{trase}(M) = n - K$. То есть $q|X \sim \chi^2(n - K)$.

(2) b и e — линейные функции от ε (из (1.2.14) и того факта, что $e = M\varepsilon$), следовательно, они совместно нормально распределены при условии X . Кроме того, они некоррелированы при условии X (см. часть (d) утверждения 1.1). Значит, b и e независимо распределены при условии X . Однако z_k — функция от b , а q — функция от e . Таким образом, z_k и q распределены независимо при условии X ³. ■

Правило принятия решений на основании t -теста

Тестирование нулевой гипотезы с помощью t -статистики называется t -тестом и проводится следующим образом:

Шаг 1: При заданном предполагаемом значении, $\bar{\beta}_k$, истинного значения β_k постройте t -отношение по формуле (1.4.5). Слишком большое отклонение t_k от 0 является признаком нарушения нулевой гипотезы. Следующий шаг уточняет, что именно означает «слишком большое» отклонение.

¹ Факт: Если $x \sim N(0, 1)$, $y \sim \chi^2(m)$ и если x и y независимы, то отношение $x/\sqrt{y/m}$ имеет t -распределение с m степенями свободы.

² Факт: Если $x \sim N(0, I_n)$, и матрица A симметрична и идемпотентна, то $x'Ax$ имеет χ^2 распределение с количеством степеней свободы, равным $\text{rank}(A)$.

³ Факт: если A — идемпотентная и симметричная матрица, то $\text{rank}(A) = \text{trase}(A)$.

⁴ Факт: Если x и y независимо распределены, то же можно сказать и о $f(x)$ и $g(y)$.

Шаг 2: Составьте таблицу критических значений t -распределения (большинство учебников по эконометрике и статистике содержит эти таблицы) и посмотрите на вход этой таблицы, соответствующий $n - K$ степеней свободы. Найдите критическое значение (*critical value*), $t_{\alpha/2}(n - K)$, такое, что площадь под кривой t -распределения справа от $t_{\alpha/2}(n - K)$ была бы равна $\alpha/2$. Это иллюстрировано на рис. 1.3. (Если, например, $n - K = 30$, и $\alpha = 5\%$, то $t_{\alpha/2}(n - K) = 2.042$.) Так как t -распределение симметрично относительно нуля,

$$\text{Prob}(-t_{\alpha/2}(n - K) < t < t_{\alpha/2}(n - K)) = 1 - \alpha.$$

Шаг 3: H_0 не отвергается, если $-t_{\alpha/2}(n - K) < t_k < t_{\alpha/2}(n - K)$ (то есть, если $t_k < t_{\alpha/2}(n - K)$), где $t_k = t$ статистика из урав. 1. В противном случае H_0 отвергается. Поскольку $t_k \sim t(n - K)$ при H_0 , то вероятность отвержения H_0 , когда H_0 верна, равна α . То есть размер (уровень значимости) теста равен α .

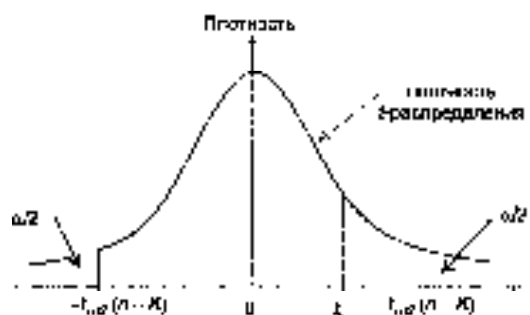


Рис. 1.3. t -распределение

Удобным свойством t -теста является то, что критические значения не зависят от X : нет необходимости вычислять критические значения для каждой выборки.

Доверительный интервал

Шаг 3 может быть переформулирован в терминах b_k и $SE(b_k)$. Поскольку t_k определена соотношением (1.4.5), H_0 не отвергается, если

$$-t_{\alpha/2}(n - K) < \frac{\hat{\beta}_k - \beta_k}{SE(b_k)} < t_{\alpha/2}(n - K)$$

или

$$\hat{\beta}_k - SE(b_k) \cdot t_{\alpha/2}(n - K) < \beta_k < \hat{\beta}_k + SE(b_k) \cdot t_{\alpha/2}(n - K).$$

Поэтому мы не отвергаем H_0 тогда и только тогда, когда тестовое значение $\hat{\beta}_K$ попадает в интервал:

$$[\hat{\beta}_K - SE(\hat{\beta}_K) \cdot t_{\alpha/2}(n - K), \hat{\beta}_K + SE(\hat{\beta}_K) \cdot t_{\alpha/2}(n - K)]. \quad (1.4.6)$$

Этот интервал называется **доверительным интервалом** с уровнем доверия $(1 - \alpha)$, или **$(1 - \alpha)$ -доверительным интервалом** (*level 1 - α confidence interval*). Доверительный интервал тем уже, чем меньше стандартная ошибка. Таким образом, чем меньше стандартная ошибка, тем более точна оценка.

примечание

Правила принятия решений t -теста можно переформулировать в терминах **p -значения** (*p-value*).

Шаг 1: Тот же.

Шаг 2: Вместо нахождения критического значения $t_{\alpha/2}(n - K)$ вычислите

$$p = \text{Prob}(t > |t_K|) \times 2.$$

Так как t -распределение симметрично относительно 0,

$$\text{Prob}(t > |t_K|) = \text{Prob}(t < -|t_K|),$$

так что

$$\text{Prob}(-|t_K| < t < |t_K|) = 1 - p. \quad (1.4.7)$$

Шаг 3: H_0 не отвергается, если $p > \alpha$. В противном случае нулевая гипотеза отвергается.

Чтобы увидеть эквивалентность двух этих правил принятия решений! (одного, основанного на критических значениях $t_{\alpha/2}(n - K)$, другого — на p -значении), обратимся к рис. 1.4. Если $\text{Prob}(t > |t_K|)$ больше, чем $\alpha/2$ (как на рисунке), то есть p -значение больше, чем α , то тогда $|t_K|$ должен быть слева от $t_{\alpha/2}(n - K)$. Это значит (из шага 3), что нулевая гипотеза не отвергается. Таким образом, когда p велико (отношение нечислительной велика для случайной величины из t -распределения), чем меньше p , тем более вероятно решение об отвержении.

Примеры применения t -теста p можно найти в главах 1.7

Линейные гипотезы

Нулевая гипотеза, которую мы хотим протестировать, может отличаться от только что рассмотренного ограничения на индивидуальный коэффициент. Часто речь идет о линейных комбинациях коэффициентов,

и гипотеза записывается в виде системы линейных уравнений:

$$H_0: R\beta = \tau, \quad (1.4.8)$$

где значения R и τ известны и определяются проверяемой гипотезой. Обозначим число уравнений (размерность τ) через $\# \tau$. Тогда R имеет размер $\# \tau \times K$. Эти $\# \tau$ уравнений — ограничения на коэффициенты в поддерживаемой гипотезе. Это называется линейной гипотезой, так как каждое уравнение линейно. Чтобы удостовериться в том, что в системе нет излишних уравнений и что уравнения не противоречат друг другу, потребуем, чтобы $\text{rank}(R) = \# \tau$ (то есть чтобы матрица R имела полный строчный ранг, равный количеству строк). Однако не стоит придавать слишком большое значение этому разговору условно, в конкретных случаях заметить нарушение этого рангового условия, если таковое имеет место, очень легко.

Пример 1.5 (продолжение примера 1.2): Возьмем 4 уравнения ларайнтовой платы из примера 1.2, где $K = 4$. Мы могли бы задаться вопросом, гипотеза о том, что образование и производительность работника в результате табочем месте имеют одинаковое влияние на заработную плату, и общий стаж на заработную плату не влияет. Эта гипотеза состоит из двух уравнений (то есть $\# \tau = 2$):

$$\beta_2 = \beta_3 \quad \text{и} \quad \beta_4 = 0.$$

То же самое можно переписать и в виде $R\beta = \tau$, если определить R и τ следующим образом:

$$R = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \tau = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Так как две строки R линейно независимы, ранговое условие выполнено.

Но предположим, что мы добавляем гипотезу еще одним ограничением:

$$\beta_2 = \beta_3 = \beta_4.$$

Это ограничение избыточно, так как оно выполнялось всегда, когда были выполнены первые два ограничения. Учитывая это ограничение, получаем систему из трех уравнений ($\# \tau = 3$), и

$$R = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & -1 \end{bmatrix}, \quad \tau = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Поскольку третья строка R представляет собой разность двух первых, R теперь не имеет полного строчного ранга. Следовательно добавления

необычного ограничения является то, что матрица R больше не удовлетворяет условию полного стробного ранга

В качестве примера рассмотрим уравнений добавим к первым двум ограничениям еще одно $\beta_2 = 0.5$. Очевидно, коэффициент β_1 не может быть равен 0 и 0,5 одновременно. Гипотеза несостоятельна, так как не существует такого вектора β , который бы одновременно удовлетворял всем условиям. Если мы все же добавим это условие, то матрица R и r будут выглядеть так:

$$R = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad r = \begin{bmatrix} 0 \\ 0 \\ 0.5 \end{bmatrix}.$$

И опять условие полного стробного ранга не выполняется, так как ранг R равен 2, в то время как $\#r = 3$.

F-тест

Для проверки линейных гипотез наиболее мощным тестом статистика, которая имела бы известное распределение при нулевой гипотезе.

Утверждение 1.4 (распределение F-отношения): Пусть предположения 1.1–1.5 выполнены. При нулевой гипотезе $H_0: R\beta = r$, где R — матрица размерности $\#r \times K$, а $\text{rank}(R) = \#r$, «F-отношение» (F-ratio), определяемое как

$$F = \frac{(Rb - r)' [R(X'X)^{-1}R']^{-1} (Rb - r) / \#r}{(Rb - r)' [RV\widehat{\text{cov}}(b; X) R']^{-1} (Rb - r) / \#r} \quad \text{из (1.3.4)} \quad (1.4.9)$$

распределено как $F(\#r, n - K)$ (F-распределение Фишера с $\#r$ и $n - K$ степенями свободы).

Как и в утверждении 1.3, достаточно показать, что распределение условно относительно X , — это $F(\#r, n - K)$. Поскольку F-распределение не зависит от X , оно совпадает с безусловным распределением.

Доказательство: Так как $\sigma^2 = n^{-1}n/(n - K)$, мы можем записать:

$$F = \frac{w/\frac{n}{\sigma^2}}{q/(n - K)},$$

где

$$w \equiv (Rb - r)' (\sigma^2 \cdot R(X'X)^{-1}R')^{-1} (Rb - r) \quad \text{и} \quad q \equiv \frac{\sigma^2 n}{\sigma^2}.$$

Теперь нам необходимо показать,

- (1) $w|X \sim \chi^2(\#r)$;
 (2) $e|X \sim \chi^2(n-K)$ (это часть 1) доказательства утверждения 1.3);
 (3) w и e независимо распределены при условии X .

Тогда, по определению F -распределения, F -отношение $\sim F(\#r, n-K)$.

- 1) Пусть $e = Hb - \tau$ (при $Hb, Hb - \tau = H(b - \theta)$). Значит, из (1.4.2), условное относительно X распределение e порождаемо с математическим ожиданием 0 и дисперсией

$$\begin{aligned}\text{Var}(e|X) &= \text{Var}(H(b - \beta)|X) = H \text{Var}(b - \beta|X) H' = \\ &= \sigma^2 \cdot H(X'X)^{-1} H',\end{aligned}$$

которая является, по тем или иным, как обратной матрицей в средней матрице n -квadraticной форме для w . Значит, w можно записать как $w' \text{Var}(e|X)^{-1/2} e$. Поскольку H — матрица полного строкового ранга и $X'X$ — невырожденная, $\sigma^2 H(X'X)^{-1} H'$ невырождена. (Почему? Показать это потребуется в одном из контрольных вопросов.) Поэтому, по определению χ^2 -распределения, $w|X \sim \chi^2(\#r)$.

- 3) e является функцией от b , а σ — функция от e . Но b и e независимо распределены при условии X , как показано в части (2) утверждения 1.3. То есть w и σ независимо распределены при условии X .

■

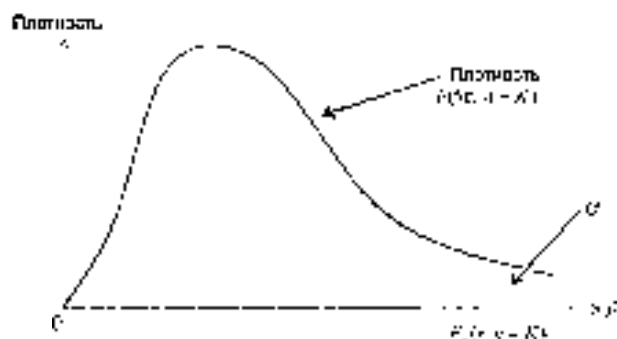
Если нулевая гипотеза $H\beta = \tau$ верна, то мы ожидаем, что $R\beta - \tau$ будет небольшим. То есть большие значения F должны восприниматься как свидетельство неплатежеспособности нулевой гипотезы. Это означает, что мы сможем только на правый хвост распределения F -статистики. Правильно принятом решении на основании F -статистики на уровне значимости α сформулированы так:

Шаг 1: Вычисляйте F -отношение по формуле (1.4.9).

Шаг 2: Обратитесь к таблице критических значений для F -распределения и посмотрите на вход для $\#r$ (количество степеней свободы числителя) и $n-K$ (количество степеней свободы знаменателя). Найдите критическое значение $F_{\alpha, \#r, n-K}$, для которого вероятность попасть в правый хвост F -распределения равна α . Это проиллюстрировано на рис. 1.1. Например, когда $\#r = 3$, $n-K = 30$ и $\alpha = 5\%$, критическое значение $F_{0.05}(3,30) = 2.92$.

¹Факт: Пусть x — n -мерный случайный вектор. Если $x \sim N(\mu, \Sigma)$, где матрица Σ невырожденная, то $(x - \mu)' \Sigma^{-1} (x - \mu) \sim \chi^2(n)$.

Шаг 3: Нулевая гипотеза не отвергается, если F -отношение из шага 1 меньше, чем $F_{\alpha}(k_1, k_2)$ (рис. 1.4). В противном случае нулевая гипотеза отвергается.

Рис. 1.4. F -распределение

Это правило принятия решений может также быть сформулировано в терминах p -значения:

Шаг 1: Тот же, что и был.

Шаг 2: Вычислите

p — площадь правого хвоста F -распределения справа от F -отношения.

Шаг 3: Нулевая гипотеза не отвергается, если $p > \alpha$. В противном случае H_0 отвергается.

Таким образом, малое p -значение является сигналом принятия нулевой гипотезы.

Более удобное выражение для F

Процесс вывода F -отношения, изложенный выше, базируется на **принципе Вальда** (*Wald principle*), поскольку в его основе лежит оценка без ограничений (то есть не учитывающая ограничения, накладываемые нулевой гипотезой). Вычисление F -отношения по формуле (1.4.9) требует обращения матрицы и умножения матриц. Однако существует альтернативная формула, требующая только подсчета двух сумм квадратов остатков: максимизированной суммы квадратов остатков, полученной по формуле (1.2.1) (которую обозначим как SSR_0), и максимизированной суммы квадратов остатков, рассчитанной с учетом ограничений (ее обозначим как SSR_H), получаемой как

$$\min_{\beta} SSR(\beta), \text{ при ограничениям } R\beta = r. \quad (1.4.10)$$

Отыскание такой оценки $\tilde{\beta}$, которая удовлетворяет задаче минимизации с ограничениями, называется **регрессией с ограничениями**, или **методом наименьших квадратов с ограничениями** (*restricted regression, restricted least squares*). В качестве аналитического выражения читателем предлагается показать, что F -отношение можно представить в следующем виде:

$$F = \frac{(SSR_R - SSR_U)/q}{SSR_U/(n - k)}, \quad (1.4.11)$$

то есть как разность двух целевых функций, деленную на оценку дисперсии ошибки. Такой способ выведения F -отношения похож на выведение статистики отношения максимальных правдоподобий при оценивании методом максимального правдоподобия (эта статистика рассчитывается как разница логарифмов правдоподобия с ограничениями, накладываемыми нулевой гипотезой, и без них). Но этот принцип и выраде F -отношения в такой форме говорят, что он основывается на **принципе отношения правдоподобий** (*Likelihood-Ratio principle*). Выражение для OLS-оценки с ограничениями можно найти аналитически. Это оставлено читателем в качестве аналитического упражнения. Вычисление OLS-оценки с ограничениями будет иллюстрировано в параграфе 1.7 на эмпирическом примере.

Сравнение t и F

Поскольку гипотезы относительно отдельных коэффициентов регрессии являются линейными, t -тест $H_0: \beta_k = \bar{\beta}_k$ является частным случаем F -теста. Чтобы увидеть это, заметим, что такая гипотеза может быть переписана в форме $R\beta = r$, если определить R и r следующим образом:

$$R = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & & \end{pmatrix}_{(1 \times k)}, \quad r = \bar{\beta}_k.$$

Поэтому, в соответствии с (1.4.9), F -отношение равно:

$$F = (\hat{\beta}_k - \bar{\beta}_k) [s^2 \cdot (k, k) \text{ элемент матрицы } (X'X)^{-1}]^{-1} (\hat{\beta}_k - \bar{\beta}_k),$$

что равно квадрату t -отношения (1.4.5). Поскольку случайная величина, распределенная по $F(1, n - k)$, представляет собой квадрат случайной величины, распределенной по $t(n - k)$, то F - и t -тесты дают одинаковые результаты.

Интересная гипотеза состоит в том, что несколько коэффициентов регрессии принимают конкретные значения. Например, предположим, что $k = 2$, и рассмотрим нулевую гипотезу:

$$H_0: \beta_1 = 1 \quad \text{и} \quad \beta_2 = 0.$$

Это можно записать в виде линейной гипотезы $H_0: \beta = \tau$, где $K = I_2$ и $\tau = (1, 0)'$. То есть можно использовать F -тест. Однако заманчивой является идея использования t -тестов для каждого коэффициента по отдельности. Мы могли бы не отвергать H_0 , когда оба ограничения $\beta_1 = 1$ и $\beta_2 = 0$ проходят t -тест. Это сводится к использованию доверительной области

$$\{(\beta_1, \beta_2) | b_1 - SE(b_1) \cdot t_{\alpha/2}(n-K) < \beta_1 < b_1 + SE(b_1) \cdot t_{\alpha/2}(n-K), \\ b_2 - SE(b_2) \cdot t_{\alpha/2}(n-K) < \beta_2 < b_2 + SE(b_2) \cdot t_{\alpha/2}(n-K)\},$$

которая представляет собой прямоугольную область на плоскости (β_1, β_2) , как показано на рис. 1.5. Если точка $(1, 0)$ (точка на гипотезе (β_1, β_2) , специфицированная нулевой (иногда) догадкой) попадает в эту область, нулевая гипотеза не отвергается. С другой стороны, доверительной областью для F -теста является:

$$\{(\beta_1, \beta_2) | (b_1 - \beta_1, b_2 - \beta_2)' (\text{Var}(\hat{b}|X))^{-1} (b_1 - \beta_1, b_2 - \beta_2) < 2F_{\alpha}(2, n-K)\}.$$

Поскольку матрица $\text{Var}(\hat{b}|X)$ положительно определена, доверительная область для F -теста представляет собой эллипс в плоскости (β_1, β_2) . Два найденные доверительные области обычно выглядят примерно так, как изображено на рис. 1.5.

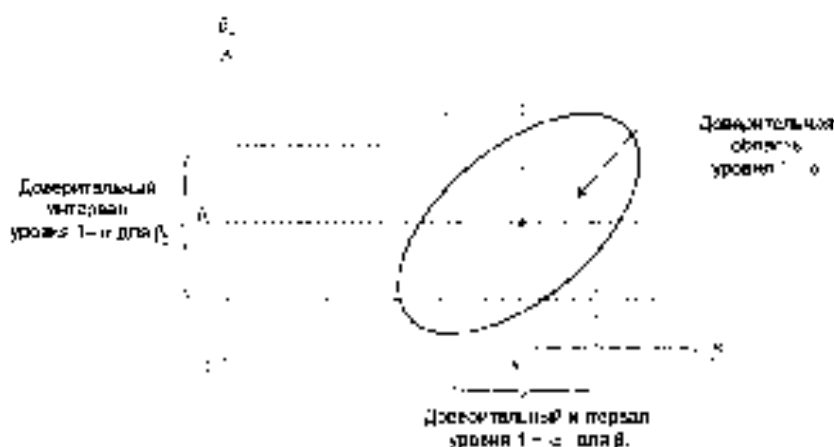


Рис. 1.5. Сравнение t - и F -тестов

Использование F -теста по сравнению с двумя индивидуальными t -тестами предпочтительнее по двум причинам. Во-первых, если размер (уровень значимости) каждого из этих t -тестов равен α , то результирующий размер (вероятность того, что точка $(1, 0)$ находится за пределами прямоугольной области) не равен α . Во-вторых, как будет отмечено в следующем параграфе (см. § 5.19), F -тест является тестом однородности

правдоподобий, а тесты отношения правдоподобий обладают определенными желательными свойствами. То есть, даже если уровень значимости каждого t -теста подобран так, чтобы результирующий уровень значимости был равен α , такой путь все равно менее предпочтителен, чем использование F -теста¹.

Пример тестовой статистики, распределение которой зависит от X

Чтобы написать текущий параграф в более широкую картину, необходимо отметить, что существуют некоторые статистики, условные распределение которых зависят от X . Рассмотрим знаменитую статистику Дарбина — Уотсона (*Durbin — Watson statistic*):

$$\frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}.$$

Условное распределение этой статистики и, следовательно, ее критические значения зависят от X . Однако Дарбин и Уотсон показали, что эти критические значения попадают в определенный интервал (границы которого зависят от размера выборки, количества регрессоров и того, включена ли в модель константа). Поэтому критические значения безусловного распределения также попадают в эти границы.

Эта статистика предназначена для проверки гипотезы об отсутствии сериальной корреляции в ошибках. Таким образом, нулевая гипотеза отражает предположение 1.4, в то время как поддерживаемая гипотеза представляет прочие предположения классической регрессионной модели (включая строгую экзотенность) и нормальность. Однако, как уже подчеркивалось в параграфе 1.1, предположение о строгой экзотенности не выполняется в моделях временных рядов, а сериальная корреляция может возникнуть только в таких моделях. Таким образом, статистика Дарбина — Уотсона бесполезна в эконометрике. Тесты на сериальную корреляцию, основанные на асимптотический теории, будут изложены в следующей главе.

Контрольные вопросы

1. (Условное и безусловное распределение.) Следует ли из предположений 1.1–1.5, что предельное (безусловное) распределение b нормально? [Ответ: Нет.] Распределены ли статистики t_k (см. (1.4.3)), t_k и F независимо от X ? [Ответ: Да, так как их условные распределения относительно X не зависят от X .]
2. (Вычисление тестовых статистик.) Проверьте, что $SE(b_k)$, b , SSR , s^2 и R^2 можно вычислять на основе следующих выборочных средних: S_{xx} , S_{xy} , $\sum y_i^2/n$ и $\bar{y}$.

¹ Взаимосвязь между t - и F -тестами более подробно изложена в [Scheffe, 1959, p. 46].

3. Чтобы формулы (1.4.9) для F были определены, матрица $R(X'X)^{-1}R'$ должна быть невырожденной. Докажите более строгий результат, что эта матрица должна быть положительно определенной. Укажите: $X'X$ — положительно определенная матрица. Матрица, обратная к положительно определенной матрице, также положительно определена. Так как матрица $R(\frac{1}{n}I \times K)$ является строковым ранга, то для любого ненулевого $\frac{1}{n}I$ ненулевого вектора z $R'z \neq 0$.
4. (Односторонний t -тест) t -тест, описанный выше, является **двусторонним t -тестом** (*two-tailed t -test*), так как уровень значимости α поровну распределен между двумя хвостами t -распределения. Предположим, что альтернатива односторонняя и записывается как $H_1: \beta_1 > \beta_0$. Разработайте следующую модификацию правила принятия решений на основании t -теста

Шаг 1: Тот же, что и ранее

Шаг 2: Найдите такое критическое значение t_α , чтобы площадь под кривой плотности t -распределения справа от t_α равнялась бы α . Обратите внимание на отличие от двустороннего теста: левый хвост распределения игнорируется и площадь размера α приписывается только правому хвосту.

Шаг 3: Нулевая гипотеза не отвергается, если $t_d \leq t_\alpha$. В противном случае нулевая гипотеза отвергается.

Покажите, что размер (уровень значимости) такого **одностороннего t -теста** равен α .

5. (Безысходность между $F(1, n-K)$ и $t(n-K)$) Воспользуйтесь таблицами критических значений для F - и t -распределений и проверьте, что $F_\alpha(1, n-K) = (t_{\alpha/2}(n-K))^2$, а в действительности таблица и уровень значимости по вашему выбору.
6. (Связь ограничения в F -тесте) «Бессмысленно тестировать гипотезу, состоящую из большого количества ограничений и членов функции, так как t -тест скорее всего, отвергнет по крайней мере некоторые из этих ограничений». Можете ли вы ошибиться в этом утверждении?
7. (Дисперсия s^2 .) Покажите, что при предположениях 1.1–1.3

$$\text{Var}(s^2|X) = \frac{2\sigma^4}{n-K}.$$

Указание: Если случайная величина распределена как $\chi^2(m)$, то ее математическое ожидание равно m , а дисперсия равна $2m$.

1.5. Связь с методом максимального правдоподобия

После того как мы специфицировали распределение вектора ошибок ϵ , для оценивания параметров модели (β, σ^2) можно использовать **метод максимального правдоподобия** (*maximum likelihood principle — ML*)¹

¹ Более подробный метод максимального правдоподобия описан в главе 6.

В этом параграфе мы покажем, что OLS-оценка вектора β ($\hat{\beta}$) совпадает с ML-оценкой, а OLS-оценка σ^2 лишь слегка отличается от ML-оценки при предположении нормальности ошибок. Мы также покажем, что $\hat{\beta}$ достигает **нижней границы Крамера — Рао** (*Cramer — Rao lower bound*).

Метод максимального правдоподобия

Как вы, возможно, помните из вводного курса статистики, основная идея метода максимального правдоподобия состоит в том, что оценки параметров вычисляются таким образом, чтобы максимизировать вероятность получения имеющейся выборки. Более формально, мы предполагаем, что плотность выборки (y, X) принадлежит семейству плотностей с конечномерным вектором параметров ζ : $f(y, X; \zeta)$. (Это называется **параметризацией** (*parameterizing*) функцией плотности.) Эта функция, если рассматривать ее как функцию от гипотетического вектора параметров ζ , называется **функцией правдоподобия** (*likelihood function*). При истинном значении вектора параметров ζ функция плотности (y, X) равна $f(y, X; \zeta)$. ML-оценкой неизвестного вектора параметров ζ называется вектор $\hat{\zeta}$, который максимизирует функцию правдоподобия с учетом имеющейся выборки (y, X) .

Условная и безусловная функции правдоподобия

Поскольку (совместная) функция плотности вероятности равна произведению частной функции плотности вероятности на условную, плотность (y, X) можно записать как

$$f(y, X; \zeta) = f(y|X; \theta) \cdot f(X; \psi), \quad (1.5.1)$$

где θ — подмножество вектора параметров ζ , которое определяет условную функцию плотности, а ψ — подмножество, определяющее частную функцию плотности. Интерпретируя нас параметры заключены в векторе θ . Для линейной модели регрессии с нормально распределенными ошибками $\theta = (\beta', \sigma^2)'$ и $f(y|X; \theta)$ определена в (1.5.4) ниже по тексту.

Пусть $\tilde{\zeta} = (\tilde{\theta}', \tilde{\psi}')'$ — гипотетическое значение $\zeta = (\theta', \psi')'$. Тогда (безусловная, или совместная) функция правдоподобия имеет вид

$$f(y, X; \tilde{\zeta}) = f(y|X; \tilde{\theta}) \cdot f(X; \tilde{\psi}). \quad (1.5.2)$$

Если бы параметрическая форма $f(X; \tilde{\psi})$ была известна, то мы могли бы максимизировать эту совместную функцию правдоподобия по всему гипотетическому вектору параметров ζ и ML-оценка θ была бы просто подмножеством вектора ML-оценок ζ . Мы не можем пойти по такому пути, так как модель классической линейной регрессии не специфицирует функцию плотности вероятности $f(X; \tilde{\zeta})$. Однако при отсутствии

функциональной зависимости между $\tilde{\theta}$ и $\tilde{\psi}$ (например, возможно, что подмножество $\tilde{\theta}$ является функцией от $\tilde{\psi}$) максимум (1.5.2) по $\tilde{\zeta}$ достигается при неэтаксимой максимизации $f(y|X;\tilde{\theta})$ по $\tilde{\theta}$ и $f(X;\tilde{\psi})$ по $\tilde{\psi}$. Таким образом, МЛ-оценка $\hat{\theta}$ также доставляет максимум условной функции правдоподобия (conditional likelihood) $f(y, X; \hat{\zeta})$.

Логарифмическое правдоподобие для регрессионной модели

Как мы уже видели, из совокупности предположений 1.5 (о нормальности ошибок), 1.2 и 1.4 следует, что распределение y при условии X нормально $N(X\beta, \sigma^2 I_n)$ (см. (1.4.1)). Учитывая, что $y = X\beta + \varepsilon$, по предположению 1.1, мы получаем

$$y|X \sim N(X\beta, \sigma^2 I_n). \quad (1.5.3)$$

Таким образом, условная плотность y при условии X равна

$$f(y|X) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right]. \quad (1.5.4)$$

Заменяя конкретные значения параметров (β, σ^2) их гипотетическими значениями $(\tilde{\beta}, \tilde{\sigma}^2)$, и беря логарифм, получаем логарифмическую функцию правдоподобия (log likelihood function):

$$\log L(\tilde{\beta}, \tilde{\sigma}^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\tilde{\sigma}^2) - \frac{1}{2\tilde{\sigma}^2}(y - X\tilde{\beta})'(y - X\tilde{\beta}). \quad (1.5.5)$$

Поскольку взятие логарифма — монотонное преобразование, МЛ-оценкой (β, σ^2) является пара $(\tilde{\beta}, \tilde{\sigma}^2)$, максимизирующая логарифм функции правдоподобия.

Максимизация правдоподобия с помощью концентрированной функции правдоподобия

Полезным упражнением является нахождение максимумов логарифмической функции правдоподобия в два этапа. Сначала найти максимум выражения по $\tilde{\beta}$ для любого заданного значения $\tilde{\sigma}^2$. В общем случае $\tilde{\beta}$, при котором достигается максимум, может зависеть от $\tilde{\sigma}^2$ (хотя бы в данном случае это не так при выполнении предположений 1.1–1.5). После этого найдите максимум по $\tilde{\sigma}^2$, принимая во внимание, что $\tilde{\beta}$ (найденный

¹Экономисты элементарной теории вероятностей часто утверждают, что функции плотности и вероятности для нормального распределения с математическим ожиданием μ и ковариационной матрицей Σ равны

$$f(y) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(y - \mu)' \Sigma^{-1}(y - \mu)\right]$$

Чтобы получить (1.5.4), просто подставьте $\mu = X\beta$ и $\Sigma = \sigma^2 I_n$.

на предыдущем шаге) может зависеть от $\hat{\sigma}^2$. То ириформная функция правдоподобия, в которой $\hat{\beta}$ представляет собой выражение из первого шага процедуры, называется **концентрированной (по $\hat{\beta}$) логарифмической функцией правдоподобия** (*concentrated (with respect to $\hat{\beta}$) log likelihood function*). Для нормально логарифмического правдоподобия (1.5.5) первый шаг сводится к минимизации суммы квадратов $(y - X\hat{\beta})'(y - X\hat{\beta})$. Значение вектора $\hat{\beta}$, при котором достигается минимум, совпадает с OLS-оценкой $\hat{\beta}$, а значение суммы квадратов в точке минимума равно $e'e$. Таким образом, концентрированное логарифмическое правдоподобие равно

концентрированное логарифмическое правдоподобие =

$$= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} e'e. \quad (1.5.6)$$

Оно представляет собой функцию только от $\hat{\sigma}^2$, а значения $\hat{\sigma}^2$, максимизирующее концентрированное логарифмическое правдоподобие, — это ML-оценка σ^2 . В рассматриваемом случае классической линейной регрессионной модели как минимум можно найти непосредственно, так как $e'e$ не зависит от σ^2 и, следовательно, может рассматриваться как константа. Однако взятие производной по $\hat{\sigma}^2$ (а не по $\hat{\sigma}$) может быть затруднительным. Но проблем можно избежать, если обозначить $\hat{\sigma}^2$ за $\tilde{\sigma}$. Беря производную от (1.5.6) по $\tilde{\sigma}$ ($\equiv \hat{\sigma}^2$) и приравнявая ее нулю, получаем следующий результат.

Утверждение 1.5 (ML-оценка (β, σ^2)): Пусть предположения 1.1–1.5 выполнены. Тогда ML-оценка β совпадает с OLS-оценкой $\hat{\beta}$, и

$$\text{ML-оценка } \sigma^2 = \frac{1}{n} e'e = \frac{SSR}{n} = \frac{n-K}{n} s^2. \quad (1.5.7)$$

Из утверждения 1.2 нам известно, что s^2 не смещена. Так как s^2 умножается на $(n-K)/n$ (множитель, отличный от единицы), то ML-оценка σ^2 является смещенной, хотя смещение стремится к нулю при количестве наблюдений n , стремящемся к бесконечности, для любого фиксированного K .

Чтобы воспользоваться этим в дальнейшем, выпишем значения функции правдоподобия в точке максимума. Подставляя (1.5.7) в (1.5.6), получаем

значение логарифма правдоподобия в точке максимума =

$$= -\frac{n}{2} \log\left(\frac{2\pi}{n}\right) - \frac{n}{2} - \frac{n}{2} \log(SSR).$$

так что значение функции правдоподобия в точке максимума равно

$$\max_{\vec{\theta}} L(\vec{\theta}; \vec{z}) = \left(\frac{2\pi}{n}\right)^{-n/2} \exp\left(-\frac{n}{2}\right) (SSR)^{-n/2}. \quad (1.5.8)$$

Граница Крамера — Рао для классической регрессионной модели

Чтобы освежить ваши знания статистики, напомним старый классическую модель линейной регрессии и приведем без доказательства неравенства Крамера — Рао для ковариационной матрицы произвольной несмещенной оценки. Для этого определим вектор *ошибки* (или *остаток*) для статистического значения параметра θ как градиент (вектор частных производных) логарифмической функции правдоподобия:

$$\text{ошибка: } e(\vec{\theta}) = \frac{\partial \log L(\vec{\theta})}{\partial \vec{\theta}}. \quad (1.5.9)$$

Неравенство Крамера — Рао: Пусть $\mathbf{z}$ — вектор случайных величин (необязательно независимых), совместная функция плотности вероятности которых задана как $f(\mathbf{z}; \theta)$, где θ — m -мерный вектор параметров в m -мерном пространстве параметров Θ . Пусть $L(\vec{\theta}) = f(\mathbf{z}; \vec{\theta})$ — функция правдоподобия, $\hat{\theta}(\mathbf{z})$ — несмещенная оценка θ с ковариационной матрицей. Тогда если выполнены *классические условия регулярности* $f(\mathbf{z}; \theta)$ (неформулируемые здесь), то

$$\text{Var}[\hat{\theta}(\mathbf{z})] \geq I(\theta)^{-1} \quad (\text{= нижняя граница Крамера — Рао}),$$

(max m)

где $I(\theta)$ — **информационная матрица**, определенная как

$$I(\theta) \equiv E_{\theta}[e(\theta) e(\theta)']. \quad (1.5.10)$$

(Обратите внимание, что в формуле стоит *значение* выража в точке истинного значения параметра θ .) Также при классических условиях регулярности ковариационная матрица равна матрице, обратной к матричному ожиданию матрицы Гессе (матрице вторых производных) логарифмической функции правдоподобия:

$$I(\theta) = -E \left[\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'} \right]. \quad (1.5.11)$$

Эта выражение называется равенством для информационных матриц (*information matrix equality*).

По методу формулировки условий регулярности и доказательства соответствующего утверждения см., например, [Амелтува, 1985, Секция 1.3.2]. Условия регулярности обеспечивают то, что операции дифференцирования и взятия математического ожидания можно поменять местами. Поэтому, например,

$$E[\partial L(\theta)/\partial \theta] = \partial E[L(\theta)]/\partial \theta.$$

Вернемся к модели классической линейной регрессии (предположения 1.1–1.5 выше). Функция правдоподобия $L(\theta)$ в выражении Кракера — Рао равна условной плотности (1.3.4), так что дисперсия и непостоянство дисперсии дисперсии при условии X . Можно показать, что условия регулярности выполняются для функции плотности нормального распределения (1.5.4) (см., например, [Амелтува, 1985, Секция 1.3.2 или 1.3.3]). В оставшейся части этой главы рассмотрим как вычислять информационную матрицу (*information matrix*) или (1.5.4). Вектор параметров θ равен $(\beta, \sigma^2)'$. Поэтому $\tilde{\theta} = (\tilde{\beta}, \tilde{\sigma}^2)'$, а матрица вторых производных равна:

$$\frac{\partial^2 \log L(\theta)}{\partial \tilde{\theta} \partial \tilde{\theta}'} = \begin{bmatrix} \frac{\partial^2 \log L(\theta)}{\partial \tilde{\beta} \partial \tilde{\beta}'} & \frac{\partial^2 \log L(\theta)}{\partial \tilde{\beta} \partial \tilde{\sigma}^2} \\ \frac{\partial^2 \log L(\theta)}{\partial \tilde{\sigma}^2 \partial \tilde{\beta}'} & \frac{\partial^2 \log L(\theta)}{\partial \tilde{\sigma}^2 \partial \tilde{\sigma}^2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sigma^2} X'X & 0 \\ 0 & -\frac{n}{\sigma^4} \end{bmatrix}. \quad (1.5.12)$$

Первая и вторая производные логарифмической функции правдоподобия (1.5.5) по $\tilde{\theta}$ в точке истинного значения вектора θ равны:

$$\frac{\partial \log L(\theta)}{\partial \tilde{\beta}} = \frac{1}{\sigma^2} X'(y - X\beta), \quad (1.5.13a)$$

$$\frac{\partial \log L(\theta)}{\partial \tilde{\sigma}^2} = \frac{n}{2\sigma^2} - \frac{1}{2\sigma^4} (y - X\beta)'(y - X\beta), \quad (1.5.13b)$$

$$\frac{\partial^2 \log L(\theta)}{\partial \tilde{\beta} \partial \tilde{\beta}'} = -\frac{1}{\sigma^2} X'X, \quad (1.5.14a)$$

$$\frac{\partial^2 \log L(\theta)}{\partial \tilde{\sigma}^2{}^2} = \frac{n}{2\sigma^4} - \frac{1}{2\sigma^6} (y - X\beta)'(y - X\beta), \quad (1.5.14b)$$

$$\frac{\partial^2 \log L(\theta)}{\partial \tilde{\beta} \partial \tilde{\sigma}^2} = -\frac{1}{\sigma^4} X'(y - X\beta). \quad (1.5.14c)$$

Поскольку эти производные вычислены при истинном значении параметров, в этих выражениях $y - X\beta = e$. Подставляя (1.5.14) в (1.5.12) и используя $E[e'X] = 0$ (предположение 1.2), $E[e'e'X] = 0$ (следствие из предположения 1.4), также используя в производной замене $\sigma^2 = \sigma^2$, несложно вывести

$$I_1(\theta) = \begin{bmatrix} \frac{1}{\sigma^2} X'X & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}. \quad (1.5.15)$$

Здесь используется условное математическое ожидание относительно X , так как функция правдоподобия (1.5.4) представляет собой условную плотность вероятности относительно X . Обращая эту блочно-диагональную матрицу, получаем границу Крамера — Рао:

$$\text{граница Крамера — Рао} \equiv I(\theta)^{-1} = \begin{bmatrix} \sigma^2 \cdot (X'X)^{-1} & \mathbf{0} \\ \mathbf{0}' & \frac{2\sigma^4}{n} \end{bmatrix}. \quad (1.5.16)$$

Поэтому несмещенная оценка b с дисперсией $\sigma^2(X'X)^{-1}$ (по утверждению 1.1) достигает границы Крамера — Рао. Таким образом, мы доказали

Утверждение 1.6 (b — наилучшая несмещенная оценка (BUE)): При предположениях 1.1–1.5 OLS-оценка b неизвестного вектора параметров β является **наилучшей несмещенной оценкой (Best Unbiased Estimator)** в том смысле, что любая другая несмещенная оценка (но необязательно линейная) обладает большей (в матричном смысле) условной дисперсией.

Этот результат не следует путать с теоремой Гаусса — Маркова, которая утверждает, что b обладает наименьшей дисперсией в классе несмещенных и линейных по y оценок. Утверждение 1.6 говорит, что b обладает наименьшей дисперсией в более широком классе оценок, который включает в себя линейные несмещенные оценки. Более сильное утверждение мы получили с использованием предположения нормальности (предположение 1.5), которого нет в теореме Гаусса — Маркова. Иными словами, теорема Гаусса — Маркова не исключает возможности существования нелинейной оценки с меньшей дисперсией, чем дисперсия OLS-оценки, однако такая возможность исключается при введении предположения нормальности.

Как мы уже видели, ML-оценка s^2 является смещенной, следовательно, неравенство Крамера — Рао неприменимо. Однако МНК-оценка дисперсии s^2 является несмещенной. Достигает ли она границы Крамера — Рао? В одном из контрольных вопросов к предыдущему параграфу мы показали, что

$$\text{Var}(s^2 | X) = \frac{2\sigma^4}{n - K}$$

при том же самом наборе предположений, что и в утверждении 1.6. Поэтому s^2 не достигает границы Крамера — Рао, равной $2\sigma^4/n$. Однако можно показать, что не существует несмещенной оценки σ^2 с дисперсией, меньшей чем $2\sigma^4/(n - K)$ (см., например: [Рао, 1973, р. 319]).

F-тест как тест отношения правдоподобий

Тест отношения правдоподобий (*likelihood ratio test*) для некоторой нулевой гипотезы сравнивает значения L_U (максимальное значение функции правдоподобия, полученное без учета ограничений, накладываемых нулевой гипотезой) и L_R (максимальное значение функции правдоподобия, полученное с учетом ограничений, накладываемых нулевой гипотезой). Если значение отношения правдоподобий $\lambda = L_U/L_R$ слишком велико, то это должно быть сигналом к тому, что нулевая гипотеза неверна. *F*-тест, проверяющий нулевую гипотезу $H_0: R\beta = r$, рассмотренный в предыдущем параграфе, является тестом отношения правдоподобий, поскольку *F*-отношение является монотонным преобразованием отношения правдоподобий λ . Для рассматриваемой сейчас модели L_U вычисляется по формуле (1.5.8), где SSR (сумма квадратов остатков, минимизированная без учета ограничений H_0) равна SSR_U в (1.4.11). Значение L_R получается путем замены SSR на сумму квадратов остатков при ограничении, SSR_R . Таким образом,

$$L_U = \max_{\beta, \sigma^2: \text{при } H_0} L(\beta, \sigma^2) = \left(\frac{2\pi}{n}\right)^{n/2} \cdot \exp\left(-\frac{n}{2}\right) \cdot (SSR_R)^{-n/2}. \quad (1.5.17)$$

и отношение правдоподобий равно:

$$\lambda = \frac{L_U}{L_R} = \left(\frac{SSR_U}{SSR_R}\right)^{n/2}. \quad (1.5.18)$$

Сравнивая эту формулу с формулой (1.4.11) для *F*-отношения, можно увидеть, что *F*-отношение является монотонным преобразованием отношения правдоподобий λ :

$$F = \frac{n}{k-r} \frac{K}{\lambda^{2/n} - 1}, \quad (1.5.19)$$

следовательно, эти два теста эквивалентны.

Метод квазимаксимального правдоподобия

Все результаты, полученные до сих пор, базируются на предположении о чистоте ошибок. Без этого предположения нет гарантии того, что ML-оценка вектора параметра β совпадает с OLS-оценкой (утверждение 1.5) или что OLS-оценка b достигает границы Крамера-Рао (утверждение 1.6). Однако из утверждения 1.5 следует, что b является оценкой метода квази- (или псевдо-) максимального правдоподобия (*quasi- (pseudo-) maximum likelihood estimator — quasi-ML*). Так называется оценка, максимизирующая неправильно специфицированную функцию правдоподобия. В нашем случае это функция правдоподобия для нормального распределения. Следовательно, результаты

параграфы 1.3 могут быть интерпретированы как дающие свойства квази-ML-оценок на конечных выборках, когда мы некорректно предполагаем нормальность ошибок.

Контрольные вопросы

1. (Использование условий регулярности.) Предполагая, что можно менять местами операции взятия математического ожидания (то есть взятия интеграла) и дифференцирования, докажите, что математическое ожидание выража в (1.5.9), оцененное при истинном значении вектора θ , равно нулю.
Указание: Необходимо показать, что

$$\int \frac{\partial \log f(z; \theta)}{\partial \theta} f(z; \theta) dz = 0.$$

Так как $f(z; \theta)$ — функция плотности, то $\int f(z; \theta) dz = 1$ для любого θ . Дифференцируйте обе части по θ и используйте условия регулярности, позволяющие менять местами операции интегрирования и дифференцирования, чтобы получить $\int [\partial f(z; \theta)] dz = 0$. Также вспомните, что

$$\frac{\partial \log f(z; \theta)}{\partial \theta} = \frac{1}{f(z; \theta)} \frac{\partial f(z; \theta)}{\partial \theta}.$$

2. (Максимизация логарифма совместной функции правдоподобия.) Рассмотрим задачу максимизации логарифма совместной функции правдоподобия (1.5.2) для классической модели регрессии, где $\tilde{\theta} = (\beta', \sigma^2)'$, а $\log f(y|X; \tilde{\theta})$ задан в (1.5.5). Как необходимо параметризовать маргинальную функцию правдоподобия $f(X; \tilde{\psi})$ и взять логарифм от (1.5.2), чтобы получить целевую функцию, которую, в свою очередь, необходимо будет максимизировать, по $\zeta = (\beta', \sigma^2)'$? Что является ML-оценкой $\hat{\theta} = (\hat{\beta}', \hat{\sigma}^2)'$? (Ответ: Она будет такой же, как и оценка из утверждения 1.5.) Выведите границу Крамера — Рао для $\hat{\beta}$. **Указание:** Из равенства для информационных матриц

$$I(\zeta) = -E \left[\frac{\partial^2 \log L(\zeta)}{\partial \zeta \partial \zeta'} \right],$$

кроме того, $\partial^2 \log L(\zeta) / (\partial \tilde{\theta} \partial \tilde{\theta}') = 0$.

3. (Логарифм концентрированной по σ^2 функции правдоподобия.) Если обозначить $\tilde{\sigma}^2$ как $\tilde{\gamma}$, получаем логарифм функции правдоподобия для классической регрессионной модели в виде:

$$\log L(\hat{\beta}, \tilde{\gamma}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\tilde{\gamma}) - \frac{1}{2\tilde{\gamma}} (y - X\hat{\beta})'(y - X\hat{\beta}).$$

При использовании двухшаговой процедуры максимизации, описанной в тексте, мы сначала максимизировали эту функцию по $\hat{\beta}$. Сейчас же на первом шаге найдем максимум по $\tilde{\gamma}$, считая $\hat{\beta}$ заданными. Покажите,

что логарифм концентрированной $QO(\beta) \equiv \bar{Q}^2$ функции правдоподобия равен:

$$-\frac{n}{2} [1 + \log(2\pi)] - \frac{n}{2} \log \left(\frac{(y - X\bar{\beta})'(y - X\bar{\beta})}{n} \right).$$

4. (Разности для нонинформированных матриц в классической регрессионной модели.) Проверьте (1.5.11) для линейной регрессионной модели.

5. (Уравнения правдоподобия для классической регрессионной модели.)

Мы использовали двухшаговую процедуру, чтобы вывести ML-оценку для классической регрессионной модели. Альтернативный подход поиска ML-оценки состоит в приравнивании нулю первых частей (1.5.13) и решения полученной системы (эти условия первого порядка для логарифмической функции правдоподобия называются **уравнениями правдоподобия** (*likelihood equations*)). Проверьте, что ML-оценка из утверждения 1.5 является решением этой системы уравнений.

1.6. Обобщенный метод наименьших квадратов (GLS)

Предположение 1.4 (о сферичности возмущений ошибок) утверждает, что матрица Σ размера $n \times n$ условных вторых моментов $E(\varepsilon\varepsilon'|X) (= \text{Var}(\varepsilon|X))$ пропорциональна единичной матрице. Без этого предположения каждый элемент матрицы $n \times n$ является в общем случае нелинейной функцией от X . Если ошибки (условно) гомоскедастичны, значения диагональных элементов $E(\varepsilon\varepsilon'|X)$ равны, и если присутствует корреляция ошибок между наблюдениями (случай серийной корреляции для моделей временных рядов), то значения недиагональных элементов не равны нулю. Для любого заданного положительного скаляра σ^2 определим $V(X) = E(\varepsilon\varepsilon'|X)/\sigma^2$ и предположим, что $V(X)$ невырожденная и известная матрица. То есть

$$E(\varepsilon\varepsilon'|X) = \sigma^2 V(X), \quad V(X) =$$

(1.6.6)

вырожденная и известная матрица. (1.6.1)

Причина проведенной декомпозиции $E(\varepsilon\varepsilon'|X)$ (на компоненту σ^2 , общую для всех элементов матрицы $E(\varepsilon\varepsilon'|X)$, и оставляющей часть $V(X)$) состоит в том, что нам не нужно знать значение σ^2 для нахождения эффективной оценки. Модель, которую мы получаем при замене предположения 1.4 на (1.6.1), которое лишь предполагает невырожденность матрицы условных вторых моментов $E(\varepsilon\varepsilon'|X)$, называется **обобщенной регрессионной моделью** (*generalized regression model*).

Последствия ослабления предположения 1.4

Те из полученных в предыдущих параграфах результатов, которые используют предположение 1.4, не сохраняются для обобщенной регрессионной модели. А именно:

- Теорема Гаусса – Маркова больше не применяется для OLS оценки

$$b \equiv (X'X)^{-1}X'y.$$

BLUE теперь является другой оценкой

- t -критерий не более не имеет распределение Стьюдента. Значит, t -тест более неприменим. Те же замечания относятся и к F -тесту.
- Однако OLS-оценка все еще является несмещенной, так как для выведения этого свойства (утверждение 1.1(a)) предположение 1.4 не использовалось.

Эффективные оценивающие при известной матрице V

Если значение $V(X)$ известно, существует ли BLUE для обобщенной регрессионной модели? Ответ положительный, и такая оценка называется оценкой обобщенного метода наименьших квадратов (*generalized least squares GLS estimator*). Сейчас мы выведем эту оценку. Выбор будет построен на базе системы обобщенной линейной модели (которая состоит из предположений 1.1–1.3 и {1.6.1}) в модели классической линейной регрессии (удалив при этом все предположения, включая предположение 1.4).

Для упрощения обозначений далее будем использовать V вместо $V(X)$. Так как V — по построению симметричная и положительно определенная матрица, существует такая невырожденная матрица C размера $n \times n$, что

$$V^{-1} = C'C. \quad (1.6.2)$$

Эта декомпозиция неединственна (с более чем одним выбором матрицы C), однако, как будет очевидно ниже, выбор C не имеет никакой роли. Создадим теперь новую регрессионную модель для переменных, преобразуя (y, X, ε) с помощью матрицы C

$$\tilde{y} \equiv Cy, \quad \tilde{X} \equiv CX, \quad \tilde{\varepsilon} \equiv C\varepsilon. \quad (1.6.3)$$

Тогда предположение 1.1 о линейности модели для (y, X, ε) ведет к тому, что для преобразованных переменных $(\tilde{y}, \tilde{X}, \tilde{\varepsilon})$ свойство линейности сохраняется.

$$\tilde{y} = \tilde{X}\beta + \tilde{\varepsilon}. \quad (1.6.4)$$

Модель для преобразованных переменных удовлетворяет также и остальным предположениям классической линейной регрессионной модели. Требование строгой экзотенности выполнено, так как

$$E(\tilde{\varepsilon}|\tilde{X}) = E(\varepsilon|X)$$

(поскольку матрица C невырождена, X и $\tilde{X}$ содержат одну и ту же информацию)

$$= E(C\varepsilon|X) =$$

$$= C E(\varepsilon|X)$$

(в силу линейности усиленного математического ожидания)

$$= 0 \quad (\text{по оценке } E(\varepsilon|X) = 0 \text{ в предположении 1.2})$$

Так как V положительно определено, предположение об отсутствии мультиколлинеарности также выполнено (доказательство этого остается в качестве контрольного задания). Предположение 1.4 и удовлетворено для преобразованной модели

$$E(\tilde{\varepsilon}\tilde{\varepsilon}'|\tilde{X}) = E(\varepsilon\varepsilon'|X)$$

(так как $\tilde{X}$ и X содержат одинаковый объем информации)

$$= C E(\varepsilon\varepsilon'|X) C' \quad (\text{поскольку } \tilde{\varepsilon}\tilde{\varepsilon}' = C\varepsilon\varepsilon'C') =$$

$$= C \cdot \sigma^2 \cdot VC' \quad (\text{из (1.6.1)}) =$$

$$= \sigma^2 CVC' =$$

$$= \sigma^2 I_n$$

(так как $(C')^{-1}V^{-1}C^{-1} = I_n$ или $CVC' = I_n$ из (1.6.2))

То есть дисперсия преобразованного вектора ошибок $\tilde{\varepsilon}$ на всем поле сферична. Наконец, распределение $\tilde{\varepsilon}|\tilde{X}$ нормально, так как распределение $\tilde{\varepsilon}|\tilde{X}$ совпадает с распределением $\varepsilon|X$, а $\tilde{\varepsilon}$ является линейным преобразованием ε . На этом проверка выполнения предположений 1.1–1.5 для модели с преобразованными переменными закончена.

Из теоремы Гаусса — Маркова для модели с преобразованными переменными следует, что BLUE для β в обобщенной линейной регрессионной модели — это OLS-оценка, примененная к модели (1.6.4):

$$\begin{aligned} \hat{\beta}_{OLS} &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y} = \\ &= (CX)'(CX)^{-1}(CX)'Cy = \\ &= (X'C'CX)^{-1}(X'C'C)y = \\ &= (X'V^{-1}X)^{-1}X'V^{-1}y \quad (\text{из (1.6.2)}), \end{aligned} \quad (1.6.5)$$

Это GLS-оценка. Ее условная дисперсия равна:

$$\begin{aligned}\text{Var}(\tilde{\beta}_{\text{GLS}}|\mathbf{X}) &= \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\text{Var}(\mathbf{y}|\mathbf{X})\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}(\sigma^2\mathbf{V})\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \\ &\text{(так как } \text{Var}(\mathbf{y}|\mathbf{X}) = \text{Var}(\mathbf{e}|\mathbf{X})) \\ &= \sigma^2 \cdot (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}.\end{aligned}\quad (1.6.6)$$

Поскольку матрица $\mathbf{X}$ на $\sigma^2\mathbf{X}$ ($= \text{Var}(\mathbf{e}'\mathbf{X})$) в (1.6.5) не приводит к количественному изменению оценки, то GLS-оценку можно также представить в виде:

$$\tilde{\beta}_{\text{GLS}} = [\mathbf{X}'\text{Var}(\mathbf{e}'\mathbf{X})^{-1}\mathbf{X}]^{-1}\mathbf{X}'\text{Var}(\mathbf{e}|\mathbf{X})^{-1}\mathbf{y}.$$

Как уже отмечалось выше, OLS-оценка $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ является искаженной при отсутствии предположения 1.4, однако несмотря на это GLS-оценка (при известной $\mathbf{V}$) предпочтительнее, так как она является более эффективной в том смысле, что ее дисперсия меньше в матричном смысле. Выигрыш в эффективности достигается за счет учета гетероскедастичности и корреляции между наблюдениями в ошибках. На практике это означает, что надо вставить матрицу, обратную к $\text{Var}(\mathbf{e}|\mathbf{X})$ (или пропорциональную ей) в формулу для GLS-оценки, как это было сделано в (1.6.5). Это рассуждение можно обобщить следующим образом.

Утверждение 1.7 (свойства GLS в конечных выборках):

- (a) (Несмещенность.) При предположениях 1.1–1.3: $E(\tilde{\beta}_{\text{GLS}}|\mathbf{X}) = \beta$
 (b) (Формулы для дисперсий.) При предположениях 1.1–1.3 и предположении (1.6.1) о том, что условная дисперсия пропорциональна $\mathbf{V}(\mathbf{X})$,

$$\text{Var}(\tilde{\beta}_{\text{GLS}}|\mathbf{X}) = \sigma^2 \cdot (\mathbf{X}'\mathbf{V}(\mathbf{X})^{-1}\mathbf{X})^{-1}.$$

- (c) (Эффективность GLS.) При том же наборе предположений, что и в пункте (b), GLS-оценка является эффективной: условная дисперсия любой несмещенной оценки, основанной на $\mathbf{y}$, больше или равна $\text{Var}(\tilde{\beta}_{\text{GLS}}|\mathbf{X})$ в матричном смысле.

Частный случай: взвешенный метод наименьших квадратов (WLS)

Идея корректноности из ковариационную матрицу ошибок надбавляет измерения в случае, когда между ошибками нет корреляции (то есть матрица $\mathbf{V}$ диагональна). Обозначая через $v_i(\mathbf{X})$ i -й диагональный элемент матрицы $\mathbf{V}(\mathbf{X})$, тогда

$$E(\varepsilon_i^2|\mathbf{X}) (= \text{Var}(\varepsilon_i|\mathbf{X})) = \sigma^2 \cdot v_i(\mathbf{X}).$$

Несложно заметить, что матрица C также диагональна, причем ее i -й диагональный элемент равен квадратному корню из $1/\sigma_i^2(X)$. Таким образом, компоненты $(\tilde{y}, \tilde{X})$ равны:

$$\tilde{y}_i = \frac{y_i}{\sqrt{\sigma_i^2(X)}}, \quad \tilde{X}_i = \frac{x_i}{\sqrt{\sigma_i^2(X)}} \quad (i = 1, 2, \dots, n).$$

Поэтому эффективное оценивание при гетероскедастичности известной формы состоит в: ridании каждому наблюдению веса, обратного квадратному корню из дисперсии этого наблюдения $\sigma_i^2(X)$, и последующем применении OLS. Это называется **взвешенной регрессией**, или **взвешенным методом наименьших квадратов** (*weighted regression, weighted least squares* — **WLS**).

Другой важный частный случай — это случайная выборка, в которой $\{y_i, x_i\}$ i.i.d. по i . Как уже было замечено в параграфе 1.1, ошибки здесь безусловно гомоскедастичны (то есть $E(\varepsilon_i^2)$ не зависит от i), однако GLS все равно можно использовать для повышения эффективности оценки, поскольку ошибки могут быть условно гетероскедастичными. Условный момент второго порядка $E(\varepsilon_i^2|X)$ в ситуации случайной выборки зависит только от x_i , и функциональная форма $E(\varepsilon_i^2|X)$ одинакова для всех i . Таким образом,

$$\sigma_i^2(X) = v(x_i) \quad \text{для случайных выборок.} \quad (1.67)$$

То есть знание матрицы $V(\cdot)$ равносильно знанию одной функции K переменных, $v(\cdot)$.

Ограничивающее свойство GLS

Все сделанные ранее радужные выводы о свойствах GLS в конечных выборках основаны на предположении, что регрессоры в обобщенной регрессионной модели строго экзогенны ($E(\tilde{\varepsilon}|X) = 0$). Это ограничивает сферу применения GLS. Предположим, как в случае с моделями временных рядов, что регрессоры не удовлетворяют свойству строгой экзогенности и ошибки серьезно коррелированы. Тогда ни OLS, ни GLS-оценки не обладают хорошими свойствами, такими как несмещенность. Однако, как будет показано в следующей главе, OLS-оценка, игнорирующая наличие сериальной корреляции в ошибках, будет обладать рядом хороших свойств в больших выборках (такими как «состоятельность» и «асимптотическая нормальность»), при условии, что регрессоры являются «предопределенными» (это требование слабее, чем строгая экзогенность). GLS-оценка, напротив, не обладает такими подкупающими свойствами. То есть, если регрессоры не строго экзогенны, ни являются предопределенными, GLS-процедура, корректирующая на сериальную корреляцию, может дать неостоятельную оценку (см. параграф 6.7).

Процедура, позволяющая учесть сериальную корреляцию и при этом сохранить состоятельность, будет представлена в главе 6.

Несмотря на то что GLS-процедура применима для коррекции гетероскедастичности, она все равно может быть использована для коррекции гетероскедастичности, если ошибки сериально не коррелированы с диагональной матрицей $V(X)$ (в форме WLS). Однако это только при условии, что матрица $V(X)$ известна. Очевидно, редко в каком распоряжении имеется априорная информация относительно значений диагональных элементов $V(X)$, которые необходимы для взвешивания наблюдений. Имея дело со случайной выборкой, в которой не может возникнуть сериальная корреляция, знание $V(X)$ сводится, как мы только что видели, к знанию одной функции K переменных, $\sigma^2(x)$. Однако даже в этом случае такая функция неизвестна в большинстве приложений.

Если мы не знаем $V(X)$, мы можем считать ее функциональную форму по имеющейся выборке. Такой подход называется **доступным обобщенным методом наименьших квадратов** (*Feasible Generalized Least Squares* — **FGLS**). Однако если $V(X)$ оценивается по выборке, то ее значение V становится случайной величиной, что влияет на распределение GLS-оценки. Для малых выборок о свойствах FGLS известно немного. Асимптотические свойства FGLS будут рассмотрены в следующей главе.

В заключение стоит отметить одну положительную сторону GLS: большинство линейных процедур оценивания (включая 2SLS, 3SLS и оценку в модели со случайными эффектами, которые будут введены позднее) можно представить как GLS-оценку при некотором способе формирования матрицы данных. Однако все эти оценки и OLS могут быть так же интерпретированы как оценки обобщенного метода моментов (GMM), а GMM-интерпретация более полезна для вывода асимптотических свойств.

Контрольные вопросы

1. (Подтверждение об отсутствии мультиколлинеарности в модели с преобразованными переменными) предположим 1.3 для модели с преобразованными переменными состоит в том, что $\text{rank}(C'X) = K$. Это условие выполнено, так как C' — порожденная матрица, а X — матрица полного столбцового ранга. Покажите это. **Указание:** Так как X — матрица полного столбцового ранга, то для любого K -мерного вектора $c \neq 0$, $Xc \neq 0$.
2. (Обобщенная сумма квадратов остатков.) Покажите, что при $\hat{\beta}_{GLS}$ достигается минимум $(y - X\hat{\beta})'V^{-1}(y - X\hat{\beta})$.
3. Выведите выражение для $\text{Var}(b|X)$ для обобщенной регрессионной модели. Как оно соотносится с $\text{Var}(\hat{\beta}_{OLS}|X)$? Проверьте, что из утвержде-

ния 1.7(c) следует, что

$$(X'X)^{-1}X'VX(X'X)^{-1} \geq (X'V^{-1}X)^{-1}$$

4. (Ошибка GLS.) Покажите, что $\hat{\beta}_{GLS} - \beta = (X'V^{-1}X)^{-1}X'V^{-1}e$.

1.7. Приложение: отдача от масштаба в производстве и распределении электроэнергии

Работа Нерлова [Nerlove, 1963] представляет собой классический пример изучения отдачи от масштаба в регулируемой отрасли. Она также идеальна для применения методов, изученных в этой главе, и ознакомления читателя с некоторыми другими, еще не рассматривавшимися.

Отрасль производства и распределения электроэнергии

В то время, когда Нерлов писал свою статью, отрасль производства и распределения электроэнергии в США характеризовалась следующими особенностями:

- (1) Местные частные монополии поставляли электроэнергию по заявкам.
- (2) Тарифы (цена электроэнергии) устанавливались Комиссией по коммунальному обслуживанию.
- (3) Цены факторов производства (например, ставка заработной платы) устанавливались для фирмы либо на совершенно конкурентном рынке, либо посредством долгосрочных контрактов с профсоюзами.

Перечисленные особенности будут важны для нас при обсуждении вопроса, правомерно ли использование в таких условиях OLS¹.

Данные

Нерлов использовал кросс-секционные данные по 145 фирмам в 44 штатах за 1955 год. В модели были использованы следующие переменные: общие издержки, цены факторов (зарплата, цена топлива, стоимость аренды капитала) и выпуск. Хотя фирмы обладают своим капиталом (например, электростанциями, оборудованием, сооружениями), в соответствии со стандартной теорией издержек [Jorgenson, 1963], фирма должна вести себя (при предположении, что издержек изменения размера

¹Подобствие дерегулирования отрасли, произошедшего после выхода статьи Нерлова, на одном местном рынке было разрешено конкурировать нескольким фирмам, и строгий контроль за ценами был отменен во многих штатах. То есть, первые две особенности в настоящее время не имеют места.

капитала (δ) так, как будто она арендует капитал у себя же по рыночной ставке (цена называется «стоимостью использования капитала»), которая равна $(1 + \delta) \cdot r$, где r — реальная ставка процента (далее мы будем использовать букву r для обозначения отдачи от масштаба). δ — норма амортизации, p_0 — цена капитала. По этой причине к капиталу можно относиться так, как будто его количество фирма может легко изменить (так же как количество используемого труда и сырья).

Приложение В статьи [Кетович, 1963] содержит тщательное обсуждение истинности данных. Данные по расходам на материалы и на труд (которые в сумме с затратами капитала составляют общие затраты) были получены от Федеральной комиссии по энергетике (1956). В качестве ставки заработной платы использовались средние по штату заработная плата работников коммунальных компаний. В идеале стоимость капитала рассчитывается как произведение восстановительной стоимости на стоимость использования. Однако из-за отсутствия соответствующих данных этот показатель рассчитывался как сумма процентных платежей и амортизационных отчислений, которые были получены из бухгалтерских отчетов.

Для чего нам нужна эконометрика

Зачем нам использовать мощные эконометрические модели (например, OLS) для установления отдачи от масштаба? Почему мы не можем быть довольны и просто построить график зависимости средних издержек (которые могут быть легко вычислены из данных как отношение общей суммы затрат к выпуску) от выпуска в зависимости от выпуска — издержки и, наоборот, является ли наклон кривой АС («средних издержек», от англ. *average cost*) отрицательным? Причина состоит в том, что фирмы имеют разные кривые АС. Если цены факторов производства не однозначны для разных фирм, то для фирм с идентичными ценами факторов средние издержки будут ниже. Предположим, в том, что все фирмы в выборке сталкиваются с одинаковыми ценами на факторы производства в заданный момент времени, обычно является адекватным предположением, но не в случае с электроэнергетикой США, где цены на ресурсы значительно варьируются от региона к региону. Поэтому нужно каким-то способом изолировать влияние цен факторов на кривую АС. Подход, который использовался см Нерлов, ставший привычным в прикладных эконометрических исследованиях, состоит в описании параметризованной функции издержек.

Другую фактором, сдвигающим индивидуальную кривую АС, является уровень эффективности производства. Если более эффективные фирмы выпускают большие продукты, то возможно, что индивидуальные кривые АС имеют положительный наклон, в то время как кривая, соеди-

няющая наибольшее комбинации средних издержек и выпуска, имеет отрицательный наклон. В качестве иллюстрации этого рассмотрим конкурентную отрасль, изображенную на рис. 1.6, где кривые AC и MC (предельные издержки, от англ. *marginal cost*) изображены для двух фирм, работающих на одном рынке. Чтобы сконцентрироваться на связи эффективности и выпуска, предположим, что цены факторов производства для всех фирм одинаковы. То есть единственная причина различия кривых AC и MC двух фирм — это различие в эффективности производства. AC и MC имеют положительные наклоны, что отражает убывающую отдачу от масштаба. Кривые AC и MC для фирмы A лежат выше кривых фирмы B , так как фирма A менее эффективна, чем B . Поскольку отрасль совершенно конкурентна, обе фирмы продают свой выпуск по одной и той же цене p . Так как выпуск фирмы определяется точкой пересечения кривой MC и рыночной цены, то выпуски и средние издержки двух фирм соответствуют на графике точкам A и B . Кривая, которую мы получим, если соединим эти две точки, имеет отрицательный наклон, давая нам ложную иллюзию *положительной* отдачи от масштаба.

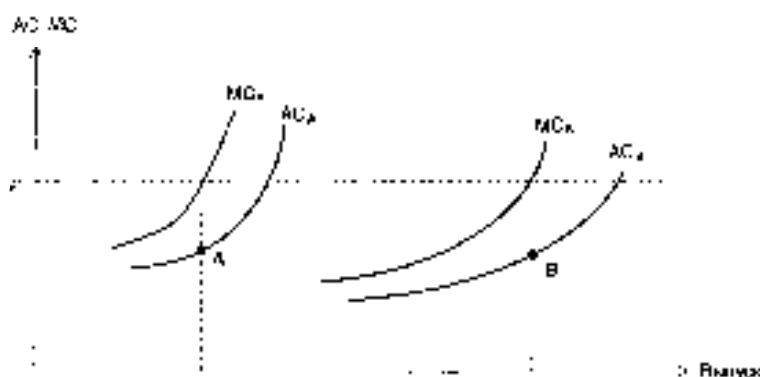


Рис. 1.6. Определение выпуска

Технология Кобба—Дугласа

Чтобы вывести ассиметричную функцию издержек, начнем с произвольной функции Кобба—Дугласа:

$$Q_i = A_i x_{i1}^{\alpha_1} x_{i2}^{\alpha_2} x_{i3}^{\alpha_3}, \quad (1.7.1)$$

где Q_i — выпуск фирмы i , x_{i1} — количество труда, используемое фирмой i , x_{i2} — капитал, x_{i3} — топливо. Параметр A_i отражает незначительные различия в эффективности производства (это часть называется *гетерогенностью фирм* (*firm heterogeneity*)). Сумма коэффициентов $\alpha_1 + \alpha_2 + \alpha_3 = \gamma$ характеризует отдачу от масштаба. Такая

образом, мы *a priori* предполагаем, что отдала от масштаба постоянна (это не следует путать с постоянной отдалой от масштаба, что означает, что $\gamma = 1$). Так как в выборке фирмы по производству электроэнергии являются частными, логично предположить, что каждая из них решает задачу минимизации издержек (посмотрите, однако, обсуждение в конце этого параграфа). Из микроэкономики нам известно, что функция издержек, соответствующая производственной функции Кобба — Дугласа, также имеет форму Кобба — Дугласа:

$$TC_i = c \cdot (A_i \alpha_1^{\alpha_1} \alpha_2^{\alpha_2} \alpha_3^{\alpha_3})^{-1/\gamma} Q_i^{1/\gamma} p_{01}^{\alpha_1/\gamma} p_{02}^{\alpha_2/\gamma} p_{03}^{\alpha_3/\gamma}, \quad (1.7.2)$$

где TC_i — общие издержки фирмы i . Беря логарифм от обеих частей, получаем линейное в логарифмах соотношение:

$$\log(TC_i) = \mu + \frac{1}{\gamma} \log(Q_i) + \frac{\alpha_1}{\gamma} \log(p_{01}) + \frac{\alpha_2}{\gamma} \log(p_{02}) + \frac{\alpha_3}{\gamma} \log(p_{03}), \quad (1.7.3)$$

где $\mu = \log[c \cdot (A_i \alpha_1^{\alpha_1} \alpha_2^{\alpha_2} \alpha_3^{\alpha_3})^{-1/\gamma}]$. Это уравнение называется **логдлинным** (*log-dlinear*), так как и зависимая, и независимые переменные стоят в логарифмах. Коэффициенты в линейном уравнении интерпретируются как эластичности. Например, коэффициенты при $\log(p_{0j})$ показывают эластичность общих издержек по ставке заработной платы, то есть на сколько процентов изменятся общие издержки при росте заработной платы на 1 процент. Отдала от масштаба, которая в уравнении (1.7.3) дана, эластично, обратной эластичности выпуска по общим издержкам, не зависит от уровня выпуска.

Пусть $\alpha \equiv E(\alpha_j)$ и отдала от масштаба γ , а $\mu_j = p_{0j}$, так что $E(\mu_j) = 0$. Этот α_j представляет собой величину, обратную к отдала от масштаба эффективности фирмы к средней эффективности отрасли. Фирмы с положительными α_j менее эффективны. Применяя эти обозначения, переищем (1.7.3) как

$$\log(TC_i) = \beta_1 + \beta_2 \log(Q_i) + \beta_3 \log(p_{01}) + \beta_4 \log(p_{02}) + \beta_5 \log(p_{03}) + \varepsilon_i, \quad (1.7.4)$$

где

$$\beta_1 = \mu, \quad \beta_2 = \frac{1}{\gamma}, \quad \beta_3 = \frac{\alpha_1}{\gamma}, \quad \beta_4 = \frac{\alpha_2}{\gamma} \text{ и } \beta_5 = \frac{\alpha_3}{\gamma}. \quad (1.7.5)$$

Таким образом, мы преобразовали функцию издержек к формату регрессии из предположения 1.1 с $K = 5$. Мы уже отметили, что важный метод построения графика зависимости средних издержек от выпуска не может признать отрицательное влияние цен факторов. Что мы только что доказали, — то, что, если в случае производственной функции Кобба — Дугласа влияние цен факторов принимается во внимание посредством включения в функцию издержек логарифмов цен. Так как уравнение

выведено из явного предположения о функциональном виде технологической ф-ки. Ошибки и коэффициенты регрессии имеют ясную экономическую интерпретацию.

Опять же мы знаем, что действительно имеем дело с технологией Кобба — Дугласа?

Технология Кобба — Дугласа, конечно, является удобным способом параметризации произвольной ф-ки. Но откуда мы знаем, что истинной произвольной ф-кой является именно та произвольная ф-ка Кобба — Дугласа? Фирма Кобба — Дугласа удивительно редко ошибается (например, дает убывающую предельную производительность), темпозатраты которых мы обычно требуем от произвольной ф-ки. Что же такое ф-ка Кобба — Дугласа? Почему придет на ум другие произвольные ф-ки? Однако ф-ка Кобба — Дугласа, как мы уже отметили, проста, на единицы хорошо нормирована сама. Статьи Нерлова входят в то же самое число тех, в которых «описанием» спецификация ф-ки Кобба — Дугласа признается недостаточной. Однако этот факт лишь подчеркивает важность ф-ки Кобба — Дугласа как ориентира, который дает возможность обдумывать подходящие обобщения.

Выполнены ли предположения OLS?

Чтобы обосновать возможность применения метода наименьших квадратов, необходимо удостовериться, что для уравнения (1.7.4) выполнены предположения 1.2–1.4. Очевидно, что предположение 1.1 (линейность) выполнено.

$$y_i = \log(Y_i) = \alpha_0 + (1, \log(Q_i), \log(r_{1i}), \log(r_{2i}), \log(r_{3i}))' \beta.$$

Нет никаких оснований предполагать, что регрессоры в (1.7.4) строго мультиколлинеарны. Кроме того, в данных Нерлова $\text{rank}(X) = 5$ (табл. 1.4), следовательно, предположение 1.3 (об отсутствии мультиколлинеарности) также выполняется.

При проверке предположения о строгой экзогенности регрессоров (предположение 1.2) актуальны особенности рынка электроэнергии, описанные выше. Можно предположить, что, как и в большинстве кросс-секционных данных, α_i не зависит от i , для $i \neq j$. Но остается вопрос: зависит ли α_i от i ? Если нет, то $E\alpha_i(X) = 0$. В соответствии с третьей особенностью рынка цены факторов не зависят от эффективности фирмы, поэтому можно предположить, что цены зависят от ошибок.

А что насчет выпуска? Так как фирмы отпускают электроэнергию по заявкам (первая особенность рынка), выпуск зависит от цены электроэнергии, которая (в свою очередь) устанавливается Комиссией по цен

мультиному обслуживанию (вторая особенность). Если регулятор устанавливает цену без учета эффективности той или иной фирмы, то тогда $\log(Q_i) \in \epsilon$, независимо. С другой стороны, если цена устанавливается таким образом, чтобы она покрывала средние издержки, то тогда эффективность фирм влияет на выпуск через влияние цены электричества на спрос. В таком случае выпуск эндогенен (коррелирован с ошибкой). В конце мы очень кратко вернемся к этому вопросу, а до этого будем игнорировать возможную эндогенность выпуска. Это нельзя было бы сделать, если бы мы имели дело с совершенно конкурентной отраслью. Так как фирмы с высокими издержками производят меньше, наблюдалось бы отрицательная корреляция между $\log(Q_i)$ и ϵ_i , что сделало бы OLS неприемлемым.

Что касается предположения 1.4, то предположение об отсутствии корреляции между ошибками разных фирм (наблюдений) не выполняется бы, если бы, например, фирмы перенимали технологические новшества у других фирм, расположенных неподалеку. Для рассматриваемого примера этого эффекта, скорее всего, нет.

Не существует априорной причины считать, что условие гомосkedастичности выполнено. Кроме того, график остатков, приведенный ниже, свидетельствует в пользу того, что это предположение не выполнено. Большая часть работы Нерлова посвящена рассмотрению способов преодоления этой проблемы.

Метод наименьших квадратов с ограничениями

Уравнение (1.7.4) сверхидентифицировано (*overidentified*) в том смысле, что, хотя это коэффициенты, будучи функцией от четырех параметров технологии ($\alpha_1, \alpha_2, \alpha_3$ и β), не являются свободными параметрами. Из (1.7.5) легко видеть, что $\beta_1 = \alpha_2 + \beta$, $\beta_2 = 1$ (вспомогательное: $\epsilon = \alpha_1 + \alpha_2 + \alpha_3 + 1$). Это является отражением общего свойства функции издержек: линейность и однородность по факторам производства. Кроме того, умножение общих издержек TC и цен факторов производства (p_1, p_2, p_3) на одну и ту же величину не приводит к изменению функции издержек (1.7.4), если и только если $\beta_1 + \beta_2 + \beta_3 = 1$.

Оценивание уравнения методом наименьших квадратов при наличии ограничений ограничивается только коэффициентом пересчета каждого из данных квадратов с ограничениями. Это может быть сделано репарировано путем получения из исходного уравнения регрессии другого уравнения регрессии, в котором учтены ограничения. В нашем примере чтобы ввести ограничение однородности $\beta_1 + \beta_2 + \beta_3 = 1$ на функцию издержек, нужно взять любую из цен факторов производства (например,

¹ См. табл. 1.7.

p_{13}) и вычлеть $\log(p_{13})$ из обеих частей (1.7.4). В результате получаем:

$$\log\left(\frac{TC_i}{p_{13}}\right) = \beta_1 + \beta_2 \log(Q_i) + \beta_3 \log\left(\frac{p_{11}}{p_{13}}\right) + \beta_4 \log\left(\frac{p_{12}}{p_{13}}\right) + \varepsilon_i. \quad (1.7.6)$$

Теперь в регрессии осталось четыре коэффициента, на основании которых можно восстановить значения четырех параметров технологии. Оценка OLS с ограничениями коэффициентов $(\beta_1, \dots, \beta_4)$ представляет собой обычную OLS-оценку коэффициентов уравнения (1.7.6). Оценка OLS с ограниченными коэффициентами β_3 вычисляется исходя из оценок $(\beta_1, \dots, \beta_4)$ и указанного ограничения.

Тестирование однородности функции издержек

Прежде чем переходить к оцениванию модели с ограничением (1.7.6), протестируем ограничение однородности производственной функции $\beta_3 + \beta_4 + \beta_5 = 1$. Для этого сначала оценим модель без ограничений (1.7.4). При использовании данных из статьи Нерлова, которые доступны в печатном виде, получаем следующую OLS-оценку.

$$\begin{aligned} \log(TC_i) = & -3.5 + 0.72 \log(Q_i) + 0.44 \log(p_{11}) \\ & (1.8) \quad (0.017) \quad (0.29) \\ & + 0.22 \log(p_{12}) + 0.13 \log(p_{13}) \\ & (0.31) \quad (0.10) \\ R^2 = & 0.926, \text{ среднее значение переменной} = 1.72, \\ SER = & 0.392, SSR = 21.352, n = 145. \end{aligned} \quad (1.7.7)$$

В скобках приведены стандартные ошибки OLS-оценок коэффициентов. Поскольку $\beta_5 = 1/r$, оценка отдачи от масштаба, рассчитанная на основании OLS-оценок коэффициента β_5 , приблизительно равна 1.4 (= $1/0.72$). OLS-оценка β_3 = α_2/r имеет неправильный знак. Как отметил Нерлов, это может быть следствием того, что цена использования капитала p_{12} завышена плохо. Это можно объяснить, потому что коэффициент b_1 определен столь точно (т.е. есть или стандартная ошибка эмпирика по отношению к оценке коэффициента), что гипотеза о его равенстве нулю $\beta_3 = 0$ не отвергается со значением t -отношения, равным $-0.65 (= -0.22/0.34)^1$.

Чтобы проверить гипотезу однородности $H_0: \beta_3 + \beta_4 + \beta_5 = 1$, мы можем переписать эту гипотезу в форме $R\beta = r$, где $R = (0, 0, 1, 1, 1)$, $r = 1$, а затем использовать формулу (1.4.9) для вычисления F -отношения. Поддерживаемой гипотезой является в данном случае модель

¹ Следствием ошибки в измерениях является то, что коэффициент при переменной, измеряемой с ошибкой, плохо определен, но это может повлиять на оценку коэффициента при других переменных. В этом смысле данная проблема будет обсуждаться в контексте асимптотической теории для заданных регрессоров в главе 3.

без ограничений (1.4.9) (то есть предположения 1.1–1.3, причем уравнение в предположении 1.1 имеет спецификацию (1.7.4)). То есть оценки коэффициентов β и оценки дисперсии $\hat{\sigma}^2$ в F -отношении должны быть получены из OLS-оценивания уравнения (1.7.4). Однако можно пойти другим путем: взять формулу (1.4.11) для F -отношения. Модель без ограничений, из которой берется $SSR_{\text{огн}}$ — это (1.7.4), а модель с ограничениями (которая накладывает ограничения нулевой гипотезы на модель без ограничений), из которой получаем $SSR_{\text{огн}}$ — это модель (1.7.6). Ниже представлена OLS-оценка (1.7.6).

$$\begin{aligned}\log\left(\frac{TC_i}{P_{\text{ав}}}\right) &= -4.7 + 0.72 \log(Q_i) \\ &\quad (0.88) \quad (0.017) \\ &\quad - 0.59 \log(p_{\text{м}}/p_{\text{ав}}) - 0.007 \log(p_{\text{л}}/p_{\text{ав}}) \\ &\quad (0.20) \quad (0.19) \\ R^2 &= 0.912, \text{ среднее зависимой переменной} = -1.48, \\ SER &= 0.49, SSR = 21.640, n = 145.\end{aligned}\quad (1.7.8)$$

F -тест для гипотезы однородности выполняется следующим образом.

Шаг 1: Используя (1.4.11), можно вычислить F -отношение:

$$\frac{(21.640 - 21.552)/1}{21.552/(145 - 5)} = 0.57.$$

Шаг 2: Находим критическое значение. Количество ограничений (уравнений) накладываемых нулевой гипотезой, равно 1, а K (число коэффициентов) в модели без ограничений (которая является поддерживаемой гипотезой) равно 5. То есть нас интересует F -распределение со степенями свободы 1 и 140 (= 145 – 5). Из соответствующей таблицы находим, что критическое значение равно приблизительно 3.9.

Шаг 3: Таким образом, нулевая гипотеза однородности не отвергается. Это обнадеживающее заключение для тех, кто всерьез относится к микроэкономической теории (подобно нам).

Личное отступление: об осторожности с интерпретацией R^2

Значение R^2 , равное 0.926, не удивительно велико для кросс-секционных данных. Однако часть объясняющей силы, которую можно объяснить эффектом масштаба, обычно закрепил фирм растут с ростом размера фирмы. Чтобы оценить вклад этого эффекта в R^2 , вычтем $\log(Q_i)$

из обеих частей (1.7.4), получая эквивалентную функцию издержек:

$$\log\left(\frac{TC_i}{Q_i}\right) = \alpha_1 + (\beta_2 - 1)\log(Q_i) + \\ + \beta_3 \log(p_{H1}) + \beta_4 \log(p_{H2}) + \beta_5 \log(p_{H3}) + \varepsilon_i. \quad (1.7.4')$$

Теперь в качестве зависимой переменной выступают средние издержки, а не общие, как было ранее. Применение OLS к (1.7.4') даст для тех же данных следующие результаты:

$$\log\left(\frac{TC_i}{Q_i}\right) = -3.5 + 0.28 \log(Q_i) + \\ (3.5) \quad (0.017) \\ + 0.44 \log(p_{H1}) + 0.22 \log(p_{H2}) + 0.13 \log(p_{H3}) \\ (0.29) \quad (0.04) \quad (0.10) \\ R^2 = 0.691, \text{ среднее зависимой переменной} = 4.83, \\ SER = 0.392, SSR = 21.532, n = 145. \quad (1.7.5)$$

Как и в безразмерном варианте, коэффициент при выпуске теперь равен -0.28 (т.е. $0.72 - 1$), при этом оценки прочих коэффициентов и стандартных ошибок не изменились. R^2 изменился только из-за изменения зависимой переменной. Было бы странно утверждать, что более низкое значение коэффициента детерминации спецификации (1.7.4) предпочтительнее, чем (1.7.4'), так как эти две уравнения представляют одну и ту же модель. Мораль: при сравнении двух перекалиброванных на основе R^2 моделей, чтобы в обоих случаях зависимая переменная была одной и той же.

Тестирование постоянной ошибки от масштаба

В качестве примера применения t -теста рассмотрим тестирование гипотезы о постоянной ошибке от масштаба ($\alpha = 1$). Поддерживаемой гипотезой является модель (1.7.6). Поскольку α (коэффициент при логарифме выпуска) равен 1, если и только если $\alpha = 1$, то нулевая гипотеза может быть сформулирована как $H_0: \beta_2 = 1$. t -тест для проверки гипотезы о постоянстве ошибки от масштаба проводится следующим образом:

Шаг 1: Выпишем t -отношение для гипотезы. По оценкам модели с ограничениями имеем $b_2 = 0.72$ со стандартной ошибкой, равной 0.017, так что

$$t\text{-отношение} = \frac{0.72 - 1}{0.017} = -16.$$

Так как поддерживаемой гипотезой здесь является модель с ограничениями (1.7.6), K (количество коэффициентов) = 4

Шаг 2: Находим критическое значение для распределения Стьюдента ($t_{(51)}$). Размер теста равен 5%, критическое значение равно 1,98.

Шаг 3: Поскольку абсолютное значение t -отношения превышает критическое значение, мы отвергаем нулевую гипотезу о постоянной отдаче от масштаба.

Важность графического анализа остатков

У рассматриваемых уравнений есть проблема, и суммирование коэффициентов добавляет к анализу регрессионных коэффициентов и их стандартных ошибок. На рис. 1.7 изображены остатки модели (по горизонтальной оси отложена переменная $\log(Q_1)$).

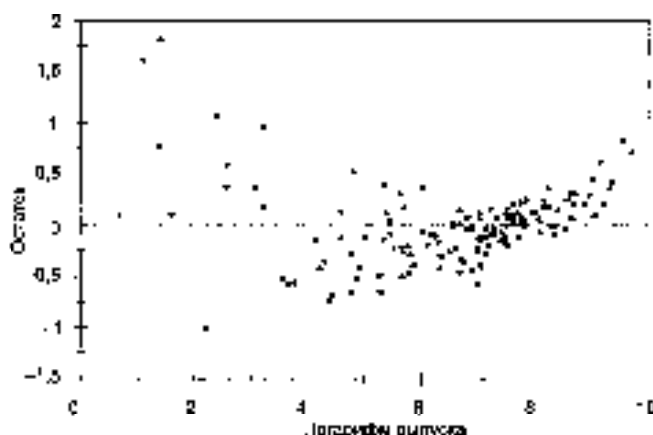


Рис. 1.7. График «логарифм выпуска — остатки»

Обратите внимание на два обстоятельства. Во-первых, с ростом объема выпуска остатки сначала имеют тенденцию быть положительными, затем отрицательными, потом вновь положительными. Это, скорее всего, свидетельствует о выходящей отдачи от масштаба (γ) непостоянна, что противоречит логической спецификации. Во-вторых, разброс остатков больше при больших значениях выпуска, что является признаком нарушения предположения о гомоскедастичности ошибок (дисперсия ошибок не зависит от регрессоров). Чтобы справиться с этими проблемами, Перлов разделил выборку из 145 фирм на 5 групп по 29 фирм (предварительно упорядочив фирмы по объему выпуска) и оценил уравнение (1.7б) отдельно для каждой группы. Эти индивидуальные коэффициенты (включая $\beta_2 = 1/\gamma$) и дисперсии ошибок были различны для пяти разных групп фирм (отличающихся по размеру). Перлов установил, что отдача от масштаба постепенно уменьшается от значения γ , значительно

большее 2, до значения, немного меньшего 1) по мере роста выпуска. В эмпирическом упражнении к этой главе читателям будет предложено повторить эти результаты и провести дополнительный анализ, используя *даты-переменные* (*dummy variables*) и известный метод наименьших квадратов.

Дальнейшая разработка темы

Одним из направлений в последующих исследованиях является обобщение технологии Кобба — Дугласа при сохранении предположения о минимизации издержек. Очевидной альтернативой рассмотренной функции Кобба — Дугласа является производственная **функция с постоянной эластичностью замещения факторов** (*Constant Elasticity of Substitution — CES*). Однако ее использование сопряжено с двумя трудностями. Во-первых, функция издержек, получаемая при CES-спецификации производственной функции, нелинейна. Эту проблему можно решить, используя нелинейный метод наименьших квадратов, представленный в главе 7. Во-вторых, из CES-спецификации производственной функции следует постоянная отдача от масштаба. Одним из главных выводов в работе Нерлова было то, что отдача от масштаба изменяется при изменении объема выпуска. [Christensen and Greene, 1976], возможно, были правы, что оценивал параметр технологии в предположении *изменяющейся* отдачи от масштаба. Используя **транслогарифмическую функцию издержек** (*translog cost function*), предложенную [Christensen, Jorgensen and Lau, 1973], они выяснили, что значимый эффект масштаба, очевидный для данных за 1953 год, перестал проявляться к 1970 году, когда большинство фирм производили намного больший объем выпуска при практически плоских кривых AC. Мы обсудим эту работу более подробно в главе 4.

Другой вопрос: действительно ли фирмы в регулируемой отрасли минимизируют издержки? Известная статья [Averch and Johnson, 1962] утверждает, что практика регулирующих органов устанавливать тарифы коммунальных предприятий таким образом, чтобы гарантировать им «справедливую норму доходности» их капитала, нарушает выбор оптимальных уровней. Так как справедливая норма доходности обычно выше рыночной ставки процента, то у коммунальных предприятий возникает стимул к повышению уровня инвестиций выше оптимального. Иными словами, фирмы минимизируют издержки, однако норма доходности в определении стоимости использования капитала — это справедливая норма доходности. Следовательно, истинный параметр технологии не может быть определен из функции издержек (если, конечно, не использовать справедливую норму доходности для расчета r_d). Однако такой способ назначения тарифов создает еще одну проблему, чтобы обеспечить фир-

мак, справедливую норму доходности, цена на электроэнергию должна быть выше на рынках, где работают предприятия с большими издержками. Полную выработку будет эндогенным.

Наконец, более свежее направление исследований — это изучение вопроса: действительно ли регулятор обладает достаточным объемом информации, чтобы его решение привело к максимизации издержек. Если фирма располагает большим объемом информации об издержках, у нее есть стимулы сообщить регулятору неверное знание своего параметра эффективности. Схемы, которые может использовать регулятор, чтобы предотвратить эту проблему, связанную со стимулированием, могут не привести к максимизации издержек. Работа [Wolck, 1994], посвященная предприятиям коммунального водоснабжения в Калифорнии, указывает, что наблюдаемый уровень издержек и выпуска лучше моделируется как результат взаимодействия между регулятором и фирмой при асимметричной информации. Волк решает проблему эндогенности выпуска путем оптимизации функции спроса одновременно с функцией издержек. Однако для этого необходима более изощренная техника, чем метод наименьших квадратов.

Контрольные вопросы

1. (Повторение теории двойственности.) Откройте ваш любимый учебник по микроэкономике и вспомните, как выводится функция издержек Кобба — Дугласа из производственной функции Кобба — Дугласа.
2. (Изменение единиц измерения.) В работе Нерова выпуск измеряется в киловатт-часах. Как изменятся оценки коэффициентов в уравнении, если выпуск будет измеряться в мегаватт-часах?
3. (Восстановление параметров технологий на основе регрессионных коэффициентов.) Покажите, что параметры производственной функции (1.7.4) могут быть однозначно определены из систем систем уравнений в (1.7.5) и определения $\bar{r} \equiv \alpha_1 + \alpha_2 + \alpha_3$. (Не используйте пятое уравнение $\hat{\beta}_3 = \alpha_3/\bar{r}$.)
4. (Восстановление коэффициентов, полученных из модели с ограничениями.) Вычислите OLS-оценку коэффициента α_2 из (1.7). Как вычислить стандартную ошибку $\hat{\sigma}_2$ на основании метода ланграна с ограничениями? Указание: Запишите $\hat{\beta}_2 = a + b\hat{\beta}_3$ для специального набора выбранных a и b , где $b = (\beta_1, \dots, \beta_4)'$ так что $\text{Var}(b; X) = \sigma^2 \text{Var}(b; X)$. Таблица результатов OLS-оценки с ограничениями должна включать $\text{Var}(b; X)$.
5. Если вы знаете $\hat{\beta}_2$ вместо $\hat{\beta}_3$ и хотите избежать $\log(\hat{\beta}_2)$ (1.7.4) $\log(\hat{\beta}_2)$, как будет выглядеть регрессия с ограничениями в таком случае? Без полного описания этой регрессии на данных Нерова, используя только результат о значении регрессии с ограничениями в тексте параграфа, можете ли вы сказать, чему будут равны оценки коэффициентов $\{\hat{\beta}_1, \dots, \hat{\beta}_4\}$, полученные OLS с ограничениями? Чему равны их стандартные ошибки? SSR? R^2 ?

6. Почему R^2 в модели (1.7.7), равный 0,926, меньше, чем в модели (1.7) (где $R^2 = 0,932$)?
7. Более реалистичным предположением о цене капитала может быть существование единого рынка капитала на территории всей страны, так что r_2 одинакова для всех фирм. В таком случае
- (а) Можем ли мы оценить параметры технологий? **Указание:** Ответ: да. Но почему? Когда r_2 постоянна, в (1.7.4) возникает проблема полной мультиколлинеарности. Но вспомним, что $(\beta_1, \dots, \beta_4)$ не являются свободными параметрами.
- (б) Можем ли мы проверить гипотезу об однородности функции издержек по ценам факторов?
8. Беря логарифмы от обеих частей производственной функции (1.7.1), можно вывести линейное в логарифмах соотношение:

$$\log(q_2) = \alpha_0 + \alpha_1 \log x_{21} + \alpha_2 \log x_{22} + \alpha_3 \log x_{23} + \varepsilon_2,$$

где ε_2 определяется как $\log(A_2) - E[\log(A_2)]$ и $\alpha_0 = E[\log(A_2)]$. Предполагая, что в отношении к данным об объемах издержек, выпуске и ценах факторов производства в нашей распоряжении есть также данные о количестве использованных факторов производства. Можем ли мы оценить коэффициенты в логлинейном соотношении с помощью метода наименьших квадратов? Почему да или почему нет? **Указание:** Зависит ли уровень факторов производства от ε_2 ?

Предложите альтернативный способ оценивания коэффициентов α . **Указание:** Посмотрите на доли факторов производства

Набор задач для главы 1

Аналитические упражнения

1. (Доказательство того, что b минимизирует SSR .) Обозначим через b OLS-оценку β . Докажите, что для любой гипотетической оценки вектора $\tilde{\beta}$, $\tilde{\beta}$ выполнено

$$(y - X\tilde{\beta})'(y - X\tilde{\beta}) \geq (y - Xb)'(y - Xb).$$

При доказательстве используйте стратегию прибавления — вычитания: возьмите разность $y - X\tilde{\beta}$, прибавьте к ней, а затем вычтите Xb . Тогда получится следующее разложение $y - X\tilde{\beta}$:

$$y - X\tilde{\beta} = (y - Xb) + (Xb - X\tilde{\beta}).$$

Указание: $(y - X\tilde{\beta})'(y - X\tilde{\beta}) = [(y - Xb) + X(b - \tilde{\beta})]'[(y - Xb) + X(b - \tilde{\beta})]$. Используя систему нормальных уравнений, покажите, что это равно:

$$(y - Xb)'(y - Xb) - (b - \tilde{\beta})'X'X(b - \tilde{\beta})$$

2. (Аннулятор для вектора, состоящего из единиц.) Обозначим через $\mathbf{1}$ n -кратный вектор-столбец, состоящий из единиц. Определим $M_1 := \mathbf{1}_n \cdot \mathbf{1}'_n (\mathbf{1}'_n \mathbf{1}_n)^{-1} \mathbf{1}'_n$. То есть M_1 — аннулятор для вектора $\mathbf{1}$. Докажите

- (a) Матрица M_1 симметричная и идемпотентная.
 (b) $M_1 \mathbf{1} = \mathbf{0}$.
 (c) $M_1 y = y - \bar{y} \cdot \mathbf{1}$, где

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

$M_1 y$ — вектор отклонений от среднего.

- (d) $M_1 X = X - \mathbf{1} \bar{X}'$, где $\bar{X} = X' \mathbf{1}_n / n$, k -й элемент вектора $\bar{X}$ размерности $K \times 1$ равен $\sum_{i=1}^n x_{ik}$.
3. (Модель в отклонениях от среднего.) Рассмотрим модель регрессии с константой. Пусть X разбивается на блоки следующим образом:

$$X_{(n \times K)} = \begin{bmatrix} \mathbf{1}_n & \cdots & X_1 \\ & & \vdots \\ & & X_K \end{bmatrix},$$

то есть первая регрессор — константа. Векторы β и b разбиваются соответственно:

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \leftarrow \begin{matrix} \text{скаляр} \\ (K-1) \times 1 \end{matrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

Обозначим $\tilde{X}_2 \equiv M_1 X$ и $\tilde{y} \equiv M_1 y$. Это отклонения от среднего для всех регрессоров, кроме константы, и зависимой переменной. Докажите

- (a) Система K скалярных уравнений имеет вид:

$$\tilde{y} - \mathbf{1}_n b_1 - \tilde{X}_2' b_2 = 0,$$

$$\text{где } \tilde{X}_2' = X_2' \mathbf{1}_n / n.$$

$$X_2' y - n \cdot b_1 x_0 = X_2' X_2 b_2 = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

- (b) $b_2 = (\tilde{X}_2' \tilde{X}_2)^{-1} \tilde{X}_2' \tilde{y}$. **Указание:** Подставьте первое уравнение системы в остальные $K-1$, чтобы избавиться от b_1 , и найдите b_2 . Это является обобщением результата, который вы доказали в контрольном вопросе 3 к параграфу 1.2.

4. (Регрессия с блочной структурой матрицы наблюдений, общей регрессии упражнения 3.) Пусть X разбита на блоки следующим образом

$$X = \begin{bmatrix} X_1 & X_2 \\ n \times k & n \times k_2 \end{bmatrix}.$$

Тогда соответствующим разложением β будет:

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} K_1 \times 1 \\ K_2 \times 1 \end{bmatrix}.$$

Поэтому регрессионную модель можно записать в виде:

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon.$$

Обозначим $M_1 \equiv X_1(X_1'X_1)^{-1}X_1'$, $M_2 \equiv I - M_1$, $\bar{X}_2 \equiv M_1X_2$ и $\bar{y} \equiv M_2y$. Таким образом, $\bar{y}$ представляет собой вектор остатков от регрессии y на X_1 , а k -й столбец $\bar{X}_2$ — вектор остатков от регрессии соответствующего k -го столбца матрицы X_2 на X_1 . Докажите:

- (a) система нормальных уравнений имеет вид:

$$X_1'X_1b_1 + X_1'X_2b_2 = X_1'y, \quad (*)$$

$$X_2'\bar{X}_2b_2 = X_2'\bar{y}. \quad (**)$$

- (b) $b_2 = (\bar{X}_2'\bar{X}_2)^{-1}\bar{X}_2'\bar{y}$. Таким образом, оценка b_2 может быть получена из регрессии остатков $\bar{y}$ на матрицу остатков $\bar{X}_2$. **Указание:** Вычленив $X_1b_1 = M_1X_2b_2 + M_1y$ из (*). Подставьте это выражение в (*), чтобы получить $X_2'M_2X_2b_2 = X_2'M_2y$. После этого используйте тот факт, что M_1 — симметричная и идемпотентная матрица. Иди, если понадобится, можете просто применить формулу (A.10) произведения A и матрицы из строк-векторов:

$$X'X = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}.$$

Покажите, что второй диагональный блок матрицы $(X'X)^{-1}$ имеет вид $(\bar{X}_2'\bar{X}_2)^{-1}$.

- (c) Остатки от регрессии $\bar{y}$ на $\bar{X}_2$ численно равны n остаткам от регрессии y на X ($\equiv (X_1'X_2)$). **Указание:** Если e — остаток от регрессии y на X , то

$$y = X_1b_1 + X_2b_2 + e$$

Умножив слева обе части на M_1 и используя $M_1 X_2 = 0$, получаем

$$\tilde{y} = \tilde{X}_2 b_2 + M_1 \epsilon.$$

Покажите, что $M_1 \epsilon = e$, а вы увидите, что b_2 равен OLS-оценке коэффициента в регрессии $\tilde{y}$ на $\tilde{X}_2$.

- (d) $b_2 = (\tilde{X}_2' \tilde{X}_2)^{-1} \tilde{X}_2' \tilde{y}$. Обратите внимание на отличие от пункта (b). Здесь в роли зависимой переменной выступает $\tilde{y}$, а не y . Совпадают ли численно остатки от регрессии y на $\tilde{X}_2$ с e ? [Ответ: нет.] Равна ли SSR регрессии y на $\tilde{X}_2$ сумме квадратов остатков регрессии $\tilde{y}$ на $\tilde{X}_2$? [Ответ: Нет.]

Результаты пунктов (c)–(d) носят название теоремы Фриша — Во — Ловелла.

- (e) Покажите:

$$\tilde{y}' \tilde{y} - e'e = \tilde{y}' X_2 (X_2' M_1 X_2)^{-1} X_2' \tilde{y}.$$

Указание: Примените общую формулу декомпозиции (1.2.15) к регрессии в пункте (c), чтобы получить

$$\tilde{y}' \tilde{y} = b_2' \tilde{X}_2' \tilde{X}_2 b_2 + e'e.$$

Затем используйте пункт (b).

- (f) Раскройте каждую из четырех регрессий:

- (1) $\tilde{y}$ на X_1 ,
- (2) $\tilde{y}$ на $\tilde{X}_2$,
- (3) $\tilde{y}$ на X_1 и $\tilde{X}_2$,
- (4) $\tilde{y}$ на X_2 .

Обозначим через SSR_1 сумму квадратов остатков в регрессии $\tilde{y}$. Покажите:

- (i) $SSR_1 = \tilde{y}' \tilde{y}$. Указание: $\tilde{y}$ построена таким образом, что $X_1' \tilde{y} = 0$, то есть X_1 не имеет объясняющей силы.
- (ii) $SSR_2 = e'e$. Указание: Используйте (d).
- (iii) $SSR_3 = e'e$. Указание: Примените теорему Фриша — Во — Ловелла к регрессии (b).

$$M_1 \tilde{y} = \tilde{y}.$$

- (iv) Проверьте на конкретном примере, что SSR_4 приближенно равно $e'e$.

- в) Регрессия с ограничениями и F-тест. При использовании метода наименьших квадратов с ограничениями суммарный остаток минимизируется при условии ограничений, накладываемых нулевой функцией $l(\tilde{\beta}) = r$. Запишем Лагранжиан:

$$L = \frac{1}{2}(y - X\tilde{\beta})'(y - X\tilde{\beta}) + \lambda'(R\tilde{\beta} - r),$$

где λ — k -мерный вектор множителей Лагранжа (вспомните: вектор R размера $k \times K$, $\tilde{\beta}$ размера $K \times 1$, r размера $k \times 1$). Обозначим через $\tilde{\beta}$ OLS-оценку параметра β в регрессии с ограничениями. Она является решением задачи минимизации с ограничениями.

- (г) Пусть b — OLS-оценка без ограничений. Покажите:

$$\begin{aligned}\tilde{\beta} &= b - (X')^{-1}R'[R(X'X)^{-1}R']^{-1}(Rb - r), \\ \lambda &= [R(X'X)^{-1}R']^{-1}(Rb - r).\end{aligned}$$

Указание: Условия первого порядка: $X'y = (X'X)\tilde{\beta} = R'\lambda$, или $X'(y - X\tilde{\beta}) = R'\lambda$. Добавьте в функцию лагранжиана $R'\tilde{\beta} = r$ и решите относительно λ и $\tilde{\beta}$.

- (д) Обозначим остаток от регрессии с ограничениями через $\tilde{e} \equiv y - X\tilde{\beta}$. Покажите:

$$\begin{aligned}SSR_0 - SSR_1 &= (b - \tilde{\beta})'(X'X)(b - \tilde{\beta}) = \\ &= (Rb - r)'[R(X'X)^{-1}R']^{-1}(Rb - r) = \\ &= \lambda'(X'X)^{-1}R'\lambda = \\ &= \tilde{e}'P\tilde{e},\end{aligned}$$

где P — проекционная матрица. Указание: Для получения первого равенства используйте структуру прибавления — вычитания.

$$\begin{aligned}SSR_0 &= (y - X\tilde{\beta})'(y - X\tilde{\beta}) = \\ &= (y - Xb) + X(b - \tilde{\beta})'[(y - Xb) + X(b - \tilde{\beta})].\end{aligned}$$

Используйте систему нормальных уравнений $X'(y - Xb) = 0$. Для второго и третьего равенств используйте (г). Для доказательства четвертого равенства самым простым путем является использование условия первого порядка, упомянутого в (г): $R'\lambda = X'\tilde{e}$.

- (е) Проверьте, что в пункте (д) вы доказали, что (1.4.9) = (1.4.11).

6. (Доказательство декомпозиции (1.2.17).) Пусть модель без ограничений — регрессия, в которой, помимо постоянной составляющей имеются и другие регрессоры, в модель с ограничениями — регрессия, в которой постоянная составляющая является единственным регрессором.

(а) Покажите, что (b) в предыдущем упражнении представляет собой декомпозицию (1.2.17) для этого случая. **Указание:** Числовая $\tilde{\beta}$ для этого случая? Покажите, что $SSR_D = \sum_i (y_i - \bar{y})^2$, и $(b - \tilde{\beta})'(X'X)(b - \tilde{\beta}) = \sum_i (\bar{y}_i - \bar{y})^2$.

(б) (R^2 и F -критерии.) Для модели, в которой константа является одним из регрессоров. Докажите, что

$$F = \frac{R^2/(K-1)}{(1-R^2)/(n-K)}$$

7. (Правило Хаусмана в панельных выборках.) Для обобщенной регрессионной модели докажите следующее. Здесь индексироваться будут математические ожидания, дисперсии и ковариации множителей случайных отклонений X .

(а) $\text{Cov}(\tilde{\beta}_{OLS}, b - \tilde{\beta}_{OLS}) = 0$. **Указание:** Вспомните, что для любых двух случайных векторов x и y

$$\text{Cov}(x, y) \equiv E[(x - E(x))(y - E(y))'].$$

Поэтому

$$\text{Cov}(Ax, By) = A \text{Cov}(x, y) B'.$$

Кроме того, так как $\tilde{\beta}$ несмещенная, то

$$\text{Cov}(\tilde{\beta}_{OLS}, b - \tilde{\beta}_{OLS}) = \text{Cov}(\tilde{\beta}_{OLS} - \beta, b - \tilde{\beta}_{OLS}).$$

(б) Пусть $\tilde{\beta}$ — линейная по y и несмещенная оценка, и определим $q \equiv \tilde{\beta} - \tilde{\beta}_{OLS}$. Пусть $\tilde{\beta}$ определена так, что матрица $V_q \equiv \text{Var}(q)$ невырождена. Докажите: $\text{Cov}(\tilde{\beta}_{OLS}, q) = 0$. (Если мы положим $\tilde{\beta} = b$, то мы вернемся к пункту (а).) **Указание:** Определим $\tilde{\beta} \equiv \tilde{\beta}_{OLS} + Hq$ для некоторой матрицы H . Покажите:

$$\text{Var}(\tilde{\beta}) = \text{Var}(\tilde{\beta}_{OLS}) + CH' + HC' + HV_qH',$$

где $C \equiv \text{Cov}(\tilde{\beta}_{OLS}, q)$. Покажите, что если $C \neq 0$, то $\text{Var}(\tilde{\beta})$ может быть меньше, чем $\text{Var}(\tilde{\beta}_{OLS})$, если положить $H = -CV_q^{-1}$. Наконец, укажите, что это противоречит утверждению 1.7(с).

- (с) (Дополнительное, только для читателей с хорошим знанием линейной алгебры) Докажите: если K столбцов X являются собственными векторами V , где $V = v \times n$ ковариационная матрица n -мерного вектора ошибок v , то тогда $b = \beta_{OLS}$. (То есть не все несмещенные линейные оценки удовлетворяют требованию невырожденности матрицы $V\text{ar}(\hat{\beta} - \beta_{OLS})$, предъявленному в пункте (b).) **Указание:** Для любой симметричной матрицы V размера $n \times n$ существует матрица H размера $n \times n$ такая, что $H'VH = I_n$ (то есть H — ортогональная матрица), и $H'VH = \Lambda$, где Λ — диагональная матрица, диагональные элементы которой являются собственными значениями V (которые являются неотрицательными, поскольку матрица V симметрична). Столбцы H называются собственными векторами H . Покажите, что

$$H^{-1} = H', \quad H'V^{-1}H = \Lambda^{-1}, \quad H'V^{-1} = \Lambda^{-1}H'.$$

Без потери общности, в качестве X можно взять первые K столбцов матрицы H . То есть $X = HF$, где

$$F = \begin{bmatrix} I_K \\ 0 \end{bmatrix}.$$

Эмпирические приложения

Прежде чем приступить к этому упражнению, прочитайте статью Марка Лерлова (Marc Nerlove) «Returns to Scale in Electricity Supply» (отдача от масштаба в производстве и распределении электроэнергии) (кроме параграфов с уравнениями (6)–(9), части параграфа 2 со с. 184 и далее и приложения А и С). Файл: NERLOVE.ASC содержит следующую информацию по 145 электросетевым компаниям за 1999 год.

Столбец 1: общие издержки (TC) в миллиардах долларов.

Столбец 2: нагрузка (Q) в миллиардах киловатт-часов.

Столбец 3: цена труда (PL).

Столбец 4: цена капитала (PK).

Столбец 5: цена энергии (PX).

Данные взяты из приложения статьи. Данные содержат 140 наблюдений, которые упорядочены по размеру, первой идет самая маленькая компания, 145-я — самая большая. Импортируя исходные данные программного обеспечения, имеющегося на сайте университета, создайте для каждой из 145 компаний переменные, необходимые для оценивания модели. Чтобы оценить, например, уравнение (1.7.4), вам понадобятся следующие переменные: $\log(TC)$, $\log(Q)$, $\log(PL)$, $\log(PK)$ и $\log(PX)$ для каждой из 145 фирм.

- (а) (Вопрос идентичности данных.) Соответствует ли цена капитала (как она использовалась Нерловым) определению стоимости использования капитала? **Указание:** Прочитайте приложение В.4 статьи Нерлова.
- (б) Оцените модель без ограничений (1.7.4) методом наименьших квадратов. Сопоставьте ли ваши результаты с оценками, приведенными в тексте?
- (в) (Метод наименьших квадратов с ограничениями.) Оцените модель с ограничениями (1.7.6) методом наименьших квадратов. Чтобы сделать это, необходимо создать дополнительные переменные для каждой фирмы в выборке. Например, зависимой переменной теперь будет $\log(TC/PP)$, а не $\log(TC)$, как раньше. Сопоставьте ли ваши результаты с оценками, приведенными в тексте? Можете ли вы проверить результаты Нерлова? Например, его оценка коэффициента β_2 равна 0.721 со стандартной ошибкой, равной 0.0174 (в статье написано, что стандартная ошибка равна 0.175, но это, скорее всего, опечатка). Где именно в статье Нерлова находятся эти результаты? А как насчет остальных коэффициентов? (Предупреждение: у вас не получится проверить результаты Нерлова в точности. Одной из причин является то, что он использовал не натуральные, а десятичные логарифмы, однако это влияет только на оценку константы. Другая причина — для его результатов использовалась исправленная версия данных, опубликованных в статье.)

Как уже отмечалось в тексте главы, график остатков свидетельствует в пользу существования линейной связи между $\log(TC)$ и $\log(Q)$. Нерлов предположил, что отдача от масштаба изменяется с размером фирмы. Следуя за ним, разбейте выборку из 145 фирм на группы по 29. (Вспомните, что данные уже отсортированы по возрастанию размера фирмы, поэтому первые 29 наблюдений представляют группу фирм с наименьшим уровнем выпуска, последние — с наибольшим.) Рассмотрите следующие три обобщения модели (1.7.6):

Модель 1: И коэффициенты (β), и дисперсии ошибок σ^2 (1.7.6) различны для разных групп.

Модель 2: Коэффициенты различны, но дисперсии ошибок одинаковы.

Модель 3: Коэффициенты β_0 и β_1 (ценовые эластичности) и дисперсии ошибок одинаковы для всех групп, а коэффициент β_2 и константа разные. Нерлов назвал модель 3 «нейтральной вариацией в отдаче от масштаба».

Для модели 1 с различными для разных групп коэффициентами и дисперсиями ошибок оценки могут быть найдены из:

$$y^{(j)} = X^{(j)}\beta^{(j)} + e^{(j)} \quad (j = 1, \dots, 5).$$

где $y^{(j)} (29 \times 1)$ – вектор значений зависимой переменной для группы j , $X^{(j)} (29 \times 4)$ – матрица значений четырех регрессоров для группы j , $\beta^{(j)} (4 \times 1)$ – вектор коэффициентов группы j , $\varepsilon^{(j)} (29 \times 1)$ – вектор ошибок. Например, второй столбец $X^{(j)}$ равен $\log(Q)$ для $j = 1, 2, \dots, 145$. Модель 1 предполагает выполнение требования условной гомоскедастичности остатков: $E(\varepsilon^{(j)} | X^{(j)}) = \sigma_j^2 I_{29}$ внутри каждой группы (но необязательно между группами).

- (d) Оцените модель 1 методом наименьших квадратов. Совпадают ли ваши результаты с оценками Нерлова? На основе ваших оценки коэффициента β_2 рассчитайте точечные оценки отдачи от масштаба в каждой из пяти групп. Каков общий характер зависимости оценочной отдачи от масштаба от уровня выпуска? Как в общих чертах изменятся оценочная дисперсия ошибок при росте выпуска?

Модель 2 добавляет к модели 1 дополнительное предположение $\sigma_j^2 = \sigma^2$ для всех j . Это ограничение, уравнивающее дисперсию, позволяет интерпретировать и модель, используя стандартизованные векторы и матрицы. Запишем эту модель в следующем виде:

$$y = X\beta + \varepsilon,$$

где

$$y = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(5)} \end{bmatrix} = \begin{bmatrix} 29 \\ \vdots \\ 29 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 116 \\ \vdots \\ 116 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 464 \\ \vdots \\ 464 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 1856 \\ \vdots \\ 1856 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 7424 \\ \vdots \\ 7424 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 29696 \\ \vdots \\ 29696 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 118784 \\ \vdots \\ 118784 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 475136 \\ \vdots \\ 475136 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 1900544 \\ \vdots \\ 1900544 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 7602176 \\ \vdots \\ 7602176 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 30408704 \\ \vdots \\ 30408704 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 121634816 \\ \vdots \\ 121634816 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 486539264 \\ \vdots \\ 486539264 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 1946157056 \\ \vdots \\ 1946157056 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 7784628224 \\ \vdots \\ 7784628224 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 31138512896 \\ \vdots \\ 31138512896 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 124554051584 \\ \vdots \\ 124554051584 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 498216206336 \\ \vdots \\ 498216206336 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 1992864825344 \\ \vdots \\ 1992864825344 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 7971459301376 \\ \vdots \\ 7971459301376 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 31885837205504 \\ \vdots \\ 31885837205504 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 127543348822016 \\ \vdots \\ 127543348822016 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 510173395288064 \\ \vdots \\ 510173395288064 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 2040693581152256 \\ \vdots \\ 2040693581152256 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 8162774324609024 \\ \vdots \\ 8162774324609024 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 32651097298436100 \\ \vdots \\ 32651097298436100 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 130604389193744400 \\ \vdots \\ 130604389193744400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 522417556774977600 \\ \vdots \\ 522417556774977600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 2089670227119910400 \\ \vdots \\ 2089670227119910400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 8358680908479641600 \\ \vdots \\ 8358680908479641600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 33434723633918566400 \\ \vdots \\ 33434723633918566400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 133738894535674265600 \\ \vdots \\ 133738894535674265600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 534955578142697062400 \\ \vdots \\ 534955578142697062400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 2139822312570788249600 \\ \vdots \\ 2139822312570788249600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 8559289250283152998400 \\ \vdots \\ 8559289250283152998400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 34237157001132611993600 \\ \vdots \\ 34237157001132611993600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 136948628004530447974400 \\ \vdots \\ 136948628004530447974400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 547794512018121791897600 \\ \vdots \\ 547794512018121791897600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 2191178048072487167590400 \\ \vdots \\ 2191178048072487167590400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 8764712192289948670361600 \\ \vdots \\ 8764712192289948670361600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 35058848769159794681446400 \\ \vdots \\ 35058848769159794681446400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 140235395076639178725785600 \\ \vdots \\ 140235395076639178725785600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 560941580306556714903142400 \\ \vdots \\ 560941580306556714903142400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 2243766321226226859612569600 \\ \vdots \\ 2243766321226226859612569600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 8975065284904907438450278400 \\ \vdots \\ 8975065284904907438450278400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 35900261139619629753801113600 \\ \vdots \\ 35900261139619629753801113600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 143601044558478519015204454400 \\ \vdots \\ 143601044558478519015204454400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 574404178233814076060817817600 \\ \vdots \\ 574404178233814076060817817600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 2297616712935256304243271270400 \\ \vdots \\ 2297616712935256304243271270400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 9190466851741025216973085081600 \\ \vdots \\ 9190466851741025216973085081600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 36761867406964100867892340326400 \\ \vdots \\ 36761867406964100867892340326400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 147047469627856403471569361305600 \\ \vdots \\ 147047469627856403471569361305600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 588189878511425613886277445222400 \\ \vdots \\ 588189878511425613886277445222400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 235275951404570245554510978089600 \\ \vdots \\ 235275951404570245554510978089600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 941103805618280982218043912358400 \\ \vdots \\ 941103805618280982218043912358400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 3764415222473123928872175649433600 \\ \vdots \\ 3764415222473123928872175649433600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 15057660889892495715488702597734400 \\ \vdots \\ 15057660889892495715488702597734400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 60230643559569982861954810390937600 \\ \vdots \\ 60230643559569982861954810390937600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 240922574238279931447819241563750400 \\ \vdots \\ 240922574238279931447819241563750400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 963690296953119725791276966255001600 \\ \vdots \\ 963690296953119725791276966255001600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 3854761187812478903165107865020006400 \\ \vdots \\ 3854761187812478903165107865020006400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 15419047551249915612660431460080025600 \\ \vdots \\ 15419047551249915612660431460080025600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 61676190204999662450641725840320102400 \\ \vdots \\ 61676190204999662450641725840320102400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 2467047608199986498025669033612804102400 \\ \vdots \\ 2467047608199986498025669033612804102400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 9868190432799945992102676134451216409600 \\ \vdots \\ 9868190432799945992102676134451216409600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 39472761731199783968410704537804865638400 \\ \vdots \\ 39472761731199783968410704537804865638400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 157891046924799135873642818151219462553600 \\ \vdots \\ 157891046924799135873642818151219462553600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 631564187699196543494571272604877850214400 \\ \vdots \\ 631564187699196543494571272604877850214400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 2526256750796786173978285090419511400857600 \\ \vdots \\ 2526256750796786173978285090419511400857600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 10105027003187144695913140361678045603430400 \\ \vdots \\ 10105027003187144695913140361678045603430400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 40420108012748578783652561446712182413721600 \\ \vdots \\ 40420108012748578783652561446712182413721600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 161680432050994315134610245786848729654886400 \\ \vdots \\ 161680432050994315134610245786848729654886400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 646721728203977260538440983147394918619545600 \\ \vdots \\ 646721728203977260538440983147394918619545600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 2586886912815909042153763932589579674478182400 \\ \vdots \\ 2586886912815909042153763932589579674478182400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 10347547651263636168615055730358318697912729600 \\ \vdots \\ 10347547651263636168615055730358318697912729600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 41390190605054544674460222921433274791650918400 \\ \vdots \\ 41390190605054544674460222921433274791650918400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 165560762420218178697840891685733109166603673600 \\ \vdots \\ 165560762420218178697840891685733109166603673600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 662243049680872714791363566742932436666414694400 \\ \vdots \\ 662243049680872714791363566742932436666414694400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 2648972198723490859165454266971729746665658777600 \\ \vdots \\ 2648972198723490859165454266971729746665658777600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 10595888794893963436661817067886918986662635110400 \\ \vdots \\ 10595888794893963436661817067886918986662635110400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 42383555179575853746647268271547675946650540441600 \\ \vdots \\ 42383555179575853746647268271547675946650540441600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 169534220718303414986589073086190703786602161766400 \\ \vdots \\ 169534220718303414986589073086190703786602161766400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 678136882873213659946356292344762815146408647065600 \\ \vdots \\ 678136882873213659946356292344762815146408647065600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 2712547531492854639785425169379051260585634588262400 \\ \vdots \\ 2712547531492854639785425169379051260585634588262400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 10850190125971418559141700677516205042342538353049600 \\ \vdots \\ 10850190125971418559141700677516205042342538353049600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 43400760503885674236566802710064820169370153412198400 \\ \vdots \\ 43400760503885674236566802710064820169370153412198400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 173603042015542696946267210840259280677480613648793600 \\ \vdots \\ 173603042015542696946267210840259280677480613648793600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 694412168062170787785068843361037122709922454595174400 \\ \vdots \\ 694412168062170787785068843361037122709922454595174400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 2777648672248683151140275373444148490839689818380697600 \\ \vdots \\ 2777648672248683151140275373444148490839689818380697600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 11110594688994732604561101493776593963358759273522790400 \\ \vdots \\ 11110594688994732604561101493776593963358759273522790400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 44442378755978930418244405975106375853435037094091161600 \\ \vdots \\ 44442378755978930418244405975106375853435037094091161600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 177769515023915721672977623900425503413740148376364646400 \\ \vdots \\ 177769515023915721672977623900425503413740148376364646400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 711078060095662886691910495601702013654960593505458585600 \\ \vdots \\ 711078060095662886691910495601702013654960593505458585600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 2844312240382651546767641982406808054619842374021834342400 \\ \vdots \\ 2844312240382651546767641982406808054619842374021834342400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 11377248961530606187070567929627232218479369496087337369600 \\ \vdots \\ 11377248961530606187070567929627232218479369496087337369600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 45508995846122424748282271718508928873917477984349349478400 \\ \vdots \\ 45508995846122424748282271718508928873917477984349349478400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 182035983384489698993129086874035715495669911937397397913600 \\ \vdots \\ 182035983384489698993129086874035715495669911937397397913600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 728143933537958795972516347496142861982679647749589591654400 \\ \vdots \\ 728143933537958795972516347496142861982679647749589591654400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 2912575734151835183890065389984571447930718586998358366617600 \\ \vdots \\ 2912575734151835183890065389984571447930718586998358366617600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 11650302936607340735560261559938285791722874347993433466470400 \\ \vdots \\ 11650302936607340735560261559938285791722874347993433466470400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 46601211746429362942241046239753143166891497391973733865881600 \\ \vdots \\ 46601211746429362942241046239753143166891497391973733865881600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 186404846985717451768964184959012572667565989567894935463526400 \\ \vdots \\ 186404846985717451768964184959012572667565989567894935463526400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 745619387942869807075856739836050290670263958271579741854105600 \\ \vdots \\ 745619387942869807075856739836050290670263958271579741854105600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 2982477551771479228303426959344201162681055833086318967416422400 \\ \vdots \\ 2982477551771479228303426959344201162681055833086318967416422400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 11929910207085916913213707837376804650724223332345275869665689600 \\ \vdots \\ 11929910207085916913213707837376804650724223332345275869665689600 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 47719640828343667652854831349507218602896893329381103478662758400 \\ \vdots \\ 47719640828343667652854831349507218602896893329381103478662758400 \end{bmatrix} \times \begin{bmatrix} 4 \\ \vdots \\ 4 \end{bmatrix} = \begin{bmatrix} 19087856331337$$

того же конечного результата с помощью дамми-переменных. Определим j -ю дамми-переменную как

$$D_{ij} = \begin{cases} 1 & \text{если } i\text{-я фирма принадлежит } j\text{-й группе, } (i = 1, \dots, 145), \\ 0 & \text{в противном случае} \end{cases}$$

Тогда второй регрессор можно построить как $D_{1i} \cdot \log(Q_i)$. Шестой — $D_{2i} \cdot \log(Q_i)$, и так далее.

- (е) Оцените модель 2 методом наименьших квадратов. Проверьте, что OLS-оценки, полученные в этом пункте, совпадают с оценками пункта (д). Также проверьте, что

$$\sum_{j=1}^3 SSR_j = SSR,$$

где SSR_j — это SSR , полученная при оценивании модели 1 в (д) для j -й группы, а SSR — это SSR в модели 2. Такое равенство случайно (то есть не является специфической чертой имеющегося набора данных). Докажите, что упомянутые выше тождества для коэффициентов и SSR выполняются вообще (без потери общности, рассмотрите случай двух групп). **Указание:** Сначала покажите, что коэффициенты модели 1 и модели 2 совпадают. Используйте формулы (A.4), (A.5) и (A.9) приложения A.

- (г) (Тест Чоу.) Модель 2 является более общей, чем модель (1.7.6), так как коэффициенты могут различаться для разных групп. Проверьте нулевую гипотезу о том, что коэффициенты одинаковы для всех групп. Сколько уравнений (ограничений) в этой гипотезе? Этот тест иногда называют **тестом Чоу на наличие структурного сдвига** (*Chow test for structural change*). Рассчитайте p -значение для F -отношения. **Указание:** Это линейная гипотеза относительно коэффициентов модели 2. То есть модель 2 — поддерживаемая гипотеза, а (1.7.6) — модель с ограничениями. Используйте формулу (1.4.11) для F -отношения.

Совет при работе в пакете Gauss: Если обозначить за x F -отношение, то команда `Gauss cdfc(x, df1, df2)` выдает площадь под кривой плотности справа от x для F -распределения со степенями свободы $df1$ и $df2$.

Совет при работе в пакете TSP: Аналогичная TSP-команда имеет вид `cdf(f, df1=df1, df2=df2) x`. Результат применения команды `OLSQ`, реализующей OLS, выдаваемый как `@SSR`, представляет SSR регрессии.

Совет при работе в пакете RATS: В RATS команда выглядит так: `cdf ftest x df1 df2`. RSS регрессии может быть получена после использования команды OLS, LINREG, с помощью %RSS.

Ограничение в модели 3, предполагающее, что ценовые эластичности одинаковы внутри групп фирм, можно ввести в модель 2, применяя преобразование, использующее лагми-переменные только для константы и логарифма выпуска. Таким образом, в новую модель включаются $12(-2 \times 5 + 2)$ переменных, причем матрица X выглядит так:

$$X = \begin{bmatrix} 1 & \log(Q_{11}) & 0 & 0 & \log(PL_{11}/PF_{11}) & \log(PK_{11}/PF_{11}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \log(Q_{21}) & 0 & 0 & \log(PL_{21}/PF_{21}) & \log(PK_{21}/PF_{21}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & \log(Q_{117}) & \log(PL_{117}/PF_{117}) & \log(PK_{117}/PF_{117}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & \log(Q_{145}) & \log(PL_{145}/PF_{145}) & \log(PK_{145}/PF_{145}) \end{bmatrix} \quad (4.8)$$

- (г) Оцените модель 3. Эта модель является частным случаем модели 2, соответствующим гипотезе о том, что две ценовые эластичности одинаковы для всех лаг групп. Протестируйте эту гипотезу на 5%-ном уровне значимости, предположая нормальность ошибок. (Замечание: F отношение на с. 183 статьи Нурдуса неверно.)

Как мы уже установили из графического анализа остатков на рис. 1.7, условный момент второго порядка $E(\varepsilon^2|X)$, скорее всего, зависит от логарифма выпуска, что является нарушением предположения об условной гомоскедастичности остатков. Сейчас мы не будем пытаться тестировать условную гомоскедастичность, так как для этого необходима асимптотическая теория, которая будет рассмотрена в следующей главе. Вместо этого мы будем делать вид, что нам известна функциональная форма зависимости условного момента второго порядка от логарифма выпуска. Эта функция, спецификация которой приведена ниже, такова, что условный момент второго порядка непрерывно зависит от логарифма выпуска, в отличие от трех моделей, рассмотренных ранее. Еще одно отличие от предыдущих моделей состоит в том, что теперь мы предполагаем, что отклоня от масштаба изменяется непрерывно с изменением выпуска. Чтобы

добиться этого, добавим в модель квадрат логарифма выпуска¹. Тогда модель 4 имеет следующий вид:

Модель 4:

$$\log\left(\frac{Y_i}{P_{i3}}\right) = \beta_1 + \beta_2 \log(Q_i) + \beta_3 [\log(Q_i)]^2 + \\ + \beta_4 \log\left(\frac{P_{i1}}{P_{i3}}\right) + \beta_5 \log\left(\frac{P_{i2}}{P_{i3}}\right) + \varepsilon_i.$$

$$E(\varepsilon_i^2 | X) = \sigma^2 \cdot \left(0.0505 + \frac{2.1377}{Q_i}\right) \quad (i = 1, 2, \dots, 145)$$

для некоторого неизвестного значения σ^2 .

- (h) Оцените модель 4 с помощью взвешенного метода наименьших квадратов для всей выборки из 145 фирм. (Будьте внимательны в отношении констант: после проведения процедуры взвешивания ни один из регрессоров не будет постоянным.) Нарисуйте график остатков. Свидетельствует ли он об условной гомоскедастичности ошибок или, как и прежде, о наличии нелинейности?

Упражнения на метод Монте-Карло

Метод **Монте-Карло** (*Monte Carlo*) состоит в генерации большого количества выборок на основе некоторой модели, чтобы изучить распределение оценок в конечных выборках. Здесь мы будем использовать этот метод для того, чтобы подтвердить два результата для малых выборок, упоминавшихся в тексте главы: несмещенность OLS-оценки коэффициентов и распределение t -отношения. Для этого рассмотрим модель простой линейной регрессии, удовлетворяющую предположениям 1.1–1.5, с количеством наблюдений $n = 32$. Регрессионные уравнения выглядят следующим образом:

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \quad (i = 1, 2, \dots, n) \\ \text{или } y = 1 \cdot \beta_1 + x \cdot \beta_2 + \varepsilon = X\beta + \varepsilon, \quad (*)$$

где $X = (1, x)$ и $\beta = (\beta_1, \beta_2)'$. Параметрами модели являются $(\beta_1, \beta_2, \sigma^2)$.

¹Мы вывели линейную в логарифмах функцию издержек из производственной функции Кобба–Дугласа. Существует ли такая произвольная функция, из которой можно было бы вывести обобщенную функцию издержек, содержащую квадрат логарифма выпуска? Это вопрос об «интегрируемости» (integrability) функций издержек, относительно обсуждения которого можно найти в: [Christensen et al., 1973].

Как уже упоминалось в тексте, модель представляет собой набор совместных распределений (y, X) . Из этого набора мы выберем конкретные совместные распределения, определяя регрессионную модель с двумя параметрами. Пусть $\beta = 1$, $\beta_2 = 0.5$ и $\sigma^2 = 1$. Распределение $x = (x_1, x_2, \dots, x_n)'$ задано следующим AR(1)-процессом:

$$x_i = \phi + \phi x_{i-1} + \eta_i \quad (i = 1, 2, \dots, n), \quad (**)$$

где $\{\eta_i\} \sim i.i.d.N(0, 1 - \phi^2)$.

$$x_i \sim N\left(\phi \left(1 + \phi + \phi^2 + \dots + \phi^{i-1}\right), \sigma^2 \left(1 + \phi^2 + \dots + \phi^{2(i-1)}\right)\right), \quad \phi = 0.5, \quad \sigma = 0.6.$$

Таким образом, задается совместное распределение (y, X) . Именно из этого распределения будет извлекаться *большое количество выборок*.

Для осуществления n итераций с помощью компьютера будет полезен предопределенный следующий код. Входящим образом уравнение пятого порядка (**), получим:

$$x_i = \phi^i x_0 + (1 - \phi + \phi^2 + \dots + \phi^{i-1})\phi + \\ + (\eta_0 - \phi\eta_1 + \phi^2\eta_2 + \dots + \phi^{i-1}\eta_i),$$

или, используя матричные обозначения,

$$\begin{pmatrix} x \\ (n \times 1) \end{pmatrix} = \begin{pmatrix} \phi \\ (n \times 1) \end{pmatrix} \cdot x_0 + \begin{pmatrix} d \\ (n \times 1) \end{pmatrix} + \begin{pmatrix} A \\ (n \times n)_{(n \times 1)} \end{pmatrix} \cdot \begin{pmatrix} \eta \\ (n \times 1) \end{pmatrix}, \quad (***)$$

где $d = (d_1, d_2, \dots, d_n)'$ и

$$d_1 = \phi, \quad d_2 = (1 - \phi)\phi, \dots, \quad d_i = (1 + \phi + \phi^2 + \dots + \phi^{i-1})\phi, \dots,$$

$$r = \begin{bmatrix} \phi \\ \phi^2 \\ \vdots \\ \phi^n \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \phi^1 & \phi & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \dots & \phi & 1 \end{bmatrix}, \quad \eta = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix}.$$

Совет при работе с пакетом StatSoft: Для построения матрицы r используйте команду `best`. Для получения матрицы A используйте `toeplitz` и `lowmat`.

(a) Проведите две симуляции методом Монте-Карло. Первая симуляция служит для вычисления $E(b|x)$ и распределения t -отношения как распределения, условного относительно x . Компьютерная программа для этой первой симуляции должна состоять из следующих шагов.

- (1) (Сгенерируйте $\mathbf{x}$ только один раз.) Используя генератор случайных чисел, создайте вектор η , состоящий из реализаций n i.i.d. $N(0, 1)$ случайных величин, и выберите ω_0 из $N(c/(1-\phi), 1/(1-\phi^2))$, рассчитайте $\mathbf{x}$ по формуле (**). ($\mathbf{x}$ также может быть рассчитан рекурсивно из (**) с использованием цикла, однако операции с векторами, подобные (**), требуют меньшего процессорного времени, чем циклы. Это замечание будет актуальным во второй симуляции, где $\mathbf{x}$ надо будет генерировать при каждом повторении.)
- (2) Создайте переменную-счетчик и положите ее равной нулю. В этой переменной будет записываться количество случаев, когда $|t| > t_{0.025}(n-2)$. Кроме того, создайте двумерный вектор, первоначально состоящий из нулей; этот вектор будет использоваться для вычисления математического ожидания OLS-оценки $\hat{\mathbf{b}}$ для $(\beta_1, \beta_2)'$.
- (3) Начните цикл для большого числа повторений (например, для миллиона). В каждом повторении выполните следующие действия:
 - (i) (Создайте $\mathbf{y}$.) Сгенерируйте n -мерный вектор ε , состоящий из реализаций n i.i.d. случайных величин $N(0, 1)$, вычислите $\mathbf{y} = (y_1, y_2, \dots, y_n)$ по формуле (*). Теперь этот $\mathbf{y}$ вместе с $\mathbf{x}$ из шага (1) составят нашу выборку $(\mathbf{y}, \mathbf{x})$.
 - (ii) По имеющейся выборке вычислите OLS-оценку $\hat{\mathbf{b}}$ и t -отношение для $H_0: \beta_2 = 0.5$.
 - (iii) Увеличьте значение счетчика на 1, если $|t| > t_{0.025}(n-2)$. Прибавьте $\hat{\mathbf{b}}$ к двумерному вектору.
- (4) После окончания цикла разделите значение переменной-счетчика на количество повторений, чтобы вычислить частоту отклонения нулевой гипотезы. Также разделите двумерный вектор, равный сумме OLS-оценок $\hat{\mathbf{b}}$ для всех повторений, на число повторений. Результат деления должен стать равным $E(\hat{\mathbf{b}}|w)$ при бесконечном количестве повторений.

Заметьте, что $\mathbf{x}$ в этой первой симуляции генерируется только один раз и остается неизменным в процессе получения всех значений $\mathbf{y}$. Во второй симуляции вычисляется безусловное распределение t -отношения. Это выполняется в несколько этапов.

- (1) Создайте переменную-счетчик, положите ее равной нулю.
- (2) Начните цикл для большого числа повторений (например, для миллиона). В каждом повторении выполните следующие действия:
 - (i) (Создайте $\mathbf{x}$.) Создайте вектор η , состоящий из реализаций n i.i.d. $N(0, 1)$ случайных величин, и выберите ω_0 из $N(c/(1-\phi), 1/(1-\phi^2))$, рассчитайте $\mathbf{x}$ по формуле (**).

(i) Создайте y . Сгенерируйте n -мерный вектор u , состоящий из реализаций n i.i.d. стандартных нормальных $N(0, 1)$, вычислите $y = (y_1, \dots, y_n)'$ по формуле (2).

(ii) По имеющейся в форме (y, x) вычислите t -отношение для $H_0: \beta_2 = 0.5$.

(iv) Увеличьте значение счетчика на 1, если $|t| > t_{0.25}(n-2)$.

(3) После окончания цикла рассчитайте значение переменной-счетчика из количества повторений.

Для этих двух симуляций проверьте, что для достаточно большого количества наблюдений:

1. Среднее OLS-оценок b из первого эксперимента неограниченно приближается к истинному значению (1, 0.5);

2. Частота отклонения нулевой гипотезы H_0 (ошибка первого рода) при различных приближается к 5% в обоих экспериментах.

(4) Являются ли регрессоры (не включенные константы) строго экзогенными в этих двух симуляциях? Насколько ли ошибка условно homoskedastic? Достаточно?

Ответы на избранные вопросы

Аналитические упражнения

$$\begin{aligned}
 1. (y - X\tilde{\beta})'(y - X\tilde{\beta}) &= \\
 &= (y - Xb)' + X(b - \tilde{\beta})'(y - Xb) - X(b - \tilde{\beta})' \\
 &\text{(приблизление и вычитание скалярных величин)} \\
 &= (y - Xb)' + (b - \tilde{\beta})'X'(y - Xb) + X(b - \tilde{\beta})' = \\
 &= (y - Xb)'(y - Xb) + (b - \tilde{\beta})'X'(y - Xb) - \\
 &- (y - Xb)'X(b - \tilde{\beta}) - (b - \tilde{\beta})'X'X(b - \tilde{\beta}) = \\
 &= (y - Xb)'(y - Xb) - 2(b - \tilde{\beta})'X'(y - Xb) + \\
 &+ (b - \tilde{\beta})'X'X(b - \tilde{\beta}).
 \end{aligned}$$

$$\text{(так как } (b - \tilde{\beta})'X'(y - Xb) = (y - Xb)'X(b - \tilde{\beta})$$

$$= (y - Xb)'(y - Xb) + (b - \tilde{\beta})'X'X(b - \tilde{\beta})$$

(поскольку $X'(y - Xb) = 0$ из системы нормальных уравнений)

$$> (y - Xb)'(y - Xb)$$

$$\text{(так как } (b - \tilde{\beta})'X'X(b - \tilde{\beta}) = z'z = \sum_{i=1}^n z_i^2 \geq 0, \text{ где } z \equiv X(b - \tilde{\beta})$$

7а. $\hat{\beta}_{OLS} = \beta = A\varepsilon$, где $A \equiv (X'V^{-1}X)^{-1}X'V^{-1}$ и $b = \hat{\beta}_{CLS} = B\varepsilon$, где $B \equiv (X'X)^{-1}X' = (X'V^{-1}X)^{-1}X'V^{-1}$. Поэтому

$$\begin{aligned} \text{Cov}(\hat{\beta}_{CLS} - \beta, b - \hat{\beta}_{CLS}) &= \\ \text{Cov}(A\varepsilon, B\varepsilon) &= \\ AVB' &= \\ = \sigma^2 AVB' \end{aligned}$$

Равенство $AVB' = 0$ доказывается напрямую

7б. Для матрицы H , определенной в указании к этому пункту,

$$\text{Var}(\hat{\beta}) - \text{Var}(\hat{\beta}_{CLS}) = -CV_0^{-1}C'.$$

Если $C' \neq 0$, то тогда существует ненулевой вектор z , такой, что $C'z \equiv v \neq 0$. Для такого z

$$z' [\text{Var}(\hat{\beta}) - \text{Var}(\hat{\beta}_{CLS})] z = -v' V_0^{-1} v < 0$$

(поскольку матрица V_0 положительно определена), получаем противоречие, так как $\hat{\beta}_{CLS}$ — эффективная оценка.

Эмпирические упражнения

- (а) Из описания Нерлова в приложении В.4 следует, что он не использовал норму амортизации 0 для расчета цены капитала.
- (б) Ваши оценки должны совпасть с (1.7.7).
- (в) Наши оценки немного отличаются от оценок Нерлова. Они отличались бы, даже если бы данные, которыми пользовался Нерлов, совпадали с теми, которые есть у вас. Причина состоит в том, что компьютеры в то время производили не такие точные вычисления и допускали больше ошибок округления.
- (г) Насколько точно вы можете повторить приведенные у Нерлова результаты? Достаточно точно. Точечные оценки от масштаба в каждой из пяти групп равны 2.5, 1.5, 1.1, 1.1 и 0.95. При увеличении выпуска оценка от масштаба снижается.
- (д) Модель 2 может быть записана в виде $y = X\beta + \varepsilon$, где y , X и ε определены так же, как и в (4). Поэтому (выбрав $j = 2$)

$$X'X = \begin{bmatrix} X^{(1)'}X^{(1)} & 0 \\ 0 & X^{(2)'}X^{(2)} \end{bmatrix}.$$

откуда следует

$$(X'X)^{-1} = \begin{bmatrix} (X^{(1)'}X^{(1)})^{-1} & 0 \\ 0 & (X^{(2)'}X^{(2)})^{-1} \end{bmatrix}.$$

Кроме того

$$X'y = \begin{bmatrix} X^{(1)'}y^{(1)} \\ X^{(2)'}y^{(2)} \end{bmatrix}$$

Поэтому

$$(X'X)^{-1}X'y = \begin{bmatrix} (X^{(1)'}X^{(1)})^{-1}X^{(1)'}y^{(1)} \\ (X^{(2)'}X^{(2)})^{-1}X^{(2)'}y^{(2)} \end{bmatrix}.$$

Таким образом, вектор OLS-оценок коэффициентов модели 2 такой же, как и в модели 1. Сумма квадратов остатков также совпадает, вследствие равенства оценок коэффициентов.

- (f) Количество ограничений равно 16. K = количество коэффициентов в модели 2 = 20. Тогда число степеней свободы должно равняться {16, 125}. $SSR_U = 12,282$ и $SSR_D = 21,840$. F -отношение = 5,97, p -значение = 0,0000. Поэтому нулевая гипотеза отвергается на любом разумном уровне значимости.
- (g) $SSR_U = 12,282$ и $SSR_D = 12,577$. Тогда $F = 0,40$ с 8 и 125 степенями свободы, p -значение равно 0,92. Нулевая гипотеза не отвергается на любом разумном уровне значимости. F -отношение в работе Нерлунга (см. с. 183, 8-я строка снизу) равно 1,576.
- (h) График остатков все равно свидетельствует о том, что условный момент второго порядка больше для меньших фирм, однако теперь нет намека на его нелинейность.

Литература

- Ameniya, T., 1985, *Advanced Econometrics*. Cambridge: Harvard University Press.
- Averch, H., and L. Johnson, 1962, "Behavior of the Firm under Regulatory Constraint," *American Economic Review*, 52, 1052-1069.
- Christensen, L., and W. Greene, 1976, "Economics of Scale in US Electric Power Generation," *Journal of Political Economy*, 84, 655-676.
- Christensen, L., D. Jorgenson, and L. Lau, 1973, "Transcendental Logarithmic Production Frontiers," *Review of Economics and Statistics*, 55, 28-45.
- Davidson, R., and J. MacKinnon, 1993, *Estimation and Inference in Econometrics*. Oxford: Oxford University Press.
- DeLong, B., and L. Summers, 1991, "Equipment Investment and Growth," *Quarterly Journal of Economics*, 99, 28-45.

- Engle, R., D. Hendry and J.-F. Richard, 1983. "Exogeneity," *Econometrica*, 51, 277-304.
- Federal Power Commission, 1955. *Statistics of Electric Utilities in the United States, 1955. Class A and B Privately Owned Companies*, Washington, D.C.
- Jorgenson, D., 1963. "Capital Theory and Investment Behavior," *American Economic Review*, 53, 247-259.
- Koopmans, T., and W. Hood, 1953. "The Estimation of Simultaneous Linear Economic Relationships," in W. Hood, and T. Koopmans (eds.), *Studies in Econometric Method*, New Haven: Yale University Press.
- Krasker, W., E. Kuhn, and R. Welsch, 1983. "Estimation for Dirty Data and Flawed Models," Chapter 11 in Z. Griliches, and M. Intriligator (eds.), *Handbook of Econometrics*, Volume 1, Amsterdam: North-Holland.
- Nerlove, M., 1953. "Returns to Scale in Electricity Supply" in C. Christ (ed.), *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Janada Grusfeld*, Stanford: Stanford University Press.
- Rao, C. R., 1973. *Linear Statistical Inference and Its Applications* (2d ed.), New York: Wiley.
- Scholle, H., 1959, *The Analysis of Variance*, New York: Wiley.
- Wolak, F., 1984. "An Econometric Analysis of the Asymmetric Information, Regulator-Utility Interaction," *Revue D'Economie et de Statistique*, 34, 13-39.

Глава 2. Теория больших выборок

В предыдущей главе мы получили точные распределения, или распределения на конечных выборках, OLS-оценки и соответствующих ей тестовых статистик. Однако в экономике предположения о точном распределении выполняются не очень часто. Теория конечных выборок теряет силу, если нарушается одно из трех следующих предположений: 1) экзогенность регрессора, 2) нормальность ошибок и 3) линейность уравнения регрессии. В настоящей главе разрабатывается альтернативный подход, в котором сохраняется только третье предположение. При таком подходе, называемом **асимптотической теорией**, или **теорией больших выборок**, аппроксимацию распределения оценки и соответствующих ей статистик получают в предположении достаточно большого объема выборки.

Вместо предположений о выборке заданного объема в теории больших выборок делаются предположения о случайном процессе, порождающем эту выборку. В первых двух параграфах этой главы вводится язык, необходимый для описания случайных процессов.

Понятия, введенные в данной главе, имеют важное значение в эконометрике рациональных ожиданий, что было продемонстрировано в классической работе Фамы о гипотезе Физера, согласно которой реальная процентная ставка постоянна. Эта гипотеза имеет очень серьезные последствия для политики: монетарная и фискальная политика не могут влиять на агрегированный спрос посредством реальной процентной ставки. Весьма удивительно, что никто не может отвергнуть данную гипотезу для Соединенных Штатов (по крайней мере, если период выборки заканчивается в начале 1970-х).

2.1. Обзор предельных теорем для последовательностей случайных величин

Материал этого параграфа касается предельного поведения последовательности случайных величин, $\{z_1, z_2, \dots\}$. Поскольку этот материал может быть уже известен вам, мы представляем его достаточно формальным рядом определений и теорем. Авторитетным источником является книга Редо [Redo, 1973, Chapter 2c], в которой приводятся доказательства всех теорем, включенных в данный параграф. В этом параграфе и в оставшейся части книги последовательность $\{z_1, z_2, \dots\}$ будет обозначаться как $\{z_n\}$.

Ранжированные виды сходимости

Сходимость по вероятности

Последовательность случайных скаляров $\{z_n\}$ *сходится по вероятности* (*converges in probability*) к константе (скалярной) α , если для любого $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \text{Prob}[|z_n - \alpha| > \epsilon] = 0. \quad (2.1.1)$$

Константа α называется *пределом по вероятности* (*probability limit*) z_n , и это записывается как $\lim_{n \rightarrow \infty} z_n = \alpha$ или $z_n \xrightarrow{p} \alpha$. Очевидно, что

$$z_n \xrightarrow{p} \alpha \text{ — то же самое, что и } z_n \xrightarrow{p} \alpha + 0.$$

Это определение сходимости по вероятности обобщается на последовательности случайных векторов или случайных матриц (матрица рассматривается как вектор, элементы которой переставлены) требованием того же ранжированной сходимости по вероятности. То есть последовательности K -мерных случайных векторов $\{z_n\}$ *сходится по вероятности к K -мерному вектору констант α* , если для любого $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \text{Prob}[|z_{nk} - \alpha_k| > \epsilon] = 0 \text{ для всех } k (= 1, 2, \dots, K). \quad (2.1.2)$$

где z_{nk} есть k -й элемент z_n и α_k есть k -й элемент α .

Сходимость почти наверное

Последовательность случайных скаляров $\{z_n\}$ *сходится почти наверное* (*converges almost surely*) к константе α , если

$$\text{Prob}\left(\lim_{n \rightarrow \infty} z_n = \alpha\right) = 1. \quad (2.1.3)$$

Мы записываем это как $z_n \xrightarrow{a.s.} \alpha$. Обобщение на случайные векторы аналогично обобщению сходимости по вероятности. Как будет упомянуто

ниже, это понятие сходимости более строгое, чем сходимость по вероятности: то есть если последовательность сходится почти наверное, то она сходится и по вероятности. Понятие, включенное в (2.1.3), назвать сложнее, поскольку указанная вероятность относится к событию, касающемуся бесконечной последовательности $\{z_1, z_2, \dots\}$. Для наших целей, однако же, что имеет значение, — это то, что сходимость почти наверное более строгая, чем сходимость по вероятности. Если мы можем доказать, что последовательность сходится почти наверное, то это один из способов доказать, что последовательность сходится по вероятности.

Сходимость в среднеквадратичном

Последовательность случайных скаляров $\{z_n\}$ **сходится в среднеквадратичном** (*converges in mean square* или *in quadratic mean*) к α (записывается как $\langle z_n \rightarrow_{r.s.} \alpha \rangle$), если

$$\lim_{n \rightarrow \infty} E\{(z_n - \alpha)^2\} = 0. \quad (2.1.4)$$

Обобщение на случайные векторы аналогично обобщению сходимости по вероятности: $z_n \rightarrow_{p.s.} \alpha$, если каждый элемент z_{ni} сходится в среднеквадратичном к соответствующей компоненте α_i .

Сходимость к случайной величине

В приведенных определениях сходимости предел является константой (то есть действительным числом). Предел может быть случайной величиной. Мы говорим, что последовательность K -мерных случайных величин $\{z_n\}$ сходится к K -мерной случайной величине z , и пишем $z_n \rightarrow_p z$, если $\{z_n - z\}$ сходится к 0:

$$\langle z_n \rightarrow_p z \rangle \quad \text{то же самое, что и} \quad \langle z_n - z \rightarrow_p 0 \rangle. \quad (2.1.5a)$$

Аналогично,

$$\langle z_n \xrightarrow{p.s.} z \rangle \quad \text{то же самое, что и} \quad \langle z_n - z \xrightarrow{p.s.} 0 \rangle, \quad (2.1.5b)$$

$$\langle z_n \xrightarrow{p} z \rangle \quad \text{то же самое, что и} \quad \langle z_n - z \xrightarrow{p} 0 \rangle. \quad (2.1.5c)$$

Сходимость по распределению

Пусть $\{z_n\}$ — последовательность случайных скаляров, а F_n — кумулятивная функция распределения (с.д.ф.) z_n . Мы говорим, что $\{z_n\}$ **сходится по распределению** (*converges in distribution*) к случайному скаляру z , если с.д.ф. F_n от z_n сходится к с.д.ф. F от z каждой

точке непрерывности F . Мы пишем « $z_n \rightarrow_d z$ » или « $z_n \rightarrow_d z$ » и называем P **асимптотическим** (*asymptotic*) (или **предельным** (*limit* или *limiting*)) **распределением** z . Иногда мы пишем « $z_n \rightarrow_d F$ », когда распределение P хорошо известно. Например, « $z_n \rightarrow_d N(0,1)$ » следует читать так: « $z_n \rightarrow_d z$ и распределение z является $N(0,1)$ (нормальное распределение со средним 0 и дисперсией 1)». Пользуясь определением, можно показать, что сходимость по вероятности является более строгой, чем сходимость по распределению, то есть

$$\langle z_n \rightarrow_p z \rangle \Rightarrow \langle z_n \rightarrow_d z \rangle. \quad (2.1.6)$$

В частном случае сходимости по распределению z является константой (тривиальная случайная величина).

Обобщение на последовательности случайных векторов простое: $z_n \rightarrow_d z$, если совместная с.в.д. F_n случайного вектора z_n сходится к совместной с.в.д. F случайного вектора z в каждой точке непрерывности F . Заметим, однако, что, в отличие от других понятий сходимости, в случае сходимости по распределению поэлементная сходимость не обязательно означает сходимость для последовательности векторов. То есть из того, что «каждый элемент $z_{ni} \rightarrow_d z_i$ к соответствующему элементу z_i » не обязательно вытекает, что « $z_n \rightarrow_d z$ ». Распространенным способом установить связь между сходимостью скаляров и сходимостью векторов по распределению является

Теорема о многомерной сходимости по распределению: (изложенная в [Kao, 1973, p. 128]). Пусть $\{z_n\}$ — последовательность K -мерных случайных векторов. Тогда,

$$\langle z_n \rightarrow_d z \rangle \Leftrightarrow \langle X' z_n \rightarrow_d X' z \rangle$$

для любого K -мерного вектора действительных чисел X' .

Связи между сходимостью по распределению и сходимостью моментов

Стоит подчеркнуть, что моменты предельного распределения z не обязательно равны пределам моментов z_n . Например, из « $z_n \rightarrow_d z$ » не обязательно вытекает « $\lim_{n \rightarrow \infty} E(z_n) = E(z)$ ». Однако

¹ Не забывайте внимания на определение «в каждой точке непрерывности F ». По большей части, за исключением случаев, став о дискретной или с граничной зависимой переменной, соответствующее распределение является непрерывным, а функция распределения непрерывна во всех точках.

Лемма 2.1 (сходимость по распределению и сходимость моментов): Пусть α_m является m -м моментом γ_n в $\text{Fin}_{A, \infty}$; $\alpha_m = \alpha_m$, где α_m коническое (то есть действительное число). Предположим, что для некоторого $\delta > 0$ $E(|z_n|^{2+\delta}) < M < \infty$ для всех n . Тогда:

$$\gamma_n \rightarrow \gamma \Rightarrow \alpha_n \rightarrow \alpha \text{ есть } m \text{ й момент } \gamma.$$

Таким образом, например, если дисперсия последовательности случайных величин, сходящихся по распределению, сходится к некоторому конечному числу, то тогда это число является дисперсией предельного распределения.

Связи между видами сходимости

Некоторые виды сходимости являются более слабыми, чем другие. Следующая теорема устанавливает соотношение между четырьмя видами сходимости.

Лемма 2.2 (соотношение между четырьмя видами сходимости):

(a) $\gamma_n \rightarrow_m \gamma, \alpha_n \Rightarrow \gamma_n \rightarrow_p \gamma$. Аналогично $\gamma_n \rightarrow_{m,2} \gamma \Rightarrow \gamma_n \rightarrow_p \gamma$.

(b) $\gamma_n \rightarrow_{p,2} \gamma \Rightarrow \gamma_n \rightarrow_p \gamma$. Аналогично $\gamma_n \rightarrow_{p,2} \gamma \Rightarrow \gamma_n \rightarrow_p \gamma$.

(c) $\gamma_n \rightarrow_p \gamma \Leftrightarrow \gamma_n \rightarrow_d \gamma$ (То есть, если предельная случайная величина является константой (тривиальной случайной величиной), сходимости по распределению равносильна сходимости по вероятности.)

Три полезных результата

Определая виды сходимости, мы можем установить три результата, важных для разработки теории больших выборок.

Лемма 2.3 (сохранение сходимости для непрерывных преобразований): Предположим, что $\alpha(\cdot)$ — векторнозначная непрерывная функция¹, которая не зависит от n .

(a) $\gamma_n \rightarrow_p \gamma \Rightarrow \alpha(\gamma_n) \rightarrow_p \alpha(\gamma)$. Иными словами,

$$\lim_{n \rightarrow \infty} \alpha(\gamma_n) = \alpha(\lim_{n \rightarrow \infty} \gamma_n)$$

при условии существования $\lim_{n \rightarrow \infty} \gamma_n$.

¹Часть (a) требует непрерывности только в α , в то время как часть (b) требует непрерывности всюду. — Прим. авт.

$$(b) *x_n \rightarrow_d x* \Rightarrow *n(z_n) \rightarrow_d n(x)*.$$

Немедленным следствием Леммы 2.3(а) является то, что обычные арифметические операции сохраняют сходимость по вероятности. Например:

$$*x_n \xrightarrow{p} \beta, y_n \xrightarrow{p} \gamma* \Rightarrow *x_n + y_n \xrightarrow{p} \beta + \gamma*.$$

$$*x_n \xrightarrow{p} \beta, y_n \xrightarrow{p} \gamma* \Rightarrow *x_n y_n \xrightarrow{p} \beta \gamma*.$$

$$*x_n \xrightarrow{p} \beta, y_n \xrightarrow{p} \gamma* \Rightarrow *x_n/y_n \xrightarrow{p} \beta/\gamma* \text{ при условии, что } \gamma \neq 0.$$

$$*Y_n \xrightarrow{p} \Gamma* \Rightarrow *Y_n^{-1} \xrightarrow{p} \Gamma^{-1}*.$$

при условии, что Γ является обратной.

Следующий результат о комбинациях сходимости по вероятности и сходимости по распределению будет неоднократно использоваться для получения асимптотического распределения оценок.

Лемма 2.4:

$$(a) *x_n \rightarrow_d x, y_n \rightarrow_p 0* \Rightarrow *x_n + y_n \rightarrow_d x + 0*.$$

$$(b) *x_n \rightarrow_d x, y_n \rightarrow_p 0* \Rightarrow *y_n' x_n \rightarrow_p 0*.$$

$$(c) *x_n \rightarrow_d x, A_n \rightarrow_p A* \Rightarrow *A_n x_n \rightarrow_d A x* \text{ при условии, что размеры } A_n \text{ и } x_n \text{ согласованы. В частности, если } x \sim N(0, \Sigma), \text{ то } A_n x_n \rightarrow_d N(0, A \Sigma A').$$

$$(d) *x_n \rightarrow_d x, A_n \rightarrow_p A* \Rightarrow *x_n' A_n^{-1} x_n \rightarrow_d x' A^{-1} x* \text{ при условии, что размеры } A_n \text{ и } x_n \text{ согласованы, и } A \text{ невырождена.}$$

Части (а) и (с) иногда называют **теоремой Slutsky's Theorem**). Если положить $\alpha = 0$, то из части (а) вытекает:

$$*x_n \xrightarrow{d} x, y_n \xrightarrow{p} 0* \Rightarrow *x_n + y_n \xrightarrow{d} x*. \quad (2.1.7)$$

То есть если $z_n = x_n + y_n$ и $y_n \rightarrow_p 0$ (то есть если $z_n - x_n \rightarrow_d 0$), то тогда асимптотическое распределение z_n то же самое, что и у x_n . Если $z_n - x_n \rightarrow_d 0$, мы иногда (но не всегда) говорим, что эти две последовательности **асимптотически эквивалентны** (*asymptotically equivalent*) и записываем это как

$$*z_n \sim x_n* \text{ или } *z_n = x_n + o_p*.$$

где o_p — некоторая подходящая случайная величина (здесь y_n), которая сходится к нулю по вероятности.

Стандартный трюк получения асимптотического распределения последовательности случайных величин заключается в нахождении асимптотически эквивалентной последовательности, асимптотическое распределение которой получить легче. В частности, заменяя в явности (1) скаляр y_n на $y_n - \alpha$, мы получаем:

$$*x_n \xrightarrow{d} x, y_n \xrightarrow{d} \alpha \Rightarrow *y'_n x_n \underset{n}{\sim} \alpha' x_n \text{ или } *y'_n x_n = \alpha' x_n + o_p* \quad (2.1.8)$$

Здесь o_p есть $(y_n - \alpha)' x_n$. Следовательно, замена y_n на его предел по вероятности не изменяет асимптотическое распределение $y'_n x_n$ при условии, что x_n сходится по распределению к некоторой случайной величине.

Третий результат позволяет нам тестировать нелинейные гипотезы при заданном асимптотическом распределении оценки.

Лемма 2.5 («дельта-метод»): Предположим, что $\{x_n\}$ — последовательность K -мерных случайных векторов, таких, что $x_n \rightarrow_p \beta$ и

$$\sqrt{n}(x_n - \beta) \xrightarrow{d} z,$$

и предположим, что $a(\cdot): \mathbb{R}^K \rightarrow \mathbb{R}^r$ имеет непрерывные первые производные в $A(\beta)$; обозначает $r \times K$ матрицу первых производных, вычисленных в β :

$$A(\beta) = \frac{\partial a(\beta)}{\partial \beta}, \quad r \times K.$$

Тогда

$$\sqrt{n}[a(x_n) - a(\beta)] \xrightarrow{d} A(\beta)z.$$

В частности:

$$*\sqrt{n}(x_n - \beta) \xrightarrow{d} N(0, \Sigma)* \Rightarrow *\sqrt{n}[a(x_n) - a(\beta)] \xrightarrow{d} N(0, A(\beta)\Sigma A(\beta)')*.$$

Доказательство этого является хорошим способом научиться тому, как использовать результаты, которые изучались до сих пор.

Доказательство. По теореме о среднем из математического анализа (см. параграф 7.3 по поводу формулировки теоремы) существует K -мерный вектор y_n между x_n и β , такой, что

$$a(x_n) - a(\beta) = \underset{r \times K}{A(y_n)} \underset{K \times 1}{(x_n - \beta)}.$$

Умножая обе части на $\sqrt{n}$, мы получаем:

$$\sqrt{n}[\alpha(x_n) - \alpha(\beta)] = A(y_n)\sqrt{n}(x_n - \beta).$$

Поскольку y_n находится между x_n и β и поскольку $x_n \rightarrow_p \beta$, мы знаем, что $y_n \rightarrow_p \beta$. Кроме того, первая производная $A(\cdot)$ непрерывна по предположению. Таким образом, по лемме 2.3(a):

$$A(y_n) \xrightarrow{p} A(\beta).$$

По лемме 2.4(c), этот факт и гипотеза о том, что $\sqrt{n}(x_n - \beta) \rightarrow_d z$, влекут за собой:

$$A(y_n)\sqrt{n}(x_n - \beta) (= \sqrt{n}[\alpha(x_n) - \alpha(\beta)]) \xrightarrow{d} A(\beta)z. \quad \blacksquare$$

Взгляд на оценки как на последовательности случайных величин

Пусть $\hat{\theta}_n$ — оценка вектора параметров θ , основанная на выборке размера n . Последовательность $(\hat{\theta}_n)$ является примером последовательности случайных величин, так что понятия, введенные в этом параграфе для последовательностей случайных величин, применимы к $\{\hat{\theta}_n\}$. Мы говорим, что оценка $\hat{\theta}_n$ является **состоятельной для θ** (*consistent for θ*), если

$$\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta \quad \text{или} \quad \hat{\theta}_n \xrightarrow{p} \theta.$$

Асимптотическое смещение (*asymptotic bias*) $\hat{\theta}_n$ определяется как⁵ $\lim_{n \rightarrow \infty} \hat{\theta}_n - \theta$. Таким образом, если оценка состоятельна, ее асимптотическое смещение равно нулю. Состоятельная оценка $\hat{\theta}_n$ является **асимптотически нормальной** (*asymptotically normal*), если

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \Sigma).$$

Такая оценка называется **$\sqrt{n}$ -состоятельной** (*$\sqrt{n}$ -consistent*). Акроним, иногда используемый для «состоятельности и асимптотической нормальности», — **CAN**. Дисперсионная матрица Σ называется **асимптотической дисперсией** (*asymptotic variance*) и обозначается как $\text{Avar}(\hat{\theta}_n)$. Некоторые авторы используют обозначение $\text{Avar}(\hat{\theta}_n)$ для среднего Σ/n (которое также нулю в пределе). В этой книге $\text{Avar}(\hat{\theta}_n)$ — это дисперсия предельного распределения $\sqrt{n}(\hat{\theta}_n - \theta)$.

⁵ Некоторые авторы используют термин «асимптотическое смещение» иным образом (Аллен [1985], например, определяет его для обозначения $\lim_{n \rightarrow \infty} n(\hat{\theta}_n - \theta)$).

Законы больших чисел и центральные предельные теоремы

Для последовательности случайных скаляров $\{z_n\}$ **выборочное среднее** (*sample mean*) $\bar{z}_n$ определяется как

$$\bar{z}_n = \frac{1}{n} \sum_{i=1}^n z_i.$$

Рассмотрим последовательность $\{\bar{z}_n\}$. Закон больших чисел (*Law of large numbers* — **LLN**) относится к условиям, при которых $\{z_n\}$ сходится или по вероятности, или почти наверное. LLN называется усиленным, если имеет место сходимость почти наверное, и слабым, если имеет место сходимость по вероятности. Из части (а) леммы 2.2 мы можем легко получить следующий слабый LLN.

Версия слабого LLN Чебышева:

$$\text{Если } \lim_{n \rightarrow \infty} E(\bar{z}_n) = \mu, \quad \lim_{n \rightarrow \infty} \text{Var}(\bar{z}_n) = 0 \Rightarrow \bar{z}_n \xrightarrow{p} \mu.$$

Это выполняется потому, что при данном условии легко доказать (см. аналитический вопрос), что $\bar{z}_n \rightarrow_{\text{p.s.}} \mu$. Следующий усиленный LLN предполагает, что $\{z_i\}$ является i.i.d.-последовательностью (i.i.d. — независимые и одинаково распределенные), но дисперсия необязательно должна быть конечной.

Второй усиленный закон больших чисел Колмогорова: Пусть последовательность $\{z_i\}$ есть i.i.d. с $E(z_i) = \mu^1$. Тогда $\bar{z}_n \rightarrow_{\text{p.s.}} \mu$.

Эти LLNs легко обобщаются на случайные векторы требованием поэлементной сходимости.

Центральные предельные теоремы (*Central Limit Theorems* — **CLT**) касаются предельного поведения разности между $\bar{z}_n$ и $E(\bar{z}_n)$ (которое равно $E(z_i)$, если $\{z_i\}$ является i.i.d.), увеличенной в $\sqrt{n}$ раз. Единственная центральная предельная теорема, которая нам нужна в случае последовательности i.i.d., такова:

CLT Линдберга — Леви: Пусть $\{z_i\}$ есть i.i.d. с $E z_i = \mu$ и $\text{Var} z_i = \Sigma$. Тогда

$$\sqrt{n}(\bar{z}_n - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (z_i - \mu) \xrightarrow{d} N(0, \Sigma).$$

¹Таким образом, среднее существует и конечно (действительное число). Когда мы скажем на некоторый момент (например, среднее), как здесь, то косвенно предположится, что этот момент существует и конечен.

Это читается так: последовательность случайных векторов $\{\sqrt{n}(\bar{z}_n - \mu)\}$ сходится по распределению к случайному вектору, некоему распределению $N(0, \Sigma)$. (Обычно CLT Линдберга—Левя используется для последовательности скалярных случайных величин. Векторная версия, показанная выше, выводится из скалярной версии следующим образом. Пусть $\{z_i\}$ есть i.i.d. $z_i \in E(z_i) = \mu$ и $\text{Var } z_i = \Sigma$ и пусть λ есть любой вектор действительных чисел той же самой размерности. Тогда $\{\lambda' z_n\}$ является последовательностью скалярных случайных переменных с $E(\lambda' z_n) = \lambda' \mu$ и $\text{Var } \lambda' z_n = \lambda' \Sigma \lambda$. Из скалярной версии Линдберга—Левя вытекает тогда, что

$$\sqrt{n}(\lambda' \bar{z}_n - \lambda' \mu) = \lambda' \sqrt{n}(\bar{z}_n - \mu) \rightarrow_d N(0, \lambda' \Sigma \lambda).$$

Но это предельное распределение есть распределение $\lambda' x$, где $x \sim N(0, \Sigma)$. Таким образом, по теореме о многомерной сходимости по распределению, сформулированной выше, $\{\sqrt{n}(\bar{z}_n - \mu)\} \rightarrow_d x$, что и утверждается в векторной версии Линдберга—Левя.)

Контрольные вопросы

1. (Общая сходимость почти по вероятности.) Последовательность действительных чисел является тривиальным примером последовательности случайных величин. Верно ли, что $\lim_{n \rightarrow \infty} z_n = a \Leftrightarrow \text{pr}(\lim_{n \rightarrow \infty} z_n = a) = 1$? **Указание:** Посмотрите на определение $\text{pr}(\lim_{n \rightarrow \infty} z_n = a)$. [Знаете ли вы, что для достижимой близости ϵ ,
2. (Альтернативное определение сходимости для последовательностей векторов.) Проверьте, что данное в тексте определение $\{z_n\} \rightarrow_d z$ эквивалентно

$$\lim_{n \rightarrow \infty} E[(z_n - z)'(z_n - z) | z_n = z] = 0.$$

Указание: $E[(z_n - z)'(z_n - z) | z_n = z] = E[(z_{n1} - z_1)^2 | z_{n1} = z_1] + \dots + E[(z_{nK} - z_K)^2 | z_{nK} = z_K]$, где K — размерность z .

Аналогично проверьте, что данное в тексте определение $\{z_n\} \rightarrow_p a$ эквивалентно

$$\lim_{n \rightarrow \infty} \text{Prob}(\|z_n - a\| | z_n = a) > \epsilon) = 0 \text{ для всех } \epsilon > 0.$$

3. Докажите лемму 2.4(c), используя пункты (a) и (b) этой леммы. **Указание:** $A_n x_n = (A_n - A)x_n + Ax_n$. Из (b) $(A_n - A)x_n \rightarrow_p 0$.

4. Предположим, что $\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow_d N(0, \sigma^2)$. Следует ли из этого, что $\hat{\theta}_n \rightarrow_p \theta$? **Указание:**

$$\hat{\theta}_n - \theta = \frac{1}{\sqrt{n}} \cdot \sqrt{n}(\hat{\theta}_n - \theta), \quad \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} = 0.$$

- б) (Объедините дефати-метод с Линдбергом — Левин) Пусть $\{z_i\}$ — последовательность i.i.d. (независимых и одинаково распределенных) случайных величин с $E(z_i) = \mu \neq 0$ и $Var(z_i) = \sigma^2$, и пусть $\bar{z}_n$ — выборочное среднее. Покажите, что

$$\sqrt{n} \left(\frac{1}{\bar{z}_n} - \frac{1}{\mu} \right) \rightarrow N \left(0, \frac{\sigma^2}{\mu^4} \right)$$

Указание: В лемме 2.5 положите $\beta = \mu$, $a(\beta) = 1/\mu$, $\mathbf{z}_n = \bar{z}_n$.

2.2. Фундаментальные понятия в анализе временных рядов

В этом параграфе мы определяем самые основные понятия в анализе временных рядов, которые будут неотъемлемой частью нашего языка. Основным понятием является **случайный процесс** (*stochastic process*), который является только красивым названием последовательности случайных величин. Если индекс случайных величин интерпретируется как представляющий время, случайный процесс называется **временным рядом** (*time series*). Если $\{z_i\}$ ($i = 1, 2, \dots$) является случайным процессом, его **реализация** (*realization*) или **выборочная траектория** (*sample path*) приписывает каждому i некоторое допустимое значение z_i . Таким образом, реализация $\{z_i\}$ есть последовательность действительных чисел. Мы будем часто использовать термин «временной ряд» и для обозначения реализации процесса, и для обозначения процесса, реализацией которого временной ряд является.

Необходимость эргодической стационарности

Фундаментальной проблемой в анализе временных рядов является то, что мы можем наблюдать реализацию процесса только один раз. Например, выборка годового темпа инфляции США за период с 1946 по 1995 год является строкой из 50 определенных чисел, являющейся только одним из возможных результатов лежащего в основе случайного процесса для темпа инфляции. Если бы история пошла другим путем, мы бы получили другую выборку. Если бы мы могли наблюдать историю много раз, мы могли бы собрать много выборок, каждая из которых содержит, возможно, другую строку из 50 чисел. Среднее темпа инфляции для, скажем, 1995 года можно тогда оценить, беря среднее темпа инфляции 1995 года (50-й элемент строки) по этим выборкам. Популяционное среднее, оцененное таким образом, иногда называется **средним по ансамблю** (*ensemble mean*). На языке общей теории вероятностей в экономике среднее по ансамблю является средним по всем возможным состояниям природы для любого заданного календарного времени.

Конечно, невозможно наблюдать множество различных альтернативных историй. Но если распределение темпа инфляции остается неизменным (на это свойство будем ссылаться как на **стационарность** (*stationarity*)), определенную строку из 50 чисел, которую мы наблюдаем, можно рассматривать как 50 различных значений из *одного и того же* распределения. Кроме того, если процесс не слишком инерционен (такое свойство имеет то, что называется **эргодичностью** (*ergodicity*)), каждый элемент строки будет содержать некоторую информацию, не доступную из других элементов. И, как показано ниже, среднее по времени по всем элементам единственной строки будет соответствовать среднему по ансамблю).

Различные классы случайных процессов¹

Стационарные процессы

Случайный процесс $\{z_i\}$ ($i = 1, 2, \dots$) является (строго) стационарным, если для любого заданного конечного целого r и для любого множества индексов $i_1, i_2, \dots, i_r$ совместное распределение $(z_{i_1}, z_{i_2}, \dots, z_{i_r})$ зависит только от $i_1 - i, i_2 - i, \dots, i_r - i$, но не от i . Например, совместное распределение (z_1, z_5) то же самое, что и для (z_{10}, z_{15}) . Что является важным для распределения — это относительное местоположение в последовательности. В частности, распределение z_i не зависит от абсолютного местоположения z_i , то есть i . Таким образом, среднее, дисперсия и другие более высокие моменты, если они существуют, остаются одинаковыми. Из этого и очевидно также следует, что любое преобразование (функция) стационарного процесса z само стационарно, то есть если $\{z_i\}$ является стационарным, то и $\{f(z_i)\}$ также². Например, последовательность $z_i z_i'$ стационарна, если стационарна последовательность z_i .

Пример 2.1 (i.i.d.-последовательности): Последовательность независимых и одинаково распределенных случайных величин является стационарным процессом, который не проявляет сериальной зависимости.

Пример 2.2 (ряд из одинаковых констант): Извлеките z_1 из некоторого распределения, а затем положите $z_i = z_1$ ($i = 2, 3, \dots$). Таким образом, значение процесса заморожено начальной датой. Так построенный процесс $\{z_i\}$ является стационарным процессом, проявляющим максимальную сериальную зависимость.

¹ Многие из понятий, приведенных в этом параграфе, также можно найти в Section 4.7 книги [Davidson and MacKinnon, 1993].

² Функция $f(\cdot)$ должна быть «измерима», так чтобы $f(z_i)$ была корректно определенной случайной величиной. Любая не прерывная функция является измеримой. В дальнейшем мы не утруждаем себя в добавлении определения «измеримой», когда функция f от случайной величины понимается как случайная величина.

Очевидно, что если векторный процесс $\{x_t\}$, построенный таким образом, является стационарным процессом, то тогда каждый момент этого вектора образует одномерный стационарный процесс. Однако обратное утверждение неверно.

Пример 2.3 (противопоставление поэлементной и совместной стационарности): Пусть $\{z_t\}$ ($t = 1, 2, \dots$) — скалярный ИД-процесс. Построим на его основе двумерный процесс $\{x_t\}$, определяя $x_{1t} = z_t$ и $x_{2t} = \delta z_t$. Скалярный процесс $\{z_t\}$ стационарный (это процесс из примера 2.1). Скалярный процесс $\{\delta z_t\}$ также стационарный (процесс из примера 2.2). Векторный процесс $\{x_t\}$, однако, не является (совместно) стационарным, поскольку (совместное) распределение $x_1' = (z_1, \delta z_1)'$ отличается от распределения $x_2' = (z_2, \delta z_2)'$.

Многие зарегистрированные временные ряды, такие как БВП, не являются стационарными, поскольку они проявляют временный тренд. Более очевидным примером нестационарности являются междугородние обменные курсы, которые, как утверждается, имеют возрастающую дисперсию. Но многие временные ряды с трендом могут быть редуцированы к стационарным процессам. Процесс называется **стационарным относительно тренда** (*trend stationary*), если он является стационарным после удаления из него (обычно линейной) функции времени (которое представлено индексом t). Если процесс не является стационарным, но его первая разность, $\Delta z_t = z_t - z_{t-1}$, стационарна, то процесс $\{z_t\}$ называется **стационарным в разностях** (*differenced stationary*). Стационарные относительно тренда и стационарные в разностях процессы будут рассмотрены в главе 9.

Совместная стационарность и ковариация (слабо-стационарные) процессы

Случайный процесс $\{x_t\}$ является **слабо (или в ковариациях, ковариационно) стационарным** (*weakly (или covariance) stationary*), если

- (i) $E(x_t)$ не зависит от t ;
- (ii) $Cov(x_t, x_{t+h})$ существует, конечна и зависит только от h , но не от t (например, $Cov(x_1, x_2)$ равна $Cov(x_{12}, x_{13})$).

Для среднего и ковариации ковариационно-стационарного процесса имеют значение не абсолютные, а относительные жесткие значения и последовательности¹. Очевидно, что если исследователя есть интерес (строго) стационарной и если дисперсия и ковариации конечны, то тогда эти последовательности являются слабо стационарными (отсюда и термин «стро-

¹ Для среднего слабого процесса жесткое значение не имеет никакого значения. — *Иван, когда вы вернетесь?*

нале). Пример стационарного ковариационного не строго стационарного процесса будет дан ниже, в примере 2.4.

Автоковариация j -го порядка (j -th order autocovariance), обозначаемая Γ_j , определяется как

$$\Gamma_j = \text{Cov}(x_t, x_{t-j}) \quad (j = 0, 1, 2, \dots).$$

Термин «авто» объясняется тем, что входящие в определение две случайные величины взяты из одного и того же процесса. Γ_0 не зависит от t (поскольку стационарность ковариаций). Также из стационарности ковариаций Γ_j удовлетворяет

$$\Gamma_j = \Gamma_j^* \quad (2.2.1)$$

(Обозначение этого является контрольным вопросом, приведенным ниже.) Ковариация 0-го порядка является дисперсией $\Gamma_0 = \text{Var}(x_t)$. Процессы в примерах 2.1 и 2.2 являются стационарными ковариационными, если эта дисперсия существует и конечна. Для процесса из примера 2.1 Γ_j является дисперсией распределения, и $\Gamma_j = 0$ для $j \neq 0$. Для процесса из примера 2.2 $\Gamma_j = \Gamma_j^*$.

Для скалярного стационарного ковариационного процесса $\{x_t\}$ автоковариация j -го порядка является скаляром. Если γ_j является этой ковариацией, то удовлетворяет

$$\gamma_j = \gamma_{-j} \quad (2.2.2)$$

Возьмем цепочку из n последовательных значений $(x_1, x_2, \dots, x_{n-1})$ скалярного процесса. Вследствие стационарности ковариаций его $n \times n$ ковариационная матрица имеет той же формой, что и для $(x_1, x_2, \dots, x_n)$, и является матрицей линейного спектра (*band spectrum matrix*):

$$\text{Var}(x_1, x_2, \dots, x_{n-1}) = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \dots & \gamma_{n-2} \\ \vdots & \gamma_1 & \gamma_2 & \ddots & \vdots \\ \gamma_{n-2} & \dots & \gamma_1 & \gamma_0 & \gamma_1 \\ \gamma_{n-1} & \dots & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix}.$$

Она называется **автоковариационной матрицей (*autocovariance matrix*)** процесса. Коэффициент автокорреляции j -го порядка (j -th order autocorrelation coefficient), ρ_j , определяется как

коэффициент автокорреляции j -го порядка =

$$\equiv \rho_j = \frac{\gamma_j}{\gamma_0} = \frac{\text{Cov}(x_t, x_{t-j})}{\text{Var}(x_t)} \quad (j = 1, 2, \dots). \quad (2.2.3)$$

Для $j = 0$, $\rho_j = 1$. Функция корреляции $\{\rho_j\}$ при $j = 0, 1, 2, \dots$ называется **корреляционной (*correlation*)**.

Процессы белого шума

Очень важным классом слабо стационарных процессов является **процесс белого шума** (*white noise process*), то есть с нулевым средним и отсутствием корреляций корреляции:

стационарный в корреляциях процесс $\{z_t\}$ является **белым шумом** (*white noise*), если $E(z_t) = 0$ и $\text{Cov}(z_t, z_{t-j}) = 0$ для $j \neq 0$.

Ясно, что последовательность независимых и одинаково распределенных случайных величин (i.i.d.) с нулевым средним и конечной дисперсией является частным случаем процесса белого шума. По этой причине она называется **независимым процессом белого шума** (independent white noise process).

Пример 2.4 (процесс белого шума, не являющийся строго стационарным¹): Пусть ω — случайная величина, равномерно распределенная на интервале $[0, 1]$, определим

$$z_t = \cos(2\pi\omega) \quad (t = 1, 2, \dots).$$

Можно показать, что $E(z_t) = 0$, $\text{Var}(z_t) = 1/2$ и $\text{Cov}(z_t, z_s) = 0$ для $t \neq s$. Поэтому $\{z_t\}$ является белым шумом. Однако ясно, что это не независимый процесс белого шума. Он даже не строго стационарный.

Эргодичность

Говорят, что стационарный процесс $\{z_t\}$ является **эргодическим** (*ergodic*), если для любых двух ограниченных функций $f: \mathbb{R}^{k-1} \rightarrow \mathbb{R}$ и $g: \mathbb{R}^{k-1} \rightarrow \mathbb{R}$

$$\begin{aligned} \lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{t=1}^n f(z_1, \dots, z_{t-1}) g(z_t, \dots, z_{t+k-1}) \right| = \\ = E[f(z_1, \dots, z_{k-1})] E[g(z_1, \dots, z_{k-1})]. \end{aligned}$$

Эмпирически стационарный процесс является эргодическим, если он асимптотически независим, то есть если две случайные величины, расположенные в последовательности далеко друг от друга, почти независимо распределены. Стационарный процесс, который может быть эргодическим, будет называться **эргодически стационарным** (*ergodic stationary*). Эргодическая стационарность будет неотъемлемой частью разработки теории больших выборок из-за следующего свойства.

¹Заметим, что в [Билл и Экпортс 7.8 и Андерсон, 1971, с. 379].

Эргодическая теорема (см., например, Theorem 9.5.5 в [Karlin and Taylor, 1975]): Пусть $\{x_i\}$ является стационарным и эргодическим процессом с $E\{x_i\} = \mu$. Тогда

$$\bar{z}_n = \frac{1}{n} \sum_{i=1}^n x_i \xrightarrow{a.s.} \mu.$$

Таким образом, эргодическая теорема является существенным обобщением LLN Колмогорова. Серийная зависимость, которая исключается предположением i.i.d. в LLN Колмогорова, в эргодической теореме допускается при условии, что она исчезает в долгосрочной перспективе. Поскольку для любой (измеримой) функции $f(\cdot)$ процесс $\{f(x_i)\}$ является эргодически стационарным всякий раз, когда эргодически стационарен x_i , из этой теоремы вытекает, что любой момент стационарного и эргодического процесса (если он существует и конечен) состоятельно оценивается выборочным средним. Предположим, например, что процесс x_i является стационарным и эргодическим, и $E(x_i x_i')$ существует и конечно. Тогда среднее $\frac{1}{n} \sum_i x_i x_i'$ состоятельно для $E(x_i x_i')$.

Простейшим примером эргодически стационарного процесса является процесс независимого белого шума. (Процесс белого шума, в котором независимость ослаблена до отсутствия серийной корреляции, не обязательно эргодичен; см. пример 2.4 выше.) Другой важный пример — AR(1) — процесс, удовлетворяющий

$$x_i = \rho + \rho x_{i-1} + \varepsilon_i, \quad |\rho| < 1,$$

где $\{\varepsilon_i\}$ является независимым белым шумом.

Мартингалы

Пусть x_i является элементом $\mathcal{F}_i$. Скалярный процесс $\{x_i\}$ называется **мартингалом по отношению к $\{\mathcal{F}_i\}$** (*martingale with respect to $\{\mathcal{F}_i\}$*), если

$$E(x_i | \mathcal{F}_{i-1}, x_1, x_2, \dots, x_{i-1}) = x_{i-1} \quad \text{для } i \geq 2. \quad (2.2.4)$$

Множество $\{x_{i-1}, x_{i-2}, \dots, x_1\}$, стоящее в условии, часто называется **информационным множеством** (*information set*) в точке (на дату) $i-1$.

¹Так что среднее предполагается существующим и конечным.

²Если процесс начинается в бесконечном прошлом, так что i приобретает значения от $-\infty$ до ∞ , предельным значением является $E(x_i | x_{i-1}, x_{i-2}, \dots) = x_{i-1}$, и ограничение $i \geq 2$ не нужно. Начинается ли процесс с бесконечного прошлого или с даты $i-1$, для теории больших выборок, которая будет излагаться ниже, не имеет значения. Что будет иметь значение, так это то, что процесс начинается ранее периода выборки.

Процесс $\{x_i\}$ называется просто **мартингалом** (*martingale*), если информационное множество образовано его собственными прошлыми значениями $(x_{i-1}, x_{i-2}, \dots, x_1)$. Если z_i включает x_i , то $\{x_i\}$ является мартингалом, поскольку

$$\begin{aligned} E(x_i | x_{i-1}, x_{i-2}, \dots, x_1) &= \\ &= E[E(x_i | z_i) | x_{i-1}, x_{i-2}, \dots, x_1] \\ &\quad \text{(закон повторных математических ожиданий)} \\ &= E(x_{i-1} | x_{i-1}, x_{i-2}, \dots, x_1) = x_{i-1}. \end{aligned}$$

Векторный процесс $\{z_i\}$ называется **мартингалом**, если

$$E(z_i | z_{i-1}, \dots, z_1) = z_{i-1} \text{ для } i \geq 2. \quad (2.2.5)$$

Пример 2.5 (мартингальная гипотеза Холла (*Hall's Martingale Hypothesis*)): Пусть z_i — вектор, содержащий совокупность макроэкономических переменных (таких как предложение денег или ВВП), включая агрегированное потребление c_i за период i . Мартингальная гипотеза в [Hall, 1978] состоит в том, что потребление является мартингалом относительно z_i :

$$E(c_i | z_{i-1}, z_{i-2}, \dots, z_1) = c_{i-1}.$$

Это форма, имеет понятие теории, потребления, называемое «сглаживанием потребления»: потребитель, желающий избежать колебаний в уровне жизни, корректирует потребление на дату $i-1$ до такого уровня, чтобы не ожидать никаких изменений в последующем уровне потребления.

Случайные блуждания

Важным примером мартингалов является **случайное блуждание** (*random walk*). Пусть $\{g_i\}$ — векторный процесс независимого белого шума (таким образом, он является i.i.d. со средним 0 и конечной дисперсионной матрицей). Случайное блуждание, $\{z_i\}$, является последовательностью накопленных сумм:

$$z_1 = g_1, z_2 = g_1 + g_2, \dots, z_i = g_1 + g_2 + \dots + g_i, \dots \quad (2.2.6)$$

Если задана последовательность $\{z_i\}$, лежащую в ее основе последовательность независимого белого шума, $\{g_i\}$, можно восстановить, без потери точности:

$$g_1 = z_1, g_2 = z_2 - z_1, \dots, g_i = z_i - z_{i-1}, \dots \quad (2.2.7)$$

Таким образом, первая разность случайного блуждания является независимым белым шумом. Случайное блуждание является мартингалом.

поскольку

$$\begin{aligned}
 E(z_i | z_{i-1}, \dots, z_1) &= E(z_i | g_{i-1}, \dots, g_1) \\
 (\text{так как } \{z_{i-1}, \dots, z_1\} \text{ и } \{g_{i-1}, \dots, g_1\} \text{ содержат одинаковую} \\
 &\text{информацию, как мы только что видели}) \\
 &= E(g_i + g_{i-1} + \dots - g_i | g_{i-1}, \dots, g_1) \\
 &= E(g_i | g_{i-1}, \dots, g_1) - (g_{i-1} + \dots + g_1) = \\
 &= g_i + \dots - g_{i-1} (E(g_i | g_{i-1}, \dots, g_1) = 0 \\
 &\text{так как } \{g_i\} \text{ является независимым белым шумом}) = \\
 &= z_{i-1} \quad (\text{по определению } z_{i-1}).
 \end{aligned} \tag{2.2.8}$$

Последовательность мартингал-разностей

Векторный процесс $g_i \in E(g_i) = 0$ называется **последовательностью мартингал-разностей** (*martingale difference sequence, m.d.s.*) или **мартингал-разностью** (*martingale differences*), если математическое ожидание, условное относительно его прошлых значений, также равно нулю:

$$E(g_i | g_{i-1}, g_{i-2}, \dots, g_1) = 0 \quad \text{для } i \geq 2. \tag{2.2.9}$$

Процесс называется так потому, что накопленная сумма z_i , созданная на основе последовательности мартингал-разностей g_i , является мартингалом: доказательство этого аналогично (2.2.8). Обратное, если z_i — мартингал, то его первая разность, построенная, как в (2.2.7), является последовательностью мартингал-разностей.

Последовательность мартингал-разностей не имеет сериальной корреляции (то есть $\text{Cov}(g_i, g_{i-j}) = 0$ для всех i и $j \neq 0$). Доказательство этого утверждения проводится следующим образом.

Доказательство. Прежде всего заметим, что мы можем, без потери общности, предположить, что $j > 1$. Так как среднее нулевое, достаточно показать, что $E(g_i g'_{i-j}) = 0$. Поэтому рассмотрим это выражение, переписав его следующим образом:

$$\begin{aligned}
 E(g_i g'_{i-j}) &= \\
 &= E[E(g_i g'_{i-j} | g_{i-j})] \quad (\text{по закону полных математических ожиданий}) \\
 &= E[E(g_i | g_{i-j}) g'_{i-j}]
 \end{aligned}$$

(вследствие линейности условных математических ожиданий).

Далее, так как $j \geq 1$, $(g_{i-1}, \dots, g_{i-j}, \dots, g_1)$ включает g_{i-1} . Следовательно,

$$\begin{aligned} E(g_i | g_{i-j}) &= \\ &= E(F(g_i | g_{i-1}, \dots, g_{i-j}, \dots, g_1) | g_{i-j}) \\ &\text{(по закону повторных математических ожиданий)} \end{aligned}$$

0.

Последнее равенство вытекает отсюда, что

$$E(g_i | g_{i-1}, \dots, g_{i-j}, \dots, g_1) = 0.$$

ARCH-процессы

Примером картин зависимости, часто используемым в анализе доходностей акций, является процесс с **авторегрессионной условной гетероскедастичностью** (*autoregressive conditional heteroskedasticity* — **ARCH**), введенный в [Engle 1982]. Процесс $\{g_t\}$ называется **ARCH-процессом** порядка 1 (ARCH process of order 1) (ARCH(1)), если он может быть записан в виде:

$$g_t = \sqrt{\zeta + \alpha g_{t-1}^2} \cdot \varepsilon_t, \quad (2.2.10)$$

где $\{\varepsilon_t\}$ есть i.i.d. с нулевым средним и единичной дисперсией. Если g_1 — известное начальное значение процесса, мы можем использовать (2.2.10) для вычисления значений g_t . Например,

$$g_2 = \sqrt{\zeta + \alpha g_1^2} \cdot \varepsilon_2.$$

В более общем случае $g_t (t \geq 2)$ есть функция от g к $\{\varepsilon_2, \varepsilon_3, \dots, \varepsilon_t\}$. Следовательно, ε_t не зависит от $(g_1, g_2, \dots, g_{t-1})$. Тогда легко показать, что $\{g_t\}$ есть m.l.s., поскольку

$$\begin{aligned} E(g_t | g_{t-1}, g_{t-2}, \dots, g_1) &= \\ &= E(\sqrt{\zeta + \alpha g_{t-1}^2} \cdot \varepsilon_t | g_{t-1}, g_{t-2}, \dots, g_1) = \\ &= \sqrt{\zeta + \alpha g_{t-1}^2} \cdot E(\varepsilon_t | g_{t-1}, g_{t-2}, \dots, g_1) = \\ &= \sqrt{\zeta + \alpha g_{t-1}^2} \cdot E(\varepsilon_t) \quad (\text{так как } \varepsilon_t \text{ не зависит от } (g_1, g_2, \dots, g_{t-1})) \\ &= 0 \quad (\text{так как } E(\varepsilon_t) = 0). \end{aligned} \quad (2.2.11)$$

Аналогичным образом (2.2.10) приводит к

$$E(g_t^2 | g_{t-1}, g_{t-2}, \dots, g_1) = \zeta + \alpha g_{t-1}^2. \quad (2.2.12)$$

Поэтому условный второй момент (который равен условной дисперсии, так как $E(g_t | g_{t-1}, g_{t-2}, \dots, g_1) = 0$) является функцией от собственной

истории процесса. В этом смысле указанный процесс выражает *субстантивную условную гетероскедастичность* (*strict conditional heteroskedasticity*). Можно показать (см., например [Engle, 1982]), что такой процесс является строго стационарным и эргодическим, если $\alpha < 1$, при условии, что y_t возникает из соответствующего распределения. Если при условии, что процесс начинается в бесконечном прошлом. Если процесс y_t стационарный, то можно найти его *базисный* или *первый* момент. Если предположить математическое ожидание от обеих сторон (2.2.13) и замечая, что

$$E[y_t^2 | y_{t-1}, y_{t-2}, \dots, y_1] = E[y_t^2] + E[y_t^2 - y_t^2 | y_{t-1}, y_{t-2}, \dots, y_1],$$

если y_t стационарный, мы получаем:

$$E[y_t^2] = \zeta + \alpha E[y_t^2] \quad \text{для} \quad E[y_t^2] = \frac{\zeta}{1-\alpha}. \quad (2.2.14)$$

Если $\alpha > 0$, эта модель отражает характерное свойство, найденное для дождливостей: значения за фиксированные интервалы, как правило, имеют также более высокие значения. (Для более подробной информации об ARCH-процессах см., например [Hamilton, 1994, Section 21.1].)

Другая формулировка отсутствия сериальной зависимости

Очевидно, что процесс независимого белого шума является стационарной последовательностью картингал-разностей с конечной дисперсией. И, как уже было видно, последовательность картингал-разностей не имеет сериальной корреляции. Таким образом, у нас есть три формулировки отсутствия сериальной зависимости для стационарного и ковариационного процесса с нулевым средним. Они, в порядке ослабления, таковы:

$$\begin{aligned} (1) \text{ «} \{y_t\} \text{ независимый белый шум} \text{»} &\Rightarrow \\ \Rightarrow (2) \text{ «} \{y_t\} \text{ стационарный ф.с. с конечной дисперсией} \text{»} &\Rightarrow \quad (2.2.15) \\ \Rightarrow (3) \text{ «} \{y_t\} \text{ белый шум} \text{»}. \end{aligned}$$

Условие (1) более сильное, чем (2) и поэтому существуют процессы, удовлетворяющие (2), но не (1). Примером является ARCH(1)-процесс (2.2.10) с $\alpha < 1$. Рис. 2.1 показывает, как реализация процесса, удовлетворяющего (1), обычно отличается от реализации процесса, удовлетворяющего (2). Нижний график (а) изображает реализацию последовательности независимых и локально распределенных случайных величин со средним 0 и единичной дисперсией. Нижний график (б) изображает ARCH(1)-процесс (2.2.10) с $\zeta = 0.2$ и $\alpha = 0.8$ (так что безусловная дисперсия, $\zeta/(1-\alpha)$ равна единице, как и на графике (а)), где $\{y_t\}$ — последовательность y_t в (2.2.10) (т.е. мы не использовали y_t в части (а)).

рис. 2.1, так что знаки в (а) и (б) одинаковы в каждой точке. Ряд в (б) в целом менее волатилен, чем в (а), но в некоторых точках он намного более волатилен. Тем не менее этот ряд стационарный.

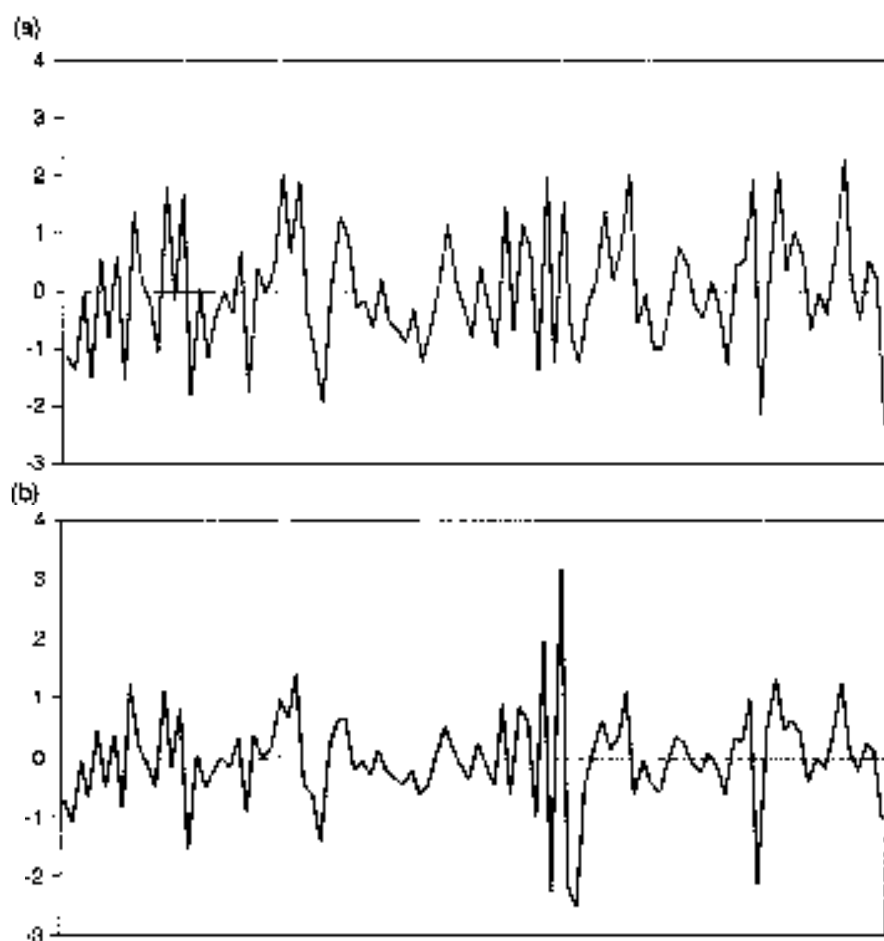


Рис. 2.1. Графики сериально некоррелированных временных рядов: (а) последовательность i.i.d. $N(0, 1)$, (б) ARCH(1) с шумами, взятыми из (а)

Условие (2) более сильное, чем (3): процесс в примере 2.4, являющийся большим шумом, но (как вы покажете в контрольном вопросе) он не удовлетворяет (2).

***CLT для эргодических стационарных
последовательностей мартингал-разностей***

Следующая CLT обобщает CLT Линдберга — Левы на стационарную и эргодическую m.d.s.

СЛТ для эргодических стационарных мартингал-разностей [Billingsley, 1961]: Пусть $\{g_t\}$ — векторная последовательность мартингал-разностей, которая является стационарной и эргодической с $E(g_t g_t') = \Sigma^1$, и пусть $\bar{g} \equiv \frac{1}{n} \sum_{t=1}^n g_t$. Тогда

$$\sqrt{n} \bar{g} = \frac{1}{\sqrt{n}} \sum_{t=1}^n g_t \xrightarrow{d} N(0, \Sigma).$$

В отличие от СЛТ Линдберга — Левы, здесь нет необходимости вычитать среднее из g_t , поскольку безусловное среднее п.д.с. по определению равно нулю. По той же причине Σ также равен $\text{Var}(g_t)$. Эта СЛТ, будучи применимой не только к последовательности i.i.d., но и к стационарным мартингал-разностям, таким как ARCH(1)-процессы, является более общей, чем СЛТ Линдберга — Левы.

Мы представили LLN для серийно коррелированных процессов в форме эргодической теоремы. Центральная предельная теорема для серийно коррелированных процессов будет представлена в главе 6.

Контрольные вопросы

1. Докажите, что $\Gamma_t = \Gamma_t'$. **Указание:**

$$\text{Cov}(z_t, z_{t-j}) = E[(z_t - \mu)(z_{t-j} - \mu)']$$

где $\mu = E(z_t)$. В силу стационарности и ковариаций,

$$\text{Cov}(z_t, z_{t-j}) = \text{Cov}(z_{t+j}, z_t).$$

2. (Прогнозирование белого шума.) Для процесса белого шума в примере 2.4 $E(z_t) = 0$. Каково значение $E(z_t, z_t)$ для $t > 2$? **Указание:** Мы должны быть в состоянии прогнанизовать будущее точно, если знаем значение z_1 . Является ли процесс m.d.s.? [Ответ: Нет.]

3. (Отсутствие ожидаемых изменений в мартингалах.) Предположим, что $\{z_t\}$ — мартингал относительно $\{z_t\}$. Покажите, что

$$E(z_{t+j} | z_{t-1}, z_{t-2}, \dots, z_t) = z_{t-1} \text{ и } E(z_{t+j+1} - z_{t+j} | z_{t-1}, z_{t-2}, \dots, z_t) = 0$$

для $j = 0, 1, \dots$. **Указание:** Используйте закон повторных математических ожиданий.

4. Пусть $\{z_t\}$ — последовательность действительных чисел, которые изменяются по i , и $\{z_t\}$ — последовательность i.i.d. случайных величин со средним 0 и конечной дисперсией. Является ли процесс $\{z_t, z_t\}$ i.i.d.? [Ответ: Нет.] Является ли он серийно независимым? [Ответ: Да.] А m.d.s.? [Ответ: Да.] Стационарным? [Ответ: Нет.]

¹Так как исследователи ют, $\{g_t\}$ стационарна, эта матрица поперечных моментов не зависит от t . Также поскольку матрица поперечных моментов указана, то подразумевается, что все поперечные моменты существуют и конечны.

5. Покажите, что случайные блуждания нестационарны. **Указание:** Проверьте дисперсию.
6. (Первая разность марков для процесса машинно-дизернестного.) Пусть $\{x_i\}$ — мартигал. Покажите, что процесс $\{y_i\}$, построенный по (2.2.7), является m.d.s. Укажите $(y_1, \dots, y_i)$ и $(x_1, \dots, x_i)$ используют одну и ту же информацию.
7. (m.d.s. не являющаяся независимым белым шумом.) Пусть $y_i = \varepsilon_i + \varepsilon_{i-1}$, где $\{\varepsilon_i\}$ — процесс независимого белого шума. Очевидно, что $\{y_i\}$ — m.d.s. Убедитесь, что $\{y_i\}$ ($i = 2, 3, \dots$) есть m.d.s. **Указание:**

$$(y_1, y_2, \dots, y_i) = (E(\varepsilon_1), \varepsilon_1 + E(\varepsilon_2), \varepsilon_1 + \varepsilon_2 + E(\varepsilon_3), \dots, \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_{i-1} + \varepsilon_i).$$

8. (Перемножен попарной является m.d.s.) Пусть $\{y_i\}$ — процесс такой, что $E(y_1, y_2, \dots, y_i)$ существует и конечно, и определит

$$r_i = E(y_1, y_2, \dots, y_i) = E(y_i, y_1, y_2, \dots, y_i).$$

Тем же образом, r_{i+1} есть наименьшее ожиданий при добавлении еще одного слагаемого к информационному множеству. Покажите, что $\{r_i\}$ ($i \geq 2$) есть m.d.s.

9. (CLT Заллэнгадан сильнее CLT Линдберга—Левя) Пусть $\{x_i\}$ — последовательность i.i.d. с $E(x_i) = \mu$ и $\text{Var}(x_i) = \Sigma$, как в CLT Линдберга—Левя. Используйте CLT для мартигал разностей, чтобы доказать утверждение CLT Линдберга—Левя, а именно что $\sqrt{n}(\bar{x}_n - \mu) \rightarrow_d N(0, \Sigma)$. **Указание:** $\{\bar{x}_n - \mu\}$ является процессом независимого белого шума и, следовательно, эргодически стационарной m.d.s.

2.3. Распределение OLS-оценки на больших выборках

Важность в эконометрике процедуры OLS, первоначально разработанной для классической регрессионной модели (глава 1), заключается в том, что она имеет хорошие асимптотические свойства для класса моделей, которые полезны в экономике, но отличны от классической. Среди этих моделей модель, представленная в этом параграфе, имеет, вероятно, наиболее широкий диапазон экономических приложений. Для получения асимптотического распределения OLS-оценки не требуется никаких специфических предположений о распределении (таких как нормальность ошибок). Требования теории конечных выборок, чтобы регрессоры были строго экзогенными или «фиксированными», заменяется ее более слабое требование того, чтобы они были «предопределенными» (Для полноты изложения, в приложении развивается параллельная асимптотическая теория для модели с «фиксированными» регрессорами.)

Модель

Мы используем термин «процесс порождения данных» (*data generating process* — DGP) для случайного процесса, порождающего конечную выборку (y, X) . Следовательно, если мы специфицируем DGP, то можно определить совместное распределение конечной выборки (y, X) . В теории конечных выборок, где размер выборки фиксированный и конечный, мы определяем модель как множество совместных распределений (y, X) . В теории больших выборок модель формулируется как множество DGP. Модель, которую мы рассматриваем, — это множество DGP, удовлетворяющих следующей совокупности предположений.

Предположение 2.1 (линейность):

$$y_i = x_i' \beta + \varepsilon_i, \quad (i = 1, 2, \dots, n).$$

где x_i есть K -мерный вектор объясняющих переменных (регрессоров), β — K -мерный вектор коэффициентов и ε_i — невозмущаемая ошибка.

Предположение 2.2 (эргодическая стационарность): $(K+1)$ -мерный векторный случайный процесс $\{y_i, x_i\}$ является совместно стационарным и эргодическим.

Предположение 2.3 (предопределенные регрессоры): Все регрессоры являются *предопределенными* (*predetermined*) в том смысле, что они ортогональны текущей ошибке. $E(x_i \varepsilon_i) = 0$ для всех i и j ($i = 1, 2, \dots, K$)¹. Это может быть записано как

$$E[x_i \cdot (y_i - x_i' \beta)] = 0 \quad \text{или, эквивалентно,} \quad E(y_i) = 0.$$

где $y_i \equiv x_i' \beta + \varepsilon_i$.

Предположение 2.4 (ранговое условие): $K \times K$ матрица $E(x_i x_i')$ является невырожденной (и, следовательно, конечной). Мы обозначаем эту матрицу Σ_{xx} .

Предположение 2.5 (y_i является последовательностью мартингал-разностей с конечными вторыми моментами): $\{y_i\}$ есть последовательность мартингал-разностей (относительно последовательности F_j), $E(y_i) = 0$. Матрица

¹ Наше определение термина «предопределенность» неуклюже. Некоторые авторы полагают, что регрессоры предопределены, если $E(x_{i+j} \varepsilon_i) = 0$ для всех $j \geq 0$, а не только для $j = 0$. То есть ошибка ортогональна не только текущему, но и будущим значениям регрессора. В [Зоргева и др. Найд, 1993] регрессоры называются предопределенными, если ε_i не зависит от x_{i+j} для всех $j \geq 0$. Наше определение такое же, как в [Nagrich, 1994].

размера $K \times K$ перекрестных моментов, $E(g_i g_j)$, чепухажена. Мы используем S для $A_{\text{var}}(g)$ (дисперсия асимптотического распределения $\sqrt{n} \bar{g}$, где $\bar{g} = \frac{1}{n} \sum_i g_i$). По предположению 2.2 и CLT для эргодической стационарной мартигал-разности $S = E(g_i g_i')$.

Первое предположение — это про то, как риньявание предположение 1.1. Оставшиеся предположения требуют длительных комментариев.

- (Эргодическая стационарность.) Тривиальный, но важный специальный случай эргодической стационарности состоит в том, что $\{y_i, x_i\}$ есть i.i.d., то есть выборка является случайной¹. В большинстве существующих микроданных о домохозяйствах являются случайными выборками, с наблюдениями, случайно выбранными из генеральной совокупности национальных домохозяйств. Таким образом, мы ни в коем случае не исключаем модели, которые используют кросс-секционные данные.
- (Модель адаптирована к условной гетероскедастичности.) Если $\{g_i, x_i\}$ стационарный, то ошибка $\varepsilon_i = y_i - x_i' \beta$ также стационарна. Поэтому из предположения 2.2 вытекает, что безусловный второй момент $E(x_i^2)$ — если он существует и конечен — постоянен по i . То есть ошибка безусловно гомоскедастична. Тем не менее ошибка может быть условно гетероскедастичной — условный второй момент, $E(\varepsilon_i^2 | x_i)$, может зависеть от x_i . Пример, в котором ошибка гомоскедастична безусловно, но не условно, включен в параграф 2.6, где будут изучены последствия наложения на модель условной гомоскедастичности (что $E(\varepsilon_i^2 | x_i) = \sigma^2$).
- ($E(x_i \cdot \varepsilon_i) = 0$ против $E(\varepsilon_i | x_i) = 0$.) Иногда вместо условия ортогональности $E(x_i \cdot \varepsilon_i) = 0$ предполагается, что ошибки не смещены в том смысле, что $E(\varepsilon_i | x_i) = 0$. Последнее условие более строгое, чем условие ортогональности, поскольку предполагает, что для любой (измеримой) функции f от x_i , $f(x_i)$ ортогональна ε_i :

$$E[f(x_i) \varepsilon_i] = E[E(f(x_i) \varepsilon_i | x_i)] = E[f(x_i) E(\varepsilon_i | x_i)] = 0.$$

Это более строгое условие удовлетворяется в моделях рациональных ожиданий, но для целей разработки асимптотической теории нам нужно только более слабое предположение условия ортогональности.

¹ На самом деле, как только сделаны предположения о независимости, то же самое результаты для больших выборок могут быть доказаны для более общего случая, в котором $\{y_i, x_i\}$ независимы, но не одинаково распределены (i.i.d.), при условии, что они удовлетворяют некоторым условиям на более высокие моменты совместного распределения (y_i, x_i) . Мы не будем рассматривать это обобщение, потому что предположение i.i.d. мы стараемся в большинстве случаев выбирать микроданных.

- (Проблема эндогенности строгих экзогенных регрессоров.) От регрессора не требуется быть строго экзогенным. Как мы отмечали в параграфе 1.1, в мультирегрессионной экзогенности (предположение 1.2) важно, что для k -го регрессора $E(x_{it}y_{it}) = 0$ для всех i и j , а не только для $i = j$, что исключает возможность того, что текущая ошибка u_{it} коррелирована с будущими регрессорами x_{it+j} для $j \geq 1$. Предположение 2.3, ограничивавшее только однократными связями каждую ошибку и регрессоры, не исключает этой возможности. Например, AR(1)-процесс, который не удовлетворяет предположению экзогенности классической регрессионной модели, может быть вменен в модель этой главы. Это более слабое предположение о предопределенных регрессорах будет позже ослаблено в следующей главе.
- (Ранговое условие как отсутствие мультиколлинеарности в пределе.) Так как $E(x_{it}x_{it}')$ согласно предположению 2.4, $\lim_{n \rightarrow \infty} S_{xx} = \frac{1}{n} \mathbf{X}'\mathbf{X}$ (где $S_{xx} \equiv \frac{1}{n} \sum_{i=1}^n x_{it}x_{it}'$) с вероятностью единицы по ординарской мере. Поэтому для достаточно больших n выборочный перекастный момент регрессоров S_{xx} , который может быть записан как $\frac{1}{n} \mathbf{X}'\mathbf{X}$, невырожден, по предположениям 2.2 и 2.4. Так как $\frac{1}{n} \mathbf{X}'\mathbf{X}$ невырожден, если и только если $\text{rank}(\mathbf{X}) = K$, предположение 1.3 (отсутствие мультиколлинеарности) удовлетворяется с вероятностью единицы для достаточно больших n . В OLS-формуле $b = S_{xx}^{-1} S_{xy}$ (где $S_{xy} \equiv \frac{1}{n} \sum_{i=1}^n x_{it}y_{it}$) S_{xx} необходимо обращать. Если S_{xx} вырождена на конечных выборках (так что она не может быть обращена), мы просто присваиваем b произвольное значение так, чтобы OLS-оценка была определена для любой выборки.
- (Достаточное условие для того, чтобы $\{y_t\}$ была m.i.s.) Так как m.i.s. (независимость мультиинкрементной) имеет нулевое среднее по ординарному предположению 2.5 более строгое, чем предположение 2.3. Предположение 2.5 будет нужно нам, чтобы доказать асимптотическую нормальность OLS-оценки. Предположение о произведении регрессора и ошибки может быть сложно интерпретировать. Достаточное условие, которое интерпретировать легче, таково:

$$E(x_1|x_2, \dots, x_{K-1}, x_{K+1}, \dots, x_{K+L}, x_{K+L+1}, \dots, x_{K+L+M}) = 0. \quad (2.3.1)$$

Заметим, что в идентификации множество включено как экзогенное, так и эндогенные регрессоры. Из этого условия вытекает, что ошибка ординарно некоррелирована, а также не коррелирована с текущими и прошлыми регрессорами (доказательство очень похоже на доказательство в предыдущем параграфе того, что m.i.s.

серьезно некоррелирована). То, что (2.3.1) достаточно для того, чтобы $\{g_i\}$ был п.д.с., можно увидеть следующим образом. Мы имеем

$$\begin{aligned} E(g_i | g_{i-1}, \dots, g_1) &= \\ &= E[E(g_i | \varepsilon_1, \varepsilon_2, \dots, \varepsilon_i, x_1, x_2, \dots, x_i) | g_{i-1}, \dots, g_1]. \end{aligned}$$

Это выражение является повторных математических ожиданий, поскольку во внутреннем информационном множестве $\{\varepsilon_1, \dots, \varepsilon_i, x_1, \dots, x_i, x_{i+1}, \dots, x_n\}$ содержится больше информации, чем во внешнем информационном множестве $\{g_1, \dots, g_i\}$. Следовательно,

$$\begin{aligned} E(g_i | g_1, \dots, g_i) &= \\ &= E(x_i | E(\varepsilon_1 | \varepsilon_2, \dots, \varepsilon_i, x_1, x_2, \dots, x_i) | g_{i-1}, \dots, g_i) \\ &\quad (\text{в силу линейности условных математических ожиданий}) \\ &= 0 \quad (i \geq (2.3.1)) \end{aligned} \quad (2.3.2)$$

- (Когда регрессоры включают константу.) Практически во всех приложениях регрессоры включают константу. Если регрессоры включают константу, так что $x_{i0} = 1$ для всех i , то предположение 2.5 и преобразованных регрессоров может быть сформулировано более привычным образом: средние ошибки нулевые (что вытекает из условия $E(x_{ik}x_{i0}) = 0$ для $k = 1$), одномерная корреляция между какой-либо переменной и регрессорами нулевая (что вытекает из условий $E(x_{ik}x_{i0}) = 0$ для $k \neq 1$ и $E(\varepsilon_i) = 0$). Также, поскольку первый момент K -мерного вектора $g_i (= x_{i0}, \dots, x_{ik})$ есть ε_i , из предположения 2.5 вытекает, что

$$E(\varepsilon_i | g_1, g_2, \dots, g_i) = 0.$$

Тогда, по закону повторных математических ожиданий, ε_i является скалярной п.д.с.

$$E(\varepsilon_i | \varepsilon_1, \varepsilon_2, \dots, \varepsilon_i) = 0. \quad (2.3.3)$$

Следовательно, из предположения 2.5 вытекает, что ошибка сама является п.д.с. и, следовательно, серьезно некоррелирована.

- (S является матрицей четвертых моментов.) Поскольку $g_i \equiv x_{i0}, \dots$, матрица S в предположении 2.5 может быть записана как $E(\varepsilon_i^2 x_{i0} x_{i0}')$. Ее (k, j) элементом является $E(x_{i0}^2 x_{ik} x_{ij})$. Поэтому S является матрицей четвертых моментов (математических ожиданий произведений четырех различных переменных). Следовательно, оценка по S будет требовать дополнительного предположения, которое будет специфицировано в параграфе 2.5.

- (Отсутствие предположения 2.5 приводит к другому выражению для матрицы S .) Благодаря предположению о том, что $\{g_i\}$ есть п.д.з., $S(= \text{Avar}(\bar{g}))$ равна $E(g_i g_i')$. Без этого предположения, как мы увидим в главе 6, выражение для S будет более сложным и включает автокорреляции g_i .

Асимптотическое распределение OLS-оценки

Мы теперь докажем, что OLS-оценка является состоятельной и асимптотически нормальной. В остальной части этой главы следует иметь в виду, что OLS-оценка b зависит от размера выборки n (хотя эта зависимость не делается явной из-за нашего выбора не снабжать b индексом n) и что число регрессоров K остается фиксированным, когда мы отслеживаем последовательность OLS-оценок, снабженных индексом n . На данный момент мы предполагаем, что для $S(= \text{Avar}(\bar{g}) = E(g_i g_i') = E(\varepsilon_i^2 x_i x_i'))$ доступна некоторая состоятельная оценка, обозначаемая как $\hat{S}$. Задача состоятельного оценивания S будет рассмотрена позже.

Утверждение 2.1 (асимптотическое распределение OLS-оценки):

- (a) (Состоятельность b для β) При предположениях 2.1–2.4, имеет место равенство $\lim_{n \rightarrow \infty} b = \beta$. (Поэтому предположение 2.5 не является необходимым для состоятельности.)
- (ii) (Асимптотическая нормальность b .) Если предположение 2.3 усилить до предположения 2.5, то тогда

$$\sqrt{n}(b - \beta) \rightarrow N(0, \text{Avar}(b)) \text{ при } n \rightarrow \infty,$$

где

$$\text{Avar}(b) = \Sigma_{xx}^{-1} S \Sigma_{xx}^{-1}. \quad (2.3.4)$$

(Напомним, $\Sigma_{xx} = E(x_i x_i')$, $S = E(g_i g_i')$, $g_i = x_i \cdot \varepsilon_i$.)

- (c) (Состоятельная оценка $\text{Avar}(b)$.) Предположим, что доступны состоятельная оценка $\hat{S}$ для S ($K \times K$). Тогда при предположении 2.2, $\text{Avar}(b)$ состоятельно оценивается посредством

$$\widehat{\text{Avar}}(b) = \Sigma_{xx}^{-1} \hat{S} \Sigma_{xx}^{-1}. \quad (2.3.5)$$

где Σ_{xx} есть выборочное среднее от $x_i x_i'$:

$$\Sigma_{xx} = \frac{1}{n} \sum_{i=1}^n x_i x_i' = \frac{1}{n} X'X, \quad (2.3.6)$$

Доказательство является ярким примером всех стандартных трюков в асимптотике. Для доказательства (a) и (b) будут использоваться три трюка. 1) записать рассматриваемый объект в терминах выборочных средних. 2) применить соответствующие LLN (Эргодическая теорема в данном контексте) и CLT (CLT для эргодической стационарной мартигал-значности) к выборочным средним и 3) использовать лемму 2.4(c) для получения асимптотического распределения. Доказательство части (c) не привнесет, поскольку оно непосредственно следует из эргодической стационарности.

Доказательство (части (a) и (c)).

- (1) Сначала мы запишем выборочную ошибку $b - \beta$ в терминах выборочных средних.

$$\begin{aligned} b - \beta &= (X'X)^{-1}X'\varepsilon = \\ &= \left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}X'\varepsilon\right) = \\ &= \left(\frac{1}{n}\sum_{i=1}^n x_i x_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n x_i \cdot \varepsilon_i\right) \equiv \\ &\equiv S_{xx}^{-1}\bar{g}. \end{aligned} \quad (2.3.7)$$

$$\text{где} \quad \bar{g} = \frac{1}{n}\sum_{i=1}^n g_i, \quad g_i = x_i \cdot \varepsilon_i.$$

Выборочные средние S_{xx} и $\bar{g}$ зависят от размера выборки n , хотя обозначения не делают это явным.

- (2) (Состоятельность.) Так как, по предположению 2.2, процесс $\{x_{it}\}$ является эргодически стационарным, $S_{xx} \rightarrow_p \Sigma_{xx}$. (Сходимости здесь фактически почти наверное, но из сходимости почти наверное вытекает сходимость по вероятности.) Поскольку матрица Σ_{xx} обратима, по предположению 2.4, $S_{xx}^{-1} \rightarrow_p \Sigma_{xx}^{-1}$ по лемме 2.3(a). Аналогично $\bar{g} \rightarrow_p E(g)$, что, по предположению 2.3, есть 0. Поэтому, по лемме 2.3(a), $S_{xx}^{-1}\bar{g} \rightarrow_p \Sigma_{xx}^{-1}0 = 0$. Следовательно, $\text{plim}_{n \rightarrow \infty}(b - \beta) = 0$, откуда вытекает, что $\text{plim}_{n \rightarrow \infty} b = \beta$.

- (3) (Асимптотическая нормальность.) Перепишем (2.3.7) в виде:

$$\sqrt{n}(b - \beta) = S_{xx}^{-1}(\sqrt{n}\bar{g}). \quad (2.3.8)$$

Как увидим при формализовке предложения 2.5, $\sqrt{n}\bar{g} \rightarrow_d N(0, S)$. Поэтому, по лемме 2.4(c), $\sqrt{n}(b - \beta)$ сходит к нормальному распределению со средним 0 и дисперсией $S_{xx}^{-1}S(S_{xx}^{-1})'$. Но так как матрица Σ_{xx} является симметричной, эти выражения равно (2.3.4). ■

Этот результат говорит о том, что распределение умноженной на $\sqrt{n}$ выборочной ошибки сходится условно точно аппроксимируется нормальным распределением, когда размер выборки достаточно большой. Естественный вопрос заключается в том, насколько большой «большой» насколько большим должен быть размер выборки, для того чтобы асимптотическая аппроксимация была обоснованной? Асимптотический результат, который мы только что получили, справедлив для всех DGP, удовлетворяющих предположениям модели. Однако размер выборки, необходимый для достижения той или иной меры близости к асимптотическому распределению, зависит от DGP. Мы затронем эту проблему в этой главе в эксперименте Монт-Карло.

s^2 состоятельна

Обратим внимание на OLS-оценку дисперсии ошибок, s^2 .

Утверждение 2.2 (состоятельная оценка дисперсии ошибок): Пусть $e_i = y_i - x_i'b$ есть OLS-остаток для наблюдения i . При предположениях 2.1–2.4,

$$s^2 = \frac{1}{n-K} \sum_{i=1}^n e_i^2 \rightarrow_p E(e_i^2)$$

при условии, что $E(e_i^2)$ существует и конечно.

Если бы мы могли наблюдать ошибку, e_i , то тогда очевидной оценкой были бы выборочное среднее от e_i^2 . Она является состоятельной в силу эргодической стационарности. Основываясь на утверждении 2.2 заключаемся в том, что замена действительной ошибки, e_i , OLS-остатком, e_i , не нарушает состоятельности. Давайте внимательно рассмотрим пункт за пунктом схему доказательства, поскольку именно тогда, как справиться с расхождением между e_i и его оценкой e_i , будет полезным и в других контекстах. Поскольку

$$e^2 = \frac{1}{n-K} \left(\sum_{i=1}^n e_i^2 \right),$$

достаточно показать, что выборочное среднее от e_i^2 , $\frac{1}{n} \sum_{i=1}^n e_i^2$, сходится по вероятности к $E(e_i^2)$. Связь между e_i и e_i задается как

$$\begin{aligned} e_i &= y_i - x_i'b \\ y_i &= x_i'\beta + e_i(b - \beta), \quad (\text{добавляя и вычитая } x_i'\beta) \\ e_i &= x_i'(b - \beta). \end{aligned} \quad (2.3.9)$$

так что

$$e_i^2 = e_i^2 - 2(b - \beta)'x_i + (b - \beta)'x_i x_i'(b - \beta). \quad (2.3.10)$$

Суммируя по i , мы получаем:

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 &= \frac{1}{n} \sum_{i=1}^n z_i^2 - 2(b - \beta)' \frac{1}{n} \sum_{i=1}^n x_i \cdot z_i + \\ &\quad - (b - \beta)' \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right) (b - \beta) = \\ &= \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 - 2(b - \beta)' \bar{z} - (b - \beta)' S_{xx}(b - \beta).\end{aligned}\quad (2.3.11)$$

Остаток доказательства, который должен показать, что plim двух последних слагаемых равен нулю, так что $\text{plim} \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 = \text{plim} \frac{1}{n} \sum_{i=1}^n e_i^2$, остается в качестве контрольного вопроса. Если вы будете рассматривать этот вопрос, зах будет ясно, что все, что требуется от оценки коэффициентов, — это ее состоятельность; если для формирования остатков используется другая состоятельная оценка, а не OLS-оценка $\hat{b}$, указанная оценка дисперсии ошибок остается состоятельной для $E(\varepsilon_i^2)$.

Контрольные вопросы

- Предположим, что $E(y_i|x_i) = x_i'\beta$, то есть предположим, что регрессия y_i на x_i является линейной функцией по x_i . Определим $\varepsilon_i = y_i - x_i'\beta$. Покажите, что ε_i ортогонален x_i . **Указание:** Сначала покажите, что $E(\varepsilon_i|x_i) = 0$.
- (Предполагается ли $E(\varepsilon_i^2)$ конечным?)
 - Вытекает ли из предположений 2.1–2.6, что $E(\varepsilon_i^2)$ существует и конечен? **Указание:** Вторые моменты строго стандартизированных переменных могут не быть конечными.
 - Если один из регрессоров в линейной модели является константой, то дисперсия ошибки конечна. Докажите это. **Указание:** Если $x_{i1} = 1$, то элемент $(1,1)$ матрицы $g_i g_i'$ равен ε_i^2 .
- Альтернативное выражение для S . Пусть $f(x_i) \equiv E(\varepsilon_i^2|x_i)$. Покажите, что $S(\equiv E(x_i x_i' \varepsilon_i^2))$ можно записать как

$$S = E\{f(x_i)x_i x_i'\}.$$

Указание: закон полных математических ожиданий.

- Завершите доказательство утверждения 2.1. **Указание:** Мы уже доказали для утверждения 2.1, что $\text{plim} \bar{y} = 0$, $\text{plim} S_{xx} = S_{xx}$ и $\text{plim}(\hat{b} - \beta) = 0$ при предположениях 2.1–2.4. Используйте лемму 2.3(a), чтобы показать, что $\text{plim}(\hat{b} - \beta)' \bar{y} = 0$ и $\text{plim}(\hat{b} - \beta)' S_{xx}(\hat{b} - \beta) = 0$.
- (Утверждение 2.2 с состоятельной $\hat{\beta}$) Докажите следующее обобщение утверждения 2.2

Пусть $\hat{\varepsilon}_i = y_i - x_i'\hat{\beta}$, где $\hat{\beta}$ — любая состоятельная оценка β . При предположениях 2.1, 2.2 и предположении того, что $E(x_1 \cdot x_1')$ и $E(x_1 x_1')$ конечны, $\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 \rightarrow_p E(\varepsilon_1^2)$.

Поэтому регрессоры не обязаны быть ортогональными ошибкам.

2.4. Тестирование гипотез

Статистические выводы в теории больших выборок основаны на тестовых статистиках, асимптотическое распределение которых известно в случае правильности нулевой гипотезы. Получить распределения тестовых статистик легче, чем в теории конечных выборок, поскольку нас интересует только аппроксимация точного распределения на больших выборках. В этом параграфе мы получаем тестовые статистики, предполагая, что для $S(\equiv L(g, g'))$ доступна состоятельная оценка, $\hat{S}$. Проблема состоятельного оценивания S будет рассмотрена в следующем параграфе.

Тестируемые линейные гипотезы

Рассмотрим тестирование гипотезы о значении k -го коэффициента β_k . Из утверждения 2.1 вытекает, что при $H_0: \beta_k = \bar{\beta}_k$

$$\sqrt{n}(\hat{\beta}_k - \bar{\beta}_k) \xrightarrow{d} N(0, \text{Avar}(\hat{\beta}_k)) \quad \text{и} \quad \widehat{\text{Avar}}(\hat{\beta}_k) \xrightarrow{p} \text{Avar}(\hat{\beta}_k),$$

где $\hat{\beta}_k$ есть k -й элемент $\hat{\beta}$ и $\text{Avar}(\hat{\beta}_k)$ есть (k, k) -й элемент $K \times K$ матрицы $\text{Avar}(\hat{\beta})$. Поэтому лемма 2.4(c) гарантирует, что

$$t_k = \frac{\sqrt{n}(\hat{\beta}_k - \bar{\beta}_k)}{\sqrt{\widehat{\text{SE}}^*(\hat{\beta}_k)}} = \frac{\hat{\beta}_k - \bar{\beta}_k}{\widehat{\text{SE}}^*(\hat{\beta}_k)} \xrightarrow{d} N(0, 1), \quad (2.4.1)$$

где

$$\widehat{\text{SE}}^*(\hat{\beta}_k) \equiv \sqrt{\frac{1}{n} \cdot \widehat{\text{Avar}}(\hat{\beta}_k)} = \sqrt{\frac{1}{n} \left(\hat{S}_{kk} - \hat{S}' \hat{S}_{kk}^{-1} \hat{S}_{kk} \right)_{kk}}$$

Замечательное t -отношение, $\widehat{\text{SE}}^*(\hat{\beta}_k)$, называется **состоятельной при гетероскедастичности стандартной ошибкой** (*heteroskedasticity-consistent standard error*), **робастной (к гетероскедастичности) стандартной ошибкой** (*heteroskedasticity-robust standard error*), или **стандартной ошибкой в форме Уайта** (*White's standard error*). Причиной для такой терминологии является то, что ошибка может быть условно гетероскедастичной; вспомните, что мы не предполагали условную гомоскедастичность (то есть что $E(\varepsilon_1^2 | x_1)$ не зависит от x_1) для получения асимптотического распределения $\hat{\beta}$. Это t -отношение называется **робастным (robust) t -отношением**, чтобы отличать его от t -отношения из главы 1. Связь между этими двумя видами t -отношения будет обсуждаться

в параграфе 2.6. Проверив нулевую гипотезу $H_0: \beta_2 = 0$, на уровне значимости α , ориентируясь на это t -отношение, осуществляется следующий образ:

Шаг 1: Вычисляем t_0 по формуле (2.4.1).

Шаг 2: Смотрим в таблицу $N(0, 1)$, чтобы найти критическое значение $t_{\alpha/2}$, которое оставляет $\alpha/2$ верхнюю хвосту стандартного нормального распределения¹. (Примечание: если $\alpha = 5\%$, $t_{\alpha/2} = 1.96$.)

Шаг 3: Нулевую гипотезу принимаем, если $|t_0| < t_{\alpha/2}$; в противном случае отвергаем ее.

Отличия от t -теста на конечных выборках: 1) другой способ вычисления стандартной ошибки, 2) мы используем таблицу $N(0, 1)$, а не $t(n - K)$; 3) **фактический размер** (actual size), или **точный размер** (exact size), теста (вероятность ошибки 1 рода при заданном объеме выборки) равен номинальному размеру (nominal size) (то есть желаемому уровню значимости α) только приблизительно, хотя аппроксимация становится очень хорошей с увеличением объема выборки. Разница между точным размером и номинальным размером теста называется **искажением размера** (size distortion). Поскольку t_0 асимптотически стандартная нормальная, искажение размера t -теста сходится к нулю при стремлении объема выборки n к бесконечности.

Таким образом, мы доказали первую половину следующего утверждения:

Утверждение 2.3 (робастное t -отношение и статистика Вальда):

Предположим, что выполняются предположения 2.1–2.5, и предположим, что существует доступная состоятельная оценка $\hat{S}$ для $S (= E[g_i g_i'])$. Как и прежде, пусть

$$\widehat{Avar}(b) \equiv S_{xx}^{-1} \hat{S} S_{xx}^{-1}.$$

Тогда

(a) При нулевой гипотезе $H_0: \beta_2 = \beta_0$, t_0 , определенная в формуле (2.4.1), $\rightarrow_d N(0, 1)$.

(b) При нулевой гипотезе $H_0: R\beta = r$, где R есть $\frac{\#r}{n} \times K$ матрица ($\frac{\#r}{n}$ — размерность r , число уравнений) выбрана r -около нуля,

$$W \equiv n \cdot (R\hat{b} - r)' \{ R \widehat{Avar}(b) R' \}^{-1} (R\hat{b} - r) \xrightarrow{d} \chi^2(\frac{\#r}{n}). \quad (2.4.2)$$

¹Обратите внимание на то, что здесь $t_{\alpha/2}$ — это α -квантиль уровня $\alpha/2$ стандартного уровня $1 - \alpha/2 = 1 - \text{Pr}(Z > \text{crit. value})$.

Что остается показать — так это то, что $W \rightarrow_d \chi^2(\#r)$, а это есть непосредственное применение леммы 2.4(d).

Доказательство (продолжение). Запишем W как

$$W = c_0' Q_0^{-1} c_0, \text{ где } c_0 = \sqrt{n}(Nb - r) \text{ и } Q_0 = R \text{Avar}(\bar{b}) R'.$$

При H_0 имеем: $c_0 = R\sqrt{n}(\bar{b} - \beta)$. Поэтому, по предположению 2.1,

$$c_0 \xrightarrow{d} c, \text{ где } c \sim N(0, R \text{Avar}(\bar{b}) R').$$

Также по предположению 2.1

$$Q_0 \xrightarrow{p} Q, \text{ где } Q \equiv R \text{Avar}(\bar{b}) R'.$$

Поскольку R имеет полный строковый ранг и $\text{Avar}(\bar{b})$ положительно определена, Q обратима. Следовательно, по лемме 2.4(d),

$$W \xrightarrow{d} c' Q^{-1} c.$$

Так как $\#r$ -мерный случайный вектор c распределен нормально со средним 0 и так как Q равна $\text{Var}(c)$, $c' Q^{-1} c \sim \chi^2(\#r)$. ■

Эта статистика хи-квадрат W является статистикой Вальда, поскольку она основана на неограниченных оценках (здесь это $\bar{b}$ и $\text{Avar}(\bar{b})$), не ограниченных нулевой гипотезой H_0 . Тестирование H_0 на уровне значимости α осуществляется следующим образом.

Шаг 1: Вычисляем статистику W по формуле (2.4.2).

Шаг 2: Смотрим в таблицу распределения $\chi^2(\#r)$, чтобы найти критические значения $\chi^2_{\alpha}(\#r)$, которые дают α верхнему хвосту распределения $\chi^2(\#r)$.

Шаг 3: Если $W < \chi^2_{\alpha}(\#r)$, то H_0 принимаем; в противном случае отвергаем ее.

Вероятность ошибки I рода приближается к α при увеличении выборки. Как будет ясно из параграфа 2.6, эта статистика Вальда тесно связана с известным F -тестом при условной гомоскедастичности.

Тест является состоятельным

Вспомним из основ статистики, что **мощность** (power) теста на конечных выборках есть вероятность отвергнуть нулевую гипотезу, когда она ложна, при заданной выборке конечного объема (то есть мощность есть I минус вероятность ошибки II рода). Мощность будет, очевидно, зависеть от DGP (то есть фактически сгенерированных данных),

рассмотренного как альтернатива, так же как и от размера (уровня значимости) теста. Для примера рассмотрим DGP $\{y_i, x_i\}$, удовлетворяющий предположениям 2.1–2.5, но не нулевой гипотезе $H_0: \beta_k = \bar{\beta}_k$. Мощность t -теста с размером α против этой альтернативы есть

$$\text{мощность} = \Pr\{|\hat{b}_k| > t_{\alpha/2}\},$$

которая зависит от DGP в вопросе, поскольку DGP контролирует распределение $\hat{b}_k$. Мы хотим, чтоб тест являлся **состоятельным** (*consistent*) против множества DGP, для один из которых не удовлетворяет нулевой гипотезе, если его мощность против любого конкретного элемента этого множества приближается к единице при $n \rightarrow \infty$ для любого предписанного уровня значимости.

То, что t -тест является состоятельным против множества альтернатив (DGP), удовлетворяющих предположениям 2.1–2.5, можно увидеть из следующего. Посмотрим на воспроизводимое здесь выражение (2.4.1) для t -отношения:

$$t_k \equiv \frac{\sqrt{n}(\hat{b}_k - \bar{\beta}_k)}{\sqrt{\widehat{\text{Avar}}(\hat{b}_k)}}.$$

Знаменатель, с одной стороны, сходится к $\sqrt{\widehat{\text{Avar}}(\hat{b}_k)}$ вопреки тому, что DGP не удовлетворяет нулевой гипотезе (помним, что все части утверждения 2.1 выполняются независимо от того, верна ли нулевая гипотеза, при условии выполнения предположений 2.1–2.5). С другой стороны, числитель стремится к $+\infty$ или $-\infty$, поскольку $\hat{b}_k$ сходится по вероятности к β_k на DGP, отличающемся от $\bar{\beta}_k$. Поэтому мощность стремится к единице при стремлении объема выборки n к бесконечности, откуда вытекает, что t -тест в утверждении 2.3 состоятелен против тех альтернатив, DGP, которые не удовлетворяют нулевой гипотезе. То же самое верно для теста Вальда.

Асимптотическая мощность

Для последующего использования в следующей главе мы определим здесь **асимптотическую мощность состоятельного теста** (*asymptotic power of a consistent test*). Как замечено выше, мощность t -теста приближается к единице при увеличении объема выборки, а то время как DGP, взятый в качестве альтернативы, остается фиксированным. Но если при увеличении объема выборки DGP становится все ближе и ближе к нулевой гипотезе, мощность может не сходиться к единице. Последовательность таких DGP называется **последовательностью локальных альтернатив** (*sequence of local alternatives*). Для регрессионной модели и для нулевой гипотезы $H_0: \beta_k = \bar{\beta}_k$ эта последовательность DGP

такая, что (i) n -й DGP $\{y_i^{(n)}, x_i^{(n)}; i = 1, 2, \dots\}$ удовлетворяет предположениям 2.1–2.5 и сходится в обычном смысле к фиксированному DGP $\{y_i, x_i\}^1$ и (ii) значения t_k n -го DGP, $t_k^{(n)}$, сходятся к $\bar{t}_k$. Предположим далее, что $\beta_k^{(n)}$ удовлетворяет

$$\beta_k^{(n)} = \bar{\beta}_k + \frac{\gamma}{\sqrt{n}} \quad (2.4.3)$$

для некоторого заданного $\gamma \neq 0$. Таким образом, $\beta_k^{(n)}$ приближается к $\bar{\beta}_k$ со скоростью, пропорциональной $1/\sqrt{n}$. Этот частный случай последовательности локальных альтернатив называется **Питменовским сдвигом** (*Pitman drift*) или **Питменовской последовательностью** (*Pitman sequence*). Подставляя (2.4.3) в (2.4.1), t -отношение здесь можно записать так

$$t_k = \frac{\sqrt{n}(b_k - \beta_k^{(n)})}{\sqrt{\text{Avar}(\hat{b}_k)}} = \frac{\gamma}{\sqrt{\text{Avar}(\hat{b}_k)}}. \quad (2.4.4)$$

Если выборка объема n порождается n -м DGP Питменовского сдвига, сходятся ли t_k к нетривиальному распределению? Поскольку n -й DGP удовлетворяет предположениям 2.1–2.5, первая компонента в правой части (2.4.4) сходится по распределению к $N(0, 1)$ согласно частям (b) и (c) утверждения 2.1. В силу части (c) утверждения 2.1 и того факта, что $\{y_i^{(n)}, x_i^{(n)}\}$ «сходится» к фиксированному DGP, вторая компонента сходится по вероятности к

$$\mu = \frac{\gamma}{\sqrt{\text{Avar}(\hat{b}_k)}}, \quad (2.4.5)$$

где $\text{Avar}(\hat{b}_k)$ вычисляется при фиксированном DGP. Следовательно, $t_k \rightarrow_d N(\mu, 1)$ вдоль этой последовательности локальных альтернатив. Если уровень значимости равен α , мощность сходится к

$$\text{Power}(\mu, \alpha) = 1 - \Phi(-t_{\alpha/2}), \quad (2.4.6)$$

где $x \sim N(\mu, 1) \Leftrightarrow t_{\alpha/2} \sim$ критическое значение для уровня значимости α . Эта вероятность называется **асимптотической мощностью** (*asymptotic power*). Она является мерой способности теста обнаруживать малые отклонения от нулевой гипотезы. Очевидно, что чем больше $|\mu|$, тем выше асимптотическая мощность для заданного размера α . Исходя из элементарных аргументов, легко показать, что статистика Вальда сходится к распределению, называемому **нецентральным хи-квадрат-распределением** (*noncentral chi-square*).

¹См., например, предположение 1 в [Newey, 1985] для точной формулировки.

Тестирование нелинейных гипотез

Статистика Вальда может быть обобщена для тестирования множества нелинейных ограничений на β . Рассмотрим нулевую гипотезу в форме:

$$H_0: \alpha(\beta) = 0.$$

Здесь α — векторозначная функция с непрерывными первыми производными. Пусть $\# \alpha$ — размерность $\alpha(\beta)$ (так что нулевая гипотеза эквивалентна ограничению), а $A(\beta)$ является $\# \alpha \times K$ -матрицей первых производных, вычисленных в β : $A(\beta) = \partial \alpha(\beta) / \partial \beta'$. Для того чтобы убедиться, что статистика была определена, мы предполагаем, что $A(\beta)$ имеет полный строковый ранг (это является обобщением требования линейной гипотезы $H_0: \pi = 0$ тем, что R имеет полный строковый ранг). Из леммы 2.3(а) и предложения 2.1(а) вытекает, что

$$\sqrt{n}[\alpha(\hat{\beta}) - \alpha(\beta)] \rightarrow_d c, \quad c \sim N(0, A(\beta)' \text{Avar}(b) A(\beta)). \quad (2.4.7)$$

Так как $\alpha(\beta) = 0$ при H_0 , (2.4.7) принимает вид

$$\sqrt{n} \alpha(\hat{\beta}) \rightarrow_d c, \quad c \sim N(0, A(\beta)' \text{Avar}(b) A(\beta)). \quad (2.4.8)$$

Поскольку $\hat{b} \rightarrow_p \beta$ по утверждению 2.1(а), лемма 2.3(а) влечет $A(\hat{b}) \rightarrow_p A(\beta)$. По предположению 2.1(с), $\widetilde{\text{Avar}}(\hat{b}) \rightarrow_p \text{Avar}(b)$. Поэтому, по лемме 2.3(а),

$$A(\hat{b}) \widetilde{\text{Avar}}(\hat{b}) A(\hat{b})' \rightarrow_p A(\beta) \text{Avar}(b) A(\beta)' = \text{Var}(c). \quad (2.4.9)$$

Поскольку $A(\beta)$ имеет полный строковый ранг и $\text{Avar}(b)$ положительно определена, $\text{Var}(c)$ обратима. Тогда лемма 2.4(а), (2.4.8) и (2.4.9) влекут

$$\sqrt{n} \alpha(\hat{b})' \{A(\hat{b}) \widetilde{\text{Avar}}(\hat{b}) A(\hat{b})'\}^{-1} \sqrt{n} \alpha(\hat{b}) \rightarrow_d \text{Var}(c)^{-1/2} c \sim \chi^2(\# \alpha). \quad (2.4.10)$$

Объединяя два $\sqrt{n}$ в (2.4.10) и учитывая (а), мы доказали

Утверждение 2.3 (предложение):

(а) При нулевой гипотезе $c \equiv \# \alpha$ ограничений $H_0: \alpha(\beta) = 0$, таких, что $A(\beta)$, матрица размера $\# \alpha \times K$ непрерывных первых производных $\alpha(\beta)$, имеет полный строковый ранг, мы имеем:

$$W \equiv n \cdot \alpha(\hat{b})' \{A(\hat{b}) \widetilde{\text{Avar}}(\hat{b}) A(\hat{b})'\}^{-1} \alpha(\hat{b}) \xrightarrow{d} \chi^2(\# \alpha). \quad (2.4.11)$$

Часть (а) является обобщением (b), при $\alpha(\beta) = R\beta - r$, (2.4.11) сводится к (2.4.2), статистика Вальда для линейных ограничений.

Выбор $\alpha(\cdot)$ для предоставления заданного множества ограничений не является единственным. Например, равенство $\beta_1\beta_2 = 1$ можно записать как $\alpha(\beta) = 0$ с $\alpha(\beta) = \beta_1\beta_2 - 1$ или с $\alpha(\beta) = \beta_1 - 1/\beta_2$. В то время как часть (с) утверждения гарантирует, что на больших выборках результат теста Вальда одинаково, независимо от выбора функции α , численное значение статистики Вальда W в действительности зависит от выбранного представления, и результат теста может быть различным на конечных выборках. В вышеприведенном примере второе представление, $\alpha(\beta) = \beta_1 - 1/\beta_2$, не удовлетворяет требованию непрерывности производных при $\beta_2 = 0$. Действительно, исследование Мунте-Карло (Gregory and Veall, 1985) показывает, что, когда β_2 близко к нулю, тест Вальда, основанный на втором представлении, отвергает нулевую гипотезу на малых выборках слишком часто.

Контрольные вопросы

- Верно ли, что $SE^*(b_k) \rightarrow 0$ при $n \rightarrow \infty$?
- (Стандартная ошибка логистической функции.) Для простоты пусть $K = 1$ и пусть b — OLS-оценка для β . Стандартной ошибкой для b является $\sqrt{\text{Avar}(b)/n}$. Предположим, что $\lambda = -\log(b)$. Оценка λ , порожденная OLS-оценкой для β , имеет вид: $\hat{\lambda} = -\log(b)$. Проверьте, что стандартная ошибка $\hat{\lambda}$ равна $(1/b) \sqrt{\text{Avar}(b)/n}$.
- (Инвариантность (или ее отсутствие) статистики Вальда.) Не существует однозначного способа записать линейную гипотезу $R\beta = r$, поскольку для любой невырожденной матрицы F размера $\frac{1}{2}n \times \frac{1}{2}n$ то же самое множество ограничений может быть представлено как $\tilde{R}\beta = \tilde{r}$ с $\tilde{R} \equiv FR$ и $\tilde{r} \equiv Fr$. Влияют ли различные выборы R и r на асимптотическое распределение W ? На распределение на конечных выборках? На численное значение?

2.5. Состоятельное оценивание $E(\varepsilon_i^2|x_i, x'_i)$

Теория, разработанная ранее, предполагает, что существует доступная состоятельная оценка, $\hat{S}$, для $S(= E(g_i g'_i) = E(\varepsilon_i^2|x_i, x'_i))$, которая будет использоваться для вычисления оцененной асимптотической дисперсии, $\widehat{\text{Avar}}(\hat{\theta})$. Этот параграф объясняет, как получить $\hat{S}$ по выборке (y, X) .

Использование остатков вместо ошибок

Если бы ошибка была наблюдаема, то тогда выборочное среднее $\varepsilon_i^2|x_i, x'_i$, очевидно, было бы состоятельно в силу эргодической стационарности. Но мы не наблюдаем ошибку, и подстановка вместо нее некоторой ее

состоятельной оценки дает:

$$\hat{S} = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 x_i x_i', \quad (2.5.1)$$

где $\hat{\varepsilon}_i \equiv y_i - x_i' \hat{\beta}$, и $\hat{\beta}$ — некоторая состоятельная оценка β . (Хотя очевидным кандидатом для состоятельной оценки β является OLS-оценка b , мы используем здесь $\hat{\beta}$, а не b , чтобы отметить, что результаты этого параграфа справедливы для любой состоятельной оценки.) Чтобы эта оценка была состоятельной для S , нам нужно сделать предположение относительно четвертых моментов регрессоров.

Предположение 2.6 (конечные четвертые моменты регрессоров):

$E\{x_i x_i x_i x_i'\}^2$ существует и конечно для всех i, j ($= 1, 2, \dots, K$).

Утверждение 2.4 (состоятельная оценка для S): Предположим, что оценка коэффициента $\hat{\beta}$, использованная для вычисления остатков $\hat{\varepsilon}_i$ для $\hat{S}$ в (2.5.1), является состоятельной, и предположим, что $S = E\{g_i g_i'\}$ существует и конечно. Тогда при предположениях 2.1, 2.2 и 2.6 $\hat{S}$, заданная в (2.5.1), является состоятельной оценкой для S .

Чтобы указать, почему необходимо предположение о четвертом моменте для регрессоров, мы даем набросок доказательства для частного случая $K = 1$ (только один регрессор). Так что x_i сейчас скаляр x_i , g_i — скаляр $g_i = x_i \varepsilon_i$ и (2.3.10) (с $b = \hat{\beta}$ и $\varepsilon_i = \hat{\varepsilon}_i$) упрощается до

$$\hat{\varepsilon}_i^2 = \varepsilon_i^2 - 2(\hat{\beta} - \beta)x_i \varepsilon_i + (\hat{\beta} - \beta)^2 x_i^2. \quad (2.5.2)$$

Докажем обе стороны, на x_i^2 и суммируя по i ,

$$\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 x_i^2 - \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 x_i^2 = -2(\hat{\beta} - \beta) \frac{1}{n} \sum_{i=1}^n \varepsilon_i x_i^3 + (\hat{\beta} - \beta)^2 \frac{1}{n} \sum_{i=1}^n x_i^4. \quad (2.5.3)$$

Сейчас мы можем видеть, почему требуется предположение о конечном четвертом моменте для x_i : если четвертый момент $E\{x_i^4\}$ конечен, то тогда по эргодической стационарности выборочное среднее x_i^4 сходится по вероятности к некоторому конечному числу, поэтому последний член в (2.5.3) исчезает (сходится к 0 по вероятности), если $\hat{\beta}$ состоятельна для β . Можно также показать (см. аналитическое упражнение 4 для доказательства), комбинируя то же предположение о четвертых моментах регрессоров и предположение о конечности $E\{g_i g_i'\} (= E\{\varepsilon_i^2 x_i x_i'\})$, что выборочное среднее $x_i^2 \varepsilon_i$ сходится по вероятности к некоторому конечному числу, так что другая составляющая в правой части (2.5.3) также исчезает.

Согласно утверждению 2.1(a), предположения, сделанные в утверждении 2.3, являются достаточными, чтобы гарантировать состоятельность $\hat{b}$, поэтому мы можем положить $\tilde{\theta} = b$ в (2.5.1) и использовать OLS-остатки для вычисления $\tilde{S}$. К тому же предположение, сделанное в утверждении 2.4, о том, что $\text{tr}(g_i g_i')$ конечно, является частью предположения 2.5, которое предполагается выполненным в утверждении 2.3. Следовательно, подразумеваемый смысл утверждения 2.4:

Если к гипотезе в утверждении 2.3 добавить предположение 2.6, то тогда $\tilde{S}$, заданная в (2.5.1) с $b = \tilde{\theta}$ (так что $\tilde{\varepsilon}_i$ — OLS-остатки e_i) может быть использована в (2.3.5) для вычисления оценочной асимптотической дисперсии:

$$\widehat{\text{Avar}}(\hat{b}) = S_{xx}^{-1} \left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 x_i x_i' \right) S_{xx}^{-1}, \quad (2.5.4)$$

которая является состоятельной для $\text{Avar}(\hat{b})$.

Представление $\tilde{S}$ в терминах матриц данных

Если B — диагональная матрица размера $n \times n$, i -й диагональный элемент которой есть ε_i^2 , то тогда $\tilde{S}$ в (2.5.1) можно представить в терминах матриц данных как

$$\tilde{S} = \frac{X' B X}{n} \quad (2.5.1')$$

$$B = \begin{bmatrix} \varepsilon_1^2 & & \\ & \ddots & \\ & & \varepsilon_n^2 \end{bmatrix}.$$

Поэтому (2.5.4) может быть записано (с $\tilde{\varepsilon}_i$ в B равным ε_i) как

$$\widehat{\text{Avar}}(\hat{b}) = n^{-1} (X' X)^{-1} (X' B X) (X' X)^{-1}. \quad (2.5.4')$$

Эти выражения, хотя они и полезны для некоторых целей, не следует использовать для вычислений, поскольку $n \times n$ матрица B будет занимать слишком много оперативной памяти компьютера, особенно когда размер выборки велик. Для вычисления $\tilde{S}$ по выборке формула (2.5.1) более полезна, чем (2.5.1').

Рассмотрение конечных выборок

Разумеется, на конечных выборках мощность вполне может быть намного ниже единицы против конкретных альтернатив. Кроме того, вероятность отвергнуть нулевую гипотезу, когда DGP действительно удовлетворяет нулевой гипотезе (ошибка I рода) может очень отличаться от предполагаемого уровня значимости. Дэвидсон и Мак-Киннон [Davidson and MacKinnon, 1993, Section 16.3] сообщают, что, по крайней мере для наблюдавшихся ими симуляций Монте-Карло, робастное t -отношение, основанное на (2.5.1), отвергает нулевую гипотезу слишком часто и что простая замена знаменателя n в (2.5.1) на количество степеней свободы $n - K$ или, эквивалентно, домножение (2.5.1) на $n/(n - K)$ (это будет *коррекцией на число степеней свободы* (*degrees of freedom correction*)) уменьшает проблему слишком частого отвержения. Они также сообщают, что робастное t -отношение, основанное на следующей корректировке $\hat{S}$, работает даже лучше:

$$\hat{S} = \frac{1}{n} \sum_{i=1}^n \frac{e_i^2}{(1 - p_i)^d} x_i x_i', \quad d = 1 \text{ или } 2, \quad (2.5.5)$$

где p_i есть p_i , определенное в контексте анализа влияния в главе 1: это есть i -й диагональный элемент проекционной матрицы P , то есть

$$p_i \equiv x_i'(X'X)^{-1}x_i = \frac{x_i'S_{mm}^{-1}x_i}{n}.$$

Контрольные вопросы

1. (Вычисление робастных стандартных ошибок.) В контрольном вопросе 2 параграфа 1.4 мы наблюдали, что стандартные ошибки OLS-оценок коэффициентов могут быть получены из S_{xx} , s_{xy} (выборочное среднее x , y), $y'y/n$ и $\bar{y}$, поэтому выборочные моменты должны быть вычислены только один раз. Верно ли то же самое для робастных стандартных ошибок, где $\hat{S}$ вычисляется по формуле (2.5.1) с $\hat{e}_i = e_i$?
2. Дисперсия на конечных выборках OLS оценки в обобщенной регрессионной модели главы 1 равна $\text{Var}(b; X) = (X'X)^{-1}X'(n^2V)X(X'X)^{-1}$. Сравните ее с (2.5.4'). Каковы различия?

2.6. Последствия условной гомоскедастичности

Тестовые статистики, разработанные в параграфах 2.4 и 2.5, отличаются от их аналогов на конечных выборках из параграфа 1.4, предназначенных для тестирования той же самой нулевой гипотезы. Как они соотносятся? Каково асимптотическое распределение t - и F -статистик из главы 1? Данный параграф отвечает на эти вопросы.

Оказывается, что робастное t -отношение численно равно t -отношению из параграфа 1.4 для конкретного выбора $\hat{S}$. Следовательно, асимптотическое распределение t -отношения из параграфа 1.4 то же самое, что и у робастного t -отношения, если этот конкретный выбор является состоятельным для S . Аналогичное соотношение выполняется между F -отношением из параграфа 1.4 и статистикой Вальда W данной главы. При предположении об условной гомоскедастичности, сформулированном ниже, этот конкретный выбор действительно является состоятельным.

Противопоставление условной и безусловной гомоскедастичности

Предположение об условной гомоскедастичности таково:

Предположение 2.7 (условная гомоскедастичность):

$$E(\varepsilon_i^2 | x_i) = \sigma^2 > 0 \quad (2.6.1)$$

Из этого предположения вытекает, что безусловный второй момент $E(\varepsilon_i^2)$ равен σ^2 по закону полных математических ожиданий. Чтобы прояснить различие между безусловной и условной гомоскедастичностью, рассмотрим следующий пример.

Пример 2.6 (безусловно гомоскедастичные, но условно гетероскедастичные ошибки): Как уже отмечалось, если $\{y_i, x_i\}$ стационарен, то стационарен и $\{\varepsilon_i\}$, и ошибка безусловна гомоскедастична по той причине, что $E(\varepsilon_i^2)$ не зависит от i . Для иллюстрации того, что ошибка может быть тем не менее условно гетероскедастична, предположим, что ε_i задается как $\varepsilon_i \equiv \eta_i f(x_i)$, где $\{\eta_i\}$ имеет нулевое среднее $E(\eta_i) = 0$ и не зависит от x_i . Условный второй момент ε_i зависит от x_i , поскольку

$$\begin{aligned} E(\varepsilon_i^2 | x_i) &= E(\eta_i^2 f(x_i)^2 | x_i) \quad (\text{так как } \varepsilon_i \equiv \eta_i f(x_i)) = \\ &= f(x_i)^2 E(\eta_i^2 | x_i) \\ &(\text{по линейности условных математических ожиданий}) \\ &= f(x_i)^2 E(\eta_i^2) \quad (\text{так как } \eta_i \text{ не зависит от } x_i \text{ по предположению}), \end{aligned}$$

где правая часть изменяется с i из-за изменчивости $f(x_i)$ в зависимости от i .

Редукция к формулам для конечных выборок

Для изучения распределений на больших выборках t - и F -статистик из главы 1 при дополнительном предположении 2.7, мы сначала изучим алгебраическую связь с их робастными аналогами. Рассмотрим следующий выбор для оценки S :

$$\hat{S} = s^2 S_{\text{плн}},$$

где s^2 — OLS-оценка для σ^2 . (Мы очень скоро покажем, что эта оценка состоятельна при условной гомоскедастичности.) Тогда выражение (2.3.5) для $\widehat{\text{Avar}}(\hat{b})$ принимает вид:

$$\widehat{\text{Avar}}(\hat{b}) = s^2 S_{xx}^{-1} = n \cdot s^2 \cdot (X'X)^{-1}. \quad (2.6.2)$$

Подставляя это выражение в (2.4.1), мы видим, что робастная стандартная ошибка становится равной

$$\sqrt{s^2}, \text{ умноженная на } (k, k) \text{ элемент матрицы } (X'X)^{-1}, \quad (2.6.3)$$

а это есть обычная стандартная ошибка в теории конечных выборок. Поэтому робастное t -отношение численно идентично обычному t -отношению для конечных выборок, когда мы полагаем $\tilde{S} = s^2 S_{xx}$. Аналогично, подставляя (2.6.2) в выражение для статистики Вальда (2.4.2), мы получаем:

$$\begin{aligned} W &= n \cdot (R\hat{b} - r)' \{ R' n \cdot s^2 \cdot (X'X)^{-1} R' \}^{-1} (R\hat{b} - r) = \\ &= (R\hat{b} - r)' \{ R' s^2 \cdot (X'X)^{-1} R' \}^{-1} (R\hat{b} - r) \quad (\text{для } n \text{ сокращаются}) \\ &= (R\hat{b} - r)' \{ R' (X'X)^{-1} R' \}^{-1} (R\hat{b} - r) / s^2 = \\ &= \#r \cdot F \quad (\text{по определению (1.4.9) } F\text{-отношения}) \\ &= (SSR_R - SSR_U) / s^2 \quad (\text{из (1.4.11)}). \end{aligned}$$

Таким образом, когда мы полагаем $\tilde{S} = s^2 S_{xx}$, статистика Вальда W численно идентична $\#r \cdot F$ (где $\#r$ есть число ограничений в нулевой гипотезе).

Распределение на больших выборках t - и F -статистик

Из утверждения 2.3 тогда следует, что t -отношение (2.4.1) асимптотически $N(0, 1)$ и $\#r \cdot F$ асимптотически $\chi^2(\#r)$, если $s^2 S_{xx}$ состоятельна для S . То, что $s^2 S_{xx}$ состоятельна для S , можно видеть из следующего. При условной гомоскедастичности матрица четвертых моментов S может быть выражена как произведение вторых моментов:

$$\begin{aligned} S &= E(g_i g_i') = E(x_i x_i' \varepsilon_i^2) \quad (\text{так как } g_i = x_i \cdot \varepsilon_i) \\ &= E[E(x_i x_i' \varepsilon_i^2 | x_i)] \quad (\text{по закону о цепных математических ожиданиях}) \\ &= E[x_i x_i' E(\varepsilon_i^2 | x_i)] \\ & \quad (\text{по линейности условных математических ожиданий}) \\ &= E(x_i x_i' \sigma^4) \quad (\text{по предположению 2.7}) \\ &= \sigma^4 E(x_i x_i') = \sigma^4 \Sigma_{xx}. \end{aligned} \quad (2.6.4)$$

Это предствление имеет несколько важных следствий.

- (Σ_{xx} является невырожденной.) Так как, по предположению 2.5, S является невырожденной, это разложение S имеет $\sigma^2 > 0$, и Σ_{xx} является невырожденной. Следовательно, отсюда вытекает предположение 2.4 (ранговое условие).
- (Необходимость в предположении о четвертом моменте.) В силу эргодической стационарности, $S_{xx} \rightarrow_p \Sigma_{xx}$. Из утверждения 2.2 вытекает, что n^2 состоятельна для σ^2 при предположениях 2.1–2.4. Таким образом, $n^2 S_{xx} \rightarrow_p \sigma^2 \Sigma_{xx} = S$. Для нас нет необходимости требовать предположение о четвертом моменте (предположение 2.6) для состоятельности.

В качестве другого следствия из (2.6.4) заметим, что выражение для $\text{Avar}(b)$ можно упростить, подставив (2.6.4) в (а) из предположения 2.1, при этом выражение для $\text{Avar}(b)$ имеет вид:

$$\text{Avar}(b) = \sigma^2 \Sigma_{xx}^{-1}. \quad (2.6.5)$$

Таким образом, мы доказали

Утверждение 2.3 (свойства на больших выборках b , t и F при условной гомоскедастичности): Предположим, что выполняются предположения 2.1–2.5 и 2.7. Тогда

- (Асимптотическое распределение b .) OLS-оценка b или β является состоятельной и асимптотически нормальной с $\text{Avar}(b) = \sigma^2 X_{xx}^{-1}$.
- (Состоятельное оценивание асимптотической дисперсии.) При том же множестве предположений $\text{Avar}(b)$ состоятельно оценивается посредством $\widehat{\text{Avar}}(b) = s^2 S_{xx}^{-1} = n \cdot s^2 \cdot (X'X)^{-1}$.
- (Асимптотическое распределение t - и F -статистик теории конечных выборок.) При $H_0: \beta_k = \beta_{k_0}$ обычное t -отношение (1.4.5) асимптотически нормально распределено как $N(0,1)$. При $H_0: R\beta = r$, $\beta \in R$ имеет асимптотическое распределение $\chi^2(r/\#r)$, где R есть F -статистика из (1.4.9), и $\beta \in r$ равно числу ограничений в H_0 .

Варианты асимптотических тестов при условной гомоскедастичности

Согласно данному результату, мы должны посмотреть таблицу $N(0,1)$, чтобы найти критическое значение для сравнения с t -отношением (1.4.5) и таблицу χ^2 для статистики $\#r \in R$, полученной из (1.4.9). Некоторые исследователи заменяют s^2 в (1.4.5) и (1.4.9) на $\frac{1}{n} \sum e_i^2$. Это так, поскольку степень свободы $n - K$ заменяется на n , или коррекция степеней

свободы, скрытая в χ^2 , устраняется. Отличие, возникающее при этой замене, исчезает на больших выборках, поскольку $\lim_{n \rightarrow \infty} [n/(n-K)] = 1$. Следовательно, независимо от того, какой тест использовать, результат теста будет одинаковым, если объем выборки достаточно большой.

Другой вариант заключается в том, чтобы оставить количество степеней свободы $n - K$, но использовать таблицу $t(n - K)$ для t -отношения и таблицу $F(\#r, n - K)$ для F , что и в точности предписано в теории конечных выборок. Это также асимптотически обосновано, поскольку при $n - K$, стремящемся к бесконечности (что и происходит, когда $n \rightarrow \infty$ с фиксированным K), распределение $t(n - K)$ сходится к $N(0, 1)$ (просто сравните таблицу t -распределения для большого количества степеней свободы с таблицей стандартного нормального распределения), а $F(\#r, n - K)$ сходится к $\chi^2(\#r)/\#r$. Иными словами, даже если ошибка не распределена нормально и регрессоры всего лишь предопределенные (остаточными ошибками) и не строго экзогенные, распределение t -отношения (1.4.5) хорошо аппроксимируется распределением $t(n - K)$, а распределение F -отношения — распределением $F(\#r, n - K)$.

Все эти варианты асимптотически эквивалентны в том, что разница в их значениях исчезает на больших выборках n , следовательно (по лемме 2.4(a)), их асимптотическое распределение одинаково. Однако, когда объем выборки только умеренно большой, аппроксимация распределения на конечных выборках, или точного распределения, тестовых статистик может быть лучше при использовании $t(n - K)$ и $F(\#r, n - K)$, а не $N(0, 1)$ и $\chi^2(\#r)$. Поскольку точное распределение зависит от DGP, но не просто от указания на то, какой вариант работает лучше на конечных выборках. Вопрос о том, какую таблицу — $N(0, 1)$ или $t(n - K)$ — нужно использовать для умеренных объемов выборки, будет рассмотрен в упражнении на метод Монте-Карло в этой главе.

Контрольные вопросы

1. (Необходимость формул для конечных выборок без условий гомоскедастичности.) Без предположения 2.7 $\text{Avar}(b)$ определяется посредством (b) и утверждения 2.1. Оценивается ли она consistently посредством (2.6.2) без предположения 2.7? [Ответ: Нет. Почему?] Является ли t -отношение (1.4.5) асимптотически стандартным нормальным без такого предположения? [Ответ: Нет.]
2. (Преимущество формул для конечных выборок при условии гомоскедастичности.) Наоборот, при предположении 2.7 $\text{Avar}(b)$ определяется в (2.6.5). Оценивается ли она consistently посредством (2.3.5) при предположении 2.7? Если предположение 2.7 выполняется, то как вы думаете, есть ли преимущество использования (2.6.2) над (2.3.5) для оценивания асимптотической дисперсии? (Замечание: Свойства оценки на конечных выборках, в общем случае, тем лучше, чем меньше число подлежащих параметров, оценива-

емых для формирования этой оценки. Как много дополнительных параметров необходимо оценить для формирования (2.3.5)? (2.6.2)?

3. (Отношение F к χ^2 .) Найдите 5%-е критические значения для $F(10, \infty)$ и сравните его с 5%-ным критическим значением для $\chi^2(10)$. Как они соотносятся между собой?
4. Без условной гомоскедастичности имеет ли $(SSR_H - SSR_U)/s^2$ асимптотически распределение $\chi^2(K-1)$? (Ответ: Нет.)
5. (nR^2 .) Для регрессии с константой рассмотрим нулевую гипотезу о том, что коэффициенты при $K-1$ регрессорах, не являющихся константой, все равны нулю. Покажите, что $nR^2 \rightarrow \chi^2(K-1)$ при гипотезе в утверждении 2.5. **Указание:** Вы доказали в аналитическом упражнении главы 1, что алгебраическое соотношение между F -отношением для нулевой гипотезы и R^2 есть

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

Можете ли вы использовать статистику nR^2 , когда ошибки не являются условно гомоскедастичными? (Ответ: Нет.)

2.7. Тестирование условной гомоскедастичности

С добавлением собственных стандартных ошибок, позволяющих нам делать выводы без спецификации условного второго момента $E(\varepsilon_i^2|x_i)$, тестирование условной гомоскедастичности стало не столь важным, как раньше. Этот параграф представляет только самый популярный тест, предложенный в работе [White, 1980] для случая случайных выборок¹.

Напомним, что $\hat{S}$, заданная в (2.3.1) (с $\hat{\varepsilon}_i = e_i$), является состоятельной для S (утверждение 2.4), и $s^2 S_{xx}$ является состоятельной для $\sigma^2 \Sigma_{xx}$ (как следствие утверждения 2.2). Но при условной гомоскедастичности $S = \sigma^2 \Sigma_{xx}$ (см. (2.6.4)), поэтому разница между двумя состоятельными оценками должна исчезать:

$$\begin{aligned} \hat{S} - s^2 S_{xx} &= \frac{1}{n} \sum_{i=1}^n e_i^2 x_i x_i' - s^2 \frac{1}{n} \sum_{i=1}^n x_i x_i' = \\ &= \frac{1}{n} \sum_{i=1}^n (e_i^2 - s^2) x_i x_i' \xrightarrow{p} 0. \end{aligned} \quad (2.7.1)$$

Пусть ψ_i будет вектором, собирающим уникальные и непостоянные элементы симметричной матрицы $x_i x_i'$ размерности $K \times K$. (Построение

¹См., например, [Judge et al., 1980b, Section 11.3] или [Greene, 1997, Section 12.4] относительно других тестов.

ψ_i из $\mathcal{H}_1$ будет аппроксимировано в примере 2.3 ниже.) Тогда (2.7.1) имеет:

$$s_n = \frac{1}{n} \sum_{i=1}^n (e_i^2 - E^2) \psi_i \rightarrow 0 \quad (2.7.2)$$

Это s_n является выборочным средним, сходящимся к нулю. При некоторых условиях, соответствующих применимости центральной предельной теоремы, мы можем ожидать, что $\sqrt{n}s_n$ сходится по распределению к нормальному распределению с нулевым средним и некоторой асимптотической дисперсией B , поэтому для любой состоятельной оценки $\hat{B}$ для B :

$$n \cdot s_n' \hat{B}^{-1} s_n \rightarrow \chi^2_1(n). \quad (2.7.3)$$

где n есть размерность s_n . Для конкретной выборки $\hat{B}$ для статистики может быть вычислена как nR^2 из следующей множественной регрессии:

$$\text{регрессия } e_i^2 \text{ на константу и } \psi_i. \quad (2.7.4)$$

В [White, 1980] эта аргументация реализована строго¹ для доказательства следующего утверждения.

Утверждение 2.6 (тест Уайта для условной гетероскедастичности): B допустимо к предположениям 2.1 и 2.4 предположим, что а) (y_i, x_i) есть i.i.d. с конечным $E(e_i^2 | x_i)$ (таким образом, усиливая предположения 2.2 и 2.5), б) e_i независимы от x_i (таким образом, усиливая предположение 2.3 и условную гомоскедастичность) и в) выполняются определенные условия на моменты e_i и x_i . Тогда

$$nR^2 \rightarrow_0 \chi^2_1(n).$$

где R^2 есть R^2 множественной регрессии (2.7.4), а n есть размерность ψ_i .

Пример 2.7 (регрессии в тесте Уайта nR^2): Рассмотрим функцию издержек Кобба–Дугласа из параграфа 1.7:

$$\log \left(\frac{Y_i}{P_{1i}} \right) = \beta_1 + \beta_2 \log(Q_i) + \beta_3 \log \left(\frac{P_{2i}}{P_{1i}} \right) + \beta_4 \log \left(\frac{P_{3i}}{P_{1i}} \right) + \varepsilon_i.$$

¹Оригинальная формулировка теоремы Уайта более общая (т.е. она содержит случай, где $\{e_i, x_i\}$ независимы, но не одинаково распределены (1.1.6)).

Здесь $\mathbf{x}_i' = (1, \log(Q_i), \log(p_{i1}/p_{i3}), \log(p_{i2}/p_{i3}))$ — четырехмерный вектор. Матрица $\mathbf{a}_i \mathbf{x}_i'$ содержит 10 ($= 4 \cdot 5/2$) уникальных элементов:

$$\begin{aligned} & 1, \log(Q_i), \log\left(\frac{p_{i1}}{p_{i3}}\right), \log\left(\frac{p_{i2}}{p_{i3}}\right), \\ & \log(Q_i)^2, \log(Q_i) \cdot \log\left(\frac{p_{i1}}{p_{i3}}\right), \log(Q_i) \cdot \log\left(\frac{p_{i2}}{p_{i3}}\right), \\ & \left[\log\left(\frac{p_{i1}}{p_{i3}}\right)\right]^2, \log\left(\frac{p_{i1}}{p_{i3}}\right) \cdot \log\left(\frac{p_{i2}}{p_{i3}}\right), \left[\log\left(\frac{p_{i2}}{p_{i3}}\right)\right]^2. \end{aligned}$$

$\boldsymbol{\psi}_i$ есть двенадцатимерный вектор, исключивший константу из этого списка. Поэтому m в утверждении 2.6 равно 9.

Если тест Уайта принимает нулевую гипотезу об условной гомоскедастичности, то тогда применяются результаты параграфа 2.6, и статистические выводы могут быть основаны на t - и F -отношениях из главы 1. В противном случае статистические выводы должны основываться на робастных статистиках t и Вальда из утверждения 2.3.

Поскольку регрессоры $\boldsymbol{\psi}_i$ во вспомогательной регрессии имеют много элементов, состоящих из квадратов и перекрестных произведений элементов $\mathbf{x}_i$, тест должен быть состоятельным (то есть его мощность приближается к единице при $n \rightarrow \infty$) против большинства альтернатив гетероскедастичности, но может потребоваться достаточно большая выборка, чтобы получить мощность, близкую к единице. Если исследователь знает, что некоторые элементы $\boldsymbol{\psi}_i$ не влияют на условный второй момент, то их можно исключить из вспомогательной регрессии, и мощность может возрасти на конечных выборках. Недостаток этого, конечно, заключается в том, что, если такое знание ложно, тест не будет иметь никакой мощности против альтернатив о гетероскедастичности, которые соотносят условный второй момент к тем элементам, которые исключены из вспомогательной регрессии.

Контрольные вопросы

1. (Размерность $\boldsymbol{\psi}_i$.) Предположим, что $\mathbf{x}_i = (1, q_i, q_i^2, p_i)'$, четырехмерный вектор. Как много члестоупных и уникальных элементов в $\mathbf{x}_i \mathbf{x}_i'$? (Ответ: 8.)

2.8. Оценивание параметризованной условной гетероскедастичности (дополнительно)

Даже когда обнаружено, что ошибка является условно гетероскедастичной, OLS-оценка все еще состоятельна и асимптотически нормальна, и обоснованные статистические выводы могут производиться с робастными

стандартными ошибками и робастными статистиками Вальда. Однако в (довольно маловероятном) случае *априорного* знания функциональной формы условного второго момента $E(\varepsilon_i^2|x_i)$ возможно получить более точные оценки с меньшей асимптотической дисперсией. Действительно, в теории конечных выборок WLS (взвешенный метод наименьших квадратов, *weighted least squares*) может включать в себя такие знания для повышения эффективности в смысле меньшей дисперсии на конечных выборках. Переносится ли этот результат для конечных выборок на теорию больших выборок? Этот параграф посвящен свойствам WLS-оценки на больших выборках. Чтобы упростить обсуждение, в этом параграфе мы усилимся предположения 2.2 и 2.5, предполагая, что $\{y_i, x_i\}$ является i.i.d. Такое предположение является естественным, поскольку оно обычно в контексте кросс-секционного анализа, где привлекается WLS.

Функциональная форма

Параметрическая функциональная форма для условного второго момента, которую мы рассмотрим, имеет вид:

$$E(\varepsilon_i^2|x_i) = z_i'\alpha, \quad (2.8.1)$$

где z_i — некоторая функция от x_i .

Пример 2.8 (параметрическая форма условного второго момента): Функциональная форма, использованная в WLS-оценивании в эмпирическом упражнении в главе 1, имеет вид

$$E\left[\varepsilon_i^2 \mid \log(Q_i), \log(Q_i)^2, \log\left(\frac{p_{11}}{p_{12}}\right), \log\left(\frac{p_{22}}{p_{21}}\right)\right] = \alpha_1 + \alpha_2 \cdot \left(\frac{1}{Q_i}\right).$$

Поскольку элементы z_i могут быть нелинейными функциями от x_i , эта спецификация является более гибкой, чем могло бы показаться на первый взгляд, но все еще исключает некоторые нелинейности. Например, функциональная форма

$$E(\varepsilon_i^2|x_i) = \exp(z_i'\alpha) \quad (2.8.2)$$

может быть более привлекательной, поскольку ее значение гарантированно положительно. Мы рассмотрим линейную спецификацию (2.8.1) только потому, что оценивание параметров в нелинейных моделях, таких как (2.8.2), требует использования нелинейного метода наименьших квадратов.

WLS с известными α

Чтобы исключить осложнения, возникающие из-за того, что приходится оценивать неизвестный вектор параметров α , мы сначала изучим распределение на больших выборках WLS-оценки $\hat{\alpha}$ известным α . Если α известен, условный второй момент может быть вычислен по данным как $\mathbf{z}_i' \alpha$, и WLS применяется точно так же, как и в параграфе 1.6: делением обеих сторон оцениваемого уравнения $y_i = \mathbf{x}_i' \beta + \varepsilon_i$ на квадратный корень из $\mathbf{z}_i' \alpha$, чтобы получить

$$\tilde{y}_i = \tilde{\mathbf{x}}_i' \beta + \tilde{\varepsilon}_i, \quad (2.8.3)$$

где

$$\tilde{y}_i \equiv \frac{y_i}{\sqrt{\mathbf{z}_i' \alpha}}, \quad \tilde{\mathbf{x}}_i \equiv \frac{\mathbf{x}_i}{\sqrt{\mathbf{z}_i' \alpha}}, \quad \tilde{\varepsilon}_i \equiv \frac{\varepsilon_i}{\sqrt{\mathbf{z}_i' \alpha}},$$

а затем применить OLS. Для дальнейшего использования запишем результатующую WLS-оценку как $\hat{\beta}(V)$. Она задается как

$$\begin{aligned} \hat{\beta}(V) &= \left(\frac{1}{n} \sum_{i=1}^n \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n \tilde{\mathbf{x}}_i \tilde{y}_i = \\ &= \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{\mathbf{z}_i' \alpha} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n \frac{1}{\mathbf{z}_i' \alpha} \mathbf{x}_i y_i = \\ &= (X' V^{-1} X)^{-1} X' V^{-1} y. \end{aligned} \quad (2.8.4)$$

с

$$V = \begin{bmatrix} \mathbf{z}_1' \alpha & & \\ & \ddots & \\ & & \mathbf{z}_n' \alpha \end{bmatrix}.$$

Если предположение 2.3 усилить условием, что

$$E(\varepsilon_i | \mathbf{x}_i) = 0, \quad (2.8.5)$$

то $E(\tilde{\varepsilon}_i | \tilde{\mathbf{x}}_i) = 0$. Чтобы увидеть это, заметим, во-первых, что, поскольку $\tilde{\varepsilon}_i$ есть функция от ε_i ,

$$E(\tilde{\varepsilon}_i | \tilde{\mathbf{x}}_i) = E\left(\frac{\varepsilon_i}{\sqrt{\mathbf{z}_i' \alpha}} \mid \mathbf{x}_i\right) = \frac{1}{\sqrt{\mathbf{z}_i' \alpha}} E(\varepsilon_i | \mathbf{x}_i) = 0. \quad (2.8.6)$$

Во-вторых, поскольку $\tilde{\mathbf{x}}_i$ есть функция от $\mathbf{x}_i$, а $\tilde{\varepsilon}_i$ информации не несет, как и ε_i . Поэтому по закону итерации математических ожиданий мы имеем:

$$E(\tilde{\varepsilon}_i | \tilde{\mathbf{x}}_i) = E[E(\tilde{\varepsilon}_i | \mathbf{x}_i) | \tilde{\mathbf{x}}_i] = 0.$$

Следовательно, при условии, что $E(\tilde{x}_i \tilde{x}_i')$ невырождена, предположения 2.1–2.3 удовлетворяются для уравнения (2.8.3). Кроме того, по построению, ошибка $\tilde{\varepsilon}_i$ условно гомосkedастична: $E(\tilde{\varepsilon}_i^2 | \tilde{x}_i) = 1$. Поэтому утверждение 2.5 применимо с $\sigma^2 = 1$: WLS-оценка является состоятельной и асимптотически нормальной, а асимптотическая дисперсия равна

$$\begin{aligned} \text{Avar}(\hat{\beta}(V)) &= E(\tilde{x}_i \tilde{x}_i')^{-1} \quad (\text{так как дисперсия ошибок есть } 1) = \\ &= \text{plim} \left(\frac{1}{n} \sum_{i=1}^n \tilde{x}_i \tilde{x}_i' \right)^{-1} \quad (\text{по эргодической стационарности}) = \\ &= \text{plim} \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i' \alpha} x_i x_i' \right)^{-1} \quad (\text{так как } \tilde{x}_i = x_i / \sqrt{x_i' \alpha}) = \\ &= \text{plim} \left(\frac{1}{n} X' V^{-1} X \right)^{-1}. \end{aligned} \quad (2.8.7)$$

Поэтому $(\frac{1}{n} X' V^{-1} X)^{-1}$ является состоятельной оценкой $\text{Avar}(\hat{\beta}(V))$.

Регрессия α^2 на x ; дает состоятельную оценку α

Если α неизвестен, его можно оценить, прогоняя отдельную регрессию. Соотношение (2.8.1) говорит, что $x_i' \alpha$ есть регрессия ε_i^2 . Если мы определим $\eta_i = \varepsilon_i^2 - E(\varepsilon_i^2 | x_i)$, то (2.8.1) может быть записано как уравнение регрессии:

$$\varepsilon_i^2 = x_i' \alpha + \eta_i. \quad (2.8.8)$$

По построению $E(\eta_i | x_i) = 0$, что вместе с тем фактом, что x_i есть функция от x_i , влечет ортогональность регрессора x_i ошибке η_i . Следовательно, при условии, что $E(x_i x_i')$ невырождена, утверждение 2.1 применимо к этой вспомогательной регрессии (2.8.8): OLS-оценка для α состоятельна и асимптотически нормальна.

Конечно, мы не можем сделать это, потому что мы не наблюдаем ошибку. Однако, так как OLS-оценка b для первоначальной регрессии $y_i = x_i' \beta + \varepsilon_i$ состоятельна, несмотря на наличие условной гетеросkedастичности, OLS-остатки e_i дают состоятельные оценки для ε_i . Вам остается в качестве аналитического упражнения показать, что, когда ε_i заменяется на e_i в регрессии (2.8.8), OLS-оценка, назовем ее $\hat{\alpha}$, состоятельна для α .

WLS с оцененным α

WLS-оценка первоначального уравнения $y_i = x_i' \beta + \varepsilon_i$ с оцененным α есть $\hat{\beta}(\hat{V})$, где $\hat{V}$ есть диагональная матрица $n \times n$, i -й диагональный элемент которой равен $x_i' \hat{\alpha}$. При соответствующих дополнительных

условиях (см., например, [Аллегруа, 1997] относительно точной формулировки таких условий) можно показать, что

- (a) $\sqrt{n}(\hat{\beta}(V) - \beta)$ и $\sqrt{n}(\hat{\beta}(\hat{V}) - \beta)$ асимптотически эквивалентны в том смысле, что их разность сходится по вероятности к нулю при $n \rightarrow \infty$. Следовательно, по лемме 2.4(а), асимптотическое распределение $\sqrt{n}(\hat{\beta}(\hat{V}) - \beta)$ то же самое, что и у $\sqrt{n}(\hat{\beta}(V) - \beta)$. Поэтому $\text{Avar}(\hat{\beta}(\hat{V}))$ равно $\text{Avar}(\hat{\beta}(V))$, которая, в свою очередь, задается (2.8.7);
- (b) $\text{plim}_n \frac{1}{n} \mathbf{X}'\hat{V}^{-1}\mathbf{X} = \text{plim}_n \frac{1}{n} \mathbf{X}'V^{-1}\mathbf{X}$. Поэтому $(\frac{1}{n} \mathbf{X}'\hat{V}^{-1}\mathbf{X})^{-1}$ состоятельна для $\text{Avar}(\hat{\beta}(\hat{V}))$.

Все это может показаться сложным, но смысл его для WLS-оценивания уравнений $y_i = x_i'\beta + \varepsilon_i$ очень ясен:

Шаг 1: Оцениваем уравнение $y_i = x_i'\beta + \varepsilon_i$ OLS и вычисляем OLS-остатки ε_i .

Шаг 2: Регрессируем ε_i^2 на x_i , чтобы получить OLS-оценку коэффициента α .

Шаг 3: Повторно оцениваем уравнение $y_i = x_i'\beta + \varepsilon_i$, используя WLS с $1/\sqrt{x_i'\hat{\alpha}}$ в качестве веса для наблюдения i .

Так как правильная оценка асимптотической дисперсии, $(\frac{1}{n} \mathbf{X}'\hat{V}^{-1}\mathbf{X})^{-1}$ в (b), есть умноженная на n дисперсионная матрица, обычно распечатываемая в стандартных регрессионных пакетах для шага 3, вычислять соответствующие тестовые статистики довольно просто: для статистических выводов можно использовать стандартные t - и F -отношения из шага 3 регрессии.

Противопоставление OLS и WLS

Таким образом, у нас есть две состоятельные и асимптотически нормальные оценки: OLS- и WLS-оценки. Мы говорим, что состоятельная и асимптотически нормальная оценка **асимптотически более эффективна** (*asymptotically more efficient*), чем другая состоятельная и асимптотически нормальная оценка того же самого параметра, если асимптотическая дисперсия первой не больше, чем последней. Вам остается в качестве аналитического упражнения показать, что WLS-оценка асимптотически более эффективна, чем OLS-оценка.

Преимущество WLS над OLS, однако, опирается на предположку о том, что объем выборки достаточно велик и функциональная форма условного второго момента правильно специфицирована. Если же функциональная форма специфицирована неправильно, WLS-оценка будет все еще состоятельной, но ее асимптотическая дисперсия может

быть, а может и не быть меньше, чем $A_{WLS}(b)$. На конечных выборках, даже если функциональная форма специфицирована верно, аппроксимация для больших выборок будет, вероятно, работать для WLS-оценок хуже, чем для OLS-оценок, поскольку в WLS-процедуре включено оценивание дополнительных параметров (σ^2).

Контрольные вопросы

1. Докажите: « $E(q_n(x)) = 0$ $\Leftrightarrow$ q_n является функцией от x_n » $\Rightarrow$ « $E(x_n \cdot q_n) = 0$ »
Укажите: закон полных математических ожиданий
2. Являются ли ошибки условно homoskedастичными во вспомогательной регрессии (2.5.8)? Если да, имеют ли они влияние на асимптотическое распределение WLS-оценок?

2.9. Проекция методом наименьших квадратов

Что делать, если предположения, обеспечивающие свойства OLS-оценок для больших выборок (за исключением эргодической стационарности), не выполняются, то мы тем не менее действуем по своему усмотрению и применяем OLS к выборке? Что тогда мы оцениваем? Ответ на этот вопрос заключается в следующем: OLS дает оценку лучшего способа линейно объединить объясняющие переменные для предсказания зависимой переменной. Такая линейная комбинация называется **проекцией методом наименьших квадратов** (*least squares projection*).

Оптимальное предсказание значения зависимой переменной

Мы были обеспокоены оцениванием неизвестных параметров по выборке. Давайте временно приоткроем роль эконометристов и попытаемся решить следующую задачу. Существуют случайный скаляр y и случайный вектор x . Мы знаем совместное распределение (y, x) и значение x . На основе этой информации хотим предсказать y . Поэтому предиктор есть функция $f(x)$ от x с функциональной формой $f(\cdot)$, определенной совместным распределением (y, x) . Естественно, мы выбираем функцию $f(\cdot)$ таким образом, чтобы минимизировать некоторый показатель, который измеряет «ошибку» или «ошибку прогноза» (*forecast error*) $y - f(x)$. Мы берем в качестве функции потерь **среднеквадратичную ошибку** (*mean squared error*) $E[(y - f(x))^2]$, поскольку она кажется столь же разумной функцией потерь, как и какая-нибудь другая, и что более важно, поскольку она приводит к следующему удобному результату:

Утверждение 2.7: $E(y|x)$ есть наилучший предиктор для y в том смысле, что он минимизирует среднеквадратичную ошибку.

Мы видели в главе 1, что стратегия прибавления — вычитания эффективна для того, чтобы показать, что вариант решения минимизирует квадратичную функцию. Давайте применим здесь эту стратегию к квадратам ошибок. Пусть $f(x)$ будет любым прогнозом. Прибавим $E(y|x)$ к ошибке прогноза $y - f(x)$ и затем вычтем его для получения разложения

$$y - f(x) = (y - E(y|x)) + (E(y|x) - f(x)). \quad (2.9.1)$$

Поэтому квадрат ошибки прогноза равен

$$(y - f(x))^2 = (y - E(y|x))^2 + 2(y - E(y|x))(E(y|x) - f(x)) + (E(y|x) - f(x))^2 \quad (2.9.2)$$

Берем математические ожидания от обеих сторон и получаем:

$$\begin{aligned} \text{среднеквадратичная ошибка} &= E[(y - f(x))^2] = \\ &= E[(y - E(y|x))^2] + 2E[(y - E(y|x))(E(y|x) - f(x))] + \\ &+ E[(E(y|x) - f(x))^2]. \end{aligned} \quad (2.9.3)$$

Непосредственным применением закона больших математических ожиданий можно показать, что средняя компонента, являющаяся ковариацией между ошибкой оптимального прогноза и разностью между прогнозами, равна нулю (контрольный вопрос). Следовательно,

$$\begin{aligned} \text{среднеквадратичная ошибка} &= E[(y - E(y|x))^2] + \\ &+ E[(E(y|x) - f(x))^2] \geq E[(y - E(y|x))^2], \end{aligned} \quad (2.9.4)$$

что показывает, что среднеквадратичная ошибка ограничена снизу величиной $E[(y - E(y|x))^2]$, и эта нижняя граница достигается указанным математическим ожиданием.

Наилучший линейный предиктор

Для вычисления $E(y|x)$, которое может быть сильно нелинейным, требуется знание совместного распределения (y, x) . Мы теперь ограничимся предикторами, которые являются линейными функциями от x и зададимся вопросом: какой предиктор является наилучшим (в смысле минимизации среднеквадратичной ошибки) линейным предиктором для y , зависящим от x ? Для этой цели рассмотрим β^* , который удовлетворяет условию ортогональности

$$E[x \cdot (y - x'\beta^*)] = 0 \quad \text{или} \quad E(xx'\beta^*) = E(xy). \quad (2.9.5)$$

Идея состоит в том, чтобы выбрать β^* таким образом, чтобы ошибка прогноза $y - x'\beta^*$ была ортогональна x . Если $E(x x')$ невырождена, условие ортогональности может быть разрешено относительно β^* :

$$\beta^* = [E(x x')]^{-1} E(x \cdot y). \quad (2.9.6)$$

Проекция методом наименьших квадратов (для линейной) (least squares (or linear) projection) y на x , обозначаемая $\hat{y}(x)$, определяется как $x'\beta^*$, где β^* удовлетворяет (2.9.6) и называется **коэффициентом проекции методом наименьших квадратов (least squares projection coefficient)**.

Утверждение 2.8: Проекция методом наименьших квадратов $\hat{y}(x)$ есть наилучший линейный предиктор y в том смысле, что он минимизирует среднеквадратичную ошибку.

Операция прибавления ** вычитания ** деления ** деления

Доказательство

Для любого линейного предиктора $x'\tilde{\beta}$

$$\begin{aligned} \text{среднеквадратичная ошибка} &= E[(y - x'\tilde{\beta})^2] \\ &= E\{(y - x'\beta^*) + x'(\beta^* - \tilde{\beta})\}^2 \} \\ &\text{(по стратегии прибавления — вычитания)} \\ &= E[(y - x'\beta^*)^2] + 2(\beta^* - \tilde{\beta})' E[x \cdot (y - x'\beta^*)] + E[(x'(\beta^* - \tilde{\beta}))^2] = \\ &= E[(y - x'\beta^*)^2] + E[(x'(\beta^* - \tilde{\beta}))^2] \\ &\text{(по условию ортогональности (2.9.5))} \\ &\geq E[(y - x'\beta^*)^2]. \end{aligned}$$

В отличие от предыдущего предиктора, который является условным математическим ожиданием, наилучший линейный предиктор требует для своего вычисления только знания второго момента совместного распределения (y, x) для вычисления (см. (2.9.6)).

Если один из регрессоров x является константой, коэффициенты проекции методом наименьших квадратов могут быть записаны в терминах дисперсий и ковариаций. Пусть $\bar{x}$ будет вектором постоянных регрессоров, так что

$$x = \begin{bmatrix} 1 \\ \bar{x} \end{bmatrix}.$$

В аналитическом упражнении к этой главе вас просят доказать, что

$$\hat{E}^*(y|x) \equiv \hat{E}^*(y|1, \hat{x}) = \mu + \gamma' \hat{x}, \quad (2.9.7)$$

где

$$\gamma = \text{Var}(\hat{x})^{-1} \text{Cov}(\hat{x}, y), \quad \mu = E(y) - \gamma' E(\hat{x}).$$

Эта формула является популяционным аналогом формулы регрессии в отклонениях от среднего (*deviations from the mean*) из главы 1 (аналитическое упражнение 3).

OLS consistently оценивает коэффициенты проекции

Давайте возвратимся теперь к роли эконометриста и рассмотрим оценивание β^* . Предположим, что у нас есть выборка объема n , извлеченная из эргодической стационарной случайного процесса $\{y_t, x_t\}$ с совместным распределением $\{y_t, x_t\}$ (не зависящим от t из-за стационарности), которое совпадает с распределением (y, x) выше. Поэтому, например, $E(x_t x_t')$ — $E(xx')$. По эргодической теореме вторые моменты в (2.9.6) могут быть consistently оценены соответствующими выборочными вторыми моментами. Таким образом, состоятельной оценкой вектора коэффициентов проекции β^* является:

$$\left(\frac{1}{n} \sum_{i=1}^n x_i x_i'\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n x_i \cdot y_i\right) = (X'X)^{-1} X'y$$

которая есть не что иное, как OLS-оценка b . То есть при предположении 2.2 (эргодическая стационарность) и предположении 2.4, гарантирующем невырожденность $E(xx')$, OLS-оценка всегда состоятельна для вектора коэффициентов проекции β^* , который удовлетворяет условию ортогональности (2.9.5).

Контрольные вопросы

1. (Неприводимость ошибок прогноза.) Для прогноза с минимальным квадратом ошибки прогноза $E(y|x)$ ошибка прогноза ортогональна любой функции $\phi(x)$ от x . То есть $E(\phi(x)\epsilon) = 0$, где $\epsilon = y - E(y|x)$. Докажите это. **Указание:** Замените под знаком математических ожиданий.
Покажите, что средний член в правой части (2.9.3) равен нулю, используя $\phi(x) = E(y|x) - f(x)$.
2. (Прогнозирование белого шума.) Предположим, что $\{\epsilon_t\}$ — белый шум. Что есть $\hat{E}^*(\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-m})$? Что есть $\hat{E}^*(\epsilon_t | 1, \epsilon_{t-1}, \dots, \epsilon_{t-m})$? Верно ли, что $\text{Cov}(\epsilon_t, \epsilon_{t-1}, \dots, \epsilon_{t-m}) = 0$? **Указание:** Является ли он нулем для процесса в примере 2.4?

3. (Устойчивые метастатические состояния, которые линейны.) Предположим, что $E(y|\tilde{x}) = \mu + \gamma'\tilde{x}$. Покажите, что $E'(q|1, \tilde{x}) = E(y|\tilde{x})$.
4. (Ближайшая проекция.) Рассмотрим модель $y_i = x_i(\beta - z_i\delta - \varepsilon_i) + \varepsilon_i$, $\varepsilon_i \in E(x_i, z_i) = 0$, $E(z_i, \varepsilon_i) \neq 0$ и $E(z_i x_i') = 0$. Таким образом, z_i не предопределен (то есть не ортогонален ошибке), но не является предопределенным регрессором x_i в том смысле, что перекрестные моменты равны нулю.
- (a) Покажите, что коэффициент проекции метода наименьших квадратов для x_i в проекции y_i на x_i и z_i равен β . Указание: Выделите $E'(z_i|x_i, x_i)$.
- (b) Чему равен коэффициент проекции метода наименьших квадратов для z_i в проекции y_i на x_i ? Указание: Трактуйте $z_i\delta - \varepsilon_i$ как ошибку.
- (c) Какую проекцию вы бы использовали для оценивания β ? Указание: Вы хотите, чтобы дисперсия ошибки была поменьше.

2.10. Тестирование сериальной корреляции

Как было отмечено в параграфе 2.3 (см. (2.3.3)), если регрессоры включают константу (это так практически во всех известных приложениях), из предположения 2.5 вытекает, что ошибки являются скалярной последовательностью маргинал-разностей (m.d.s.), поэтому если обнаружена сериальная корреляция ошибок, это является указанием на нарушение предположения 2 б. Сериальная корреляция традиционно была важной темой в эконометрике, и имеется необходимое количество тестов на сериальную корреляцию (то есть проверка нулевой гипотезы об отсутствии сериальной корреляции у ошибок). Некоторые из них, однако, требуют, чтобы регрессоры были строго экзогенными. Тест, который будет представлен в этом параграфе, не требует экзогенности. Поскольку проблема сериальной корреляции возникает только в моделях временных рядов, мы используем в данном параграфе (и в следующих) индекс «t» вместо «i». Всяду в этом параграфе мы предполагаем, что регрессоры включают константу.

Было бы хорошо обобщить эти тесты для учета сериальной корреляции в g_t ($\equiv x_t + z_t$), но таких тестов, которые получили бы признание, не было предложено. Это пробел в литературе, но не серьезный, поскольку в настоящее время исследователи знают, как жить с сериальной корреляцией в g_t . То есть, как будет показано в главе 6, имеется способ сделать вывод о присутствии сериальной корреляции в g_t . Ранее в этой главе мы изучали, как вычислять стандартные ошибки, робастные к условной гетероскедастичности. В главе 6 эти стандартные ошибки будут сделаны также робастными к сериальной корреляции.

Статистика Бокса — Пирса и Льюнга — Бокса

Прежде чем перейти к тестам на сериальную корреляцию ошибок, мы временно переходим из рамки регрессионной основы и рассмотрим сериальную корреляцию наблюдённых временных ошибок. Предположим, что у нас есть выборка объёма n , $\{z_1, \dots, z_n\}$, качественно из скалярного стационарного и дифференцируемого процесса. В параграфе 2.2 мы определили (популярно-нон) автокорреляцию j -го порядка как γ_j . **Выборочная автоковариация j -го порядка** (*sample j th order autocovariance*) определяется соответствием:

$$\hat{\gamma}_j \equiv \frac{1}{n} \sum_{t=j+1}^n (z_t - \bar{z}_n)(z_{t-j} - \bar{z}_n) \quad (j = 0, 1, \dots), \quad (2.10.1)$$

где

$$\bar{z}_n = \frac{1}{n} \sum_{t=1}^n z_t.$$

(Если предельное среднее $E(z_t)$ известно, можно заменять им выборочное среднее; для малых n можно улучшить свойства на конечных выборках.) Здесь, даже если и нужно использовать только $n - j$ наблюдений, выходящих за пределы n , а не $n - j$. Делится ли сумма на $n - j$ или на n — это не влияет на распределение на больших выборках. Оно верно для выборок умеренного объёма численно разделение может быть существенным, и мы должны всегда быть точны в том, что используется в знаменателе. **Выборочный коэффициент автокорреляции j -го порядка** (*sample j -th order autocorrelation coefficient*), $\hat{\rho}_j$, определяется как

$$\hat{\rho}_j \equiv \frac{\hat{\gamma}_j}{\hat{\gamma}_0} \quad (j = 1, 2, \dots). \quad (2.10.2)$$

Если $\{z_t\}$ порядково стационарный, то легко показать тем, что $\hat{\gamma}_j$ является состоятельной для γ_j ($j = 0, 1, 2, \dots$). Следовательно, по лемме 2.3(a), $\hat{\rho}_j$ является состоятельной для ρ_j ($j = 1, 2, \dots$). В частности если $\{z_t\}$ сериально некоррелирован, то тогда все выборочные коэффициенты автокорреляции сойдётся к 0 по вероятности. Для тестирования сериальной корреляции т.ч. однако, надо знать асимптотическое распределение $\sqrt{n}\hat{\rho}_j$. Это обеспечивается

Утверждение 2.8 (частный случай теоремы 6.7 [Hall and Heyde, 1980]): Предположим, что $\{z_t\}$ может быть записан как $\mu + \varepsilon_t$, где ε_t — стационарная последовательность марковских разностей с «собственными» условиями компактности:

(собственные условия компактности)

$$\hat{\gamma}_j^2 \hat{\gamma}_0^{-1} \varepsilon_{t-1} \varepsilon_{t-2} \dots \varepsilon_{t-j} \rightarrow 0 \quad (\sigma^2, \sigma'^2 > 0),$$

Первое выборочная автокорреляция $\hat{\rho}_1$ определяется, как в (2.10.1) и (2.10.2). Тогда

$$\sqrt{n}\hat{\rho}_1 \rightarrow_d N(0, \sigma^4 I_p) \quad \text{и} \quad \sqrt{n}\hat{\rho} \rightarrow_d N(0, I_p),$$

где $\hat{\rho} = (\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_p)'$ и $\hat{\rho} = (\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_p)'$.

Здесь от процесса x_t не требуется быть эргодическим. Доказательство этого при дополнительном условии эргодичности остается в качестве аналитического упражнения. Таким образом, умноженные на $\sqrt{n}$ автокорреляции асимптотически являются i.i.d. и имеют распределение $N(0, 1)$. Процесс $\{x_t\}$ предполагается здесь более общим, чем независимый процесс белого шума, но условием, который должен быть выполнен, является то, что этому результату, например, не удовлетворяет ARCH-процесс.

Один из способов тестирования серийной корреляции в x_t — проверить, равна ли ρ автокорреляция первого порядка, ρ_1 . Утверждение 2.9 гласит

$$\sqrt{n}\hat{\rho}_1 = \frac{\hat{\rho}_1}{1/\sqrt{n}} \rightarrow_d N(0, 1). \quad (2.10.3)$$

Поэтому t -статистика, сформированная как отношение $\hat{\rho}_1$ к «стандартной ошибке» $1/\sqrt{n}$, является асимптотически стандартной нормальной.

Мы можем также тестировать на одновременное равенство нулю целой группы автокорреляций. Пусть $\hat{\rho} = (\hat{\rho}_1, \dots, \hat{\rho}_p)'$ является p -мерным вектором, содержащим первые p выборочных автокорреляций. Так как компоненты $\sqrt{n}\hat{\rho}$ асимптотически независимы и одинаково распределены как стандартные нормальные, их сумма квадратов, называемая **Q Бокса — Пирса** (Box — Pierce Q), поскольку она была впервые предложена в [Box and Pierce, 1970], имеет квантильное распределение хи-квадрат:

$$\text{статистика } Q \text{ Бокса — Пирса} \equiv n \sum_{j=1}^p \hat{\rho}_j^2 = \sum_{j=1}^p (\sqrt{n}\hat{\rho}_j)^2 \rightarrow_d \chi^2(p). \quad (2.10.4)$$

Легко доказать (см. контрольный выпуск), что следующая модифицированная версия **Q Льюнга — Бокса** (Ljung — Box Q) является асимптотически эквивалентной той же величине, что ее аналог от Q Бокса — Пирса исчезает на больших выборках. Поэтому, по лемме 2.4(a), она также имеет асимптотическое распределение хи-квадрат:

Статистика Q Льюнга — Бокса $\equiv$

$$n(n+2) \sum_{j=1}^p \frac{\hat{\rho}_j^2}{n-j} = \sum_{j=1}^p \frac{n+2}{n-j} (\sqrt{n}\hat{\rho}_j)^2 \rightarrow_d \chi^2(p). \quad (3.10.5)$$

Эта модификация часто дает лучшую аппроксимацию распределения хи-квадрат для умеренных объемов выборки (она будет предложена ниже, когда это в упражнении Монт-Калло). Для каждой из статистик нет четкого руководства по выбору p . Если p слишком мало, то есть опускать прочесть существующей автокорреляции выходного процесса, то если p слишком велико относительно объема выборки, распределение статистики на конечных выборках, вероятно, ухудшится, значительно удаляясь от expected chi-квадрат.

Выборочные попарные корреляции, вычисленные с помощью остатков

Вспомогательная регрессионная модель, описанной предположениями 2.1–2.5. Если бы ошибка ε_t была независимой, мы могли бы вычислять выборочные автокорреляции как

$$\hat{\rho}_j \equiv \frac{\sum_{t=1}^n \varepsilon_t \varepsilon_{t-j}}{n}, \quad (j = 1, 2, \dots), \quad (2.10.6)$$

где

$$\hat{\rho}_j = \frac{1}{n} \sum_{t=1}^n \varepsilon_t \varepsilon_{t-j}, \quad (j = 0, 1, 2, \dots). \quad (2.10.7)$$

(Нет необходимости вычитать выборочное среднее, поскольку популяционное среднее равно нулю, как результат исключения константы в число регрессии.) Так как $\{\varepsilon_t, \varepsilon_{t-j}\}$ рассматривается стационарным, по предположению 2.2, $\hat{\rho}_j$ сходится по вероятности к соответствующему популяционному среднему, $E(\varepsilon_t \varepsilon_{t-j})$, для всех j , и $\hat{\rho}_j$ является состоятельной для истинного j -го коэффициента автокорреляции для ε_t .

Далее рассмотрим более реалистичный случай, когда мы не знаем истинной ошибки. Мы можем заменить ε_t в формуле выше на его OLS-оценку $\hat{\varepsilon}_t$ и вычислять выборочную автокорреляцию как

$$\hat{\rho}_j \equiv \frac{\sum_{t=1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-j}}{n}, \quad (j = 1, 2, \dots), \quad (2.10.8)$$

где

$$\hat{\rho}_j \equiv \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-j}, \quad (j = 0, 1, 2, \dots). \quad (2.10.9)$$

(Поскольку регрессионная константа константа, то и модельные уравнения, соответствующей константе, галитерфет, что выборочное среднее $\hat{\varepsilon}_t$ равно нулю. Поэтому нет необходимости вычитать выборочное среднее.) Правильно ли использовать $\hat{\rho}_j$ (вычисленное с помощью остатков) вместо $\hat{\rho}_j$,

и основанную на остатках статистику Q , полученную из $\{\hat{\rho}_j\}$, для тестирования сериальной корреляции? [Ответ: Да. Но только если регрессоры строго экзогенны.]

Вспомним выражение, связывающее остаток e_t и действительную ошибку ε_t , которое было дано в (2.3.9). Используя его, можно записать:

$$\begin{aligned}\hat{\rho}_j &= \frac{1}{n} \sum_{t=j+1}^n e_t e_{t-j} \\ &= \frac{1}{n} \sum_{t=j+1}^n [y_t - x_t'(b - \beta)][y_{t-j} - x_{t-j}'(b - \beta)] \\ \hat{\rho}_j &= \frac{1}{n} \sum_{t=j+1}^n (x_{t-j} \cdot \varepsilon_t - x_t \cdot \varepsilon_{t-j})'(b - \beta) + \\ &+ (b - \beta)' \left(\frac{1}{n} \sum_{t=j+1}^n x_t x_{t-j}' \right) (b - \beta).\end{aligned}\quad (2.10.10)$$

Если $E(x_t \cdot x_{t-1}')$, $E(x_{t-2} \cdot x_t')$ и $E(x_t x_{t-j}')$ все конечны, то вторая и третья компоненты исчезают (сходятся к нулю по вероятности), поскольку $b - \beta \rightarrow_p 0$. Следовательно,

$$\hat{\rho}_j \rightarrow_p \hat{\rho}_j^* = 0 \quad (j = 0, 1, 2, \dots),$$

и, таким образом, различие между $\hat{\rho}_j$ и $\hat{\rho}_j^*$ также исчезает на больших выборках.

Однако это будет не так, если обе эти величины умножить на $\sqrt{n}$. $\sqrt{n}\hat{\rho}_j$ и $\sqrt{n}\hat{\rho}_j^*$ могут быть записаны как

$$\sqrt{n}\hat{\rho}_j = \frac{\sqrt{n}\hat{\rho}_j^*}{\hat{\sigma}_\varepsilon^2} \quad \text{и} \quad \sqrt{n}\hat{\rho}_j = \frac{\sqrt{n}\hat{\rho}_j}{\hat{\sigma}_\varepsilon^2}.\quad (2.10.11)$$

Мы уже видели, что для обоих $\hat{\rho}_j$ и $\hat{\rho}_j^*$ знаменатель $\rightarrow_p \sigma^2$. Поэтому различие между $\hat{\rho}_j$ и $\hat{\rho}_j^*$ будет исчезать, если будет исчезать значение между $\hat{\rho}_j$ и $\hat{\rho}_j^*$ (если вы не уверены, см. контрольный вопрос 3 ниже). Теперь, умножая обе стороны (2.10.10) на $\sqrt{n}$, мы получаем:

$$\begin{aligned}\sqrt{n}\hat{\rho}_j &= \sqrt{n}\hat{\rho}_j^* - \frac{1}{n} \sum_{t=j+1}^n (x_{t-j} \cdot \varepsilon_t - x_t \cdot \varepsilon_{t-j})' \sqrt{n}(b - \beta) + \\ &+ \sqrt{n}(b - \beta)' \left(\frac{1}{n} \sum_{t=j+1}^n x_t x_{t-j}' \right) (b - \beta).\end{aligned}\quad (2.10.12)$$

Поскольку $\sqrt{n}(\hat{\beta} - \beta)$ сходится к случайной величине (распределение которой нормальное), третья компонента в правой части исчезает, по лемме 2.4(b). Что касается второй компоненты, мы имеем:

$$\frac{1}{n} \sum_{i=j+1}^n (x_{t-j} - \varepsilon_t + x_t - \varepsilon_{t-j}) \xrightarrow{p} E(x_{t-j} \cdot \varepsilon_t) + E(x_t - \varepsilon_{t-j}). \quad (2.10.13)$$

Если регрессоры строго экзогенны в том смысле, что $E(x_t \cdot \varepsilon_s) = 0$ для всех t, s , то

$$E(x_{t-j} \cdot \varepsilon_t) + E(x_t - \varepsilon_{t-j}) = 0, \quad (2.10.14)$$

и, по лемме 2.4(b), вторая компонента сходится к нулю по вероятности. Поэтому различие между $\sqrt{n}\hat{\rho}_j$ и $\sqrt{n}\bar{\rho}_j$ исчезает, и это означает, что статистика Q , вычисленная с помощью остатков $\{\varepsilon_t\}$, также имеет асимптотическое распределение хи-квадрат, и мы можем использовать эту основанную на остатках статистику Q для тестирования серийной корреляции. Если, с другой стороны, регрессоры не являются строго экзогенными, то тогда нет гарантии того, что (2.10.14) выполняется. Следовательно, основанная на остатках статистика Q может не иметь асимптотического распределения хи-квадрат.

**Тестирование с предопределенными,
но не строго экзогенными регрессорами**

Таким образом, если регрессоры не являются строго экзогенными, нам необходимо модифицировать статистику Q , чтобы восстановить ее асимптотическое распределение. Для этой цели рассмотрим два ограничения:

(более сильная форма предопределенности)

$$E(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, x_t, x_{t-1}, \dots) = 0. \quad (2.10.15)$$

(более сильная форма условной гомоскедастичности)

$$E(\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, x_t, x_{t-1}, \dots) = \sigma^2 > 0 \quad (2.10.16)$$

Первое условие просто воспроизводит (2.3.1) (только теперь с t в качестве индекса). Как было показано в параграфе 2.3, оно более сильное, чем предположение 2.3, и из него вытекает, что $g_t (\equiv x_t \cdot \varepsilon_t)$ является m.d.s. Условие (2.10.16) с условным множеством, включающим x_t , очевидно более сильное, чем предположение 2.7 (условная гомоскедастичность). Оно также более сильное, чем предположение о собственной условной гомоскедастичности и утверждение 2.9, поскольку условное множество включает текущие и прошлые x , так же как и прошлые ε . Следующий результат показывает, что при этих дополнительных условиях существует соответствующая модификация статистики Q .

Утверждение 2.10 (тестирование серийной корреляции с предопределенными регрессорами): Предположим, что удовлетворяются предположения 2.1, 2.2, 2.4, (2.10.15) и (2.10.16). Пусть выборочная автокорреляция OLS-остатков, $\hat{\rho}_j$, определяется, как в (2.10.8). Тогда

$$\sqrt{n}\tilde{\gamma} \rightarrow N(0, \sigma^4(I_p - \Phi)) \text{ и } \sqrt{n}\tilde{\rho} \rightarrow N(0, I_p - \Phi), \quad (2.10.17)$$

где $\tilde{\gamma} = (\tilde{\gamma}_1, \tilde{\gamma}_2, \dots, \tilde{\gamma}_p)'$, $\tilde{\rho} = (\tilde{\rho}_1, \tilde{\rho}_2, \dots, \tilde{\rho}_p)'$ и ϕ_{jk} (элемент $\equiv (j, k)$ матрицы Φ размерности $p \times p$) задается как

$$\phi_{jk} = E(x_i \cdot x_{i-j})' E(x_i x_i')^{-1} E(x_i \cdot x_{i-k}) / \sigma^2, \quad (2.10.18)$$

Доказательство, хотя и не очень сложное, перенесено в приложение 2.B. По эргодической теореме, матрица Φ состоятельно оценивается своим выборочным аналогом:

$$\tilde{\Phi} \equiv (\tilde{\phi}_{jk}), \quad \tilde{\phi}_{jk} \equiv \hat{\rho}_j' S_{xx}^{-1} \bar{\mu}_k / s^2 \quad (j, k = 1, 2, \dots, p), \quad (2.10.19)$$

где

$$s^2 \equiv \frac{1}{n} \sum_{i=1}^n e_i^2, \quad \bar{\mu}_j \equiv \frac{1}{n} \sum_{i=j+1}^n x_i \cdot x_{i-j}.$$

Из этого и из предположение 2.10 следует, что

$$\text{модифицированная } Q \text{ Бокса-Пиркса} \equiv n \cdot \tilde{\rho}' (I_p - \tilde{\Phi})^{-1} \tilde{\rho} \xrightarrow{d} \chi^2(p). \quad (2.10.20)$$

Пока регрессоры предопределены и ошибки условно гомоскедастичны в смысле (2.10.15) и (2.10.16), эта модифицированная статистика Q может быть использована, когда регрессоры не являются строго экзогенными.

Тест, основанный на вспомогательной регрессии

Хотя вычисление этой модифицированной статистики Q выполняется непосредственно с помощью матричного программного обеспечения, полезно найти асимптотически эквивалентную статистику, которая может быть вычислена в регрессионных пакетах. Для этой цели рассмотрим вспомогательную регрессию:

$$\text{регрессия } e_t \text{ на } x_t, e_{t-1}, e_{t-2}, \dots, e_{t-p}. \quad (2.10.21)$$

Чтобы признать эту вспомогательную регрессию для $t = 1, 2, \dots, n$, нам нужны данные о $(e_t, e_{t-1}, \dots, e_{t-p})$. Асимптотически, не имеет значения,

какое конкретное значение им присвоить, но кажется разумным положить их равными 0, их ожидаемому значению¹. Из этой вспомогательной регрессии мы можем вычислить F -статистику для гипотезы о том, что все p коэффициентов при $e_{t-1}, e_{t-2}, \dots, e_{t-p}$ равны нулю. Учитывая утверждение 2.5(с), вполне естественно задать вопрос: имеет ли $p \cdot F$ асимптотическое распределение $\chi^2(p)$? Эта догадка действительно верна; при гипотезе в утверждении 2.10 модифицированная статистика Q (2.10.20) асимптотически эквивалентна $p \cdot F$ (то есть разность между ними сходится к нулю по вероятности при $n \rightarrow \infty$), поэтому $p \cdot F$ также имеет асимптотическое распределение хи-квадрат (демонстрации этого посвящено аналитическое упражнение).

Эта $p \cdot F$ -статистика оказывается асимптотически эквивалентной nR^2 из вспомогательной регрессии, что может быть показано следующим образом. Вспомним алгебраический результат из главы 1 о том, что F -отношение может быть вычислено на основе разности сумм квадратов остатков от неограниченной и ограниченной регрессий. Неограниченная регрессия в текущем контексте есть (2.10.21), в то время как ограниченная регрессия есть

$$\text{регрессия } e_t \text{ на } x_t. \quad (2.10.22)$$

Следовательно, если SSR_{it} и SSR_{jt} являются SSR из (2.10.21) и (2.10.22), соответственно, мы имеем:

$$p \cdot F = \frac{SSR_{jt} - SSR_{it}}{SSR_{it} / (n - \#x_t - p)} = (n - \#x_t - p) \frac{SSR_{jt} - SSR_{it}}{SSR_{it}}. \quad (2.10.23)$$

где $\#x_t$ — число переменных в x_t . Однако поскольку e_t является остатком от первоначальной регрессии (регрессии y_t на x_t), регрессоры x_t в (2.10.22) не имеют объясняющей силы. Поэтому $SSR_{jt} = e'e$, где e — n -мерный вектор остатков из первоначальной регрессии, а (2.10.23) численно идентична

$$(n - \#x_t - p) \frac{R_{it}^2}{1 - R_{it}^2},$$

где R_{it}^2 — нецентрированный R^2 для вспомогательной регрессии (2.10.21)². Но поскольку, по построению, выборочное среднее e_t равно нулю (это потому, что x_t включает константу), R_{it}^2 численно идентичен R^2 для вспо-

¹ Другой асимптотически эквивалентный выбор заключается в проведении первоначальной регрессии для $t = p+1, p+2, \dots, n$.

² Вспомогая из (1.2.16), что нецентрированный R^2 для регрессии y на x определяется как

$$R^2 = \frac{y'y - e'e}{y'y},$$

в текущем контексте $y'y$ в этой формуле есть $e'e$, а $e'e$ — $e'e$ в формуле есть SSR_{jt} .

минимальной регрессии. Таким образом, мы доказали следующий алгебраический результат:

$$p \cdot F = (n - \sum_{t=1}^n x_t - p) \frac{H^2}{1 + \frac{p \cdot F}{n - \sum_{t=1}^n x_t - p}},$$

где R^2 равен H^2 по минимальной регрессии (2.10.21). Решая это уравнение относительно H^2 и умножая обе стороны на n , получаем:

$$n R^2 = \frac{n}{n - \sum_{t=1}^n x_t - p} \cdot \frac{1}{1 + \frac{p \cdot F}{n - \sum_{t=1}^n x_t - p}} \cdot p \cdot F.$$

Поскольку $\frac{p \cdot F}{n - \sum_{t=1}^n x_t - p} \rightarrow 0$ при $n \rightarrow \infty$ (как следствие леммы 2.4(b)), это уравнение показывает, что $p \cdot F \sim n R^2 \rightarrow 0$. Следовательно, $n R^2$ из минимальной регрессии (2.10.21) имеет асимптотическое распределение $\chi^2(p)$. Тест, основанный на $n R^2$, называется **тестом Брейша — Годфри на сериальную корреляцию** (*Breusch—Godfrey test for serial correlation*). Когда $p = 1$, эта тестовая статистика асимптотически эквивалентна квадрату статистики, известный как **статистика h Дарбина** (*Durbin's h*)¹.

Контрольные вопросы

1. (Состоятельность выборочных автокорреляций.) Покажите: если $\{z_t\}$ эргодически стационарный, то тогда выборочная автоковариация $\hat{\gamma}_1$, заданная в (2.10.1), является состоятельной для γ_1 , попарной автоковариации. Укажите: $\gamma_1 = E(z_t z_{t-1}) - E(z_t) E(z_{t-1})$. Перепишите $\hat{\gamma}_1$ как

$$\hat{\gamma}_1 = \frac{1}{n} \sum_{t=2}^n z_t z_{t-1} - \bar{z}_n \frac{1}{n} \sum_{t=2}^n z_{t-1} - \bar{z}_n \frac{1}{n} \sum_{t=2}^n z_t + n \cdot \bar{z}_n^2 = (\bar{z}_n)^2.$$

2. (Асимптотическая эквивалентность двух Q .) Докажите, что разность между Q Бокса — Пирса и Q Льюнга — Бокса сходится к 0 по вероятности при $n \rightarrow \infty$ (потому что асимптотические распределения одинаковы). Укажите: Пусть

$$x_n = \left((\sqrt{n} \hat{\rho}_1)^2, \dots, (\sqrt{n} \hat{\rho}_p)^2 \right)'$$

Найдите p -мерный вектор a_n , такой, чтобы разность между двумя Q была $a_n' x_n$. Покажите, что $a_n \rightarrow 0$. Асимптотическое свойство, которое нужно использовать, — лемма 2.4(b).

¹См. [Breusch, 1978] и [Godfrey, 1978]. Тест Брейша — Годфри был первоначально предложен как тест множителей Лагранжа на нулевую попарную корреляцию ошибок, состоящих из дискретизированных регрессоров x и дискретизированной зависимой переменной. Обобщенная версия показывается, что тест применим к более общему случаю рассмотренному в утверждении 2.10.

3. Рассмотрим $\sqrt{n}\hat{\rho}_j$ и $\sqrt{n}\hat{\rho}_j^*$ в (2.10.11). Мы показали, что $\hat{\gamma}_0 \rightarrow \sigma^2$ и $\hat{\gamma}_0^* \rightarrow \sigma^2$. По утверждению 2.9 $\sqrt{n}\hat{\gamma}_j \rightarrow_d N(0, \sigma^4)$. Принимая эти утверждения как доказанные, покажите, что $\sqrt{n}\hat{\rho}_j - \sqrt{n}\hat{\rho}_j^* \rightarrow_p 0$, если $\sqrt{n}\hat{\gamma}_j - \sqrt{n}\hat{\gamma}_j^* \rightarrow_p 0$.
Указание: Если $\sqrt{n}\hat{\gamma}_j - \sqrt{n}\hat{\gamma}_j^* \rightarrow_p 0$, то, по лемме 2.4(a), $\sqrt{n}\hat{\gamma}_j \rightarrow_d N(0, \sigma^4)$.

$$\frac{\hat{\gamma}_j}{\hat{\gamma}_0} - \frac{\hat{\gamma}_j^*}{\hat{\gamma}_0^*} = \hat{\gamma}_j \left(\frac{1}{\hat{\gamma}_0} - \frac{1}{\sigma^2} \right) - \hat{\gamma}_j^* \left(\frac{1}{\hat{\gamma}_0^*} - \frac{1}{\sigma^2} \right) + \frac{1}{\sigma^2} (\hat{\gamma}_j - \hat{\gamma}_j^*).$$

Умножьте обе стороны на $\sqrt{n}$ и примените лемму 2.4(b).

2.11. Приложение: эконометрика рациональных ожиданий

В динамических экономических моделях и в других контекстах ожидания относительно будущего, естественно, играют важную роль. **Эконометрика рациональных ожиданий** (*rational expectations econometrics*) относится к уравнениям, включающим ожидания. Ожидания обычно ненаблюдаемы, но, если мы предполагаем, что экономические агенты формируют ожидания рационально, мы можем преодолеть проблему ненаблюдаемых ожиданий, используя технику этой главы. Приложением, которое мы рассмотрим, является **гипотеза эффективности рынка** (*efficient market hypothesis*) Фэми. По его словам, «эффективным рынком капитала является рынок, который эффективен при обработке информации». <...> В эффективном рынке цены «в полной мере отражают доступную информацию» [Fama, 1976, p. 133]. Слова «в полной мере отражают» могут быть точно формализованы. Мы будем делать это для конкретного рынка капитала и тестировать следствия из этой гипотезы.

Гипотеза эффективности рынка

Рынок капитала, изученный в [Fama, 1973], является рынком казначейских векселей США. В этом параграфе мы сосредоточимся на ставках по одномесячным казначейским векселям, наблюдаемых на основе месячного базиса. (Позднее в этой книге мы также рассмотрим казначейские векселя с различными сроками погашения.) Для формализации гипотезы эффективности рынка для рынка казначейских векселей нам нужно ввести некоторые понятия из макроэкономики и финансов. Поскольку в данном параграфе мы имеем дело исключительно с временными рядами, в качестве индекса мы используем «t», а не «i». Определим,

v_t = цена одномесячных казначейских векселей на начало месяца t .

R_t = одномесячная номинальная процентная ставка за месяц t ,

то есть номинальная доходность по векселю за месяц, равная

$(1 - v_t)/v_t$, так что $v_t = 1/(1 + R_t)$.

¹Из приведенных ниже формул следует, что индекс t ввиду векселя номинальной стоимостью 1.

P_t — значение индекса потребительских цен (ИПЦ),

который является нашей мерой уровня цен на начало месяца t .

π_{t+1} — темп инфляции за месяц t (то есть с начала месяца t по начало месяца $t+1$) $= (P_{t+1} - P_t)/P_t$.

π_{t+1} — ожидаемый темп инфляции за месяц t , ожидание формируется в начале месяца t .

η_{t+1} — ошибка прогноза инфляции $= \pi_{t+1} - \pi_{t+1}$.

R_t — ex post реальная ставка процента за месяц

$$t = \frac{1/P_{t+1} - r_t/P_t}{r_t/P_t} = \frac{1 + R_t}{1 + \pi_{t+1}} - 1 \approx R_t - \pi_{t+1},$$

r_{t+1} — ex ante реальная ставка процента $t = \frac{1 - R_t}{1 + \pi_{t+1}} - 1 \approx$

$$\approx R_t - \pi_{t+1} = \frac{1 - R_t}{1 + \pi_{t+1}} - 1 \approx R_t - \pi_{t+1}.$$

В эконометрике рациональных ожиданий очень важна синхронизация — указание того, значение какой переменной наблюдается в какой момент времени. Наше правило для датированных переменных (которое отличает от Фамы) заключается в том, что переменная имеет индекс t , если ее значение является первым наблюдаемым на начало периода (месяца) t . Например, поскольку π_{t+1} , темп инфляции за месяц t зависит от P_{t+1} , его индексом является $t+1$, а не t . По той же причине r_{t+1} , ex post реальная процентная ставка за месяц t , имеет индекс $t+1$. В противоположность этому номинальная процентная ставка R_t за месяц t имеет индекс t , а не $t+1$, поскольку она определена ценой казначейского векселя на начало месяца t , r_t . Рис. 2.2 показывает, значение какой переменной наблюдается в какой момент времени.

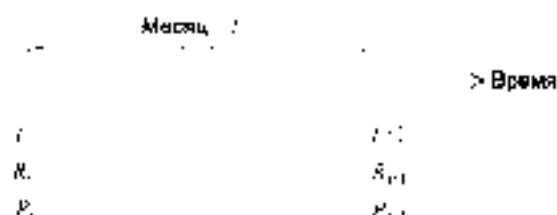


Рис. 2.2. Что когда наблюдается

Мы берем в качестве периода месяц только потому, что наши данные ежемесечные. Заметим, что срок погашения ценной бумаги (один месяц) совпадает с интервалом выборки (месяц). Если интервал выборки меньше, чем срок погашения, что имеет место, например, когда у нас

В [Галпа, 1975] наша π_{t+1} есть $\bar{\Delta}_t$, r_{t+1} есть $\bar{r}_{t+1}$ и R_t есть $\bar{R}_t$.

есть данные по ставке трехмесячных казначейских векселей, то тогда сроки погашения перекрываются и нам нужна более сложная техника, приводимая в главе 6.

Гипотеза эффективности рынка есть совместная гипотеза, объединяющая:

Рациональные ожидания. Инфляционные ожидания рациональны: $\pi_{t+1} = E(\pi_{t+1}|I_t)$, где I_t — информация, доступная на начало месяца t и включающая $\{R_t, R_{t-1}, \dots, \pi_t, \pi_{t-1}, \dots\}$. К тому же $I_t \supset I_{t-1} \supset \dots \supset I_{t-2}, \dots$. То есть агенты, участвующие в рынке, не забывают прошлое.

Постоянные реальные ставки. Ех ante реальная процентная ставка постоянна: $r_{t+1} = r$.

Тестируемые следствия

Мы получим два тестируемых следствия гипотезы эффективности рынка, используя следующие ключевые наблюдения относительно ошибки прогноза инфляции при рациональных ожиданиях:

- (a) $E(\eta_{t+1}|I_t) = 0$, то есть ошибка прогноза инфляции является последовательностью мартингал-разностей относительно информационного множества.
- (b) $\{\eta_t, \eta_{t-1}, \dots\}$, так же как и $\{R_t, R_{t-1}, \dots, \pi_t, \pi_{t-1}, \dots\}$, включены в I_t (известны на начало месяца t). То есть агенты помнят все свои прошлые ошибки.
- (a) Имеет нулевую ий смысл; если для прогнозирования темпа инфляции люди используют всю доступную информацию, ошибка прогноза, которая известна только постфактум, не имеет какой-либо систематической связи с тем, что люди знали, когда они формировали ожидания. Это можно легко доказать:

$$\begin{aligned} E(\eta_{t+1}|I_t) &= \\ &= E(\pi_{t+1} - E(\pi_{t+1}|I_t)) \quad (\text{поскольку } \eta_{t+1} \equiv \pi_{t+1} - E(\pi_{t+1}|I_t)) = \\ &= E(\pi_{t+1}|I_t) - E(E(\pi_{t+1}|I_t)|I_t) = \\ &= E(\pi_{t+1}|I_t) - E(E(\pi_{t+1}|I_t)|I_t) \\ &\quad (\text{поскольку } E(\pi_{t+1}|I_t) = E(E(\pi_{t+1}|I_t)|I_t) \text{ в силу рациональности ожиданий}) \\ &= E(\pi_{t+1}|I_t) - E(\pi_{t+1}|I_t) = 0. \end{aligned} \quad (2.11.1)$$

- (b) Вытекает из того, что агенты помнят прошлые прогнозы инфляции. Поскольку $\pi_{t-j-1} = E(\pi_{t-j}|I_{t-j-1})$ является функцией от I_{t-j-1} , она включена в I_{t-j} , (то есть известна на начало месяца $t-j-1$) и, следовательно, в I_t (известна на начало месяца t) для $j \geq 0$.

Таким образом, $\pi_{t-j} \equiv \lambda_{t-j} - \lambda_{t-j-1}$ включена в I_t для всех $j \geq 0$. Наблюдения (a) и (b) вместе с законом континуума математических означают, что

(c) $\{\pi_t\}$ является m.d.s.

Поэтому он является процессом с нулевым средним и сериально некоррелированным.

Это непротиворечиво, поскольку ожидания дефляционы, по-прежнему основанные на предположении о постоянной реальной ставке, они влезут для тестируемых следствия относительно темпа инфляции и процентной ставки, которые уже наблюдаемы.

Следствие 1: Ex post реальная процентная ставка имеет постоянное среднее и сериально некоррелирована.

Это следует из того, что

$$\begin{aligned} r_{t+1} &\equiv R_t - \pi_{t+1} = \\ &= (R_t + \pi_{t+1}) - (\pi_{t+1} + r_{t+1}) = \\ &= \pi_{t+1} - \eta_{t+1} \quad (\text{по определению}) \\ &= -r - \eta_{t+1} \quad (\text{так как } \pi_{t+1} = r \text{ по предположению}). \end{aligned} \quad (2.11.2)$$

В силу (c), $\{\pi_t\}$ имеет среднее π и сериально некоррелирована.

Следствие 2: $E(\pi_{t+1}|I_t) = r + R_t$.

То есть, наилучшая (в смысле минимальной среднеквадратичной ошибки) предикторка инфляции на основе всей доступной информации является номинальная процентная ставка: номинальная ставка R_t , которая определена ценой казначейского векселя π_t , объединяет всю текущую доступную информацию, имеющую отклонение к прогнозированию будущей инфляции. Это является формализацией того, что цены на активы «полностью отражают» доступную информацию. Указанное утверждение также может быть легко получено. Решая (2.11.2) относительно π_{t+1} , получаем: $\pi_{t+1} = -r + R_t + \eta_{t+1}$. Поэтому

$$\begin{aligned} E(\pi_{t+1}|I_t) &= E(-r + R_t + \eta_{t+1}|I_t) = \\ &= -r + R_t + E(\eta_{t+1}|I_t) \\ & \quad (\text{поскольку } r \text{ есть константа, а } R_t \text{ включена в } I_t) \\ &= -r + R_t \quad (\text{из (a)}). \end{aligned} \quad (2.11.3)$$

В оставшейся части этого параграфа мы тестируем эти два следствия эффективности рынка.

Тестирование на сериальную корреляцию

Рассмотрим сначала следствие, указывающее, что *ex post* реальная процентная ставка не имеет сериальной корреляции, которое Фама тестирует, используя результат из утверждения 2.9. Мы используем рассмотренный в [Mishkin, 1992] набор месячных данных по одномесечным казначейским векселям и месячному темпу инфляции ИПЦ, указанному в процентах годового темпа¹. Рис. 2.3 изображает данные. *Ex post* реальная процентная ставка, определенная как разность между ставкой по казначейским векселям и темпом инфляции ИПЦ, изображена на рис. 2.4.

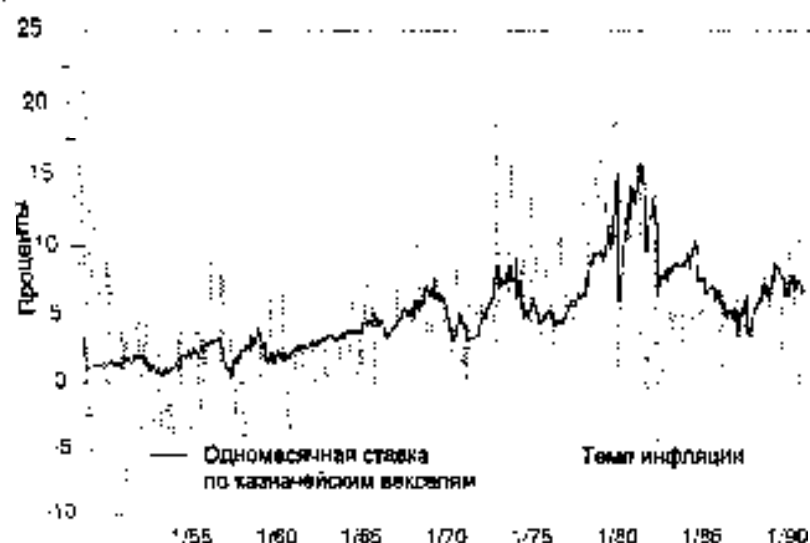


Рис. 2.3. Инфляция и процентная ставка

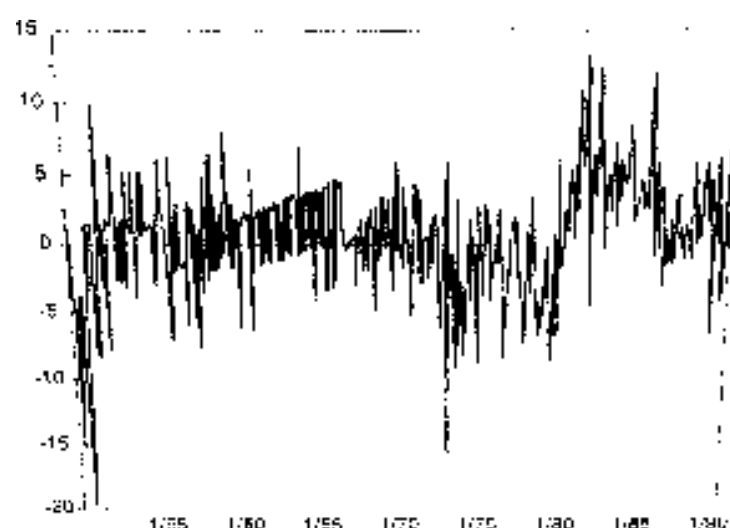


Рис. 2.4. Реальная процентная ставка

¹ Этот набор данных описан подробнее в главе пятом утверждения. Ссылка Фама, мера инфляции, использованная в тексте, вычислена на основе стран индекса CPI. Мера инфляции, рассмотренная в [Mishkin, 1992], использует композитный CPI, относящийся к стоимости домов. См. также (x) американского управления Б.

Для повторения результатов Фэма мы берем левую выборку тем же, что и в [Гаста, 1975], с января 1958 по июль 1971 г. Объем выборки, таким образом, равен 223. Свойства временного ряда реальной процентной ставки суммированы в табл. 2.1.

Таблица 2.1
Реальная процентная ставка, июль 1958 г. — июль 1971 г.

	среднее = 0,58%, стандартное отклонение = 2,387%, объем выборки = 223										
	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$	$j = 7$	$j = 8$	$j = 9$	$j = 10$	$j = 11$
$\bar{R}_j$	-0,101	0,172	0,019	-0,024	0,054	-0,021	-0,082	0,095	0,043	0,019	0,104
стан. ошибка	0,057	0,067	0,067	0,067	0,067	0,067	0,067	0,067	0,067	0,067	0,067
Q-критерий	2,0	9,3	2,1	10,1	10,1	10,2	12,1	14,2	16,3	16,4	26,5
р-значение, %	12,8	1,1	2,8	6,8	7,3	11,7	9,6	7,6	6,1	8,9	12,8

При вычислении выборочных автокорреляций $\hat{\rho}_j = \hat{\gamma}_j / \hat{\gamma}_0$ мы используем формулу (2.10.1) для $\hat{\gamma}_j$, где знаменатель равен объему выборки, а не $n - j$. Аналогично в качестве формулы для стандартной ошибки $\hat{\rho}_j$ берется $1/\sqrt{n}$, а не $1/\sqrt{n - j}$ (так что такая стандартная ошибка не зависит от j). Q-статистика Дьюнга — Боке в таблице для, скажем, $j = 2$ есть (2.10.5) с $p = 2$. Результаты в таблице показывают, что ни одна из автокорреляций не является сильно значимой индивидуально или как группа, в соответствии с первым следствием гипотезы эффективного рынка.

Является ли номинальная процентная ставка оптимальным предиктором?

Является ли номинальная процентная ставка оптимальным предиктором?

Мы можем тестировать следствие 2, оценивая соответствующую регрессию и проводя t -тест утверждения 2.3. Примечателен тот факт о гипотезе эффективности рынка, что, несмотря на свою простоту, она специфична, так что из нее вытекают важные части предположений, обосновывающих t -тест¹. Чтобы увидеть это, напомним $\eta_t = \pi_t - 1$, $x_t \equiv (1, R_t)'$, $\varepsilon_t \equiv \eta_{t+1}$ (ошибка прогноза инфляции) и перепишем (2.11.2) как

$$\pi_{t+1} = \beta_1 - \beta_2 R_t + \eta_{t+1} \quad \text{или} \quad \eta_t = x_t' \beta + \varepsilon_t, \quad (2.11.4)$$

так что предположение 2.1 очевидно удовлетворяется с

$$\beta = (-\tau, 1)'. \quad (2.11.5)$$

Другие предположения модели проверяются следующим образом:

• **Предположение 2.3 (предопределенные регрессоры)**

$g_t \equiv x_t \varepsilon_t$ в текущем контексте есть $(\eta_{t+1}, R_t \eta_{t+1})'$. Что нам нужно показать, так это то, что $E(\eta_{t+1}) = 0$ и $E(R_t \eta_{t+1}) = 0$. Первое вытекает из (а) и закона полных математических ожиданий:

$$E(\eta_{t+1}) = E[E(\eta_{t+1} | I_t)] = 0, \quad (2.11.6)$$

Второе выполняется потому, что

$$\begin{aligned} E(R_t \eta_{t+1}) &= \\ &= E[E(R_t \eta_{t+1} | I_t)] \end{aligned}$$

(по закону полных математических ожиданий)

¹Техническая часть — то, что матрица четвертых моментов $E(g_t g_t')$ существует и конечна и что регрессоры имеют конечные четвертые моменты, не предполагается гипотезой эффективности рынка и должно быть предположено.

$$= E[R_i | F_i(t_{i-1}, I_i)]$$

(из-за линейности условных математических ожиданий и $R_i \in I_i$)
 $= 0$ (по (а)). (2.11.7)

- **Предположение 2.5.** Из пункта (b) следует, что I_i включает в себя $\{e_i, (1 - \alpha)e_i, 2e_i, \dots\}$, так как $e_i \in (e_i, 2e_i) \cup \dots$. (a) тогда исходит (2.5.1), что является достаточным условием для того, чтобы $\{g_i\}$ были m.d.s.
- **Предположение 2.4.** Матрица $E(x_i x_i')$ здесь имеет вид:

$$E(x_i x_i') = \begin{bmatrix} 1 & E(R_i) \\ E(R_i) & E(R_i^2) \end{bmatrix}. \quad (2.11.8)$$

Ее определитель равен $E(R_i^2) - [E(R_i)]^2 = \text{Var}(R_i)$ и должен быть положительным, если бы дисперсия была нулевой, мы бы не наблюдали флуктуации процентной ставки i_t .

- **Предположение 2.3 (эргодическая стационарность).** Оно требует, чтобы $\{\pi_t, R_t\}$ был эргодически стационарным. График на рис. 2.3 показывает участок восходящего движения, особенно для номинальной процентной ставки, хотя за этим участком следует возращение на более низкий уровень. Тестирование стационарности ряда будет рассмотрено в примере 9.2 главы 9. В настоящее время, несмотря на это достаточно случайное свидетельство об обратном, мы исходим из предположения стационарности (но исследование наличия тренда в ряде будет затронуто в заключительном подраздразре). Что касается эргодичности, мы просто должны предположить ее.

Что мы показали, так это то, что при выполнении гипотезы эффективности рынка $\{y_t, x_t\}$ (где $y_t = \pi_{t+1}$ и $x_t = (1, R_t)'$) принадлежит множеству DGP, удовлетворяющих предположениям 2.1–2.5.

Проверим требуемые предположения, мы теперь перейдем к анализу. Используя наши данные, получая оцененную регрессию для $t = 1, \dots, 223$ ($t = 1$ соответствует январю 1953 г.)

$$\pi_{t+1} = -0.868 - 1.015 R_t \\ (0.431) (0.112)$$

$R^2 = 0.24$, среднее зависимой переменной = 2.35.

$$SER = 2.84\%, \quad n = 223. \quad (2.11.9)$$

Робастные и гетеросkedастичности стандартные ошибки, вычисленные как в (2.4.11) (где $\hat{S}$ задается (2.5.1) с $Z_t = e_t$), указаны в скобках. Как

показано в (2.11.5), эффективность рынка означает равенство 1 коэффициента при R_t^1 . Robustное t -приложение для нулевой гипотезы о том, что коэффициент при R_t^1 равен 1, равно $(1.015 - 1)/0.112 = 0.13$. Соответствующее ему p -значение равно примерно 90%. Таким образом, мы можем легко принять нулевую гипотезу.

Существуют и другие способы тестировать следствие о том, что номинальная процентная ставка объединяет всю текущую доступную информацию. Мы можем включить в регрессию больше объясняющих переменных и посмотреть, имеют ли они дополнительную объяснительную мощность сверх номинальной ставки. Однако дополнительные переменные, которые будут введены, должны быть частью текущего информационного множества I_t , поскольку очевидно, что гипотеза эффективности рынка согласуется с существованием некоторой переменной, в настоящее время не известной, предсказывающей будущую инфляцию лучше, чем текущая номинальная ставка. Формально, если z_t включает переключные, но исключенные в I_t , доказательства, которое мы использовали для обоснования предположения 2.3, больше недействительно.

R_t — множитель ставки заданной

Как было подчеркнуто, одно из важных преимуществ теории больших выборок из теории конечных выборок (*finite-sample theory*) из главы 1 заключается в том, что регрессоры обязательно должны быть строго экзогенными, пока они используются как источник ошибок. Мы сейчас покажем из примера, что R_t не является строгим экзогенным, так что теория конечных выборок неприменима к регрессии Фэми (2.11.4). Предположим, что ток и уровень инфляции π_t являются $AR(1)$ -процессом

$$\pi_t = c + \rho\pi_{t-1} + \eta_t, \quad \{\eta_t\} \text{ — независимый белый шум.}$$

Если π_{t+1} не зависит от любого элемента I_t , то

$$\begin{aligned} E(\pi_{t+1}|I_t) &= \\ &= c + \rho\pi_t + E(\eta_{t+1}|I_t) \quad (\text{поскольку } \eta_t \text{ содержится в } I_t) \\ &= c + \rho\pi_t \\ & \quad (\text{поскольку } E(\pi_{t+1}|I_t) = E(\pi_{t+1}) \text{ и } E(\pi_{t+1}) = 0 \text{ по предположению}). \end{aligned}$$

Поэтому π_{t+1} действительно является ошибкой прогноза инфляции для π_{t+1} , и

$$R_t = c + \rho\pi_t$$

(поскольку $R_t = c + E(\pi_{t+1}|I_t)$ при эффективности рынка). Тогда легко видеть, что константа ошибок ε_t в регрессии Фэми (π_{t+1} на константу

¹Напомним, что константа равна $-c$, но это не приводит к тестированию о равенстве нулю, поскольку π_t и ε_t — эффективные рынки не специфицируют значение c .

и R_t), которая есть ошибка прогноза инфляции π_{t+1} , может быть коррелирована с будущими регрессорами. Например,

$$\begin{aligned}\text{Cov}(u_t, R_{t+1}) &= \\&= \text{Cov}(u_t, \tau + \epsilon + \rho\pi_{t+1}) = \\&= \rho \text{Cov}(u_t, \pi_{t+1}) \\&= \rho \text{Var}(u_{t-1}) \quad (\text{поскольку } \pi_{t+1} = \epsilon + \rho\pi_t + u_{t+1} \text{ и } \text{Cov}(u_{t+1}, \pi_t) = 0),\end{aligned}$$

что не является нулем.

Хотя теория конечных выборок не применима, ее рецепты статистических выводов близки к правильным при условии, что ошибки условно homoskedастичны и размер выборки достаточно велик, о чем говорит параграф 2.6.

Последующее развитие

Хотя работа Фэми является, пожалуй, лучшим примером циклометрии (раскаленных ожиданий), события после ее публикации и большое количество эмпирических исследований, на которые она ссылалась, доказали, что близкое к взаимно однозначному соотношению между инфляцией и процентными ставками ограничивается определенным периодом до 1979 года. Такая тесная связь не может быть найдена для предвоенного периода или для многих других стран. Кроме того, увеличение реальных процентных ставок после скандала с операционной процедурой ФРС, которое состоялось в октябре 1979 года, противоречит предпосылке регрессии Фэми, что ожидаемая реальная процентная ставка не меняется с течением времени. Когда выборка ограничена периодом после октября 1970 года, коэффициент при процентной ставке в регрессии Фэми равен почти единице¹.

Эти результаты ставят вопрос о том, почему сильная связь происходит только в определенные периоды времени, а не в другие. Объяснение Митчелла [Mitchell, 1992] заключается в том, что институциональные правила включают в процентные ставки ожидания. В периоды, когда темп инфляции имеет относительно краткосрочные колебания, процентные ставки не реагируют на инфляцию. Однако устойчивое движение инфляции отражается на процентных ставках. Период, когда наблюдается сильная связь (которая включает в себя период выборки Фэми, см. рис. 2.3) является именно тем, на котором инфляция показывала устойчивое долгосрочное движение.

29 января 1997 года Министерство финансов США продало с аукциона 7 миллиардов долларов десятилетних облигаций с фиксированной процентной облигацией (T-10), как возможность исследователям выявить ее влия-

¹См., например, [Mitchell, 1992, Table 1]. Митчелл утверждает, что часть (1) долгосрочного уравнения (1).

реальную процентную ставку как доходность по индексированным облигациям. Данные по Великобритании, где индексированные облигации были доступны с начала 1980-х, показывают, что доходности по индексированным облигациям хотя и гораздо менее волатильны, чем доходности по обычным облигациям, но не являются постоянными во времени. Предположение о постоянной реальной ставке, которое мы сделали, было вполне разумным предположением, посредством которого мы можем проверить, отражают ли цены на облигации в полной мере доступную информацию. Теперь мы можем сделать это без вспомогательного предположения, регрессируя фактический темп инфляции на разность доходностей между обычными и индексированными облигациями. Мы также можем отбросить предположение, невозможное делавшееся до сих пор, о том, что разность доходностей равна ожидаемому темпу инфляции. Инвесторы могут не быть нейтральными к риску, и инфляционная премия за риск сверх ожидаемого уровня темпа инфляции может быть необходима для того, чтобы побудить их держать обычные облигации. Если это так, то ожидаемый темп инфляции снова становится ненаблюдаемым. Эконометрика рациональных ожиданий будет оставаться полезным инструментом для работы с ненаблюдаемыми ожиданиями.

Контрольные вопросы

1. Пусть $R_t = 5\%$ и $\pi_{t-1} = 2\%$. Вычислите r_{t+1} , используя две формулы: $(1 - R_t)/(1 + \pi_t) - 1$ и $R_t - \pi_t$. Могли ли возникнуть? Что будет, если $R_t = 20\%$ и $\pi_{t+1} = 17\%$?
2. Покажите, что при допущении совершенности рынка $\tilde{r}_t^e(1 + 1 - R_t) = r_t - R_t$. Указание: Используйте (2.9.7).
3. Если бы ошибки прогноза инфляции $\tilde{\pi}_t$ наблюдались, как бы вы тестировали эффективность рынка? Нужно ли нам предположение о постоянной реальной ставке?
4. Если темп инфляции и процентные ставки измерять в доллах, а не в процентах (то есть 0.08 вместо 8%), то как изменится регрессионный результат (2.11.9)? А если темп инфляции измерять в промиллах, но не в процентах, и процентную ставку в процентах за год? Указание: Если π — темп инфляции за месяц, а y — годовая темп, то $1 + y = (1 + \pi)^{12}$, так что $y \approx 12\pi$.
5. Предположим, что (2.11.4), что $\tilde{\pi}_t$ включает третью переменную, которая не входит в I_t . Каким образом аргументация, использованная при анализе предположения 2.1, терпит неудачу? Указание: Простой элемент $\tilde{\pi}_t$ является третьей переменной, умноженной на π_{t+1} .
6. Предложите пример неэффективности рынка такой, что следствие 1 выполняется, но следствие 2 — нет. Указание: Предположите, что π_t и R_t отрицательно коррелированы, и добавьте член к лагранжеву лагранжиану.

2.12. Регрессия на время

На протяжении всей этой главы мы предполагаем, что (y_t, x_t) является стационарным. Это предположение, однако, не всегда выполняется в регрессионных временных рядах. В этой главе мы изучаем два случая, в которых регрессия не является стационарной и все же применимы OLS. Первый рассматривается здесь, а то время как другой случай, называемый «интегрированными регрессиями», рассматривается в главе 13.

Регрессия, которую мы рассматриваем, записывается как

$$y_t = \alpha + \beta \cdot t + \varepsilon_t, \quad (2.12.1)$$

где $\{\varepsilon_t\}$ — нулевой белый шум. Эта регрессия может быть записана как $y_t = x_t' \beta + \varepsilon_t$,

$$x_t = (1, t)', \quad \beta = (\alpha, \beta)'. \quad (2.12.2)$$

Ясно, что x_t не является стационарным, поскольку математическое ожидание второго элемента равно t , и оно возрастает с течением времени. Аналогично нестационарным является y_t . Линейная функция $\alpha + \beta \cdot t$ называется **временным трендом** (*time trend*) y_t . Мы говорим, что процесс **стационарный относительно тренда** (*trend stationary*), если он может быть записан как сумма временного тренда и стационарного процесса. Процесс $\{\varepsilon_t\}$ здесь является частным случаем процесса, стационарного относительно тренда, стационарная компонента которого является тем самым белым шумом.

Асимптотическое распределение OLS-оценок

Пусть b является OLS-оценкой для β , основанной на выборке размера n :

$$b \equiv \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \left(\sum_{t=1}^n x_t x_t' \right)^{-1} \left(\sum_{t=1}^n x_t \cdot y_t \right). \quad (2.12.3)$$

Ошибка этой оценки может быть записана как

$$b - \beta = \begin{bmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \end{bmatrix} = \left(\sum_{t=1}^n x_t x_t' \right)^{-1} \left(\sum_{t=1}^n x_t \cdot \varepsilon_t \right). \quad (2.12.4)$$

Используя алгебраические результаты $\sum_{t=1}^n t = n \cdot (n+1)/2$ и $\sum_{t=1}^n t^2 = n \cdot (n+1)(2n+1)/6$, легко доказать, что

$$\sum_{t=1}^n x_t x_t' = \begin{bmatrix} n & n \cdot (n+1)/2 \\ n \cdot (n+1)/2 & n \cdot (n+1)(2n+1)/6 \end{bmatrix}. \quad (2.12.5)$$

Поэтому, в отличие от стационарного случая, $\sum_{t=1}^n x_t x_t' / n$ не сходится по вероятности к не вырожденной матрице, чья диагональная диагональ

Оказывается, что OLS-оценки $\hat{\alpha}$ и $\hat{\beta}$ являются состоятельными, но имеют различные скорости сходимости. Как и в стационарном случае, скорость сходимости для $\hat{\alpha}$ есть $\sqrt{n}$. В противоположность этому скорость для $\hat{\beta}$ есть $n^{3/2}$. То есть $n^{3/2}(\hat{\beta} - \beta)$ сходится к невырожденному распределению (распределению непостоянной случайной величины). Чтобы присвоить эти различные скорости сходимости элементам $b = \beta$, умножим обе части (2.12.4) на матрицу:

$$\Upsilon_n = \begin{bmatrix} \sqrt{n} & 0 \\ 0 & n^{3/2} \end{bmatrix}. \quad (2.12.5)$$

Это приводит к

$$\begin{aligned} \Upsilon_n(b - \beta) &= \left[\frac{\sqrt{n}(\hat{\alpha} - \alpha)}{n^{3/2}(\hat{\beta} - \beta)} \right] = \Upsilon_n \left(\sum_{t=1}^n x_t x_t' \right)^{-1} \left(\sum_{t=1}^n x_t \cdot \varepsilon_t \right) = \\ &= \Upsilon_n \left(\sum_{t=1}^n x_t x_t' \right)^{-1} \Upsilon_n^{-1} \left(\sum_{t=1}^n x_t \cdot \varepsilon_t \right) = \\ &= \left[\Upsilon_n^{-1} \left(\sum_{t=1}^n x_t x_t' \right) \Upsilon_n^{-1} \right]^{-1} \left(\Upsilon_n^{-1} \sum_{t=1}^n x_t \cdot \varepsilon_t \right) \\ &\equiv Q_n^{-1} v_n. \end{aligned} \quad (2.12.7)$$

где

$$Q_n \equiv \Upsilon_n^{-1} \left(\sum_{t=1}^n x_t x_t' \right) \Upsilon_n^{-1} \quad \text{и} \quad v_n \equiv \Upsilon_n^{-1} \sum_{t=1}^n x_t \cdot \varepsilon_t. \quad (2.12.8)$$

Подставляя (2.12.5) и (2.12.6) в эти выражения для Q_n и v_n , получаем:

$$\begin{aligned} Q_n &= \begin{bmatrix} 1 & (n+1)/(2n) \\ (n+1)/(2n) & (n+1)(2n+1)/(6n^2) \end{bmatrix}, \\ v_n &= \begin{bmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^n \varepsilon_t \\ \frac{1}{n^{3/2}} \sum_{t=1}^n t \varepsilon_t \end{bmatrix}. \end{aligned} \quad (2.12.9)$$

Ясно, что мы имеем

$$Q_n \rightarrow Q \equiv \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}. \quad (2.12.10)$$

Для v_n можно показать (см., например: [Hampel, 1994, pp. 458-460]), что

$$v_n \xrightarrow{d} N(0, \sigma^2 Q). \quad (2.12.11)$$

Если $\{\varepsilon_t\}$ является независимым белым шумом с $E(\varepsilon_t^2) = \sigma^2$ и $E(\varepsilon_t^4) < \infty$. Таким образом, асимптотическое распределение (2.12.7) является нормальным со средним 0 и дисперсией $Q^{-1}(\sigma^2 Q) = Q^{-1} \cdot \sigma^2 Q^{-1}$. Продолжив предыдущее обсуждение, получаем

Утверждение 2.11 (OLS-оценивание регрессии на время): Рассмотрим регрессию на время (2.12.1), где ε_t — независимый белый шум с $E(\varepsilon_t^2) = \sigma^2$ и $E(\varepsilon_t^4) < \infty$, и пусть $\hat{\alpha}$ и $\hat{\delta}$ являются OLS-оценками для α и δ . Тогда

$$\left[\sqrt{n}(\hat{\alpha} - \alpha) \right] \xrightarrow{d} N\left(0, \sigma^2 \cdot \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}^{-1}\right).$$

Как и в стационарном случае, $\hat{\alpha}$ является $\sqrt{n}$ -состоятельной, поскольку $\sqrt{n}(\hat{\alpha} - \alpha)$ сходится к (нормальной) случайной величине. OLS-оценка коэффициента при времени, $\hat{\delta}$, также является состоятельной, но скорость сходимости выше: она $n^{3/2}$ -состоятельна в том смысле, что $n^{3/2}(\hat{\delta} - \delta)$ сходится к случайной величине. В этом смысле $\hat{\delta}$ является **гиперсостоятельной** (*hyperconsistent*)¹.

Тестирование гипотез в регрессиях на время

Таким образом, OLS-коэффициенты в регрессии на время асимптотически нормальны при условии, что ошибка оценки правильно анкалирована. Мы сейчас покажем, что деление ошибки оценки на стандартную ошибку приводит к шкалированию, которое делает результирующее отношение — t -значение — асимптотически нормальным.

Сначала рассмотрим t -значение для нулевой гипотезы $\alpha = \alpha_0$. Замечая, что

$$\text{элемент } (1, 1) \text{ матрицы } \left(\sum_{i=1}^n x_i x_i' \right)^{-1} = [1 \quad 0] \left(\sum_{i=1}^n x_i x_i' \right)^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad (2.12.12)$$

t -значение можно записать как

$$\begin{aligned} t &= \frac{\hat{\alpha} - \alpha_0}{\sqrt{s^2 \cdot [1 \quad 0] \left(\sum_{i=1}^n x_i x_i' \right)^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}}} = \\ &= \frac{\sqrt{n}(\hat{\alpha} - \alpha_0)}{\sqrt{s^2 \cdot [1 \quad 0] \left(\sum_{i=1}^n x_i x_i' \right)^{-1} \begin{bmatrix} \sqrt{n} \\ 0 \end{bmatrix}}} = \\ &= \frac{\sqrt{n}(\hat{\alpha} - \alpha_0)}{\sqrt{s^2 \cdot [1 \quad 0] \mathbf{Y}_n \left(\sum_{i=1}^n x_i x_i' \right)^{-1} \mathbf{Y}_n \begin{bmatrix} 1 \\ 0 \end{bmatrix}}} \\ &\quad (\text{поскольку } [\sqrt{n} \quad 0] = [1 \quad 0] \mathbf{Y}_n) \end{aligned}$$

¹ Оценка, которая n -состоятельна, называется **суперсостоятельной** (*superconsistent*). Известный пример суперсостоятельных оценок возникает при оценивании процессов с единичными корнями: см. главу 9. Некоторые авторы используют термин «суперсостоятельность» более широко, для обозначения оценок, которые n^{γ} -состоятельны, с $\gamma > \frac{1}{2}$. На их языке $\hat{\delta}$ в тексте является суперсостоятельной.

$$\begin{aligned}
 &= \frac{\sqrt{n}(\hat{\delta} - \alpha_0)}{\sqrt{\kappa^2 \cdot [1 - \theta] \{ \Upsilon_n' \cdot (\sum_{i=1}^n x_i x_i' \Upsilon_n^{-1})^{-1} [1] \}}} \\
 &= \frac{\sqrt{n}(\hat{\delta} - \alpha_0)}{\sqrt{\kappa^2 \cdot [1 - \theta] Q_n^{-1} [1]}} \quad (\text{из (2.12.8)}), \quad (2.12.13)
 \end{aligned}$$

Можно непосредственно доказать (см. контрольный вопрос 1), что κ^2 постоянна для σ^2 . И из (2.12.10) $Q_n \rightarrow_p Q$. Таким образом, по лемме 2.4(с) и утверждению 2.11,

t -значение для нулевой гипотезы $\alpha = \alpha_0$

$$t = \frac{\text{1-й элемент } z}{\sqrt{\sigma^2 \cdot \text{элемент (1,1) матрицы } Q^{-1}}}, \quad (2.12.14)$$

где $z \sim N(0, \sigma^2 Q^{-1})$. Поэтому t -значение для α асимптотически $N(0, 1)$, как и в стационарном случае.

Используя тот же самый трюк и замечая, что $[0 \quad \mu^{N/2}] = [0 \quad 1] \Upsilon_n$, мы можем записать t -значение для нулевой гипотезы $\delta = \delta_0$ как

$$t = \frac{\sqrt{n/2}(\hat{\delta} - \delta_0)}{\sqrt{\kappa^2 \cdot [1 \quad 1] Q_n^{-1} [1]}}. \quad (2.12.15)$$

Поэтому t -значение для коэффициента при времени также является асимптотически $N(0, 1)$. Статистические выводы относительно α или δ могут быть получены аналогично стационарному случаю.

Контрольные вопросы

1. κ^2 постоянна для σ^2 . Покажите, что $\kappa^2 \rightarrow_p \sigma^2$. Указание: Исходя из формулы (2.3.10) получите

$$\epsilon_t^2 = z_t^2 - 2(\hat{b} - \beta)' \Upsilon_n \Upsilon_n^{-1} z_t + z_t' (\hat{b} - \beta)' \Upsilon_n \Upsilon_n^{-1} z_t / \Upsilon_n^{-1} \Upsilon_n (\hat{b} - \beta).$$

Суммируйте по t , чтобы получить

$$\frac{1}{n} \sum_{t=1}^n \epsilon_t^2 = \frac{1}{n} \sum_{t=1}^n z_t^2 - \frac{2}{n} (\hat{b} - \beta)' \Upsilon_n z_n + \frac{1}{n} (\hat{b} - \beta)' \Upsilon_n Q \Upsilon_n (\hat{b} - \beta).$$

2. (Серия как некоррелированные ошибки.) Предположим, что $\{z_t\}$ — белый стационарный процесс, а ϵ_t — процесс независимого белого шума, но предположим, что $z_n \rightarrow_p N(0, V)$, где $V \neq \sigma^2 Q$. Являются ли OLS оценки $\hat{\alpha}$ и $\hat{\delta}$ состоятельными? Сходятся ли $\Upsilon(\hat{b} - \beta)$ к нормальному случайному вектору с нулевым средним? Если да, то какова дисперсия возможного распределения? [Ответ: $Q^{-1} V Q^{-1}$.] Являются ли t -значения асимптотически нормальными? [Ответ: Да. Но дисперсия не будет единичной.]

Приложение 2.А. Асимптотика с фиксированными регрессорами

Чтобы доказать результаты, изложенные в выводах утверждения 2.5, мы добавим к предположениям 2.1 следующие.

Предположение 2.А.1: X является детерминированной матрицей. При $n \rightarrow \infty$ $S_{xx} \rightarrow \Sigma_{xx}$, невырожденной матрице (здесь обычная сходимость, поскольку X не является случайной).

Предположение 2.А.2: (ε_i) является i.i.d. с $E(\varepsilon_i) = 0$ и $\text{Var}(\varepsilon_i) = \sigma^2$.

Утверждение 2.А.1: При предположениях 2.1, 2.А.1 и 2.А.2 левые предположения на $\{y_i\}$, называемое *условием Линдберга* (где $g_i \equiv y_i - x_i'$), сохраняются те же самые выводы, что и в утверждении 2.5.

Вместо того чтобы сформулировать условие Линдберга, мы объясним, почему такое условие предпочтительнее. Для доказательства асимптотической нормальности b мы могли бы взять (2.3.8) и показать, что $\sqrt{n}b$ сходится к нормальному распределению. Техническая сложность заключается в том, что последовательность $\{g_i\}$ не является стационарной или независимой, так что $\{g_i\}$ является последовательностью независимых векторов. Например,

$$\text{Var}(g_i) = E(x_i' x_i \varepsilon_i^2) = E(x_i' x_i) E(\varepsilon_i^2) = \sigma^2 x_i' x_i$$

не остается постоянным при переходе от наблюдений к наблюдению. Следовательно, не CLT Линдберга. Тем не менее CLT для мультипараметрических применимы. Однако, по сути, существует обобщение CLT Линдберга — Лева — к стационарным процессам с асимптотической дисперсией, так называемая **CLT Линдберга — Феллера** (Lindeberg — Feiler CLT), которая устанавливает техническое ограничение на хвост распределения g_i , чтобы наблюдения с большими $|g_i|$ не могли доминировать в распределении $\sqrt{n}g$.

Приложение 2.В. Доказательство утверждения 2.10

В этом приложении приводится доказательство утверждения 2.10. Мы начинаем с выражения для $\hat{\beta}_n$, данного в (2.10.12). Первая часть доказательства заключается в нахождении выражения, которое является асимптотически эквивалентным.

$$\sqrt{n}(\hat{\beta}_1 - \sqrt{n}\beta_1) = \frac{1}{n} \sum_{i=1}^n (x_{i1} - x_{i1} + x_{i1} - x_{i1}) \sqrt{n}(\hat{\beta}_1 - \beta_1) =$$

$$= \sqrt{n}(\hat{\beta}_1 - \beta_1) \left(\frac{1}{n} \sum_{i=1}^n (x_{i1} - x_{i1}) \right)$$

$$\stackrel{a.s.}{\rightarrow} \sqrt{n}(\hat{\beta}_1 - \beta_1) \cdot \frac{1}{n} \sum_{i=1}^n (x_{i1} - x_{i1} + x_{i1} - x_{i1}) \sqrt{n}(\hat{\beta}_1 - \beta_1)$$

(по закону Slutsky законно выше не писать)

$$\stackrel{a.s.}{\rightarrow} \sqrt{n}(\hat{\beta}_1 - \beta_1) \cdot \frac{1}{n} \sum_{i=1}^n (x_{i1} - x_{i1} + x_{i1} - x_{i1}) \sqrt{n}(\hat{\beta}_1 - \beta_1)$$

$$(\text{по закону } \frac{1}{n} \sum_{i=1}^n (x_{i1} - x_{i1} + x_{i1} - x_{i1}) \rightarrow \frac{1}{n} \sum_{i=1}^n (x_{i1} - x_{i1} + x_{i1} - x_{i1}))$$

$$= \sqrt{n}(\hat{\beta}_1 - \beta_1) \sqrt{n}(\hat{\beta}_1 - \beta_1) =$$

т.е. $\hat{\beta}_1 = \beta_1 + o_p(1)$ (по закону $E(x_{i1} - x_{i1}) = 0$ по (2.3), (2.4))

$$= \sqrt{n}(\hat{\beta}_1 - \beta_1) \sqrt{n}(\hat{\beta}_1 - \beta_1) \frac{1}{n} \sum_{i=1}^n x_{i1} - x_{i1} =$$

$$= \sqrt{n}(\hat{\beta}_1 - \beta_1) \frac{1}{n} \sum_{i=1}^n x_{i1} - x_{i1} - \mu_1 \sqrt{n}(\hat{\beta}_1 - \beta_1) \frac{1}{n} \sum_{i=1}^n x_{i1} - x_{i1} \text{ по (2.4), (2.7)}$$

$$\stackrel{a.s.}{\rightarrow} \sqrt{n}(\hat{\beta}_1 - \beta_1) \frac{1}{n} \sum_{i=1}^n x_{i1} - x_{i1} - \mu_1 \sqrt{n}(\hat{\beta}_1 - \beta_1) \frac{1}{n} \sum_{i=1}^n x_{i1} - x_{i1}$$

$$(\text{по закону } \sqrt{n}(\hat{\beta}_1 - \beta_1) \frac{1}{n} \sum_{i=1}^n x_{i1} - x_{i1} = \sqrt{n}(\hat{\beta}_1 - \beta_1) \frac{1}{n} \sum_{i=1}^n x_{i1} - x_{i1})$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n x_{i1} - x_{i1} = o_p(1) \text{ по закону (2.3)}$$

$$\stackrel{a.s.}{\rightarrow} \sqrt{n}(\hat{\beta}_1 - \beta_1) \frac{1}{n} \sum_{i=1}^n x_{i1} - x_{i1} - \mu_1 \sqrt{n}(\hat{\beta}_1 - \beta_1) \frac{1}{n} \sum_{i=1}^n x_{i1} - x_{i1}$$

(по закону $\sqrt{n}(\hat{\beta}_1 - \beta_1) \rightarrow \sqrt{n}(\hat{\beta}_1 - \beta_1)$)

$$= o_p(1) \sqrt{n}(\hat{\beta}_1 - \beta_1) = o_p(1)$$

$$(\hat{\beta}_1 - \beta_1) = \left[\frac{1}{\sum_{i=1}^n x_{i1}^2} \sum_{i=1}^n x_{i1} y_{i1} - \beta_1 \right]$$

$$= \frac{1}{n} \sum_{i=1}^n x_{i1} y_{i1} - \beta_1 = \left| \frac{\sum_{i=1}^n x_{i1} y_{i1}}{n} - \beta_1 \right| \quad (2.41)$$

Таким образом, полагая

$$\underset{(p \times 1)}{\gamma} \equiv \begin{bmatrix} \tilde{\gamma}_1 \\ \vdots \\ \tilde{\gamma}_p \end{bmatrix},$$

мы доказали, что

$$\sqrt{n}\tilde{\gamma} \underset{n}{\sim} C' \sqrt{n}\tilde{g},$$

где

$$\underset{(p(K+1) \times p)}{C} \equiv \begin{bmatrix} c_1 & & \\ & \ddots & \\ & & c_p \\ & & & c_p \end{bmatrix},$$

$$\underset{(p(K+1) \times 1)}{\tilde{g}} \equiv \frac{1}{n} \sum_{t=1}^n g_t,$$

$$\underset{(p(K+1) \times 1)}{g_t} \equiv \begin{bmatrix} g_{1t} \\ \vdots \\ g_{pt} \end{bmatrix}.$$

Вторая часть доказательства состоит в том, чтобы показать, что g_t является мартингал-разностью, а именно что

$$E(g_{jt}|g_{t-1}, g_{t-2}, \dots) = 0 \quad (\text{для } j = 1, 2, \dots, p).$$

Но это легко следует из закона повторных математических ожиданий, того факта, что $(g_{t-1}, g_{t-2}, \dots)$ содержит меньше информации, чем $(x_t, x_{t-1}, \dots, x_{t-1}, x_{t-2}, \dots)$. (2.10.15).

Ясно, что g_t эргодически стационарна. Следовательно, по CLT для эргодически стационарных мартингал-разностей, $\sqrt{n}\tilde{g} \rightarrow_d N(0, E(g_t g_t'))$. Следующий шаг состоит в вычислении математического ожидания $E(\mu_t g_t')$, (j, k) -й блок которого имеет вид:

$$\underset{(K+1) \times (K+1)}{E(g_{jt} g_{kt}')} = \begin{bmatrix} E(x_t^j x_{t-j}^k) & E(x_t^j x_{t-1}^k) \\ E(x_t^j x_{t-1}^k) & E(x_t^j x_t^k) \end{bmatrix}.$$

Используя закон полных математических ожиданий² и (2.10.16), легко показать, что

$$\underset{(K+1) \times (K+1)}{E(g_{jt} g_{kt}')} = \begin{bmatrix} \sigma^2 \delta_{jk} & \sigma^2 \mu_j' \\ \sigma^2 \mu_k & \sigma^2 \Sigma_{xx} \end{bmatrix}, \quad (2.B.2)$$

где δ_{jk} — «дельта-символ Кронекера», принимающий значение 1, если $j = k$, и 0 в противном случае.

Поскольку, как показано выше,

$$\sqrt{n}\tilde{\gamma} \xrightarrow{d} C'\sqrt{n}\tilde{g} \text{ и } \sqrt{n}\tilde{g} \xrightarrow{d} N(0, E(g_1 g_1')),$$

мы имеем:

$$\text{Avar}(\tilde{\gamma}) \equiv \text{дисперсия предельного распределения } \sqrt{n}\tilde{\gamma}) = C' E(g_1 g_1') C.$$

Ее (j, k) -й элемент есть

$$(j, k)\text{-й элемент } \text{Avar}(\tilde{\gamma}) = c_j' E(g_{1j} g_{1k}') c_k.$$

Подставляя сюда (2.В.2) и используя определение c_j в (2.В.1), после некоторых простых вычислений получаем, что

$$(j, k)\text{-й элемент } \text{Avar}(\tilde{\gamma}) = \sigma^4 \cdot [\delta_{jk} - \mu_j' \Sigma_{xx}^{-1} \mu_k / \sigma^2].$$

Поэтому

$$\sqrt{n}\tilde{\gamma} \xrightarrow{d} N(0, \sigma^4 I_n - \sigma^4 \Phi),$$

где Φ определена и утверждение 2.10. Поскольку $\sqrt{n}\tilde{g} \xrightarrow{d} n\tilde{\gamma}/\sigma^4$, предельное распределение для $\sqrt{n}\tilde{g}$ то же самое, что и предельное распределение для $\sqrt{n}\tilde{\gamma}/\sigma^2$, которое является $N(0, I_p - \Phi)$. ■

Набор задач для главы 2

Аналитические утверждения

1. (Взято из примера 3.4.2 в [Домгуги, 1985].) Пусть ε_n определяется как

$$\varepsilon_n = \begin{cases} 0 & \text{с вероятностью } (n-1)/n, \\ n^2 & \text{с вероятностью } 1/n. \end{cases}$$

Покажите, что $\lim_{n \rightarrow \infty} \varepsilon_n = 0$, но $\lim_{n \rightarrow \infty} E(\varepsilon_n) = \infty$.

2. Докажите слабый LLN Чебышева, показывая, что $\bar{\varepsilon}_n \rightarrow_{\text{пл}} \mu$. **Указание:**

$$\bar{\varepsilon}_n - \mu = (\bar{\varepsilon}_n - E(\bar{\varepsilon}_n)) + (E(\bar{\varepsilon}_n) - \mu).$$

Поэтому

$$(\bar{\varepsilon}_n - \mu)^2 = (\bar{\varepsilon}_n - E(\bar{\varepsilon}_n))^2 + 2(\bar{\varepsilon}_n - E(\bar{\varepsilon}_n))(E(\bar{\varepsilon}_n) - \mu) + (E(\bar{\varepsilon}_n) - \mu)^2.$$

Покажите, что $E[(\bar{\varepsilon}_n - E(\bar{\varepsilon}_n))(E(\bar{\varepsilon}_n) - \mu)] = 0$.

3. (Состоятельность и асимптотическая нормальность OLS для случайных выборок.) Рассмотрите задачу предположения 2.2 на

Предположение 2.2': $\{y_i, x_i\}$ является случайной выборкой,

и предположения 2.5 на

Предположение 2.6': $S = E(g_i g_i')$ существует и конечно.

Докажите следующую упрощенную версию утверждения 2.1:

Утверждение 2.1 для случайных выборок: При предположениях 2.1, 2.3, 2.4 и 2.2' OLS-оценка $\hat{b}$ является состоятельной. При дополнительном предположении 2.6' OLS-оценка $\hat{b}$ является асимптотически нормальной с $\text{Asy}(\hat{b})$, заданной в (1).

(а) Покажите, что это является частным случаем утверждения 2.1 в тексте.

(б) Дайте прямое доказательство этого утверждения.

Указание: Используйте результаты Колмогорова и Лиласберга-Лева.

4. (Доказательство утверждения 2.4.) Мы хотим доказать утверждение 2.4. Чтобы избежать ненужных усложнений, мы предположим, как и в тексте, что $K = 1$ (только один регрессор). Что остается показать — это то, что выборочные средние первой компоненты правой части (2.5.3) сходятся по вероятности к конечному числу. Докажите это. **Указание:** Пусть $f \equiv x_1$, а $h \equiv x_1^2$. **Неравенство Коши — Шварца (Cauchy — Schwarz inequality)** утверждает:

$$E[(f \cdot h)] \leq \sqrt{E(f^2)E(h^2)}.$$

Поэтому

$$E(x_1^3) \leq \sqrt{E(x_1^2)E(x_1^4)}.$$

По предположению, $E(x_1^2)$ конечно и, по предположению 2.6, $E(x_1^4)$ также конечно. Поэтому $E(x_1^3)$ ($= E(y_1^3)$) конечно, и $\frac{1}{n} \sum_{i=1}^n x_{1i}^3 \rightarrow 0$ с вероятностью 1 к конечному числу в силу эргодической стационарности.

5. (Прямое доказательство того, что аддитивное SSR_n деленное на s^2 имеет асимптотическое расщепление χ^2 .) В параграфе 2.5 мы доказали, что

$$(SSR_n - SSR_0)/s^2 \rightarrow \chi^2(\#r)$$

как следствие утверждения 2.3. Дайте прямое доказательство, сначала показывая, что

$$SSR_n - SSR_0 = (\sqrt{ng})' S_{xx}^{-1} R' (R S_{xx}^{-1} R')^{-1} R S_{xx}^{-1} \sqrt{ng},$$

$$\bar{y} \equiv \frac{1}{n} \sum_{i=1}^n y_i = \varepsilon_i.$$

6. (Дополнительно, состоятельность $\hat{\alpha}$.) В этом упражнении мы докажем утверждение, сделанное в параграфе 2.8, о том, что $\hat{\alpha}$, OLS-оценка для регрессии ε_i^2 на z_i , состоятельна. Пусть $\hat{\alpha}$ будет OLS-оценкой из

$$\varepsilon_i^2 = z_i' \alpha + \eta_i, \quad \eta_i = \varepsilon_i^2 - E(\varepsilon_i^2 | x_i).$$

- (а) Сделайте соответствующие предположения и дополнение к предположениям 2.1 и 2.2, чтобы показать, что $\hat{\alpha}$ является состоятельной. **Указание:** Вам нужно ранговое условие для z_i .

- (б) Получите:

$$\hat{\alpha} - \alpha = \left(\frac{1}{n} \sum_{i=1}^n z_i z_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n z_i \cdot \eta_i \quad (*)$$

с $\eta_i = -2(b - \beta)' x_i \cdot \varepsilon_i + (b - \beta)' x_i x_i' (b - \beta)$. **Указание:** Расхождение между ε_i^2 и квадратом OLS-остатка e_i^2 из OLS-оценивания исходного уравнения $y_i = x_i' \beta + \varepsilon_i$ задано в (2.3.10). Подстановка в (2.8.8) дает: $\varepsilon_i^2 = z_i' \alpha + (\eta_i - e_i)$

- (с) Чтобы избежать несущественных усложнений, предположим, что x_i является скаляром x_i . Тогда $\frac{1}{n} \sum_{i=1}^n z_i \cdot \eta_i$ в (*) принимает вид:

$$2(b - \beta) \frac{1}{n} \sum_{i=1}^n x_i \varepsilon_i z_i + (b - \beta)' \frac{1}{n} \sum_{i=1}^n x_i^2 z_i. \quad (**)$$

Покажите, что plim первой компоненты равен нулю. **Указание:**

$$E(x_i \varepsilon_i z_i) = E(x_i \cdot z_i \cdot E(\varepsilon_i | x_i)).$$

Покажите, что plim второй компоненты равен нулю, если $E(x_i^2 z_i)$ существует и конечно.

- (д) Таким образом, мы доказали, что $\hat{\alpha} - \alpha$ стремится к нулю. Докажите более сильный результат о том, что $\sqrt{n}(\hat{\alpha} - \alpha)$ стремится к нулю. **Указание:** Используйте лемму 2.4(й).

7. (Дополнительно, докажите, что WLS асимптотически эффективнее, чем OLS.) В параграфе 2.8 WLS-оценка обозначена как $\hat{\beta}(\hat{V})$. Мы хотим доказать, что она асимптотически эффективнее OLS, а именно

$$\text{Avar}(\hat{\beta}(\hat{V})) \leq \text{Avar}(b).$$

Но, поскольку $\text{Avar}(\hat{\beta}(\hat{V})) = \text{Avar}(\hat{\beta}(V))$, достаточно доказать, что

$$\text{Avar}(\hat{\beta}(V)) \leq \text{Avar}(b).$$

Докажите последнее матричное неравенство. Указание: В главе 1 мы доказали алгебраический результат о том, что

$$(X'X)^{-1}(X'VX)(X'X)^{-1} \geq (X'V^{-1}X)^{-1} \quad (*)$$

для любой положительной определенной матрицы V . (Это следует из того факта, что GLS эффективнее OLS, т.е. дисперсия (умноженная на σ^2) дисперсия OLS-оценки для обобщенной регрессионной модели, а то же самое как дисперсия (умноженная на σ^2) дисперсия GLS-оценки.) Умножение обеих сторон (*) на n и взятие предела по вероятности приводят к:

$$\left(\lim_{n \rightarrow \infty} \frac{1}{n} X'X\right)^{-1} \left(\lim_{n \rightarrow \infty} \frac{1}{n} X'VX\right) \left(\lim_{n \rightarrow \infty} \frac{1}{n} X'X\right)^{-1} \geq \\ \geq \lim_{n \rightarrow \infty} \left(\frac{1}{n} X'VX\right)^{-1}. \quad (**)$$

Правая часть есть $\text{Avar}(\hat{\beta}(V))$ (см. (2.8.7)). Покажите, что левая часть равна $\text{Avar}(\hat{\beta}) = \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1}$.

8. Докажите (2.9.7). Указание: Если (μ, γ) является вектором коэффициентов регрессии наклонных квадратов, то

$$E(xx')^{-1} \begin{bmatrix} \mu \\ \gamma \end{bmatrix} = E(x \cdot y), \quad (**)$$

или

$$\begin{bmatrix} \mu \\ \gamma \end{bmatrix} = [E(xx')]^{-1} E(x \cdot y), \quad (***)$$

где

$$E(xx') = \begin{bmatrix} 1 & E(\tilde{x})' \\ E(\tilde{x}) & E(\tilde{x}\tilde{x}') \end{bmatrix}, \quad E(x \cdot y) = \begin{bmatrix} E(y) \\ E(\tilde{x} \cdot y) \end{bmatrix}$$

Первый слагаемый в (*) является $\mu + E(\tilde{x})'\gamma = E(y)$. Используйте его для удаления μ из остальных уравнений (**), а затем найдите решение относительно γ . При этом используйте $\text{Var}(\tilde{x}) = E(\tilde{x}\tilde{x}') - E(\tilde{x})E(\tilde{x})'$, $\text{Cov}(\tilde{x}, y) = E(\tilde{x} \cdot y) - E(\tilde{x})E(y)$. Аналогичным образом используйте формулу для обращения блочных матриц (см. (A.10) приложение A), чтобы получить

$$E(xx')^{-1} = \begin{bmatrix} -E(\tilde{x})' \text{Var}(\tilde{x})^{-1} E(\tilde{x}) & -E(\tilde{x})' \text{Var}(\tilde{x})^{-1} \\ -\text{Var}(\tilde{x})^{-1} E(\tilde{x}) & \text{Var}(\tilde{x})^{-1} \end{bmatrix}.$$

Затем подставьте это в (**).

9. (Доказательство утверждения 2.9 для $p = 1$.) Предположим, что $\{\varepsilon_t\}$ — эргодически стационарная последовательность мартигал-разностей и что $E(\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_1) = \sigma^2 < \infty$. Пусть $y_t = \varepsilon_t + \varepsilon_{t-1}$ и

$$\bar{\varepsilon}_j = \frac{1}{n} \sum_{t=j+1}^n \varepsilon_t + \varepsilon_{t-j}, \quad (j = 0, 1), \quad \bar{\rho}_1 = \frac{\bar{\varepsilon}_0}{\bar{\varepsilon}_1}.$$

- (a) Покажите, что $\{\rho_t\}$ ($t = 2, 3, \dots$) является последовательностью мартигал-разностей.
 (b) Покажите, что $1/\bar{\rho}_1^2 = \sigma^{-4}$.
 (c) Покажите, что $\sqrt{n}\bar{\rho}_1 \rightarrow_d N(0, \sigma^4)$ при $n \rightarrow \infty$.
 (d) Покажите, что $\sqrt{n}\bar{\rho}_1 \rightarrow_d N(0, 1)$. **Указание:** Сначала покажите, что $\sqrt{n}\bar{\varepsilon}_1/\sigma^2 \rightarrow_d N(0, 1)$. Затем используйте лемму 2.4(c).
 10. (Асимптотика выборочного среднего для МА(2).) Пусть $\{\varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \dots\}$ является Л.Б. с нулевым средним и дисперсией σ_ε^2 . Рассмотрим процесс $\{y_{-1}, y_0, y_1, y_2, \dots\}$, порожденный как

$$y_{t-1} = \varepsilon_{t-1}, \quad y_0 = \varepsilon_0 + \theta_1 \varepsilon_{-1}, \quad y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \quad (t = 1, 2, \dots).$$

(Этот процесс называется процессом скользящего среднего второго порядка (*second-order moving average process*, МА(2)), см. главу 6.

- (a) Покажите, что $\{y_{-1}, y_2, \dots\}$ является стационарным в ковариациях. Получите выражение для автоковариаций γ_j ($j = 0, 1, 2, \dots$).
 (b) Пусть

$$\begin{aligned} \gamma_j &= E(y_t y_{t-j}) = E(y_{-1} y_{-1-j}) = \dots = E(y_{t-j} y_{t-j-1}) \\ &= E(y_{-j} y_{-j-1}) = \dots = E(y_{-j} y_{-j-1}) \\ &\quad (j = 1, 2, \dots; j = 0, 1, 2, \dots), \end{aligned}$$

Покажите, что

$$\gamma_0 = \sigma_\varepsilon^2(1 + \theta_1 + \theta_2), \quad \gamma_1 = \theta_1 \sigma_\varepsilon^2, \quad \gamma_2 = \theta_2 \sigma_\varepsilon^2, \quad \gamma_3 = 0, \quad \gamma_4 = 0, \dots$$

Указание: Существует каноническое отображение между $\{\varepsilon_{-1}, \varepsilon_0, \dots, \varepsilon_n\}$ и $\{y_{-1}, y_0, \dots, y_n\}$, поэтому $E(y_t y_{t-j}, \dots, y_{t-1}) = E(\varepsilon_t \varepsilon_{t-j}, \dots, \varepsilon_{t-1})$.

- (c) Пусть

$$\bar{y}_n \equiv \frac{1}{n}(y_{-1} + y_0 + \dots + y_{n-1})$$

Покажите, что

$$\text{Var}(\sqrt{n}\bar{y}_n) = \sigma_\varepsilon^2 \left[\left(1 + \frac{1}{n}\right) \gamma_0 + \left(1 + \frac{2}{n}\right) \gamma_1 \right].$$

(4) Поскольку $\{y_n\}$ не является i.i.d., CLT Леплера - Леви неприменима. Однако в главе 6 будет показано, что $\sqrt{n}y_n$ сходится по распределению к нормальной случайной величине. Что является средним предельного распределения? Что является дисперсией предельного распределения (то есть $\text{Var}(\sqrt{n}y_n)$)? Предположите, что лемма 2.1 применима к $\sqrt{n}y_n$. **Указание:** В лемме 2.1 положите $c_n = \sqrt{n}d_n$, $\lim_{n \rightarrow \infty} (1 - q/n) = 1$ для любого фиксированного q .

11. Дополнительно, тест Бройша - Годфре на сериальную корреляцию. В этом упражнении мы доказываем, что модифицированная Q Бокса - Пирса асимптотически эквивалентна rF из многократной регрессии (2.10.21), где r является числом лагов, а F является F -отношением для гипотезы о том, что все коэффициенты при $c_{1n}, c_{2n}, \dots, c_{rn}$ равны нулю.

Сначала установим обозначения. Пусть

$$X_{(n \times K)} = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \\ x_{n+1}' \end{bmatrix}, \quad E_{(n \times p)} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ c_1 & 0 & \dots & 0 \\ c_2 & c_1 & \dots & 0 \\ c_3 & c_2 & \dots & c_1 \\ c_4 & c_3 & \dots & c_2 \\ \vdots & \vdots & \dots & \vdots \\ c_{n-p+1} & c_{n-p+2} & \dots & c_{n-p} \end{bmatrix}, \quad e_{(n \times 1)} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \\ e_{n+1} \end{bmatrix},$$

$$B_{(K+1) \times (K+1)} = \begin{bmatrix} \frac{1}{n} X'X & \frac{1}{n} X'E \\ \frac{1}{n} E'X & \frac{1}{n} E'E \end{bmatrix}, \quad \hat{B} := \begin{bmatrix} \hat{B}^{(1)}_{(n \times K)} & \hat{B}^{(2)}_{(n \times p)} \\ \hat{B}^{(3)}_{(p \times K)} & \hat{B}^{(4)}_{(p \times p)} \end{bmatrix},$$

$$\hat{\gamma}_{(p \times 1)} = \begin{bmatrix} \hat{\gamma}_1 \\ \vdots \\ \hat{\gamma}_p \end{bmatrix},$$

где e_t является белым шумом сходящейся регрессии $y_t = x_t'\theta + e_t$, $\hat{\gamma}_j$ является выбранной диагональной матрицей j -го лага, определенной в (2.10.8), вычисленной по остаткам.

- (a) В многократной регрессии (2.10.21) имеет $K + p$ регрессоров. Пусть $\hat{\alpha}$ является вектором $K + p$ коэффициентов. Покажите, что

$$\hat{\alpha} = \hat{B}^{-1} \begin{bmatrix} 0 \\ \hat{\gamma} \end{bmatrix},$$

Указание: Поскольку e является вектором остатков из исходной регрессии с X в качестве регрессора, $X'e = 0$.

(b) Покажите:

$$\tilde{U} = U + \begin{bmatrix} \Sigma_{ee} & H \\ H' & \sigma^2 I_p \end{bmatrix},$$

где

$$H_{(n \times p)} \equiv [E(x_1 \cdot e_{t-1}) \quad \dots \quad E(x_1 \cdot e_{t-p})].$$

Указание: p -й элемент $\frac{1}{n} X' E$ равен $\frac{1}{n} \sum_{t=1}^n x_1 \cdot e_{t-p}$. Используйте (2.3.9), чтобы показать, что он сходится к дисперсии e в $E(x_1 \cdot e_{t-p})$.

(c) (Очень легко.) Покажите, что $\tilde{\sigma}^2 \rightarrow_p 0$.

(d) Пусть SSR здесь является суммой квадратов остатков в искусственной, а не исходной регрессии. Покажите:

$$\frac{SSR}{n - K - p} \rightarrow_p \sigma^2,$$

где σ^2 является дисперсией e_t , ошибки из оригинальной регрессии.

Указание: $\frac{1}{n} K'e = \tilde{\gamma}$. Покажите, что $\frac{SSR}{n} = \frac{1}{n} e'e = \tilde{\gamma}' \begin{bmatrix} 0 \\ \tilde{\gamma} \end{bmatrix}$.

(e) Покажите:

$$R^2 = \frac{n \tilde{\gamma}' \tilde{B}^{22} \tilde{\gamma}}{SSR/(n - K - p)}.$$

Указание: Примените формулу для R -отношения в (1.4.9) к вспомогательной регрессии.

(f) Используйте фигуру для обращения блочных матриц (см. (A.10) приложения A), чтобы показать

$$\tilde{B}^{22} = \left[\frac{1}{n} K'E - \left(\frac{1}{n} E'X \right) S_{xx}^{-1} \left(\frac{1}{n} X'E \right) \right]^{-1}.$$

(g) Пусть $\tilde{\mu}$ и $\tilde{\Phi}$ будут такими же, как в (2.10.5) и (2.10.19). Покажите, что модифицированная Q Бокса — Пирса, определенная в (2.10.20), асимптотически эквивалентна $\mu P'$ (то есть разность между ними сходится к нулю по вероятности). **Указание:** Покажите, что $\mu P'$ и модифицированная Q Бокса — Пирса асимптотически эквивалентны

$$n \tilde{\gamma}' (I_p - \tilde{\Phi})^{-1} \tilde{\gamma} / \sigma^4,$$

где $\tilde{\Phi}$ определена в (2.10.18).

12. (Дополнительно, тест Чоу на структурный сдвиг при больших выборках.) Рассмотрим применение теста Чоу на структурный сдвиг к регрессионной модели с условной гомоскедастичностью, описанной в утверждении 2.5. Поскольку проблема структурного сдвига возникает главным образом в моделях временных рядов, мы используем « t » в качестве индекса. Мы обобщаем предложение 2.1, допуская изменение вектора коэффициентов в момент сдвига τ :

$$\beta_t = \begin{cases} \alpha'_t \beta_1 + \varepsilon_t & \text{для } t = 1, 2, \dots, \tau, \\ \alpha'_t \beta_2 + \varepsilon_t & \text{для } t = \tau + 1, \tau + 2, \dots, n. \end{cases}$$

Мы предполагаем, что дата сдвига τ известна. (Когда дата сдвига неизвестна, ситуация более сложная и является объектом недавних исследований. См.: [Stock, 1994, Section 6] в качестве примера.) Тестируемая нулевая гипотеза: $\beta_1 = \beta_2$. Это множество из K ограничений. Пусть SSR_1 является суммой квадратов остатков для первого периода ($t = 1, 2, \dots, \tau$), SSR_2 является SSR для второго периода ($t = \tau + 1, \tau + 2, \dots, n$), а SSR_R является SSR для всей выборки при ограничении $\beta_1 = \beta_2$. Статистика Чоу была определена в главе 1 как

$$F = \frac{[SSR_R - (SSR_1 + SSR_2)]/K}{(SSR_1 + SSR_2)/(n - 2K)}.$$

Пусть b_1 является OLS-оценкой для β_1 , полученной по первому периоду, а b_2 является OLS-оценкой для β_2 по второму периоду.

(а) Покажите, что KF численно равно:

$$\sqrt{n}(b_1 - b_2)' \left[\left(\frac{\frac{1}{n} \sum_{t=1}^{\tau} \mathbf{x}_t \mathbf{x}_t'}{s^2} \right)^{-1} - \left(\frac{\frac{1}{n} \sum_{t=\tau+1}^n \mathbf{x}_t \mathbf{x}_t'}{s^2} \right)^{-1} \right]^{-1} \sqrt{n}(b_1 - b_2),$$

где $s^2 = (SSR_1 + SSR_2)/(n - 2K)$. Указание: Приведенные выше n уравнений могут быть записаны в матричном виде как $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, где

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{pmatrix}, \quad \mathbf{X}_1 = \begin{bmatrix} \mathbf{x}'_1 \\ \vdots \\ \mathbf{x}'_{\tau} \end{bmatrix},$$

$$\mathbf{X}_2 = \begin{bmatrix} \mathbf{x}'_{\tau+1} \\ \vdots \\ \mathbf{x}'_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}.$$

Нулевая гипотеза H_0 или $H_1 = H_0$, может быть принята или $H_0 \neq H_1$, где $H_1 = [I_K : -I_K]$ и $r = 0$. Используйте утверждение 1.4.

- (b) Пусть $\lambda = r/\alpha$. Покажите, что при $n \rightarrow \infty$ с фиксированным λ

$$\frac{1}{n} \sum_{t=1}^n x_t x_t' \xrightarrow{p} \lambda \Sigma_{xx}, \quad \frac{1}{n} \sum_{t=r+1}^n x_t x_t' \xrightarrow{p} (1 - \lambda) \Sigma_{xx},$$

где $\Sigma_{xx} \equiv E(x_t x_t')$.

- (c) Покажите, что при $n \rightarrow \infty$ с фиксированным λ

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n x_t \cdot z_t \xrightarrow{d} N(0, \lambda \sigma^2 \Sigma_{xx}),$$

$$\frac{1}{\sqrt{n}} \sum_{t=r+1}^n x_t \cdot z_t \xrightarrow{d} N(0, (1 - \lambda) \sigma^2 \Sigma_{xx}).$$

Можно доказать (см.: [Stock, 1994, Section 5]), что $\frac{1}{\sqrt{n}} \sum_{t=1}^n x_t z_t$ и $\frac{1}{\sqrt{n}} \sum_{t=r+1}^n x_t z_t$ асимптотически некоррелированы. Поэтому

$$\left[\frac{1}{\sqrt{n}} \sum_{t=1}^n x_t \cdot z_t \right] \xrightarrow{d} N \left(0, \begin{bmatrix} \lambda \sigma^2 \Sigma_{xx} & 0 \\ 0 & (1 - \lambda) \sigma^2 \Sigma_{xx} \end{bmatrix} \right).$$

- (d) Покажите, что $\text{Asy}(\hat{b}_1 - \hat{b}_2) = (\lambda^{-1} \sigma^2 \Sigma_{xx}^{-1} - (1 - \lambda)^{-1} \sigma^2 \Sigma_{xx}^{-1})$. Укажите: Пусть

$$\hat{b} \equiv \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix}.$$

Покажите, что

$$\text{Asy}(\hat{b}) = \begin{bmatrix} \lambda^{-1} \sigma^2 \Sigma_{xx}^{-1} & 0 \\ 0 & (1 - \lambda)^{-1} \sigma^2 \Sigma_{xx}^{-1} \end{bmatrix}.$$

$\hat{b}_1 - \hat{b}_2 = [I_K : -I_K] \hat{b}$. Используйте лемму 2.4(c).

- (e) В заключение покажите, что $KF \xrightarrow{p} \chi^2(K)$ при $n \rightarrow \infty$ с фиксированным $\lambda = r/\alpha$.

Эмпирические приложения

1. Перед выполнением этого упражнения прочитайте введение и параграфы 1–14 в [Fama, 1975]. В файле данных MJSLIN.ASC месячные данные предоставлены в виде:

Столбец 1: год.

Столбец 2: месяц.

Столбец 3: одномесячный темп инфляции (в процентах, годовой темп; назовите его *PA11*).

Столбец 4: трехмесячный темп инфляции (в процентах, годовой темп; назовите его *PA13*).

Столбец 5: одномесячная ставка по казначейским векселям (в процентах, годовой темп; назовите его *TBI*).

Столбец 6: трехмесячная ставка по казначейским векселям (в процентах, годовой темп; назовите его *TBI3*).

Столбец 7: CPI для городских потребителей, все составляющие (среднее за 1982–1984 годы приравнивается к 100; назовите его *CPI*).

Период выборки составляют наблюдения с февраля 1950 г. по декабрь 1990 г. (491 наблюдение). Данные по *PA11*, *PA13*, *TBI* и *TBI3* те же самые, которые использовались в [Mishkin, 1992]) и стали доступными для нас благодаря ему. Данные по казначейским векселям получены из Center for Research in Security Prices (CRSP) Чикагского университета. Ставкой по казначейским векселям за конкретный месяц является ставка на последний рабочий день предыдущего месяца (и, поэтому, она может быть взята как процентная ставка на начало месяца). Построение *PA11* и *PA13* будет интуитивно в конце этого упражнения; в течение некоторого времени мы будем использовать инфляцию, полученную из *CPI*.

- (а) (Библиотечная/интернет-работа.) Чтобы проверить правильность данных в *MISHKIN.ASC*, найдите соответствующие таблицы из прошлых номеров *Treasury Bulletin* или *Federal Reserve Bulletin* с распечатанными данными о ставках по казначейским векселям. Или посетите веб-сайты Board of Governors (www.bog.frb.fed.us) или Treasury Department (www.ustreas.gov) для достижения той же цели. Можете ли вы найти одномесячные ставки по казначейским векселям? [Ответ: Вероятно, нет.] Являются ли ставки в *MISHKIN.ASC* близкими к соответствующим таблицам? Похожи ли они на ставки на начало месяца?
- (б) (Библиотечная/интернет-работа.) Найдите соответствующие таблицы из прошлых номеров *Monthly Labor Review* (или посетите веб-сайт Бюро статистики труда, www.bls.gov) для проверки корректности графиков *CPI* и *MISHKIN.ASC*. Проверьте, что датировка переменных такова, что январское значение *CPI* является *CPI* за этот месяц. Относительно определения *CPI* проверьте следующее. 1) *CPI* является *CPI* для городских потреба-

телен, для всех составляющих, включающих продукты питания и жилье, и без сезонной коррекции. 2) Измерение цен всех составляющих индекса производится в течение месяца. Когда объявляется CPI за месяц?

- (с) CPI является индексом с фиксированными весами или переменными весами? **Указание:** Размните ваш старый учебник по макроэкономике промежуточного уровня; учебник по макроэкономике для магистрантов не понадобится.

Одномесечная ставка по казначейским векселям за месяц t в наборе данных соответствует периоду с начала месяца t до конца этого месяца (как вы только что проверили). В идеале, если бы у нас были данные по урону цен на начало каждого периода, мы могли бы вычислить темп инфляции за тот же самый период как $(P_{t+1} - P_t)/P_t$, где P_t — уровень цен на начало периода. Мы используем CPI за месяц $t-1$ для P_t (то есть полагаем $P_t = CPI_{t-1}$). Поскольку компоненты CPI собраны в различные моменты в течение месяца, возникает неизбежное смещение меры инфляции и данных по процентным ставкам. Другая проблема заключается в сроках опубликования данных о CPI. Из теории эффективности рынка следует, что P_t^e известна рынку на начало месяца t , когда установлены ставки на казначейские векселя на месяц. Однако CPI за месяц $t-1$, который мы берем в качестве P_t , не объявляется до некоторого момента следующего месяца (месяца t). Таким образом, мы предполагаем, что люди знают CPI за месяц $t-1$ на начало месяца t .

Совет при работе с пакетом TSP: При загрузке данных и вычислении темпа инфляции вы должны использовать способность TSP обрабатывать календарные даты. Начальная часть вашей программы в TSP могла бы выглядеть так:

```
? The data are monthly, 1950:2 thru 1993:12
freq m;smpl 50:2 90:12;
? Read in the ASCII data
read(file=?mishkin.asc?); year month pail pai3
tb1 tb3 cpi;
? Calculate inflation rate and the real rate
smpl 50:3 90:12;
pai=((cpi/cpi(-1))^12-1)*100;
r = tb1-pai;
```

Совет при работе с пакетом RATS: Аналогично начальная часть вашей программы RATS может выглядеть так:

```
* The data are monthly, 1950:2 thru 1990:12
cal 50 2 12:all 0 90:12
* Read in the ASCII data
open data mishkin.asc
data(org=oba) /year month pail pai3 tb1 tb3 cpi
* Calculate inflation rate and real rate
set pai = (cpi{t}/cpi{t-1})*12-1)*100
set r = tb1{t)-pai{t)
```

- (d) Воспроизведите результаты в табл. 2.1. Поскольку ставка по казначейским векселям указывается в процентах в расчете за год, темп инфляции должен измеряться в тех же единицах. Вычислите r_{t+1} , которая должна быть согласована с TB_1 (одномесячная ставка по казначейским векселям за месяц t), то есть вычисляться по формуле сложных процентов:

$$\left[\left(\frac{1+r}{1} \right)^{12} - 1 \right] \times 100$$

Совет при работе с пакетом GAUSS: Один из способов вычисления выборочных автоковариаций по формуле (2.10.1) и выборочных автокорреляций по (2.10.2) таков. Пусть z числится n -мерным вектором, i -я компонента которого — z_i , а $nobs$ является размером выборки, n .

```
z=z-trans(z); /* de-mean the series */
nlag=12; /* lag length */
rho=zeros(nlag,1);
j=1;do until j>nlag;
rho[j]=z'*shiftr(z',j,0)/nobs;
/* sample autocovariances */
j=j+1;endc;
rho=rho./(z'*z/nobs);
/* sample autocorrelations */
```

Совет при работе с пакетом TSP: Используйте команду `WIDENT`. Она не вычисляет r -значения, потому что (если вы не готовы учиться, как использовать матричные команды TSP) откажитесь от идеи получения r -значений.

Совет при работе с пакетом TSP: Используйте `CORRELATE` или `BOXJENK`.

- (e) Можете ли вы воспроизвести (2.11.0), которая использует стандартные ошибки? Какова интерпретация константы?

Совет при работе с пакетом GAUSS: Для робастных и гетероскедастичности стандартных ошибок мы должны вычитать $\hat{S}$ по формуле (2.5.1) с $\hat{\varepsilon}_i = u_i$ (OLS-остаток). Пакет Gauss-кода заключается в следующем. Пусть $x = n \times K$ матрица данных и $eobs = n$ -мерный вектор OLS-остатков. Если g является $n \times K$ -матрицей, i -я строка которой — $x_i' - e_i$, Gauss-код для порождения $g = g'g^{-1}$ — `g = x' * eobs / nobs`. Gauss-код для вычисления $\hat{S} = g'g / nobs$, где $nobs$ является объектом выборки. Это, однако, не использует тот факт, что $\hat{S}$ симметрична. Более эффективная кофакторная команда является `moment(g, 0) / nobs`.

Совет при работе с пакетом TSP: Воспользуйтесь опцией `DSTYPE = 0` для OLSQ.

Совет при работе с пакетом GAUSS: Используйте объект ROBUSTERRORES от LINREG.

- (f) (Коррекция Давидсона — Мак-Кинчона для оценки Уайта.) Скопости на конечных выборках робастности t -отношения могут быть улучшены коррекцией Давидсона — Мак-Кинчона, которая раскрывается в основном тексте. Вычислите робастные стандартные ошибки, используя три формулы, которые обсуждались в основном тексте. Первая — это коррекция на степени свободы умножением $\hat{S}$ на $n/(n - K)$. Вторая — вычитание $\hat{S}$ по формуле (2.5.5) с $d = 1$. Третья — (2.5.5) с $d = 2$. Они соответствуют опциям `HCTYPE = 1, 2, 3` программы OLSQ для TSP.

Совет при работе с пакетом GAUSS: Пусть $p = n$ -мерный вектор для сохранения p_i в (2.5.5), `exhinv = Sinv-1`, $x = n \times K$ матрица данных X . Один из способов вычислить p таков:

```
p=zeros(nobs,1);
i=1;do until i>nobs;
p[i]=x[i,]*exhinv*x[i,]'/nobs;
i=i+1;enddo;
```

- (g) (Оценивание при условной гомоскедастичности.) Проверьте эффективность рынка, регрессируя π_{t+1} на константу и TBI_t при условной гомоскедастичности (2.6.1.) Сравните ваши результаты с полученными в (e) и (f). В какой части имеются отличия?
- (h) (Тест Бройша — Голфри.) Для спецификации в (g) проведите тест Бройша — Голфри на серийную корреляцию с $p = 12$. (Значение статистики dh^2 должно быть около 27.0). Пусть e_t ($t = 1, 2, \dots, n$) является OLS-остатком из регрессии в (g). Для выполнения теста Бройша — Голфри, как описано в тексте, вам нужно положить e_t ($t = 0, -1, \dots, -11$) равными нулю при проведении вспомогательной регрессии для $t = 1, \dots, n$.

Совет при работе с пакетом TSP: Код TSP для вычисления статистики nR^2 таков. resid является OLS-остатком из исходной регрессии.

```
? Part (h): Breusch-Godfrey
resid=@res;
smpl 52:1 52:12;
resid=0;
smpl 53:1 71:7;
olsq resid c tbl resid(-1) resid(-2) resid(-3)
               resid(-4) resid(-5) resid(-6)
               resid(-7) resid(-8) resid(-9)
               resid(-10) resid(-11)
               resid(-12);
```

```
set w=@nob*@rsq;cdf(chisq,df=12) w;
```

Здесь w является статистикой nR^2 .

Совет при работе с пакетом RATS: Код RATS для вычисления статистики nR^2 следующий. resid является OLS-остатком из исходной регрессии.

```
* Part (h): Breusch-Godfrey
set resid 53:1 71:7 = pai-%beta(1)-%beta(2)*tbl
set resid 52:1 52:12 = 0
linreg resid 53:1 71:7;
# constant tbl resid1 to 12
compute w = %nobs*%rsquared
cdf chisquared w 12
```

Здесь w является статистикой nR^2 .

Совет при работе с пакетом GAUSS: Как и в кодах Gauss для (d), полезной является команда shiftr. Пусть y (223×1) и x (223×2), соответственно, являются вектором значений зависимой переменной и матрицей регрессоров из предыдущей регрессии. Пусть b (2×1) является вектором оцененных коэффициентов из предыдущей регрессии. Ваша программа Gauss для создания y и x для вспомогательной регрессии может выглядеть так:

```
"@ Part (h): Breusch-Godfrey @";
y=y-x*b; /* residual vector from the previous
regression */
j=1;do until j>12;
x=x shiftr(y',j,0)';
j=j+1;endo;
```

- (.) (Реконструкция Фамы.) Каковы собственные точечные оценки и стандартные ошибки коэффициента при номинальной процентной ставке у Фамы? Идентичны ли они вашим результатам? (Фама использует другие обозначения, поэтому мы должны переработать его результаты в наши обозначения.) Почему его оценка константы отличается от вашей? (Одна очевидная причина заключается в том, что наши данные другие, но существуют и другие причины.) Дополнительно: Каковы различия между его данными и нашими?
- (j) (Сезонные дампы.) CPI, использованный для вычисления темпа инфляции, не является сезонно скорректированным. Чтобы принять во внимание сезонные факторы при продолжении использования сезонно нескорректированных данных в регрессии, определим двенадцать месячных дампов, $M1, M2, \dots, M12$, и используем их вместо константы в регрессии в (g). Приводит ли это к изменению ваших результатов? Что случится, если вы включите константу вместе с двенадцатью месячными дампами в регрессию? Дополнительно: Есть ли какая-нибудь причина для предпочтения сезонно нескорректированных данных сезонно скорректированным?

Совет при работе с пакетом TSP: Используйте команду `DUMMY` для порождения месячных дампов.

Совет при работе с пакетом RATS: Используйте команду `SEASONAL` для порождения месячных дампов.

Хорошо известная проблема с рядом CPI заключается в том, что его компонента, связанная с жилищным строительством, представляет расходы на проценты по ипотеке до 1983 года и более подходящую меру — «эквивалент арендной платы» — с 1983 года. Переменные инфляции $PA11$ и $PA13$ в `MISHKIN ASC` вычислены по индексу цен, который использует «эквивалент арендной платы» для всех периодов. Для получения дополнительной информации об этом см.: Section II [Huiizuga and Mishkin, 1984]. Датировка переменных такова, что наблюдение за январь для $PA11$ вычислено по декабрьским и январским данным индекса цен, а $PA13$ — по декабрьским и мартовским данным индекса цен. Поэтому при предположении о том, что сезонный индекс соответствует концу месяца, январское наблюдение для $PA11$ является темпом инфляции в течение этого месяца, а январское наблюдение для $PA13$ является темпом инфляции с начала января по конец марта.

- (k) Оцените регрессию Фамы для 1/53–7/71, используя лучшую меру одномесечного темпа инфляции, и сравните результаты с тем, которые получены в (e).

- (b) Оцените регрессию Фокса за период после октября 1979 года. На сколько изменились номинальный процентная ставка? (Коэффициент должен снизиться до 0,364.)

2. (Продолжение упражнения Нерлова в главе 1, с. 102.)

- (a) Для модели 4 выполните тест « R^2 Уайта на условную гетеросkedастичность. (Значение χ^2 должно быть равным 66,546).
- (b) (Байес-статистика.) Оцените Модель 4 посредством OLS. Это модель 1 процедуры лагарафа 2.8. Выполните шаг 2, оценивая следующую вероятностную регрессию: регрессируйте $\hat{\epsilon}_i^2$ на константу и $1/\hat{\sigma}_i^2$. Проверьте, что шаг 3 состоит тем, что вы делали в части (b).
- (c) Вычислите стандартные ошибки Уайта для модели 4

Упражнения на метод Монте-Карло

1. (Коррекция на степень свободы.) В модели упражнения на метод Монте-Карло в главе 1 мы предположили, что ошибки были нормально или. Другим важным аспектом того, что мы предположили, что ошибки равномерно между -0.5 и 0.5 .
- (a) Проверьте, что DGP второй симуляции (где X различна для симулируемых выборок) удовлетворяет предположениям 2.1–2.5 и 2.7. (Можно показать, что предположение 2.2 [уточненная стационарность] также удовлетворяется. Это является следствием утверждения 5.1(d).)
- (b) Проведите вторую симуляцию из упражнения на метод Монте-Карло в главе 1 с нормальным распределением ошибок. В каждом повторении вычислите обычное t -отношение, как ранее, и сравните его с двумя критическими значениями. Первое является тем же самым критическим значением (2.042), которое указывалось экспрессивным $t(30)$ и второе является критическим значением (1.96) для $N(0, 1)$. Вычислите частоту отвержения для каждого критического значения. Какие из них ближе к номинальному размеру 5%?
2. (Противопоставление статистик Бокса — Пирса и Льюнга — Бокса.) Мы хотим проверить утверждение о том, что свойства на конечных выборках Q -статистики Льюнга — Бокса приближаются свойства на конечных выборках Q -статистики Бокса — Пирса. Создайте строку из 50 случайных чисел i.i.d. со средним 0. (Выберите распределение, которое вам нравится.) Беря эту строку в качестве данных, сделайте следующее:

- (1) Вычислите статистики Q Бокса — Пирса и Q Льюнга — Бокса (см. (2.10.4) и (2.10.5)) с $p = 4$. (Среднее по ансамблю равно 0 по построению. Однако истинное дисперсионное среднее ряда тогда вычислите автокорреляции.)
- (2) Для каждой статистики примите или отвергните нулевую гипотезу об отсутствии сериальной корреляции на уровне значимости 5%, предполагая, что статистика распределена как $\chi^2(4)$.

Продавайте это большое количество раз, каждый раз используя различные строки из 50 i.i.d. случайных чисел, порожденных законом. Для каждой Q -статистики вычислите частоту отвержения нулевой гипотезы. Если это распределение частоты на выборках хорошо аппроксимируется $\chi^2(4)$, эта частота должна быть близка к 5%. Какая статистика дает вам частоту, более близкую к номинальному размеру 5%? Не слишком ли часто мы не отвергаем нулевую гипотезу, если используем Q Бокса — Пирса?

Ответы на избранные вопросы

Аналитические упражнения

3. Из (2.3.7) в тексте $b + \beta = S_{xx}^{-1} \bar{y}$. По Кошкаторезу, $S_{xx} \rightarrow_p E(x_i x_i')$ и $\bar{y} \rightarrow_p \mu$. Поэтому b сходятся к μ СЛТ Ливингтона — Ланга, $\sqrt{n}(b - \mu) \rightarrow_c N(0, S)$. Оставшаяся часть доказательства теперь должна быть рутинной.
7. Что остается показать — это то, что левая часть (34) равна:

$$A_{\text{OLS}}(b) = \Sigma_{xx}^{-1} S \Sigma_{xx}^{-1}, \quad \Sigma_{xx} = \lim_{n \rightarrow \infty} \frac{1}{n} X'X$$

и

$$\begin{aligned} S &= E[\varepsilon_i^2 x_i x_i'] = \\ &= E[E[\varepsilon_i^2 | x_i] x_i x_i'] = \\ &= E[x_i(\sigma^2) x_i x_i'] \quad \text{по (2.8.1)} = \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (x_i' \Delta) x_i x_i' \quad (\text{по эргодической стационарности}) = \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} X' V X \quad (\text{по предельности (2.8.4) для } V). \end{aligned}$$

Эмпирические упражнения

- 1c. CPI, вероятно, наиболее широко использующийся индекс цен с фиксированными весами.
- 1d. Постоянная составляющая, взятая со знаком «минус», равна τ (постоянная ex ante процентная ставка).
- 1e. Точечные оценки те же самые. Отличаются только стандартные ошибки.
- 1f. Посмотрите на таблицу 3 Фамы, его коэффициент при R_t , когда период выборки с 1/53 по 7/71, равен 0,98 с с.в. 0,10. Очень близко, но не точно то же самое. Есть два возможных объяснения этой разницы между нашими оценками и оценками Фамы. Во-первых, в вычислении темпа инфляции за месяц t , $\pi_{t-1} = (P_{t+1} - P_t)/P_t$, из Фамы ясно, что для P_t используется CPI_{t-1} , а не CPI_t . Во-вторых, нес для CPI во время его написания мог быть взят за 1958 год. Все же для CPI в наших данных взят за 1982–1984 годы. Наша оценка константы (–0,868) отличается от оценки Фамы (0,00070), поскольку, во-первых, у него зависимой переменной является темп инфляции, взятый со знаком «минус», и, во-вторых, темп инфляции и ставка по казначейским векселям берутся у него в расчете на месяц.
- 1j. Сезонно скорректированные ряды, выпускаемые BLS, являются своего рода двусторонними скользящими средними. Таким образом, например, сезонно скорректированный CPI_t зависит от сезонно нескорректированных значений *будущего* CPI , которые не содержатся в I_t . Значит, если нет причин принимать во внимание сезонные факторы в соотношении между темпом инфляции и номинальной процентной ставкой, нужно использовать сезонно нескорректированные данные. Если есть необходимость принять во внимание сезонные факторы, лучше включить сезонные дамми в регрессию, а не использовать сезонно скорректированные данные. Включение сезонных дамми не изменяет результаты сколько-нибудь существенно образом.
- 1k. Коэффициент R_t снижается с 1,014 до 0,807.

Библиография

- Americano, T., 1977, "A Note on a Heteroscedastic Model," *Journal of Econometrics*, 6, 365-370.
- , 1986, *Advanced Econometrics*, Cambridge: Harvard University Press.
- Anderson, T. W., 1971, *The Statistical Analysis of Time Series*, New York: Wiley.
- Billingsley, P., 1961, "The Lindeberg-Levy Theorem for Martingales," in *Proceedings of the American Mathematical Society*, Volume 12, 788-792.
- Bax, G., and D. Pierce, 1970, "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, 65, 1509-1526.
- Breusen, T., 1978, "Testing for Autocorrelation in Dynamic Linear Models," *Australian Economic Papers*, 17, 334-355.
- Davidson, R., and J. MacKinnon, 1993, *Estimation and Inference in Econometrics*, Oxford: Oxford University Press.
- Durbin, J., 1970, "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors are Lagged Dependent Variables," *Econometrica*, 38, 410-421.
- Engle, R., 1982, "An Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1001.
- Fama, E., 1975, "Short-Term Interest Rates as Predictors of Inflation," *American Economic Review*, 65, 266-282.
- , 1970, *Foundations of Finance*, Basic Books.
- Goodrey, L., 1978, "Testing against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables," *Econometrica*, 46, 1283-1301.
- Greene, W., 1997, *Econometric Analysis* (3d ed.), New York: Prentice Hall.
- Gregory, A., and M. Vellal, 1988, "Formulating Wald Tests of Nonlinear Restrictions," *Econometrica*, 56, 1465-1468.
- Hall, P., and C. Heyde, 1980, *Martingale Limit Theory and Its Application*, New York: Academic.
- Hall, R., 1978, "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, 86, 971-987.
- Hamilton, J., 1994, *Time Series Analysis*, Princeton: Princeton University Press.
- Huizinga, J., and F. Mishkin, 1984, "Inflation and Real Interest Rates on Assets with Different Risk Characteristics," *Journal of Finance*, 39, 609-714.
- Judge, G., W. Griffiths, R. Hill, H. Lutkepohl, and T. Lee, 1985, *The Theory and Practice of Econometrics* (2nd ed.), New York: Wiley.
- Karlin, S., and H. M. Taylor, 1975, *A First Course in Stochastic Processes* (2d ed.), New York: Academic.
- Koopmans, T., and W. Hoxd, 1953, "The Estimation of Simultaneous Linear Economic Relationships," in W. Hoxd and T. Koopmans (eds.), *Studies in Econometric Method*, New Haven: Yale University Press.
- Mishkin, F., 1992, "Is the Fisher Effect for Real?" *Journal of Monetary Economics*, 30, 195-215.

- Newey, W., 1984, "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29, 229–250.
- Rao, C. R., 1973. *Linear Statistical Inference and Its Applications* (2nd ed.). New York: Wiley.
- Snick, J., 1994. "Unit Roots, Structural Breaks and Trends," in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*. Volume IV, New York: North Holland.
- White, H., 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimation and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

Глава 3. Одномерный GMM

Главным допущением метода наименьших квадратов является ортогональность случайных ошибок и регрессоров. Без этого допущения МНК-оценки не будут даже состоятельными. Поскольку во многих важных приложениях условие ортогональности не выполняется, необходимо уметь работать с эндогенными регрессорами. Решение дает метод оценивания, называемый **обобщенным методом моментов (GMM)** (*Generalized Method of Moments*), который включает в себя МНК как частный случай. В этой главе представлен GMM для моделей с одним уравнением, а в следующей главе представлено обобщение этого метода на несколько уравнений. Эти главы касаются только линейных уравнений. Оценка обобщенным методом моментов нелинейных уравнений будет описана в главе 7. В главе 8 будет представлена основная альтернатива GMM, метод максимального правдоподобия (ML).

Отражая преобладание в экономике эндогенных регрессоров, данная глава начинается с нескольких примеров. Общее определение эндогенных регрессоров приведено в параграфе 3.3. В параграфе 3.4 представлена GMM-процедура для модели из параграфа 3.3. Параграфы 3.5–3.7 представляют асимптотические свойства оценок GMM и соответствующих тестовых статистик. При наличии условной гомоскедастичности выведенные в данных параграфах формулы можно упростить. Упрощенные формулы собраны в параграфе 3.8. В частности, двухшаговый МНК, метод, разработанный изначально для оценивания систем одновременных уравнений, является частным случаем GMM. ML-альтернатива двухшагового МНК, называемая максимальным правдоподобием с ограниченной информацией (LIML), будет рассмотрена в параграфе 8.6.

Эмпирические упражнения в этой главе связаны с наиболее часто оцениваемым в экономике уравнением — уравнением заработной платы. Это уравнение связывает ставку заработной платы с различными характеристиками индивида, такими как образование и способности. Поскольку образование — это выбор индивида и поскольку способности оцениваются плохо, регрессоры, по-видимому, являются эндогенными. Применяя к этому уравнению описанные в данной главе методики, мы убеждаемся в том, что

оценки параметров зависят от того, произведём ли мы коррекцию на эндогенность.

3.1. Смещение из-за эндогенности: пример Уоркинга

Система одновременных уравнений рыночного равновесия

Классическим примером смещений, создаваемых эндогенными регрессорами, является пример Уоркинга (1927). Рассмотрим следующую простую модель спроса и предложения:

$$q_t^d = \alpha_0 + \alpha_1 p_t + u_t \quad (\text{уравнение спроса}), \quad (3.1.1a)$$

$$q_t^s = \beta_0 + \beta_1 p_t + v_t \quad (\text{уравнение предложения}), \quad (3.1.1b)$$

$$q_t^d = q_t^s \quad (\text{рыночное равновесие}). \quad (3.1.1c)$$

Здесь q_t^d — объём спроса на рассматриваемый товар (допустим, кофе) в период t , q_t^s — объём предложения, p_t — цена. Ошибка u_t отражает иные факторы помимо цены, влияющие на спрос на кофе, например, отношения к кофе. В зависимости от значения u_t кривая спроса в координатах «цена — объём» смещается вверх или вниз. Аналогично v_t отражает иные факторы предложения помимо цены. Предположим, что $E(u_t) = 0$ и $E(v_t) = 0$ (в противном случае свободные члены α_0 и β_0 изменятся на ненулевую величину). Для того чтобы избежать непринципиальных усложнений, предположим также, что $\text{Cov}(u_t, v_t) = 0$. Если предположить, что $q_t = q_t^d = q_t^s$, систему из трёх уравнений (3.1.1) можно свести к системе из двух уравнений:

$$q_t = \alpha_0 + \alpha_1 p_t + u_t \quad (\text{уравнение спроса}), \quad (3.1.2a)$$

$$q_t = \beta_0 + \beta_1 p_t + v_t \quad (\text{уравнение предложения}). \quad (3.1.2b)$$

Регрессор называется **эндогенным** (*endogenous*), если он не является предопределённым (то есть не ортогонален вектору ошибок) и, таким образом, не удовлетворяет условию нулевой корреляции. При наличии n уравнений n -мерной составленной системы регрессионная модель нарушается, и, соответственно, регрессор является эндогенным тогда и только тогда, когда регрессор коррелирован с ошибкой. В данном примере регрессор p_t с необходимостью является эндогенным в обоих уравнениях. Чтобы понять почему, рассмотрим (3.1.2) как систему одновременных уравнений и решим её относительно (p_t, q_t) :

$$p_t = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} = \frac{v_t - u_t}{\beta_1 - \alpha_1}, \quad (3.1.3a)$$

$$q_t = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 v_t - \beta_1 u_t}{\alpha_1 - \beta_1}. \quad (3.1.3b)$$

Таким образом, цена является функцией двух указанных ошибок. Используя (3.1.3а), можно вычислить ковариацию регрессора y_2 со сдвигом спроса u_1 и сдвигом предложения v_2 :

$$\text{Cov}(y_2, u_1) = -\frac{\text{Var}(u_1)}{\alpha_1 - \beta_2}, \quad \text{Cov}(y_2, v_2) = \frac{\text{Var}(v_2)}{\alpha_1 - \beta_1}. \quad (3.1.4)$$

не равные нулю (кроме случая $\text{Var}(u_1) = 0$ и $\text{Var}(v_2) = 0$). Таким образом, если кривая спроса имеет отрицательный наклон ($\alpha_1 < 0$), а кривая предложения — положительный ($\beta_1 > 0$), цена положительно коррелирована со сдвигом спроса и отрицательно со сдвигом предложения. В данном примере эндогенность является следствием рыночного равновесия.

Смещение из-за эндогенности

Что получается в результате одностороннего регрессии объема на константу и цену: кривая спроса или предложения? Ответ: ни то ни другое, так как цена эндогенна и в уравнении спроса, и в уравнении предложения. Из параграфа 2.3 следует, что оценки МНК соответствуют коэффициентам регрессии наименьших квадратов. В проекции q_1 на константу и p , коэффициент при p равен $\text{Cov}(y_2, q_1) / \text{Var}(q_1)$ ¹, таким образом,

$$\text{r-line МНК-оценки коэффициента цены} = \frac{\text{Cov}(y_2, q_1)}{\text{Var}(q_1)}. \quad (3.1.5)$$

Чтобы разрешить это соотношение относительно коэффициента цены для кривой спроса (α_1), воспользуемся уравнением спроса (3.1.2а) для расчета $\text{Cov}(y_2, q_1)$:

$$\text{Cov}(y_2, q_1) = \alpha_1 \text{Var}(q_1) + \text{Cov}(p_2, q_1). \quad (3.1.6)$$

Подставив (3.1.6) в (3.1.5), получим выражение для асимптотического смещения α_1 :

$$\text{r-line МНК-оценки коэффициента цены} = \alpha_1 + \frac{\text{Cov}(p_2, q_1)}{\text{Var}(q_1)}. \quad (3.1.7)$$

Аналогично асимптотическое смещение для коэффициента цены для кривой предложения β_1 равно:

$$\text{r-line МНК-оценки коэффициента цены} = \beta_1 + \frac{\text{Cov}(p_2, v_2)}{\text{Var}(q_2)}. \quad (3.1.8)$$

¹ Пусть γ — вектор коэффициентов МНК в уравнении $\hat{y} = \gamma'x + \varepsilon$, $\varepsilon = y - \hat{y}$. Тогда $\gamma = \text{Var}(x)^{-1} \text{Cov}(y, x)$. Чтобы убедиться в этом факте, можно использовать упражнение 2 главы 2.

Но, как показано в (3.1.4), $Cov(p_i, u_i) \neq 0$ и $Cov(p_i, v_i) \neq 0$, поэтому МНК-оценки не соответствуют ни α_1 , ни β_1 . Это явственно известно как **смещение из-за эндогенности** (*endogeneity bias*). Оно также известно как **смещение из-за одновременности уравнений** (*simultaneous equations bias*), или **смещение из-за одновременности** (*simultaneity bias*), так как регрессор и случайная ошибка часто связаны друг с другом посредством системы одновременных уравнений, как в данном примере.

В случае отсутствия сдвигов спроса ($u_i = 0$ для всех i) $Cov(p_i, u_i) = 0$ формула (3.1.7) показывает, что МНК-оценка соответствует параметру α_1 . В этом случае кривая спроса не сдвигается, и, как показано на рис. 3.1(а), все наблюдаемые при сдвиге кривой предложения комбинации цены и объема попадают на кривую спроса. В случае отсутствия сдвига предложения все наблюдаемые при сдвиге кривой спроса комбинации цены и объема попадают на кривую предложения (см. рис. 3.1(б)). В общем случае сдвига обеих кривых МНК-оценка соответствует взвешенному среднему α_1 и β_1 . Это можно показать аналитически следующим образом: выражение для предела по вероятности МНК-оценки:

$$\text{Рлим. МНК-оценки коэффициента цены} = \frac{\alpha_1 \text{Var}(u_i) + \beta_1 \text{Var}(v_i)}{\text{Var}(u_i) + \text{Var}(v_i)}. \quad (3.1.9)$$

Доказательство этого предлагается как домашнее задание.



Рис. 3.1. Кривые спроса и предложения

Наблюдаемые сдвиги предложения

Причина, по которой кривые спроса и предложения не оцениваются состоятельно в общем случае, заключается в том, что мы не можем определить из данных, чем вызвано изменение цены и объема: сдвигом спроса или сдвигом предложения. Можно предположить, что оценить кривую спроса (предложения) возможно, если наблюдаются некоторые из факторов, сдвигающих кривую предложения (спроса). Предположим,

что сдвиг предложения y_1 можно разложить на фактор x_2 и на независимый фактор ζ_1 , коррелируемый с x_1 ¹:

$$y_1 = \beta_1 + \beta_2 x_2 + \zeta_1, \quad \text{где } \beta_2 \neq 0 \text{ (предположение)}. \quad (3.1.2b')$$

Предположим теперь, что наблюдаемый сдвиг предложения y_1 является предопределенным в уравнении спроса (то есть не коррелирует с ошибкой): x_2 можно считать, например, температурой в регионах выращивания кофе. Если температура x_2 не коррелирует с ценой кофе y_2 , то должна существовать возможность выделить из движения цен компоненту, связанную с температурой, но не связанную со сдвигом спроса. Тогда возможно оценить кривую спроса как связь между потреблением кофе и этой компонентой цены.

Для рассматриваемого уравнения предопределенная переменная, коррелирующая с эндогенным регрессором, называется **инструментальной переменной** или **инструментом** (*instrumental variable or instrument*). Ее часто называют **релевантным инструментом**² (*valid instrument*), чтобы подчеркнуть ее ненулевую корреляцию с эндогенным регрессором. В данном примере наблюдаемый сдвиг предложения y_1 может служить инструментом для уравнения спроса. Это можно легко увидеть, решив систему одновременных уравнений (3.1.2a) и (3.1.2b') относительно (y_2, q_2) :

$$y_2 = \frac{\beta_0 - \alpha_2}{\alpha_2 - \beta_1} + \frac{\beta_2}{\alpha_1 - \beta_2} x_2 + \frac{\zeta_1 - u_1}{\alpha_1 - \beta_1}, \quad (3.1.10a)$$

$$q_2 = \frac{\alpha_1 \beta_0 - \alpha_2 \beta_1}{\alpha_2 - \beta_1} + \frac{\alpha_1 \beta_2}{\alpha_1 - \beta_1} x_2 + \frac{\alpha_1 \zeta_1 - \beta_1 u_1}{\alpha_1 - \beta_1}. \quad (3.1.10b)$$

Поскольку $\text{Cov}(x_2, \zeta_1) = 0$ по предположению и $\text{Cov}(x_2, u_1) = 0$ по предположению, из (3.1.10a) следует, что

$$\text{Cov}(y_2, x_2) = \frac{\beta_2}{\alpha_1 - \beta_2} \text{Var}(x_2) \neq 0.$$

Поэтому y_2 — релевантный инструмент.

Имея релевантный инструмент, можно оценить коэффициент α_1 и кривую предложения. Воспользуемся уравнением (3.1.2a) для объяснения $\text{Cov}(y_2, q_2)$ (а не $\text{Cov}(y_1, q_2)$), как в (8.1.6):

$$\begin{aligned} \text{Cov}(y_2, q_2) &= \alpha_1 \text{Cov}(x_2, q_2) + \text{Cov}(u_1, q_2) = \\ &= \alpha_1 \text{Cov}(x_2, y_2) - (\text{Cov}(y_1, q_2) = 0 \text{ по предположению}). \end{aligned}$$

¹ Это разложение всегда возможно. Если предположить, что x_2 на константе и y_1 равен $\beta_1 + \beta_2 x_2$, то означало бы $\zeta_1 = y_1 - \beta_1 - \beta_2 x_2$, тем самым $\zeta_1 = \zeta_1 + \beta_1 - \beta_2 x_2$, но определено, не коррелирует с x_2 . Подставим это выражение в (3.1.2b) и получим $y_1 = \beta_1 + \beta_2 x_2$.

² В литературе иногда используется термин *excluded instrument*, однако в литературе в этом значении чаще используется термин *exclusion* («исключение»). Термин же «valid» («валидный») употребляется часто как указание на некоррелируемость инструмента с ошибкой. — Прим. пер. ред. перевода.

Как мы только что убедились, $Cov(x_2, y_2) \neq 0$. Поэтому можем разделить обе части равенства на $Cov(x_2, y_2)$ и получить

$$\alpha_1 = \frac{Cov(x_2, y_2)}{Cov(x_2, x_2)}. \quad (3.1.11)$$

Естественной оценкой, которая напрашивается сама собой, является

$$\hat{\alpha}_{1,IV} = \frac{\text{выборочная ковариация } x_2 \text{ и } y_2}{\text{выборочная дисперсия } x_2 \text{ в } x_2}. \quad (3.1.12)$$

Эта оценка называется оценкой по методу инструментальных переменных (IV) (*instrumental variables (IV) estimator*) с x_2 в качестве инструмента. Мы иногда говорим, что эндогенный регрессор y_2 инструментизируется посредством x_2 .

Другой популярной процедурой состоятельного оценивания α_1 является двухшаговый МНК (2SLS) (*two-stage least squares (2SLS)*). Она называется так потому, что эта процедура состоит из оценивания двух регрессий. На первом шаге оценивается регрессия эндогенного регрессора y_2 на предопределенную переменную x_2 для получения подобранных значений $\hat{y}_2$. На втором шаге оценивается регрессия зависимой переменной y_1 на константу и $\hat{y}_2$. Использование $\hat{y}_2$ вместо y_2 отличает 2SLS от безысходного применения МНК для оценивания уравнения спроса. 2SLS-оценки коэффициента α_1 — это МНК-оценки коэффициента при $\hat{y}_2$ на втором шаге, поэтому она равна¹:

$$\hat{\alpha}_{1,2SLS} = \frac{\text{выборочная ковариация } \hat{y}_2 \text{ и } y_1}{\text{выборочная дисперсия } \hat{y}_2}. \quad (3.1.13)$$

Чтобы оценить регрессию на втором шаге с уравнением спроса, переписем (3.1.1a) как

$$y_1 = \alpha_0 + \alpha_1 \hat{y}_2 + (y_1 - \alpha_0 - \alpha_1 \hat{y}_2), \quad (3.1.14)$$

Регрессия на втором шаге оценивает это уравнение, tratando выражение в скобках как ошибку. Получаемое при этом МНК-оценки α_1 является оценкой по следующему уравнению. Если $\hat{y}_2$ — подобранный значение $\hat{y}_2$ было в точности равно МНК-оценке $\hat{y}_2^*(y_2, x_2)$, то мы могли бы $(y_1 - \alpha_0 - \alpha_1 \hat{y}_2)$ не были бы коррелированы с $\hat{y}_2$ из-за того, что она не коррелирована с x_2 и $\hat{y}_2$ — линейная функция x_2 и $(y_2 - \hat{y}_2)$ потому, что она является ошибкой проекции. Подбранное значение $\hat{y}_2$ не равно точно $\hat{y}_2^*(y_2, x_2)$, но эта разница исчезает по мере увеличения объема

¹ В регрессии y_1 на константу и $\hat{y}_2$ МНК-оценки коэффициента при $\hat{y}_2$ совпадают с оценками ковариации между y_1 и $\hat{y}_2$ и выборочной дисперсией $\hat{y}_2$. Доказательство этого было контрпримером к параграфу 1.2.

выборки. Таким образом, $\hat{\beta}$ асимптотически не коррелирует с ошибкой в уравнении (3.1.14), что делает 2SLS-оценку состоятельной.

В данном примере IV- и 2SLS-оценки численно совпадают (это будет доказано позднее в более общем случае). В общем, 2SLS-оценке может быть записана как IV-оценка с надлежащим образом выбранными инструментами: а IV-оценка, в свою очередь, является частным случаем GMM-оценки.

Контрольные вопросы

1. В системе одновременных уравнений (3.1.2) предположим, что $\text{Cov}(u_i, v_i)$ не обязательно равно 0. Обязательно ли, что цена и спрос сдвигаются в положительном направлении, если $\alpha_1 < 0$ и $\beta_1 > 0$? [Ответ: Нет.] Почему?
2. (3.1.7) показывает, что МНК-оценка коэффициента при цене в регрессии объема на константу и цену смещена. Смещена ли МНК-оценка постоянной составляющей α_0 ? **Указание:** Предел по левосторонности МНК-оценки постоянной составляющей равен:

$$E(q_i) - \frac{\text{Cov}(p_i, q_i)}{\text{Var}(p_i)} E(p_i).$$

Но из (3.1.2а) следует $E(y_i) = \alpha_0 + \alpha_1 E(p_i)$.

3. Выведите (3.1.9). **Указание:** Покажите, что

$$\begin{aligned}\text{Var}(p_i) &= \frac{\text{Var}(q_i) + \text{Var}(q_i)}{(\alpha_1 - \beta_1)^2}, \\ \text{Cov}(p_i, q_i) &= \frac{\alpha_1 \text{Var}(u_i) + \beta_1 \text{Var}(v_i)}{(\alpha_1 - \beta_1)^2}.\end{aligned}$$

4. Для модели рыночного равновесия (3.1.2а), (3.1.2б), в которой

$$\text{Cov}(x_i, \zeta_i) = 0, \quad \text{Cov}(x_i, u_i) = 0, \quad \text{Cov}(x_i, v_i) = 0,$$

проверьте, что в уравнении спроса цена экзогенна. А в уравнении предложения? **Указание:** Посмотрите на (3.1.10а).

Необходимо ли предположение о том, что сдвиги спроса и предложения некоррелированы ($\text{Cov}(u_i, v_i) = 0$) для состоятельности оценок $\hat{\beta}_1$ и $\hat{\beta}_{1,2SLS}$? **Указание:** Является ли x_i релевантным инструментом без этого предположения?

3.2. Другие примеры

Смещение из-за эндогенности возникает в самых различных ситуациях. Мы рассмотрим еще несколько примеров.

Простая макроэкономическая модель

Предельно простой макроэкономической моделью является иллюстративная модель Ханзелло (1943):

$$C_t = \alpha_0 + \alpha_1 Y_t + u_t, \quad 0 < \alpha_1 < 1 \quad (\text{функция потребления}), \\ Y_t = C_t + I_t \quad (\text{уравнение ВВП}).$$

где C_t — совокупное потребление в году t , Y_t — ВВП, I_t — инвестиция, α_1 — предельная склонность к потреблению (**MPC** — *Marginal Propensity to Consume*). Как известно из начального курса макроэкономики, равновесное значение ВВП:

$$Y_t = \frac{\alpha_0}{1 - \alpha_1} + \frac{I_t}{1 - \alpha_1} + \frac{u_t}{1 - \alpha_1}. \quad (3.2.1)$$

Если инвестиции предопределены, так что $\text{Cov}(I_t, u_t) = 0$, то из (3.2.1) следует, что

$$\begin{aligned} \text{Cov}(Y_t, u_t) &= \frac{\text{Var}(u_t)}{1 - \alpha_1} > 0, \\ \text{Cov}(I_t, Y_t) &= \frac{\text{Var}(I_t)}{1 - \alpha_1} > 0. \end{aligned}$$

Таким образом, доход эндогенен в функции потребления, но инвестиции являются релевантным инструментом для этого эндогенного регрессора. Простые выкладки, аналогичные используемым для вывода (3.1.7), показывают, что МНК-оценка для MPC, полученная из регрессии потребления на совокупный доход, будет статистически смещена:

$$\text{plim } \hat{\alpha}_{OLS} - \alpha_1 = \frac{\text{Cov}(Y_t, u_t)}{\text{Var}(Y_t)} = \frac{1 - \alpha_1}{1 + \frac{\text{Var}(I_t)}{\text{Var}(u_t)}} > 0. \quad (3.2.2)$$

Это, возможно, самый очевидный пример смещения из-за эндогенности. Отличие от теоремы Уоркина заключается в том, что здесь второе уравнение является тождеством, помогающим увидеть эндогенность регрессора. Асимптотическое смещение может быть скорректировано использованием инвестиций в качестве инструмента для дохода в функции потребления. Инвестиции здесь играют роль наблюдаемого сдвига предложения в примере Уоркина.

Ошибки в переменных

Термин «ошибки в переменных» (*errors in variables*) относится к случаю, когда предопределенная и эндогенность переменных с необходимостью ставит под сомнение — когда ее измерения проводятся с ошибками

Эта проблема встречается весьма часто, особенно в микроэкономических данных по домохозяйствам. Например, в известном исследовании **динамики доходов (PSID — Panel Study of Income Dynamics)** информация о переменных, таких как потребление продуктов питания и доход, собиралась по телефону. Было бы слишком оптимистичным предполагать, что опрошиваемый сможет с ходу назвать сумму, потраченную на продукты питания в определенный период времени.

Кросс-секционным версия гипотезы **постоянного дохода (Permanent Income Hypothesis)** Фришмана (1957) может быть аккуратно сформулирована как задача с ошибками с переменными. Эта гипотеза утверждает, что «потребление (потребление) C_t^* домохозяйства пропорционально «постоянному доходу» Y_t^* :

$$C_t^* = kY_t^*, \quad 0 < k < 1. \quad (3.2.3)$$

Предполагается, что измеренное потребление C_t и измеренный доход Y_t являются значениями постоянного дохода и потребления, измеренными с ошибкой:

$$C_t = C_t^* + e_t \quad \text{и} \quad Y_t = Y_t^* + y_t. \quad (3.2.4)$$

Предполагается, что ошибки измерения e_t и y_t имеют нулевое среднее и не коррелированы с постоянным потреблением и доходом:

$$E(e_t) = 0, \quad E(y_t) = 0, \quad E(e_t y_t) = 0, \quad (3.2.5)$$

$$E(C_t^* e_t) = 0, \quad E(Y_t^* y_t) = 0, \quad E(C_t^* y_t) = 0, \quad E(Y_t^* e_t) = 0. \quad (3.2.6)$$

Подстановка (3.2.4) в (3.2.3) указанное соотношение можно переписать в терминах измеренных потребления и доходов:

$$C_t = kY_t + (1-k)e_t + y_t, \quad 0 < k < 1. \quad (3.2.7)$$

Данный пример отличается от предыдущих тем, что уравнение не содержит постоянной составляющей. Поэтому для определения того, является ли переменная предопределенной, следует использовать смешанный момент $E(Y_t y_t)$, а не ковариацию $\text{Cov}(Y_t, y_t)$. Из (3.2.4)–(3.2.7) нетрудно вывести, что

$$E(Y_t y_t) = -k E(y_t^2) < 0.$$

Таким образом, измеренный доход является в функции потребления (3.2.7). В отличие от предыдущих примеров, эта эндогенность является следствием ошибок измерения. С учетом того факта, что оценка МНК коэффициента при Y_t в (3.2.7) соответствует коэффициенту проекции наименьших квадратов $E(C_t Y_t) / E(Y_t^2)$, из (3.2.4)–(3.2.7) можно также получить, что

$$\text{plim } \tilde{\beta}_{OLS} = \frac{k E(y_t^2)^2}{E[(Y_t^2)^2] + E(y_t^2)} < k. \quad (3.2.8)$$

То есть оценка регрессии измеренного потребления на измеренный доход (без постоянной составляющей) дает заниженную оценку k . Фридман использовал эти результаты для объяснения того, почему оценка МРС, полученная с помощью кросс-секционных регрессии потребления на доход, меньше аналогичной оценки МРС, полученной на основе наблюдений агрегированных временных рядов.

Предположим на мгновение, что существует релевантный инструмент x_i , так что $E(x_i a_i) = 0$ и $E(x_i Y_i) \neq 0$. Аналогично выводу (3.1.11) можно получить, что

$$k = \frac{E(x_i C_i)}{E(x_i Y_i)}. \quad (3.2.9)$$

Поэтому состоятельной IV-оценкой будет

$$\hat{k}_{IV} = \frac{\text{выборочное среднее } x_i C_i}{\text{выборочное среднее } x_i Y_i}. \quad (3.2.10)$$

Но существует ли релевантный инструмент? Да, и это постоянная величина. Подставляя $x_i = 1$ в (3.2.10), мы получим состоятельную оценку для k , являющуюся отношением выборочного среднего измеренного потребления к выборочному среднему измеренного дохода. Именно так Фридман и оценивал k .

Производственная функция

Во многих ситуациях, случайная компонента модели включает в себя факторы, которые известны рассматриваемому экономическому агенту, но не наблюдаются эконометристом. Эндогенность возникает, когда регрессоры являются решениями, принимаемыми агентом на основе этих факторов. Рассмотрим кросс-секционную выборку компаний, выбирающих объем потребляемого труда для максимизации прибыли. Производственная функция фирмы i

$$Q_i = A_i \cdot (L_i)^{\phi_1} \cdot \exp(\eta_i), \quad 0 < \phi_1 < 1. \quad (3.2.11)$$

где Q_i — выпуск фирмы, L_i — потребляемый труд, A_i — известный фирме уровень ее эффективности, η_i — технологический шок. В отличие от A_i , η_i не наблюдаем для фирмы в момент выбора объема труда L_i . Ни A_i , ни η_i не наблюдаемы для эконометриста и независимы друг от друга.

Предположим, что для каждой фирмы i , сериально независим. Таким образом, $B = E[\exp(\eta_i)]$ одинаково для всех фирм¹, и ожидаемый фирмой

¹ Если η_i для фирмы i коррелированы во времени, фирма может использовать прошлые значения η_i для оценки текущего значения η_i , таким образом, B будет различным для разных фирм. Также поскольку математическое ожидание значения нелинейной функции от случайной величины с нулевым математическим ожиданием необязательно равно нулю, то B не обязан быть равным нулю, даже если $E(\eta_i) = 0$.

уровень выпуска равен:

$$A_i \cdot (L_i)^{\alpha} \cdot B.$$

Пусть p — цена продукции, w — уровень заработной платы. Для упрощения предположим, что все фирмы из одной и той же конкурентной отрасли, то есть p и w постоянны для всех фирм. Цель фирмы — выбрать объем L_i , максимизирующий ожидаемую прибыль:

$$p \cdot A_i \cdot (L_i)^{\alpha} \cdot B - w \cdot L_i.$$

Взяв производную этого выражения по L_i , приравняв ее к нулю и решая полученное уравнение относительно L_i , получим оптимальный объем затрат труда L_i :

$$L_i = \left(\frac{w}{p}\right)^{\frac{1}{1-\alpha}} (A_i B \phi_i)^{\frac{1}{1-\alpha}}. \quad (3.2.12)$$

Пусть α_i — отклонение фирмы i от среднего логарифма эффективности: $\alpha_i = \log(A_i) - E[\log(A_i)]$, а $\phi_i = E[\log(A_i)]$ (так что $E(\alpha_i) = 0$ и $A_i = \exp(\phi_i + \alpha_i)$). Тогда производственную функцию (3.2.11) и функцию спроса на труд (3.2.12) можно записать в логарифмическом виде

$$\log(y_i) = \phi_i + \alpha_i \log(L_i) + (\alpha_i + u_i), \quad (3.2.13)$$

$$\log(L_i) = \beta_0 + \frac{1}{1-\alpha_i} u_i, \quad (3.2.14)$$

где

$$\beta_0 = \frac{1}{\alpha_i - 1} [\log(w/p) - \phi_i - \log(\phi_i B)].$$

Поскольку для всех фирм цены одинаковы, β_0 постоянна для всех фирм. Формула (3.2.14) показывает, что в логлинейной производственной функции (3.2.12) $\log(L_i)$ — эндогенный регрессор, положительно коррелированный с ошибкой $(\alpha_i + u_i)$ через u_i . Таким образом, МНК-оценка ϕ при оценивании логлинейной производственной функции смешивает вклад в выпуск u_i со вкладом труда. Этот пример показывает еще один источник экзогенности — учет факторов, которые не наблюдаются экономистом, но оказываются значимым, принимающим решение.

Контрольные вопросы

1. Рассмотрите регрессию C_i на константу и 1, в рамках гипотезы постоянного дохода Фридриха. Выведите предел по вероятности МНК-оценку коэффициента при Y_i в этой модели. Указание: Предел по вероятности равен отношению $\text{Cov}(C_i, Y_i)$ к $\text{Var}(Y_i)$. Покажите, что оно равно

$$\frac{k \text{Var}(Y_i)}{\text{Var}(Y_i^*) + \text{Var}(y_i)}.$$

2. В примере с производственной функцией покажите, что $\lim_{n \rightarrow \infty} \hat{\beta}_{1,OLS} = 1$, где $\hat{\beta}_{1,OLS}$ — МНК-оценка β_1 на (3.2.13). **Указание:** Исключите α_1 из уравнения (3.2.13) с помощью функции строга на труа (3.2.14).
3. В примере с производственной функцией предположим, что перед принятием решения о затратах труда фирма может наблюдать π_{it} и α_{it} . Как изменится функция строга на труа (3.2.12)? Покажите, что $\log(Q_{it})$ и $\log(L_{it})$ абсолютно коррелированы. **Указание:** $\log(Q_{it})$ будет точной линейной функцией от $\log(L_{it})$ без случайных ошибок.

3.3. Общая постановка задачи

Дадим теперь общую постановку задачи. Модель описывается следующим набором предположений и является обобщением модели из главы 2.

Регрессоры и инструменты

Предположение 3.1 (линейность): *Оцениваемое уравнение линейно:*

$$y_i = z_i' \delta + \varepsilon_i \quad (i = 1, 2, \dots, n),$$

где z_i — L -мерный вектор регрессоров, δ — L -мерный вектор коэффициентов, а ε_i — невозмущаемая ошибка.

Предположение 3.2 (эргодическая стационарность):

Пусть x_i — K -мерный вектор, рассматриваемый как вектор инструмента, и пусть w_i — уникальные (различные) и непостоянные элементы $\{w_i, z_i, x_i\}$. $\{w_i\}$ совместно стационарен и эргодичен.

Предположение 3.3 (условия ортогональности): *Все K переменных x и w_i «предопределены» (predetermined) в том смысле, что они ортогональны текущей ошибке: $E(x_{ik} \varepsilon_i) = 0$ для всех i и k ($k = 1, 2, \dots, K$)².*

Это может быть записано как

$$E[x_i' (y_i - z_i' \delta)] = 0 \quad \text{или} \quad E(y_i) = 0,$$

где $y_i \equiv x_i + \varepsilon_i$.

¹ Далее будут применены понятия *то*.

² Как было отмечено в параграфе 2.3, такое использование термина «предопределенность» может быть не совсем строгим для части прогнозных. Это не имеет значения, — это ортогональность текущих значений ошибок и текущих значений регрессоров. Мы не требуем ортогональности текущего значения ошибок и прошлых значений регрессоров.

Мы отметили в параграфе 2.3, что если в состав регрессоров входит константа, то условие ортогональности эквивалентно условию $E(\varepsilon_i) = 0$ и что регрессоры не коррелированы с ошибкой. Теперь условия ортогональности касаются инструментальных переменных. То есть если один из инструментов — константа, то предположение 3.3 эквивалентно тому, что $E(z_{is}) = 0$; что непостоянные инструменты не коррелированы с ошибкой.

Расширенные в двух предыдущих параграфах примеры могут быть переписаны в этом общем виде. Например,

Пример 3.1 (пример Уоркинга с наблюдаемым сдвигом предложения): Рассмотрим модель рыночного равновесия (3.1.2а) и (3.1.2б). Предположим, что оцениваемое уравнение — уравнение спроса. Пример можно привести к общему виду, глася:

$$y_i = \alpha_0 + \alpha_1 x_i + \varepsilon_i, \quad L = 1, \quad \delta = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}, \quad z_i = \begin{bmatrix} 1 \\ x_i \end{bmatrix}, \quad \varepsilon_i = \varepsilon_i, \quad K = 2, \quad x_i = \begin{bmatrix} 1 \\ x_i \end{bmatrix}$$

и т.д. $(y_i, z_i, x_i)'$. Поскольку математическое ожидание ошибки равно нулю, константа ортогональна ошибке и может быть включена в z_i . Также получаем, что x_i , которое, по предположению, не коррелирует с ошибкой, удовлетворяет условию ортогональности $E(x_i \varepsilon_i) = 0$. Поэтому оно тоже может быть включено в z_i .

Другие рассмотренные примеры также могут быть записаны как частные случаи общей модели.

Два отдельных символа, z_i и x_i , могут создать впечатление, что регрессоры и инструменты не содержат одних и тех же переменных, но это не всегда так. Более того, в приведенном примере регрессоры и инструменты содержат общую переменную — константу. Инструменты, являющиеся также и регрессорами, называются **предопределенными регрессорами** (*predetermined regressors*), а остальные регрессоры, которые не включены в z_i , — **эндогенными регрессорами** (*endogenous regressors*). Характерным примером вышесказанного является

Пример 3.2 (уравнение заработной платы): Упрощенная версия уравнения заработной платы, которое будет оцениваться позднее в этой главе,

$$\ln W_i = \beta_0 + \beta_1 S_i + \beta_2 EXPR_i + \beta_3 IQ_i + \varepsilon_i,$$

где $\ln W_i$ — логарифм заработной платы индивида i , S_i — количество законченных лет обучения, $EXPR_i$ — количество лет опыта, IQ_i — уровень IQ, ε_i — ненаблюдаемые характеристики индивида, влияющие на уровень заработной платы. Мы предположим, что $E(\varepsilon_i) = 0$ (в простейшем случае, конечно, математическое ожидание ε_i в β_0). В одной из следующих глав, которую мы отложим позднее, предположим, что β_2 является

предопределенной переменной, но IQ_i является эндогенной переменной из-за ошибок измерения. Мы также предполагаем, что $EXPR_i$, AGE_i (возраст индивида) и MED_i (продолжительность образования матери в годах) предопределены. Возраст исключает из уравнения заработной платы, отражая влияние и ошибок предположения о том, чем при контроле стажа возраст не будет влиять на заработную плату. В терминах общей модели:

$$y = LW, \quad L = 1, \quad x_1 = \begin{bmatrix} 1 \\ S_i \\ (EXPR_i) \\ IQ_i \end{bmatrix}, \quad K = 3, \quad x_2 = \begin{bmatrix} 1 \\ S_i \\ EXPR_i \\ AGE_i \\ MED_i \end{bmatrix}$$

и $w = (W, S_i, S_i, EXPR_i, IQ_i, AGE_i, MED_i)'$. Как и в примере 3.1, мы можем исключить константу в x_1 , так как $E(x_1) = 0$. В этом примере x_2 и x_1 включают три общие переменные ($1, S_i, EXPR_i$). Если бы, например, S_i не была бы включена в x_2 , не можно было бы рассмотреть как эндогенную.

Если, как в этом примере, некоторые регрессоры x_2 являются предопределенными, их следует включать в число инструментов x_1 . GMM-оценки вектора параметров использует информацию, предоставленную условиями ортогональности. Игнорирование предопределенных регрессоров в число инструментов отображает условия ортогональности, которые могли бы быть использованы.

Идентификация

Как видно из примеров предыдущих двух глав, идеальным инструментом должен быть не только предопределенным (ортогональным к ошибке), но также коррелированным с регрессорами. В противном случае IV-оценка не может быть определена (см., например, 18.1–1.1). Обобщением на случай более чем одного регрессора и более одной предопределенной переменной является

Предположение 3.4 (ранговое условие идентификации): $E(x_1 x_1')$, матрица размера $K \times K$, имеет полный ранг (то есть ее ранг равен K , число столбцов). Мы обозначаем эту матрицу как Σ_{xx} .

Чтобы показать, что это действительно обобщение, рассмотрим пример 3.1. Если $x_1 = (1, p_1)'$, $x_2 = (1, x_1)'$, то матрица Σ_{xx} равна:

$$\Sigma_{xx} = \begin{bmatrix} 1 & E(p_1) \\ E(p_1) & E(p_1 p_1) \end{bmatrix}$$

¹ То есть σ_y должен и как скаляр, и как вектор. Если σ_y является скаляром, то условие предопределяется, то есть момент существует и конечен.

Определитель этой матрицы отличен от нуля (матрица имеет полный столбчатый ранг) тогда и только тогда, когда $\text{Cov}(x_i, p_i) = E(x_i p_i) - \bar{x}_i \bar{p}_i \neq 0$.

Предположение 3.4 называется **ранговым условием идентификации** (*rank condition for identification*) по следующей причине. Чтобы подчеркнуть зависимость g_i ($\equiv x_i z_i$) от данных и вектора параметров, перепишем его как

$$g_i = g(x_i; \delta) \equiv x_i \cdot (y_i - x_i' \delta). \quad (3.3.1)$$

Тогда условие ортогональности можно переписать в виде:

$$E[g(x_i; \delta)] = \mathbf{0}_{(K \times 1)}. \quad (3.3.2)$$

Пусть $\tilde{\delta}$ ($L \times 1$) — гипотетическое значение δ . Рассмотрим систему K одновременных уравнений от L неизвестных (L элементов $\tilde{\delta}$):

$$E[g(x_i; \tilde{\delta})] = \mathbf{0}_{(K \times 1)}. \quad (3.3.3)$$

Условие ортогональности (3.3.2) означает, что истинное значение вектора коэффициентов δ является решением этой системы K одновременных уравнений (3.3.3). Поэтому сделанные нами предположения гарантируют, что решение системы одновременных уравнений существует. Мы говорим, что вектор коэффициентов (или уравнение) **идентифицируем** (*identified*), если $\tilde{\delta} = \delta$ — единственное решение.

Поскольку оцениваемое уравнение в нашей модели линейно, то функция $g(x_i; \tilde{\delta})$ линейна по $\tilde{\delta}$, так как она может быть записана в виде $x_i \cdot y_i - x_i x_i' \tilde{\delta}$. Поэтому (3.3.3) — это система из K линейных уравнений:

$$E(x_i \cdot y_i) - E(x_i x_i') \tilde{\delta} = 0 \quad \text{или} \quad \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i x_i' \tilde{\delta} = \sum_{i=1}^N x_i y_i, \quad (3.3.4)$$

где

$$\sigma_{xy} \equiv E(x_i \cdot y_i), \quad \Sigma_{xx} \equiv E(x_i x_i').$$

Необходимое и достаточное условие того, что $\tilde{\delta} = \delta$ является единственным решением этой системы, можно вывести из следующего результата матричной алгебры¹.

Предположим, что существует решение системы линейных одновременных уравнений относительно x : $Ax = b$. Необходимым и достаточным условием того, что это решение единственно, является полный столбчатый ранг матрицы A .

¹ См., например: [Sargol, 1983 pp. 235-234].

Таким образом, $\hat{\beta} = \beta$ — единственное решение (3.3.4) тогда и только тогда, когда матрица Σ_{xx} имеет полный столбцовый ранг, что и позволяет предположение 3.4.

Порядковое условие идентифицируемости

Поскольку $\text{rank}(\Sigma_{xx}) \leq L$, если $K < L$, то необходимым условием идентифицируемости будет

$$K (= \# \text{предопределенных переменных}) \geq L (= \# \text{регрессоров}). \quad (3.3.5a)$$

Это называется **порядковым условием идентифицируемости** (*order condition for identification*). Это можно трактовать по-разному. Поскольку K является также числом условий нормальности, а L — числом параметров, то порядковое условие можно считать эквивалентно как

$$\# \text{условий нормальности} \geq \# \text{параметров}. \quad (3.3.5b)$$

В частном случае для предопределенных регрессоров в обеих строках неравенства получим другое эквивалентное утверждение.

$$\begin{aligned} \# \text{предопределенных переменных, не включенных в уравнение} &\geq \\ &\geq \# \text{включенных регрессоров}. \end{aligned} \quad (3.3.5c)$$

В зависимости от выполнения порядкового условия мы получим, что уравнение

- **сверхидентифицируемо** (*overidentified*), если порядковое условие выполнено и $K > L$,
- **точно идентифицируемо** (*just identified*), если порядковое условие выполнено и $K = L$,
- **неидентифицируемо** (*underidentified*), если порядковое условие не выполняется (имеет $K < L$).

Поскольку порядковое условие является необходимым, его невыполнение означает, что уравнение неидентифицируемо.

Предположение асимптотической нормальности

Как и в главе 2, нам нужно усилить условия ортогональности, чтобы оценки были асимптотически нормальными.

Предположение 3.5 (g_t — мартингал-разность с конечными моментами второго порядка): Пусть $y_t \equiv x_t \cdot \varepsilon_t$, $\{g_t\}$ — мартингал-разность (так что $E(g_t) = 0$). Матрица смешанных моментов размера $K \times K$, $E(g_t g_t')$, невырождена. Используем обозначение S для $\text{Avar}(\bar{g})$ (то есть дисперсии, определенной соотношением $\sqrt{n} \bar{g}$, где $\bar{g} = \frac{1}{n} \sum_{t=1}^n g_t$). По предположению 3.2 и ЦИТ для стационарных эргодических мартингал-разностей, $S = E(g_t g_t')$.

Это аналог предположения 2.3, к которому применимы те же примечания:

- Если инструменты включают в себя константу, то из этого предположения вытекает, что ошибка является мартингал-разностью (и тем самым, строгой коррелирована).
- Достаточным и, вероятно, более простым для понимания условием для предположения 3.5 является

$$E(\varepsilon_t | \varepsilon_{t-1}, \dots, \varepsilon_1, x_t, x_{t-1}, \dots, x_1) = 0, \quad (3.3.6)$$

означает, что ошибка не просто является мартингал-разностью, но также ортогональна не только текущему, но и прошлым значениям и инструментам.

- Поскольку $g_t g_t' = \varepsilon_t^2 x_t x_t'$, S является матрицей четвертых моментов. Состоятельное ссылавание S потребует условия на моменты четвертого порядка, которое сформулировано ниже в предположении 3.6.
- Если $\{g_t\}$ строго коррелирована, то S (определенная как $\text{Avar}(\bar{g})$) не равна $E(g_t g_t')$ и имеет более сложную форму, что будет доказано в главе 6.

Контрольные вопросы

1. Является ли уравнение спроса в примере Зоркяна (пример 3.1) идентифицируемым? Сверхидентифицируемым? Идентифицируемо ли уравнение предложения?
2. Пусть рядовое условие удовлетворяется для уравнения заработной платы в примере 3.2. Является ли это уравнение сверхидентифицируемым?
3. В примере с произвольной функцией не определено никаких инструментов помимо константы. Поскольку $K = 1$ и $L = 2$, Порядковое условие говорит, что уравнение неидентифицируемо. Запишите условие ортогональности и удостоверьтесь, что существует бесконечно много пар (ϕ_1, ϕ_2) , удовлетворяющих условию ортогональности.

4. Проверьте, что примеры из параграфа 3.2 являются частными случаями общей модели этого параграфа, определив (g_1, g_2, g_3) и выписав ранговое условие для каждого примера.
5. Покажите, что из указанной модели следует, что

$$\text{rank}(\Sigma_{xx}) = \text{rank}(\Sigma_{xx}; \pi_{xy}).$$

Указание: $\delta = 0$ — решение (3.3.4). (Некоторые учебники добавляют это уравнение в ранговое условие, однако оно вытекает из других предположений модели.)

6. (Перевисанные инструменты.) Добавьте к z еще одну переменную, назовите ее ξ . Хитрость в придумывании, она не связана с регрессорами в том смысле, что $E(\xi_i z_i) = 0$ для всех $i (= 1, 2, \dots, L)$. Продолжает ли выполняться ранговое условие? Указание: Если матрица размером $K \times L$ имеет полный столбцовый ранг, то добавление любой L -мерной строки к строкам матрицы не приводит к снижению ранга.
7. (Линейно зависящие инструменты.) В примере 3.2 предположим, что $ACE_i = -E_i X_i R_i - 5$ для всех наблюдений. Означает ли это обязательно невыполнение рангового условия? [Ответ: Нет.] Выполняется ли условие полного ранга в предположении 3.5 (то, что матрица $E(g_i g_i')$ невырождена)? Указание: Существует K -мерный вектор $\alpha \neq 0$ такой, что $\alpha' g_i = 0$. Покажите, что $\alpha' E(g_i g_i') = 0$.
8. (Тонкая комбинация инструментов.) Пусть A — матрица размера $q \times K$, имеющая полный строковый ранг (то есть $q \leq K$), такая, что $A \Sigma_{xx}$ имеет полный столбцовый ранг. Пусть $\tilde{z} = Az$ (то есть $\tilde{z}$ — это вектор из q преобразованных инструментов). Проверьте, что предположения 3.3–3.5 выполняются для $\tilde{z}$, если они выполняются для z .
9. Проверьте, что модель, построенная на предположениях 3.1–3.5, редуцируется к модели из главы 2, если $z_i = x_i$.

3.4. Обобщенный метод моментов (GMM)

Условия ортогональности утверждают, что все моменты из некоторого набора моментов генеральной совокупности равны нулю. Основной принцип «метода моментов» (*method of moments*) заключается в том, чтобы найти такую оценку параметра, при которой соответствующие выборочные моменты также были бы равны нулю. Моменты генеральной совокупности в условиях ортогональности — это $E(g_i(w_i; \delta))$. Их выборочные аналоги — это выборочные средние $\bar{g}(w_i; \delta)$ (где $\bar{g}(w_i; \delta)$ определено в (3.3.1)), вычисленные при некотором гипотетическом значении δ :

$$\bar{g}_n(\delta) \equiv \frac{1}{n} \sum_{i=1}^n g(w_i; \delta), \quad (3.4.1)$$

Применение метода моментов к нашей модели сводится к выбору $\tilde{\delta}$, являющейся решением системы из K одновременных уравнений с L неизвестными: $g_n(\tilde{\delta}) = 0$, которая является выборочным аналогом (3.3.3). Поскольку оцениваемое уравнение линейно, $g_n(\tilde{\delta})$ можно переписать как

$$\begin{aligned} g_n(\tilde{\delta}) &= \frac{1}{n} \sum_{i=1}^n x_i \cdot (y_i - z_i' \tilde{\delta}) = \quad (\text{из выражения (3.3.3) для } g(w_i; \delta)) \\ &= \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i - \left(\frac{1}{n} \sum_{i=1}^n x_i z_i' \right) \tilde{\delta} = s_{xy} - s_{xz} \tilde{\delta}, \end{aligned} \quad (3.4.2)$$

где s_{xy} и s_{xz} являются выборочными аналогами σ_{xy} и σ_{xz} :

$$s_{xy} \equiv \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i \quad \text{и} \quad s_{xz} \equiv \frac{1}{n} \sum_{i=1}^n x_i z_i',$$

Таким образом, выборочный аналог $g_n(\tilde{\delta}) = 0$ — это система K уравнений с L неизвестными:

$$s_{xz} \tilde{\delta} = s_{xy}. \quad (3.4.3)$$

Это выборочный аналог (3.3.4). Если условий ортогональности больше, чем количество параметров, система может не иметь решений. Обобщение метода моментов на такие случаи называется **обобщенным методом моментов (GMM) (generalized method of moments)**.

Метод моментов

Если уравнение идентифицируемо точно, то тогда $K = L$ и Σ_{xz} является квадратной и обратимой. Поскольку, по предположению 3.2, S_{xz} сходится к Σ_{xz} почти наверняка, S_{xz} обратима для достаточно большого объема выборки с вероятностью 1. Таким образом, при достаточно большом объеме выборки система одновременных уравнений (3.4.3) имеет единственное решение:

$$\tilde{\delta}_{IV} = s_{xz}^{-1} s_{xy} = \left(\frac{1}{n} \sum_{i=1}^n x_i z_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n x_i y_i. \quad (3.4.4)$$

Эта оценка также называется **оценкой с инструментальными переменными (instrumental variables estimator)** с z_i в качестве инструментов. Поскольку она определена в точно идентифицируемом случае, формула предполагает, что количество инструментов равно количеству регрессоров. Более того, если $z_i = x_i$ (все регрессоры предопределены или ортогональны ошибке регрессии), то $\tilde{\delta}_{IV}$ сходится к МНК-оценке. Таким образом, МНК-оценка является оценкой по методу моментов.

Обобщенный метод моментов

Если уравнение сводится к линейному, то есть $K = 1$, мы не можем в общем случае подобрать 1-мерный вектор $\tilde{\delta}$, удовлетворяющий K уравнениям (3.4.3). Если мы не можем сделать $g_n(\tilde{\delta})$ в точности равным 0, то по крайней мере мы можем подобрать $\tilde{\delta}$ так, что $g_n(\tilde{\delta})$ будет максимально близким к 0. Чтобы уточнить, что мы понимаем здесь под «близостью», определим расстояние между любыми двумя K -мерными векторами ξ и η в виде квадратичной формы $(\xi - \eta)' \tilde{W} (\xi - \eta)$, где $\tilde{W}$, иногда называемая *взвешивающей матрицей* (*weighting matrix*), симметричная и положительно определенная матрица, определяющая расстояние¹.

Определение 3.1 (GMM-оценка): Пусть $\tilde{W}$ — симметричная положительно определенная матрица размера $K \times K$, возможно, зависящая от выборки, такая, что $\tilde{W} \rightarrow_p W$ при стремлении объема выборки n к бесконечности, где W — симметричная и положительно определенная матрица. «GMM-оценка» (*GMM estimators*) $\tilde{\delta}$, обозначаемая $\tilde{\delta}(\tilde{W})$, — это

$$\tilde{\delta}(\tilde{W}) = \underset{\tilde{\delta}}{\operatorname{argmin}} J(\tilde{\delta}, \tilde{W}). \quad (3.4.5)$$

где

$$J(\tilde{\delta}, \tilde{W}) \equiv n \cdot g_n(\tilde{\delta})' \tilde{W} g_n(\tilde{\delta}).$$

(Причина, по которой расстояние $g_n(\tilde{\delta})' \tilde{W} g_n(\tilde{\delta})$ умножается на объем выборки n , будет объяснена в параграфе 3.6.) Взвешивающая матрица может быть случайной и зависящей от выборки, чтобы учесть возможность оценивания этой матрицы по выборке. Определение показывает, что GMM является частным случаем «метода минимального расстояния» (*minimum distance estimation*). В методе оценки минимального расстояния $\rho(\tilde{\delta}) \equiv g_n(\tilde{\delta}) = 0$, как и в GMM, но функция $g_n(\cdot)$ не обязана быть выборочными средними.

Поскольку $g_n(\tilde{\delta})$ линейна по $\tilde{\delta}$ (см. (3.4.2)), целевая функция квадратична по $\tilde{\delta}$, когда уравнение линейно:

$$J(\tilde{\delta}, \tilde{W}) = n \cdot (a_{xy} - S_{xz} \tilde{\delta})' \tilde{W} (a_{xy} - S_{xz} \tilde{\delta}). \quad (3.4.6)$$

Вы можете сами показать, что условием первого порядка для минимизации этой целевой функции относительно $\tilde{\delta}$ является

$$\begin{matrix} S'_{xz} & \tilde{W} & a_{xy} \\ (L \times K) & (K \times K) & (K \times 1) \end{matrix} = \begin{matrix} S'_{xz} & \tilde{W} & S_{xz} \\ (L \times K) & (K \times K) & (K \times L) \end{matrix} \begin{matrix} \tilde{\delta} \\ \\ (L \times 1) \end{matrix}. \quad (3.4.7)$$

¹ Не следует путать «взвешивание» в GMM и во взвешенном методе наименьших квадратов (WLS). В GMM взвешивание применяется к выборочному среднему $\tilde{g}$, и то время как в WLS оно применяется к каждому наблюдению.

При выполнении предположений 3.2 и 3.4 матрица S_{xx} имеет полный столбцовый ранг для достаточно больших n с вероятностью 1. Поскольку $\tilde{W}$ положительно определена, матрица $S'_{xx}\tilde{W}S_{xx}$ размера $L \times L$ невырождена. Поэтому единственное решение можно получить, умножив обе части на матрицу, обратную к $S'_{xx}\tilde{W}S_{xx}$. Это единственное решение является GMM-оценкой:

$$\text{GMM-оценка: } \hat{\delta}(\tilde{W}) = (S'_{xx}\tilde{W}S_{xx})^{-1}S'_{xx}\tilde{W}y_{\text{exp}}. \quad (3.4.8)$$

Если $K = L$, то S_{xx} является квадратной матрицей и (3.4.8) сводится к IV-оценке (3.4.4). Тем самым GMM является обобщением метода моментов.

Ошибки оценки

Для дальнейшего использования выведем выражение для ошибки оценки. Умножив обе стороны уравнения $y_i = x'_i\delta + \varepsilon_i$ слева на w_i и взяв средние, получим:

$$y_{\text{exp}} = S_{xx}\delta + \bar{y}, \quad (3.4.9)$$

где

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n w_i y_i = \frac{1}{n} \sum_{i=1}^n g(w_i; \delta) = g_*(\delta). \quad (3.4.10)$$

Подставляя (3.4.9) в (3.4.8), получаем:

$$\hat{\delta}(\tilde{W}) - \delta = (S'_{xx}\tilde{W}S_{xx})^{-1}S'_{xx}\tilde{W}\bar{y}. \quad (3.4.11)$$

Контрольные вопросы

1. Проверьте, что (3.1.12) это IV-оценка α для приложения метода моментов к уравнению спроса (3.1.2a) с (1.1.1₁) в качестве инструментов.
2. Если уравнение идентифицируемо слабо, то какое минимальное значение $J(\hat{\delta}, \tilde{W})$?
3. Даже если уравнение сверхидентифицируемо, теоретический вариант системы K уравнений (3.3.4) может не иметь решения. Почему не имеет решения выборочный вариант системы (3.4.3)? Укажите: Теоретический вариант имеет решение, так как матрица $(\Sigma_{xx}|\Sigma_{xy})$ имеет ранг L . Ранг L y доминирует над x и представляет собой набор равенств отрицательных элементов Σ_{xx} и Σ_{xy} . Даже если S_{xx} и y_{xy} сходится к Σ_{xx} и Σ_{xy} , это не означает, что ранг $(S_{xx}'y_{xy})$ равен L для достаточно большого n . Напротив, S_{xx} имеет полный столбцовый ранг для достаточно большого n , если во пределе, Σ_{xx} имеет полный столбцовый ранг. Это связано с тем, что ранговое условие для матрицы является набором неравенств для элементов этой матрицы.

4. Что не так со следующим утверждением?

Даже если уравнение сверидентифицируемо, гайта решение (3.4.3) можно без проблем. Следует умножить обе стороны уравнения на S'_{xx} и получить

$$S'_{xx} S_{xx} \delta = S'_{xx} y_{nx}. \quad (3.4.12)$$

Поскольку S_{xx} имеет полный столбцовый ранг, $S'_{xx} S_{xx}$ обратима. Поэтому решением является

$$\hat{\delta} = (S'_{xx} S_{xx})^{-1} S'_{xx} y_{nx}.$$

Указание: Это $\hat{\delta}$ опрелеленно является решением (3.4.12). Но является ли для решением (3.4.3)?

- б. (Выраженная $\tilde{W}$.) Проверьте, что GMM-оценка (3.4.8) определена для достаточно большой n , даже если $\tilde{W}$ ($\equiv \rho(n) \tilde{W}$) вырождена, если не выражена $S'_{xx} \tilde{W} S_{xx}$.

3.5. Асимптотические свойства GMM

Формула (3.4.8) определяет GMM-оценки, которые образуют множество оценок, индексированных взвешивающей матрицей $\tilde{W}$. Следует отметить, что все оценки, представленные в предыдущих главах, являются GMM-оценками для некоторых $\tilde{W}$. Задачей данного параграфа является рассмотрение асимптотической теории для GMM-оценок для любого заданного выбора $\tilde{W}$, которая может быть проведена достаточно легко с использованием методов предыдущей главы. Первая половина данного параграфа рассматривает утверждения 2.1–2.4 из главы 2 для GMM-оценок, которые предпочтительнее других GMM-оценок. Вопрос, который не возникал в этих утверждениях, — это вопрос выбора подходящей GMM-оценки, которая предпочтительнее других GMM-оценок. Это вопрос оптимального выбора $\tilde{W}$, и он будет рассмотрен во второй половине данного параграфа.

Асимптотическое распределение GMM-оценок

Асимптотическая теория для $\hat{\delta}(\tilde{W})$, справедливая для любого выбора $\tilde{W}$, такова:

Утверждение 3.1 (асимптотическое распределение оценок GMM):

- (a) (Состоятельность.) При предположениях 3.1–3.4, $\lim_{n \rightarrow \infty} \hat{\delta}(\tilde{W}) = \delta$.
 (b) (Асимптотическая нормальность.) Если предположение 3.3 усилено как утверждение 3.5, то

$$\sqrt{n}(\hat{\delta}(\tilde{W}) - \delta) \xrightarrow{d} N(0, \text{Avar}(\hat{\delta}(\tilde{W}))) \quad \text{при } n \rightarrow \infty.$$

720

$$\text{Avar}(\hat{\delta}(\tilde{W})) = (\Sigma'_{xx} W \Sigma_{xx})^{-1} \Sigma'_{xx} W S W \Sigma_{xx} (\Sigma'_{xx} W \Sigma_{xx})^{-1}. \quad (3.5.1)$$

(Напомним: $\Sigma_{xx} \equiv E(x_i x_i')$, $S \equiv E(g_i g_i') = E(z_i^2 x_i x_i')$, $W \equiv \text{plim } \tilde{W}$.)

(с) (Состоятельная оценка $\text{Avar}(\hat{\delta}(\tilde{W}))$.) Предположим, что доступна состоятельная оценка S ($K \times K$), $\tilde{S}$. Тогда, при предположении 3.2, состоятельная оценкой $\text{Avar}(\hat{\delta}(\tilde{W}))$ является

$$\widehat{\text{Avar}}(\hat{\delta}(\tilde{W})) = (S'_{xx} \tilde{W} S_{xx})^{-1} S'_{xx} \tilde{W} \tilde{S} \tilde{W} S_{xx} (S'_{xx} \tilde{W} S_{xx})^{-1}. \quad (3.5.2)$$

где S_{xx} — выборочное среднее $x_i x_i'$:

$$S_{xx} \equiv \frac{1}{n} \sum_{i=1}^n x_i x_i'.$$

Не самое красивое выражение для асимптотической дисперсии станет гораздо приятнее при оптимальном выборе взвешивающей матрицы. Доказательство утверждения 2.1.1 может доказать утверждение 3.1. Ключевыми компонентами являются:

1. $S_{xx} \rightarrow_p \Sigma_{xx}$ (согласно свойству эргодической стационарности).
2. $\bar{g}(\bar{z} \equiv \frac{1}{n} \sum_{i=1}^n z_i) \rightarrow_p 0$ (согласно свойству эргодической стационарности к условиям ортогональности).
3. $\sqrt{n} \bar{g} \rightarrow_p N(0, S)$ (по предположению 3.5).

Состоятельность сразу получается в результате применения пунктов 1 и 2 к лемме 2.3(a) к выражению для ошибки оценки (3.4.11). Для доказательства асимптотической нормальности следует умножить обе части (3.4.11) на $\sqrt{n}$.

$$\sqrt{n}(\hat{\delta}(\tilde{W}) - \delta) = (S'_{xx} \tilde{W} S_{xx})^{-1} S'_{xx} \tilde{W} \sqrt{n} \bar{g} \quad (3.5.3)$$

и использовать пункт 3, леммы 2.3(a) и 2.4(c). Часть (c) утверждения 3.1 непосредственно вытекает из леммы 2.3(a).

Оценочная дисперсия ошибки

В параграфе 2.3 была доказана состоятельность МНК-оценки дисперсии ошибки. Было отмечено, что этот результат сохраняется, если остаться только на основании состоятельной оценки вектора коэффициентов. То же самое справедливо и здесь.

Утверждение 3.2 (состоятельная оценка дисперсии ошибки): Для любой состоятельной оценки $\hat{\beta}$, $\hat{\delta}$, определим $\hat{\varepsilon}_i \equiv y_i - \mathbf{z}_i' \hat{\beta}$. При предположениях 3.1, 3.2 и дополнительном предположении, что $E(\mathbf{z}_i \mathbf{z}_i')$ (вторые моменты регрессоров) существуют и конечны,

$$\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 \xrightarrow{p} E(\varepsilon_i^2),$$

если $E(\varepsilon_i^2)$ существует и конечен.

Доказательство очень похоже на доказательство утверждения 2.2. Связь между $\hat{\varepsilon}_i$ и ε_i дается формулой

$$\hat{\varepsilon}_i \equiv y_i - \mathbf{z}_i' \hat{\beta} = \varepsilon_i - \mathbf{z}_i' (\hat{\beta} - \beta), \quad (3.5.4)$$

так что

$$\hat{\varepsilon}_i^2 = \varepsilon_i^2 - 2(\hat{\beta} - \beta)' \mathbf{z}_i \varepsilon_i + (\hat{\beta} - \beta)' \mathbf{z}_i \mathbf{z}_i' (\hat{\beta} - \beta). \quad (3.5.5)$$

Суммируя по i , получаем:

$$\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 - 2(\hat{\beta} - \beta)' \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \varepsilon_i + (\hat{\beta} - \beta)' \left(\frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i' \right) (\hat{\beta} - \beta). \quad (3.5.6)$$

Как обычно, $\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 \xrightarrow{p} E(\varepsilon_i^2)$. Поскольку $\hat{\beta}$ является состоятельной и $\frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i'$ сходится по вероятности к некоторой конечной матрице по предположению, последнее слагаемое стремится к нулю. Легко показать, что $E(\mathbf{z}_i \varepsilon_i)$ существует и конечно¹. Тогда, из свойства эргодической стационарности,

$$\frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \varepsilon_i \xrightarrow{p} \text{некоторый конечный вектор.}$$

Таким образом, второе слагаемое в правой части (3.5.6) также стремится к нулю.

Тестирование гипотез

Из утверждений 3.1b и 3.1c очевидным образом выводится

Утверждение 3.3 (робастная t -статистика и статистика Вальда):

Пусть выполнены предположения 3.1–3.5, и пусть доступна состоятельная оценка $\hat{S}$ матрицы $S \equiv \text{Avar}(\bar{g}) = E(g_i g_i')$. Пусть

$$\widehat{\text{Avar}}(\hat{\beta} | \bar{W}) \equiv (S'_{wx} \bar{W} S_{zx})^{-1} S'_{wx} \bar{W} \hat{S} \bar{W} S_{wx} (S'_{wx} \bar{W} S_{wx})^{-1}.$$

Тогда:

¹ Согласно неравенству Коши – Шварца $E(|\mathbf{z}_i \varepsilon_i|) \leq \sqrt{E(\mathbf{z}_i \mathbf{z}_i') E(\varepsilon_i^2)}$, где ε_i — i -й элемент $\mathbf{z}_i$. $E(\mathbf{z}_i \mathbf{z}_i')$ и $E(\varepsilon_i^2)$ конечны по предположению.

(a) При нулевой гипотезе $H_0: \delta_0 = \bar{\delta}_0$,

$$z_1 = \frac{\sqrt{n}(\hat{\delta}_0(\bar{W}) - \bar{\delta}_0)}{\sqrt{(\text{Avar}(\hat{\delta}_0(\bar{W})))_0}} = \frac{\hat{\delta}_0(\bar{W}) - \bar{\delta}_0}{SE_{\hat{\delta}_0}} \rightarrow_d N(0, 1),$$

где $(\text{Avar}(\hat{\delta}_0(\bar{W})))_0$ означает (i, i) элемент $\text{Avar}(\hat{\delta}_0(\bar{W}))$ (являющийся $\text{Avar}(\hat{\delta}_0(\bar{W}))$) и

$$SE_{\hat{\delta}_0}^2 \text{ (робастная стандартная ошибка)} = \sqrt{\frac{1}{n} (\text{Avar}(\hat{\delta}_0(\bar{W})))_0}. \quad (3.5.7)$$

(b) При нулевой гипотезе $H_0: R\delta = r$, где $\#r$ — количество ненулевых (размерность r), и R ($\#r \times L$) имеет полный строковый ранг

$$W = n \cdot (R\hat{\delta}(\bar{W}) - r)' \{R(\text{Avar}(\hat{\delta}_0(\bar{W})))R'\}^{-1} (R\hat{\delta}(\bar{W}) - r) \rightarrow_d \chi^2(\#r). \quad (3.5.8)$$

(c) При нулевой гипотезе $H_0: a(\delta) = 0$ такой, что $A(\delta)$ — матрица первых производных $a(\delta)$ размера $\#a \times L$ (где $\#a$ — размерность a) — непрерывна и имеет полный строковый ранг,

$$W = n \cdot (a(\hat{\delta}(\bar{W}))' [A(\hat{\delta}(\bar{W}))(\text{Avar}(\hat{\delta}_0(\bar{W})))A(\hat{\delta}(\bar{W}))']^{-1} a(\hat{\delta}(\bar{W})) \rightarrow_d \chi^2(\#a). \quad (3.5.9)$$

Для статистики Вальда для нелинейных гипотез справедливы те же комментарии, что и для утверждения 2.6: значения статистики Вальда нечувствительны к предположениям нелинейных ограничений.

Оценивание S

Мы уже рассматривали оценивание S ($-E(g_j g_j') = E(x_j^2 x_j x_j')$) в параграфе 2.5. Продолжим наш очерк, с учетом введенного теперь уравнения $y_i = z_i' \delta + \varepsilon_i$, принимает вид:

$$\hat{S} = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 x_i x_i', \quad (3.5.10)$$

где $\hat{\varepsilon}_i = y_i - z_i' \bar{\delta}$ и $\bar{\delta}$ — состоятельная оценка δ . Для состоятельности этой оценки нужно предположение о четвертых моментах, являющееся обобщением предположения 2.6.

Предположение 3.6 (конечные четвертые моменты): $E[(x_{it}z_{it})^2]$ существует и конечен для всех k ($= 1, \dots, K$) и i ($= 1, \dots, I$).

В качестве аналитического упражнения остается доказательство следующего результата:

Утверждение 3.4 (состоятельное оценивание S): Предположим, что оценка $\hat{S}$, используемая для вычисления остатка $\hat{\varepsilon}_i$ для $\hat{S}$ в (3.5.10), состоятельна, и предположим, что $S = E(g_i g_i')$ существует и конечна. Тогда, с учетом предположений 3.1, 3.2 и 3.6, $\hat{S}$ из выражения (3.5.10) является состоятельной оценкой для S .

Эффективная GMM-оценка

Естественно, хотелось бы выбрать среди GMM-оценок, индексированных посредством $\tilde{W}$, такую оценку, которая имела бы наименьшую асимптотическую дисперсию. Следующее утверждение обеспечивает выбор W , минимизирующей асимптотическую дисперсию.

Утверждение 3.5 (оптимальный выбор взвешивающей матрицы): Нижняя граница асимптотической дисперсии GMM-оценки (3.4.8), проиндексированной посредством $\tilde{W}$, задается выражением $(\Sigma'_{xz} S^{-1} \Sigma_{xz})^{-1}$. Нижняя граница достигается, если $\tilde{W}$ такая, что¹ $W (= \text{plim } \tilde{W}) = S^{-1}$.

Поскольку асимптотическая дисперсия для любой заданной $\tilde{W}$ дается выражением (3.5.1), для доказательства этого утверждения нужно показать, что

$$(\Sigma'_{xz} W \Sigma_{xz})^{-1} \Sigma'_{xz} W S W \Sigma_{xz} (\Sigma'_{xz} W \Sigma_{xz})^{-1} \geq (\Sigma'_{xz} S^{-1} \Sigma_{xz})^{-1} \quad (3.5.11)$$

для любой симметричной положительно определенной матрицы W . Доказательство этого алгебраического результата остается в качестве упражнения.

GMM-оценка, удовлетворяющая условию эффективности $\text{plim } \tilde{W} = S^{-1}$, будет называться **эффективной** (или **оптимальной**) **GMM-оценкой** (*efficient (or optimal) GMM estimator*). Простой подстановкой S^{-1} , являющийся состоятельной оценкой S^{-1} , вместо $\tilde{W}$ в формулах утвер-

¹Условие $\tilde{W} \rightarrow S^{-1}$ достаточное, но не необходимое для эффективности оценок. Необходимым и достаточным условием является существование матрицы C такой, что $\Sigma'_{xz} \tilde{W} = C \Sigma'_{xz} S^{-1}$. См.: [Newey and McFadden, 1993, p. 2163].

ждения 3.1 поочередно:

$$\text{Эффективная GMM-оценка: } \tilde{\beta}(\tilde{S}^{-1}) = (S'_{\pi\pi} \tilde{S}^{-1} S_{\pi\pi})^{-1} S'_{\pi\pi} \tilde{S}^{-1} s_{\pi\pi}. \quad (3.5.12)$$

$$\text{Авария } \tilde{\beta}(\tilde{S}^{-1}) = (S'_{\pi\pi} \tilde{S}^{-1} S_{\pi\pi})^{-1}. \quad (3.5.13)$$

$$\text{Авария } \tilde{\beta}(\tilde{S}^{-1}) = (S'_{\pi\pi} \tilde{S}^{-1} S_{\pi\pi})^{-1}. \quad (3.5.14)$$

Аналогичная подстановка следующим образом преобразует формулу из утверждения 3.3:

$$t_i = \frac{\tilde{\alpha}_i(\tilde{S}^{-1}) - \tilde{\alpha}_i}{SE_i^2}, \quad (3.5.15)$$

где SE_i^2 — робастная стандартная ошибка, заданная выражением:

$$SE_i^2 = \sqrt{\frac{1}{n} \{ (S'_{\pi\pi} \tilde{S}^{-1} S_{\pi\pi})^{-1} \}_{ii}}.$$

и

$$W = n \alpha(\tilde{\beta}(\tilde{S}^{-1}))' \{ A(\tilde{\beta}(\tilde{S}^{-1})) (S'_{\pi\pi} \tilde{S}^{-1} S_{\pi\pi})^{-1} A(\tilde{\beta}(\tilde{S}^{-1}))' \}^{-1} \alpha(\tilde{\beta}(\tilde{S}^{-1})). \quad (3.5.16)$$

Для получения эффективной GMM-оценки необходима состоятельная оценка $\tilde{S}$. Из утверждения 3.4 следует, что $\tilde{S}$, основанная на любой состоятельной оценке $\hat{\beta}$, состоятельна. Это приводит к следующей двухступенчатой процедуре получения эффективных GMM-оценок (*two-step efficient GMM procedure*).

Шаг 1: Выберите матрицу $\tilde{W}$, сходящуюся по вероятности к симметричной положительно определенной матрице, и минимизируйте $J(\tilde{\beta}, \tilde{W})$ по $\tilde{\beta}$ для получения $\tilde{\beta}(\tilde{W})$. В таких матрицах нет проблемы (например $\tilde{W} = I$), но обычно полагают $\tilde{W} = S_{\pi\pi}^{-1}$. Итоговая оценка $\tilde{\beta}(S_{\pi\pi}^{-1})$ является знаменитой оценкой двух-этапного МНК (как мы увидим в параграфе 3.8). Используя полученную оценку для вычисления остатков $\tilde{e}_i = y_i - z_i' \tilde{\beta}(\tilde{W})$ и получения состоятельной оценки $\tilde{S}$ для S по формуле (3.5.10).

Шаг 2: Минимизируйте $J(\tilde{\beta}, \tilde{S}^{-1})$ по $\tilde{\beta}$. Значение, составляющее минимум, является эффективной GMM оценкой.

Асимптотическая мощность

Подобно оценкам коэффициентов, статистика W и статистика Вальда зависят от выбора $\tilde{W}$. Кажется нелогично очевидным, что статистики, связанные с эффективной GMM-оценкой, предпочтительны на больших выборках. Это можно показать формально в терминах асимптотической

мощности, введенной в параграфе 2.4. Возьмем для примера t -статистику для тестирования $H_0: \delta_t = \bar{\delta}_t$. Эта t -статистика записывается (см. (3.5.7)) так

$$t_t \equiv \frac{\sqrt{n}(\hat{\delta}_t(\tilde{W}) - \bar{\delta}_t)}{\sqrt{(Avar(\hat{\delta}(\tilde{W})))_{tt}}}. \quad (3.5.17)$$

Знаменатель сходится к некоторому конечному числу, даже когда эта нулевая гипотеза неверна. Напротив, предел по вероятности числителя, в случае, когда нулевая гипотеза неверна, бесконечен. Поэтому мощность при любой фиксированной альтернативе приближается к единице. Таким образом, тест состоятелен. Это справедливо для любого выбора $\tilde{W}$, поэтому состоятельность теста не может служить критерием выбора $\tilde{W}$.

Рассмотрим последовательность локальных альтернатив со сдвигом Питмена:

$$\delta_t^{(n)} = \bar{\delta}_t + \frac{\gamma}{\sqrt{n}} \quad (3.5.18)$$

для некоторого $\gamma \neq 0$. Подставляя (3.5.18) в (3.5.17), можно переписать t -статистику как

$$t_t = \frac{\sqrt{n}(\hat{\delta}_t(\tilde{W}) - \delta_t^{(n)})}{\sqrt{(Avar(\hat{\delta}(\tilde{W})))_{tt}}} + \frac{\gamma}{\sqrt{(Avar(\hat{\delta}(\tilde{W})))_{tt}}}. \quad (3.5.19)$$

Применяя те же соображения, что и для получения асимптотического распределения t -статистики (2.4.4), можно показать, что $t_t \rightarrow_d N(\mu, 1)$, где

$$\mu = \frac{\gamma}{\sqrt{(Avar(\hat{\delta}(\tilde{W})))_{tt}}}. \quad (3.5.20)$$

Если уровень значимости равен α , асимптотическая мощность задается выражением $\text{Prob}(|x| > t_{\alpha/2})$, где $x \sim N(\mu, 1)$ и $t_{\alpha/2}$ — критическое значение уровня α . Очевидно, чем больше значение $|\mu|$, тем выше асимптотическая мощность. А $|\mu|$ уменьшается с ростом асимптотической дисперсии. Таким образом, асимптотическая мощность максимальна при эффективной GMM-оценке.

Свойства в малых выборках

Сохраняются ли эти желательные асимптотические свойства эффективных GMM-оценок и связанных с ними тестовых статистик в случае конечных выборок? Эффективная GMM-оценка использует $\hat{S}^{-1}$, функцию от оцененных четвертых моментов, для выбора $\tilde{W}$. Обычно для надежного оценивания четвертых моментов требуется гораздо больший

объем выборки. Чем для оценивания первых и вторых моментов. Поэтому можно было бы ожидать от эффективных GMM-оценок худших результатов в конечных выборках по сравнению с оценками, не использующими четвертые моменты. Итальянский выпуск 1996 года журнала *Journal of Business and Economic Statistics* содержит ряд статей, исследующих распределения GMM-оценок и соответствующих тестовых статистик в конечных выборках для различных процессов порождения данных. Общий их вывод заключается в том, что равновзвешенная GMM-оценка ($\bar{W} = I$) обычно превосходит эффективную GMM-оценку по смещению и дисперсии в конечных выборках. Также было обнаружено, что размер статистики Вальда значительно превосходит в конечных выборках заявленный уровень значимости. То есть если α — заявленный уровень значимости, а c_α — соответствующее ему критическое значение, так что $\text{Prob}(\chi^2 > c_\alpha) = \alpha$, то вероятность того, что статистика Вальды в конечных выборках превысит c_α , намного превышает α . Другими словами, нулевая гипотеза будет отвергаться слишком часто. Однако, наряду с другими исследованиями, посвященными конечным выборкам, данные исследования не дают четких количественных указаний, которыми может руководствоваться практик.

Контрольные вопросы

1. Покажите, что все результаты параграфа 2.3–2.5 являются частными случаями результатов этой главы. В частности, покажите, что (1.5.1) сводится к (b). **Указание:** Σ_{xx} является квадратной, если $x_t = x_t'$.
2. (Взрожденная W) Пусть W вырождена, но $\Sigma'_{xx} W \Sigma_{xx}$ невырождена. Покажите, что все результаты данной главы сохраняются (за исключением утверждения 3.5).
3. (Асимптотически эквивалентный выбор $\bar{W}$) Предположим, что $\bar{W}_1 - \bar{W}_2 \rightarrow_p 0$. Покажите, что

$$\sqrt{n}b(\bar{W}_1) - \sqrt{n}b(\bar{W}_2) \rightarrow_p 0.$$

Указание:

$$\sqrt{n}b(\bar{W}_1) - \sqrt{n}b(\bar{W}_2) = (S'_{xx} \bar{W}_1 S_{xx})^{-1} S'_{xx} \bar{W}_1 (S'_{xx} \bar{W}_2 S_{xx})^{-1} S'_{xx} \bar{W}_2 / \sqrt{n}g.$$

$\sqrt{n}g$ сходится по распределению к случайной величине. Используйте лемму 2.4(b).

4. (Трехшаговая GMM.) Предположим, что к эффективной двухшаговой процедуре GMM-оценок добавили следующий шаг: пересчитать $\hat{\beta}$ по формуле (3.5.10), но на этот раз с использованием остатков из второго шага. Вычислите GMM-оценку с использованием этого шага. Состоятельна ли она? Асимптотически нормальна? Эффективна? **Указание:** Проверьте, что, согласно утверждению 3.4, пересчитанная $\hat{\beta}$ является состоятельной оценкой для β . Зависит ли асимптотическое распределение (GMM-оценки) от выбора $\bar{W}$, если $\text{rank}_{n \rightarrow \infty} \bar{W}$ одинакова?

5. (Когда z_1 — строгое подмножество z_2 .) Предположим, что z_1 — строгое подмножество из z_2 . То есть, z_2 содержит, в дополнение к регрессорам (которые все предопределены), и некоторые другие предопределенные переменные. Совпадают ли MНК и GMM-оценки численно? (Ответ: Нет.)
6. (Представление эффективной GMM-оценки через матрицу данных.) Пусть B — диагональная матрица размера $n \times n$, b элемент которой равен $\hat{\sigma}_i^2$, где $\hat{\sigma}_i$ является оценкой дисперсии i -й ошибки ε_i . То есть

$$B = \begin{bmatrix} \hat{\sigma}_1^2 & & \\ & \ddots & \\ & & \hat{\sigma}_n^2 \end{bmatrix}.$$

Проверьте, что

$$\hat{\beta}; \hat{S}^{-1}) = (Z'X(X'BX)^{-1}X'Z)^{-1}Z'X(X'BX)^{-1}X'y,$$

где X , y и Z — матрицы данных (см. инструмент, заданный перекрестки и регрессоры (они определены ниже в параграфе 3.8)).

7. (GLS-представление эффективного GMM.) Пусть X , y и Z такие же, как в предыдущем вопросе. То же самое уравнение можно записать в виде $y = Z\delta + \varepsilon$. Для каждой обе стороны слева на X' , чтобы получить

$$X'y = X'Z\delta + X'\varepsilon.$$

Считая S ковариационной матрицей $X'\varepsilon$, $\hat{S}$ — ее состоятельную оценку, примените FGLS. Проверьте, что FGLS-оценка является эффективной GMM-оценкой (FGLS-оценка была определена в параграфе 1.6).

8. (Линейные комбинации условий ортогональности.) Выведите эффективную GMM-оценку, используя линейную комбинацию условий ортогональности

$$A E(g_i) = 0$$

где A — матрица размера $q \times K$ с полным строковым рангом ($q \leq K$). (Ответ: Замените $\hat{S}_{\varepsilon\varepsilon}$ на $A\hat{S}_{\varepsilon\varepsilon}A'$, x_{xy} на Ax_{xy} и S на ASA' . Формально эту оценку можно записать как (3.4.8) с $\tilde{W} = A'(A\hat{S}A')^{-1}A$.) Проверьте, что при $q = K$ (так, что A невырождена) эффективная GMM-оценка численно совпадает с эффективной GMM-оценкой, соответствующей условиям ортогональности $E(g_i) = 0$.

3.6. Тестирование сверхидентифицирующих ограничений

Если уравнение идентифицируемо точно, то возможно выбрать $\hat{\delta}$ таким, что все выборочные моменты $g_n(\hat{\delta})$ будут равными нулю и расстояние

$$J(\hat{\delta}, \tilde{W}) \equiv n \cdot g_n(\hat{\delta})' \tilde{W} g_n(\hat{\delta})$$

также будет нулевым ($\bar{\delta}$ является IV-оценкой). Если уравнение сверхидентифицируемо, то тогда это расстояние нельзя сделать точно равным нулю, но можно ожидать, что минимальное расстояние будет близким к нулю. Оказывается, что если зависящая матрица $\tilde{W}$ выбирается оптимально, так что $\text{plim } \tilde{W} = S^{-1}$, то минимальное расстояние имеет асимптотическое χ^2 -распределение.

Пусть $\hat{S}$ является состоятельной оценкой S . Рассмотрим сначала случай, когда расстояние вычисляется при истинном значении параметра δ . Поскольку, по определению, $g_n(\bar{\delta}) = \bar{g} (= \frac{1}{n} \sum_i g_i)$ для $\bar{\delta} = \delta$, это расстояние равно:

$$J(\delta, \hat{S}^{-1}) = n \cdot \bar{g}' \hat{S}^{-1} \bar{g} - (\sqrt{n} \bar{g})' \hat{S}^{-1} (\sqrt{n} \bar{g}). \quad (3.6.1)$$

Поскольку $\sqrt{n} \bar{g} \rightarrow_d N(0, S)$ и $\hat{S} \rightarrow_p S$, оно имеет асимптотическое распределение $\chi^2(K)$ по лемме 2.4d. Если заменить δ на $\hat{\delta}(\hat{S})$, то число степеней свободы изменится от K до $K - L$. Очевидной причиной является то, что перед оцениванием выборочного среднего g_i необходимо оценить L параметров δ (мы сталкивались с аналогичной ситуацией в главе 1, в контексте несмещенного оценивания σ^2). Обобщим это как

Утверждение 3.6 (тест Хансена на сверхидентифицирующие ограничения [Hansen, 1982]): Пусть имеется состоятельная оценка, $\hat{S}$, для $S (= E(g_i g_i'))$. При предположениях 3.1–3.5,

$$J(\hat{\delta}(\hat{S}^{-1}), \hat{S}^{-1}) (= n \cdot g_n(\hat{\delta}(\hat{S}^{-1}))' \hat{S}^{-1} g_n(\hat{\delta}(\hat{S}^{-1}))) \rightarrow_d \chi^2(K - L).$$

Формальное доказательство оставшем как дополнительное упражнение. Поскольку $\hat{S}$, заданное (3.5.10), является состоятельной оценкой (при дополнительном предположении 3.6), минимальное расстояние, вычисляемое на втором шаге эффективной GMM-оценки, имеет асимптотическое распределение $\chi^2(K - L)$.

Следует отметить следующее при моменты, связанные с использованием и интерпретацией указанного J -теста:

- Это **тест спецификации (specification test)**, проверяющий выполнение всех ограничений модели (которые являются предположениями, принятыми в утверждении 3.6). Если J -статистика из утверждения 3.6 незначительно велика, то либо условия ортогональности (предположение 3.3), либо иные условия, либо и те и другие, скорее, не выполняются. Только при полной уверенности в этих иных условиях можно интерпретировать большое значение J как признак неадекватности некоторых из K инструментов, включенных в z .

- В отличие от рассмотренных ранее тестов, данный тест нестойтелен по отношению к некоторым нарушениям условий ортогональности. Основная причина этого — снижение числа степеней свободы с K до $K - L$. Легко показать, что $\bar{g}$ связан со своим выборочным аналогом $g_n(\hat{\beta}(\hat{S}^{-1}))$ как

$$\sqrt{n} g_n(\hat{\beta}(\hat{S}^{-1})) = \tilde{B} \sqrt{n} \bar{g}, \quad \tilde{B} \equiv I_K - S_{xz}(S'_{xz} \hat{S}^{-1} S_{xz})^{-1} S'_{xz} \hat{S}^{-1}. \quad (3.6.2)$$

Проблема в том, что, так как $\tilde{B} S_{xz} = 0$, матрица $\tilde{B}$ не имеет полного столбцового ранга. Если условия ортогональности не выполняются и $E(g_i) \neq 0$, то элементы $\sqrt{n} \bar{g}$ будут стремиться к $-\infty$ и $+\infty$. Но, так как $\tilde{B}$ не имеет полного столбцового ранга, $\tilde{B} \sqrt{n} \bar{g}$ и поэтому $J(\hat{\beta}(\hat{S}^{-1}), \hat{S}^{-1})$ могут оставаться конечными при некоторых типах нарушений ортогональности¹.

- Свойства тестов в конечных выборках стали объектом пристального внимания лишь недавно. Некоторые статьи из нумерического выпуска 1996 года *Journal of Business and Economic Statistics* показали, что реальный размер J -теста в конечных выборках значительно превышает номинальный размер (нулевая гипотеза отвергается слишком часто).

Тестирование подмножества условий ортогональности

Предположим, что мы можем разделить K инструментов на две группы: вектор x_{1i} из K_1 переменных, удовлетворяющих условиям ортогональности, и вектор x_{2i} из оставшихся $K - K_1$ переменных, находящихся под подозрением. Поскольку порядок переменных не меняет численных свойств статистик, можно предположить без потери общности, что именно последние $K - K_1$ переменных x_i являются подозреваемыми инструментами.

$$x_i = \begin{bmatrix} x_{1i} \\ x_{2i} \end{bmatrix} \begin{cases} K_1 \text{ строк,} \\ K - K_1 \text{ строк.} \end{cases} \quad (3.6.3)$$

Часть модели, которую мы хотели бы протестировать, — это

$$E(x_{2i} z_i) = 0. \quad (3.6.4)$$

Это ограничение тестируемо, если число надежных инструментов не меньше числа коэффициентов, то есть $K_1 \geq L$. Основная идея заключается в сравнении двух J -статистик двух отдельных GMM оценок одного

¹См. [Newey, 1995, Section 5] для детального рассмотрения этого вопроса.

и того же вектора δ : с использованием только вектора x_{i1} и с использованием обоих векторов, x_{i1} и x_{i2} . Если значение дополнительных инструментов увеличивает значение J , то есть основания сомневаться в их предопределенности.

В соответствии с указанным разбиением x_i выборочные условия ортогональности $g_n(\tilde{\delta})$ и S могут быть записаны как

$$g_n(\tilde{\delta}) = \begin{bmatrix} g_{1n}(\tilde{\delta}) \\ g_{2n}(\tilde{\delta}) \end{bmatrix}, \quad (K \times K) \quad S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \quad (3.6.5)$$

где

$$\begin{aligned} S_{11} &= E(z_i^2 x_{i1} x_{i1}'), & S_{12} &= E(z_i^2 x_{i1} x_{i2}'), \\ S_{21} &= E(z_i^2 x_{i2} x_{i1}'), & S_{22} &= E(z_i^2 x_{i2} x_{i2}'). \end{aligned}$$

В частности, $g_{1n}(\tilde{\delta})$ можно записать как

$$g_{1n}(\tilde{\delta}) = \frac{1}{n} \sum_{i=1}^n x_{i1} (y_i - z_i' \tilde{\delta}) = s_{x_1 y} - s_{x_1 z} \tilde{\delta},$$

где

$$s_{x_1 y} = \frac{1}{n} \sum_{i=1}^n x_{i1} y_i, \quad s_{x_1 z} = \frac{1}{n} \sum_{i=1}^n x_{i1} z_i'. \quad (3.6.6)$$

Для состоятельной оценки $\tilde{S}$ эффективная GMM-оценка, использующая все K инструментов, и соответствующая J -статистика уже были выведены в этом и предыдущем параграфах. Воспроизведем их:

$$\tilde{\delta} = (S'_{xz} \tilde{S}^{-1} S_{xz})^{-1} S'_{xz} \tilde{S}^{-1} s_{xy}, \quad (3.6.7)$$

$$J = n \cdot g_n(\tilde{\delta})' \tilde{S}^{-1} g_n(\tilde{\delta}). \quad (3.6.8)$$

Эффективная GMM-оценка того же вектора коэффициентов, но с использованием только первых K_1 инструментов и соответствующая J -статистика получаются заменой x_i на x_{i1} .

$$\tilde{\delta} = (S'_{x_1 z} (\tilde{S}_{11})^{-1} S_{x_1 z})^{-1} S'_{x_1 z} (\tilde{S}_{11})^{-1} s_{x_1 y}, \quad (3.6.9)$$

$$J_1 = n \cdot g_{1n}(\tilde{\delta})' (\tilde{S}_{11})^{-1} g_{1n}(\tilde{\delta}). \quad (3.6.10)$$

где $\tilde{S}_{11}$ — состоятельная оценка S_{11} . Тест основан на следующем утверждении, определяющем асимптотическое распределение $J - J_1$ (доказательство остается в качестве дополнительного упражнения).

Утверждение 3.7 (тестирование подмножества условий ортогональности)¹: Пусть выполнены предположения 3.1–3.5 и x_1 — известный вектор x_0 . Если выполнены предположения 3.4 и требование выполнения ранга для полной идентификации для x_1 (то есть $E(x_1 x_1')$ имеет полный столбцовый ранг). Тогда для любых состоятельных оценок $\hat{S}$ для S и $\hat{S}_1$ для S_{11} ,

$$C = (I - J_1 J_1^{-1}) x_1' (K - K_1),$$

где $K = \frac{1}{n} \sum x_1 x_1'$ (размерность x_1), $K_1 = \frac{1}{n} \sum x_{11} x_{11}'$ (размерность x_{11}), а J и J_1 определять в (3.6.8) и (3.6.10).

Выбор $\hat{S}$ и $\hat{S}_1$ асимптотически неважен, пока они состоятельны. Но в конечных выборках статистика C может быть отрицательной. Эта проблема может быть решена, и C может быть сделана неотрицательной, если вместо использования одной и той же матрицы $\hat{S}$, то есть если матрица $\hat{S}_{11}$ в (3.6.9) и (3.6.10) будет подматрицей $\hat{S}$ из (3.6.7) и (3.6.8). Это достигается следующими шагами:

1. Примените эффективный двухшаговый GMM к данным выборок инструментов x_1 для получения $\hat{S}$ на первом шаге и $\hat{J}$ и $\hat{S}_1$ на втором.
2. Извлеките подматрицу $\hat{S}_{11}$ из матрицы $\hat{S}$, вычислите $\bar{S}$ по (3.6.9) с использованием этой матрицы и J_1 по (3.6.10). Затем вычтите разность J .

Доказательство неотрицательности полученной оценки C в конечных выборках остается как дополнительное аналитическое упражнение.

Утверждение 3.7 можно использовать для тестирования экзогенности подмножества регрессоров, как в следующем примере.

Пример 3.3 (тестирование предопределенности образования в уравнении заработной платы): В уравнении заработной платы из примера 3.2 предположим, что уровень образования X_1 может быть эндогенным. Для тестирования S_1 на эндогенность разложим x_1 как

$$x_{11} = \begin{bmatrix} 1 \\ EXPR_1 \\ AGE_1 \\ UED_1 \end{bmatrix}, \quad x_{01} = S_1.$$

Вектор регрессоров x_1 тот же, что и в примере 3.2. Первый шаг — применить эффективное двухшаговое GMM-оценивание $\hat{S}$ с $x_1 \equiv (1, EXPR_1,$

¹Тест разработан в [Newey, 1985, Секция 4] и [Efronimov, Hansen, Singleton, 1999, Appendix C].

$AGE_1, ME(D), S_1)'$ в качестве инструментов. Получим J и матрицу $\hat{S}$ размера 5×5 . На втором этапе извлечем подматрицу $\hat{S}_{11}$, соответствующую x_{11} , и оценим тот же вектор δ с меньшим количеством инструментов и с использованием этой $\hat{S}_{11}$. Разность J -статистик, полученных двумя различными вариантами GMM-оценивания δ , должна иметь асимптотическое распределение $\chi^2(1)$.

Контрольные вопросы

1. Верно ли, что $J(\hat{\delta}(\hat{S}^{-1}); \hat{S}^{-1}) \sim \chi^2(K-L)$, если $\hat{S}$ и $\bar{S}$ – две разные состоятельные оценки S ?
2. Покажите, что $J(\hat{\delta}(\hat{S}^{-1}); \hat{S}^{-1}) = n s'_{21} \hat{S}^{-1} (s_{22} - s_{21} \hat{\delta}(\hat{S}^{-1}))$.
3. Может ли число степеней свободы C -статистики быть больше, чем $K-L$? [Ответ: Нет.]
4. Пусть $K_1 = L$. Зависит ли численное значение C от разбиения x_1 на x_{11} и x_{12} ?

3.7. Тестирование гипотез по принципу отношения правдоподобий

В параграфе 3.5 нами были выведены статистики χ^2 для $H_0: \alpha(\delta) = 0$ по принципу Вальда. В данном параграфе то же самое будет сделано по принципу отношения правдоподобий (LR), заключающемуся в сравнении значений целевой функции с учетом нулевой гипотезы и без. Вывод тестовых статистик по принципу множителей Лагранжа для GMM и обобщение на случай нелинейных уравнений даны в параграфе 7.4.

В эффективном GMM-оценивании целевой функцией является $J(\hat{\delta}, \bar{S}^{-1})$ для некоторой состоятельной оценки $\bar{S}$ для S . **Эффективная GMM-оценка с ограничением** (*restricted efficient GMM estimator*) определится как

$$\begin{aligned} \text{эффективная GMM-оценка с ограничением: } \bar{\delta}(\bar{S}^{-1}) \equiv \\ \equiv \arg \min_{\delta} J(\hat{\delta}, \bar{S}^{-1}) \text{ с учетом } H_0. \end{aligned} \quad (3.7.1)$$

LR-принцип подразумевает, что разность

$$LR \equiv J(\bar{\delta}(\bar{S}^{-1}), \bar{S}^{-1}) - J(\hat{\delta}(\hat{S}^{-1}), \hat{S}^{-1}) \quad (3.7.2)$$

должна иметь асимптотическое распределение χ^2 . Это действительно так.

Утверждение 3.8 (тестовая статистика по принципу LR): Пусть выполнены предположения 3.1–3.5 и пусть доступна состоятельная оценка $\hat{S}$ матрицы S ($= E(g_i g_i')$). Рассмотрим нулевую гипотезу из $\#a$ ограничений $H_0: a(\delta) = 0$ таких, что $A(\delta) \leftarrow$ матрица первых производных размер $\#a \times L$, непрерывна и имеет полный строковый ранг. Определим две статистики, W и LR , согласно (3.5.16) и (3.7.2) соответственно. Тогда, при нулевой гипотезе, верно следующее:

- (a) Обе указанные статистики асимптотически эквивалентны, так как имеют одинаковые асимптотические распределения (а именно $\chi^2(\#a)$).
- (b) Обе статистики асимптотически эквивалентны в более строгом смысле, так что их разность сходится к нулю по вероятности: $LR - W \rightarrow_p 0$. (По лемме 2.4(a), этот результат сильнее, чем (a).)
- (c) Более того, если указанная гипотеза линейна, так что ограничения могут быть записаны в виде $R\delta = r$, то тогда две указанные статистики численно равны.

Доказательство утверждения для линейного случая является алгебраическим утверждением, которое оставим в качестве аналитического утверждения. Относительно полного доказательства смотрите параграф 7.4. Несколько комментариев к утверждению 3.8:

- С одной стороны, преимущество LR перед W является инвариантностью: численное значение LR не зависит от формы представления нелинейных ограничений в виде функции $a(\cdot)$. С другой стороны, при нелинейной гипотезе явно необходимо написать компьютерную программу для поиска эффективной GMM-оценки с ограничениями.
- Утверждение 3.8 требует, чтобы матрица расстояний $\hat{W}$ удовлетворяла условию эффективности $\lim \hat{W} = S^{-1}$. В противном случае LR не будет иметь асимптотического χ^2 -распределения. Напротив, статистика Вальда имеет асимптотическое распределение χ^2 и без выполнения этого условия.
- При вычислении LR следует использовать одну и ту же оценку S . Пусть $\tilde{S}$ и $\bar{S}$ — две различные состоятельные оценки S . Рассмотрим статистику

$$J(\tilde{\delta}(\bar{S}^{-1}), \bar{S}^{-1}) - J(\tilde{\delta}(\tilde{S}^{-1}), \tilde{S}^{-1}),$$

полученную при применении двух отдельных двухшаговых эффективных GMM-процедур. $\bar{S}$ получена на первом шаге с учетом ограничений, $\tilde{S}$ получена на первом шаге без учета ограничений. Эта

статистики асимптотически эквивалентна LR (их разность сходится по вероятности к нулю), но в конечных выборках она может быть отрицательной. Использование односторонних оценок $\hat{S}$ гарантирует неотрицательность статистики в конечных выборках. Исследователи обычно используют оценку, полученную при оценивании без учета ограничений, хотя оценка, полученная при оценивании с учетом ограничений, также допустима, поскольку она состоятельна при нулевой гипотезе.

- Часть δ утверждения (асимптотическая эквивалентность в более строгом смысле) говорит, что при достаточно большом объеме выборки и справедливости гипотезы результаты LR -теста (а не только вероятность отвержения или принятия) будут такими же, как и у теста W , потому что вероятность того, что статистики отличаются даже на малую величину, равна нулю на достаточно больших выборках.
- Для численного совпадения $LR = W$ в линейном случае, одна и та же оценка $\hat{S}$ должна быть использована для расчета не только LR , но и W . В противном случае статистики будут только асимптотически эквивалентными.

LR-статистика для модели регрессии

Поскольку модель регрессии из главы 2 является частным случаем GMM-модели из этой главы, может быть полезно узнать, как LR будет выглядеть в этом частном случае. Поскольку в этой модели регрессии $x_i = z_i$ эффективной GMM-оценкой (не учитывающий ограничения) является MLK и $J(\bar{\delta}(\hat{S}^{-1}), \hat{S}^{-1}) = 0$. Таким образом,

$$LR = J(\bar{\delta}(\hat{S}^{-1}), \hat{S}^{-1}), \quad (3.7.3)$$

где $\bar{\delta}(\hat{S}^{-1})$ — ограниченная эффективная GMM-оценка. По предположению 3.8, эта статистика имеет асимптотическое распределение χ^2 и численно равна статистике Вальда при линейной нулевой гипотезе. Как будет показано ниже, при условной гомоскедастичности, эта статистика может быть записана как разность сумм квадратов остатков, сформированных на дисперсию ошибки.

Тест добавления переменной (необязательный материал)

В предыдущем параграфе мы рассмотрели тест спецификац. ин. основанный на S -статистике, для проверки эндогенности подмножества $x_{1,2}$ инструментом x_1 при ус. свий предопределенности остальных инструментов.

тов x_i . Иногда мы сталкиваемся с частным случаем:

$$y_i = x_i' \beta + \varepsilon_i. \quad (3.7.4)$$

Популярным способом проверить предопределенность инструментов x_i является оценивание расширенной модели:

$$y_i = x_{i1}' \beta + x_{i2}' \alpha + \varepsilon_i + x_i' \gamma + \varepsilon_i, \quad \text{с } \gamma = \begin{bmatrix} \delta \\ \alpha \end{bmatrix} \quad (3.7.5)$$

и тестирование нулевой гипотезы $H_0: \alpha = 0$. Тестирование приходится либо с использованием статистики Вальда, либо эквивалентной ей L -статистики. Этот тест часто называют тестом *добавления переменных* (*variable addition test*). Как он связан с C -тестом из утверждения 3.7?.

Для расчета L - H необходимо получить две эффективные GMM-оценки γ с одним и тем же набором инструментов x_i с ограничением $\alpha = 0$ и без него. Эффективной GMM-оценкой без ограничений будет МНК оценка уравнения (3.7.5), не учитывающего ограничений. Соответствующая J -статистика равна нулю. Пусть

$$\hat{S} \equiv \frac{1}{n} \sum_{i=1}^n e_i^2 x_i x_i', \quad (3.7.6)$$

где e_i — остатки МНК оценивания уравнения (3.7.5) без ограничений. При использовании этой оценки $\hat{S}$ эффективная GMM-оценка γ , учитывающая ограничения, минимизирует функцию

$$J(\tilde{\gamma}; \hat{S}^{-1}) = n(x_{10}' - S_{\alpha\alpha}\tilde{\gamma})' \hat{S}^{-1}(x_{10}' - S_{\alpha\alpha}\tilde{\gamma})$$

при $\alpha = 0$. Очевидно, что вместо $\hat{S}$, учитывающую ограничения, можно записать следующим образом

$$\pi = \begin{bmatrix} \tilde{\delta}(\hat{S}^{-1}) \\ 0 \end{bmatrix},$$

где $\tilde{\delta}(\hat{S}^{-1})$ — эффективная GMM-оценка уравнения (3.7.4), учитывая этого ограничения, с инструментами x_i . Итак,

$$\begin{aligned} J(R) &= n(x_{10}' - S_{\alpha\alpha}\tilde{\gamma})' \hat{S}^{-1}(x_{10}' - S_{\alpha\alpha}\tilde{\gamma}) \\ &\quad (\text{так как } J = 0 \text{ для GMM без ограничений}) \\ &= n(x_{10}' - S_{\alpha\alpha}\tilde{\delta}(\hat{S}^{-1}))' \hat{S}^{-1}(x_{20}' - S_{\alpha\alpha}\tilde{\delta}(\hat{S}^{-1})) \\ &\quad (\text{так как } S_{\alpha\alpha}\tilde{\gamma} = S_{\alpha\alpha}\tilde{\delta}(\hat{S}^{-1})). \end{aligned}$$

Это не что иное, как J -статистика Хансена для уравнения (3.7.4) без ограничений с инструментами $\bar{z}_i$. Эта статистика, в свою очередь, равна C -статистике, так как J из утверждения 3.7 в настоящем случае равно нулю. Таким образом, все статистики W из регрессии без ограничений, LR , C — численно равны J Хансена, при условии использования одной и той же $\bar{S}$. То есть тест добавления переменных численно эквивалентен тесту Хансена на сверхидентифицирующие ограничения.

Контрольные вопросы

1. (LR для модели регрессии.) Проверьте, что $\hat{\alpha}$ модель регрессии с $\bar{z}_i = \bar{z}_i$,

$$LR = y'(X_1(\bar{S})^{-1}X_1'y - 2y'X_1(\bar{S})^{-1}X_1'X_2)\bar{\delta} - \bar{\delta}'(X_1'X_1)(\bar{S})^{-1}X_1'X_2\bar{\delta},$$

2. (Выбор $\bar{S}$ в тесте добавления переменных.) Предположим, что $\bar{S}$ формируется из остатков, полученных при регрессии уравнения с ограничениями (3.7.4). $\bar{e}$ используется для получения W , LR и C . Равны ли они численно? Почему ли они асимптотическое χ^2 распределение? Укажите: Составитель для матрицы $\bar{S}$ при предельной регрессии?

3.8. Приложения условной гомоскедастичности

До сих пор при изводе асимптотик для GMM-оценок мы не предполагали наличия условной гомоскедастичности. В этом параграфе рассматриваются последствия введения следующего предположения.

Предположение 3.7 (условная гомоскедастичность):

$$E(\varepsilon_i^2 | \mathbf{x}_i) = \sigma^2.$$

При условной гомоскедастичности матрица четвертых моментов S (равная $E(y_i g_i') - E(\varepsilon_i^2 | \mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i'$) может быть записана как произведение вторых моментов:

$$S = \sigma^2 \Sigma_{xx}, \quad (3.8.1)$$

где $\Sigma_{xx} = E(\mathbf{x}_i \mathbf{x}_i')$. Как и в лемме 2 данное разложение имеет несколько следствий:

- Поскольку S невырождена, по предположению 3.5, из данного разложения следует, что $\sigma^2 > 0$ и Σ_{xx} невырождена.
- Оценка, использующая данное разложение S , имеет вид:

$$\bar{S} = \sigma^2 \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' = \sigma^2 S_{xx}, \quad (3.8.2)$$

где $\hat{\sigma}^2$ — некоторая состоятельная оценка, определяемая ниже. Из эргодической стационарности, $S_{xx} \rightarrow_s \Sigma_{xx}$. Таким образом, при состоятельной $\hat{\sigma}^2$ нам не нужны требование состоятельности четвертых моментов (предположение 3.6), чтобы обеспечить состоятельность $\hat{S}$.

Излишне говорить, что все представленные до сих пор результаты верны при этом дополнительном предположении условной гомоскедастичности. Но многие результаты и формулы могут быть упрощены заменой S на $\sigma^2 \Sigma_{xx}$, а выражения (3.5.10) для $\hat{S}$ на (3.8.2). Эти упрощения собраны в данном параграфе.

Эффективный GMM становится 2SLS

В эффективном GMM-оценивании взвешивающей матрицей является $\hat{S}^{-1}$. Если взять $\hat{S}$ из (3.8.2), то GMM-оценка станет разной:

$$\begin{aligned}\hat{\delta}(\hat{S}^{-1}) &= [S'_{xx}(\hat{\sigma}^2 S_{xx})^{-1} S_{xz}]^{-1} S'_{xz}(\hat{\sigma}^2 S_{xx})^{-1} s_{zy} = \\ &= (S'_{xx} S_{xx}^{-1} S_{xz})^{-1} S'_{xz} S_{xx}^{-1} s_{zy} = \\ &= \hat{\delta}(S_{xx}^{-1}) = \hat{\delta}_{2SLS}\end{aligned}\quad (3.8.3)$$

и не зависящей от $\hat{\sigma}^2$. В общем случае основной задачей первого шага эффективной двухшаговой процедуры GMM является получение состоятельной оценки S . При условной гомоскедастичности в первом шаге необходимости нет, так как второй шаг сводится к $\hat{\delta}(S_{xx}^{-1})$ — оценке GMM с S_{xx} вместо $\hat{S}$. Эта оценка, $\hat{\delta}_{2SLS}$, называется **двухшаговой МНК** (*Two-Stage Least Squares*) (2SLS или TSLS)¹. Это же уравнение можно оценить методом максимального правдоподобия. Параграф 8.6 будет включать в себя ML аналог 2SLS, называемый оценкой максимального правдоподобия с ограниченной информацией.

Выражение для $\text{Avar}(\hat{\delta}_{2SLS})$ может быть получено подстановкой (3.8.1) в (3.5.13):

$$\text{Avar}(\hat{\delta}_{2SLS}) = \sigma^2 \cdot (\Sigma'_{xx} \Sigma_{xx}^{-1} \Sigma_{xz})^{-1}. \quad (3.8.4)$$

Естественной оценкой этого является

$$\widehat{\text{Avar}}(\hat{\delta}_{2SLS}) = \hat{\sigma}^2 \cdot (S'_{xx} S_{xx}^{-1} S_{xz})^{-1}. \quad (3.8.5)$$

Что касается $\hat{\sigma}^2$, рассмотрим выборочную дисперсию ошибок 2SLS:

$$\hat{\sigma}^2 \equiv \frac{1}{n} \sum_{i=1}^n (y_i - \mathbf{z}'_i \hat{\delta}_{2SLS})^2 \quad (3.8.6)$$

¹ Оценка впервые предложена в [Theil, 1953].

(некоторые авторы при вычислении $\hat{\sigma}^2$ делят сумму квадратов на $n - L$, а не на n). В силу утверждения 3.2, (3.8.6) $\rightarrow_p \hat{\sigma}^2$, если $E(z_i z_i')$ существует и конечна. Тем самым $\tilde{S}$, определенная в (3.8.2), с данной $\hat{\sigma}^2$ является состоятельной для S .

Подставляя (3.8.2) в (3.8.15) и (3.5.16), получим следующие выражения для t -статистики и статистики Вальда:

$$t_f = \frac{\tilde{\delta}_{2SLS} - \bar{\delta}_f}{SE_f} \quad \text{с} \quad SE_f = \sqrt{\frac{\hat{\sigma}^2}{n} \cdot ((S'_{xz} S_{zz}^{-1} S_{xz})^{-1})_{ff}} \quad (3.8.7)$$

$$W = n \cdot \frac{\alpha(\tilde{\delta}_{2SLS})' [A(\tilde{\delta}_{2SLS})(S'_{xz} S_{zz}^{-1} S_{xz})^{-1} A(\tilde{\delta}_{2SLS})']^{-1} \alpha(\tilde{\delta}_{2SLS})}{\hat{\sigma}^2} \quad (3.8.8)$$

J становится статистикой Саргана

Зная $\tilde{W} = (\hat{\sigma}^2 S_{zz})^{-1}$, получим следующее выражение для расстояния (3.4.6):

$$J(\tilde{\delta}, (\hat{\sigma}^2 S_{zz})^{-1}) = n \cdot \frac{(z_{zy} - S_{zx} \tilde{\delta})' S_{zz}^{-1} (z_{zy} - S_{zx} \tilde{\delta})}{\hat{\sigma}^2} \quad (3.8.9)$$

Из утверждения 3.6 тогда следует, что это расстояние, вычисленное для эффективной GMM-оценки при условной гомоскедастичности, $\tilde{\delta}_{2SLS}$, имеет асимптотическое распределение χ^2 . Это расстояние называется **статистикой Саргана** (*Sargan's statistic*) [Sargan, 1958]:

$$\text{статистика Саргана} = n \cdot \frac{(z_{zy} - S_{zx} \tilde{\delta}_{2SLS})' S_{zz}^{-1} (z_{zy} - S_{zx} \tilde{\delta}_{2SLS})}{\hat{\sigma}^2} \quad (3.8.10)$$

Резюмируем полученные нами до этого результаты.

Утверждение 3.9 (асимптотические свойства 2SLS):

(a) При предположениях 3.1–3.4, оценка 2SLS (3.8.3) состоятельна. При добавлении предположения 3.5, эта оценка асимптотически нормальна с асимптотической дисперсией (3.5.1), где $W = (\sigma^2 \Sigma_{zz})^{-1}$. При добавлении к предположениям 3.1–3.5 предположения 3.7 (условная гомоскедастичность) эта оценка будет эффективной GMM-оценкой.

Более того, если $E(z_i z_i')$ существует и конечна¹, то

(b) (3.8.5) — состоятельная оценка асимптотической дисперсии.

(c) t_f в (3.8.7) $\rightarrow_d N(0, 1)$, W в (3.8.8) $\rightarrow_d \chi^2(k)$ и

¹Это дополнительное предположение требуется для состоятельности $\hat{\sigma}^2$, см. утверждение 3.2.

(d) статистика Саргана (3.8.10) $\rightarrow_d \chi^2(K \cdot L)$.

Утверждение 3.8 показывает, что LR -статистика, являющаяся разностью значений J , полученных с учетом нулевой гипотезы и без, имеет асимптотическое распределение χ^2 . Поскольку J можно записать как (3.8.9), мы получаем:

$$LR = n \times \frac{(s_{xy} - S_{xz}\bar{\delta})' S_{xx}^{-1} (s_{xy} - S_{xz}\bar{\delta}) \cdot (s_{xy} - S_{xz}\hat{\delta}_{2SLS})' S_{xx}^{-1} (s_{xy} - S_{xz}\hat{\delta}_{2SLS})}{\hat{\sigma}^2}, \quad (3.8.11)$$

где $\bar{\delta}$ — ограниченная 2SLS-оценка, минимизирующая (3.8.9) при нулевой гипотезе¹. В утверждении 3.8 использование одноклаковых $\hat{\sigma}^2$ гарантирует неотрицательность статистики в конечных выборках. Здесь деление на одну и ту же $\hat{\sigma}^2$ гарантирует неотрицательность статистики. Если гипотеза линейна, то тогда эта LR -статистика численно равна статистике Вальда W .

Свойства 2SLS на конечных выборках

Имеется довольно большое количество литературы, посвященной распределению оценки 2SLS в конечных выборках (см., например: [Judge et al., 1985, Section 15.4] и [Staiger and Stock, 1997, Section 1]). Некоторые исследования выводят точное распределение этих оценки в конечных выборках, и то время как другие используют метод Монте-Карли для различных DGP. Аналитические результаты, однако, неприменимы в эмпирических исследованиях, так как они выводятся при жестких ограничениях фиксированности инструментов и нормальности ошибок, а аналитические выражения не поддаются вычислениям.

Для случая единственного регрессора и единственного (стохастического) инструмента с нормальными ошибками [Nelson and Starz, 1990] выводят точное распределение оценки 2SLS в конечных выборках, относительно простое и легко вычисляемое. Они также показывают, что, если инструмент «слабый», в смысле низкой объяснительной силы на первом шаге регрессии регрессора на инструмент, основанная на этом распределении оценка ошибки оценки $\hat{\sigma}_{2SLS}^2 - \hat{\sigma}^2$ располагается далеко от нуля при недостаточно большом объеме выборки.

Их работа показывает необходимость сообщать значение R^2 для первого шага оценивания; если R^2 мал, то следует ожидать плохую аппроксимацию распределения оценки 2SLS (пример приведен в части (g)).

¹ Эта статистика впервые представлена в: [Gallant and Jorgensen, 1979].

эмпирического упрощения). Недавно [Staiger and Stock, 1997] предложили альтернативное асимптотическое приближение распределения оценок 2SLS в конечных выборках, других оценках и статистиках для случая «слабых» инструментов. Их метод заключается в рассмотрении подпоследовательности моделей, в которых коэффициенты регрессии инструментов сходятся к нулю (аналитическое упрощение 10 павидит данных) для асимптотики для простого случая одного регрессора и одного инструмента).

Альтернативные оценки 2SLS

Если определить матрицы данных как

$$\mathbf{X}_{(n \times K)} = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{bmatrix}, \quad \mathbf{Z}_{(n \times L)} = \begin{bmatrix} z_1' \\ z_2' \\ \vdots \\ z_n' \end{bmatrix}, \quad \mathbf{y}_{(n \times 1)} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

то можно увидеть, что оценки 2SLS и соответствующие статистики могут быть переписаны как

$$\bar{\delta}_{2SLS} = [Z'X(X'X)^{-1}X'Z]^{-1}Z'X(X'X)^{-1}X'y - \\ - (Z'PZ)^{-1}Z'Py, \quad (3.8.3')$$

где $P \equiv X(X'X)^{-1}X'$ — проекционная матрица

$$Asym(\bar{\delta}_{2SLS}) = n \cdot \bar{\sigma}^2 \cdot [Z'X(X'X)^{-1}X'Z]^{-1} = n \cdot \bar{\sigma}^2 \cdot (Z'PZ)^{-1}, \quad (3.8.5')$$

$$\bar{\sigma}^2 = \frac{\bar{\varepsilon}'\bar{\varepsilon}}{n}, \quad \text{где } \bar{\varepsilon} = y - Z\bar{\delta}_{2SLS}, \quad (3.8.6')$$

$$t_k = \frac{\bar{\delta}_{2SLS,k}}{\sqrt{\bar{\sigma}^2 \cdot ([Z'X(X'X)^{-1}X'Z]^{-1})_{kk}}} = \frac{\bar{\delta}_k}{\sqrt{\bar{\sigma}^2 \cdot ([Z'PZ]^{-1})_{kk}}}, \quad (3.8.7')$$

$$W = \frac{\alpha(\bar{\delta}_{2SLS})' \left[A(\bar{\delta}_{2SLS}) (Z'X(X'X)^{-1}X'Z)^{-1} A(\bar{\delta}_{2SLS})' \right]^{-1} \alpha(\bar{\delta}_{2SLS})}{\bar{\sigma}^2} = \quad (3.8.8')$$

$$J(\bar{\delta}, (\bar{\sigma}^2, S_{xx})^{-1}) = (y - Z\bar{\delta})' P (y - Z\bar{\delta}), \quad (3.8.9')$$

$$\text{Статистика Саргана} = \frac{\bar{\varepsilon}' P \bar{\varepsilon}}{\bar{\sigma}^2}. \quad (3.8.10')$$

Используя эти формулы, можно получить для других выводов оценки 2SLS.

2SLS как IV-оценка

Пусть $\tilde{z}_i (L \times 1)$ — вектор L инструментов (которые будут сгенерированы из π_1 по описанному ниже способу) для L регрессоров; и пусть $\tilde{Z}$ — матрица этих инструментов размера $n \times L$. То есть i -я строка $\tilde{Z}$ — это $\tilde{z}_i'$. IV-оценка δ с $\tilde{z}_i$ в качестве инструментов есть, по (3.4.4),

$$\hat{\delta}_{IV} = \left(\frac{1}{n} \sum_{i=1}^n \tilde{z}_i z_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n \tilde{z}_i y_i = (\tilde{Z}' Z)^{-1} \tilde{Z}' y. \quad (3.8.12)$$

Теперь мы сгенерируем эти инструменты следующим образом: i -й инструмент является подобранным значением регрессии z_{ℓ} (ℓ -го регрессора) на π_1 . Вектор длины n из подобранных значений равен $X(X'X)^{-1}X'z_{\ell}$, где z_{ℓ} — вектор длины n значений ℓ -го регрессора (ℓ -й столбец матрицы Z). Таким образом, $n \times L$ матрица инструментов имеет вид

$$\tilde{Z} = [X(X'X)^{-1}X'z_1, \dots, X(X'X)^{-1}X'z_L] = X(X'X)^{-1}X'Z = PZ, \quad (3.8.13)$$

где P — проекционная матрица [подстановка этого выражения в (3.8.12) дает 2SLS-оценку (см. (3.8.3))].

2SLS как две регрессии

Вместо подстановки сгенерированных инструментов $\tilde{z}_i$ в выражение для IV-оценки на уравнения $y_i = z_i'\delta + \varepsilon_i$, рассмотрим регрессию y_i на $\tilde{z}_i$. Оценка коэффициентов имеет вид

$$(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'y = (Z'P'PZ)^{-1}Z'P'y = (Z'PZ)^{-1}Z'P'y,$$

поскольку P симметрична и идемпотентна.

Это 2SLS-оценка. Поэтому оценки 2SLS могут быть получены за два шага: оценкой регрессии ℓ регрессора на π_1 и получением подобранных значений $\tilde{z}_i$; и последующей оценкой регрессии y_i на эти подобранные значения.

Для predetermined регрессоров, входящих в π_1 , нет необходимости в первом шаге, так как их значения π_1 будут самими этими регрессорами. Действительно, если z_{ℓ} задан/предель и включен в π_1 как k -й инструмент, вектор длины n подобранных значений для регрессора ℓ равен Pz_{ℓ} , где z_{ℓ} — вектор длины n , чьим k -м элементом является z_{ℓ} . Но так как z_{ℓ} также является k -м столбцом X , $Pz_{\ell} = P\pi_k$. Поскольку P — проекционная матрица, то $P\pi_k = \pi_k$.

Этот вывод 2SLS полезен, так как он оправдывает название оценки. Но есть одна подводная камушка. При выполнении регрессии y_i на $\tilde{z}_i$ на первом этапе стандартные ошибки автоматически рассчитываются МНК-программой на основе ошибок $y - \tilde{Z}\hat{\delta}_{2SLS}$, которые не равны

остаткам $2SLS \hat{y} - Z\hat{\delta}_{2SLS}$. Поэтому соответствующие стандартные ошибки и асимптотические дисперсии не могут использоваться для статистических выводов.

Когда регрессоры predetermined

Когда все регрессоры являются predetermined, а ошибки условно гомоскедастичны, существует тесная связь между функцией расстояния J эффективной GMM-оценки и суммой квадратов остатков (SSR). Из (3.8.9'):

$$\begin{aligned} J(\tilde{\delta}, (\hat{\sigma}^2, S_{\text{зм}})^{-1}) &= \frac{(y - Z\tilde{\delta})' P (y - Z\tilde{\delta})}{\hat{\sigma}^2} = \\ &= \frac{y' P y - 2y' P Z \tilde{\delta} - \tilde{\delta}' Z' P Z \tilde{\delta}}{\hat{\sigma}^2} = \\ &= \frac{y' P y}{\hat{\sigma}^2} - \frac{2y' Z \tilde{\delta}}{\hat{\sigma}^2} - \frac{\tilde{\delta}' Z' Z \tilde{\delta}}{\hat{\sigma}^2} \\ &\quad (\text{так как } PZ = Z, \text{ когда } z_i \in \pi_i) \\ &= \frac{(y - Z\tilde{\delta})' (y - Z\tilde{\delta})}{\hat{\sigma}^2} - \frac{y'y - y' P y}{\hat{\sigma}^2} = \\ &= \frac{(y - Z\tilde{\delta})' (y - Z\tilde{\delta})}{\hat{\sigma}^2} - \frac{(y - \hat{y})' (y - \hat{y})}{\hat{\sigma}^2}, \quad (3.8.14) \end{aligned}$$

где $\hat{y} = Py$ — вектор подобранных значений МНК-оценки без ограничений. Поскольку вычитаемое справа слагаемое не зависит от $\tilde{\delta}$, минимизация J сводится к минимизации суммы квадратов остатков (SSR) $(y - Z\tilde{\delta})' (y - Z\tilde{\delta})$. Из этого следует, что

- Эффективной GMM-оценкой является МНК-оценка (что справедливо и без условий гомоскедастичности так как $z_i \in \pi_i$).
- Ограниченной GMM-оценкой с учетом ограничений нулевой гипотезы является МНК-оценка с ограничениями (с SSR, а не J в качестве целевой функции).
- Статистика Вальда, численно равная LR-статистике, может быть рассчитана как нормированная на $\hat{\sigma}^2$ разность SSR с учетом ограничений нулевой гипотезы и без. Данный результат подтверждает получение статистики Вальда в соответствии с принципом отклонения правдоподобия в параграфе 2.6.

Тестирование подмножества условий ортогональности

В параграфе 3.6 нами была представлена C -статистика для тестирования подмножества условий ортогональности. Она использовала две разные эффективные GMM-оценки одного и того же уравнения: одну с использованием полного набора инструментов $\mathbf{z}_1$, другую — с использованием подмножества инструментов $\mathbf{z}_1$. Для получения выражения для этой статистики при условной гомоскедастичности предположим, что $\mathbf{X}_1$ — матрица размера $(n \times K_1)$, i -я строка которой равна $\mathbf{z}'_{i1}$. Поскольку при условной гомоскедастичности эффективной GMM-оценкой является 2SLS-оценка, эти две GMM-оценки задаются выражениями

$$\tilde{\delta} = (\mathbf{Z}'\mathbf{P}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{P}\mathbf{y} \quad \text{с} \quad \mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}', \quad (3.8.16)$$

$$\bar{\delta} = (\mathbf{Z}'\mathbf{P}_1\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{P}_1\mathbf{y} \quad \text{с} \quad \mathbf{P}_1 = \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1. \quad (3.8.17)$$

А C -статистика становится разностью двух статистик Саргана:

$$C = \frac{\tilde{\varepsilon}'\mathbf{P}\tilde{\varepsilon} - \bar{\varepsilon}'\mathbf{P}_1\bar{\varepsilon}}{\hat{\sigma}^2}, \quad (3.8.17)$$

где

$$\tilde{\varepsilon} \equiv \mathbf{y} - \mathbf{Z}\tilde{\delta}, \quad \bar{\varepsilon} \equiv \mathbf{y} - \mathbf{Z}\bar{\delta}, \quad \hat{\sigma}^2 = \frac{\tilde{\varepsilon}'\tilde{\varepsilon}}{n}.$$

Как показано в случае без условной гомоскедастичности, C гарантированно будет неотрицательной при использовании одной и той же матрицы $\tilde{\mathbf{S}}$, что при условной гомоскедастичности эквивалентно использованию одной и той же оценки $\hat{\sigma}^2$, для нормировки как $\tilde{\varepsilon}'\mathbf{P}\tilde{\varepsilon}$, так и $\bar{\varepsilon}'\mathbf{P}_1\bar{\varepsilon}$ (см. (3.8.17)). По утверждению 3.7, C имеет асимптотическое распределение $\chi^2(K - K_1)$.

Вероятно, более популярной является статистика, предложенная [Hansen and Taylor, 1980] и детально изученная [Newey, 1985]. $\bar{\delta}$ асимптотически более эффективна, чем $\tilde{\delta}$, так как использует больше условий ортогональности. Тем самым $\text{Avar}(\bar{\delta}) \leq \text{Avar}(\tilde{\delta})$. Более того, как нам нужно будет показать (аналитическое упражнение 9), при той же гипотезе, гарантирующей асимптотическое распределение χ^2 у C -статистики,

$$\text{Avar}(\bar{\delta} - \tilde{\delta}) = \text{Avar}(\bar{\delta}) - \text{Avar}(\tilde{\delta}). \quad (3.8.18)$$

(Если мы уже сталкивались с тестом Хаусмана в контексте ML , вы можете узнать о нем GMM версию принципа Хаусмана). По утверждению 3.9, $\text{Avar}(\bar{\delta})$ состоятельно оценивается как

$$\overline{\text{Avar}}(\bar{\delta}) \equiv n\bar{\sigma}^2 \cdot (\mathbf{Z}'\mathbf{P}\mathbf{Z})^{-1}. \quad (3.8.19)$$

Похожим образом, $\text{Avar}(\tilde{\delta})$ состоятельно оценивается как

$$\overline{\text{Avar}}(\tilde{\delta}) \equiv n\tilde{\sigma}^2 \cdot (\mathbf{Z}'\mathbf{P}_1\mathbf{Z})^{-1}. \quad (3.8.20)$$

Здесь, как при оценивании C -статистики, используется одна и та же оценка $\hat{\sigma}^2$ для гарантирования неотрицательности статистики. Итоговая оценка $\text{Avar}(\bar{\delta} - \hat{\delta})$ равна:

$$\widehat{\text{Avar}}(\bar{\delta} - \hat{\delta}) = \hat{\sigma}^2 \cdot n \cdot \{ (Z' P_1 Z)^{-1} - (Z' P Z)^{-1} \}. \quad (3.8.21)$$

Хаусман и Тейлор [Hausman and Taylor, 1980] показали, что

- в конечных выборках эта матрица положительно полуопределена (неотрицательно определена), но необязательно невырождена;
- ♦ для любой обобщенной обратной¹ к этой матрице статистика Хаусмана (*Hausman statistic*)

$$H = \sqrt{n}(\bar{\delta} - \hat{\delta})' \{ \hat{\sigma}^2 \cdot n \cdot \{ (Z' P_1 Z)^{-1} - (Z' P Z)^{-1} \} \}^{-1} \sqrt{n}(\bar{\delta} - \hat{\delta}) = \\ = \frac{(\bar{\delta} - \hat{\delta})' [(Z' P_1 Z)^{-1} - (Z' P Z)^{-1}] (\bar{\delta} - \hat{\delta})}{\hat{\sigma}^2} \quad (3.8.22)$$

инвариантна относительно выбора обобщенной обратной и имеет асимптотическое распределение χ^2 с числом степеней свободы, равным $\text{tr}[(K - K_1) \cdot I - s]$, где

$$s = \sum_{j=1}^J z_j' \Omega z_j =$$

= количество регрессоров, используемых в качестве инструментов.

Какова связь между C и H при условной гомоскедастичности? Можно показать (см., [Newey, 1985]), что: если $K - K_1 \leq L - s$, так что H и C имеют одинаковое число степеней свободы, то тогда H и C численно равны. В противном случае эти статистики не равны и имеют разное число степеней свободы.

Частный случай, в котором $K - K_1 \leq L - s$, — это когда z_{j2} является подмножеством z_j , то есть подозрительные инструменты являются регрессорами. В этом случае (3.8.17) и (3.8.22) — два способа рассчитать одну и ту же статистику. В случае $K - K_1 > L - s$ число степеней свободы H меньше, чем у C . По этой причине тест Хаусмана, в отличие от C -теста, не состоятелен против некоторых локальных альтернатив².

¹Обобщенной обратной, A^+ , для матрицы A называется любая матрица, такая, что $AA^+A = A$. Если матрица A квадратная и невырожденная, то A^+ единственное решение A^{-1} .

²Статистику Хаусмана можно обобщить на случай условной гетероскедастичности, но это не применяется на практике, так как степень свободы зависит от неизвестных значений матриц Σ_{xx} и S . (См., [Newey, 1985].)

Тестирование условной гомоскедастичности

Как было показано в параграфе 2.7, для случая МНК с предопределенными регрессорами существует удобный тест условной гомоскедастичности nR^2 . В текущем случае статистика nR^2 , полученная регрессированием квадрата 2SLS остатков на константу и моменты второго порядка инструментальных переменных, не имеет ожидаемого асимптотического распределения. Тестовая статистика, имеющая асимптотическое распределение χ^2 , существует, но чрезвычайно сложна. (См.: [White, 1982].)

Тестирование на наличие автокорреляции

Для случая МНК в параграфе 2.10 были разработаны тесты на наличие автокорреляции в ошибке. Точнее, при выполнении предположения (2.10.15) (которое сильнее, чем предположения 2.3 и 3.3) и (2.10.16) (которое сильнее предположений 2.7 и 3.7 условной гомоскедастичности), модифицированная статистика Бокса — Пирса Q (2.10.20) может быть использована для тестирования нулевой гипотезы об отсутствии автокорреляции в ошибке, и эта статистика асимптотически эквивалентна (в более строгом смысле сходимости разности к нулю по вероятности) статистике nR^2 из регрессии (2.10.21). Может ли этот тест быть распространен на случай эндогенных регрессоров z_t ? Если инструменты x_t удовлетворяют (2.10.15) и (2.10.16), то тогда доказательство в приложении 2.В к главе 2 может быть обобщено для получения модифицированной статистики Q , которая будет иметь асимптотическое распределение χ^2 при нулевой гипотезе об отсутствии автокорреляции. Однако выражение для такой статистики более сложное, чем (2.10.20), и поэтому не представлено здесь. При этом непохоже, чтобы эта модифицированная Q -статистика была асимптотически эквивалентной статистике nR^2 из регрессии (2.10.21).

Контрольные вопросы

1. (GMM с условной гомоскедастичностью.) В эффективном двухшаговом GMM-оценке $\hat{\beta}$ классифицируется на втором шаге до (3.5.10) с использованием остатков первого шага. Выведите асимптотическую дисперсию двухшаговой оценки для случая условной гомоскедастичности. Равна ли она (3.8.4)? **Указание:** При условной гомоскедастичности (3.5.10) $\rightarrow_p \frac{1}{n} \sigma^2 \sum x_{it} x_{it}'$.
2. (2SLS без условной гомоскедастичности.) Состоятельна ли оценка 2SLS при отсутствии условной гомоскедастичности? Выведите plim (3.8.2) и $\text{Avar}(\hat{\beta})$ без предположения об условной гомоскедастичности. Является ли 2SLS столь же эффективным, как и двухшаговый GMM (то есть асимптотическая дисперсия у него так же мала, как и (3.5.13))?

Указание: 2SLS-оценка является GMM-оценкой с выбором $\bar{W}$, и она не обязана быть эффективной при отсутствии условной гомоскедастичности.

3. (GLS-интерпретация 2SLS.) Проверьте, что 2SLS-оценку можно записать как GLS-оценку, если $S_{\text{те}}$ и $y_{2\text{т}}$ интерпретировать как матрицу значений регрессоров и вектор значений зависимой переменной, а $S_{\text{те}}$ — как ковариационную матрицу вектора ошибок.
4. Приведите два примера, в которых $\bar{\delta}_{2SLS}$ и $\bar{\delta}(\bar{S}^{-1})$ асимптотически эквивалентны, то есть

$$\sqrt{n}(\bar{\delta}_{2SLS} - \bar{\delta}(\bar{S}^{-1})) \rightarrow 0$$

Указание: Ключевые слова «условная гомоскедастичность» и «точно идентифицируема».

5. Предположим, что уравнение идентифицируемо точно. Покажите, что 2SLS (3.8.3) сводится к IV (3.4.4).
6. (Статистика Саргана как R^2 .) Докажите, что статистика Саргана (3.8.10') равна $nR^2_{\text{ин}}$, где $R^2_{\text{ин}}$ — нецентрированный коэффициент детерминации из регрессии $\bar{e}$ на X . **Указание:** Контрольный вопрос 8 из параграфа 1.3.
7. ($\mathcal{X}$ как строгое подмножество $\mathcal{X}_1$.) Предположим, что $\mathcal{X}$ является строгим подмножеством $\mathcal{X}_1$. То есть $\mathcal{X}$ содержит в дополнение к предопределенным регрессорам несколько других предопределенных переменных. Мы показали, что эффективная GMM-оценка является МНК при условной гомоскедастичности. Сохранится ли этот результат без условной гомоскедастичности? **Указание:** Контрольный вопрос 5 из параграфа 3.3.

3.9. Применение: отдача от образования

Начиная с работы [Mincer, 1958] связь между уровнем заработной платы и образованием была объектом большого количества теоретических и эмпирических исследований. Такое внимание может показаться странным, так как объяснение положительного влияния выглядит очевидным: образование увеличивает производительность труда индивида. Однако есть и другие объяснения. Например, модель сигналов на рынке труда Стенси утверждает, что более образованные кандидаты получают большую заработную плату исключительно по причине того, что образование используется в качестве сигнала о более высоких способностях; несмотря на то что образование не увеличивает вероятность более высокого заработка, между этими показателями есть корреляция по причине действия из них третьей переменной — способностей. В данном параграфе показано, как отделить влияние образования от влияния способностей с помощью методик на данной главе. Одной из работ, посвященных этой проблеме, является работа Гринхаса [Griliches, 1976].

Данные NLS-Y

Данные, использованные Гринхиссом, — Young Man's Cohort из National Longitudinal Survey (NLS-Y). Опись в данной когорте проводился в 1966 году среди 5225 респондентов: мальчики в 14–24 лет и был повторен через один и через два года. К 1969 году радио-петлевая панельной выборки было потеряно, но 2026 респондентов были (или уровень своей зарплаты в 1969 году и имели достаточно хорошие данные для выбора необходимых для исследования переменных. Предельной проблемой этих данных было включение в них двух показателей способностей. Первый — оценка в тесте *Knowledge of the World of Work*, проведенной респондентами в 1966 году. Второй — уровень IQ. Все участники прошли, закончившие 9-й класс в 1966 году, подписывали документ, разрешающий администрации школ передавать организаторам теста данные по успеваемости. Итоговый опрос, проведенный в 1968 году, принял данные по установленным способностям респондентов, структурированные в аналогии уровня IQ. Из указанных 2026 респондентов уровень IQ имеется только для 1382, что говорит о том, что итоговый опрос покрывает только две трети исходной выборки. Наше обсуждение касается результатов Гринхисса, основанных на этой меньшей выборке. В табл. 3.1 приведены средние и стандартные отклонения ключевых переменных:

Таблица 3.1

Характеристики респондентов из NLS-Y

Средние и стандартные отклонения (в скобках)

Переменная	
Размер выборки	1,362
Возраст в 1966 году	22.3 (1.2)
Образование в 1966 году (6)	12.5 (1.0)
Потери заработной платы (в пентах) в 1968 году (2W)	3.08 (9.16)
Оценка в тесте <i>Knowledge of the World of Work</i> (KWW)	45.1 (10.0)
Уровень IQ (IQ)	97.7 (15.3)
Возраст в 1969 году	23.7 (2.6)

Источник: [G. Jones, 1976, Table 1].

Полупартиципационные уравнения заработной платы

В литературе обычно оценивается следующее уравнение для заработной платы:

$$\ln W = \alpha + \beta S + \gamma A + \delta' h + \varepsilon, \quad (3.9.1)$$

где $\ln W$ — логарифм заработной платы индивида, S — образование в годах, A — мера способностей, h — вектор наблюдаемых характеристик индивида (например, опыт работы и место жительства), δ — соответствующий вектор коэффициентов, ε — независимая ошибка с нулевым математическим ожиданием (в данном параграфе индекс индивида i будет опускаться для упрощения записи). Полупартиципационная спецификация для S часто оправдывается линейной связью между логарифмом заработной платы и образованием, установленной в качестве эмпирического факта на больших поперечных данных (например, Census Population Survey)¹.

Коэффициент β показывает процентное изменение заработной платы при увеличении срока образования индивида на один год. Тем самым он показывает «предельную отдачу» от инвестирования в человеческий капитал, которая должна быть величиной того же порядка, что и доходность от инвестирования в финансовые активы. Мы предполагаем, что переменные S , A , h не коррелируют с ошибкой регрессии ε ; поэтому МНК является подходящим методом оценки при включении способностей A в модель наряду с образованием S и h . В оставшейся части параграфа мы рассмотрим то, что Грэхам назвал «смещением из-за способностей», — смещение МНК-оценки коэффициента β при включении A в уравнение или включение вместо нее какой-то неполноценной меры способностей.

Смещение из-за исключения переменной

Иногда наборы данных не включают в себя меры способностей (например, Census Population Survey). Каковы последствия игнорирования и исключения способностей в уравнение заработной платы? Из параграфа 2.9 нам известно, что регрессия $\ln W$ на константу, S и h даст относительно точную оценку соответствующей МНК-проекции:

$$\begin{aligned} \hat{E}^1(\ln W | 1, S, h) &= \hat{E}^1(\alpha + \beta S + \gamma A + \delta' h + \varepsilon | 1, S, h) \quad (\text{из (3.9.1)}) = \\ &= \alpha + \beta S + \delta' h + \gamma \hat{E}^1(A | 1, S, h) + \hat{E}^1(\varepsilon | 1, S, h). \end{aligned} \quad (3.9.2)$$

Поскольку все регрессоры в (3.9.1) предопределены, имеем $E(\varepsilon) = 0$, $E(S\varepsilon) = 0$ и $E(h\varepsilon) = 0$. Поэтому $\hat{E}^1(\varepsilon | 1, S, h) = 0$. Если обозначить

¹ Вироста функциональной формы рассматриваются в [Card, 1995].

$\hat{E}^*(A|1, S, h) = \theta_1 + \theta_S S + \theta_h' h$, то (3.9.2) преобразуется в

$$\hat{E}(LW|1, S, h) = (\alpha + \gamma\theta_1) + (\beta + \gamma\theta_S)S + (\delta + \gamma\theta_h)'h.$$

Тем самым МНК-оценка коэффициента может оказаться асимптотически смещенной для всех включенных регрессоров. Это явление известно как **смещение из-за не включения переменных** (*omitted variable bias*). В частности,

$$\text{plim } \hat{\beta}_{OLS} = \beta + \gamma\theta_S.$$

То есть МНК-оценка коэффициента при образовании S включает непрямой эффект способностей через образование ($\gamma\theta_S$) наряду с прямым эффектом β . Если θ_S положительна, то оценка коэффициента $\hat{\beta}_{OLS}$ смещена вверх.

Используя описанную выше выборку из NLS-Y, состоящую из 1362 индивидов, Гримальес оценил уравнение заработной платы для 1969 года. Список переменных, включенных в h , здесь представлен не будет; достаточно сказать, что он включает в себя опыт работы в годах и некоторые демографические переменные для региона и размера города. Оценка Гримальеса для коэффициента при уровне образования S в случае игнорирования способностей представлена в строке 1 табл. 3.2. Оценка смотрится хорошо: эта точечная оценка похожа на доходность, которую можно получить от вложения в финансовый актив; причем эта оценка достаточно точна, о чем говорит высокое значение t -статистики.

IQ как мера способностей

Как уже было сказано, в NLS-Y представлены две меры способностей: IQ и KWW (оценка в тесте *Knowledge of the World of Work* в 1966 году). Поскольку респонденты NLS-Y имели в 1966 году возраст не менее 14 лет, результаты данного теста отражают влияние уже полученного на тот момент образования и не могут служить мерой чистых способностей. Уровень IQ не имеет этого недостатка. Если бы оценка IQ была идеальной мерой способностей индивида, то можно было бы подменить состоятельную оценку уравнения (3.9.1) простой подстановкой IQ вместо A . МНК-оценки Гримальеса при включении IQ в модель представлены в строке 2 табл. 3.2. Теперь коэффициент при образовании ниже, что подтверждает наше предположение о том, что коэффициент при образовании при не включении в модель меры способностей включает в себя также влияние способностей на размер заработной платы.

Таблица 3.2

Оценки коэффициентов

Уравнение: $LMV = \alpha + \beta S + \gamma IQ + \text{прочие переменные}$

Строка	Метод оценки	Коэффициент		SER	R^2	Эндогенные регрессоры	Исключенные предопределенные регрессоры
		S	IQ				
1	MHK	0,065 (13,2)		0,332	0,309	нет	
2	MLK	0,059 (10,7)	0,0019 (2,8)	0,331	0,313	нет	
3	2SLS	0,062 (7,0)	0,0038 (2,4)	0,332		IQ	MED, KWW, возраст, возраст в квадрате, общие характеристики

Источники: строка 1 — уравнение (B1) в: [Gizilichens, 1976, Table 2];

строка 2 — уравнение (H4) в табл. 2 там же;

Ошибки в переменных

Разумеется, значение IQ не может быть безлинейной мерой способности. Если η — ошибка измерения, то IQ связан с A как

$$IQ = \varphi + A + \eta, \quad (3.9.3)$$

с $E(\eta) = 0$. Интерпретируя η как ошибку измерения, позволяет предположить, что она не коррелирует с A , S , h и ошибкой регрессии ε . Подставляя (3.9.3) в (3.9.1), получаем:

$$LW = (\alpha - \gamma\varphi) + \beta S + \gamma IQ + \delta h + (\varepsilon - \gamma\eta). \quad (3.9.4)$$

Проиллюстрируем теперь на этом примере, что если хотя бы одна из включенных в модель переменных измерена с ошибкой, то оценки коэффициентов у всех переменных могут быть асимптотически смещенными.

Для изучения последствий включения в модель измеренной с ошибкой меры IQ рассмотрим следующую МНК-проекцию:

$$\begin{aligned} \hat{E}^*(LW|1, S, IQ, h) &= (\alpha - \gamma\varphi) + \beta S + \gamma IQ + \delta h \\ &= \hat{E}^*(\varepsilon|1, S, h) + \gamma \hat{E}^*(\eta|1, S, IQ, h). \end{aligned} \quad (3.9.5)$$

Рассмотрим $\hat{E}^*(\varepsilon|1, S, IQ, h)$ в (3.9.5). Поскольку $E(\varepsilon) = 0$, и S и h не коррелированы с ε , $(1, S, h)$ ортогональны ε , IQ также ортогонален ε , так как

$$\begin{aligned} E(IQ\varepsilon) &= E[(\varphi + A + \eta)\varepsilon] \quad (\text{из (3.9.3)}) = \\ &= E(\eta\varepsilon) \quad (\text{поскольку } E(\varphi) = 0, \text{Cov}(A, \varepsilon) = 0) = \\ &= 0. \end{aligned}$$

Тем самым $\hat{E}^*(\varepsilon|1, S, IQ, h) = 0$ в (3.9.5).

Таким образом, смещение, если таковые есть, равно $-\gamma \hat{E}^*(\eta|1, S, IQ, h)$ в (3.9.5). Пусть $(\beta_S, \beta_{IQ}, \beta_h)$ — коэффициенты проекции (S, IQ, h) в $\hat{E}^*(\eta|1, S, IQ, h)$. По формуле (2.9.7) имеем 2

$$\begin{bmatrix} \beta_S \\ \beta_{IQ} \\ \beta_h \end{bmatrix} = \begin{bmatrix} \text{Var}(S) & \text{Cov}(S, IQ) & \text{Cov}(S, h) \\ \text{Cov}(IQ, S) & \text{Var}(IQ) & \text{Cov}(IQ, h) \\ \text{Cov}(h, S) & \text{Cov}(h, IQ) & \text{Var}(h) \end{bmatrix}^{-1} \begin{bmatrix} \text{Cov}(S, \eta) \\ \text{Cov}(IQ, \eta) \\ \text{Cov}(h, \eta) \end{bmatrix}. \quad (3.9.6)$$

В этом выражении $\text{Cov}(S, \eta)$ и $\text{Cov}(h, \eta)$ известны (по предположению). Однако $\text{Cov}(IQ, \eta)$ не имеет значения, так как

$$\begin{aligned} \text{Cov}(IQ, \eta) &= E(IQ\eta) \quad (\text{поскольку } E(\eta) = 0) = \\ &= E[(\varphi + A + \eta)\eta] \quad \text{из (3.9.3)} = \\ &= E(\eta^2) = E(A\eta) + E(\eta^2) \\ &= \text{Var}(\eta) \quad (\text{так как } E(\eta) = 0 \text{ и } \text{Cov}(A, \eta) = 0). \end{aligned}$$

То есть ошибка измерения, не коррелируя с истинным значением, с неоднородностью коррелирует с измеренным значением. С учетом того, что

$$\text{Cov}(S, \eta) = 0, \quad \text{Cov}(IQ, \eta) = \text{Var}(\eta), \quad \text{и} \quad \text{Cov}(h, \eta) = 0,$$

коэффициенты регрессии можно записать как

$$\begin{bmatrix} \theta_S \\ \theta_{IQ} \\ \theta_h \end{bmatrix} = \text{Var}(\eta) \cdot \alpha,$$

где

$$\alpha = \text{второй столбец} \begin{bmatrix} \text{Var}(S) & \text{Cov}(S, IQ) & \text{Cov}(S, h) \\ \text{Cov}(IQ, S) & \text{Var}(IQ) & \text{Cov}(IQ, h) \\ \text{Cov}(h, S) & \text{Cov}(h, IQ) & \text{Var}(h) \end{bmatrix}^{-1}.$$

Тем самым в регрессии LW на константу, S , IQ и h

$$\text{plim} \hat{\beta}_{OLS} = \beta + \gamma \cdot \text{Var}(\eta) \cdot (1\text{-й элемент } \alpha), \quad (3.9.7a)$$

$$\text{plim} \hat{\beta}_{OLS} = \gamma + \gamma \cdot \text{Var}(\eta) \cdot (2\text{-й элемент } \alpha) \quad (3.9.7b)$$

Поскольку второй элемент α положительный (это диагональный элемент матрицы, обратной к ковариационной матрице), МНК-оценка коэффициента при способности будет смещена вниз. Однако направление смещения коэффициента при образовании зависит от знака первого элемента α . Обычно $\text{Cov}(S, IQ)$ положительна, так что, за исключением незначительно сильного корреляции (S, IQ) с h , первый элемент α будет отрицательным. Тем самым можно предположить, что $\hat{\beta}_{OLS}$ для образования будет смещена вверх.

Коррекция смещения с помощью 2SLS

Для учета смещения Грисхекс применяет к уравнению заработной платы 2SLS. Чтобы сделать это, следует указать список инструментов. В набор инструментов можно включить предопределенные регрессоры (1. S , h). Дополните данные инструменты для IQ , включив в набор: — возраст, возраст в квадрате, $LHWH$, образовательный уровень и некоторые другие характеристики человека (например, зачисленность в вуз). Эти переменные также рассматриваются как предопределенные. 2SLS-оценка Грисхекса для этой спецификации приведена в строке 3 табл. 3.2¹. В соответствии с нашим предсказанием того, что МНК-оценка для коэффициента при мере способностей смещена вниз, МНК-оценка из строки 2, равная 0,0019, меньше, чем приведенная в строке 3 2SLS-оценка, равная 0,0038. Наше

¹ R^2 не приведено, так как сумму квадратов ошибок в случае 2SLS, в отличие от случая МНК, нельзя разделить между «объясненной» и «необъясненной» изменчивостью.

приходит к выводу о том, что оценка β_{OLS} коэффициента при объяснении при включении IQ смещена вверх, также подтверждается данными, так как оценка β , равная 0,639, в строке 2 выше, чем оценка 0,052 в строке 3.

В итоге «смещение из-за способностей», исследуемое Гриндхесом, заключается в смещении оценки коэффициента при объяснении различий в заработках IQ и смещении ее в неизвестном направлении (вероятно, вверх) при неточном измерении объясненной переменной (в данном случае IQ). 2SLS даст решение, но этот метод зависит от предположения, что использованные инструменты (такие как MED и AFW) не коррелируют с ненаблюдаемыми детерминантами заработной платы ε и q . Эти детерминанты могут включать в себя такие персональные характеристики, как умения. Представляется небезосновательным предположить, что AFW зависит от таких характеристик. Если это так, то AFW не может быть инструментом. Если пойти дальше, то можно утверждать, что образованные матери также непредопределенны потенциалом, так как некоторые характеристики матери, влияющие на уровень ее образования и меняющиеся на рынке труда, могут быть несогласованы детьми. Не являются ли данные персональные предопределенными, можно спорить с использованием статистики Сарджина, но в статье Гриндхеса эти статистики не приведены (и сама работа написана задолго до того, как это стало общей практикой).

Дальнейшее развитие

Возможно, из-за того, что выбор инструментов, коррелированных с IQ , но не коррелированных с ненаблюдаемыми характеристиками, является сложной задачей, учет «смещения из-за способностей» продолжает привлекать большое внимание в литературе. Существует довольно много литературы, начиная с [Weissman et al., 1980], в которой производится сравнение однояйцевых близнецов с разным уровнем образования для учета генетических характеристик и влияния семьи. Можно также рассматривать одного и того же человека в разные моменты времени. Необходимой для объяснения таких случаев является, однако, фиксированными эффектами, будет рассмотрена в главе 5.

Другой важной проблемой является эндогенность уровня образования S . Если предположить, что ошибка включает в себя набор ненаблюдаемых характеристик индивиду, которые могут повлиять на выбор уровня образования, то S уже следует рассматривать как эндогенную переменную. Но, опять же, выбор подходящих инструментов для S , не коррелированных с ненаблюдаемыми характеристиками, но коррелированных с уровнем образования, очень сложен. Последние работы в этой области можно рассматривать как поиск детерминантов уровня образования, имеющих мало общего с персональными и ненаблюдаемыми

характеристикам. Например, [Angrist and Krueger, 1991] не обнаружили переменную, показывающую, воспользовались ли индивиды законом об обязательном школьном образовании.

Наконец, снова привлекают внимание работы, посвященные связи между качеством школы и заработком (см., например, [Card and Krueger, 1995]).

Контрольные вопросы

1. Перечислите все предположения о средних и ковариациях (S.A.IQ), AICN и A...), сделанные в данном параграфе. Научите, как Гринман формулирует эти предположения. Какие из них являются ключевыми для состоятельности 2SLS-оценки?
2. Предположим, что зависимость между IQ и A задается (3.9.3), а выражением

$$IQ = \beta + \pi A + \eta.$$

Остается ли состоятельной 2SLS-оценка коэффициента β при уровне образования? (Ответ: Да.)

3. В модели сигналов на рынке труда Спенкс индивиды разделены на две группы: малоспособных, выбирающих низкий уровень образования, и очень способных, выбирающих высокий уровень образования. Образование не увеличивает трудоспособность индивидов, так что размер заработной платы определяется только способностями индивида. Если бы мы имели данные по заработной плате, способностям и образованию для выборки, включающей индивидов обеих типов, могли бы мы точно проверить гипотезу о том, что уровень образования сам по себе не приводит к более высокой заработной плате? Указание: Возможно наличие проблемы корреляции

Набор задач для главы 3

Аналитические упражнения

1. Докажите, что симметричная и идемпотентная матрица положительно полуопределена.
2. Мы хотим доказать утверждение 3.4. Чтобы избежать ненужных осложнений, рассмотрим случай $K = 1$ и $L = 1$ (так что π_i и δ_i скаляры). То есть (3.5.5) принимает вид:

$$\tilde{z}_i^2 = z_i^2 - 2(\tilde{\delta} - \delta)z_i + (\tilde{\delta} - \delta)^2 z_i^2,$$

а $\tilde{S}$ в (3.5.10) принимает вид:

$$\frac{1}{n} \sum_{i=1}^n z_i^2 z_i^2 - 2(\tilde{\delta} - \delta) \frac{1}{n} \sum_{i=1}^n z_i^2 z_i + (\tilde{\delta} - \delta)^2 \frac{1}{n} \sum_{i=1}^n z_i^2 z_i^2. \quad (*)$$

- (a) Покажите, что первая составляющая в (*) сходится по вероятности к S .
- (b) Используя неравенство Коши — Шварца

$$E(|X \cdot Y|) \leq \sqrt{E(X^2) \cdot E(Y^2)},$$

покажите, что $E(z_1 z_1' z_1)$ существует и конечен. **Указание:** $z_1 z_1' z_1$ является произведением $z_1 z_1'$ и $z_1 z_1$.

Факт: $E(\cdot)$ существует и конечно тогда и только тогда, когда $E(\|x\|) < \infty$. Где используется предположение 3.6²

- (c) Покажите, что вторая составляющая в (*) сходится к нулю по вероятности.
- (d) Покажите, что третья составляющая в (*) сходится к нулю по вероятности.
3. Мы хотим доказать асимптотический результат (3.5.11). Для этого требуется доказать, что $A - B$ является положительно полуопределенной матрицей, где

$$A = (\Sigma_{xx}' W \Sigma_{xx})^{-1} \Sigma_{xx}' W S W \Sigma_{xx} (\Sigma_{xx}' W \Sigma_{xx})^{-1}$$

и

$$B \equiv (\Sigma_{xx}' S^{-1} \Sigma_{xx})^{-1}.$$

Из матричной алгебры известно, что если A и B положительно определенные и симметричные матрицы, то $A - B$ будет положительно полуопределенной тогда и только тогда, когда $B^{-1} - A^{-1}$ положительно полуопределена. Тем самым нужно доказать, что

$$Q \equiv \Sigma_{xx}' S^{-1} \Sigma_{xx} - \Sigma_{xx}' W \Sigma_{xx} (\Sigma_{xx}' W S W \Sigma_{xx})^{-1} \Sigma_{xx}' W \Sigma_{xx}$$

является положительно полуопределенной.

- (a) Поскольку S ($K \times K$) положительно определена, то существует невырожденная матрица C ($K \times K$) такая, что $C'C = S^{-1}$. Проверьте, что Q можно записать в виде:

$$Q = H' M_C H,$$

где

$$H = C \Sigma_{xx}, \quad M_C = I_K - G(G'G)^{-1}G', \quad G = C^{-1} W \Sigma_{xx}.$$

Указание: $C^{-1}C^{-1} = S$.

(д) Покажите, что Q положительно полуопределена. **Указание:** Сначала покажите, что M_Q симметрична и идемпотентна. Как было показано в упражнении 1, симметричная идемпотентная матрица положительно полуопределена.

4. Предположим, что $\mathbf{y}$ является строгим единичством $\mathbf{x}$. Что можно сказать о матричном смещении асимптотической дисперсии МНК-оценки или асимптотической дисперсии эффективной двухшаговой ГММ-оценки? **Указание:** Без потери общности можно предположить, что $\mathbf{x}_1$ — первые L элементов K -мерного вектора $\mathbf{x}$. Если матрица $\mathbf{W}$ ($K \times K$) дана как

$$\mathbf{W} = \begin{bmatrix} \mathbf{A} & \mathbf{a} \\ \mathbf{0} & 0 \end{bmatrix},$$

где

$$\mathbf{A} \equiv E(x_1^2 x_1 x_1') - \text{квадратная } (L \times L) \text{ подматрица матрицы } S,$$

то асимптотическая дисперсия МНК-оценки может быть записана как в формуле (3.5.1). Если вы сделали упражнение 3, то можете проверить, что (3.5.1) не требует обязательной определенности $\mathbf{W}$ если $\Sigma'_{xx} \mathbf{W} \Sigma_{xx}$ невырождена.

5. (Обязательное упражнение) Докажите утверждение 3.6, предположив следующие факты:

(а) Покажите, что $\hat{\beta}_n(\hat{S}(\hat{S}^{-1})) = \hat{B}\hat{\beta}$, где

$$\hat{B} \equiv \mathbf{I}_K - S_{xx}(S'_{xx}\hat{S}S_{xx})^{-1}S'_{xx}\hat{S}^{-1}.$$

(б) Поскольку $\hat{S}$ положительно определен, существует невырожденная матрица $\mathbf{C}$ ($K \times K$), такая, что $\mathbf{C}'\mathbf{C} = \hat{S}^{-1}$. Пусть $\mathbf{A} = \mathbf{C}S_{xx}$. Покажите, что $\hat{B}'\hat{S}^{-1}\hat{B} = \mathbf{C}'\mathbf{M}\mathbf{C}$, где $\mathbf{M} = \mathbf{I}_K - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$. Какой ранг у $\mathbf{M}$? **Указание:** Ранг идемпотентной матрицы равен ее следу.

(в) Пусть $\mathbf{v} \equiv \sqrt{n}\mathbf{C}\hat{\beta}$. Покажите, что $\mathbf{v} \rightarrow_d N(0, \mathbf{I}_K)$.

(г) Покажите, что $\mathbf{A}'\hat{B}'(\hat{S}^{-1})\hat{S}^{-1} = \mathbf{v}'\mathbf{M}\mathbf{v}$. Каково его асимптотическое распределение?

6. Покажите, что I -статистика из утверждения 3.6 численно равна:

$$n \cdot \mathbf{e}'_{xy} \hat{B}' \hat{S}^{-1} \hat{B} \mathbf{e}_{xy},$$

где $\hat{B}$ определена в упражнении 5а). **Указание:** Из упражнения 5 следует $\mathbf{I} = n\hat{\beta}'\hat{B}'\hat{S}^{-1}\hat{B}\hat{\beta}$. Покажите, что $\hat{B}\hat{\beta} = \hat{B}\mathbf{e}_{xy}$.

7. (Необязательное упражнение.) Докажите утверждение 3.7, применив следующие шаги. Обязательны, не включая в этом упражнении, те же самые, что и в упражнении 5.

(a) Докажите, что $g_{11}(\tilde{\theta}) = \tilde{B}_1 \tilde{g}_1 + \tilde{B}_1'(\tilde{S}_{11})^{-1} \tilde{B}_1 = C_1' M_1 C_1$,

$$\text{где } \tilde{g}_1 \equiv \frac{1}{n} \sum_{i=1}^n w_i \cdot \varepsilon_i,$$

$$\tilde{B}_1 \equiv I_{K_1} - S_{xx} S_{x\varepsilon}' (\tilde{S}_{11})^{-1} S_{x\varepsilon} S_{x\varepsilon}' (\tilde{S}_{11})^{-1},$$

$$C_1' C_1 \equiv (\tilde{S}_{11})^{-1},$$

$$M_1 \equiv I_{K_1} - A_1 (A_1' A_1)^{-1} A_1',$$

$$A_1 \equiv C_1 S_{x,\varepsilon}$$

(b) Докажите, что $\text{rank}(M_1) = K - L$, $\text{rank}(\tilde{M}_1) = K_1 - L$.

(c) В предыдущем упражнении было доказано, что $J = v' M v$, где $v = \sqrt{n} C g$. Покажите, что $J_1 = v_1' M_1 v_1$, где $v_1 = \sqrt{n} C_1 \tilde{g}_1$.

Для продолжения определим F ($K \times K_1$) как

$$F = \begin{bmatrix} I_{K_1} \\ 0 \end{bmatrix} \begin{cases} K_1 \text{ строк,} \\ K - K_1 \text{ строк.} \end{cases}$$

$$\text{Тогда } \pi_{11} = F' \pi_1, S_{\alpha,\varepsilon} = F' S_{\alpha\varepsilon}, \tilde{g}_1 = F' \tilde{g}_1.$$

(d) Докажите, что $J = J_1 + v'(M - D)v$, где

$$D = C'^{-1} F C_1' M_1 C_1 F' C^{-1}.$$

(e) Докажите, что D симметрична и идемпотентна с рангом $K_1 - L$.

Указание:

$$F' C^{-1} C'^{-1} F = F' (C C')^{-1} F = F' \tilde{S} F = \tilde{S}_{11},$$

$$C_1' \tilde{S}_{11} C_1' = C_1' (C_1' C_1)^{-1} C_1' = I_{K_1}.$$

(f) Докажите, что $A'D = 0$. Указание:

$$A'D = (S_{xx}' C') (C^{-1} F C_1' M_1 C F' C^{-1}) = (S_{xx}' F C_1' M_1) (C_1 F' C^{-1}), \\ S_{xx}' F C_1' = A_1'.$$

так что

$$S_{xx}' F C_1' M_1 = A_1' M_1.$$

(g) Докажите, что $M - D$ симметрична и идемпотентна с рангом $K - K_1$.

- (h) Докажите желаемый результат. $J = J_1 \rightarrow_1 \chi^2(K - K_1)$.
- (i) На шаге d) устанавливается, что $C = \alpha \tilde{g}' C' (M - D) C \tilde{g}$. Покажите, что C можно переписать в виде:

$$C = S'_{xy} C' (M - D) C S_{xy}.$$

Указание: $g_n(\tilde{\delta}) = \tilde{B} S_{xy}$. Таким образом, численное значение C было бы таким же, если бы мы изначально заменили $\tilde{g}$ на S_{xy} и α .

- (j) Покажите, что $M - D$ положительно полуопределена. Поэтому квадратичная форма C будет нестрогосохраняющей на любой выборке.
8. (Необязательное упражнение, доказательство следующей эквивалентности и утверждения 3.8.) В эффективной GMM с ограничениями функция J минимизируется при ограничении $H\tilde{\delta} = \tau$. Сейчас же будем требовать, чтобы $\tilde{W}$ была равной $\tilde{S}^{-1}$, в записанном Лагранжиане виде.

$$L = \lambda \cdot (S_{xy} - S_{xz}\tilde{\delta})' \tilde{W} (S_{xy} - S_{xz}\tilde{\delta}) + \lambda' (R\tilde{\delta} - \tau),$$

где λ — $\frac{m}{2}$ -мерный вектор множителей Лагранжа (векминимум τ имеет размер $\frac{m}{2} \times 1$, $H = \frac{m}{2} \times L$, $\tilde{\delta} = L \times 1$). Пусть $\tilde{\delta}$ — решение задачи минимизации с ограничениями (если $\tilde{W} = \tilde{S}^{-1}$, то $\tilde{\delta}$ — эффективная GMM-оценка).

- (a) Пусть $\tilde{\delta}(\tilde{W})$ — GMM-оценка без ограничений, минимизирующая $J(\tilde{\delta}, \tilde{W})$. Покажите, что

$$\begin{aligned} \tilde{\delta} &= \tilde{\delta}(\tilde{W}) = (S'_{xz} \tilde{W} S_{xz})^{-1} H' (R' S'_{xz} \tilde{W} S_{xz})^{-1} R^{-1} [R\tilde{\delta}(\tilde{W}) - \tau], \\ \lambda &= \partial \lambda \cdot [R' S'_{xz} \tilde{W} S_{xz})^{-1} H^{-1} [R\tilde{\delta}(\tilde{W}) - \tau]. \end{aligned}$$

Указание: Условия первого порядка имеют вид:

$$-2\alpha S'_{xz} \tilde{W} S_{xz} + 2\lambda (S'_{xz} \tilde{W} S_{xz}) \tilde{\delta} - R' \lambda = 0$$

Совмещая их с ограничением $H\tilde{\delta} = \tau$, получим линейную систему относительно λ и $\tilde{\delta}$.

- (b) Покажите, что

$$J(\tilde{\delta}, \tilde{W}) = J(\tilde{\delta}(\tilde{W}), \tilde{W}) + \tau \cdot (\tilde{\delta}(\tilde{W}) - \tilde{\delta})' (S'_{xz} \tilde{W} S_{xz}) (\tilde{\delta}(\tilde{W}) - \tilde{\delta}).$$

Указание:

$$\begin{aligned} J(\tilde{\delta}, \tilde{W}) &= \alpha \cdot (S_{xy} - S_{xz}\tilde{\delta})' \tilde{W} (S_{xy} - S_{xz}\tilde{\delta}) = \\ &= \alpha \cdot \left[(S_{xy} - S_{xz}\tilde{\delta}(\tilde{W})) + S_{xz}(\tilde{\delta}(\tilde{W}) - \tilde{\delta}) \right]' \tilde{W} \times \\ &\times \left[(S_{xy} - S_{xz}\tilde{\delta}(\tilde{W})) + S_{xz}(\tilde{\delta}(\tilde{W}) - \tilde{\delta}) \right]. \end{aligned}$$

Используйте условия первого порядка для GMM без ограничений:

$$S'_{xx}\tilde{W}(\sigma_{xy} - S_{xx}\delta(\tilde{W})) = 0.$$

- (с) Пусть $\tilde{W} = \hat{S}^{-1}$. Покажите, что статистика Вальда W (3.5.8) имеет пределную LH (3.7.2).

9. (Принцип Хаусмана в GMM.) Пусть $\hat{\delta}_1 = \hat{\delta}(\tilde{W}_1)$ и $\hat{\delta}_2 = \hat{\delta}(\tilde{W}_2)$ — две GMM-оценки с разными взвешивающими матрицами $\tilde{W}_1$ и $\tilde{W}_2$.

- (а) Покажите, что

$$\begin{bmatrix} \sqrt{n}(\hat{\delta}_1 - \delta) \\ \sqrt{n}(\hat{\delta}_2 - \delta) \end{bmatrix} \rightarrow N \left(0, \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \right),$$

где

$$A_{11} = Q_1^{-1} \Sigma'_{xx} W_1 S W_1 \Sigma_{xx} Q_1^{-1},$$

$$A_{12} = Q_1^{-1} \Sigma'_{xx} W_1 S W_2 \Sigma_{xx} Q_1^{-1},$$

$$A_{21} = Q_2^{-1} \Sigma'_{xx} W_2 S W_1 \Sigma_{xx} Q_2^{-1},$$

$$A_{22} = Q_2^{-1} \Sigma'_{xx} W_2 S W_2 \Sigma_{xx} Q_2^{-1},$$

$$Q_1 = \Sigma'_{xx} W_1 \Sigma_{xx},$$

$$Q_2 = \Sigma'_{xx} W_2 \Sigma_{xx},$$

$$\lim_{n \rightarrow \infty} \tilde{W}_j = W_j \quad (j = 1, 2)$$

- (b) Пусть $q = \hat{\delta}_1 - \hat{\delta}_2$. Покажите, что $\sqrt{n}q \rightarrow_d N(0, \text{Avar}(q))$, где

$$\text{Avar}(q) = A_{11} + A_{22} - A_{12} - A_{21}.$$

- (с) Пусть $\tilde{W}_2 = \hat{S}^{-1}$, так что $\hat{\delta}_2$ — эффективная GMM-оценка. Покажите, что

$$\text{Avar}(q) = A_{11} - (\Sigma'_{xx} S^{-1} \Sigma_{xx})^{-1} \cdot \text{Avar}(\hat{\delta}(\tilde{W}_1)) - \text{Avar}(\hat{\delta}(\hat{S}^{-1})).$$

10. (Слабые инструменты, вопрос, восходящий к M. Уотсону.) Рассмотрим модель

$$y_i = \pi_1 x_i + \pi_2 \quad (i = 1, 2, \dots, n),$$

где y_i, π_1 и π_2 — скалярные случайные величины. [Пусть x_i — скалярная инструментальная переменная, связанная с z_i следующим уравнением:

$$z_i = \pi_1 x_i + \pi_2.$$

Пусть

$$\eta_i = \begin{bmatrix} z_i \\ y_i \end{bmatrix}, \quad \theta_i = \eta_i - \pi_i \equiv \begin{bmatrix} \eta_{1i} \\ \eta_{2i} \end{bmatrix} \equiv \begin{bmatrix} \pi_1 x_i \\ \pi_1 x_i + \pi_2 \end{bmatrix}.$$

Предположим, что

- (1) $\{x_t, g_t\}$ — эргодический стационарный процесс.
- (2) g_t — последовательность мартингал-разностей с положительно определенной матрицей $E(g_t g_t') = S$.
- (3) $E(x_t^2) = \sigma_x^2 > 0$.
- (4) $\beta \neq 0$.

Пусть $\tilde{\beta} \equiv x_{t+1}' x_t x_t' / \sum_{t=1}^n$

$$\hat{\alpha}_{x1} = \frac{1}{n} \sum_{t=1}^n x_t x_t', \quad \hat{\alpha}_{x2} = \frac{i}{n} \sum_{t=1}^n x_t x_t'$$

- (a) Докажите, что $\sigma_{x2} = E(x_t x_t') \neq 0$ (то есть ранговое условие идентификации выполнено).
- (b) Докажите, что $\tilde{\beta} \rightarrow_p \beta$.

Мы хотим сейчас рассмотреть случай, когда ранговое условие удовлетворяется «визуально», то есть σ_{x2} очень близко к нулю. Одним из способов добиться этого является замена предположения 4) на 4') $\beta = 1/\sqrt{n}$.

В оставшейся части этой задачи будем считать, что выполнены предположения (1)–(3) и 4') (это предположение и заключение попросту оспаривался [Sargent and Stock, 1997]). Замечим, что $\tilde{\beta} - \beta = x_{t+1}' \hat{\beta}_1$, где $\hat{\beta}_1$ — выборочное среднее $g_t = x_t g_t$.

- (c) Покажите, что $\hat{\alpha}_{x2} \rightarrow_p 0$.
- (d) Покажите, что $\sqrt{n} \hat{\alpha}_{x2} \rightarrow_d \sigma_x^2 + a$, где a имеет распределение $N(0, \alpha_{22})$ и $\alpha_{22} = (2.2)$ момент θ .
- (e) Покажите, что $\tilde{\beta} - \beta \rightarrow_d (\sigma_x^2 + a)^{-1/2}$, где (a, b) имеют совместное нормальное распределение с нулевым средним и ковариационной матрицей S .
- (f) Состоятельна ли $\tilde{\beta}$?

Эмпирические упражнения

Перед выполнением прочтите [Griliches, 1976]. Мы будем оценивать уравнение заработной платы, оцененное Грелихом, используя выборку из NLS-Y [Blau and Newman, 1992]¹. NLS-Y представляет панельные данные (panel data), в которых различные моменты времени были опрошены одни и те же молодые люди (мы не будем, однако, использовать в этом упражнении панельную структуру этого множества данных).

¹См. [Blau and Newman, 1992, Section II] для более детального описания данных.

Выборка содержит информацию об этих индивидах (для момента времени) в самом раннем году с доступной информацией о заработной плате и в 1980 году. В файле GRILIC.ASC представлены данные *RNS*, *RNS80*, *MRT*, *MRT80*, *SMSA*, *SMSA80*, *MED*, *IQ*, *KWW*, *YEAR*, *AGE*, *AGE80*, *S*, *S80*, *EXPR*, *EXPR80*, *TENURE*, *TENURE80*, *LW* и *LW80* (в таком порядке, столбцы представляют переменные). Переменная *YEAR* — год, соответствующий первому моменту времени. Переменные без «80» относятся к первому моменту времени, с «80» — к 1980 году. Определения переменных для второго момента

RNS — данные для приращения в южных штатах;

MRT — данные для семейного статуса (1 для женатых);

SMSA — данные для проживания в крупных городах;

MED — образование матери в годах;

KWW — оценка в тесте *Knowledge of the World of Work*;

IQ — уровень IQ;

AGE — возраст индивида;

S — образование в годах;

EXPR — опыт работы в годах;

TENURE — срок пребывания в должности в годах;

LW — логарифм заработной платы.

Выборка Блэкманн — Нейкиркс содержит 816 наблюдений после исключения тех же самых индивидов. Также удалены случаи с отсутствующей информацией об образовании матери, что сократило выборку до 755 наблюдений.

- (a) Рассчитайте средние и стандартные отклонения всех представленных переменных (в том числе для 1980 года) и подготовьте таблицу подобных табл. 1 Приложения. Также рассчитайте корреляцию между *IQ* и *S* (используйте команду *KSD(CORR)* для TSP или *SDOM (CENT, CORR)* для RATS);

Поскольку год, который был выбран для заработной платы различен для всех индивидов, величина заработной платы будет включать эффект года. Создайте фиктивные переменные для каждого года с 1986-го по 1973-й. (Замечание: в 1972 году наблюдений нет.) Они будут включены в уравнение заработной платы для учета эффекта года.

- (b) Рассмотрим уравнение заработной платы (опуская индекс индивиду-
да):

$$\ln W = \alpha S + \beta IQ + \gamma H + \epsilon,$$

где $\ln W$ — логарифм заработной платы, S — образование в годах, и

$$H \equiv (\text{EXPER, TENURE, RNS, SMSA, указанные данные}).$$

(Включать константу нет необходимости, так как она является для линейной комбинацией экзогенных переменных.) Подготовьте таблицу, по-
добную табл. 3.2. При 2SLS-оценивании уравнения заработной платы
выполните инструментальный регресс с помощью из группы разделенных регрес-
сий (S и H) в исходных данных, предопределенных переменных (MPH ,
 KWW , MRT , AGE). Сопоставьте из отнесенные значения трех
оценок коэффициентов для образования с предсказаниями, основан-
ными на оценках, полученных при инструментарии и ошибками
и регрессии.³

Совет при работе с пакетом TSP: Для выполнения 2SLS используйте команду `INST` или `2SLS`.

Совет при работе с пакетом RATS: Используйте команду `LINREG`
с опцией `INST`. Для определения списка инструментов используйте
команду `INSTRUMENTS`.

- (c) Для второго 2SLS-оценивания уравнения заработной платы из пункта
b) рассчитайте статистику Саргана (она должна быть равна 87,653).
Каким должно быть число степеней свободы? Выпишите результат.

Совет при работе с пакетом TSP: Статистика Саргана может быть
рассчитана как $\text{@rho1}/(\text{@kvar}/\text{@nobs})$, где переменные с «@» явля-
ются результатами выполнения команд TSP. @kvar — число наблюдений,
 $\text{@rho1} = \hat{\epsilon}'X(X'X)^{-1}X'\hat{\epsilon}$, где $\hat{\epsilon}$ — остатки 2SLS-оценивания,
 $\text{@ssr} = \hat{\epsilon}'\hat{\epsilon}$.

Совет при работе с пакетом RATS: Статистика Саргана может
быть рассчитана как $\text{@dvar}/(\text{@kvar}/\text{@nobs})$, где переменные с «@»
являются результатами команды RATS. @nobs — число наблюдений,
 $\text{@dvar} = \hat{\epsilon}'X(X'X)^{-1}X'\hat{\epsilon}$, $\text{@ssr} = \hat{\epsilon}'\hat{\epsilon}$.

- (d) Получите 2SLS-оценку, реально оценивая две регрессии. Проверьте,
что полученные на втором шаге стандартные ошибки отличаются
от полученных в b).
- (e) Грэхэм упоминает, что образование также может быть эндогенной
переменной. Как он аргументирует это? Оцените уравнение зар-
ботной платы методом 2SLS, рассматривая S и IQ как эндогенные
переменные. Что происходит с коэффициентом при образовании? Как

можно объяснить различие оценок коэффициента при образовании здесь и в пункте б)? Вычислите статистику Саргана (она должна быть равна 19.268) и ее значение.

- б) Оцените уравнение заработной платы с помощью GMM, рассматривая образование как предопределенную переменную аналогично пункту б). Меньше ли стандартные ошибки 2SLS, чем стандартные ошибки GMM? Проверьте предопределенность образования с помощью C -статистики. Для оценки C нужно получить две GMM-оценки: с и без образования в качестве инструмента. Используйте одно и то же $\hat{S}$. Для расчета $\hat{S}$ используйте оценки 2SLS из пункта б). C -статистика должна быть равной 58.168

Совет при работе с пакетом TSP: Вычислить $\hat{S}$ довольно затруднительно. Следующий код, хотя и не совсем идеален, выполняет необходимый расчет. Здесь `res0` — остатки 2SLS из пункта б).

```

spr1=1/758;
x0=res0*a;
x1=res0*expr;
x2=res0*tenure;
x3=res0*yns;
x4=res0*smza;
x5=res0*y66;
...
x11=res0*y73;
x12=res0*med;
x13=res0*kww;
x14=res0*mrtr;
x15=res0*age;
msd(print, moment) x0 x15;
```

@mom — результат выполнения команды MSD — и есть $\hat{S}$. Для реализации GMM используйте команду GMM. Опция `COVOC=matrix` намекает эту команду использовать заданную матрицу как $\hat{S}$. J -статистика рассчитывается как @rob*@phi.

Совет при работе с пакетом RATS: LINREG с опциями INST и ROBUSTERRORS реализует GMM для линейных уравнений. Опция `WMATRIX=matrix` намекает использовать $\hat{H}^{-1}$, а не $\hat{S}$. Для вычисления $\hat{S}^{-1}$ сделайте следующее. Здесь `res0` — остатки 2SLS из пункта б).

```

mccov / res0
* s expr tenure gns smza y66 ... y73
```

```
med kww mrt age
compute zinv=inv(1+comet)
zinv является  $\hat{S}^{-1}$ . J-статистика задается как varzmv
```

- (g) Вернитесь к уравнению заработной платы, которое вы оценили в пункте е) методом 2SLS, считая уровень обучения эндогенным. Препятствует беспомощности является большее значение статистики Саджита (и J-статистики из пункта f). Исключая MED и KWW из списка инструментов. Удалены ли теперь другие порядковые условия идентификации? Что происходит с 2SLS-оценками? (Коэффициент при образовании будет равен -0.5293% .) В чем, по-вашему, проблема? **Указание:** Имеется две альтернативные регрессии $\hat{S}$ и IQ MIM и ACF, одностепенные эндогенные предопределенные переменные. Если MIM и ACF не имеют предельно сильной связи с регрессией второго шага $\hat{S}$ и IQ не инструментальны, то тогда полученные значения $\hat{S}$ и IQ будут весьма близки к линейным комбинациям включенных предопределенных переменных. Это приведет к проблеме мультиколлинеарности в регрессии второго шага. Убедитесь в предельно сильной связи MIM и ACF в регрессии первого шага.

Ответы на избранные вопросы

Аналитические упражнения

4. Асимптотическая дисперсия MIM-оценки разна (после подстановки $\alpha = \alpha_0$ в (3.5.1))

$$\Sigma_{xx}^{-1} A \Sigma_{xx}^{-1} + (\Sigma_{xx} A^{-1} \Sigma_{xx})^{-1},$$

где $A \equiv E(z_i^2 z_i z_i')$. Асимптотическая дисперсия двухшагового GMM (согласно (3.6.13)) разна

$$(\Sigma'_{xx} S^{-1} \Sigma_{xx})^{-1},$$

где $S \equiv E(z_i^2 z_i z_i')$. Если определить W , как в указании, то

$$W' S W = W' I - \Sigma'_{xx} W' \Sigma_{xx} = \Sigma_{xx} A \Sigma_{xx}.$$

Таким образом, (3.6.13) сводится к асимптотической дисперсии MIM-оценки. Но (3.5.11) слагаемые $(\Sigma'_{xx} S^{-1} \Sigma_{xx})^{-1}$ — асимптотическая дисперсия двухшаговой GMM-оценки.

Таблица 3.3

Оценки параметров, выборка Близюма — Ньюмарка

Уравнение, зависимая переменная — *LW*, регрессоры: *S*, *IQ*, *EXPR*, *TENURE* и другие контрольные переменные.

№	Метод оценки	Коэффициент при				SER	R ²	Тест на гетеросkedастичность	Эндогенные переменные	Исключенные предопределенные переменные, кроме региональных данных
		<i>S</i>	<i>IQ</i>	<i>EXPR</i>	<i>TENURE</i>					
1	MHK	0,070 (0,0067)	—	0,030 (0,0065)	0,043 (0,0075)	0,328	0,425	—	нет	—
2	MHK	0,062 (0,0073)	0,0027 (0,0010)	0,031 (0,0065)	0,042 (0,0075)	0,326	0,430	—	нет	—
3	2SLS	0,069 (0,013)	0,0002 (0,0039)	0,030 (0,0066)	0,043 (0,0076)	0,328	—	87,6 (<i>p</i> = 0,0000)	<i>IQ</i>	<i>MED</i> , <i>KWW</i> , <i>AGE</i> , <i>MRT</i>
4	2SLS	0,172 (0,021)	-0,008 (0,0047)	0,049 (0,0082)	0,042 (0,0088)	0,380	—	13,3 (<i>p</i> = 0,00131)	<i>S</i> , <i>IQ</i>	<i>MED</i> , <i>KWW</i> , <i>AGE</i> , <i>MRT</i>
5	GMM	0,176 (0,021)	-0,009 (0,0049)	0,050 (0,0080)	0,043 (0,0085)	0,379	—	11,6 (<i>p</i> = 0,00000)	<i>S</i> , <i>IQ</i>	<i>MED</i> , <i>KWW</i> , <i>AGE</i> , <i>MRT</i>
6	2SLS	0,117 (0,027)	0,002 (0,0050)	0,033 (0,0052)	0,051 (0,0029)	0,380	—	14,9 (<i>p</i> = 0,00058)	<i>SIX</i> , <i>IQ</i>	<i>MED</i> , <i>KWW</i> , <i>AGE80</i> , <i>MRT80</i>

Указание: стандартные ошибки в скобках. Строка 6 для 1980 года.

Эмпирические упражнения

- (b) См. табл. 3.8, строка 5.
- (e) Он указывает на проблемы эффективности образования. Во-первых, ошибки измерения η могут быть связаны с успеваемостью обучения. Во-вторых, так как аргументированно ранее, набор S зависит от предбазисовских характеристик. В-третьих, образование может быть измерено с ошибкой. Общая формула для асимптотического смещения 2SLS-оценки такая

$$\text{plim } \hat{\beta}_{2SLS} - \beta = (\Sigma'_{\alpha\alpha} \Sigma_{\alpha\alpha}^{-1} \Sigma_{\alpha\alpha})^{-1} \Sigma'_{\alpha\alpha} \Sigma_{\alpha\alpha}^{-1} E(\varepsilon_1 \cdot \varepsilon_1).$$

Если S экзогенно, то набор инструментов z_i для 2SLS-оценки ошибочно включает в себя не предопределенную переменную. Поэтому один из элементов $E(\varepsilon_1 \cdot z_1)$ не равен нулю. Тогда из формулы следует, что асимптотическое смещение может быть по крайней мере у некоторых коэффициентов.

- (g) Без MED и WVW в качестве инструментов 2SLS-оценка становится неточной. Порядковое условие удовлетворяется как равенство. Вина лежит на низкой предсказательной силе MRT (посмотрите на ее значение в регрессиях первого этапа для S и IQ).

Литература

- Angrist, J., and A. Krueger, 1991, "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 106, 979-1014.
- Beitman, J., Z. Hrubec, P. Taubman, and T. Wales, 1980, *Socioeconomic Success: A Study of the Effects of Genetic Endowments, Family Environment, and Schooling*, Amsterdam: North Holland.
- Blackburn, M., and D. Neumark, 1992, "Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials," *Quarterly Journal of Economics*, 107, 1421-1436.
- Card, D., 1995, "Earnings, Schooling, and Ability Revisited," *Research in Labor Economics*, 14, 23-49.
- Card, D., and A. Krueger, 1996, "Labor Market Effects of School Quality: Theory and Evidence," Working Paper 5440, NBER.
- Fickenshain, M., L. Hansen, and K. Singleton, 1988, "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty," *Quarterly Journal of Economics*, 103, 51-76.
- Friedman, M., 1957 *A Theory of the Consumption Function*, Princeton: Princeton University Press.
- Gallant, R., and D. Jorgenson, 1979, "Statistical Inference for a System of Simultaneous, Non-linear, Implicit Equations in the Context of Instruments Variable Estimation," *Journal of Economics*, 11, 273-302.

- Griliches, Z., 1976, "Wages of Very Young Men," *Journal of Political Economy*, 84, 569-585.
- Lagarias, J., 1943, "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, 11, 1-12.
- Hansen, L., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.
- Hausman, J., and W. Taylor, 1980, "Comparing Specification Tests and Classical Tests," MIT Working Paper No. 266.
- Judge, G., W. Griffiths, R. Hill, H. Lutkepohl, and T. Lee, 1985, *The Theory and Practice of Econometrics* (24 ed.), New York: Wiley.
- Mincer, J., 1958, "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66, 281-302.
- Nelson, C., and R. Startz, 1980, "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967-976.
- Newey, W., 1985, "Generalized Method of Moments Specification Testing," *Journal of Economics*, 29, 229-255.
- Newey, W., and D. McFadden, 1994, "Large Sample Estimation and Hypothesis Testing," Chapte. 56 in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Volume IV, New York: North-Holland.
- Sargan, D., 1958, "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393-415.
- Searle, S., 1982, *Matrix Algebra Useful for Statistics*, New York: Wiley.
- Staiger, D., and J. Stock, 1997, "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557-586.
- Tsai, H., 1955, "Repeated Least-Squares Applied to Complete Equation Systems," M.S., The Hague: Central Planning Bureau.
- White, H., 1982, "Instrumental Variables Regression with Independent Observations," *Econometrica*, 50, 483-489.
- , 1984, *Asymptotic Theory for Econometricians*, Academic.
- Working, L. J., 1927, "What Do 'Statistical Demand' Curves Show?" *Quarterly Journal of Economics*, 41, 219-235.

Глава 4. Обобщенный метод моментов для систем уравнений

Эта глава посвящена оцениванию обобщенным методом моментов нескольких уравнений одновременно. Зная, как оценивать обобщенным методом моментов (GMM) одно уравнение, остается сделать лишь небольшой шаг для перехода к системе уравнений, несмотря на кажущуюся сложность системы обозначений, предназначенной для контроля за каждым уравнением. Это определится тем, что GMM-оценка нескольких уравнений может быть представлена в виде оценки обобщенным методом моментов одного уравнения, если надлежащим образом специфицировать матрицы и векторы, входящие в формулу для GMM-оценки единственного уравнения. В таком случае мы можем разрабатывать асимптотическую теорию оценивания систем уравнений обобщенным методом моментов в почти готовом виде.

Использование GMM-оценки для нескольких уравнений дает значительные выгоды. В случае условной гомосkedастичности она редуцируется к инструментальной оценке с полной информацией (LIVE), которая, в свою очередь, редуцируется к оценке трехшаговым методом наименьших квадратов (3SLS), если набор инструментов является общим для всех уравнений. Если также предположить, что все регрессоры являются предопределенными, то тогда 3SLS редуцируется к оценке кажущихся несвязанными регрессий (SUR), которая, в свою очередь, редуцируется к оценке многомерной регрессии, если все уравнения имеют один и тот же набор регрессоров.

Мы также покажем, что систему нескольких уравнений можно записать в виде системы с перекрестными ограничениями равенства коэффициентов во всех уравнениях. GMM-оценка для такой системы опять является частным случаем обычной GMM-оценки для одного уравнения. GMM-оценка, в случае когда все регрессоры являются предопределенными, а ошибки — условно гомосkedастичными, носит название оценки со случайными эффектами (RE-оценки). Таким образом, SUR и RE являются эквивалентными оценками.

Прикладной раздел данной главы посвящен оцениванию модели спроса на взаимосвязанные факторы посредством SUR/RE. Си-

система функций спроса получается из транслогарифмической функции издержек, которая является обобщением линейной и логарифмической функций издержек, рассмотренной в главе 1.

Здесь мы не рассматриваем оценивание систем уравнений методом максимального правдоподобия (ML). Подробно изложение данного метода представлено в параграфе 5.5.

Если вам неясна значимость обозначений в этой главе, вы можете просто подожить M (число уравнений) равным 2. Это никак не тряснет суть общности любого из результатов данной главы.

4.1. Модель с несколькими уравнениями

Чтобы иметь возможность работать с более чем одним уравнением, нам необходимо усложнить обозначения, введя дополнительный подстрочный индекс, обозначающий конкретное уравнение. Например, i наблюдение зависимой переменной m -го уравнения будет обозначаться y_{im} .

Линейность

Рассмотрим систему из M линейных уравнений, каждое из которых аналогично линейному уравнению в предположении 3.1:

Предположение 4.1 (линейность): *Нмеется M линейных уравнений*

$$y_{im} = \mathbf{z}_{im}' \boldsymbol{\beta}_m + \varepsilon_{im} \quad (m = 1, 2, \dots, M; i = 1, 2, \dots, N), \quad (4.1.1)$$

где N — размер выборки, $\mathbf{z}_{im}$ — k_m -мерный вектор регрессоров, $\boldsymbol{\beta}_m$ — соответствующий вектор коэффициентов, ε_{im} — случайная ошибка в уравнении m .

Модель не включает в себя никаких предположений о корреляциях ошибок между уравнениями (*inter-equation*) или одновременной корреляции (*contemporaneous correlation*). Также не включаются никаких *одноровневых* ограничений на коэффициенты на разных уравнениях. То есть модель не предполагает *перекрестных* ограничений на коэффициенты. Чтобы проиллюстрировать это, рассмотрим

Пример 4.1 (уравнение заработной платы): В примере Гринхуса из главы 3 мы оценивали уравнение заработной платы. В одной из спецификаций Гринхус добавил к ней уравнение для $KTVW$ (результат теста на знание рынка труда):

$$\text{LWPS} = \phi_1 + \beta_1 S_1 + \gamma_1 IQ_1 - \pi \text{EXPR} + \varepsilon_{1i},$$

$$KTVW_i = \phi_2 + \beta_2 S_i + \gamma_2 IQ_i + \varepsilon_{2i}.$$

где (для отслеживания вашей памяти) LW_i — логарифм заработной платы i -го индивида в первый год опроса, S_i — количество лет обучения в школе, $IQ_i \sim IQ$, а $EXPR_i$ — стаж, измеренный в годах. В этом примере $M = 2$, $L_1 = 4$, $L_2 = 3$, $y_{i1} = LW_i$, $y_{i2} = KWW_i$ и

$$z_{i1} = \begin{bmatrix} 1 \\ S_i \\ IQ_i \\ EXPR_i \end{bmatrix}, \quad z_{i2} = \begin{bmatrix} 1 \\ S_i \\ IQ_i \end{bmatrix}, \quad \delta_1 = \begin{bmatrix} \phi_1 \\ \beta_1 \\ \gamma_1 \\ \pi \end{bmatrix}, \quad \delta_2 = \begin{bmatrix} \phi_2 \\ \beta_2 \\ \gamma_2 \end{bmatrix}. \quad (4.1.2)$$

Ошибки ε_{i1} и ε_{i2} могут быть коррелированными. Корреляция может возникнуть, к примеру, если какая-то ненаблюдаемая переменная влияет на заработную плату и на результат теста. Также в модели нет никаких перекрестных ограничений типа $\beta_1 = \beta_2$.

Перекрестные ограничения естественно возникают в случае **панельных данных** (*panel data*), когда одно и то же соотношение может оцениваться для различных моментов времени.

Пример 4.2 (уравнение заработной платы для двух лет): Данные NLS-Y, которые были использованы в примере Грисзеса, в действительности содержат информацию и для 1980 года, так что мы можем оценить два уравнения: одно — для первого момента времени (скажем, для 1969 года) и второе — для 1980 года:

$$LW69_i = \phi_1 + \beta_1 S69_i + \gamma_1 IQ_i + \pi EXPR69_i + \varepsilon_{i1},$$

$$LW80_i = \phi_2 + \beta_2 S80_i + \gamma_2 IQ_i + \pi EXPR80_i + \varepsilon_{i2}.$$

На языке нескольких уравнений $S69_i$ (образование в 1969 году) и $S80_i$ (образование в 1980 году) суть две разные переменные. Было бы естественно допустить случай, когда коэффициенты остаются неизменными во времени: $\phi_1 = \phi_2$, $\beta_1 = \beta_2$, $\gamma_1 = \gamma_2$. Это есть набор линейных перекрестных ограничений.

Как будет показано ниже, в параграфе 4.3, тестирование (линейных и нелинейных) перекрестных ограничений является достаточно несложным делом. Оценивание при наличии ограничений, сводящегося к равенству коэффициентов в разных уравнениях, будет рассмотрено в параграфе 4.6.

Стационарность и эргодичность

Пусть x_{it} есть вектор из K_x инструментов для уравнения it . Хотя это и нечасто встречается на практике, количество инструментов K_x может меняться от уравнения к уравнению.

Пример 4.3 (продолжение примера 4.1): Предположим, что переменная IQ_i является эндогенной в обоих уравнениях, а также что имеется переменная MED (образование матери), являющаяся предопределенной в первом уравнении (так что $E(MED_{t+1}) = 0$). Тогда инструментами для уравнения для LW являются $(L, S, EXPR_i, MED)$. Если эти инструменты ортогональны также и к ε_{2t} , то уравнение для KWW может быть связано с использованным того же набора инструментов, что и уравнение для LW . В данном примере $K_1 = K_2 = 1$ и

$$z_t = x_{2t} \begin{bmatrix} 1 \\ S_t \\ EXPR_t \\ MED_t \end{bmatrix}. \quad (4.13)$$

Модификация предположения 3.2 для системы из нескольких уравнений выглядит следующим образом.

Предположение 4.2 (эргодическая стационарность):

Пусть $\{z_t\}$ образован различными и не связанными константами элементами $\{y_1, \dots, y_m; z_1, \dots, z_m; x_1, \dots, x_m\}$. Тогда $\{z_t\}$ является совместной стационарной и эргодической.

Данное предположение является более сильным, чем предположение о стационарности и эргодичности для каждого уравнения системы в отдельности. Даже если процесс $\{y_m, z_m, x_m\}$ является стационарным и эргодическим для каждого уравнения в отдельности, то, строго говоря, из этого необязательно следует, что более крупный процесс $\{z_t\}$, являющийся объединением отдельных процессов, является (совместно) стационарным и эргодическим (см. пример 3.2 из главы 3, в котором рассмотрен частный случай векторный процесс, компоненты которого являются стационарными и эргодическими процессами). На практике эту задачу можно избежать, так как обычно уравнения состоят из одних и тех же элементов.

Пример 4.4 (продолжение примера 4.1): Поскольку $z_1, z_2, z_3, z_4, z_5, z_6$, например, имеют несколько общих элементов, то z_t будет состоять всего из шести элементов: $(LW_t, KWW_t, S_t, IQ_t, EXPR_t, MED_t)$. Предположение об эргодичности и стационарности для каждого уравнения состоит в том, что $\{LW_t, S_t, IQ_t, EXPR_t, MED_t\}$ является совместной стационарной и эргодической. Предположение 4.2 требует, чтобы совместный процесс включал в себя KWW_t .

Условия ортогональности

Условия ортогональности для системы из M уравнений есть просто совокупность условий ортогональности для отдельных уравнений.

Предложение 4.3 (условия ортогональности): Для каждого уравнения h K_{hm} переменных x_{hm} являются предопределенными. То есть выполнены следующие условия ортогональности.

$$E(x_{hm} \cdot \varepsilon_{hm}) = 0 \quad (h = 1, 2, \dots, M).$$

Таким образом, в итоге мы имеем $\sum_m K_{hm}$ условий ортогональности. Если мы положим

$$g_h = \frac{1}{\sum_{m=1}^M K_{hm} + 1} \begin{bmatrix} \varepsilon_{h1} \cdot \varepsilon_{h1} \\ \vdots \\ \varepsilon_{hM} \cdot \varepsilon_{hm} \end{bmatrix}, \quad (4.1.4)$$

то все эти условия ортогональности можно записать компактно как $E(g_h) = 0$. Заметим, что мы не предполагаем никаких «перекрестных» ортогональностей. Например, ε_{h1} не обязан быть ортогональным ε_{h2} , хотя требуется, чтобы он был ортогональным ε_{h1} . Однако если переменная экзогенна и в x_{h1} и в x_{h2} , то она будет ортогональна и ε_{h1} и ε_{h2} .

Пример 4.5 (продолжение примера 4.1): В данном примере $\sum_m K_{1m} = 5$ ($= 4 + 1$) к g_1 есть

$$g_1 = \begin{bmatrix} \varepsilon_{11} \\ S\varepsilon_{11} \\ EXPR_{1|2|3} \\ MED_{1|2|3} \\ \varepsilon_{12} \\ S\varepsilon_{12} \\ EXPR_{1|2} \\ MED_{1|2} \end{bmatrix}.$$

Поскольку векторы x_{h1} и x_{h2} имеют один и тот же набор инструментов, каждый инструмент ортогонален и ε_{h1} и ε_{h2} .

Идентифицируемость

Установив условия ортогональности для системы из нескольких уравнений, мы можем получить условие идентифицируемости в каждом случае одного уравнения. Модификация равенства (3.3.1) для случая нескольких уравнений выглядит следующим образом:

$$g(\eta; \delta) = \begin{bmatrix} x_{h1} \cdot (\eta_{h1} - z'_{h1}\delta_1) \\ \vdots \\ x_{hM} \cdot (\eta_{hM} - z'_{hM}\delta_M) \end{bmatrix} \quad (4.1.5)$$

где w_1 соответствует предположению 4.2, а вектор δ без индекса есть штабелированный вектор коэффициентов

$$(\sum_{i=1}^M \delta_{i0} > 0) \quad - \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_M \end{bmatrix}. \quad (4.1.6)$$

Таким образом, условия ортогональности могут быть записаны как $E[g(w_1; \delta)] = 0$. Вектор коэффициентов является идентифицируемым, если $\tilde{\delta} = \delta$ есть единственное решение системы уравнений

$$E[g(w_1; \tilde{\delta})] = 0. \quad (4.1.7)$$

Используя определение (4.1.6), мы можем переписать левую часть много уравнения, $E[g(w_1; \tilde{\delta})]$, как

$$\begin{aligned} E[g(w_1; \tilde{\delta})] &= \begin{bmatrix} E[x_{11} \cdot (y_{11} - z'_{11} \tilde{\delta}_1)] \\ \vdots \\ E[x_{1M} \cdot (y_{1M} - z'_{1M} \tilde{\delta}_M)] \end{bmatrix} \\ &= \begin{bmatrix} E(x_{11} \cdot y_{11}) \\ \vdots \\ E(x_{1M} \cdot y_{1M}) \end{bmatrix} - \begin{bmatrix} E(x_{11} z'_{11}) \tilde{\delta}_1 \\ \vdots \\ E(x_{1M} z'_{1M}) \tilde{\delta}_M \end{bmatrix} \\ &= \begin{bmatrix} E(x_{11} \cdot y_{11}) \\ \vdots \\ E(x_{1M} \cdot y_{1M}) \end{bmatrix} - \begin{bmatrix} E(x_{11} z'_{11}) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & E(x_{1M} z'_{1M}) \end{bmatrix} \begin{bmatrix} \tilde{\delta}_1 \\ \vdots \\ \tilde{\delta}_M \end{bmatrix} \equiv \\ &\equiv \sigma_{zy} - \Sigma_{zx} \tilde{\delta}. \end{aligned} \quad (4.1.8)$$

т.е.

$$\sigma_{zy} \equiv \begin{bmatrix} E(x_{11} \cdot y_{11}) \\ \vdots \\ E(x_{1M} \cdot y_{1M}) \end{bmatrix}, \quad \Sigma_{zx} \equiv \begin{bmatrix} E(x_{11} z'_{11}) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & E(x_{1M} z'_{1M}) \end{bmatrix}. \quad (4.1.9)$$

(Третье слагаемое в (4.1.8) получено применением формулы (A.4) из приложения A для умножения блочных матриц.) Следовательно, система уравнений, определяющая $\tilde{\delta}$, может быть записана как

$$\Sigma_{zx} \tilde{\delta} = \sigma_{zy}, \quad (4.1.10)$$

что эквивалентно записанному по форме равенству (3.3.4), определяющему $\tilde{\delta}$ в случае единственного уравнения!

Из обсуждения идентифицируемости в случае одного уравнения следует, что необходимым и достаточным условием является полный ранг матрицы Σ_{zx} . Но так как матрица Σ_{zx} является блочно-диагональной, то это условие эквивалентно следующему

Предположение 4.4 (ранговое условие идентифицируемости): Для каждого $m (= 1, 2, \dots, M)$ матрица $E(\mathbf{a}_{im}\mathbf{a}_{im}') (K_m \times L_m)$ имеет полный столбцовый ранг.

Не сложно понять, почему условие идентифицируемости выглядит именно так. Мы можем единственным образом определить все векторы коэффициентов $(\beta_1, \dots, \beta_M)$ тогда и только тогда, когда вектор β_m определен единственным образом, что может случиться, если и только если предположение 3.4 выполнено для всех уравнений. Ранговое условие в данном случае выглядит просто, так как нет никаких априорных неэкономических ограничений. Идентифицируемость для случая, когда коэффициенты одинаковы во всех уравнениях, будет рассмотрена в параграфе 4.6.

Предположение для асимптотической нормальности

Как и для случая одного уравнения, для обеспечения асимптотической нормальности условия функциональности должны быть усилены.

Предположение 4.5 (g_t является мартингал-разностью с конечными вторыми моментами): $\{g_t\}$ является совместной мартингал-разностью. Матрица $E(g_t g_t')$ невырождена.

Те же самые рассуждения, которые мы делали относительно совместной эрмитической стационарности, применимы и здесь: приведенное предположение является более слабым, чем просто требование выполнения асимптотического предположения для всех уравнений по отдельности. Как и в предыдущих главах, мы используем символ S для $\text{Var}(\bar{y})$, асимптотической дисперсии $\bar{y}$ (то есть дисперсии предельного распределения $\sqrt{n}\bar{y}$). Согласно ЦПТ для стационарных и эрмитических мартингал-разностей, вся дисперсия равна $E(g_t g_t')$. Эта матрица имеет блочную структуру:

$$\begin{aligned} S &= E(g_t g_t') = \\ &= \begin{bmatrix} E(g_{1t} g_{1t}') & \dots & E(g_{1t} g_{Mt}') \\ \vdots & & \vdots \\ E(g_{Mt} g_{1t}') & \dots & E(g_{Mt} g_{Mt}') \end{bmatrix}, \quad (4.1.1) \end{aligned}$$

То есть блок (m, k) матрицы S есть $E(g_{mt} g_{kt}') (m, k = 1, 2, \dots, M)$.

Таким образом, модель из «многоуравнений» представляет собой систему уравнений, для каждого из которых мы делаем те же предположения, которые мы делали для модели единственного уравнения, добавляя в надлежащих случаях требование совместности (такое, как совместная стационарность).

Связь с «полной» системой одновременных уравнений

Модель из нескольких уравнений, рассматриваемая во многих учебниках и называемая «полной» системой одновременных уравнений¹, включает в себя некоторые дополнительные предположения. Эти предположения не требуются для оценки методом GMM, однако если это представляет интерес, удовлетворить его можно в первых двух подпараграфах параграфа 8.5.

4.2. GMM для системы уравнений

Получение выражения для GMM-оценки для нескольких уравнений может быть проведено аналогично случаю одного уравнения. Пусть $\tilde{\delta}$ — гипотетическое значение параметра δ , и определим $g_n(\tilde{\delta})$ посредством (3.4.1). Определение GMM-оценки в данном случае такое же, как и в (3.4.5), однако взвешивающая матрица $\tilde{W}$ имеет теперь размер $\sum_{i=1}^n K_{i1} \times \sum_{i=1}^n K_{i1}$. В предыдущем параграфе мы могли переписать $E[g(w_i; \delta)]$ для случая нескольких уравнений как линейную функцию от $\tilde{\delta}$ (см. (4.1.8)). Мы можем сделать то же самое и сейчас для выборочного аналога $g_n(\tilde{\delta})$.

$$\begin{aligned}
 g_n(\tilde{\delta}) &= \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1} \cdot (y_{i1} - z'_{i1} \tilde{\delta}_1) \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{iM} \cdot (y_{iM} - z'_{iM} \tilde{\delta}_M) \end{bmatrix} = \\
 &= \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1} \cdot y_{i1} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{iM} \cdot y_{iM} \end{bmatrix} - \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1} z'_{i1} \tilde{\delta}_1 \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{iM} z'_{iM} \tilde{\delta}_M \end{bmatrix} = \\
 &= \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1} \cdot y_{i1} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{iM} \cdot y_{iM} \end{bmatrix} - \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1} z'_{i1} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{iM} z'_{iM} \end{bmatrix} \times \\
 &\quad \times \begin{bmatrix} \tilde{\delta}_1 \\ \vdots \\ \tilde{\delta}_M \end{bmatrix} = s_{xy} - S_{xz} \tilde{\delta}. \quad (4.2.1)
 \end{aligned}$$

¹ Полная система одновременных уравнений содержит столько уравнений, сколько имеется эндогенных переменных. — *Прим. член. ред. перевода.*

где

$$s_{xy} \equiv \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1} \cdot y_{i1} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{iM} \cdot y_{iM} \end{bmatrix}, \quad s_{xx} \equiv \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1} x'_{i1} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{iM} x'_{iM} \end{bmatrix}. \quad (4.2.2)$$

Исходя из формы это тоже самое, что и выражение для $g_n(\delta)$ для случая одного уравнения.

Отсюда следует, что результаты параграфа 3.4, касающиеся вывода ГММ-оценки, применимы и в случае нескольких уравнений. В частности, простое перемножение (3.4.5) $\times$ (3.4.11) дает:

ГММ-оценка для системы уравнений:

$$\hat{\delta}(\hat{W}) = (S'_{xx} \hat{W} S_{xx})^{-1} S'_{xx} \hat{W} s_{xy}. \quad (4.2.3)$$

$$\text{ее ошибка оценивания: } \hat{\delta}(\hat{W}) - \delta = (S'_{xx} \hat{W} S_{xx})^{-1} S'_{xx} \hat{W} \bar{g}, \quad (4.2.4)$$

Особенности формулы ГММ-оценки системы уравнений таковы: (i) s_{xy} является штабелированным вектором, как в (4.2.2); (ii) S_{xx} является блочно-диагональной матрицей, как в (4.2.2); (iii) соответственно, взвешивающая матрица $\hat{W}$ имеет размер $\sum_m K_m \times \sum_m K_m$ и (iv) $\bar{g}$ в (4.2.4), выборочное среднее для g_i , является штабелированным вектором:

$$\bar{g} \equiv \frac{1}{n} \sum_{i=1}^n g_i = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n w_{i1} \cdot \varepsilon_{i1} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{iM} \cdot \varepsilon_{iM} \end{bmatrix} = g_n(\delta). \quad (4.2.5)$$

Используя все эти особенности, необходимо будет переписать формулу для ГММ-оценки для системы уравнений (4.2.3). Если $\hat{W}_{mh} (K_m \times K_h)$ есть (m, h) блок матрицы $\hat{W}$ ($m, h = 1, 2, \dots, M$), то тогда (4.2.3) можно записать подробно как

$$\hat{\delta}(\hat{W}) = \begin{bmatrix} \hat{\delta}_1(\hat{W}) \\ \vdots \\ \hat{\delta}_M(\hat{W}) \end{bmatrix} = \begin{bmatrix} \left[\frac{1}{n} \sum_{i=1}^n z_{i1} x'_{i1} \right. & \dots & \left. \frac{1}{n} \sum_{i=1}^n z_{iM} x'_{iM} \right] \\ \vdots & & \vdots \\ \left[\frac{1}{n} \sum_{i=1}^n z_{iM} x'_{i1} \right. & \dots & \left. \frac{1}{n} \sum_{i=1}^n x_{iM} x'_{iM} \right] \end{bmatrix} \times$$

$$\times \begin{bmatrix} \hat{W}_{11} & \dots & \hat{W}_{1M} \\ \vdots & & \vdots \\ \hat{W}_{M1} & \dots & \hat{W}_{MM} \end{bmatrix} \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1} x'_{i1} & \dots & \frac{1}{n} \sum_{i=1}^n x_{i1} x'_{iM} \\ \vdots & & \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{iM} x'_{i1} & \dots & \frac{1}{n} \sum_{i=1}^n x_{iM} x'_{iM} \end{bmatrix}.$$

$$\begin{aligned}
& \times \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1} x'_{i1} & & \\ & \ddots & \\ & & \frac{1}{n} \sum_{i=1}^n x_{iM} x'_{iM} \end{bmatrix} \begin{bmatrix} \tilde{W}_1 & \cdots & \tilde{W}_{1M} \\ & \ddots & \\ \tilde{W}_{M1} & \cdots & \tilde{W}_{MM} \end{bmatrix} \times \\
& \cdot \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1} \cdot y_i \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{iM} \cdot y_i \end{bmatrix} = \begin{bmatrix} A_{11} & \cdots & A_{1M} \\ \vdots & \ddots & \vdots \\ A_{M1} & \cdots & A_{MM} \end{bmatrix}^{-1} \begin{bmatrix} B_1 \\ \vdots \\ B_M \end{bmatrix}, \quad (4.2.6)
\end{aligned}$$

где

$$A_{11} = \left(\frac{1}{n} \sum_{i=1}^n x_{i1} x'_{i1} \right) \tilde{W}_{11} \left(\frac{1}{n} \sum_{i=1}^n x_{i1} x'_{i1} \right),$$

$$A_{1M} = \left(\frac{1}{n} \sum_{i=1}^n x_{i1} x'_{i1} \right) \tilde{W}_{1M} \left(\frac{1}{n} \sum_{i=1}^n x_{iM} x'_{iM} \right),$$

$$A_{M1} = \left(\frac{1}{n} \sum_{i=1}^n x_{iM} x'_{iM} \right) \tilde{W}_{M1} \left(\frac{1}{n} \sum_{i=1}^n x_{i1} x'_{i1} \right),$$

$$A_{MM} = \left(\frac{1}{n} \sum_{i=1}^n x_{iM} x'_{iM} \right) \tilde{W}_{MM} \left(\frac{1}{n} \sum_{i=1}^n x_{iM} x'_{iM} \right),$$

$$\begin{aligned}
B_1 &= \left(\frac{1}{n} \sum_{i=1}^n x_{i1} x'_{i1} \right) \tilde{W}_{11} \left(\frac{1}{n} \sum_{i=1}^n x_{i1} y_i \right) + \\
&+ \cdots + \left(\frac{1}{n} \sum_{i=1}^n x_{i1} x'_{i1} \right) \tilde{W}_{1M} \left(\frac{1}{n} \sum_{i=1}^n x_{iM} y_i \right),
\end{aligned}$$

$$\begin{aligned}
B_M &= \left(\frac{1}{n} \sum_{i=1}^n x_{iM} x'_{iM} \right) \tilde{W}_{M1} \left(\frac{1}{n} \sum_{i=1}^n x_{i1} y_i \right) + \\
&+ \cdots + \left(\frac{1}{n} \sum_{i=1}^n x_{iM} x'_{iM} \right) \tilde{W}_{MM} \left(\frac{1}{n} \sum_{i=1}^n x_{iM} y_i \right).
\end{aligned}$$

Для перехода к подвешенной строке используйте формулы (A.6) и (A.7) для блочных матриц из приложения А. Если вы находите эту работу слишком сложной, просто положите $M = 2$.

4.3. Теория больших выборок

Теория больших выборок для обобщенного метода моментов (ОММ) для одного уравнения была сформулирована в утверждениях 3.1–3.5. Перенос ее на модель с несколькими уравнениями — это просто вопрос механического алгебры. При условии что δ , Σ_{xx} , δ , η , S_{xx} и S_{xy} определены, как в предыдущих параграфах, и что в предположении 3.х

зависимости от предположения 4.4), утверждения 3.1, 3.3 и 3.5–3.8 справедливы и для случая систем нескольких уравнений. Нужно лишь сделать несколько комментариев относительно их интерпретации.

(Тестирование гипотез.) В рассмотренном сейчас случае линейных уравнений δ представляет собой штабелированный вектор, составленный из коэффициентов разных уравнений. Суть утверждений 3.3 и 3.8 о тестировании гипотез состоит в том, что мы можем тестировать перекрестные ограничения.

Пример 4.8 (продолжение примера 4.2): В случае состоящей из двух уравнений системы уравнений заработной платы для двух лет штабелированный вектор коэффициентов δ есть

$$\delta = (\phi_1, \beta_1, \gamma_1, \pi_1, \phi_2, \beta_2, \gamma_2, \pi_2)'$$

Было бы интересно проверить гипотезу $H_0: \beta_1 = \beta_2$ и $\pi_1 = \pi_2$ или, что проще для образования и ставя неэквивалентно во времени. Эту гипотезу можно записать в виде $R\delta = r$, где

$$R = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}, \quad r = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Можно также проверить и нелинейные перекрестные ограничения; для этого следует просто использовать пункт (с) утверждения 3.3 или утверждение 3.8.

(Тест на сверхдоговорные ограничения.) Количество условий ортогональности равно $\sum_m K_m$, а количество коэффициентов равно $\sum_m L_m$. Соответственно, число степеней свободы J -статистики из утверждения 3.3 равно $\sum_m K_m - \sum_m L_m$, а для G -статистики из утверждения 3.7 оно же есть общее количество независимых гомогенных инструментов из разных уравнений.

Утверждение 3.2 не может быть напрямую использовано для случая системы из нескольких уравнений, однако очевидно, как его можно адаптировать.

Утверждение 4.1 (состоятельные оценщики перекрестных моментов одновременных ошибок): Пусть $\tilde{\delta}_{nt}$ — n -вектор вида δ_{nt} и пусть $\tilde{\delta}_{nt} = y_{nt} - x'_{nt}\tilde{\beta}_{nt}$ — порезукаемый остаток для $n = 1, 2, \dots, M$. В предположениях 4.1 и 4.2, а также в предположении, что $E(x_{nt}\delta'_{nt})$ существует и конечно для всех n ($n = 1, 2, \dots, M$),

$$\tilde{\delta}_{ntk} \xrightarrow{p} \sigma_{ntk}, \quad \text{где} \quad (4.3.1)$$

$$\tilde{\sigma}_{ntk} \equiv \frac{1}{n} \sum_{i=1}^n \xi_{it} \tilde{\delta}_{itk} \quad \text{и} \quad \sigma_{ntk} \equiv E(x_{ntk}\delta_{ntk})$$

при условии, что $E(\tilde{\delta}_{ntk}\delta_{ntk})$ конечен и положителен.

Доказательство весьма похоже на доказательство утверждения 3.2 и весьма утомительно, поэтому мы оставляем его в качестве самостоятельного упражнения.

Остается утверждение 4.2, касающееся состоятельности оценивания S . Предположение, соответствующее предположению 3.6 (о конечности четвертых моментов), таково:

Предположение 4.6 (конечные четвертые моменты): $E[(x_{ijk} \cdot z_{ijl})^2]$ существует и конечно для всех $k (= 1, 2, \dots, K_m)$, $j (= 1, 2, \dots, L_k)$, $m, n (= 1, 2, \dots, M)$, где x_{ijk} есть k -й элемент x_{im} и z_{ijl} есть j -й элемент z_{in} .

Формула (3.5.10) для состоятельной оценки матрицы S в случае системы нескольких уравнений имеет вид:

$$\hat{S} = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{i1} \hat{\varepsilon}_{i1} x_{i1} x'_{i1} & \cdots & \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{i1} \hat{\varepsilon}_{in} x_{i1} x'_{in} \\ \vdots & & \vdots \\ \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{in} \hat{\varepsilon}_{i1} x_{in} x'_{i1} & \cdots & \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{in} \hat{\varepsilon}_{in} x_{in} x'_{in} \end{bmatrix} \quad (4.3.2)$$

для некоторой состоятельной оценки $\hat{\varepsilon}_{in}$ ошибки ε_{in} .

Утверждение 4.2 (состоятельная оценка матрицы S , асимптотическая дисперсия вектора $\hat{g}$): Пусть $\hat{\delta}_m$ — состоятельная оценка вектора δ_m и пусть $\hat{\varepsilon}_{in} = y_{in} - x'_{in} \hat{\delta}_m$ представляет собой подразумеваемый остаток для $m = 1, 2, \dots, M$. В предположениях 4.1, 4.2 и 4.6 $\hat{S}$ из (4.3.2) является состоятельной оценкой для матрицы S .

Мы не будем приводить доказательство, поскольку метод доказательства очень схож с использованным при доказательстве утверждения 2.4.

Поскольку $\hat{S}$ — состоятельная оценка S , GMM-оценка с использованием $\hat{S}^{-1}$ в качестве взвешивающей матрицы, $\hat{\delta}(\hat{S}^{-1})$, является эффективной GMM-оценкой с минимальной асимптотической дисперсией. Для удобства здесь воспроизводятся формулы (3.5.13) и (3.5.14) для асимптотической дисперсии и ее оценки:

$$\text{Avar}(\hat{\delta}(\hat{S}^{-1})) = (\Sigma'_{xz} S^{-1} \Sigma_{xz})^{-1}. \quad (4.3.3)$$

$$\widetilde{\text{Avar}}(\hat{\delta}(\hat{S}^{-1})) = (S'_{xz} \hat{S}^{-1} S_{xz})^{-1}. \quad (4.3.4)$$

Резюмируя, GMM-оценка системы уравнений имеет следующие специфические свойства: Σ_{xz} — блочно-диагональная матрица из (4.1.9), S задается уравнением (4.1.11), s_{xz} и S_{xz} берутся из уравнения (4.2.2), а $\hat{S}$ — из уравнения (4.3.2). Для получения начальной оценки $\hat{\delta}_m$, необходимой для вычисления $\hat{\varepsilon}_{in}$ и $\hat{S}$, мы можем использовать FIVE-оценку (будет рассмотрена позже) или эффективную GMM-оценку для одного

уравнения, примененную к каждому уравнению в отдельности. Разумеется, что при условии состоятельности выбор начальной оценки никак не влияет на асимптотическое распределение эффективной GMM-оценки.

В табл. 4.1 сведены все основные результаты данного параграфа.

Таблица 4.1

GMM для системы уравнений в формате случая единственного уравнения

Выборочный аналог:

условий ортогональности: $q_n(\delta) = y_{nx} - S_{xx}\delta = 0$.

GMM-оценка: $\hat{\delta}(\tilde{W}) = (S'_{xx}\tilde{W}S_{xx})^{-1}S'_{xx}\tilde{W}y_{nx}$.

Ее выборочная ошибка: $\tilde{\delta}(\tilde{W}) = (S'_{xx}\tilde{W}S_{xx})^{-1}S'_{xx}\tilde{W}u_{nx}$.

Асимптотическая дисперсия

оценки при GMM-оценке: $\text{Avar}(\hat{\delta}(\tilde{S}^{-1})) = (S'_{xx}S^{-1}\Sigma_{xx})^{-1}$.

Ее оценка: $\text{Avar}(\hat{\delta}(\tilde{S}^{-1})) = (S'_{xx}\tilde{S}^{-1}S_{xx})^{-1}$.

J-статистика: $J(\hat{\delta}(\tilde{S}^{-1}), \tilde{S}^{-1}) = n \cdot q_n(\hat{\delta}(\tilde{S}^{-1}))' \tilde{S}^{-1} q_n(\hat{\delta}(\tilde{S}^{-1}))$.

GMM для одного уравнения, применяемого к интересующему нас уравнению		GMM для системы уравнений
q_i	$x_i - \alpha_i$	(4.1.4)
δ	δ	(4.1.6)
y_{nx}	$\frac{1}{n} \sum_{i=1}^n x_i \cdot y_i$	(4.2.2)
S_{xx}	$\frac{1}{n} \sum_{i=1}^n x_i x_i'$	(4.2.2)
Ранг W	$K \times K$	$\sum_m K_m \times \sum_m K_m$
Σ_{xx}	$E(x_i x_i')$	(4.1.9)
$S (= \text{Avar}(\tilde{\delta}))$	$E(\tilde{z}_i \tilde{z}_i')$	(4.1.11)
$\tilde{S}$	$\frac{1}{n} \sum_{i=1}^n \tilde{z}_i \tilde{z}_i'$	(4.3.2)
При каких условиях оценка состоятельна?	3.1–3.4	4.1–4.4
При каких условиях оценка асимптотически нормальна?	3.1–3.5	4.1–4.5
При каких условиях $\tilde{S} \rightarrow_p S$?	4.1, 3.2–3.6, $E(q_i q_i')$ конечно	4.1–4.2, 4.6, $E(q_i q_i')$ конечно
Степень свободы J	$K - 1$	$\sum_m (K_m - 1)$

4.4. Сравнение оценивания одного уравнения и системы уравнений

Очевидной альтернативой GMM для системы уравнений, оценивающему штабелированный вектор коэффициентов $\delta(\Sigma_m L_m \times 1)$ совместно, является применение GMM для одного уравнения к каждому уравнению в отдельности и последующее штабелирование оценок коэффици-

ентов. Если взвешивающая матрица в случае блочного GMM для m -го уравнения есть $\tilde{W}_{mm} (K_m \times K_m)$, то получившийся в результате оценивания каждого уравнения в отдельности штабелированный вектор коэффициентов δ может быть записан как GMM-оценка для системы уравнений $\delta(\tilde{W})$ из (4.2.3), где взвешивающая матрица $\tilde{W}$ размерности $\sum_m K_m \times \sum_m K_m$ есть блочно-диагональная матрица, m -й блок которой есть $\tilde{W}_{mm}$:

$$\tilde{W} = \begin{bmatrix} \tilde{W}_{11} & & \\ & \ddots & \\ & & \tilde{W}_{MM} \end{bmatrix}. \quad (4.4.1)$$

Это равенство следует из того, что обе матрицы, $S_{xz}(\sum_m K_m \times \sum_m L_m)$ и $\tilde{W}$, являются блочно-диагональными. (Читателю следует подставить выражение (4.2.2) для S'_{xz} и (4.4.1) для $\tilde{W}$ в (4.2.3), чтобы проверить, что $S'_{xz}\tilde{W}S_{xz}$ и, следовательно, $(S'_{xz}\tilde{W}S_{xz})^{-1}S'_{xz}\tilde{W}$ являются блочно-диагональными, используя формулы (A.5) и (A.9) из приложения A. Для простоты можно положить $M = 2$.)

Таким образом, разница между использованием GMM для системы уравнений и GMM для одного уравнения при оценивании штабелированного вектора δ состоит в выборе взвешивающей матрицы $\tilde{W}$ размера $\sum_m K_m \times \sum_m K_m$. Иначе говоря, GMM, применяемый последовательно к каждому уравнению, является частным случаем GMM для системы уравнений.

Когда они эквивалентны?

Как видно из главы 3, если уравнение точно идентифицируемо, то выбор взвешивающей матрицы $\tilde{W}$ не имеет значения, поскольку GMM-оценка, независимо от выбора $\tilde{W}$, численно равна IV-оценке. То же самое верно и для GMM для системы уравнений: если каждое уравнение системы точно идентифицируемо, то есть $L_m = K_m$ для всех m , то тогда S_{xz} из (4.2.3) — квадратная матрица и GMM-оценка при любом выборе $\tilde{W}$ становится IV-оценкой системы уравнений $S_{xz}^{-1}S_{xy}$ (и, так как матрица S_{xz} является блочно-диагональной, полученная оценка представляет собой совокупность IV-оценок для каждого уравнения).

С другой стороны, если хотя бы одно из M уравнений сверхидентифицируемо, выбор матрицы $\tilde{W}(\sum_m K_m \times \sum_m K_m)$ влияет на численное значение GMM-оценки. Рассмотрим эффективную GMM-оценку для $\delta(\sum_m L_m < 1)$, полученную оцениванием каждого уравнения в отдельности, причем каждое уравнение оценено при оптимальном выборе $\tilde{W}_{mm}$ ($m = 1, 2, \dots, M$). Эффективна ли эта оценка так же, как и оценка GMM всей системы уравнений одновременной? Оценку, полученную в результате GMM-оценивания каждого уравнения в отдельности, можно

записать как $\hat{\delta}(\tilde{W})$, где взвешивающая матрица $\tilde{W}$ размера $\sum_m K_m \times \sum_m L_m$ удовлетворяет соотношению:

$$\lim_{n \rightarrow \infty} \tilde{W} = \begin{bmatrix} (E(z_{11}^2 x_1 x_1'))^{-1} & & \\ & \ddots & \\ & & (E(z_{MM}^2 x_M x_M'))^{-1} \end{bmatrix}. \quad (4.4.2)$$

Поскольку этот предел по вероятности не обязательно равен S^{-1} (если только не выполняется условие (4.4.3) ниже), данная оценка, как правило, менее эффективна, чем одновременная GMM-оценка системы уравнений $\hat{\delta}(\hat{S}^{-1})$, где $\hat{S}$ есть состоятельная оценка S . Но если уравнения являются «несвязанными»¹ в том смысле, что

$$E(z_{im} z_{ih} x_{im} x_{ih}') = 0 \quad \text{для всех } m \neq h \quad (= 1, 2, \dots, M), \quad (4.4.3)$$

то S становится блочно-диагональной и для матрицы $\tilde{W}$, определенной выше, $\lim \tilde{W} = S^{-1}$. Так как в этом случае $\tilde{W} - S^{-1} \rightarrow_p 0$, мы имеем (как было доказано в контрольном вопросе 3 из параграфа 3.5)

$$\sqrt{n} \hat{\delta}(\tilde{W}) - \sqrt{n} \hat{\delta}(\hat{S}^{-1}) \rightarrow 0.$$

Из леммы 2.4(a) следует, что асимптотическое распределение $\hat{\delta}(\tilde{W})$ такое же, как и у $\hat{\delta}(\hat{S}^{-1})$, поэтому эффективная GMM-оценка каждого уравнения поочередно есть не что иное, как эффективная GMM-оценка всей системы одновременно.

Подытожим сказанное выше:

Утверждение 4.3 (эквивалентность GMM-оценивания поочередно каждого уравнения и всей системы одновременно):

- (a) Если все уравнения точно идентифицируемы, то GMM-оценивание поочередно каждого уравнения и GMM-оценивание всей системы одновременно дают численно одинаковые оценки, эквивалентные IV-оценке.
- (b) Если хотя бы одно из уравнений сверхидентифицируемо, но уравнения «не связаны» в смысле (4.4.3), то GMM-оценивание поочередно каждого уравнения и GMM-оценивание всей системы одновременно асимптотически эквивалентны в том смысле, что умноженная на $\sqrt{n}$ разность этих оценок сходится к нулю по вероятности.

¹ При условной гомоскедастичности это условие превращается в более знакомое условие $E(z_{im} z_{ih}) = 0$. См. следующий параграф.

В двух описанных случаях совместное оценивание не даст никакого выигрыша в эффективности. Более того, если *априорно* известно, что уравнения «не связаны», то тогда предпочтительнее оценивать каждое уравнение в отдельности, так как использование априорного знания, выражающегося в занулении всех элементов матрицы $\tilde{W}$ вне диагональных блоков, весьма вероятно, приведет к улучшению свойств оценки на конечных выборках.

Совместное оценивание может быть рискованным

За исключением случаев (а) и (б), совместное оценивание асимптотически было эффективно. Даже если интерес представляет только оценивание одного конкретного уравнения (как, например, уравнения для LM на примере 4.1), вы можете обычно выиграть в асимптотической эффективности, комбинируя его с другими уравнениями (см., однако, аналитическое уравнение 10 в качестве примера, когда совместное оценивание не влечет за собой увеличение асимптотической эффективности, даже если добавляемые уравнения не являются «несвязанными»).

Следует, однако, сделать некоторые оговорки. Во-первых, оценки коэффициентов интересующего нас уравнения могут иметь лучшие свойства на конечных выборках без совместного оценивания с другими уравнениями. Во-вторых, асимптотические результаты предполагают, что модель **правильно специфицирована** (*correctly specified*), то есть все предположения модели выполнены. Если модель специфицирована неправильно, то ни GMM-оценка каждого уравнения поочередно, ни GMM-оценка всей системы одновременно не гарантируют даже состоятельности. А шансы неправильности спецификации возрастают при добавлении уравнений к системе.

Чтобы проиллюстрировать второе замечание, допустим, что предположение 4.3 не выполняется из-за невыполнения условий ортогональности для M -го уравнения и только для него: $E(x_{LM} \varepsilon_M) \neq 0$. Это может произойти, например, в случае пропуска переменной, которая должна быть включена в качестве регрессора. Чтобы увидеть, что это ведет к возможной потере состоятельности во всех уравнениях, посмотрим на формулу (4.2.4) для ошибки оценивания, где $\bar{g}$ задается выражением (4.2.5). M -й блок матрицы $\text{plim}_{n \rightarrow \infty} \bar{g}$ не равен нулю. Поскольку $\text{plim}_{n \rightarrow \infty} S_{xx} = \Sigma_{xx}$ при условии эргодической стационарности и $\tilde{W} = W$, по предположению, асимптотическое смещение равно:

$$\text{plim}_{n \rightarrow \infty} \tilde{\delta}(\tilde{W}) - \delta = (\Sigma'_{xx} W \Sigma_{xx})^{-1} \Sigma'_{xx} W \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} E(x_{LM} \varepsilon_M). \quad (4.4.4)$$

У эффективной GMM-оценки системы уравнений $\mathbf{W}$, и, следовательно, $(\Sigma'_{xz} \mathbf{W} \Sigma_{xz})^{-1} \Sigma'_{xz} \mathbf{W}$, в общем случае не являются блочно-диагональными, так что любой элемент из $\lim_{n \rightarrow \infty} \hat{\delta}(\hat{\mathbf{W}}) = \delta$ может быть отличным от нуля. Даже для коэффициентов интересующих нас правильно специфицированных уравнений асимптотическое смещение может быть ненулевым. То есть в случае совместного оценивания смещение, вызванное неправильной спецификацией одного уравнения, распространяется на всю систему. Эта проблема не возникает в случае GMM-оценивания каждого уравнения поочередно, при котором матрица $\hat{\mathbf{W}}$ имеет блочно-диагональную структуру.

Контрольные вопросы

1. Выразите 2SLS-оценку каждого уравнения поочередно в форме (4.2.3), специфицируя подходящим образом $\hat{\mathbf{W}}_{\text{class}}$ в (4.4.1).

[Ответ: $\hat{\mathbf{W}}_{\text{class}} = (\frac{1}{n} \sum_{i=1}^n \mathbf{x}_{mi} \mathbf{x}'_{mi})^{-1}$.]

4.5. Частные случаи GMM-оценивания системы уравнений: FIVE, 3SLS и SUR

Условная гомоскедастичность

В случае условной гомоскедастичности можно легко получить несколько важных оценок в качестве частных случаев GMM-оценки системы уравнений. Модификация условий гомоскедастичности для систем уравнений такова:

Предположение 4.7 (условная гомоскедастичность):

$$E(\varepsilon_{mh} \varepsilon_{bh} | \mathbf{x}_{mh}, \mathbf{x}_{bh}) = \sigma_{mh}$$

для всех $m, h = 1, 2, \dots, M$.

В этом случае безусловный перекрестный момент $E(\varepsilon_{mh} \varepsilon_{bh})$ равен σ_{mh} в соответствии с формулой полного математического ожидания. Вам должно быть несложно показать, что

блок (m, h) из $\Gamma(g, g')$, представленный в (4.1.1), =

$$= E(\varepsilon_{mh} \varepsilon'_{bh} \mathbf{x}_{mh} \mathbf{x}'_{bh}) =$$

$$= E[E(\varepsilon_{mh} \varepsilon'_{bh} \mathbf{x}_{mh} \mathbf{x}'_{bh} | \mathbf{x}_{mh}, \mathbf{x}_{bh})]$$

(по закону полного математического ожидания) =

$$\begin{aligned}
&= E[E(\varepsilon_{imh} | x_{i1}, x_{i2}) x_{im} x'_{ih}] \\
&\text{(ввиду линейности условного математического ожидания)} - \\
&= E[\sigma_{imh} x_{im} x'_{ih}] \\
&\text{(ввиду условной гомоскедастичности)} = \\
&= \sigma_{imh} E(x_{im} x'_{ih}) \quad \text{(ввиду линейности математического ожидания)}.
\end{aligned} \tag{4.5.1}$$

Тогда S из (4.1.1) можно записать как

$$S = \begin{bmatrix} \sigma_{11} E(x_{11} x'_{11}) & \cdots & \sigma_{1M} E(x_{11} x'_{1M}) \\ \vdots & & \vdots \\ \sigma_{M1} E(x_{M1} x'_{11}) & \cdots & \sigma_{MM} E(x_{M1} x'_{1M}) \end{bmatrix}. \tag{4.5.2}$$

Поскольку, по предположению 4.3, S конечна, из данной декомпозиции вытекает, что $E(x_{im} x'_{ih})$ существует и конечно для всех m, h ($m, h = 1, 2, \dots, M$).

Инструментальная оценка с полной информацией (FIVE)

Оценка для S , использующая структуру четвертых моментов, показанную в (4.5.2), такова:

$$\hat{S} = \begin{bmatrix} \hat{\sigma}_{11} \cdot \left(\frac{1}{n} \sum_{i=1}^n x_{i1} x'_{i1} \right) & \cdots & \hat{\sigma}_{1M} \cdot \left(\frac{1}{n} \sum_{i=1}^n x_{i1} x'_{iM} \right) \\ \vdots & & \vdots \\ \hat{\sigma}_{M1} \cdot \left(\frac{1}{n} \sum_{i=1}^n x_{iM} x'_{i1} \right) & \cdots & \hat{\sigma}_{MM} \cdot \left(\frac{1}{n} \sum_{i=1}^n x_{iM} x'_{iM} \right) \end{bmatrix}, \tag{4.5.3}$$

где для некоторой состоятельной оценки $\hat{\delta}_{mh}$ вектора δ_{mh}

$$\hat{\sigma}_{mh} \equiv \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{im} \hat{\varepsilon}_{ih}, \quad \hat{\varepsilon}_{im} \equiv y_{im} - x'_{im} \hat{\delta}_{im} \quad (m, h = 1, 2, \dots, M). \tag{4.5.4}$$

Согласно утверждению 4.1, $\hat{\sigma}_{mh} \rightarrow_p \sigma_{mh}$ при условии (в дополнение к предположениям 4.1 и 4.2), что $E(x_{im} x'_{ih})$ конечно. Из эргодической стационарности следует, что $\frac{1}{n} \sum_{i=1}^n x_{im} x'_{ih}$ сходится по вероятности к $E(x_{im} x'_{ih})$, которое, как было только что отмечено, существует и конечно. Следовательно, (4.5.3) состоятельна для S без условия на четвертые моменты (предположение 4.6).

FIVE-оценка (FIVE estimator) вектора δ , обозначаемая $\hat{\delta}_{FIVE}$, есть

$$\hat{\delta}_{FIVE} \equiv \hat{\delta}(\hat{S}^{-1}).$$

где $\hat{S}$ задается посредством (4.5.3), чтобы использовать условную гомоскедастичность¹. Будучи GMM-оценкой для системы уравнений при некоторой выбранной $\hat{W}$, она является состоятельной и асимптотически нормальной согласно утверждению 3.1, адаптированному к случаю системы уравнений. Поскольку $\hat{S}$ — состоятельная оценка для S , она является эффективной согласно утверждению 3.3. Таким образом, мы доказали

Утверждение 4.4 (свойства FIVE-оценок на больших выборках): Предположим, что выполняются предположения 4.1–4.5 и 4.7. Кроме того, предположим, что $E(z_m z_m')$ существует и конечна для всех m, n ($= 1, 2, \dots, M$).² Пусть S и $\hat{S}$ соответствуют (4.5.2) и (4.5.3). Тогда

- (a) $\hat{S} \rightarrow_p S$.
- (b) $\hat{\delta}_{FIVE}$, определяемая как $\hat{\delta}(\hat{S}^{-1})$, является состоятельной, асимптотически нормальной и эффективной с $\text{Avar}(\hat{\delta}_{FIVE})$ из (4.3.3).
- (c) Оценка асимптотической дисперсии, приведенная в (4.3.4), является состоятельной для $\text{Avar}(\hat{\delta}_{FIVE})$.
- (d) (Статистика Саргана.)

$$J(\hat{\delta}_{FIVE}, \hat{S}^{-1}) \equiv n \cdot g_n(\hat{\delta}_{FIVE})' \hat{S}^{-1} g_n(\hat{\delta}_{FIVE}) + \gamma^2 \left(\sum_m (K_m - I_m) \right),$$

где $g_n(\cdot)$ задается в (4.2.1).

Обычно начальная оценка $\hat{\delta}_0$, необходимая для получения $\hat{S}$, является 2SLS-оценкой.

Трехшаговый метод наименьших квадратов (3SLS)

Когда набор инструментов один и тот же для всех уравнений, формула для FIVE-оценки может быть упрощена. Упрощенная формула определяет 3SLS-оценку (*3SLS-estimator*), обозначаемую как $\hat{\delta}_{3SLS}$ ³. Для этого положим

$$\epsilon_{(N)} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_M \end{bmatrix}. \quad (4.5.b)$$

¹Это оценка из [Bellamy and Jorgenson, 1971].

²Это предположение не необходимо для состоятельности $\hat{\delta}_{FIVE}$. См. утверждение 4.1.

³Это оценка из [Zellner and Theil, 1962].

Тогда матрица перекрестных моментов для ε_i размера $M \times M$, обозначаемая как Σ , может быть записана как

$$\Sigma_{(M \times M)} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1M} \\ \vdots & & \vdots \\ \sigma_{M1} & \cdots & \sigma_{MM} \end{bmatrix} = E(\varepsilon_i \varepsilon_i'). \quad (4.5.6)$$

Чтобы состоятельно оценить Σ , нам требуется начальная состоятельная оценка $\hat{\delta}_{in}$ для получения остатка $\hat{\varepsilon}_{in}$. Термин 3SLS связан с тем, что в качестве начальной оценки для δ_{in} используется 2SLS-оценка. При наличии остатков, полученных таким путем, естественной оценкой для Σ является

$$\hat{\Sigma} = \begin{bmatrix} \hat{\sigma}_{11} & \cdots & \hat{\sigma}_{1M} \\ \vdots & & \vdots \\ \hat{\sigma}_{M1} & \cdots & \hat{\sigma}_{MM} \end{bmatrix} = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i \hat{\varepsilon}_i'. \quad (4.5.7)$$

Эта матрица, в которой элемент (m, h) является оцененным перекрестным моментом, заданным в (4.5.4).

Если $w_i = (w_{i1} \ w_{i2} \ \cdots \ w_{iK})'$ — общий набор инструментов размерности K , то g_i из (4.1.6), S из (4.3.2) и $\tilde{S}$ из (4.3.3) могут быть компактно записаны с использованием произведения Кронекера¹ как

$$g_i = E_i \otimes x_i, \quad (4.5.8)$$

$(M \times K \times 1)$

$$S_{(M \times K \times M \times K)} = \Sigma_{(M \times M)} \otimes E(x_i x_i'), \quad (4.5.9)$$

$(K \times K)$

так что $S^{-1} = \Sigma^{-1} \otimes [E(x_i x_i')]^{-1}$,

$$\tilde{S} = \hat{\Sigma} \otimes \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right), \quad (4.5.10)$$

так что $\tilde{S}^{-1} = \hat{\Sigma}^{-1} \otimes \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1}$. Из представления в виде произведения Кронекера (4.5.9) для S и невырожденности S (согласно предположению 4.5) вытекает, что обе матрицы Σ и $E(x_i x_i')$ невырождены.

Для дальнейшего преобразования формулы для 3SLS-оценки мы должны вернуться к соотношению (4.2.6), которое, при $\tilde{W} = \tilde{S}^{-1}$, описывает эффективную GMM-оценку для системы уравнений. Формула (4.5.10) предполагает, что $\tilde{W}$ в (4.2.6) таково, что

$$\tilde{W}_{mh} \equiv (m, h) \text{ блок } \tilde{W} = \sigma^{mh} \cdot \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1}. \quad (4.5.11)$$

¹ Если вы не знакомы с данным произведением Кронекера, посмотрите формулы (A.1.1) и (A.15) приложения А.

где $\hat{\sigma}^{ab}$ есть элемент (a, b) матрицы $\hat{\Sigma}^{-1}$. Подставляя это в (4.2.6), получаем:

$$\hat{\delta}_{\text{LS}} = \begin{bmatrix} \hat{\sigma}^{11} \hat{A}_{11} & \dots & \hat{\sigma}^{1M} \hat{A}_{1M} \\ \vdots & & \vdots \\ \hat{\sigma}^{M1} \hat{A}_{11} & \dots & \hat{\sigma}^{MM} \hat{A}_{MM} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\sigma}^{11} \hat{z}_{11} + \dots + \hat{\sigma}^{1M} \hat{z}_{1M} \\ \vdots \\ \hat{\sigma}^{M1} \hat{z}_{M1} + \dots + \hat{\sigma}^{MM} \hat{z}_{MM} \end{bmatrix}, \quad (4.5.12)$$

где

$$\hat{A}_{ab} = \left(\frac{1}{n} \sum_{i=1}^n z_{ai} x_i' \right) \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left(\frac{1}{p} \sum_{j=1}^p x_j z_{bj}' \right), \quad (4.5.13)$$

$$\hat{z}_{ab} = \left(\frac{1}{n} \sum_{i=1}^n z_{ai} x_i' \right) \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n e_i \cdot y_{ib} \right). \quad (4.5.14)$$

Аналогично, подставляя в (4.3.3) выражение для Σ_{xx} из (4.1.9) и выражение для S из (4.5.9) и используя формулу (А.6) из приложения А, получим выражение для $\text{Avar}(\hat{\delta}_{\text{LS}})$:

$$\text{Avar}(\hat{\delta}_{\text{LS}}) = \begin{bmatrix} \sigma^{11} A_{11} & \dots & \sigma^{1M} A_{1M} \\ \vdots & & \vdots \\ \sigma^{M1} A_{11} & \dots & \sigma^{MM} A_{MM} \end{bmatrix}^{-1}, \quad (4.5.15)$$

где

$$A_{ab} \equiv E(z_{ai} x_i' (E(x_i x_i')^{-1} E(x_i x_{ib}'))$$

и σ^{ab} есть элемент (a, b) матрицы Σ^{-1} . Эта асимптотическая дисперсия действительно оценивается по следующему

$$\widehat{\text{Avar}}(\hat{\delta}_{\text{LS}}) = \begin{bmatrix} \hat{\sigma}^{11} \hat{A}_{11} & \dots & \hat{\sigma}^{1M} \hat{A}_{1M} \\ \vdots & & \vdots \\ \hat{\sigma}^{M1} \hat{A}_{11} & \dots & \hat{\sigma}^{MM} \hat{A}_{MM} \end{bmatrix}^{-1}. \quad (4.5.17)$$

(Этот выборочный вывод можно также получить из формулы поворота (4.2.2) и (4.5.10) в (4.3.4).)

В большинстве учебников приводятся даже еще более привлекательные выражения с использованием матриц данных (таких как матрица X размера $n \times K$). Ввиду таких выражений мы отпущим в качестве аналитического упрощения в этой главе. Подставим полученные результаты:

Утверждение 4.5 (Свойства 3SLS-оценки на большой выборке):

Допустим, что предположения 4.1–4.5 и 4.7 выполняются, и предположим, что $x_{in} \rightarrow x$ (общий набор инструментов). Предположим, далее,

что $E(z_{1m}z'_{1h})$ существует и конечно для всех m, h ($= 1, 2, \dots, M$). Пусть $\tilde{S}$ есть матрица размера $M \times M$ оцененных перекрестных моментов ошибок, полученная в (4.5.7) с использованием 2SLS-остатков. Тогда

- (a) $\hat{\delta}_{2SLS}$ из (4.5.12) является состоятельной, асимптотически нормальной и эффективной оценкой с асимптотической дисперсией $\text{Avar}(\hat{\delta}_{2SLS})$, заданной в (4.5.15).
- (b) Оценка асимптотической дисперсии (4.5.17) является состоятельной оценкой для $\text{Avar}(\hat{\delta}_{2SLS})$.
- (c) (Статистика Саргана.)

$$J(\hat{\delta}_{2SLS}, \hat{S}^{-1}) \equiv n \cdot g_n(\hat{\delta}_{2SLS})' \hat{S}^{-1} g_n(\hat{\delta}_{2SLS}) \rightarrow \chi^2(MK - \sum_{m=1}^M L_m),$$

где $\hat{S} = \tilde{S} \otimes \left(\frac{1}{n} \sum_i x_i x_i' \right)$, K — число общих инструментов и $g_n(\cdot)$ задается в (4.2.1).

Каждые не связанными регрессии (SUR)

2SLS-формула может быть упрощена, если

$$x_i = \text{объединение } (z_{i1}, \dots, z_{iM}), \quad (4.5.18)$$

Это равносильно условию:

$$E(z_{im} \cdot \varepsilon_{ih}) = 0 \quad (m, h = 1, 2, \dots, M) \quad (4.5.18')$$

То есть предопределенные регрессоры должны удовлетворять условию перекрестной ортогональности: они должны быть предопределенными не только в своем уравнении ($E(z_{im} \cdot \varepsilon_{im}) = 0$), но также и для других уравнений ($E(z_{im} \cdot \varepsilon_{ih}) = 0$ для $m \neq h$). Эта упрощенная формула называется **SUR-оценкой** (*SUR estimator*) и обозначается как $\hat{\delta}_{SUR}^1$.

Пример 4.7 (пример 4.1 с измененной спецификацией инструментов): Если (4.5.18) выполняется для системы из двух уравнений из примера 4.1, x_i есть объединение z_{i1} и z_{i2} , то тогда

$$x_i = \begin{bmatrix} 1 \\ S_i \\ IQ_i \\ EXPR_i \end{bmatrix}. \quad (4.5.19)$$

¹Эта оценка была утверждена в [Zellner, 1962].

Условия ортогональности $E(x_1 \cdot \varepsilon_{12}) = 0$ и $E(x_1 \cdot \varepsilon_{22}) = 0$ превращаются в

$$E \begin{bmatrix} \varepsilon_{11} \\ S_1 \varepsilon_{11} \\ IQ_1 \varepsilon_{11} \\ EXPR_1 \varepsilon_{11} \end{bmatrix} = 0 \quad \text{и} \quad E \begin{bmatrix} \varepsilon_{22} \\ S_1 \varepsilon_{22} \\ IQ_1 \varepsilon_{22} \\ EXPR_1 \varepsilon_{22} \end{bmatrix} = 0, \quad (4.5.20)$$

что отличается от декларирования того, что регрессоры являются предопределенными в обоих уравнениях:

$$E \begin{bmatrix} \varepsilon_{11} \\ S_1 \varepsilon_{11} \\ IQ_1 \varepsilon_{11} \\ EXPR_1 \varepsilon_{11} \end{bmatrix} = 0 \quad \text{и} \quad E \begin{bmatrix} \varepsilon_{12} \\ S_1 \varepsilon_{12} \\ IQ_1 \varepsilon_{12} \end{bmatrix} = 0. \quad (4.5.21)$$

Разница между (4.5.20) и (4.5.21) состоит в том, что уравнение KIVW сверифицифицируется с $EXPR$ как дополнительный инструмент.

Так как SUR есть частный случай 3SLS, формулы (4.5.12), (4.5.15) и (4.5.17) для $\hat{\beta}_{3SLS}$ применимы к $\hat{\beta}_{SUR}$. Из (4.5.18) следует, что 3 полученных ранее выражений для $\tilde{A}_{mm}$, $\tilde{C}_{mh}$ и A_{mh} не исчезают:

$$\tilde{A}_{mm} = \frac{1}{n} \sum_{i=1}^n x_{mi} x'_{mi}, \quad (4.5.13')$$

$$\tilde{C}_{mh} = \frac{1}{n} \sum_{i=1}^n x_{mi} \cdot y_{hi}, \quad (4.5.14')$$

$$A_{mh} = E(x_{mi} x'_{hi}). \quad (4.5.16')$$

Докажем (4.5.16'). Без ограничения общности предположим, что x_{mi} есть L_{mi} -элементов из K -векторов вектора x_i . Пусть $D(K \times \times L_{mi})$ — первые L_{mi} столбцов I_K . Тогда

$$x_{mi} = D' x_i, \quad E(x_{mi} x'_{mi}) = D' E(x_i x'_i), \quad D' E(x_i x'_i) = E(x_{mi} x'_{mi}).$$

Подставляя это в (4.5.16), получим (4.5.16'). Доказательство равенств (4.5.13') и (4.5.14'), касающихся выборочных моментов, проводится аналогично случаю с $x_{mi} = D' x_i$. (В одной из задач предлагается привести альтернативное доказательство с использованием оператора проектирования.)

В методе 3SLS в качестве начальной состоятельной оценки $\tilde{\Sigma}$ использовалась 2SLS-оценка. Но 2SLS в случае вложенности регрессоров во множество инструментов превращается в OLS (как показано в контрольном вопросе 7 параграфа 3.8). Таким образом, для SUR начальной оценкой служит OLS-оценка. Из вышесказанного следует

Утверждение 4.6 (свойства SUR на больших выборках): Пусть предположения 4.1–4.5 и 4.7 выполняются при $\mathbf{x}_i = \text{объединение } (z_{i1}, \dots, z_{iM})$. Пусть $\hat{\Sigma}$ — матрица оцененных кросс-моментов ошибок размерности $M \times M$, полученная в (4.5.7) с использованием OLS-остатков. Тогда

(a) $\hat{\delta}_{\text{SUR}}$ из (4.5.12) вместе с $\hat{A}_{mh}$ и $\hat{c}_{mh}$ из (4.5.13) и (4.5.14) для $m, h = 1, \dots, M$ есть состоятельная, асимптотически нормальная и эффективная оценка с $\text{Avar}(\hat{\delta}_{\text{SUR}})$ из (4.5.15), где A_{mh} задается в (4.5.16).

(b) Оценка асимптотической дисперсии (4.5.17), где $\hat{A}_{mh}$ задается в (4.5.13), есть состоятельная оценка для $\text{Avar}(\hat{\delta}_{\text{SUR}})$.

(c) (Статистика Саргана.)

$$J(\hat{\delta}_{\text{SUR}}, \hat{S}^{-1}) = n \cdot g_n(\hat{\delta}_{\text{SUR}})' \hat{S}^{-1} g_n(\hat{\delta}_{\text{SUR}}) \xrightarrow{d} \chi^2 \left(MK - \sum_m L_m \right),$$

где $\hat{S} = \hat{\Sigma} \otimes (\frac{1}{n} \sum_i \pi_i \mathbf{x}_i' \mathbf{x}_i)$, K — количество общих инструментов и $g_n(\cdot)$ задается в (4.2.1).

(В противоположность утверждению 4.5 здесь, благодаря (4.5.18), не требуется условия конечности $E(\mathbf{x}_{ih} \mathbf{x}_{ih}')$, поскольку это следует из условия невырожденности $E(\mathbf{x}_i \mathbf{x}_i')$, что, в свою очередь, следует из невырожденности S [из предположения 4.5] и того факта, что S может быть записана как $\Sigma \otimes (\frac{1}{n} \sum_i \pi_i \mathbf{x}_i' \mathbf{x}_i)$.) Статистика Саргана для SUR редко рассматривается на практике.

SUR и OLS

Поскольку регрессоры predetermined, систему можно оценить с помощью OLS для каждого уравнения в отдельности. Тогда почему SUR лучше OLS? Как мы только что видели, при условной гомоскедастичности эффективная GMM оценка системы уравнений есть FIVE-оценка, которая численно равна SUR-оценке при условии (4.5.18). Как показано в главе 3, при условной гомоскедастичности эффективная GMM-оценка одного уравнения есть 2SLS-оценка, которая, в свою очередь, численно эквивалентна OLS-оценке, так как регрессоры являются predetermined-ными при (4.5.18). Это соотношение показано на рис. 4.1. Следовательно, соотношение между SUR и OLS для каждого уравнения в отдельности точно такое же, как и соотношение между GMM для системы уравнений и GMM для каждого уравнения в отдельности. Как обсуждалось в параграфе 4.4, следует рассмотреть два случая.

(a) Каждое уравнение идентифицируемо точно. Поскольку общий набор инструментов является объединением всех регрессоров, то это может

быть возможным, только если во всех уравнениях од-и / те же регрессоры, то есть $z_{itn} = x_i$ для всех n . SUR в таком случае называется **многомерной регрессией** (*multivariate regressions*). Мы показали в параграфе 4.4, что оба эффективных GMM для системы уравнений и для каждого уравнения в отдельности, т.е. эквивалентны оценке методом IV для точно идентифицируемой системы. Так как регрессоры предопределены, мы можем заключить, что GMM-оценка модели многомерной регрессии есть OLS-оценка каждого уравнения в отдельности¹. Это можно проверить, подставив (4.5.13') и (4.5.14') (с $z_{itn} = x_i$) в выражение для точечной оценки (4.5.12). А подставив (4.5.16') и (4.5.13') (опять же с $z_{itn} = x_i$) в (4.5.15) и (4.5.17), мы получим, что для многомерной регрессии выполнено:

$$\text{Avar}(\hat{\delta}) = \Sigma \times [E(x_i x_i')]^{-1}. \quad (4.5.22)$$

$$\widehat{\text{Avar}}(\hat{\delta}) = \hat{\Sigma} \times \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1}. \quad (4.5.23)$$

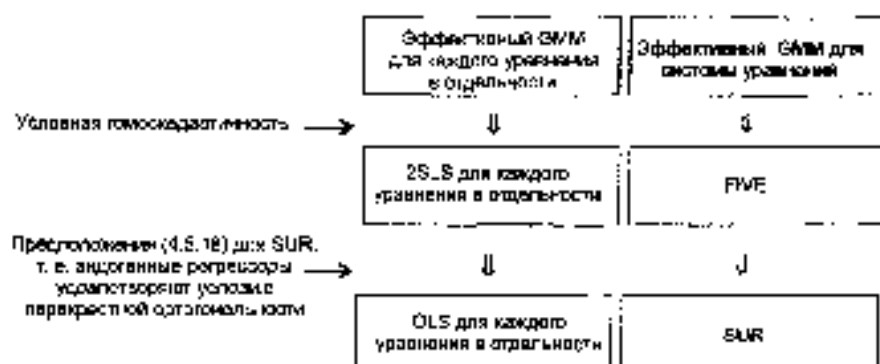


Рис. 4.1. OLS и GMM

- (b) По крайней мере одно уравнение сверхидентифицируемо. Тогда SUR-оценивание более эффективно, чем оценивание методом OLS каждого уравнения в отдельности, если только уравнения не являются «не связанными» друг с другом в смысле (4.4.3). В случае условной гомоскедастичности и общего набора инструментов (4.4.3) преобразуется в

$$\sigma_{mh} E(x_i x_i') = 0 \text{ для всех } m \neq h.$$

Поскольку $E(x_i x_i')$ не может быть нулевой матрицей (она предполагается невырожденной), уравнения являются «не связанными» друг

¹ В главе 5 будет показано, что эта оценка является также простой максимального правдоподобия.

с другом, только если $\sigma_{mh} = 0$. Следовательно, SUR более эффективна, чем OLS, если $\sigma_{mh} \neq 0$ для некоторой пары (m, h) . Если $\sigma_{mh} = 0$ для всех (m, h) , то эти две оценки асимптотически эквивалентны (разность оценок, умноженная на $\sqrt{n}$, стремится к нулю).

Эффективность SUR можно также увидеть, если посмотреть на эту модель как на многомерную регрессию с априорными исключяющими ограничениями. В качестве примера расширим систему из двух уравнений из примера 4.3 следующим образом:

$$\begin{aligned}LW_i &= \phi_1 + \beta_1 S_i + \gamma_1 IQ_i + \pi EXPR_i + \varepsilon_{1i}, \\KW_i &= \phi_2 + \beta_2 S_i + \gamma_2 IQ_i + \pi EXPR_i + \varepsilon_{2i}.\end{aligned}$$

Здесь общий набор инструментов $(1, S_i, IQ_i, EXPR_i)$. Поэтому эта система из двух уравнений есть многомерная регрессия с одинаковыми регрессорами в обоих уравнениях. Но если накладывается априорное ограничение в виде равенства π_2 нулю и, следовательно, $EXPR_i$ исключается из второго уравнения, то указанная модель превращается в SUR-модель. SUR-оценка более эффективна по сравнению с оценкой многомерной регрессии, поскольку она использует исключяющие ограничения.

Сравнение между различными методами оценивания, рассмотренными в данном параграфе, показано в табл. 4.2.

Контрольные вопросы

1. (FIVE без условной гомоскедастичности.) Состоятельна ли оценка $\hat{S}$ в (4.5.3) для S без требования условной гомоскедастичности? [Ответ: Нет. Почему?] Состоятельна ли оценка (4.3.2) для (4.5.2) при условной гомоскедастичности? [Ответ: Да. Почему?]
2. (FIVE без условной гомоскедастичности.) Является ли оценка FIVE состоятельной и асимптотически нормальной без требования условной гомоскедастичности? А эффективной? **Указание:** FIVE-оценка есть эффективная GMM-оценка системы уравнений (4.2.3) для некоторой $\tilde{W}$. Удовлетворяет ли $\tilde{W}$ условию эффективности $\tilde{W} = S^{-1}$?
3. (Когда FIVE-оценка и 2SLS-оценка каждого уравнения в отдельности численно одинаковы.) Предположим, что все уравнения идентифицируемы точно. Являются ли в таком случае FIVE-оценка и 2SLS-оценка каждого уравнения в отдельности численно эквивалентными (то есть умноженная на $\sqrt{n}$ оценка разности этих оценок стремится к нулю)? А численно разными? **Указание:** Обе оценки FIVE и 2SLS сводятся к IV.
4. (Когда FIVE-оценка и 2SLS-оценка каждого уравнения в отдельности асимптотически эквивалентны.) Если ошибки являются сферическими ($\sigma_{mh} = 0$ для $m \neq h$), то FIVE-оценка и 2SLS-оценка каждого уравнения в отдельности

Соотношения между оценками системы уравнений

GMM для нескольких уравнений		FIVE	2SLS	SUR	Многочленная регрессия
Модель	Предположения 4.1–4.6	Предположения 4.1–4.5, 4.7 $E(x_{im}x'_{ik})$ конечно	Предположения 4.1–4.5, 4.7 $E(x_{im}x'_{ik})$ конечно $x_{im} = x_k$ для всех m	Предположения 4.1–4.5, 4.7 $x_{im} \rightarrow x_k$ для всех m , x_k — объединенные $x_{k1}, \dots, x_{km}$	Предположения 4.1–4.5, 4.7, $x_{im} \rightarrow x_k$ для всех m , $x_{im} = x_k$ для всех m
$S(\equiv \text{Avar}(\bar{g}))$	(4.1.11)	(4.5.2)	$\Sigma \otimes b(x_k x'_k)$	$\Sigma \otimes E(x_k x'_k)$	неявно
$\bar{g}$	(4.3.2)	(4.5.3)	$\bar{\Sigma} \otimes (\pi^{-1} \Sigma_c \omega_k x'_k)$ $\bar{\Sigma}$ из 2SLS-остатков	$\bar{\Sigma} \otimes (\pi^{-1} \Sigma_c \omega_k x'_k)$ $\bar{\Sigma}$ из 1SLS-остатков	неявно
$\bar{g}(\bar{S}^{-1})$	упрощенный нет	упрощенный нет	(4.5.12) с (4.5.13), (4.5.13)	(4.5.12) с (4.5.13'), (4.5.14')	поочередное OLS-оценивание каждого уравнения
$\text{Avar} \bar{g}(\bar{S}^{-1})$	(4.3.3)	(4.3.3)	(4.5.15) с (4.5.16)	(4.5.15) с (4.5.16')	формула для OLS
$\widehat{\text{Avar}} \bar{g}(\bar{S}^{-1})$	(4.3.4)	(4.5.4)	(4.5.17) с (4.5.13)	(4.5.17) с (4.5.13')	формула для OLS

асимптотически эквивалентны (и эта разность этих оценок, умноженная на $\sqrt{n}$, стремится к нулю). **Указание:** Итогами требования условной гомоскедастичности на утверждение 4.3(b). При условной гомоскедастичности эффективный GMM для системы уравнений сводится к FIVE и эффективный GMM для каждого уравнения в отдельности сводится к 2SLS для каждого уравнения в отдельности. Также для условной гомоскедастичности уравнения являются «несвязанными» в смысле (4.4.3), если $\sigma_{m,n} = 0$.

5. (Условная гомоскедастичность при предположении SUR.) Проверьте, что при предположении SUR (4.5.18) предположения 4.7 принимают вид:

$$E(\varepsilon_m \varepsilon_n | \mathbf{z}_m, \dots, \mathbf{z}_M) = \sigma_{mn} \text{ для всех } m, n = 1, 2, \dots, M.$$

6. Что произойдет с SUR-оценкой из примера 4.7, если $\mathbf{z}_n$ будет включать в себя также $\Delta E/D$? **Указание:** $\mathbf{z}_n$ по-прежнему «исчезает» из (4.5.13), (4.5.14) и (4.5.16).

А что произойдет со статистикой Саргана?

7. (Идентифицируемость в SUR.) Из утверждения 4.6 условно идентифицируемости (предположение 4.4) не является необходимым, так как оно вытекает из других предположений. Проверьте, что предположение 4.4 вытекает из предположения 4.5 и 4.7. **Указание:** Из предположения 4.5 и разложения Крамера для $\mathbf{B}$ следует, что $E(\mathbf{z}_m \mathbf{z}_n')$ невырождена. $\mathbf{z}_{m1}$ есть подмножество $\mathbf{z}_m$.

8. (SUR без условной гомоскедастичности.)

(a) Являются ли SUR-оценки состоятельными в асимптотически нормальными без требования условной гомоскедастичности? **Указание:** SUR-оценка по-прежнему представляет собой GMM-оценку системы уравнений, так что можно использовать утверждение 3.1.

(b) Можно ли свести SUR к множественной регрессии и в случае, когда система идентифицирована точно, то есть когда идентифицировано каждое уравнение (так что $\mathbf{z}_{m1} = \mathbf{z}_m$), без требования условной гомоскедастичности? [Ответ: Да.]

9. (Роль условий «перекрестной» ортогональности.) Предположим, что вместо (4.5.18) условия ортогональности выглядят как $E(\mathbf{z}_{m1} \varepsilon_n) = 0$ для $m, n = 1, 2, \dots, M$. Является ли SUR-оценка состоятельной в этом случае? [Ответ: Необязательно.] Найдите эффективную GMM-оценку системы уравнений без требования условной гомоскедастичности. [Ответ: Это OLS.]

10. Проверьте, что статистика Саргана для кросс-секционной регрессии равна нулю.

4.6. Общие коэффициенты

На практике, особенно в работе с панельными данными, часто приходится иметь дело с частным случаем системы уравнений, в котором количество регрессоров одинаково во всех уравнениях (соответствует

с одинаковым набором коэффициентов). Такая модель представляет собой частный случай модели из нескольких уравнений из параграфа 4.1. Здесь показано, как включить ОММ в рассмотрение уравнения, задавая общие коэффициенты на коэффициенты. В итоге будет показано, что модель с допускающим наличием ограничений (*seemingly restricted model*) в действительности включает в себя как частный случай модель с несколькими уравнениями без общих коэффициентов.

Модель с общими коэффициентами

Система уравнений с общими коэффициентами представляется:

Предположение 4.1' (линейность): Предположим, имеют M линейных уравнений:

$$y_{it} = z'_{it}\delta + \varepsilon_{it}, i = 1, 2, \dots, M; t = 1, 2, \dots, n, \quad (4.6.1)$$

где n — размер выборки, z_{it} — L -мерный вектор регрессоров, δ — L -мерный вектор коэффициентов, общий для всех уравнений, а ε_{it} — независимая ошибка в i -м уравнении.

Среди других предположений модели нескольких уравнений, предположений 4.2–4.5, только предположение 4.4 (идентифицируемости) требует изменения, отражающего наличие ограничения в виде общих коэффициентов. Выражением $g(w_t; \delta)$ для нескольких уравнений является:

$$g(w_t; \delta) \equiv \begin{bmatrix} x_{1t} \cdot (y_{1t} - z'_{1t}\delta) \\ \vdots \\ x_{Mt} \cdot (y_{Mt} - z'_{Mt}\delta) \end{bmatrix}, \quad (4.6.2)$$

так что $E[g(w_t; \delta)]$ превращается в

$$E[g(w_t; \delta)] = \begin{bmatrix} E(x_{1t} \cdot y_{1t}) \\ \vdots \\ E(x_{Mt} \cdot y_{Mt}) \end{bmatrix} - \begin{bmatrix} E(x_{1t} \cdot z'_{1t})\delta \\ \vdots \\ E(x_{Mt} \cdot z'_{Mt})\delta \end{bmatrix} = \\ = \frac{\sum_{t=1}^n \sigma_{xy}}{\sum_{t=1}^n N_{t \times 1}} + \frac{\sum_{t=1}^n \sigma_{xz}}{\sum_{t=1}^n N_{t \times L}} \bar{\delta}, \quad (4.6.3)$$

где

$$\sigma_{xy} \equiv \begin{bmatrix} E(x_{1t} \cdot y_{1t}) \\ \vdots \\ E(x_{Mt} \cdot y_{Mt}) \end{bmatrix}, \quad \sigma_{xz} \equiv \begin{bmatrix} E(x_{1t} \cdot z'_{1t}) \\ \vdots \\ E(x_{Mt} \cdot z'_{Mt}) \end{bmatrix}. \quad (4.6.4)$$

То есть σ_{xz} теперь является стандартизированной, а не блочно-диагональной матрицей. С этим изменением система уравнений, определяющая $\bar{\delta}$, снова есть (4.1'–(4)), так что условием идентифицируемости является

Предположение 4.4' (идентифицируемость при общих коэффициентах): Матрица Σ_{xx} размерности $\sum_{m=1}^M K_m \times L$, определенная в (4.6 4), имеет полный ранг.

Это условие слабее, чем предположение 4.4, которое требует, чтобы каждое уравнение системы было идентифицируемо. На самом деле достаточным условием идентифицируемости является полный столбцовый ранг матрицы $\Pi(\mathbf{x}_{im}, \mathbf{z}'_{im})$ для некоторого¹ m . Коэффициенты во всех уравнениях одна и та же благодаря именно споровому ограничению. Система может быть идентифицируемой, даже если ни одно уравнение в отдельности не является идентифицируемым.

GMM-оценка

Оставляем читателю проверить, что $g_n(\hat{\delta})$ можно записать как $\varepsilon_{1n} - \delta_{22}\hat{\delta}$ с

$$S_{22} = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1} \cdot w_{i1} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{iM} \cdot w_{iM} \end{bmatrix}, \quad S_{xx} = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1} z'_{i1} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{iM} z'_{iM} \end{bmatrix}, \quad (4.6 5)$$

так что GMM-оценка со взвешивающей матрицей $\bar{W}$ размера $\sum_m K_m \times \sum_m K_m$ есть (4.2 3) с ошибкой оценки из (4.2 4). При условии, что $\Sigma_{\varepsilon\varepsilon}$ и S_{xx} перепределены соответствующим образом к $\hat{\delta}$ интерпретируются, как набор общих коэффициентов, теория больших выборок для конечных нескольких уравнений та же самая, что и для случая самостоятельных коэффициентов из параграфа 4.3 с заменой предположения 4.1 на предположение 4.1', предположения 4.4 на предположение 4.4' и оценка $\hat{\delta}$ перепределенными как $y_{im} - \varepsilon'_{im}\hat{\delta}$ (а не $y_{im} - \varepsilon'_{im}\hat{\delta}_{im}$) для некоторой самостоятельной оценки L -много вектора общих коэффициентов δ .

Чтобы связать данную GMM-оценку с другими популярными оценками, имеющимися в литературе, нужно подробно раскрыть формулу для данной GMM-оценки. Подставляя (4.6 5) в (4.2 3) получаем:

$$\hat{\delta}(\bar{W}) = \left(\begin{bmatrix} \frac{1}{n} \sum_{i=1}^n \varepsilon_i x'_{i1} & \cdots & \frac{1}{n} \sum_{i=1}^n \varepsilon_i x'_{iM} \end{bmatrix} \begin{bmatrix} \bar{W}_{11} & \cdots & \bar{W}_{1M} \\ \vdots & & \vdots \\ \bar{W}_{M1} & \cdots & \bar{W}_{MM} \end{bmatrix} \right)^{-1} \times$$

¹Дополнительно: $\Pi(\mathbf{x}_{im}, \mathbf{z}'_{im})$ имеет полный столбцовый ранг, то есть ранг равен L . Поскольку эти векторы разных категорий линейно независимых строк, $\Pi(\mathbf{x}_{1m}, \mathbf{z}'_{1m})$ имеет L линейно независимых строк. Поэтому Σ_{xx} имеет не менее L линейно независимых строк.

$$\begin{aligned}
& \times \left[\begin{array}{c} \frac{1}{n} \sum_{i=1}^n x_{i1} z'_{i1} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{iM} z'_{iM} \end{array} \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n x_{i1} x'_{i1} \quad \cdots \quad \frac{1}{n} \sum_{i=1}^n x_{iM} x'_{iM} \right] \\
& \times \left[\begin{array}{ccc} \widehat{W}_{11} & \cdots & \widehat{W}_{1M} \\ \vdots & & \vdots \\ \widehat{W}_{M1} & \cdots & \widehat{W}_{MM} \end{array} \right] \left[\begin{array}{c} \frac{1}{n} \sum_{i=1}^n x_{i1} \cdot y_{i1} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{iM} \cdot y_{iM} \end{array} \right] = \\
& = \left[\sum_{m=1}^M \sum_{h=1}^M \left\{ \left(\frac{1}{n} \sum_{i=1}^n z_{im} x'_{im} \right) \widehat{W}_{mh} \left(\frac{1}{n} \sum_{i=1}^n x_{ih} z'_{ih} \right) \right\} \right]^{-1} \times \\
& \times \left[\sum_{m=1}^M \sum_{h=1}^M \left\{ \left(\frac{1}{n} \sum_{i=1}^n z_{im} x'_{im} \right) \widehat{W}_{mh} \left(\frac{1}{n} \sum_{i=1}^n x_{ih} \cdot y_{ih} \right) \right\} \right]. \quad (4.6.6)
\end{aligned}$$

где $\widehat{W}_{mh}$ есть блок (m, h) матрицы $\widehat{W}$. (Для получения второго равенства используется формула (A.8) из приложения А.) Эффективную GMM-оценку можно получить, если заменить в данном выражении $\widehat{W}$ на матрицу, обратную матрице $\widehat{S}$, определенной в (4.3.2).

Введение требования условной гомоскедастичности

Эффективную GMM-оценку можно получить, положив в (4.6.6) $\widehat{W} = \widehat{S}^{-1}$, где $\widehat{S}$ есть состоятельная оценка матрицы S . Как и ранее, мы можем наложить требование условной гомоскедастичности, чтобы свести GMM-оценку к более известным оценкам. Если задать $\widehat{S}$ соотношением (4.5.3), то получится FIVE-оценка.

Если мы предположим еще, что набор инструментов одинаков для всех уравнений, то тогда, как и ранее, эту $\widehat{S}$ можно представить как произведение Кронекера:

$$\widehat{S} = \widehat{\Sigma} \otimes \left(\frac{1}{n} \sum_{i=1}^n x_i x'_i \right),$$

так что $\widehat{W}_{mh}$ в (4.6.6) задается соотношением (4.5.1), что дает нам так называемую 3SLS-оценку с общими коэффициентами:

$$\begin{aligned}
& \left[\sum_{m=1}^M \sum_{h=1}^M \left\{ \widehat{\sigma}^{mh} \cdot \left(\frac{1}{n} \sum_{i=1}^n z_{im} x'_i \right) \left(\frac{1}{n} \sum_{i=1}^n x_i x'_i \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n x_i z'_{ih} \right) \right\} \right]^{-1} \times \\
& \times \left[\sum_{m=1}^M \sum_{h=1}^M \left\{ \widehat{\sigma}^{mh} \cdot \left(\frac{1}{n} \sum_{i=1}^n z_{im} x'_i \right) \left(\frac{1}{n} \sum_{i=1}^n x_i x'_i \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n x_i \cdot y_{ih} \right) \right\} \right]. \quad (4.6.7)
\end{aligned}$$

где $\hat{\sigma}^{mh}$ — блок (m, h) матрицы $\hat{S}^{-1}$. Если, в дополнение, выполняется условие SUR (4.5.18), то тогда происходит «сжатывание» в (4.6.7) и эффективная GMM-оценка превращается в так называемую (по историческим причинам) *оценку со случайными эффектами* (random-effects estimator):

$$\hat{\delta}_{RE} = \left[\sum_{m=1}^M \sum_{h=1}^M \hat{\sigma}^{mh} \left(\frac{1}{n} \sum_{i=1}^n z_{im} z'_{ih} \right) \right]^{-1} \sum_{m=1}^M \sum_{h=1}^M \hat{\sigma}^{mh} \left(\frac{1}{n} \sum_{i=1}^n z_{im} \cdot y_{ih} \right). \quad (4.6.8)$$

Ее асимптотическая дисперсия

$$\text{Avar}(\hat{\delta}_{RE}) = \left[\sum_{m=1}^M \sum_{h=1}^M \sigma^{mh} E(z_{im} z'_{ih}) \right]^{-1}. \quad (4.6.9)$$

(Чтобы получить это выражение, вернемся к общей формуле (4.3.3) для асимптотической дисперсии эффективной GMM-оценки. Подождем Σ_{zz} как в (4.6.4), $S = \Sigma \otimes E(x_i x'_i)$, воспользуемся формулой (A.5) из приложения A для вычисления производящей матрицы, и обнаружим «сжатывание» до, что превращает (4.5.16) в (4.5.16') на с. 309 при условии (4.5.18) для SUR.) Для этой асимптотической дисперсии можно получить достаточно хорошую оценку:

$$\widehat{\text{Avar}}(\hat{\delta}_{RE}) = \left[\sum_{m=1}^M \sum_{h=1}^M \hat{\sigma}^{mh} \left(\frac{1}{n} \sum_{i=1}^n z_{im} z'_{ih} \right) \right]^{-1}. \quad (4.6.10)$$

Получаем результат об оценке случайных эффектов

Утверждение 4.7 (свойства на больших выборках оценки случайных эффектов): *Предположим, что предположения 4.1', 4.2, 4.3, 4.4', 4.5 и 4.7 выполняются в отношении объектов $(z_{1i}, \dots, z_{Mi})$. Пусть Σ — матрица размера $M \times M$, элемент (m, h) которой равен $E(z_{im} z'_{ih})$, и пусть $\hat{\Sigma}$ является состоятельной оценкой Σ . Тогда*

- Оценка $\hat{\delta}_{RE}$, заданная в (4.6.8), является состоятельной, асимптотически нормальной и эффективной с $\text{Avar}(\hat{\delta}_{RE})$, заданной в (4.6.9).*
- Оцененная асимптотическая дисперсия (4.6.10) является состоятельной для $\text{Avar}(\hat{\delta}_{RE})$.*
- (Статистика Саргала)*

$$J(\hat{\delta}_{RE}, \hat{S}^{-1}) \equiv n \cdot g_n(\hat{\delta}_{RE})' \hat{S}^{-1} g_n(\hat{\delta}_{RE}) \rightarrow_d \chi^2(MK - L),$$

где $\hat{S} = \hat{\Sigma} \otimes \left(\frac{1}{n} \sum_{i=1}^n x_i x'_i \right)$, K — количество объектов инструмента и $g_n(\hat{\delta}) = y_{xy} - S_{xz} \hat{\delta}$ с y_{xy} и S_{xz} задано в (4.6.5).

Нижже мы перепишем формулы (4.6.8)–(4.6.10) в более элегантном виде

OLS в модели пула

Для SUR из предположений параграфа мы построили $\hat{\Sigma}$ из остатков, полученных OLS для каждого уравнения. Состоятельность — это все, что требуется для $\hat{\Sigma}$ поэтому аналогичная процедура для $\hat{\Sigma}$ работает здесь точно так же, но распределение $\hat{\delta}_{\text{SUR}}$ из конечных выборок можно улучшить, если мы используем априорное суждение о том, что коэффициенты одинаковы во всех уравнениях в оценивании $\hat{\Sigma}$. Приравняем поэтому $\hat{W}$ в GMM-оценки на первом шаге для получения наилучшей состоятельной оценки δ

$$W \approx \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1},$$

а так

$$\hat{\Sigma}^{-1} \approx \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1}.$$

Эта оценка дается формулой (4.6.8) с $\delta^{mh} = 1$ для $m = h$ и 0 для $m \neq h$, что можно записать как

$$\begin{aligned} \hat{\delta}_{\text{pooled OLS}} &= \left[\sum_{m=1}^M \left(\frac{1}{n} \sum_{i=1}^n z_{im} z_{im}' \right) \right]^{-1} \sum_{m=1}^M \left(\frac{1}{n} \sum_{i=1}^n z_{im} \cdot y_{im} \right) = \\ &= \left(\sum_{i=1}^n \sum_{m=1}^M z_{im} z_{im}' \right)^{-1} \sum_{i=1}^n \sum_{m=1}^M z_{im} \cdot y_{im}, \end{aligned} \quad (4.6.11)$$

что является просто OLS-оценкой по выборке объема Mn , где наблюдения из разных уравнений объединяются (или *pooled across equations*). По этой причине оценка называется **объединенной OLS** (*pooled OLS*) оценкой. Из этого выражения следует, что необходимыми условиями ортогональности являются условия

$$E(z_{1m} \cdot z_{1h}) + \dots + E(z_{im} \cdot z_{ih}) = 0, \quad (4.6.12)$$

которые не включают «кросс-ортогональности» $E(z_{im} \cdot z_{jh}) = 0$ ($m \neq h$). В качестве аналитического упражнения следует показать, что объединенная OLS-оценка является GMM-оценкой, соответствующей (4.6.12).

Поскольку объединенную OLS-оценку легко вычислить, она является популярной опцией для тех исследователей, которые не хотят программировать получение оценок. Эта оценка также робастна к нарушению условий «кросс-ортогональности». Однако также иметь в виду то, что

стандартные ошибки, которые прижизнятся в пакетах программ и которые не принимают во внимание перекрестные моменты внутри уравнений (σ_{mih}), являются смещенными. Для объединенной OLS-оценки, которая является GMM-оценкой с неоптимальным выбором $\bar{W}$, корректной формулой для асимптотической дисперсии является (3.5.1) и утверждение 3.1. Полагая $\bar{W} = I_M \otimes [E\{z_i z_i'\}]^{-1}$, $S = \Sigma \otimes E\{w_i w_i'\}$, Σ_{xz} как в (4.6.4) и снова наблюдая исчезающие x , мы получаем:

$$\text{Avar}(\hat{\delta}_{\text{pooled OLS}}) = \left[\sum_{m=1}^M E(z_{im} z_{im}') \right]^{-1} \times \\ \times \sum_{m=1}^M \sum_{h=1}^M \sigma_{mih} E(z_{im} z_{ih}') \left[\sum_{m=1}^M E(z_{im} z_{im}') \right]^{-1}, \quad (4.6.13)$$

которая состоятельно оценивается посредством

$$\left(\sum_{m=1}^M \frac{1}{n} \sum_{i=1}^n z_{im} z_{im}' \right)^{-1} \sum_{m=1}^M \sum_{h=1}^M \hat{\sigma}_{mih} \frac{1}{n} \sum_{i=1}^n z_{im} z_{ih}' \left(\sum_{m=1}^M \frac{1}{n} \sum_{i=1}^n z_{im} z_{im}' \right)^{-1}. \quad (4.6.14)$$

Поскольку объединенная OLS является состоятельной оценкой, соответствующие остатки можно использовать для вычисления $\hat{\sigma}_{mih}$ ($m, h = 1, \dots, M$) в этом выражении. Корректными стандартными ошибками для объединенной OLS будут квадратные корни из (умноженных на $1/n$) диагональных элементов этой матрицы.

Приведение формул к более красивому виду

Приведенные формулы для оценки случайных эффектов и объединенной OLS-оценки вывести достаточно просто, но выглядят они с двойным и тройным суммированием довольно сложными. Их можно, однако, привести к более красивому виду — и в то же время к более полезному для программирования, — если ввести новые матричные обозначения. Пусть

$$y_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iM} \end{bmatrix}, \quad Z_i = \begin{bmatrix} z'_{i1} \\ \vdots \\ z'_{iM} \end{bmatrix}, \quad \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iM} \end{bmatrix}, \quad (4.6.15)$$

так что систему из M уравнений с общими коэффициентами в предположении 4.1' можно записать как

$$y_i = Z_i \delta + \varepsilon_i \quad (i = 1, 2, \dots, n). \quad (4.6.1')$$

Приведение к более красивому виду сложных формул использует следующие алгебраические результаты:

$$\sum_{m=1}^M z_{im} z'_{im} = Z_i' Z_i, \quad \sum_{m=1}^M z_{im} \cdot y_{im} = Z_i' y_i, \quad (4.6.16a)$$

$$\sum_{m=1}^M \sum_{k=1}^M c_{mk} \cdot z_{im} \cdot y_{ik} = Z_i' C y_i, \quad \sum_{m=1}^M \sum_{k=1}^M c_{mk} \cdot z_{im} z'_{ik} = Z_i' C Z_i, \quad (4.6.16b)$$

для любой $M \times M$ матрицы $C = (c_{mk})$. Возьмем теперь тройные суммирование в (4.6.5),

$$\begin{aligned} \sum_{m=1}^M \sum_{k=1}^M \tilde{\sigma}^{mk} \cdot \left(\frac{1}{n} \sum_{i=1}^n z_{im} \cdot y_{ik} \right) &= \\ &= \frac{1}{n} \sum_{i=1}^n \left(\sum_{m=1}^M \sum_{k=1}^M \tilde{\sigma}^{mk} \cdot z_{im} \cdot y_{ik} \right) \\ &\quad (\text{изменяя порядок суммирования}), \\ &= \frac{1}{n} \sum_{i=1}^n Z_i' \tilde{\Sigma}^{-1} y_i \quad (\text{по (4.6.16a) с } C = \tilde{\Sigma}^{-1}). \end{aligned} \quad (4.6.17)$$

Аналогично по (4.6.16b) с $C = \tilde{\Sigma}^{-1}$, другие тройные суммирования в (4.6.5) принимают вид

$$\sum_{m=1}^M \sum_{k=1}^M \tilde{\sigma}^{mk} \frac{1}{n} \sum_{i=1}^n z_{im} z'_{ik} = \frac{1}{n} \sum_{i=1}^n Z_i' \tilde{\Sigma}^{-1} Z_i. \quad (4.6.18)$$

Таким же образом можно записать двойные и тройные суммирования в (4.6.9) и (4.6.10). Мы имеем тогда переписанные формулы для случайных эффектов в виде

$$\tilde{\delta}_{RE} = \left(\frac{1}{n} \sum_{i=1}^n Z_i' \tilde{\Sigma}^{-1} Z_i \right)^{-1} \frac{1}{n} \sum_{i=1}^n Z_i' \tilde{\Sigma}^{-1} y_i, \quad (4.6.9')$$

$$\text{Avar}(\tilde{\delta}_{RE}) = \{E(Z_i' \tilde{\Sigma}^{-1} Z_i)\}^{-1}, \quad (4.6.9'')$$

$$\text{Avar}(\tilde{\delta}_{UE}) = \left(\frac{1}{n} \sum_{i=1}^n Z_i' \tilde{\Sigma}^{-1} Z_i \right)^{-1}. \quad (4.6.10')$$

Используя (4.6.16), формулы для обобщенной ОLS-оценки можно записать как

$$\hat{\delta}_{\text{обобщ. OLS}} = \left(\frac{1}{n} \sum_{i=1}^n Z_i Z_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n Z_i' y_i. \quad (4.6.11')$$

$$\text{Асим}(\hat{\delta}_{\text{обобщ. OLS}}) = E(Z_i Z_i')^{-1} E(Z_i' \Sigma Z_i) E(Z_i Z_i')^{-1}. \quad (4.6.13')$$

$$\text{Асим}(\hat{\delta}_{\text{обобщ. OLS}}) = \left(\frac{1}{n} \sum_{i=1}^n Z_i Z_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n Z_i' \hat{\Sigma} Z_i \right) \left(\frac{1}{n} \sum_{i=1}^n Z_i Z_i' \right)^{-1}. \quad (4.6.14')$$

Эти формулы не просто сложны, как будет видно в следующем параграфе, они будут полезны для обработки определенных классов панельных данных (так называемых **несбалансированных панельных (unbalanced panel)** данных).

Ограничения, которые не являются ограничивающими

Хотя заметим, что модель в этом параграфе с предложенным обобщением коэффициентов является частным случаем модели из параграфа 4.1, все же можно привести к формату модели данного параграфа при заданном первом предельном регрессоре. Рассмотрим, например, пример 4.1. Вместе с определениями x_{i1} и x_{i2} как в (4.1.2), определим их как

$$x_{i1} = \begin{bmatrix} 1 \\ S_i \\ IQ_i \\ EXPR_i \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad x_{i2} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ S_i \\ IQ_i \end{bmatrix}, \quad \text{с } \delta = \begin{bmatrix} \phi_1 \\ \beta_1 \\ 0 \\ \pi \\ \phi_2 \\ \beta_2 \\ \gamma_2 \end{bmatrix}.$$

Тогда система из двух уравнений соответствует формату предложения 4.1'. (См. контрольный вопрос 5 ниже для демонстрации того же самого в более общем случае.)

Более того, систему, в которой общие коэффициенты имеют только некоторое подмножество переменных, можно записать в виде предложения 4.1'. Рассмотрим пример 4.2. Допустим, что мы хотим *эквивалентно* предположить, что коэффициенты π и ϕ_2 являются константами, IQ и $EXPR$ остаются постоянными во времени (так что $S_1 = S_2 = S$, $\gamma_1 = \gamma_2 = \gamma$, $\pi_1 = \pi_2 = \pi$), но константы различны. Преобразование осуществляется следующим образом:

$$z_{i1} = \begin{bmatrix} 1 \\ 0 \\ S\delta\theta_i \\ IQ_i \\ EXPR\delta\theta_i \end{bmatrix}, \quad z_{i2} = \begin{bmatrix} 0 \\ 1 \\ S\delta\theta_i \\ IQ_i \\ EXPR\delta\theta_i \end{bmatrix}, \quad \delta = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \delta \\ \gamma \\ \alpha \end{bmatrix}. \quad (4.6.19)$$

Следовательно, ограничение, требующее общих коэффициентов, не является ограничивающим. Мы могли бы получить GMM-оценку для модели из параграфа 4.1 как частный случай GMM-оценки с ограничением, требующим общих коэффициентов. Но модель без такого ограничения, возможно, является более простым выражением в многомерном уравнении.

Контрольные вопросы

1. (Идентификация с общими коэффициентами.) Пусть z_{imj} является j -м элементом z_{im} . То есть z_{imj} является j -м регрессором в m -м уравнении. Предположим, что $z_{imj} = 1$ для всех i, m . Какие из предположений, сделанных для этой модели, нарушаются, если $z_{im2} = 1$ для всех i, m ? **Указание:** Посмотрите на различные уравнения. Изменится ли как-нибудь, если $z_{im2} = 10$ для всех i, m ? **Указание:** Должен измениться.

2. (Важность «кресте-ортогональности».) Предположим, что $E(z_{im} \cdot \varepsilon_{im}) = 0$ для $m = 1$, но не обязательно для $m \neq 1$. Будет ли оценка случайных эффектов состоятельной? **Указание:** Ошибка оценки случайных эффектов равна:

$$\hat{\delta}_{FE} - \delta = \left[\sum_{m=1}^M \sum_{i=1}^N \bar{\sigma}^{im} \left(\frac{1}{N} \sum_{i=1}^N z_{im} z'_{im} \right) \right]^{-1} \sum_{m=1}^M \sum_{i=1}^N \bar{\sigma}^{im} \left(\frac{1}{N} \sum_{i=1}^N z_{im} \cdot \varepsilon_{im} \right).$$

3. (Неэффективность объясняемой OLS оценки.) Выведите эффективную GMM-оценку для δ , которая использует тот факт, что $E(z_{im} \cdot \varepsilon_{im}) = 0$ для всех m . Является ли она той же самой, что и объединенная OLS-оценка? [Ответ: Нет.] Является ли она такой же, что и оценка случайных эффектов? [Ответ: Нет.]

4. (Σ_{xx} , Σ_{xz} , Σ_{zz} с общими инструментами в кросс-секционных приложениях.) Предположим, что инструменты во всех уравнениях одинаковы: $x_{im} = x_i$. Покажите, что Σ_{xx} в (4.6.4) можно записать как $E(Z_i Z_i')$, где Z_i является $M \times L$ матрицей, определенной в (4.6.15). Также $\Sigma_{xz} = \frac{1}{N} \sum_{i=1}^N y_i \cdot x_i'$, $\Sigma_{zz} = -\frac{1}{N} \sum_{i=1}^N Z_i' \cdot x_i$. **Указание:** $x_i z'_{im} = z'_{im} \cdot x_i$.

5. (Ограничение, не являющееся ограничивающим.) Чтобы исключить возможность путаницы, возникающую из того факта, что z_{im} в параграфе 4.1 и z_{im} в этом параграфе различны, запишем систему из M уравнений (4.6.11) как

$$y_{im} = z'_{im} \delta_{im} + \varepsilon_{im} \quad (m = 1, 2, \dots, M; \quad i = 1, 2, \dots, N). \quad (1)$$

где z'_{im} и δ_{im} являются L -векторы.

- (я) (Предположение 4.1 как частный случай предположения 4.1') Проверьте, что (1) можно записать как (4.6.1), если определить

$$\begin{aligned} (\Sigma_{m+1}^{M+1}, L_{m+1} \times 1) &= \begin{bmatrix} 0 \\ \vdots \\ 0 \\ z_{m+1}^* \\ 0 \end{bmatrix} \quad \left. \begin{array}{l} L_1 \text{ строк} \\ \vdots \\ L_m \text{ строк} \\ L_{m+1} \text{ строк} \\ \vdots \\ L_M \text{ строк} \end{array} \right\} \\ \text{и } (\Sigma_{m+1}^{M+1}, L_{m+1} \times 1) &= \begin{bmatrix} \delta_1^* \\ \vdots \\ \delta_m^* \end{bmatrix}. \end{aligned} \quad (2)$$

- (b) (Предположение 4.4 как частный случай предположения 4.4') Проверьте, что, когда z_{im} определяется как в (2) с z_{im}^* интерпретируемым как z_{im} в параграфе 4.1, предположение 4.4' становится предположением 4.4.
- (c) (Формула GMM.) При такой же интерпретации проверьте, что матрица S_{zz} в (4.6.5) становится блочно-диагональной, m -й блок которой равен $\frac{1}{n} \sum_i z_{im} z_{im}'$, так что (4.6.6) становится (4.2.6), если z_{im} в (4.2.6) понимается как z_{im}^* .
6. (Идентификация JRF модели.) Предположение 4.4' и утверждение 4.7 фактически не являются необходимыми. Проверьте, что из предположений 4.5 и 4.7 следует предположение 4.4'. **Указание:** Если $E(z_i x_i')$ невырождена, то предположение 4.1' выполняется, как показано в контрольном упражнении 7 параграфа 4.3. Предположение 4.4' более слабое, чем предположение 4.1.
7. (Только для тех, кто знаком с кронекеровским произведением.)
- (a) Покажите, что $E(Z_i' \Sigma^{-1} Z_i)$ в (4.6.9') из с. 321 для оценки случайных эффектов является невырожденной, если выполнены предположения утверждения 4.7. **Указание:**

$$\begin{aligned} E(Z_i' \Sigma^{-1} Z_i) &= \sum_{m=1}^M \sum_{n=1}^M \sigma^{mn} E(z_{im} z_{in}') = \\ &= \sum_{m=1}^M \sum_{n=1}^M \sigma^{mn} E(z_{im} z_{in}') [E(z_i z_i')]^{-1} E(z_i z_{in}') = \\ &\quad (\text{т. может быть использовано, поскольку } z_i \text{ включает } z_{im}, z_{in}) \\ &= \Sigma_{zz}' (\Sigma^{-1} z_i) E(z_i z_i')^{-1} \Sigma_{zz}. \end{aligned}$$

Σ_{zz} имеет полный левый столбцовый ранг по предположению 4.4'

- (b) Покажите, что $E(Z_i' Z_i)$ в (4.6.13') на с. 321 для объясненного OLS в модели Гуд является обратимой. **Указание:** Замечьте Σ^{-1} на i

4.7. Приложение. Спрос на взаимосвязанные факторы

Транслогарифмическая функция издержек (*translog cost function*) представляет собой обобщение функции издержек типа Кобба — Дугласа (логарифмической), рассмотренной в параграфе 1.7. Привлекательной чертой транслогарифмической спецификации является то, что уравнение спроса на факторы, минимизирующие издержки (преобразованные в уравнении для долей затрат на факторы и издержки), является линейным по логарифмам выпуска и ценам факторов. При этом коэффициенты представляют собой (частичные) параметры функции издержек, определяющих технопонты. Эти параметры технологии могут быть оценены из системы уравнений для долей затрат. Одним из первых эту идею реализовали [Berndt and Wood, 1975] и [Christensen and Greene, 1976]. Этот параграф посвящен изложению основ их подхода. Данный подход также может быть применен для анализа потребительского спроса.

Чтобы лучше сфокусироваться на интересующем нас вопросе и следовать логике этих основополагающих работ, дальнейшее изложение выстроено при предположении об условной гомоскедастичности. Кроме того, мы будем предполагать, что регрессоры (логарифмы цен факторов и логарифм выпуска) в системе уравнений для долей являются предопределенными. Как мы уже обсуждали в параграфе 1.7, это предположение является вполне обоснованным для отрасли производства и распределения электроэнергии США до недавнего дерегулирования отрасли. Следовательно, подходящим методом оценивания нескольких уравнений является модель многомерной регрессии.

Транслогарифмическая функция издержек

Мы уже частично описали переход от логарифмической к транслогарифмической функции издержек в интервальном утверждении главы 1: мы реализовали идею добавить квадрат логарифма выпуска в логарифмную функцию издержек (см. модель 4). Если этот подход расширить и добавить квадраты и перекрестные произведения логарифмов всех $\ln x$ -метронов, то логарифмическая функция издержек с тремя аргументами (1-74) становится транслогарифмической функцией вида:

$$\begin{aligned} \log(C) = & \alpha_0 - \sum_{j=1}^3 \gamma_{1j} \log(y_j) + \frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^3 \gamma_{ij} \log(y_i) \log(y_j) + \\ & + \alpha_Q \log(Q) + \frac{1}{2} \alpha_{QQ} (\log(Q))^2 + \sum_{j=1}^3 \gamma_{jQ} \log(y_j) \log(Q) + \varepsilon. \end{aligned} \quad (4.7.1)$$

Здесь, как и в остальной части параграфа, мы опускаем индекс наблюдений i . В этом выражении слагаемое

$$\frac{1}{2} \sum_{i=1}^3 \sum_{k=1}^3 \gamma_{ik} \log(p_i) \log(p_k)$$

имеется квадратичная форма, представляющей вклад фактора второго порядка. Без потери общности можно предположить, что матрица коэффициентов квадратичной формы размера 3×3 , (γ_{ik}) симметрична¹:

$$\gamma_{ik} = \gamma_{ki} \quad (i, k = 1, 2, 3). \quad (4.7.2)$$

В случае логлинейной функции издержек, рассмотренной в параграфе 1.7, отлеча от масштаба может быть вычислена как единица, деленная на эластичность издержек по выпуску. Если это определение применить к транслогатифической функции издержек, получим, что

$$\begin{aligned} \text{отлеча от масштаба} &= \frac{1}{\partial \log(C) / \partial \log(Q)} = \\ &= \frac{1}{\alpha_Q - \gamma_{QQ} \log(Q) - \sum_{j=1}^J \gamma_{Qj} \log(p_j)} \end{aligned} \quad (4.7.3)$$

Лемма факторов

Связь между параметрами функции издержек и спросом на факторы определяется **леммой Шепарда** (*Shephard's Lemma*) из микроэкономики. Пусть x_j — спрос на фактор j , минимизирующий функцию издержек при фиксированных ценах факторов (p_1, p_2, p_3) и выпуске Q . То есть $\sum_{j=1}^3 p_j x_j = C$. Лемма гласит, что

$$\frac{\partial C}{\partial p_j} = x_j. \quad (4.7.4)$$

Используя выражение

$$\frac{\partial \log(C)}{\partial \log(p_j)} = \frac{p_j}{C} \frac{\partial C}{\partial p_j},$$

лемму можно переформулировать как равенство частной производной логарифма функции издержек по цене фактора, а именно

$$\frac{\partial \log(C)}{\partial \log(p_j)} = \frac{p_j x_j}{C} \quad (4.7.5)$$

¹ Пусть x — n -мерный вектор. Тогда $x'Ax$ — скалярная квадратичная форма. Поскольку $x'A'x = x'A'x$, то $x'A'x = x'[(A + A')/2]x$. Таким образом, если матрица A действительная, ее можно заменить симметрической матрицей $(A + A')/2$, не изменив при этом значение квадратичной формы.

В случае транслогарифмической функции (4.7.1) частная производная логарифма функции затрат вычисляется легко:

$$\frac{\partial \log(C)}{\partial \log(p_j)} = \alpha_j + \sum_{k=1}^3 \gamma_{jk} \log(p_k) + \gamma_{jQ} \log(Q) \quad (j = 1, 2, 3). \quad (4.7.6)$$

Соединяя это с (4.7.5) и определяя доли затрат на факторы как $\gamma_j \equiv p_j x_j / C$, получим следующую систему для долей затрат на факторы для транслогарифмической функции:

$$\gamma_j = \alpha_j + \sum_{k=1}^3 \gamma_{jk} \log(p_k) + \gamma_{jQ} \log(Q) \quad (j = 1, 2, 3). \quad (4.7.7)$$

На эту систему уравнений наложены **перекрестные ограничения** (cross-equation restrictions), состоящие в том, что коэффициент перед $\log(p_k)$ в уравнении для γ_j равен коэффициенту перед $\log(p_j)$ в уравнении для γ_k для $j \neq k$. До конца параграфа мы будем называть эти ограничения ограничениями симметрии, хотя для γ не выполняются следующие предположения о симметрии (4.7.2). Эти перекрестные ограничения являются следствием выкладки: проверьте, что без предположения о симметрии коэффициент при $\log(p_k)$ в уравнении для γ_j , например, будет равен $(\gamma_{12} + \gamma_{21})/2$ и будет совпадать с коэффициентом при $\log(p_1)$ в уравнении для γ_2 .

Эластичности замещения

В литературе, посвященной производственным функциям, большое внимание уделяется оцениванию эластичностей замещения (т.е. частные эластичности замещения Хикса — Аллена) между различными факторами производства¹. Как показано в [Уизман, 1962], эластичность замещения между факторами j и k , обозначаемая как η_{jk} , связана с функцией издержек C посредством следующего соотношения:

$$\eta_{jk} = \frac{C \cdot \frac{\partial^2 C}{\partial p_j \partial p_k}}{\frac{\partial C}{\partial p_j} \cdot \frac{\partial C}{\partial p_k}}. \quad (4.7.8)$$

Для транслогарифмической функции издержек применима лемма Шенгуда дает следующий результат:

$$\eta_{jk} = \begin{cases} \frac{\gamma_{jk} + \gamma_{kj}}{\gamma_j \gamma_k} & \text{для } j \neq k \\ \gamma_{jj} + \frac{\alpha_j^2}{\gamma_j} - \alpha_j & \text{для } j = k. \end{cases} \quad (4.7.9)$$

¹ История развития данного направления изложена в [Berndt, 1991, Section 9.2].

где v_j — доля затрат на фактор j в общих издержках. В частном случае логлинейной функции издержек, где $\gamma_{jk} = 0$ для всех j, k , получаем $\eta_{jk} = 1$ для $j \neq k$, что подтверждает то, что для технологии Кобба — Дугласа эластичности замещения между любыми двумя факторами равны единице.

Свойства функции издержек

Как хорошо известно из микроэкономики (см., например, [Varian, 1992, Chapter 8]), функция издержек обладает следующими свойствами.

1. (Однородность.) Однородность первой степени по ценам факторов.
2. (Моногонность.) Неубывание по ценам факторов, для чего необходимо, чтобы $\partial C / \partial p_j \geq 0$ для всех j .
3. (Вогнутость.) Вогнутость по ценам факторов, для чего необходимо, чтобы гессен (матрица вторых производных) функции издержек C ,

$$H \equiv (\partial^2 C / \partial p_j \partial p_k),$$

был отрицательно полуопределен. Из (4.7.8) следует, что требование вогнутости выполнено если и только если матрица эластичностей замещения (η_{jk}) отрицательно полуопределена¹.

В случае транслогарифмической функции издержек эти свойства можно переформулировать следующим образом.

1. *Однородность.* Несложно проверить (см. контрольные вопросы), что при введении ограничений симметрии на (γ_{jk}) условие однородности может быть переписано как:

$$\begin{cases} \alpha_1 + \alpha_2 + \alpha_3 = 1, \\ \gamma_{11} + \gamma_{12} + \gamma_{13} = 0, \\ \gamma_{12} + \gamma_{22} + \gamma_{23} = 0, \\ \gamma_{13} + \gamma_{23} + \gamma_{33} = 0, \\ \gamma_{12} + \gamma_{22} + \gamma_{32} = 0. \end{cases} \quad (4.7.10)$$

Мы будем пользоваться предположением об однородности при оценке уравнений для долей.

¹ Пусть H — гессен функции издержек C и V — диагональная матрица, j -й элемент которой равен $\partial C / \partial p_j$. Тогда матрица эластичностей замещения (η_{jk}) , определенная в (4.7.8), может быть переписана в более компактном виде $(\eta_{jk}) = C \cdot V^{-1} H V^{-1}$. Поскольку V — диагональная матрица, то $V^{-1} H V^{-1}$ отрицательно полуопределена, если и только если отрицательно полуопределена матрица H .

2. *Монотонности*: Принимая во внимание лемму Шенарда, свойство монотонности эквивалентно требованию, что правые части уравнений для долей (4.7.7) неотрицательны для любых комбинаций цен факторов и выпуска. Для любых значений коэффициентов всегда возможно выбрать цены факторов и объем выпуска таким образом, чтобы это требование нарушалось. Следовательно, мы не можем ввести требование монотонности для уравнений для долей затрат. Однако возможно проверить, выполнено ли требование монотонности для заданных значений (*relevant range*). Иными словами, выполнено ли это требование для всех комбинаций цен факторов и объемов выпуска, присутствующих в конкретном наборе данных. Мы сделаем это для оцененных параметров ниже.

3. *Вогнутости*: Вогнутость требует, чтобы матрица эластичностей замещения (4.7.9) была отрицательно полуопределена для любой комбинации долей затрат на факторы. Можно показать, что необходимым и достаточным условием вогнутости является отрицательная полуопределенность матрицы (γ_{jk}) размера¹ 3×3 .

В частности, диагональные элементы $(\gamma_{11}, \gamma_{22}, \gamma_{33})$ должны быть неположительными. Условие отрицательной полуопределенности матрицы (γ_{jk}) представляет собой набор неравенств для ее элементов. Оценивание при введении ограничений в виде неравенств возможно (см.: [Jorgenson, 1986, Section 2.4.5] для описания методики), но мы не будем рассматривать его здесь.

Стохастические спецификации

Транслогарифмическая функция издержек (4.7.1) включает в себя (аддитивную) ошибку, ε , в то время как уравнения долей затрат, выделенные из нее, не включают ошибку. При использовании нашей базы данных уравнения для долей не выполняются в точности для каждой фирмы, следовательно, мы должны включить ошибки. Обычным является включение случайного возмущения в каждое из уравнений для долей. Существуют два способа обоснования подобной стохастической спецификации. Один из них — это ошибки оптимизации: фирмы совершают

¹ Доказательство. Обозначим через $\mathbf{a}$ вектор размера 1×1 , j -й элемент которого равен a_j , а $\mathbf{S}$ — диагональная матрица, j -й элемент которой равен s_j . Тогда матрица эластичностей замещения, определенная в (4.7.9), может быть записана как $(\eta_{jk}) = \mathbf{S}^{-1}[(\gamma_{jk}) - \mathbf{a}\mathbf{a}' - \mathbf{S}]\mathbf{S}^{-1}$. Для некоторой комбинации долей затрат выражение $\mathbf{a}\mathbf{a}' - \mathbf{S}$ равно нулю (например, для $\mathbf{a} = (1, 0, 0)'$). То есть необходимым условием отрицательной полуопределенности матрицы (η_{jk}) является отрицательная полуопределенность матрицы (γ_{jk}) . Это условие также является достаточным, так как матрица $\mathbf{a}\mathbf{a}' - \mathbf{S}$ отрицательно полуопределена, и сумма двух отрицательно определенных матриц также является отрицательно полуопределенной.

случайные ошибки при выборе комбинаций факторов производства, минимизирующих издержки. Но тогда фактически облик функции затрат, соответствующий функции издержек, поскольку при выборе функции издержек не трогосматриваются ошибки оптимизации. Другим обоснованием наличия ошибок является допущение о том, что α_j в уравнениях долей является стохастическим и различным для различных фирм. В таком случае константой в уравнении доли фактора j в общих издержках будет математическое ожидание α_j , а отклонение α_j от своего математического ожидания. Но это также означает, что α_j в функции издержек следует интерпретировать как случайную величину.

Поэтому в обоих случаях внутренняя согласованность не позволяет бы чаш включать функцию издержек в систему уравнений для долей с аддитивными ошибками, чтобы оценивать все уравнения совместно. До конца параграфа мы будем рассматривать систему уравнений, состоящую только из уравнений для долей затрат факторов. Однако это означает, что мы сможем оценить лишь часть параметров функции издержек. Для параметра $(\alpha_Q$ и α_{Q2}), которые отсутствуют в функциях спроса на факторы, не могут быть оценены. Так как отдача от масштаба зависит от них (см. (4.7.3)), она также не может быть оценена.

Природа ограничений

После добавления аддитивных ошибок в систему уравнений для долей издержек на факторы производств система уравнений (4.7.7) запишется следующим образом:

$$\begin{cases} y_1 = \alpha_1 + \alpha_{11} \log(p_1) + \alpha_{12} \log(p_2) + \alpha_{13} \log(p_3) + \alpha_{1Q} \log(Q) + \varepsilon_1, \\ y_2 = \alpha_2 + \alpha_{21} \log(p_1) + \alpha_{22} \log(p_2) + \alpha_{23} \log(p_3) + \alpha_{2Q} \log(Q) + \varepsilon_2, \\ y_3 = \alpha_3 + \alpha_{31} \log(p_1) + \alpha_{32} \log(p_2) + \alpha_{33} \log(p_3) + \alpha_{3Q} \log(Q) + \varepsilon_3. \end{cases} \quad (4.7.11)$$

В этой системе пятьдесят коэффициентов, или параметров. Перекрестные ограничения (выраженные в том, что матрица логарифмов коэффициентов при ценах размера 3×3 , α_{ij} , симметрична), и ограничения однородности (4.7.10) вместе дают восемь ограничений. Таким образом, пятнадцать параметров могут быть описаны восемью независимыми параметрами. Чтобы понять это, полезно разделить ограничения на следующие три группы.

$$\text{ограничения в виде сумм:} \quad \begin{cases} \alpha_1 + \alpha_2 + \alpha_3 = 1, \\ \alpha_{11} + \alpha_{21} + \alpha_{31} = 0, \\ \alpha_{12} + \alpha_{22} + \alpha_{32} = 0, \\ \alpha_{13} + \alpha_{23} + \alpha_{33} = 0, \\ \alpha_{1Q} + \alpha_{2Q} + \alpha_{3Q} = 0. \end{cases} \quad (4.7.12)$$

$$\text{однородности} \quad \begin{cases} \gamma_{11} + \gamma_{12} - \gamma_{13} = 0, \\ \gamma_{21} + \gamma_{22} + \gamma_{23} = 0, \\ \gamma_{31} + \gamma_{32} + \gamma_{33} = 0, \end{cases} \quad (4.7.13)$$

$$\text{симметрии} \quad \begin{cases} \gamma_{12} = \gamma_{21}, \\ \gamma_{13} = \gamma_{31}, \\ \gamma_{23} = \gamma_{32}. \end{cases} \quad (4.7.14)$$

Здесь одинаковые уравнения, но только восемь из них являются ограничительными: учитывая ограничения в виде сумм, одно из трех ограничений однородности следует из двух других; при условии ограничений в виде сумм и однородности два из трех ограничений симметрии следует из третьего (проверьте это).

Модель многомерной регрессии с перекрестными ограничениями

Задача ограничений однородности (4.7.13) осуществляется непосредственно, поскольку она не является перекрестными. Мы можем устранить три параметра, скажем $(\gamma_{13}, \gamma_{23}, \gamma_{33})$, и получить следующую систему:

$$\begin{cases} \varepsilon_1 = \alpha_1 + \gamma_{11} \log(p_1/p_0) + \gamma_{12} \log(p_2/p_0) + \gamma_{14} \log(Q) + \varepsilon_1, \\ \varepsilon_2 = \alpha_2 + \gamma_{21} \log(p_1/p_0) + \gamma_{22} \log(p_2/p_0) + \gamma_{24} \log(Q) + \varepsilon_2, \\ \varepsilon_3 = \alpha_3 + \gamma_{31} \log(p_1/p_0) + \gamma_{32} \log(p_2/p_0) + \gamma_{34} \log(Q) + \varepsilon_3. \end{cases} \quad (4.7.15)$$

Особенность этой системы (вне зависимости от введения требования однородности) в том, что сумма зависимых переменных $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ даст ошибку для всех наблюдений. Учитывая это, а также ограничения в виде сумм, получаем, что сумма ошибок $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ равна нулю для любого наблюдения. То есть ковариационная матрица ошибок $\Sigma \equiv \text{Var}(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ размера 3×3 выведена. Вложим, что в модели многомерной регрессии (которая является частным случаем SUR-модели, описанной в учреждении 4.6) матрица S , определенная в (4.1.11), может быть записана как Кросс-корреляция: $S = \Sigma \otimes I'(\log p_i')$. Таким образом, S становится проблемой, что нарушает предположение 4.5. С этой проблемой и проблемой нулевой корреляции путем исключения одного уравнения из системы и решения оставшихся уравнений (наше этот вопрос рассмотрен более подробно). Коэффициенты полученного уравнения могут быть вычислены с использованием основных коэффициентов и ограничений в виде сумм. Таким образом, учитываются как ограничения в виде сумм, так и ограничения однородности.

Как уже упоминалось выше, при учете принципов симметрии и однородности остается только одно ограничение симметрии. Напротив, если мы исключили из системы коэффициенты $(\gamma_{13}, \gamma_{23}, \gamma_{33})$ при исключении ограничений однородности (так это было сделано в (4.7.15)) и если мы опустили третье уравнение при введении ограничений в виде сумм, то тогда единственное оставшееся ограничение симметрии — это $\gamma_{12} = \gamma_{21}$.

При введении этого перекрестного ограничения система принимает вид:

$$\begin{cases} \varepsilon_1 = \alpha_1 + \gamma_{11} \log(p_1/p_3) + \gamma_{12} \log(p_2/p_3) + \gamma_{1Q} \log(Q) + \varepsilon_1, \\ \varepsilon_2 = \alpha_2 + \gamma_{21} \log(p_1/p_3) + \gamma_{22} \log(p_2/p_3) + \gamma_{2Q} \log(Q) + \varepsilon_2 \end{cases} \quad (4.7.16)$$

Поскольку регрессоры являются предопределенными и поскольку уравнения содержат общий набор регрессоров, система может быть оценена с помощью метода мно.мерной регрессии с перекрестным ограничением симметрии. Это оценивание с ограничением может быть преобразовано в безусловное оценивание после приведения к виду «с общими коэффициентами», изложенному в параграфе 4.6. То есть система (4.7.16) может быть переписана в форме «с общими коэффициентами» $y_1 = Z_1\beta + \varepsilon_1$ (см. (4.6.1') на с. 320), если мы определим матрицы следующим образом (мы все еще опускаем индекс «i» для наблюдений в $\varepsilon_1, \varepsilon_2, \log(p_1/p_3), \log(p_2/p_3), \log(Q)$).

$$y_1 = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}, \quad Z_1 = \begin{bmatrix} 1 & 0 & \log(p_1/p_3) & \log(p_2/p_3) & 0 & \log(Q) & 0 \\ 0 & 1 & 0 & \log(p_1/p_3) & \log(p_2/p_3) & 0 & \log(Q) \end{bmatrix},$$

$$\beta' = [\alpha_1 \quad \alpha_2 \quad \gamma_{11} \quad \gamma_{12} \quad \gamma_{21} \quad \gamma_{22} \quad \gamma_{1Q} \quad \gamma_{2Q}]. \quad (4.7.17)$$

Таким образом, оценивание многомерной регрессии при условии симметрии равносильно оцениванию методом случайных эффектов без ограничений.

Это оценивание методом случайных эффектов указанной системы из двух уравнений дает состоятельные оценки следующих семи параметров:

$$\alpha_1, \alpha_2, \gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \gamma_{1Q}, \gamma_{2Q}.$$

Остальные из пятнадцати параметров уравнений для долей затрат,

$$\alpha_3, \gamma_{31}, \gamma_{32}, \gamma_{33}, \gamma_{34}, \gamma_{35}, \gamma_{36}, \gamma_{37}, \gamma_{38}, \gamma_{39},$$

могут быть вычислены с использованием соотношений в виде строк (4.7.12), аналогичных (4.7.13) и симметрии (4.7.14). Например, γ_{34} можно вычислить так:

$$\gamma_{34} = -\gamma_{41} - \gamma_{42} = -\gamma_{14} - \gamma_{24} = -(\gamma_{11} + \gamma_{12}) + (\gamma_{12} + \gamma_{22}) = -\gamma_{11} + 2\gamma_{12} - \gamma_{22}.$$

То есть численная оценка γ_{34} равна

$$\hat{\gamma}_{34} = \hat{\gamma}_{11} + 2\hat{\gamma}_{12} - \hat{\gamma}_{22},$$

где $(\hat{\gamma}_{11}, \hat{\gamma}_{12}, \hat{\gamma}_{22})$ — оценки параметров, полученные по оценочным методом случайных эффектов. Стандартная ошибка $\hat{\gamma}_{34}$ может быть вычислена применением «метода-матрицы» (см. главу 2.5) к составленной

оценке $\Delta \text{var}(\hat{\gamma}_{11}, \hat{\gamma}_{12}, \hat{\gamma}_{22})$, полученной при оценивании методом случайных эффектов.

Какое уравнение исключать

Имеет ли значение, какие из уравнений системы исключать? Очевидно, что ответ был бы отрицательным, если бы не было перекрестных ограничений, так как многомерная регрессия численно эквивалентна OLS-оцениванию каждого уравнения в отдельности. Посмотрим, выполняется ли численная эквивалентность при наличии перекрестного ограничения. Пусть Σ^* — ковариационная матрица ошибок размера 2×2 для системы из двух уравнений, полученной при исключении одного уравнения из системы (4.7.13). Это соответствующая подматрица ковариационной матрицы ошибок Σ размера 3×3 . (Например, если исключить третье уравнение, то тогда Σ^* — ведущая подматрица размера 2×2 матрицы Σ размера 3×3 .) Чтобы реализовать оценивание методом случайных эффектов, нам необходимо найти состоятельную оценку ($\hat{\Sigma}^*$) матрицы Σ . Существуют два способа вычисления $\hat{\Sigma}^*$:

- (OLS-оценивание каждого уравнения.) Оцените два уравнения по отдельности, таким образом игнорируя перекрестное ограничение, а затем используйте полученные остатки для вычисления $\hat{\Sigma}^*$. Равнозначно можно произвести раздельное OLS-оценивание трех уравнений и использовать полученные при этом остатки для вычисления $\hat{\Sigma}$ (оценки ковариационной матрицы Σ размера 3×3), а затем извлечь необходимую подматрицу матрицы $\hat{\Sigma}$. Если матрица $\hat{\Sigma}^*$ вычисляется подобным образом, то численная эквивалентность гарантирована (вы проверите это в эмпирическом упражнении), и не имеет значения, какое из уравнений системы исключать.
- Получите $\hat{\Sigma}^*$ с помощью некоторого метода, учитывающего перекрестные ограничения, связывающего разные уравнения (например, объединенного OLS, примененного к системе, приведенной к форме с общими коэффициентами). Численная эквивалентность в этом случае не гарантируется.

Численная эквивалентность выполнена в случае применения OLS для каждого уравнения по отдельности, поскольку матрица $\hat{\Sigma}$, из которой извлекается $\hat{\Sigma}^*$, является вырожденной в конечных выборках. В больших выборках оценка $\hat{\Sigma}^*$ становится consistente, и, даже если $\hat{\Sigma}^*$ получается не из OLS-оценивания уравнений по отдельности, это соотношение между $\hat{\Sigma}^*$ и $\hat{\Sigma}$ выполняется асимптотически, так как матрица Σ вырожденная. То есть инвариантность выполняется асимптотически, если $\hat{\Sigma}^*$ состоятельна. В частности, асимптотическая дисперсия оценок параметров не зависит от выбора исключаемого уравнения.

Результаты

Рассмотренная в главе 1 работа Перлова использует кросс-секционные данные за 1995 год для компаний США, занимающихся производством и распределением электроэнергии [Christensen and Greene, 1976] анализируют похожие данные за 1970 год для 99 фирм, используя при этом источники данных, которые позволили им получить более точные меры для уровня заработной платы и цен на топливо (см параграф V их работы для более подробного описания). Их набор данных включает информацию о долях затрат на факторы, так что уравнения для ценой затрат можно оценить. Средние значения и стандартные отклонения для некоторых переменных из их набора данных представлены в табл. 4.3, из которой видно, что в общих затратах на производство электроэнергии значительную долю составляют затраты на топливо.

Таблица 4.3

Простые статистики (объем выборки = 99)

	Производство электроэнергии, мгдт кВт/ч	Доля затрат труда	Доля затрат капитала	Доля затрат на топливо
Среднее	0.0	0.141	0.227	0.633
Стандартное отклонение	10.3	0.059	0.062	0.192

Для расчета $\hat{\Sigma}$ мы используем остатки, полученные после OLS-оценки каждого уравнения по отдельности, так что оценки, получаемые методом случайных эффектов, численно инвариантны к выбору удаляемого уравнения. Эти оценки приведены в табл. 4.4 наряду с $\hat{\Sigma}$, полученный в результате применения OLS ко всем уравнениям по отдельности. Индекс фактора производства i равен 1 для труда, 2 — для капитала и 3 — для топлива. Некоторые (на самом деле все) диагональные элементы матрицы оценок коэффициентов при ценах ($\hat{\sigma}_{ij}$) положительны. То есть матрица не является отрицательно полуопределенной, что противоречит выпуклости.

Даже если ($\hat{\sigma}_{ij}$) не является отрицательно полуопределенной, соответствующие эластичности замещения, рассчитываемые по формуле (4.7.9), могут быть отрицательно полуопределенными. Поскольку эта формула выводится в предположении отсутствия ошибок оптимизации, для α_i и α_j в этой формуле должны использоваться оценки долей затрат, а не

¹Таким образом, выпуклость может быть формализована. Используя компьютер мы можем найти асимптотическое распределение характеристических корней ($\hat{\sigma}_{ij}$). Нулевая гипотеза отсутствия состоит в том, что все эти характеристические корни положительны.

фактические доли затрат. Эластичности, усредненные по фирмам, приведены в табл. 4.5. (Они являются внедиагональными элементами симметричной матрицы эластичностей замещения размера 3×3 , которая была найдена таким образом.) Указанные три фактора производства являются субститутами, поскольку эластичности замещения положительны. Однако степень их взаимозаменяемости намного меньше по сравнению с лесной техникой. Требование кооперности нарушается даже для имеющихся данных: для 20 фирм из 99, представленных в выборке, матрица эластичностей замещения не является отрицательно полуопределенной. Монотонность, напротив, выполняется для рассматриваемого набора данных: оцененные доли затрат неотрицательны для всех факторов производства и для всех фирм в выборке.

Таблица 4.4

Оценки методом случайных эффектов

Параметр	Точечная оценка	Стандартная ошибка	t-значение
α_1	-0.132	0.106	-1.25
α_2	0.318	0.055	3.75
α_3	0.813	0.064	8.69
γ_1	0.084	0.020	4.19
γ_{12}	-0.023	0.016	-1.46
γ_{13}	0.060	0.015	3.92
γ_{23}	0.122	0.020	6.19
γ_{24}	-0.009	0.017	-5.75
γ_{34}	0.159	0.023	6.90
γ_{10}	-0.0211	0.0026	-8.55
γ_{20}	-0.0056	0.0030	-2.87
γ_{30}	0.0297	0.0037	7.98

$\hat{\Sigma}$ полученная посредством последовательного

OLS-оценивания каждого уравнения —

$$= \begin{bmatrix} 0.00173 & -0.000171 & -0.00156 \\ -0.000171 & 0.00253 & -0.00236 \\ 0.00156 & -0.00236 & 0.00391 \end{bmatrix}.$$

Таблица 4.5

Эластичности замещения

Труд—капитал	Капитал—топливо	Труд—топливо
0.17	0.29	0.27

Контрольные вопросы

1. Выведите (4.7.6). Проверьте, что симметрия (γ_{jk}) используется при выводе.
2. Выведите ограничения однородности (4.7.10). Указание: Для того чтобы транслогарифмическая функция издержек (4.7.1) была однородной третьей степени по ценам факторов, необходимо и достаточно, чтобы для любого скаляра $\lambda \neq 0$

$$\begin{aligned} \sum_{j=1}^3 \alpha_j \log(\lambda \cdot p_j) + \frac{1}{2} \sum_{j=1}^3 \sum_{k=1}^3 \gamma_{jk} \log(\lambda \cdot p_j) \log(\lambda \cdot p_k) + \\ + \alpha_Q \log(Q) + \frac{1}{2} \gamma_{QQ} (\log(Q))^2 + \sum_{j=1}^3 \gamma_{jQ} \log(\lambda \cdot p_j) \log(Q) = \\ = \log(\lambda) + \sum_{j=1}^3 \alpha_j \log(p_j) + \frac{1}{2} \sum_{j=1}^3 \sum_{k=1}^3 \gamma_{jk} \log(p_j) \log(p_k) + \\ + \alpha_Q \log(Q) + \frac{1}{2} \gamma_{QQ} (\log(Q))^2 + \sum_{j=1}^3 \gamma_{jQ} \log(p_j) \log(Q). \end{aligned}$$

3. (Выпол ограничений) Используя ограничения в виде сумм (4.7.12), два из трех уравнений (4.7.13) и одно из трех уравнений (4.7.14), выведите для других уравнений (4.7.14).
4. (Для OLS выразимость ограничений.) Рассмотрим OLS-оценивание каждого из трех уравнений (4.7.15) по отдельности. Покажите, что полученные оценки коэффициентов удовлетворяют ограничениям в виде сумм. Покажите, что ковариационная матрица ошибок размера 3×3 , вычисленная с помощью OLS-оценивания уравнений по отдельности, является вырожденной. Указание: Пусть y_1, y_2, y_3 — векторы левых частей трех связанных уравнений (где n — размер выборки), пусть X — матрица данных размера $n \times K$, содержащая K общих регрессоров ($K = 4$ в (4.7.15)). Покажите, что $y_1 + y_2 + y_3 = 1$. Система коэффициентов, полученная при OLS-оценивании каждого уравнения по отдельности, равна $(X'X)^{-1}X'y_j$, $j = 1, 2, 3$. Три вектора остатков размера $n \times 1$ равны My_j , $j = 1, 2, 3$, где $M = I_n - X(X'X)^{-1}X'$.

Набор задач для главы 4

Аналитические упражнения

1. (Представление 3SLS в матричном виде.) В модели 3SLS набор инструментальных переменных одинаков для всех уравнений. Пусть K — количество общих инструментов. Введем следующие опреде-

предполагается:

$$\begin{aligned}
 X_{n \times K} &= \begin{bmatrix} x_{11} \\ \vdots \\ x_{n1} \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1K} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nK} \end{bmatrix}, \\
 Z_{m \times L} &= \begin{bmatrix} z_{1m} \\ \vdots \\ z_{mm} \end{bmatrix}, \quad (M = \sum_{m=1}^M L_m) = \begin{bmatrix} Z_1 & & \\ & \ddots & \\ & & Z_M \end{bmatrix}, \\
 \delta_{L \times 1} &= \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_M \end{bmatrix}, \\
 y_{(M+1) \times 1} &= \begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix}, \quad y_m = \begin{bmatrix} y_{1m} \\ \vdots \\ y_{mm} \end{bmatrix}, \\
 \varepsilon_{(M+1) \times 1} &= \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_M \end{bmatrix}, \quad \varepsilon_m = \begin{bmatrix} \varepsilon_{1m} \\ \vdots \\ \varepsilon_{mm} \end{bmatrix}.
 \end{aligned}$$

Покажите следующие:

- Предположение 4.1 можно записать как $y = Z\delta + \varepsilon$.
- $\hat{S}$ в (4.5.10) принимает вид $\hat{S} = \hat{\Sigma} \otimes \frac{1}{n} X'X$.
- $\hat{\delta}_{OLS} = [Z'(\hat{\Sigma}^{-1} \otimes P)Z]^{-1} Z'(\hat{\Sigma}^{-1} \otimes P)y$, где $P = X(X'X)^{-1}X'$.
- $\text{Asymptotic}(\hat{\Sigma}) = \Sigma \cdot [Z'(\Sigma^{-1} \otimes P)Z]^{-1}$.
- $\hat{A}_{OLS}$ в (4.5.13) принимает вид $\hat{A}_{OLS} = \frac{1}{n} Z_m' P Z_m$.
- $\hat{\varepsilon}_{OLS}$ в (4.5.14) принимает вид $\hat{\varepsilon}_{OLS} = \frac{1}{n} Z_m' P y_m$.
- $J(\hat{\delta}_{OLS}, \hat{S}^{-1}) = (y - Z\hat{\delta}_{OLS})'(\hat{\Sigma}^{-1} \otimes P)y - Z\hat{\delta}_{OLS}$.

Указание:

$$\frac{1}{n} \sum_{i=1}^n x_i x_i' = \frac{1}{n} X'X, \quad \text{а также } (I_M \otimes X)' y; S_{yz} = \frac{1}{n} (I_M \otimes X)' Z.$$

Используйте формулы (A.11) и (A.13) (A.15) приложения A.

2. (Представление в матричном виде SUR, RE и объединенного OLS.)

- (а) Докажите, что (4.5.13) сводится к (4.5.13') (я с. 309), а (4.5.14) сводится к (4.5.14') при предположении SUR (4.5.15). Указание: Если $P = X(X'X)^{-1}X'$, то $\hat{A}_{OLS} = \frac{1}{n} Z_m' P Z_m$, $\hat{\varepsilon}_{OLS} = \frac{1}{n} Z_m' P y_m$, $P Z_m = Z_m \cdot Z_m'$, если столбцы Z_m обозначены из столбцов X .

(b) Покажите, что для SUR

$$\begin{aligned}\text{оценка} &= [Z'(\tilde{\Sigma}^{-1} \otimes I_n)Z]^{-1}Z'(\tilde{\Sigma}^{-1} \otimes I_n)y, \\ \text{где } \tilde{A}^{-1}\tilde{y} &= n^{-1}[Z'(\tilde{\Sigma}^{-1} \otimes I_n)Z]^{-1}.\end{aligned}$$

(c) Пусть $\{y_{it}\}$ для этой части выписи) $y_{it} = \alpha_i$ для всех t . Тогда $Z = I_M \otimes X$. Покажите, что SUR сводится к оценке многомерной регрессии

$$[I_M \otimes (X'X)^{-1}X']y.$$

Проверьте, что m -й K -мерный блок этого MK -мерного вектора y — это $(X'X)^{-1}X'y_m$.

Теперь определим Z так:

$$Z_{(Mn \times L)} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_M \end{bmatrix}. \quad (7)$$

Тогда предположение 4.1' (линейные уравнения с общими коэффициентами) может быть переписано как $y = Z\delta + \varepsilon$, где $g = n = nk$, что δ — n -мерный вектор параметров δ , а δ — L -мерный вектор общих коэффициентов.

(d) Покажите, что для оценки методом случайных эффектов верны формулы, введенные ранее в гл. 3 (здесь для SUR, и Z определена как в (*)). Указание: Эта оценка приведена в (4.6.8), а не $\tilde{A}^{-1}\tilde{y}$ — в (4.6.10). Используйте формулу (A.8) приложения A.

(e) Покажите, что

$$\hat{\delta}_{\text{random effects в (4.6.11)}} = (Z'Z)^{-1}Z'y.$$

$$\text{Аналог } \hat{\delta}_{\text{random effects в (4.6.14)}} = n \cdot (Z'Z)^{-1}Z'(\tilde{\Sigma} \otimes I_n)Z[(Z'Z)^{-1}Z'y].$$

(f) В матрице Z размера $Mn \times L$, определенной в (*), строки упорядочены сначала по номерам уравнений (m), а затем по наблюдениям (i). Теперь упорядочим строки сначала по наблюдениям (i), что приводит к переобращению Z :

$$Z_{(Mn \times L)} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_M \end{bmatrix}.$$

где $Z_i (i = 1, 2, \dots, n)$ — матрица размера $M \times L$, определяемая в (4.6.15). Перегруппируйте ряды y и ε аналогичным образом:

$$y_{(M \times 1)} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \varepsilon_{(M \times 1)} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Как изменяются формулы для RE и объединенного OLS, которые мы вывели в пунктах (d) и (e)?

3. (OLS интерпретации SUR и RE) SUR-модель состоит из пред. выше экв. 4.1–4.6, 4.7 и (4.5.15) (что x_i — объединение $(z_{1i}, \dots, z_{Li})$). Уточним модель, учитывая предположения 4.3 (условия правильности) следующим образом:

$$E(z_{mi}, z_{1i}, \dots, z_{Li}) = 0 \quad (m = 1, \dots, M),$$

и предположение 4.2 (эргодическая стационарность) как i.i.d.

$$(y_1, \dots, y_M; z_{11}, \dots, z_{L1};$$

- (a) Проверьте, что для этих предположений матрица Σ блочна-строго диагональна.
 (b) Пусть $\varepsilon(Mn \times 1)$ и $Z(Mn \times \sum_{m=1}^M L_m)$ определяются как же, как и в статистическом утверждении 1 выше. Покажите, что в уточненной модели выполняется

$$E(\varepsilon|Z) = 0 \text{ и } E(\varepsilon\varepsilon' | Z) = \Sigma \otimes I_n.$$

Указание: Сначала необходимо показать, что $E(\varepsilon_{im}|Z) = 0$. Если (m, i) матрица $E(\varepsilon\varepsilon' | Z)$ — это $E(\varepsilon_{im}\varepsilon'_{ik} | Z)$. Необходимо показать, что $\forall i$ матрица разл. $\sigma_{imk} I_n$.

- (c) (Свойства SUR в конечных выборках.) Пусть матрица Σ известна, так что систематические оценки $\hat{\sigma}_{imk}$ в формулах SUR и утверждении 4.6 можно отождествить с истинными значениями σ_{imk} .
 (i) Проверьте, что эта модель удовлетворяет предположениям OLS параграфа 1.6 (то есть предположениям 1.1–1.3 и (1.6.11) в объеме выборки Mn).
 (ii) Проверьте, что SUR-оценки являются OLS-оценкой. Укажете Y и матрицу X в параграфе 1.6 здесь выступают Z . Покажите, что $\sigma^2 V$ здесь принимает вид $\Sigma \otimes I_n$. Исходя из этого как ответ в пункте 2(ii).

- (iii) Проверьте, что SUR-оценка является несмещенной в конечных выборках, то есть $E(\hat{\delta}_{\text{SUR}} - \delta|Z) = 0$.
- (iv) Выражение для $\text{Var}(\hat{\delta}_{\text{SUR}}|Z)$ приведено в утверждении 1.7 параграфа 1.6. Покажите, что

$$\text{Avar}(\hat{\delta}_{\text{SUR}}) = \lim_{n \rightarrow \infty} n \cdot \text{Var}(\hat{\delta}_{\text{SUR}}|Z).$$

Указание: С учетом того что $E(\epsilon\epsilon'|Z) = \Sigma \otimes I_n$, матрица $\text{Var}(\hat{\delta}_{\text{SUR}}|Z)$ является обратной к матрице, (m, h) -блок которой равен $\sigma^{mh} Z'_{in} Z_{ih}$. $Z'_{in} Z_{ih} = \sum_1 z_{in} z'_{ih}$.

- (d) (Свойства RE в конечных выборках.) Положим теперь не модель условие, состоящее в том, что вектор коэффициентов одинаков для всех уравнений, и Z определяется как в (*) аналитического упражнения 2. Как и в (c), предположим, что Σ известна, так что состоятельная оценка $\hat{\sigma}_{mh}$ в формулах RE утверждения 4.7 может быть заменена истинным значением σ_{mh} .

- (i) Проверьте, что такая модель с общими коэффициентами удовлетворяет предположениям GLS параграфа 1.6. В частности, предположение 1.1 может быть переписано как $y = -Z\delta - \epsilon$, где y и ϵ — те же, что и в аналитическом упражнении 2.
- (ii) Проверьте, что оценка методом случайных эффектов является GLS-оценкой.
- (iii) Проверьте, что оценка методом случайных эффектов является несмещенной в конечных выборках.
- (iv) Покажите, что

$$\text{Avar}(\hat{\delta}_{\text{RE}}) = \lim_{n \rightarrow \infty} n \cdot \text{Var}(\hat{\delta}_{\text{RE}}|Z).$$

Указание: $\text{Var}(\hat{\delta}_{\text{RE}}|Z)$ — обратная к матрице $\sum_m \sum_h \sigma^{mh} Z'_{in} Z_{ih}$.

- 4 (Стандартные ошибки, полученные при использовании GMM отдельно для каждого уравнения и при использовании GMM для нескольких уравнений.) В параграфе 4.4 мы показали, что Avar эффективной оценки, полученной применением GMM к каждому уравнению по отдельности, не меньше (в матричном смысле), чем Avar эффективной оценки, полученной применением GMM для нескольких уравнений. В этом упражнении мы покажем, что то же самое верно и для $\widetilde{\text{Avar}}$. Пусть $\tilde{S}$ и $S_{\pi\pi}$ те же, что и в (4.3.2) и (4.2.2) соответственно.

- (а) (Тривиально.) Проверьте, что при соответствующих предположениях $(S'_{xx}\bar{S}^{-1}S_{xx})^{-1}$ является состоятельной оценкой асимптотической дисперсии эффективной GMM-оценки для нескольких уравнений. Что эти же предположения?
- (б) Пусть $\tilde{\delta}_m$ — эффективная GMM-оценка δ_m для одного уравнения, а $\tilde{\delta}$ — вектор, сформированный как $(\tilde{\delta}_1, \dots, \tilde{\delta}_M)$. $\tilde{\delta}$ — эффективная оценка для δ , получаемая GMM для каждого уравнения в отдельности. Покажите, что $\text{Avar}(\tilde{\delta}_m)$ состоительно оценивается посредством

$$\left(\frac{1}{n} \sum_{i=1}^n z_{im} z'_{im} \tilde{W}_{im} \frac{1}{n} \sum_{i=1}^n z_{im} z'_{im} \right)^{-1}, \quad (*)$$

где

$$\tilde{W}_{im} = \left(\frac{1}{n} \sum_{i=1}^n \tilde{z}_{im}^2 z_{im} z'_{im} \right)^{-1}.$$

- (с) Проверьте, что $(*)$ — это блок (m, m) матрицы

$$(S'_{xx} \tilde{W} S_{xx})^{-1} S'_{xx} \tilde{W} \tilde{S} \tilde{W} S_{xx} (S'_{xx} \tilde{W} S_{xx})^{-1}, \quad (**)$$

где $\tilde{W}$ — блочно-диагональная матрица, m -й диагональный блок которой — это $\tilde{W}_{mm}$.

- (д) Утверждение 3.5 — это алгебраический результат, который применим не только к теоретическим моментам, так что из него следует, что

$$(**) \geq (S'_{xx} \bar{S}^{-1} S_{xx})^{-1}.$$

Считая это известным, покажите следующее: если одни и те же остатки используются при вычислении $\tilde{W}$ для GMM, применяемого к уравнениям по отдельности, и при вычислении $\tilde{S}$ для GMM для нескольких уравнений, то тогда для каждого коэффициента стандартная ошибка, полученная при применении эффективного GMM для нескольких уравнений, не превышает стандартной ошибки, полученной при применении GMM к уравнениям по отдельности. **Указание:** Стандартные ошибки — это квадратные корни из диагональных элементов оцененной асимптотической ковариационной матрицы, деленные на объем выборки.

- (е) Сохраняются ли результаты, полученные в (д), если модель неправильно специфицирована в том или ином отношении? **Указание:** То, что мы доказали, является алгебраическим результатом.

- (f) Выполните шаги (a)–(c) для FIVE и для 2SLS, применяемого для каждого уравнения отдельно. **Указание:** Для этого требуется только переопределить $\tilde{S}$ и $\tilde{W}$ для данной модели. Для $\tilde{W}_{FIVE}$ матрица $\tilde{W}$, $\tilde{W}_{2SLS}$ для 2SLS — это матрица, обратная к

$$\frac{1}{n} \sum_{i=1}^n \epsilon_{im}^2 + \frac{1}{n} \sum_{i=1}^n z_{im} \sigma_{im}^2$$

а $\tilde{S}$ для FIVE — это (4.3.31).

5. (Идентификация и определение, связанные с двумя уравнениями) Рассмотрим уравнения латентной переменной 1969 и 1980 годов

$$LW90_i = \beta_0 + \beta_1 \cdot S89_i + \beta_2 \cdot IQ_i + \epsilon_{1i},$$

$$LW89_i = \lambda_0 + \lambda_1 \cdot S89_i + \lambda_2 \cdot IQ_i + \epsilon_{2i}.$$

Условия ортогональности лагранжевых множителей выглядят:

$$E(\epsilon_{1i}) = 0, \quad E(\epsilon_{2i}) = 0, \quad E(MED_{1989} \epsilon_{1i}) = 0, \quad E(MED_{1989} \epsilon_{2i}) = 0.$$

Пусть i — индекс наблюдений является случайная выборка для переменных $\{LW90, LW89, S89, IQ, MED\}$.

- (a) Является ли уравнение для LW90 идентифицируемым, если рассмотреть его в изоляции от других? **Указание:** Проверьте для уравнения LW90 порядковое условие идентифицируемости.
- (b) Запишите указанные четыре условия ортогональности в терминах $\{b_1, b_2, b_3\}$ как систему уравнений, линейных по β . Сколько уравнений получилось? Запишите ранговое условие идентифицируемости в терминах соответствующих перекрестных моментов. (Этот пример показывает, что даже если каждое уравнение системы неидентифицируемо, если рассмотреть его в изоляции, система может быть идентифицируемой (для каждого набора перекрестных ограничений, требующих, чтобы β были скалярными коэффициентами уравнений).)
- (c) Покажите, что система не может быть идентифицируемой, если LQ и MED не коррелированы между собой.
6. (Полная идентификация) Докажите утверждение 4.1. **Указание:** Обобщите доказательства утверждения 3.2.
7. (Число параметров и эффективность 2SLS и 3SLS.) Пусть набор инструментов z системе уравнений $y_{im} = Z'_{im} \delta_m + \epsilon_{im}$ ($i = 1, \dots, n$;

$m = 1, \dots, M$) одинаковы для всех уравнений (то есть $\sigma_{\varepsilon_m} = \sigma_1$). (Значит и для ε предположим, что все условия гомосkedастичности 3SLS выполняются.)

(а) Если в правых частях всех уравнений стоят одни и те же переменные (то есть $z_{im} = z_i$), то оценки 2SLS и 3SLS численно совпадают. Докажите это. **Указание:** Пусть $B \equiv \frac{1}{M} \sum_{i=1}^M z_i z_i'$, 3SLS-оценка — это (4.2.5), где $S_{\varepsilon\varepsilon} = I_M \otimes B$, $\tilde{W} = \tilde{S}^{-1} \otimes S_{\varepsilon\varepsilon}^{-1}$, $S_{\varepsilon\varepsilon} = \frac{1}{M} \sum_{i=1}^M z_i z_i'$.

(б) Если оценки не являются составкой гомосkedастичности, то совпадают ли верным результатом пункта (а) для GMM-оценки для каждого уравнения по отдельности и для GMM-оценки для всех уравнений вместе? **Указание:** Нет. Почему? Можете ли вы записать $\tilde{S}^{-1}$ в виде Кронекерова произведения без предположений об условной гомосkedастичности?

8. (SUR — это больше, чем предположение о предопределенности регрессоров в каждой из уравнений.) Рассмотрим оценивание системы из M уравнений: $y_{im} = z_{im}'\beta_m + \varepsilon_{im}$ ($i = 1, \dots, n$; $m = 1, \dots, M$). Условия ортогональности: $E(z_{im} \cdot \varepsilon_{im}) = 0$ ($m = 1, 2, \dots, M$).

(а) Совпадают ли эффективная GMM-оценка $\hat{\beta}(\tilde{S}^{-1})$ для нескольких уравнений в OLS-оценивании каждого уравнения отдельно? Изменится ли ваш вывод, если ошибки условно гомосkedастичны? **Указание:** Каков размер $S_{\varepsilon\varepsilon}$? $\tilde{S}$?

(б) Объясните, почему эффективная GMM-оценка, которую вы вывели в пункте (а) при предположении об условной гомосkedастичности отличается от SUR? **Указание:** Разница в спецификации условий ортогональности.

9. (Доказательное опасение совместного оценивания.) Взято из [Greene, 1997, гл. 7.06]. Рассмотрим SUR-модель из двух уравнений:

$$\begin{aligned} y_1 &= \beta_1 x_1 + \varepsilon_1, \\ y_2 &= \beta_2 x_1 + \beta_3 x_2 + \varepsilon_2. \end{aligned}$$

Для простоты предположим, что $E(\varepsilon_1^2)$, $E(\varepsilon_2|\varepsilon_1)$ и $E(\varepsilon_2^2)$ известны. Предположим теперь, что вы применяете SUR, но ошибочно не включаете x_2 во второе уравнение. Какое влияние это окажет на состоятельность оценки β_3 ? **Указание:** Пусть $\varepsilon_1, \varepsilon_2, x_1, y_1, y_2, \varepsilon$,

и ε_2 — n -мерные векторы соответствующих переменных и

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}, \quad \bar{X} = \begin{bmatrix} x_1 & 0 \\ 0 & x_2 \end{bmatrix},$$

$$d = \begin{bmatrix} 0 \\ x_1 \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} E(x_1^2) & E(x_1 \varepsilon_{12}) \\ E(\varepsilon_{12} x_1) & E(\varepsilon_{12}^2) \end{bmatrix}.$$

(Так что, например, $\bar{X}$ — вектор размера $2n \times 2$, и d — вектор размера 2×1 .) Покажите, что

(а) Указанная система двух уравнений может быть записана как $y = \bar{X}\beta + \beta_1 \cdot d + \varepsilon$.

(б) SUR-оценка $\hat{\beta}$ в случае, когда первичная структура явно проигнорирована, равна:

$$\hat{\beta}_{SUR} = [\bar{X}'(\Sigma^{-1} \otimes I_n)\bar{X}]^{-1} \bar{X}'(\Sigma^{-1} \otimes I_n)y.$$

(в) Используя формулы (A.6) и (A.7) приложения А, покажите, что

$$\hat{\beta}_{SUR} = \hat{\beta} + \hat{\alpha} \cdot a + b,$$

где

$$a = \begin{bmatrix} \sigma_{11}^{-1} \frac{1}{n} \sum_{i=1}^n x_{11}^2 & \sigma_{12}^{-1} \frac{1}{n} \sum_{i=1}^n x_{11} x_{12} \\ \sigma_{12}^{-1} \frac{1}{n} \sum_{i=1}^n x_{12} x_{11} & \sigma_{22}^{-1} \frac{1}{n} \sum_{i=1}^n x_{12}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{12}^{-1} \frac{1}{n} \sum_{i=1}^n x_{11} x_{12} \\ \sigma_{22}^{-1} \frac{1}{n} \sum_{i=1}^n x_{12} x_{12} \end{bmatrix},$$

$$b = \begin{bmatrix} \sigma_{11}^{-1} \frac{1}{n} \sum_{i=1}^n x_{11}^2 & \sigma_{12}^{-1} \frac{1}{n} \sum_{i=1}^n x_{11} x_{12} \\ \sigma_{12}^{-1} \frac{1}{n} \sum_{i=1}^n x_{12} x_{11} & \sigma_{22}^{-1} \frac{1}{n} \sum_{i=1}^n x_{12}^2 \end{bmatrix} \times$$

$$\times \begin{bmatrix} \sigma_{11}^{-1} \frac{1}{n} \sum_{i=1}^n x_{11} x_{12} + \sigma_{12}^{-1} \frac{1}{n} \sum_{i=1}^n x_{12} x_{12} \\ \sigma_{12}^{-1} \frac{1}{n} \sum_{i=1}^n x_{12} x_{11} + \sigma_{22}^{-1} \frac{1}{n} \sum_{i=1}^n x_{12} x_{12} \end{bmatrix}.$$

(а) Покажите, что $\text{plim } b = 0$. Будет ли $\text{plim } a = 0$?

10. (По желанию. Добавление точно идентифицируемых уравнений не увеличивает эффективность.) Рассмотрим модель 3SLS с двумя уравнениями, в которой второе уравнение идентифицируемо точно (то есть $L_2 = K$). Пусть $\hat{\beta}_{1,3SLS}$ — это 3SLS-оценки коэффициентов первого уравнения, а $\hat{\beta}_{2,3SLS}$ — это 2SLS-оценки этих же коэффициентов.

- (а) Запишите $\text{Avar}(\tilde{\delta}_{1,2SLS})$ в терминах σ_{11} и A_{11} , определенных в (4.5.16).
- (б) Покажите, что $\text{Avar}(\tilde{\delta}_{1,2SLS}) = \text{Avar}(\tilde{\delta}_{1,2SLS})$. Указание: Используйте формулу Фробениуса для нахождения обратной матрицы для блочных матриц (см. формулу (A.10) приложения A), запишите левый верхний (размера $(J-1) \times (J-1)$) блок $\text{Avar}(\tilde{\delta}_{1,2SLS})$ как G^{-1} , где

$$G = \sigma^{11} A_{11} - \left(\frac{\sigma^{12} \cdot \sigma^{21}}{\sigma^{22}} \right) A_{12} A_{22}^{-1} A_{21}.$$

Но и соответствия в (4.5.16)

$$\begin{aligned} A_{11} A_{22}^{-1} A_{21} &= E(x_{11}x_1') [E(x_1x_1')]^{-1} E(x_1x_2') [E(x_2x_2')]^{-1} E(x_2x_1') \\ &\quad \times E(x_2x_2') [E(x_2x_2')]^{-1} E(x_1x_2') = \\ &= E(x_1x_1') [E(x_1x_1')]^{-1} E(x_1x_2') \\ &\quad (\text{так как } \#x_1 (= L_1) = \#x_2 (= K)) \\ &= A_{11}. \end{aligned}$$

Таким образом,

$$\begin{aligned} G &= \left[\sigma^{11} - \left(\frac{\sigma^{12} \cdot \sigma^{21}}{\sigma^{22}} \right) \right] A_{11} = \frac{A_{11}}{\sigma_{11}}, \\ &\quad (\text{потому что } \sigma^{11} = \left(\frac{\sigma^{12} \cdot \sigma^{21}}{\sigma^{22}} \right) = 1/\sigma_{11}). \end{aligned}$$

(Линейные комбинации условной ортогональности) Для модели с несколькими уравнениями из утверждения 4.7 выведите эффективную GMM-оценку, которая использует следующие условия ортогональности.

$$E(x_{21} \cdot \varepsilon_{11} + \dots + x_{2M} \cdot \varepsilon_{1M}) = 0 \quad (1)$$

Указание: Выборочный аналог левой части этих условий ортогональности выглядит следующим образом:

$$\begin{aligned} g_n(\tilde{\delta}) &= \frac{1}{n} \sum_{i=1}^n x_{i1} \cdot y_{i1} + \dots - \frac{1}{n} \sum_{i=1}^n x_{iM} \cdot y_{iM} \\ &\quad - \left(\frac{1}{n} \sum_{i=1}^n x_{i1} \cdot z_{i1}' + \dots + \frac{1}{n} \sum_{i=1}^n x_{iM} \cdot z_{iM}' \right) \tilde{\delta}. \end{aligned}$$

Количество условий ортогональности равно количеству коэффициентов. Эффективная GMM-оценка является IV-оценкой, которая является решением $g_n(\tilde{\delta}) = 0$.

Эмпирические упражнения

В файле GREENE.ASC содержится данные по следующим переменным:

- Столбец 1: индикатор фирмы в выборке Норлова 1966.
- Столбец 2: издержки в миллионах долларов в ценах 1970 года.
- Столбец 3: выпуск в миллионах часовых часов.
- Столбец 4: цена трудовых ресурсов.
- Столбец 5: цена капитала.
- Столбец 6: цена топлива.
- Столбец 7: доля издержек на труд.
- Столбец 8: доля издержек на капитал.

В выборке 90 наблюдений. Доля издержек на топливо может быть вычислена как единица минус сумма долей издержек на труд и на капитал. Эти данные были использованы в работе [Christensen and Greene, 1976].

В тексте даны уравнение системы (для доли затрат на топливо) не использовалось при оценивании модели со случайными эффектами. Чтобы проверить, что результаты будут численно одинаковыми, используйте второе уравнение (для доли затрат на капитал) и устраните проблему при введении ограничения однородности. Результатом этого будет система из двух уравнений:

$$\begin{cases} x_1 = \alpha_1 + \gamma_{11} \log(p_1/p_2) + \gamma_{12} \log(p_3/p_2) + \gamma_{13} \log(Q) + \varepsilon_1 \\ \text{(доля издержек на труд)}, \\ x_2 = \alpha_2 + \gamma_{21} \log(p_1/p_2) + \gamma_{22} \log(p_3/p_2) + \gamma_{23} \log(Q) + \varepsilon_2 \\ \text{(доля издержек на топливо)}. \end{cases}$$

Назовем это *системой без ограничений*. Система с *равновесными* накладывает условие симметричности $\gamma_{12} = \gamma_{21}$ на систему уравнений без ограничений.

- (а) Проверьте, что вы можете воспроизвести значения простых выборочных статистик, приведенные в тексте.
- (б) Совпадение полученных результатов? Ответьте модели со случайными эффектами (то есть многомерную регрессию при ограничении, связывающем левые уравнения системы) для системы с ограничениями. Можете ли вы воспроизвести оценки коэффициентов, приведенные в тексте? Вы должны сначала применить OLS к каждому уравнению системы без ограничений по отдельности, чтобы вычислить $\hat{\Sigma}^*$ (Замечание относительно вычислений: запрограммировать оценку же распределения гораздо проще, используя программные модули, такие как TSP и RATS. Однако эти программные модули не делают статистику Сардана для SUR.)

Совет для пакета GAUSS: Оценка модели со случайными эффектами может быть получена с помощью операций над матрицами данных, но мы советуем использовать циклы «for». Используйте (4.6.8') на с. 321.

Совет для пакета RATS: Используйте процедуру NLSYS, которая разработана для нелинейных уравнений, но которая может быть также использована для линейных уравнений. Она может также быть использована для вычисления Σ' , как это сделано в следующих кодах:

Оценивание модели с исключением уравнения для доли
затрат на капитал

```
* Определите параметры
nonlin a1 a1 g11 g1f g1f g1f g1y g1y
* Определите систему из 2 уравнений
frm1 labor a1 = a1+g11*log(p1/pk)+g1f*log(pf/pk)
+g1y*log(kwh) frm1 fuel af =
af+g1f*log(p1/pk)+g1f*log(pf/pk) +g1y*log(kwh)
* Задайте начальные значения для параметров
compute a1=af=g11=g1f=g1f=g1f=g1y=g1y=0
* Теперь оцените многомерную регрессию
* без ограничений, связывающих разные уравнения
системы.
nlsumest(nlsigma=sigmahat) / labor fuel
```

Здесь sigmahat — матрица Σ' размера 2×2 . Оценивание этой матрицы производится с помощью следующих кодов:

```
* Теперь введите ограничения, связывающие разные
уравнения системы,
* на уравнение для доли затрат на топливо (set g1f=y1f)
* Определите параметры.
nonlin a1 af g11 g1f g1f g1y g1y
* введите условие симметрии.
frm1 fuelb af = af+g1f*log(p1/pk)+g1f*log(pf/pk)
+g1y*log(kwh)
* Многомерная регрессия с ковариационной матрицей,
* вычисляемой из остатков, полученным
* при оценивании модели OLS для каждого уравнения
в отдельности.
nlsumest(nlsigma=sigmahat) / labor fuelb
```

Совет для пакета TSP. Пример кодов TSP, которые вычисляют то же самое, что и коды RATS выше:

```
? Оцените систему без уравнения
? для доли издержек на капитал.
fml labor a1=a1+q1)*log(p1/pk)+g1f*log(pf/pk)
+g1y*log(kwh);
fml fuel a1=a1+g1f*log(p1/pk)+g1f*log(pf/pk)
+g1y*log(kwh);
? Положите значения параметров равными нулю.
parax a1 a1 g1f g1f g1f g1f g1y g1y;
? Многомерная регрессия без симметрии.
sur labor fuel;
? Сгенерируйте ковариационную матрицу ошибок по остаткам.
mform sigmahat=@covu;
?
? Сгенерируйте многомерную регрессию
? с ограничением, связывающим разные уравнения системы.
? Введите условие симметрии.
fml fuel a1=a1+g1f*log(p1/pk)+g1f*log(pf/pk)
+g1y*log(kwh);
? Многомерная регрессия при условии симметрии.
lsq(maxitw=0, wscale=sigmahat) labor fuel;
Это дает RE-оценку семи свободных параметров
```

0.010302 0.013033 0.007730

Используйте, как объяснено в тексте, ограничения в виде сумм, однородности и симметрии, чтобы вычислить точечные оценки и их стандартные ошибки для оставшихся восьми параметров. Этому должна помочь программа ANALYZ в TSP.

- (с) (По желанию для пользователей программы GAUSS.) Вычислите статистику Саргана. **Указание:** Используйте формулу из утверждения 4.7. Без ограничений должен быть статистический константа $\ln(p_1/p_2)$, $\ln(p_2/p_2)$, $\ln(p_2)$. Для $\tilde{S}^*$ используйте оценки, полученные из системы без ограничений $\hat{\sigma}_{xy} = \frac{1}{n} \sum_{i=1}^n \hat{\mu}_i \otimes \hat{\sigma}_i$ и $\hat{S}_{xx} = \frac{1}{n} \sum_{i=1}^n \hat{Z}_i \otimes \hat{\sigma}_i$. Значение статистики должно быть равным 0.6313 с p -значением 0.42621.
- (д) (Проверка симметрии.) Учтите, что статистика Саргана равна 0.6313, проведите тестирование следующих гипотез. Пусть издержки на капитал — система без ограничений. Нулевой гипотезой состоит в выполнении ограничения симметрии $\gamma_{12} = \gamma_{21}$. **Указание:**

Согласно утверждению 3.8 разность между статистиками Саргана при наличии и при отсутствии симметрии имеет асимптотическое распределение хи квадрат. Чему равна статистика Саргана для системы без ограничений?

- (e) (Тест Вальда на симметрию.) Проверьте ту же нулевую гипотезу с помощью статистики Вальда. Таким образом вы проверите ограничение симметрии в системе без ограничений. Проверьте, что эта тестовая статистика численно равняется статистике Саргана для системы с ограничениями, которую вы оценивали в пункте (b).

Совет для пакета TSP: Просто вставьте строку

```
analyz wald wald=glf-gfl;
```

сразу после кодов для оценивания без ограничений.

- (i) Используя оценки параметров, полученные в пункте (b), рассчитайте среднюю по 99 фирмам численность замещенного труда капиталом. У вас получится тот же результат, что в пункте?

Ответы на избранные вопросы

Аналитические выражения

3b.

$$\begin{aligned} E(\varepsilon_{im}|\mathbf{Z}) &= \\ &= E(\varepsilon_{im}|\varepsilon_1, \dots, \varepsilon_m, \varepsilon_1, \dots, \varepsilon_2, \dots, \varepsilon_{m-1}, \dots, \varepsilon_m) = \\ &= \text{(по определению } \mathbf{Z}\text{)} \\ &= E(\varepsilon_{im}|\varepsilon_1, \dots, \varepsilon_m) \text{ (по предположению i i.d.)} \\ &= 0 \text{ (по условиям условия сбалансированности).} \end{aligned}$$

Элемент (i, j) квадратной матрицы $E(\varepsilon_{im}\varepsilon_{jk}|\mathbf{Z})$ равен $0 \times 0 = 0$ по $E(\varepsilon_{im}\varepsilon_{jk}|\mathbf{Z})$. По предположению i.i.d., он равен нулю для $j \neq i$. Для $j = i$

$$\begin{aligned} E(\varepsilon_{im}\varepsilon_{ik}|\mathbf{Z}) &= \\ &= E(\varepsilon_{im}\varepsilon_{ik}|\varepsilon_1, \dots, \varepsilon_{1M}, \varepsilon_2, \dots, \varepsilon_{2M}, \dots, \varepsilon_{m1}, \dots, \varepsilon_{mM}) = \\ &= E(\varepsilon_{im}\varepsilon_{ik}|\varepsilon_1, \dots, \varepsilon_m) \text{ (по предположению i i.d.)} \end{aligned}$$

Поскольку $\varepsilon_{im} = \varepsilon_i$, а $\varepsilon_{ik} = \varepsilon_i$ любой $(\varepsilon_2, \dots, \varepsilon_{1M})$ в модели SUR, предположение условной гомоскедастичности (предположение 4.7) гласит, что $E(\varepsilon_{im}\varepsilon_{ik}|\varepsilon_1, \dots, \varepsilon_m) = \sigma_{im}$.

5б.

$$\begin{bmatrix} 1 & E(S89) & E(IQ) \\ 1 & E(S89) & E(IQ) \\ E(MED) & E(S89 \cdot MED) & E(IQ \cdot MED) \\ E(MED) & E(S89 \cdot MED) & E(IQ \cdot MED) \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} =$$

$$= \begin{bmatrix} E(LW89) \\ E(LW89) \\ E(LW89 \cdot MED) \\ E(LW89 \cdot MED) \end{bmatrix}.$$

Условие идентифицируемости системы состоит в том, что матрица коэффициентов размера 4×3 имеет полный столбцовый ранг.

5с. Если IQ и MED не коррелированы между собой, то тогда $E(IQ \cdot MED) = E(IQ) \cdot E(MED)$ и третий столбец матрицы коэффициентов равен $E(IQ)$, умноженному на целый столбец. В этом случае матрица не может иметь полный столбцовый ранг.

6. $\hat{\varepsilon}_{im} = y_{im} - z'_{im}\hat{\delta}_{im} = z_{im} - z'_{im}(\hat{\delta}_{im} - \delta_{im})$. Поэтому

$$\frac{1}{n} \sum_{i=1}^n [z_{im} - z'_{im}(\hat{\delta}_{im} - \delta_{im})][z_{in} - z'_{in}(\hat{\delta}_{in} - \delta_{in})] = (1) + (2) + (3) + (4).$$

и

$$(1) = \frac{1}{n} \sum_{i=1}^n z_{im} z_{in}$$

$$(2) = -(\hat{\delta}_{im} - \delta_{im})' \frac{1}{n} \sum_{i=1}^n z_{im} z_{in}$$

$$(3) = -(\hat{\delta}_{in} - \delta_{in})' \frac{1}{n} \sum_{i=1}^n z_{im} z_{in}$$

$$(4) = (\hat{\delta}_{im} - \delta_{im})' \left(\frac{1}{n} \sum_{i=1}^n z_{im} z'_{in} \right) (\hat{\delta}_{in} - \delta_{in}).$$

Как обычно, по предположениям 4.1 и 4.2, $(1) \rightarrow_p \sigma_{imn} = E(z_{im} z_{in})$. Для (4), согласно предположению 4.2 и предположению о том, что $E(z_{im} z'_{in})$ конечно, $\frac{1}{n} \sum_{i=1}^n z_{im} z'_{in}$ сходится по вероятности к (конечной) матрице. Таким образом, $(4) \rightarrow_p 0$.

Что касается (2), то по неравенству Коши — Шварца

$$E[z_{im} z_{in}] \leq \sqrt{E(z_{im}^2) E(z_{in}^2)}.$$

где x_{imj} — j -й элемент x_{im} . Тогда $E(x_{im} \cdot x_{im})$ конечно и (2) $\rightarrow_p 0$, потому что $\hat{\delta}_m - \delta_m \rightarrow_p 0$. Аналогичным образом (3) $\rightarrow_p 0$.

Литература

- Berndt, E., 1991, *The Practice of Econometrics*, New York: Addison-Wesley.
- Berndt, E., and D. Wood, 1975, "Technology, Prices, and the Derived Demand for Energy," *Review of Economics and Statistics*, 57, 259-268.
- Brundy, J., and D. Jorgenson, 1971, "Efficient Estimation of Simultaneous Systems by Instrumental Variables," *Review of Economics and Statistics*, 53, 207-224.
- Christensen, L., and W. Greene, 1976, "Economics of Scale in U.S. Electric Power Generation," *Journal of Political Economy*, 84, 655-676.
- Greene, W., 1997, *Econometric Analysis* (3d ed.), New Jersey: Prentice Hall.
- Jorgenson, D., 1986, "Econometric Methods for Modeling Producer Behavior," in Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics*, Volume III, Amsterdam: Elsevier Science Publishers, 1541-1813.
- Uzawa, H., 1962, "Production Function with Constant Elasticities of Substitution," *Review of Economic Studies*, 29, 291-299.
- Varian, H., 1992, *Microeconomic Analysis* (3d ed.), New York: Norton.
- Zellner, A., 1962, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 63, 502-511.
- Zellner, A., and H. Theil, 1962, "Three-Stage Least Squares: Simultaneous Estimation of Stochastic Equations," *Econometrica*, 30, 34-75.

Глава 5. Панельные данные

Лонгитюдные (*longitudinal*), или **панельные** (*panel*), данные представляют повторные наблюдения для некоторого набора кросс-секционных единиц. Как предполагает название, панель имеет два измерения: одно — для кросс-секционных единиц, второе — для наблюдений. Второе измерение обычно является временем, но бывают и исключения. Например, кросс-секционная выборка близнецов является панелью, в которой вторую размерность образуют братья и сестры. В этой главе должно приниматься во внимание различие между кросс-секционными единицами и наблюдениями — терминами, которые до сих пор использовались как синонимы. Когда это будет уместно, мы будем использовать для кросс-секционных единиц термины «группа».

Во многих приложениях мы сталкиваемся с **моделью компонент ошибки** (*error-components model*). В таких моделях ошибка в уравнении имеет компоненту, которая является общей для всех наблюдений (инвариантной к времени, если вторая размерность в панели является временем), но которая может не быть ортогональной регрессорам. Существует доступный метод, называемый оценкой фиксированных эффектов, который позволяет нам оценивать такую модель без помощи инструментальных переменных. Эта оценка и оценка случайных эффектов из предыдущей главы являются двумя основными методами в эконометрике панельных данных.

Необязательный параграф этой главы, параграф 5.3, показывает, как модифицировать оценку фиксированных эффектов, когда число наблюдений изменяется от группы к группе. В параграфе, посвященном приложению, рассматривается эмпирика роста, тема, которая в последнее время привлекает большое внимание.

Как выясняется, оценка фиксированных эффектов и другие оценки, рассматриваемые в этой главе, являются частными случаями GMM-оценок при соответствующем преобразовании модели компонент ошибки. GMM-интерпретация оценки фиксированных эффектов продолжается в аналитических упражнениях к этой главе, где вам будет предложено найти оценку, которая будет эквив-

статистически более эффективной, чем оценка фиксированных эффектов.

5.1. Модель компонент ошибки

Наиболее отразной точкой является модель с несколькими уравнениями с одинаковыми коэффициентами из утверждения 4.7. Чтобы упростить обсуждение, мы предположим, что выборка является случайной. Это предположение выполняется для многих популярных панелей, таких как Panel Study of Income Dynamics (PSID) и National Longitudinal Survey (NLS). В этих панелях большое количество — сотни или даже тысячи — кросс-секционных единиц извлечены из генеральной совокупности случайным образом.

Для этой ситуации случайных выборок предположения утверждения 4.7 можно переформулировать (см. (4.6.15)) в отношении определений y_i ($M \times 1$), Z_i ($M \times L$), ε_i ($M \times 1$):

$$\text{(система из } M \text{ линейных уравнений)} \quad y_i = Z_i\delta + \varepsilon_i, \quad (5.1.1)$$

$$\text{(случайная выборка)} \quad \{y_i, Z_i\} \text{ являются i.i.d.} \quad (5.1.2)$$

$$\text{«предположение SUR»} \quad E(x_{ih} \cdot \varepsilon_h) = 0 \quad \text{для } i, h = 1, 2, \dots, M, \quad (5.1.3)$$

то есть $E(\varepsilon_i \otimes x_i) = 0$, где $x_i = (x_{i1}, \dots, x_{iM})$,

$$\text{(идентификация)} \quad E(Z_i \otimes x_i) \text{ имеет полный столбцовый ранг}^1, \quad (5.1.4)$$

$$\text{(условная гомоскедастичность)} \quad E(\varepsilon_i \varepsilon_i' | x_i) = E(\varepsilon_i \varepsilon_i') = \Sigma, \quad (5.1.5)$$

$$\text{(невырожденность } E(g_i g_i')) \quad E(g_i g_i') \text{ невырождена, где } g_i \equiv \varepsilon_i \otimes x_i. \quad (5.1.6)$$

Как отмечалось в параграфе 4.5, поскольку при условной гомоскедастичности $E(g_i g_i') = \Sigma \otimes E(x_i x_i')$, (5.1.6) эквивалентно условию:

$$\Sigma \text{ и } E(x_i x_i') \text{ невырождены.} \quad (5.1.6')$$

При этих предположениях оценка случайных эффектов δ_{HLS} , определенная в (4.6.8') на с. 321, является эффективной GMM-оценкой.

¹ Матрицу Σ_{Zx} в предположении 4.4' можно записать как $E(Z_i \otimes x_i)$; см. центральный вопрос 4 в параграфе 4.6.

Компоненты ошибки

Данная модель, основанная на использовании панельных данных, одна из тех, в которых предполагается, что случайная ошибка ε_{it} состоит из двух компонент:

$$\varepsilon_{it} = \alpha_i + \eta_{it} \quad (5.1.7)$$

Первая необъяснимая компонента α_i , без индекса t и, следовательно, входящая общей для всех M уравнений i -м наблюдению, называется **индивидуальным эффектом** (*individual effect*), **индивидуальной неоднородностью** (*individual heterogeneity*) или **фиксированным эффектом** (*fixed effect*). Если мы определим

$$\eta_i = \begin{bmatrix} \eta_{i1} \\ \vdots \\ \eta_{iM} \end{bmatrix},$$

матричное представление системы из M уравнений будет иметь вид:

$$\begin{bmatrix} y_{i1} \\ \vdots \\ y_{iM} \end{bmatrix} = \begin{bmatrix} z'_{i1}\delta \\ \vdots \\ z'_{iM}\delta \end{bmatrix} + \begin{bmatrix} \alpha_i \\ \vdots \\ \alpha_i \end{bmatrix} + \begin{bmatrix} \eta_{i1} \\ \vdots \\ \eta_{iM} \end{bmatrix}$$

или

$$y_i = Z_i\delta + 1_M \cdot \alpha_i + \eta_i \quad (i = 1, 2, \dots, N), \quad (5.1.8')$$

где 1_M является M -мерным вектором из единиц.

Условие ортогональности (5.1.3) выполняется, если регрессоры системы ортогональны обеим компонентам ошибки, то есть если

$$E(z_{it} \cdot \alpha_i) = 0 \quad \text{для } i = 1, 2, \dots, N \quad (5.1.8a)$$

и

$$E(z_{it} \cdot \eta_{it}) = 0 \quad \text{для } t, i = 1, 2, \dots, M \quad (5.1.8b)$$

Однако во многих приложениях, которые используют панельные данные, предположение (5.1.8a) не является разумным. Это потому, что фиксированный эффект представляет некоторую постоянную характеристику индивидуальной экономической единицы, что иллюстрируется следующими примерами:

Пример 5.1 (производственная функция с неоднородностью)
В логлинейной производственной функции (3.2.13) из параграфа 3.2.2 положим, что эффективность фирм α_i остается постоянной во времени. Тогда уравнение для года t имеет вид:

$$\log(Q_{it}) = \phi_0 + \phi_1 \log(L_{it}) + u_i + v_{it},$$

где Q_{itm} является выпуском фирмы i в год m , L_{itm} является количеством затраченного труда в год m и η_{itm} является технологическим шоком в год m . Это уравнение можно записать как (5.1.1'), полагая

$$y_{itm} = \log(Q_{itm}), \quad x_{itm} = (1, \log(L_{itm}))', \quad \delta = (\phi_1, \phi_2)', \quad \alpha_i = \alpha, \quad \eta_{itm} = \eta_{itm}.$$

Также

$$w_i = (1, \log(L_{i1}), \dots, \log(L_{iM}))'.$$

При совершенной конкуренции индивидуальный эффект α_i будет положительно коррелирован с количеством затраченного труда, поскольку эффективные фирмы, α_i которых выше, чем их среднее значение, будут нанять большее количество рабочей силы для расширения (см. формулу (3.2.14)). Если w_{itm} представляет неожиданные токи, которые являются непредвиденными для фирмы в момент выбора входных показателей, разумно предположить, что w_{itm} не коррелирована с регрессорами.

Пример 5.2 (уравнение заработной платы): Для простоты удалим стаж из уравнения заработной платы в примере 4.2, но предположим, что, в дополнение к данным за 1969 и 1980 годы, доступны также данные за 1982 год. Предположим также, что коэффициенты при образовании и при IQ остаются постоянными во времени, но константа от времени зависит (например, вследствие влияния на заработную плату делового цикла). С учетом декомпозиции ошибки (5.1.7), система из трех уравнений принимает вид:

$$\begin{cases} LW69_i = \phi_1 + \beta S69_i + \gamma IQ_i + \alpha_i + \eta_{i1}, \\ LW80_i = \phi_2 + \beta S80_i + \gamma IQ_i + \alpha_i + \eta_{i2}, \\ LW82_i = \phi_3 + \beta S82_i + \gamma IQ_i + \alpha_i + \eta_{i3}. \end{cases}$$

Как было отмечено в конце параграфа 4.6, даже если некоторое подмножество коэффициентов изменяется от уравнения к уравнению, эту систему можно записать как несколько уравнений с общими коэффициентами. В данном примере это достигается с помощью

$$Z_i = \begin{bmatrix} x_{i1}' \\ x_{i2}' \\ x_{i3}' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & S69_i & IQ_i \\ 0 & 1 & 0 & S80_i & IQ_i \\ 0 & 0 & 1 & S82_i & IQ_i \end{bmatrix}, \quad \delta' = (\phi_1, \phi_2, \phi_3, \beta, \gamma). \quad (5.1.9)$$

Также

$$x_i = (1, S69_i, S80_i, S82_i, IQ_i)'.$$

Ошибка включает в себя детерминанты заработной платы, не учтенные переменными образования и IQ . Разумно предположить, что они разделяются между перманентными чертами индивида (эта часть обозначается

как α_i), который влияет бы на выбор образования индивидом, и другие факторы (эта часть обозначается как ε_{it}) не коррелирующая с регрессорами, такими как образование и возраст, и логарифм уровня заработной платы.

Групповые средние

К счастью, существует популярная оценка, называемая **оценкой фиксированных эффектов** (*fixed-effects estimator*), которая робастна к нарушению предположения (5.1.8a), то есть корреляция является состоятельной, даже когда регрессоры не являются строго экзогенными относительно эффекту α_i . Прелюдия обсуждается в следующем параграфе, заметим, что эта оценка применяется к системе из M уравнений, преобразованной из первоначальной системы (5.1.7). Матрицы, используемая для преобразования, является аннулятором, ассоциированным с $\mathbf{1}_M$:

$$\begin{aligned} \mathbf{Q} &\equiv \mathbf{I}_M - \mathbf{1}_M(\mathbf{1}_M' \mathbf{1}_M)^{-1} \mathbf{1}_M' = \\ &= \mathbf{I}_M - \frac{1}{M} \mathbf{1}_M \mathbf{1}_M' \\ &= \mathbf{I}_M - \begin{bmatrix} 1/M & \cdots & 1/M \\ \vdots & & \vdots \\ 1/M & \cdots & 1/M \end{bmatrix}, \end{aligned} \quad (5.1.10)$$

Что эта матрица делает — исключает **отклонения от групповых средних** (*deviations from group means*). Например, умножение M -мерного вектора $\mathbf{y}$ слева на $\mathbf{Q}$ приводит к

$$\bar{\mathbf{y}}_i = \mathbf{Q} \mathbf{y}_i = \begin{bmatrix} y_{i1} - \bar{y}_1 \\ \vdots \\ y_{iM} - \bar{y}_M \end{bmatrix} = \mathbf{y}_i - \mathbf{1}_M \bar{\mathbf{y}}, \quad (5.1.11)$$

где

$$\bar{\mathbf{y}}_i = \frac{1}{M} \mathbf{1}_M' \mathbf{y}_i = \frac{1}{M} \sum_{m=1}^M y_{im}$$

является **групповым средним** (*group mean*) для зависимой переменной. Заметим, что это среднее по m , а не по i , оно специфично для каждой группы (i). Аналогично с регрессорами: каждый столбец $\mathbf{Q} \mathbf{Z}$ является вектором отклонений для соответствующего столбца $\mathbf{Z}$.

Репараметризация

Неудивительно, однако, что робастность к нарушениям (3.1.8a) имеет свою цену: некоторые из параметров модели могут стать неидентифицируемыми после такого преобразования посредством Q . Приведем два примера.

- Очевидный случай — коэффициент при переменной, удаленной из ряда уравнений. Например, IQ_1 в примере 5.2 является общим регрессором (см. (5.1.9)), так что столбец в Z_1 , соответствующий IQ_1 (пятый столбец), равен $1_M \cdot IQ_1$, и пятый столбец в QZ_1 является вектором из нулей. Следовательно, коэффициент при IQ_1 не может быть идентифицирован после такого преобразования. Чтобы избежать коллинеарности при общих регрессорах (и остальных), пусть b , является вектором общих регрессоров, и запишем тогда $M \times L$ матрицу регрессоров Z_1 как

$$Z_1 = (b' \cdot 1_M b_1') \quad (5.1.12)$$

вектор коэффициентов δ тогда им соответствует:

$$\delta = \begin{bmatrix} \beta' \\ \gamma' \end{bmatrix}.$$

Вектор коэффициентов γ не может быть идентифицирован после преобразования.

- Также может случиться, что, в дополнение к γ , некоторые коэффициенты в β неидентифицируемы, в зависимости репараметризации. Рассмотрим пример 5.2, где каждое уравнение имеет различные константы. Одну из констант нужно удалить из F , к примеру b_1 , поскольку можно создать общую константу, беря линейную комбинацию констант, специфичных для каждого уравнения. Однако репараметризация определит:

$$F = \begin{bmatrix} 1 & 0 & S69_1 \\ 0 & 1 & S80_1 \\ 0 & 0 & S82_1 \end{bmatrix}, \quad b_1 = (1, IQ_1)'. \quad (5.1.13)$$

так что

$$Z_1 = \begin{bmatrix} 1 & 0 & S69_1 & 1 & IQ_1 \\ 0 & 1 & S80_1 & 1 & IQ_1 \\ 0 & 0 & S82_1 & 1 & IQ_1 \end{bmatrix}, \quad \beta = (\alpha_1 - \alpha_3, \alpha_2 - \alpha_3, \beta_1)', \quad \gamma = (\alpha_3, \gamma_1)'. \quad (5.1.14)$$

В более общем случае условие идентификации для оценивания фиксированных эффектов состоит в том, что F_i и b_i определяются так, чтобы

$$E(QF_i \otimes x_i) \text{ имеет полный столбцовый ранг,} \quad (5.1.15)$$

где $x_i = \text{объединение } (x_{i1}, \dots, x_{iM})'$. Почему это является условием идентификации? Потому что (как вы увидите более ясно в аналитическом упрещении 2) оценка фиксированных эффектов является частным случаем оценки случайных эффектов, примененной к преобразованной регрессии Qy на QF . (5.1.15) является просто адаптацией условия идентификации (5.1.4) к преобразованной регрессии.

С Z_i , разделенной таким образом между F_i и b_i , систему из M уравнений (5.1.1') на с. 334 можно записать как

$$\begin{aligned} y_i &= F_i\beta + 1_M \cdot b_i' \gamma + 1_M \cdot \alpha_i + \eta_i \quad (i = 1, 2, \dots, n), \quad \text{или} \\ y_{im} &= f'_{im}\beta + b'_i \gamma + \alpha_i + \eta_{im} \quad (i = 1, 2, \dots, n; m = 1, \dots, M). \end{aligned} \quad (5.1.1'')$$

где f'_{im} является m -й строкой в F_i . Модель компонент ошибки (*error-component model*) состоит из предположений (5.1.1'') (M линейных уравнений), (5.1.2) (случайная выборка), (5.1.8) (предположение SUR с двумя компонентами ошибки), (5.1.4) (идентификация), (5.1.5) (условная гомоскедастичность), (5.1.6) (невырожденность $E(g_i g_i')$) и (5.1.15) (идентификация преобразованных эффектов). Мы включаем дополнительное условие идентификации (5.1.15) для оценивания с фиксированными эффектами, так что обе оценки, случайных и фиксированных эффектов определены для той же самой модели.

Контрольные вопросы

1. Проверка предположений утверждения 4.7. Проверьте, что условия утверждения 4.7 удовлетворены, когда выполнены (5.1.4)–(5.1.6). Покажите, что условие идентификации (5.1.4) выполняется, когда $E(x_i x_i')$ невырождена. Указание: x_{im} является подмножеством x_i . Таким образом, (5.1.4) является излишним.
2. (Неоднозначность репараметризации.) (5.1.13) не является единственной репараметризацией для примера 5.2. Если F_i определяется как

$$F_i = \begin{bmatrix} 0 & 0 & S\delta\theta_1 \\ 1 & 0 & S\delta\theta_2 \\ 0 & 1 & S\delta\theta_3 \end{bmatrix},$$

то как следует определять b_i и β ?

3. Без репараметризации F_i в примере 5.2 равна:

$$F_i = \begin{bmatrix} 1 & 0 & 0 & S\delta\theta_1 \\ 0 & 1 & 0 & S\delta\theta_2 \\ 0 & 0 & 1 & S\delta\theta_3 \end{bmatrix}$$

с $b_i = IQ_i$. Проверьте, что (5.1.15) не удовлетворяется.

5.2. Оценка фиксированных эффектов

Формулы

Как уже было отмечено, оценка фиксированных эффектов определяется для преобразованной системы. Умножив обе части (3.1.1') на α слева на Q и используя тот факт, что $Q1_{\Delta} = 0$, мы получаем

$$Qy_i = QF_i\beta + Q\eta_i$$

или

$$\tilde{y}_i = \tilde{F}_i\beta + \tilde{\eta}_i \quad (i = 1, 2, \dots, n), \quad (5.2.1)$$

где

$$\tilde{y}_i = Qy_i, \quad \tilde{F}_i = QF_i, \quad \tilde{\eta}_i = Q\eta_i. \quad (5.2.2)$$

(M×1) (M×K) (M×1)

Таким образом, фиксированный эффект α , в общем регрессионном выражении из преобразованных уравнений. Сформируем теперь объединенную выборку преобразованных y_i и F_i как

$$\tilde{y} = \begin{bmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_n \end{bmatrix}, \quad \tilde{F} = \begin{bmatrix} \tilde{F}_1 \\ \vdots \\ \tilde{F}_n \end{bmatrix}. \quad (5.2.3)$$

(MN×1) (MN×K)

Оценка фиксированных эффектов (*fixed-effects estimator*) $\hat{\alpha}$ в β , обозначаемая как $\hat{\beta}_{FE}$, является объединенной OLS оценкой, т.е. есть OLS-оценкой, примененной к объединенной выборке $(\tilde{y}, \tilde{F})$ объема MN . Следовательно,

$$\begin{aligned} \hat{\beta}_{FE} &\equiv (\tilde{F}'\tilde{F})^{-1}\tilde{F}'\tilde{y} = \\ &= \left(\frac{1}{n}\sum_{i=1}^n \tilde{F}_i'\tilde{F}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^n \tilde{F}_i'\tilde{y}_i = \\ &= \left[\frac{1}{n}\sum_{i=1}^n (QF_i)'(QF_i)\right]^{-1}\frac{1}{n}\sum_{i=1}^n (QF_i)'Qy_i = \\ &= \left(\frac{1}{n}\sum_{i=1}^n F_i'QQF_i\right)^{-1}\frac{1}{n}\sum_{i=1}^n F_i'Qy_i = \\ &= \left(\frac{1}{n}\sum_{i=1}^n F_i'QF_i\right)^{-1}\frac{1}{n}\sum_{i=1}^n F_i'Qy_i. \end{aligned} \quad (5.2.4)$$

(поэтому Q симметрична и идемпотентна)

Подставляя верхнее уравнение (5.2.1) в (5.2.4), можно получить ошибку оценки как

$$\begin{aligned}\hat{\beta}_{OLS} = \beta &= \left(\frac{1}{n} \sum_{i=1}^n \tilde{F}_i' \tilde{F}_i \right)^{-1} \frac{1}{n} \sum_{i=1}^n \tilde{F}_i' \tilde{y}_i = \\ &= \left(\frac{1}{n} \sum_{i=1}^n v_i' Q F_i \right)^{-1} \frac{1}{n} \sum_{i=1}^n v_i' Q y_i.\end{aligned}\quad (5.2.5)$$

Будучи основанным на отклонениях от групповых средних, оценка фиксированных эффектов известна также как **внутри-оценка** (*within estimator*), или **ковариационная оценка** (*covariance estimator*). Другое, и, возможно, более популярное, ее получение заключается в применении OLS к уравнению, то есть не в отклонениях, но со специфическими дамми для групп (5), добавленными к списку регрессоров. По этой причине эта оценка называется также **OLS-дамми** (*least squares dummy variables — LSDV*) оценкой. LSDV интерпретация позволяет нам увидеть, какие нужны дополнительные предположения для модели компонент ошибки при разработке теории конечных выборок для этой оценки. Это остается в качестве аналитического упражнения 1.

Свойства на больших выборках

Как мы докажем в аналитическом утверждении 2, оценку фиксированных эффектов можно также записать как GMM-оценку, что дает возможность непосредственной разработки теории для больших выборок для оценки фиксированных эффектов, что подытоживается в утверждении 5.1 ниже. Доказательство этого утверждения, однако, можно выполнить более просто, используя выражение (5.2.5) для ошибки оценки.

Утверждение 5.1 (свойства оценки фиксированных эффектов на больших выборках): Рассмотрим модель компонент ошибок (состоящую из (5.1.1') и с 353, (5.1.2), (5.1.8)–(5.1.4), (5.1.6) и (5.1.15)), со стандартным предположением SUR (5.1.8), требуя только, чтобы

$$E(f_{itm} | \eta_{it}) = 0 \quad \text{для всех } m, i, t (= 1, 2, \dots, M). \quad (5.2.6)$$

где f_{itm} — m -й строкой F_i . Определим $\tilde{y}_i$, $\tilde{F}_i$, и $\tilde{\eta}_i$ по формулам (5.2.2). Тогда:

(a) оценка фиксированных эффектов (5.2.4) состоятельна и асимптотически нормальна с

$$\text{Asym}(\hat{\beta}_{FE}) = [E(\tilde{F}_i' \tilde{F}_i)]^{-1} E[\tilde{F}_i' E(\tilde{\eta}_i \tilde{\eta}_i') \tilde{F}_i] [E(\tilde{F}_i' \tilde{F}_i)]^{-1}. \quad (5.2.7)$$

(b) Эта асимптотическая дисперсия состоятельно оценивается как

$$\widehat{\text{Avar}}(\tilde{\beta}_{FE}) = \left(\frac{1}{n} \sum_{i=1}^n \tilde{F}_i' \tilde{F}_i \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n \tilde{F}_i' \tilde{V} \tilde{F}_i \right) \left(\frac{1}{n} \sum_{i=1}^n \tilde{F}_i' \tilde{F}_i \right)^{-1}, \quad (5.2.8)$$

где $\tilde{V}$ является выборочной матрицей из перекрестных моментов преобразованных остатков, связанных с оценкой фиксированных эффектов:

$$\tilde{V}_{(M \times M)} = \frac{1}{n} \sum_{i=1}^n (\tilde{y}_i - \tilde{F}_i' \tilde{\beta}_{FE})(\tilde{y}_i - \tilde{F}_i' \tilde{\beta}_{FE})'. \quad (5.2.9)$$

Единственная нетривиальная часть доказательства состоит в том, чтобы показать:

- (1) $E(\tilde{F}_i' \tilde{F}_i)$ является невырожденной (и, следовательно, обратимой)
- (2) $E(\tilde{F}_i' Q \eta_i) = 0$ (что нужно для состоятельности) и
- (3) $E(\tilde{F}_i' \tilde{\eta}_i \tilde{\eta}_i' \tilde{F}_i)$ (что является Avar для $\frac{1}{n} \sum_{i=1}^n \tilde{F}_i' \tilde{\eta}_i$) $= E[\tilde{F}_i' E(\tilde{\eta}_i \tilde{\eta}_i' | \tilde{F}_i) \tilde{F}_i]$.

Остатывая часть доказательства является простым применением LLN Колмогорова и CLT Линдберга—Левы.

Доказательство того, что из (5.1.15) следует (1), остается в качестве дополнительного аналитического упражнения.

Для доказательства (2) используйте формулу (4.6.16b) и перепишите математическое ожидание как

$$\begin{aligned} E(\tilde{F}_i' Q \eta_i) &= E\left(\sum_{m=1}^M \sum_{h=1}^M q_{mh} f_{im} \cdot \eta_{ih}\right) = \quad (q_{mh} \equiv (m, h) \text{ элемент } Q) \\ &= \sum_{m=1}^M \sum_{h=1}^M q_{mh} E(f_{im} \cdot \eta_{ih}). \end{aligned} \quad (5.2.10)$$

Оно равно нулю вследствие условий ортогональности (5.2.6).

Доказательство (3) заключается в следующем. Сначала заметим, что

$$E(\tilde{F}_i' \tilde{\eta}_i \tilde{\eta}_i' \tilde{F}_i) = E[\tilde{F}_i' E(\tilde{\eta}_i \tilde{\eta}_i' | x_i) \tilde{F}_i].$$

(Это следует из закона полных математических ожиданий и того факта, что x_i — объединение $(z_{i1}, \dots, z_{iM})$, включает в себя элементы на $\tilde{F}_i$ и $\tilde{F}_i$ является функцией от F_i .) Но $E(\tilde{\eta}_i \tilde{\eta}_i' | x_i)$ равно безусловному сред-

и уу $E(\tilde{\eta}_i \tilde{\eta}_i')$, поскольку

$$\begin{aligned} E(\tilde{\eta}_i \tilde{\eta}_i' | x_i) &= E(\tilde{\varepsilon}_i \tilde{\varepsilon}_i' | x_i) \\ (\text{так как } \tilde{\varepsilon}_i &\equiv Q\varepsilon_i = Q(I_M \otimes \mathbf{1}_T + \mathbf{1}_T \mathbf{1}_T') = Q\eta_i = \tilde{\eta}_i) \\ &= Q E(\varepsilon_i \varepsilon_i' | x_i) Q = \\ &= Q E(\varepsilon_i \varepsilon_i') Q \quad (\text{скалярно } (1, 1, \tilde{\eta})) = \\ &= \Sigma(\tilde{\varepsilon}_i \tilde{\varepsilon}_i') = \\ &= E(\tilde{\eta}_i \tilde{\eta}_i') \quad (\text{так как } \tilde{\varepsilon}_i = \tilde{\eta}_i). \end{aligned} \quad (5.2.11)$$

Беглый взгляд на выражение для асимптотической дисперсии должен побудить вас предполагать, что существует какая-то другая оценка, которая является асимптотически более эффективной, потому что асимптотическую дисперсию сферической панели можно записать в виде обратной матрицы. В самом деле, как мы докажем в аналитическом упражнении 2, существует оценка, которая асимптотически более эффективна, чем оценка фиксированных эффектов.

Отступление: когда η_i является сферической

Обычно в модели компонент ошибки дополнительно предполагается, что другая компонента ошибки η_i является сферической

$$E(\eta_i \eta_i') = \sigma_\eta^2 I_n, \quad \text{так что } E(\tilde{\eta} \tilde{\eta}') = \sigma_\eta^2 Q. \quad (5.2.12)$$

Если, как мы делаем на панельных данных, мы являемся притом, то это предположение часто называется предположением об «отсутствии серийной корреляции». Отсутствие серийной корреляции в этом случае не означает путать с условием, что η_i не коррелирована с η_j для $i \neq j$. Последнее является следствием нашего поддерживаемого предположения о том, что выборки являются i.i.d.

Подставляя это в (5.2.7) и замечая, что $Q\bar{F} = \bar{F}$, мы получаем, что выражение для асимптотической дисперсии упрощается до

$$\text{Asym}(\hat{\beta}_{FE}) = \sigma_\eta^2 \cdot \left[E(\tilde{F}' \tilde{F}) \right]^{-1}. \quad (5.2.13)$$

Если существует состоятельная оценка $\hat{\sigma}_\eta^2$ для σ_η^2 , то эта асимптотическая дисперсия состоятельно оценивается как

$$\widehat{\text{Asym}}(\hat{\beta}_{FE}) = \hat{\sigma}_\eta^2 \cdot \left(\frac{1}{n} \sum_{i=1}^n \tilde{F}_i' \tilde{F}_i \right)^{-1}, \quad (5.2.14)$$

что равно σ , умноженному на $\hat{\sigma}_\eta^2 \cdot (\bar{F}' \bar{F})^{-1}$, то есть n , умноженному на обычное выражение для стандартной матрицы дисперсии, если OLS применяется к объединенной выборке $(\bar{y}, \bar{F})$ объема Mn .

Чтобы расширить аналогию с OLS, определим SSR как

$$SSR = (\tilde{y} - \tilde{F}\tilde{\beta}_{FE})'(\tilde{y} - \tilde{F}\tilde{\beta}_{FE}) = \sum_{i=1}^n (\tilde{y}_i - \tilde{F}_i\tilde{\beta}_{FE})'(\tilde{y}_i - \tilde{F}_i\tilde{\beta}_{FE}). \quad (5.2.16)$$

В объединенной OLS на отклонения от групповых средних есть $\# \beta$ коэффициентов, которые нужно оценить. Обычная OLS-оценка дисперсии ошибок равна

$$\frac{SSR}{Mn - \# \beta}. \quad (5.2.16)$$

Она, однако, не является состоятельной для σ_ε^2 . Вас попросят показать в аналитическом упражнении 3, что состоятельной оценкой является

$$\frac{SSR}{Mn - n - \# \beta}. \quad (5.2.17)$$

(Вообще-то, и $SSR/(Mn - n)$ также является состоятельной.) [Причина того, что на знаменателе (5.2.16) нужно вычесть n , заключается в том, что M преобразованных уравнений не являются линейно независимыми; для обеих частей M преобразованных уравнений сумма по i всегда равна нулю, как вы сможете увидеть, умножая обе части (5.2.1) слева на $1'_M$; и не только то, что $1'_M Q = 0$. Эффективный объект объединенной выборки равен $Mn - n$, а не Mn . Использование (5.2.16) вместо (5.2.17) является очень распространенной ошибкой, которая приводит к недооценке стандартных ошибок и терпеливое замечаний. Например, если $M = 3$, $n = 1000$ и $\# \beta = 5$, то t -значение будет переоценено примерно на 23% (= квадратному корню из $2995/1995$ минус единица).

Смешение случайных эффектов с фиксированными

Вернемся теперь в общему случаю с отсутствием ограничений на $E(\varepsilon_i \varepsilon_j')$, так что ошибки могут иметь произвольную «структуру» «серийной корреляции» вдоль n . Сравнивая (5.2.6) с (5.1.8), мы видим, что условия ортогональности, не используемые в оценке фиксированных эффектов, имеют вид:

$$E(\varepsilon_{it} \cdot \alpha_i) = 0 \text{ для всех } i, \quad E(\varepsilon_{it} \cdot \alpha_i) = 0, \quad E(\varepsilon_{it} \cdot \alpha_{im}) = 0 \text{ для всех } i, \quad (5.2.18)$$

То есть оценка фиксированных эффектов является робастной к нарушениям (5.2.18). Пусть $\tilde{\beta}_{FE}$ является элементом в $\tilde{\beta}_{FE}$: {оценка случайных эффектов, примененная к оригинальной системе (5.1.16) из M уравнений на n 368 , который соответствует $\tilde{\beta}$ }.

$$\tilde{\beta}_{FE} = \begin{bmatrix} \tilde{\beta}_{FE} \\ \tilde{\gamma}_{FE} \end{bmatrix}. \quad (5.2.19)$$

Также пусть $\text{Avar}(\hat{\beta}_{\text{GLS}})$ является асимптотической дисперсией для $\hat{\beta}_{\text{GLS}}$. Это левая часть матрицы $\text{Avar}(\hat{\beta}_{\text{GLS}})$ (заданная в (4.6.9') на с. 321), соответствующая β .

Оценка случайных эффектов $\hat{\beta}_{\text{GLS}}$ является эффективной, в то время как $\hat{\beta}_{\text{FE}}$ является состоятельной, но неэффективной. Однако, если (5.2.18) нарушается, больше нет гарантии того, что $\hat{\beta}_{\text{FE}}$ будет состоятельной, в то время как $\hat{\beta}_{\text{GLS}}$ останется состоятельной. Таким образом, естественным тестом для (5.2.18) будет рассмотрение разности между этими двумя оценками:

$$\hat{q} \equiv \hat{\beta}_{\text{FE}} - \hat{\beta}_{\text{GLS}}. \quad (5.2.20)$$

Легко доказать, что вектор $\sqrt{n}\hat{q}$ асимптотически нормален. Поинтид Хаусмана (первоначально разработанный для ML-оценок, но, как доказано в аналитическом упражнении 9 к главе 4, применимый и к GLS-оценкам) предполагает, что

$$\text{Avar}(\hat{q}) = \text{Avar}(\hat{\beta}_{\text{FE}}) - \text{Avar}(\hat{\beta}_{\text{GLS}}). \quad (5.2.21)$$

(Поэтому нет необходимости включать асимптотическую ковариацию между $\hat{\beta}_{\text{FE}}$ и $\hat{\beta}_{\text{GLS}}$.) По утверждению 4.7, левая часть матрицы (4.6.9') на с. 321 дает состоятельную оценку для $\text{Avar}(\hat{\beta}_{\text{GLS}})$. По утверждению 5.1 следует, что $\text{Avar}(\hat{\beta}_{\text{FE}})$ состоятельно оценивается посредством (5.2.8). Следовательно, состоятельная оценка для $\text{Avar}(\hat{q})$ является их разностью. Записывая как $\widehat{\text{Avar}}(\hat{q})$. В приложении доказывается, что (i) $\widehat{\text{Avar}}(\hat{q})$ является квадратичной (и, следовательно, положительно определенной) и (ii) статистика Хаусмана, определенная ниже, гарантированно является неотрицательной для любой выборки $\{y_i, Z_i\}$. Подведем итог.

Утверждение 5.2 (тест Хаусмана на спецификацию): Допустим, что выполняются предположения модели компонент ошибки ((5.1.1') (на с. 358), (5.1.2), (5.1.8), (5.1.4) (5.1.6) и (5.1.15)). Определим $\hat{q}$ и $\widehat{\text{Avar}}(\hat{q})$ как только что было описано. Тогда статистика Хаусмана

$$H \equiv n \cdot \hat{q}' (\widehat{\text{Avar}}(\hat{q}))^{-1} \hat{q} \quad (5.2.22)$$

асимптотически распределена как χ^2 -квадрат с $k-3$ степенями свободы. Она неотрицательна в конечных выборках.

Этот тест является тестом на спецификацию, поскольку он может обнаруживать нарушение (5.2.18), которое является частью поддерживаемых предположений модели компонент ошибки. Более полно рассмотрим последовательность локальных альтернатив, которые удовлетворяют всем предположениям модели компонент ошибки, за исключением (5.2.18)¹

¹Понятие локальных альтернатив вводится в параграфе 2.4

При некоторых дополнительных технических предположениях можно показать, что статистика Хаусмана сходится вдоль этих последовательностей локальных альтернатив к нецентральному χ^2 распределению с некоторым параметром нецентральности¹. То есть статистика Хаусмана имеет мощность против локальных альтернатив, при которых оценка случайных эффектов несостоятельна.

Ослабление условной гомоскедастичности

Получить распределение оценки фиксированных эффектов на больших выборках нетрудно и без условной гомоскедастичности. Как будет видно в следующем параграфе, обобщить результаты для больших выборок с целью охвата несбалансированных панелей на самом деле легче без условий гомоскедастичности.

Утверждение 5.3 (оценка фиксированных эффектов без условной гомоскедастичности): Уберем условную гомоскедастичность (5.1.5) из предположений утверждения 5.1. Определим $\hat{y}_i$, $\tilde{F}_i$ и $\hat{\eta}_i$ посредством (5.2.2). Тогда:

(а) оценка фиксированных эффектов (5.2.4) согласованна и асимптотически нормальна с

$$\text{Avar}(\hat{\beta}_{FE}) = \left[E(\tilde{F}_i' \tilde{F}_i) \right]^{-1} E(\tilde{F}_i' \hat{\eta}_i \hat{\eta}_i' \tilde{F}_i) \left[E(\tilde{F}_i' \tilde{F}_i) \right]^{-1}. \quad (5.2.23)$$

(б) Если кроме этого выполняется соответствующее предположение о конечности четвертого момента (которое является адаптацией предположения 4.6), то асимптотическая дисперсия состоятельно оценивается как

$$\widehat{\text{Avar}}(\hat{\beta}_{FE}) = \left(\frac{1}{n} \sum_{i=1}^n \tilde{F}_i' \tilde{F}_i \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n (\tilde{F}_i' \hat{\eta}_i \hat{\eta}_i' \tilde{F}_i) \right) \left(\frac{1}{n} \sum_{i=1}^n \tilde{F}_i' \tilde{F}_i \right)^{-1}. \quad (5.2.24)$$

где $\hat{\eta}_i \equiv \hat{y}_i - \tilde{F}_i \hat{\beta}_{FE}$ (это $\hat{\eta}_i$ не нужно путать с $\tilde{\eta}_i$ в (5.2.1)).

И опять, есть два способа доказать это. Запишите оценку фиксированных эффектов как ГММ-оценку и затем примените теорию больших выборок для ГММ-оценок. Или вы можете доказать это иным средством, рассматривая выражение (5.2.4) для ошибки оценки. Единственная нетривиальная часть доказательства — показать, что средняя сумма в выражении для $\widehat{\text{Avar}}(\hat{\beta}_{FE})$ сходится по вероятности к ее популяционному аналогу, $E(\tilde{F}_i' \hat{\eta}_i \hat{\eta}_i' \tilde{F}_i)$. Требующийся метод доказательства

¹См.: [Newey, 1983]. Требуемое техническое предположение является предположением 3 в: [Newey, 1983].

очень похож на тот, который используется для доказательства утверждения 4.2, и поэтому не повторяется здесь.

Контрольные вопросы

1. Должны ли мы, для того чтобы выполнялись результаты для больших выборок в этом параграфе, предполагать, что α_i и η_i ортогональны друг другу?
2. (Удаление дублирующих уравнений.) Как упоминалось в основном тексте, M преобразованных уравнений не являются линейно независимыми, в том смысле, что любое из M преобразованных уравнений является линейной комбинацией остальных $M - 1$ уравнений. Поэтому рассмотрим формирование объединенной выборки объема $Mm - n$, удаляя одно из преобразованных уравнений (скажем, последнее уравнение) для каждой группы (i). Является ли OLS-оценка в этой меньшей объединенной выборке оценкой фиксированных эффектов? [Ответ: Нет.]
3. (Важность перекрестной ортогональности.) Предположим, что (5.2.6) выполняется для $m = h$, но не обязательно для $m \neq h$. Будет ли оценка фиксированных эффектов обязательно состоятельной? [Ответ: Нет.]
4. Проверьте, что $E(\tilde{\eta}_i \tilde{\eta}_i')$ является вырожденной матрицей. (Однако матрицы $E[\tilde{F}_i' E(\tilde{\eta}_i \tilde{\eta}_i') \tilde{F}_i]$ в (5.2.7) и $E(\tilde{F}_i' \tilde{\eta}_i \tilde{\eta}_i' \tilde{F}_i)$ в (5.2.23) являются невырожденными, как вас попросят доказать в дополнительном аналитическом упражнении.)
5. (Состоятельное оценивание ковариаций ошибок.) Чтобы доказать, что (5.2.9) является состоятельной оценкой для $E(\tilde{\eta}_i \tilde{\eta}_i')$, вы должны применить утверждение 4.1 к преобразованной системе из M уравнений. Проверьте, что условия утверждения здесь выполняются. В частности, выполняется ли требование того, чтобы $E(\tilde{f}_{im} \tilde{f}_{ih}') (i \text{ где } \tilde{f}_{im} \text{ является } m\text{-й строкой } \tilde{F}_i)$ было конечным, в модели компонент ошибки? **Указание:** $\tilde{f}_{im}$ является линейным преобразованием x_i . Из (5.1.5) и (5.1.6) следует, что $E(x_i x_i')$ невырождена.
6. (Что тестирует статистика Хаусмана?) Для простоты предположим, что матрица $\Sigma (= E(\varepsilon_i \varepsilon_i'))$ известна. Также для простоты предположим, что общих регрессоров нет, так что $z_{im} = f_{im}$, и (5.2.18) редуцируется к (5.1.8a): $E(z_{im} \cdot \alpha_i) = 0$ для всех m . Это является ограничением, не используемым в оценке фиксированных эффектов. Является ли оценка случайных эффектов обязательно несостоятельной, когда (5.1.8a) не выполняется, но выполняются все другие предположения модели компонент ошибки (такие как (5.1.8b))? **Указание:** Покажите, что без (5.1.8a)

$$\text{plim } \hat{\delta}_{RE} - \delta = E(Z_i' \Sigma^{-1} Z_i)^{-1} E(Z_i' \Sigma^{-1} \varepsilon_i),$$

где $\varepsilon_i = 1 \cdot \alpha_i + \eta_i$. При условии, что (5.1.8b) выполняется, (5.1.8a) является достаточным для того, чтобы $E(Z_i' \Sigma^{-1} \varepsilon_i)$ было равно нулю. Является ли оно необходимым? Таким образом, строго говоря, тест Хаусмана не является тестом для (5.1.8a); он является тестом для $E(Z_i' \Sigma^{-1} \varepsilon_i) = 0$.

5.3. Несбалансированные панели (необязательный параграф)

Применяя модель нескольких уравнений к панельным данным, мы сделали некое, но важное предположение о том, что переменные наблюдаемы для всех it , так что число наблюдений для каждой рассматриваемой единицы одинаковое (M). Такая панель называется **сбалансированной панелью** (*balanced panel*). Однако доступных сбалансированных множеств панельных данных не существует из-за того, что одни обследуемые единицы выпадают из обследования, а другие входят в него. Неизбежно некоторые фирмы исчезают из обследуемой выборки из-за банкротств и слияний до окончания опроса, в то время как те фирмы, которых не было в начале исследования, могут быть включены в него позже. Некоторые из домохозяйств, первоначально участвовавших в опросе, не доводят его до конца или из-за исчезновения домохозяйства, или из-за того, что респондентам надоедает отвечать на одни и те же вопросы снова и снова. Панель идентичных братьев и сестер является несбалансированной, если она включает в себя, наряду с парами близнецов, также идентичные тройки (и даже семерки).

В этом параграфе показано, что оценка фиксированных эффектов легко обобщается на несбалансированные панели. Этот оптимистичный вывод, однако, зависит от предположения о том, что падающее обследование единицы в выборке не зависит от ошибок. Без этого предположения оценки уже не являются состоятельными. Этот феномен называется **смещением из-за отбора** (*selectivity bias*). Более подробное обоснование того, почему возникает смещение из-за отбора и как это можно исправить, представлено в параграфе 8.2, поскольку этот вопрос не является специфичным для панельных данных.

Обучение пропущенных наблюдений

Чтобы справиться с пропущенными наблюдениями, удобно определить дамми-переменную, d_{itn} ,

$$d_{itn} = \begin{cases} 1 & \text{если наблюдение } n \text{ находится в выборке} \\ 0 & \text{в противном случае} \end{cases} \quad (5.3.1)$$

и

$$d_i = \begin{bmatrix} d_{i1} \\ \vdots \\ d_{iM} \end{bmatrix}, \quad M_i = \sum_{n=1}^M d_{in} = \# \text{ наблюдений для группы } i. \quad (5.3.2)$$

Если наблюдение n пропущено для группы i , заполним n -е строки y_i , F_i и τ_i нулями так, чтобы они продолжали иметь фиксированный размер.

То есть

$$y_{it} = \begin{bmatrix} d_{i1} \cdot y_{it} \\ \vdots \\ d_{iM} \cdot y_{it} \end{bmatrix} = F_{it} = \begin{bmatrix} d_{i1} \cdot f_{it} \\ \vdots \\ d_{iM} \cdot f_{it} \end{bmatrix}, \quad \eta_{it} = \begin{bmatrix} d_{i1} \cdot \eta_{it} \\ \vdots \\ d_{iM} \cdot \eta_{it} \end{bmatrix}. \quad (5.3.3)$$

Таким образом, для каждого i и t либо все элементы вектора $(y_{it}, f_{it}, \eta_{it})$ наблюдаемы, либо не наблюдаемы ни один. Мы не рассматриваем случай, в котором некоторые, но не все, элементы $(y_{it}, f_{it}, \eta_{it})$ наблюдаемы.

В этих новых обозначениях, (5.1.1") и в 3.58 приобретает вид:

$$y_i = F_i \beta + d_i \cdot b_i' \gamma + d_i \cdot \alpha_i + \eta_i, \quad (i = 1, 2, \dots, N). \quad (5.3.4)$$

В случае сбалансированных панелей, т.е. $d_i = 1_M$, преобразующая матрица Q в формуле фиксированных эффектов (5.2.4) является нулевым, соответствующим 1_M (см. (5.1.10)). При наличии пропущенных наблюдений мы используем в качестве преобразующей матрицы нулевой, соответствующий

$$\begin{aligned} Q_{(M \times M)} &= I_M - d_i(d_i' d_i)^{-1} d_i' = \\ &= I_M - \frac{1}{M_i} d_i d_i' \quad (\text{поскольку } d_i' d_i = M_i). \end{aligned} \quad (5.3.5)$$

Эта матрица зависит от i , поскольку d_i для разных i различны. Однако мы оставляем эту зависимость Q от i неявной для экономности обозначений. С таким образом переопределенными y_i , F_i , η_i и Q формула для оценки фиксированных эффектов остается той же самой, что и (5.2.4). Подставляя (5.3.4) в эту формулу и замечая, что $Q d_i = 0$, мы видим, что выражение для ошибки оценки ничем не отличается от (5.2.3).

Обобщение против сжатия

Как было видно в параграфе 5.2 для случая сбалансированных панелей, оценку фиксированных эффектов можно вычислять как OLS-оценку по объединенной выборке (размера Mn) сключенной из групповых средних. Как и прежде, пусть $\tilde{y}_i$, $\tilde{F}_i$ и $\tilde{\eta}_i$ являются преобразованиями y_i , F_i , η_i через Q соответственно. Например, $\tilde{y}_i$ выглядят как

$$\tilde{y}_i \equiv Q y_i = \begin{bmatrix} d_{i1} \cdot y_{i1} - d_{i1} \cdot \bar{y}_i \\ \vdots \\ d_{iM} \cdot y_{iM} - d_{iM} \cdot \bar{y}_i \end{bmatrix},$$

где

$$\begin{aligned}\bar{y}_i &\equiv \frac{1}{\Delta_i} \sum_{t=1}^{\Delta_i} y_{it} = \\ &= \frac{1}{\Delta_i} \times (\text{сумма } y_{it} \text{ по всем } t, \text{ для которых данные доступны}). \quad (5.3.6)\end{aligned}$$

Таким образом, если наблюдение y_{it} для группы i доступно, то i имеет $\bar{y}_i$, но еще должно наблюдение y_{it} от группы i и среднее $\bar{y}_i$ по группе i среднее вычисляется из доступным наблюдением. Поскольку $y_{it} = 0$, если наблюдение не доступно, строим $\bar{y}_i$, соответствующие пропущенным наблюдениям, равны нулю. Такую же структуру имеют и столбцы $\bar{F}$. Таким образом, эти строки можно удалить из объединенной выборки без влияния на численное значение оценки фиксированных эффектов. То есть мы можем до формирования объединенной выборки «снять» $(\bar{y}_i, \bar{F}_i)$, удалив строки, соответствующие нулям.

Отсутствие смещения вследствие таблицы

Оценка фиксированных эффектов, адаптированная таким образом к несбалансированным панелям, остается состоятельной и асимптотически нормальной, если условие ортогональности (5.2.6) в утверждении 5.1 модифицируется как

(отсутствие смещения вследствие выбора)

$$E(f_{it} \cdot y_{it} | d_i) = 0 \quad (n, h = 1, 2, \dots, H) \quad (5.3.7)$$

Как показано в следующем утверждении, это условие является ключевым для гарантирования состоятельности оценки фиксированных эффектов. По закону полных математических ожиданий, это условие является более сильным, чем условие (5.2.6), требующее, чтобы было равно нулю безусловное математическое ожидание. Легко проверить пример, где (5.3.7) не выполняется, поскольку характер выбора d_i зависит от ошибки u_i . Заметим, что (5.3.7) не включает фиксированный эффект α_i . Таким образом, если зависимость выбора в выборку от ошибок реализуется только через фиксированный эффект, проблема смещения из-за такого отбора не возникает.

Если условия ортогональности (5.2.6) усиливается до (5.3.7), то тогда выводы утверждения 5.3 переносятся на несбалансированные панели.

Утверждение 5.4 (свойства на больших выборках оценки фиксированных эффектов для несбалансированной панели): Рассмотрим модель компонент ошибок, состоящей из (5.3.4), (5.1.2), (5.1.4)–(5.1.6), (5.1.15) и (5.3.7) с (y_i, F_i, u_i) , как в (5.3.3). Как и выше, определим

$(\tilde{y}_i, \tilde{F}_i, \tilde{\eta}_i)$ посредством (5.2.2), но с Q , заданной в (5.3.5). Тогда выводы утверждения 5.3 остаются в силе.

Здесь мы показываем, что $E(F_i' Q \eta_i) = 0$. (Другие части доказательства, являющиеся применением LLN Колмогорова и CLT Линдберга-Леви, те же самые, что и в доказательстве утверждения 5.3.) Поскольку

$$F_i' Q \eta_i = \sum_{m=1}^M \sum_{h=1}^M q_{mh}^{(i)} \cdot d_{im} \cdot d_{ih} \cdot f_{im} \cdot \eta_{ih}, \quad (5.3.8)$$

где $q_{mh}^{(i)}$ — (m, h) -й элемент Q , примененный к i -му уравнению, мы имеем:

$$E(F_i' Q \eta_i) = \sum_{m=1}^M \sum_{h=1}^M E(q_{mh}^{(i)} \cdot d_{im} \cdot d_{ih} \cdot f_{im} \cdot \eta_{ih}). \quad (5.3.9)$$

Если нет смещения из-за отбора, так что $E(f_{im} \cdot \eta_{ih} | d_i) = 0$, то (5.3.9) равно нулю, так как

$$\begin{aligned} E(q_{mh}^{(i)} \cdot d_{im} \cdot d_{ih} \cdot f_{im} \cdot \eta_{ih}) &= \\ &= E \left[q_{mh}^{(i)} \cdot d_{im} \cdot d_{ih} \cdot E(f_{im} \cdot \eta_{ih} | d_i) \right] \\ & \text{(поскольку } q_{mh}^{(i)} \text{ является функцией от } d_i). \end{aligned} \quad (5.3.10)$$

Интересно, что обобщение на несбалансированную панель легче без условной homoskedастичности. Чтобы (5.2.7) из утверждения 5.1 были асимптотически для средних оценки ДРЗ условный homoskedастичности, мы должны были бы предположить, что второй момент $\tilde{\eta}_i$, условный относительно x_i и d_i , равен безусловному второму моменту, а соответствующая оценка этого второго момента была бы немного сложнее.

Контрольные вопросы

1. Для $d_i = (1, 0, 1)'$ запишите Q .
2. Альтернативное получение оценки фиксированных эффектов для несбалансированных панелей. Пусть y_i^* является M_i -мерным вектором, полученным из M -мерного вектора y_i путем удаления строк, для которых нет наблюдений. Также так же создадим F_i^* ($M_i \times J$) из F_i и d_i^* из d_i (так что d_i^* является M_i -мерным вектором из единиц). Пусть Q^* ($M \times M$) является аннулятором для d_i^* . Проверьте, что оценку фиксированных эффектов можно также записать как (5.2.4) с помощью этих матриц и векторов со звездочками.
3. В параграфе 4.6 описывается объединенная OLS-оценка в уровнях для системы (4.6.1') на с. 320.

- (а) Как бы вы модифицировали оценку (3.5.11') на с. 322, чтобы охватить несбалансированные панели? **Указание:** Определите μ и Z_t ($t \neq 1$) через «облачки».
- (б) Чему равна асимптотическая дисперсия, при отсутствии условной гомосkedастичности?
[Ответ $E(Z_t Z_t')^{-1} E(Z_t \varepsilon_t \varepsilon_t' Z_t')^{-1} E(Z_t Z_t')^{-1}$].
- (в) Как бы вы определили асимптотическую дисперсию? **Указание:** Завершите и утверждения 5.3 $\tilde{F}_1$ на Z_t , и $\tilde{f}_1$ на ε_t .

5.4. Приложение: Международные различия в темпах роста

Растут ли бедные экономики быстрее, чем богатые? Если это так, то насколько быстрее? Эти вопросы рассмотрены в быстро увеличивающейся в последнее время литературе по росту¹. В этом параграфе приводится ключевое уравнение, которое было использовано для объяснения экономического роста в различных странах, и обсуждаются проблемы его оценивания. Фактическое оценивание нескольких спецификаций этого уравнения вынесено в эмпирическое упражнение.

Выход оценочного уравнения

В неоклассической теории роста выпуск на эффективную единицу труда в период t , обозначаемый как $q(t)$ и определенный более точно в (5.4.4) ниже, сходится к стационарному уровню q^* . Логлинейная аппроксимация в окрестности стационарного состояния дает

$$\frac{d \log(q(t))}{dt} = \lambda \cdot [\log(q^*) - \log(q(t))]. \quad (5.4.1)$$

Оценивание λ , скорости конвергенции (*speed of convergence*), было целью в недавней литературе по росту. Из (5.4.1) следует, что для любых двух точек t_{m-1} и t_m за бесконечное время

$$\log(q(t_m)) = (1 - \rho) \cdot \log(q^*) + \rho \cdot \log(q(t_{m-1})). \quad (5.4.2)$$

где

$$\rho = \exp[-\lambda \cdot (t_m - t_{m-1})]. \quad (5.4.3)$$

(Поскольку t и m являются i_{it} равномерно распределены по времени, ρ не зависит от m .) Выпуск на единицу эффективного труда определяется как

$$q(t) = \frac{Y(t)}{A(t) L(t)}. \quad (5.4.4)$$

¹ Здесь обширный обзор литературы является [Barro and Sala-i-Martin, 1995].

где $Y(t)$ равно агрегированному выпуску, $L(t)$ равно агрегированным обработанным часам, и $A(t)$ равно уровню трудоемкого технологического прогресса. Предполагая, что $A(t)$ растет с постоянным темпом g (так что $A(t) = A(0)\exp(gt)$), из (5.4.4) следует

$$\log(q^*t) = \log\left(\frac{Y(t)}{L(t)}\right) - \log(A(0)) - g \cdot t. \quad (5.4.5)$$

Подставляя это в (5.4.2), мы получим:

$$\log\left(\frac{Y(t_m)}{L(t_m)}\right) = \rho \cdot \log\left(\frac{Y(t_{m-1})}{L(t_{m-1})}\right) + (1 - \rho) \cdot [\log(q^*) - \log(A(0))] + \phi_m, \quad (5.4.6)$$

где

$$\phi_m \equiv g \cdot (t_m - \rho \cdot t_{m-1}). \quad (5.4.7)$$

Эквивалентную форму уравнения (5.4.6) можно получить, вычитая из обеих его частей $\log \frac{Y(t_{m-1})}{L(t_{m-1})}$:

$$\begin{aligned} \log\left(\frac{Y(t_m)}{L(t_m)}\right) - \log\left(\frac{Y(t_{m-1})}{L(t_{m-1})}\right) = \\ = (\rho - 1) \log\left(\frac{Y(t_{m-1})}{L(t_{m-1})}\right) + (1 - \rho) \cdot [\log(q^*) + \log(A(0))] + \phi_m. \end{aligned} \quad (5.4.6')$$

Поскольку $\rho < 1$ при $\lambda > 0$, это уравнение говорит, что уровень выпуска на душу населения имеет отрицательное влияние на последующий рост. Бедные страны должны расти быстрее. Эта форма уравнения интуитивно более понятна, но для целей выбора корректного метода оценивания более информативным является (5.4.6), а не (5.4.6').

Уравнение (5.4.6) должно выполняться для каждой страны ¹. В оставшейся части этого параграфа мы предполагаем, что скорость конвергенции (λ) и темп роста трудоемкого технологического прогресса (g) одинаковы для разных стран. Тогда (5.4.6) для страны i можно записать как

$$y_{it} = \phi_{it} + \rho y_{i,t-1} + \alpha_i, \quad (5.4.8)$$

где

$$\begin{aligned} y_{it} &= \log\left(\frac{Y(t_{it})}{L(t_{it})}\right) \quad \text{для страны } i, \\ \alpha_i &= (1 - \rho) \times \{\log(q^*) + \log(A(0))\} \quad \text{для страны } i. \end{aligned} \quad (5.4.9)$$

¹Мы получили уравнение, описывающее конвергенцию, предполагая закрытую экономику. Это предположение сложно оправдать, но можно получить аналогичное уравнение при свободной мобильности физического капитала, если человеческий капитал является другим фактором производства и не является свободно мобилизуемым на международном уровне. См., например: [Barro and Sala-i-Martin, 1995].

Добавление ошибок

В прикладной эконометрии довольно стандартной является практика, когда на основе теории выводится уравнение без включения ошибки, а ошибка добавляется затем для оценивания. Эмпирика по росту не является исключением: ошибка η_{it} добавляется в (5.4.8), чтобы получить

$$\Delta \ln y_{it} = \phi_{it} + \eta_{it} = \rho y_{it-1} + \alpha_i + \eta_{it}. \quad (5.4.10)$$

Одной из естественных интерпретаций η_{it} являются бизнес-циклы, которые в макроэкономике определяются как разность между фактическим выпуском и потенциальным выпуском. Теория роста говорит о потенциальном выпуске. Несоответствие между потенциальным и фактическим выпусками становится ошибкой в оцениваемом уравнении, выведенном из теории роста. Но тогда существовала бы проблема ошибок в переменных, которая делает регрессоры эндогенными¹. В оставшейся части параграфа мы следуем майнстриму в литературе по росту и игнорируем проблему эндогенности.

Другой проблемой является возможная корреляция в η_{it} между странами (i). Эта пространственная (или географическая) корреляция не возникла бы, если бы выборка была случайной. Но выборка стран не является случайной; она представляет собой реальную совокупность. Феномен, известный как международные бизнес-циклы, предполагает, что пространственная корреляция должна быть положительной. Статистические выводы, рассматривающие страны как если бы они были независимыми точками данных, завышают статистическую значимость. Эта проблема также по большей части игнорируется в литературе и будет игнорироваться и в нашем обсуждении.

Трактовка α_i

Как ясно из (5.4.9), компонента α_i в уравнении (5.4.10) зависит от стационарного уровня y^* и исходного уровня технологии $A(0)$. Как трактовать α_i в оценивании, зависит от того, какой версии неоклассической теории роста вы придерживаетесь. Согласно модели оптимального роста Касса — Купманса, единственным фактором, определяющим стационарный уровень y^* , является темп роста труда. Таким образом, если технологическое знание свободно перетекает на международном уровне, так что $A(0)$ является одинаковым для разных стран, то международные различия в α_i

¹ Пусть теория роста пытается объяснить η_{it} — логарифм логарифма малого выпуска в год t для страны i , так что (5.4.10) выполняется для y_{it-1} и не для y_{it} . Пусть логарифм фактического выпуска y_{it} соотносится с y_{it-1} как $y_{it} = \lambda y_{it-1} + \eta_{it}$. Тогда получается (5.4.10), но с ошибкой $\eta_{it} = \lambda y_{it-1} + \eta_{it}$. Так что проблему могут создавать не только ошибки в переменной, но также и то, что ошибки имеют структуру скользящего градиента.

полностью объясняются общим ростом труда. Следовательно, оставшаяся в структуре спроса ему удовлетворенности, упомянутую только что, OLS-регрессии (5.4.10) обеспечивает состоятельность, если темпы роста труда входят в качестве дополнительного регрессора для контроля α_2 . Если бы предприниматели, стимули роста Седюу — Сиза, влияли сбережения является другим фактором, определяющих g^* , поэтому сбережения должны включать в структуру сбережений парадокс с тем же ростом труда.

Большая доля современной литературы по росту под рубрикой «условия конвергенция» включает эти и другие переменные, которые могут влиять на g^* или $\Delta(0)$ — такие как мера экономической стабильности и степень финансового посредничества — чтобы контролировать α_2 . В эмпирическом уравнении (5.4.10) формирует сдвиги в g^* , следовательно, скорость конвергенции Δ , используя подход условной конвергенции.

Однако, независимо от того, насколько изобретательно выбраны переменные для контроля α_2 , некоторая компонента в α_2 , имеющая дело с особенностями, уникальными для страны, останется необъясненной и войдет в ошибку. Будучи частью α_2 , она коррелирована с регрессором $g_{i,t-1}$ в (5.4.10). OLS-регрессии, даже если уравнение включает целый ряд переменных для контроля α_2 , не обеспечит состоятельность. Альтернативный подход состоит в том, чтобы трактовать α_2 как ненаблюдаемый фиксированный эффект.

Систематические оценивающие скорости конвергенции

Если выпуск и другие переменные наблюдаются в $M+1$ моменте времени, $t_0, t_1, t_2, \dots, t_M$, то уравнение (5.4.10) доступно для $m = 1, \dots, M$. Метод фиксированных эффектов можно применить к этой системе из M уравнений, но он не даст состоятельности. Причина состоит в том, что эта система является «знаменательской», в том смысле, что некоторые из регрессоров являются зависимыми переменными для других уравнений системы. Чтобы увидеть это, посмотрим на первые два уравнения системы:

$$\begin{aligned} y_{i1} &= \alpha_1 + \alpha_2 g_{i1} + \alpha_3 + \eta_{i1}, \\ y_{i2} &= \alpha_1 + \alpha_2 g_{i2} + \alpha_3 + \eta_{i2}. \end{aligned}$$

Предполагая, что ошибки η_i ортогональны регрессорам (которые суть константа и g_{i1}) в первом уравнении, так что $E(\eta_{i1}) = 0$ и $E(\eta_{i1} g_{i1}) = 0$. Если мы сделаем еще предположение, что $E(\eta_{i1} g_{i2}) = 0$, то из первого уравнения

$$E(g_{i1} \eta_{i1}) = \text{Var}(\eta_{i1}) \neq 0.$$

Но g_{i1} является одним из регрессоров во втором уравнении. Следовательно, если ошибка η_{i1} ортогональна регрессорам для каждого уравне-

ния, «крест-ортogonalности обязательно нарушается». Как подчеркивается в параграфе 5.2 (см. контрольный вопрос 3), если «крест-ортogonalности не выполняются, оценка однокрепленных эффектов не обязательно будет состоятельной. Относительно точного выражения для смещения см. аналитические упражнения 4 в этой главе.

Один из способов достижения состоятельности таков. Возьмем первые разности M уравнений, получая $M - 1$ уравнений:

$$y_{it} - y_{it-1} = \alpha_1 + (y_{it-1} - y_{it-2})\rho + (u_{it} - u_{it-1}),$$

$$y_{it-1} - y_{it-2} = \alpha_2 + (y_{it-2} - y_{it-3})\rho + (u_{it-1} - u_{it-2}),$$

...

$$y_{it-M+1} - y_{it-M} = \alpha_{M-1} + (y_{it-M} - y_{it-M-1})\rho + (u_{it-M+1} - u_{it-M}). \quad (5.4.11)$$

где

$$\rho u_{it} \equiv u_{it+1} - u_{it}.$$

Оценка случайных эффектов и OLS для каждого уравнения те дают состоятельность, поскольку регрессоры не являются ортогональными ошибкам. Например, в первом уравнении разность $y_{it-1} - y_{it-2}$ неортогональна $u_{it} - u_{it-1}$, поскольку $E(y_{it-1}u_{it}) \neq 0$, как только что было видно. Состоятельную оценку можно осуществить посредством многомерного GMM, если есть релевантные инструменты. В дополнительном эмпирическом упражнении к этой главе вы попробуете использовать норму сбережений в качестве инструмента для эндогенных регрессоров. При условии, что норма сбережений не коррелирует с u_{it} , эта оценка является робастной к проблеме эндогенности, упомянутой выше.

Контрольные вопросы

1. Предположим, что g (темпы роста технологического прогресса) одинаковы для всех стран, необходимых для получения (5.4.8). Должны ли мы предполагать, что g является постоянным во времени? (Ответ: Нет.)
2. (Идентификация.) Для простоты пусть $M = 3$, так что (5.4.11) является системой из двух уравнений:
 - (a) Запишите (5.4.11) как систему из нескольких уравнений с общими коэффициентами, надлежащим образом определяя $q_{1t}, q_{2t}, z_{1t}, z_{2t}, \delta$ в (4.6.1). Указание: z_{2t} , например, равно $(0, 1, q_{1t} - q_{1t-1})'$.
 - (b) Если z_{1t} является инструментом для $q_{1t} = y_{it} - y_{it-1}$, вектором инструментов для m -го уравнения является $z_{1tm} = (1, y_{itm})'$ для $m = 1, 2$. Покажите, что система идентифицируема, если $\text{Cov}(z_{1tm}, y_{itm} - q_{1tm-1}) \neq 0$.

¹Это было впервые отмечено в [Nickell, 1981].

или трех мн. Указание:

$$\Sigma = \begin{bmatrix} 1 & 0 & E(y_1 - y_0) \\ E(y_1) & 0 & E(y_1(y_1 - y_0)) \\ 0 & 1 & E(y_1 - y_0) \\ 0 & E(y_2) & E(y_2(y_2 - y_0)) \end{bmatrix}.$$

Приложение 5.А. Распределение статистики Хаусмана

В этом приложении доказывается, что $\Delta_{\text{HLM}}(\hat{q})$ в (5.2.21) является положительно определенной матрицей и статистика Хаусмана (5.2.22) гарантированно неотрицательна в любых конечных выборках.

Как показано в параграфе 4.6,

$$\hat{\beta}_{\text{HLM}} = (S'_{xx} \hat{S}^{-1} S_{xx})^{-1} S'_{xz} \hat{S}^{-1} s_{xy}, \quad (5.A.1)$$

где

$$\begin{aligned} S_{xx} &= \frac{1}{n} \sum_{i=1}^n Z_i' \otimes x_i, & \hat{S} &= \hat{\Sigma} \otimes S_{xx}, \\ S_{xy} &= \frac{1}{n} \sum_{i=1}^n x_i x_i', & s_{xy} &= \frac{1}{n} \sum_{i=1}^n y_i \otimes x_i, \\ K &\equiv \frac{1}{n} \sum_{i=1}^n z_i z_i', & I &\equiv \frac{1}{n} \sum_{i=1}^n z_i z_i' \quad (\text{так что } Z_i \text{ имеет размер } M \times L). \end{aligned}$$

Пусть

$$Z_i = (F_i' 1 b_i'), \quad p \equiv \# \beta, \quad D = \begin{bmatrix} I_p \\ 0 \end{bmatrix},$$

так что $F_i = Z_i D$. Тогда можно довольно легко показать, что

$$\hat{\beta}_{\text{HLM}} = (D' S'_{xx} \tilde{W} S_{xx} D)^{-1} D' S'_{xx} \tilde{W} s_{xy}, \quad (5.A.2)$$

где

$$\tilde{W} = Q \otimes S_{xx}^{-1}, \quad Q \equiv \text{аннулятор для } 1_{Lp}.$$

Из (5.A.1) и (5.A.2) следует, что

$$\hat{q} = (D' S'_{xx} \tilde{W} S_{xx} D)^{-1} D' S'_{xx} \tilde{W} g_n(\hat{\beta}_{\text{HLM}}), \quad g_n(\hat{\beta}_{\text{HLM}}) = s_{xy} - S_{xx} \hat{\beta}_{\text{HLM}}. \quad (5.A.3)$$

(Чтобы вывести это, вспомним из параграфа 4.6, что

$$S'_{xx} \tilde{W}^{-1} S_{xx} = \frac{1}{n} \sum_{i=1}^n Z_i' Q Z_i.$$

Но поскольку $Q Z_i = (Q F_i' 0)$, это равно

$$\begin{bmatrix} \frac{1}{n} \sum_{i=1}^n F_i' Q F_i & 0 \\ 0 & 0 \end{bmatrix}, \quad (5.A.4)$$

так что $S'_{zz} \tilde{W} S_{zz} = D D' S'_{zz} \tilde{W} S_{zz} = S'_{zz} \tilde{W} S_{zz} D D'$.

Как показано в аналитическом упражнении 5.1, можно:

$$g_n(\tilde{\delta}_{RE}) = \tilde{B} \tilde{g}, \quad \tilde{B} = I_{qK} - S_{zz}(S'_{zz} \tilde{S}^{-1} S_{zz})^{-1} S'_{zz} \tilde{S}^{-1}. \quad (5.A.5)$$

Подставляя это в (5.A.3), мы получаем:

$$\sqrt{n} \hat{q} = (D' S'_{zz} \tilde{W} S_{zz} D)^{-1} D' S'_{zz} \tilde{W} \tilde{B} \sqrt{n} \tilde{g}. \quad (5.A.6)$$

Асимптотическая дисперсия этого выражения равна:

$$\begin{aligned} \text{Avar}(\hat{q}) &= (D' S'_{zz} \tilde{W} S_{zz} D)^{-1} \times \\ &\times D' S'_{zz} \tilde{W} B S B' \tilde{W} S_{zz} D (D' S'_{zz} \tilde{W} S_{zz} D)^{-1}, \end{aligned} \quad (5.A.7)$$

где $B = \text{plim } \tilde{B}$, $\tilde{W} = \text{plim } \tilde{W} = Q \otimes \Sigma_{zz}^{-1}$ и $S \equiv \text{Avar}(\tilde{g})$. Здесь $D' S'_{zz} \tilde{W} S_{zz} D$ обратна, поскольку легко показать, что

$$D' S'_{zz} \tilde{W} S_{zz} D = E(F_i' Q F_i), \quad (5.A.8)$$

которая, как доказано в аналитическом упражнении 5, невырождена. (5.A.7) состоятельно оценивается посредством

$$\begin{aligned} \widehat{\text{Avar}}(\hat{q}) &= (D' S'_{zz} \tilde{W} S_{zz} D)^{-1} \times \\ &\times D' S'_{zz} \tilde{W} \hat{B} \hat{S} \hat{B}' \tilde{W} S_{zz} D (D' S'_{zz} \tilde{W} S_{zz} D)^{-1}. \end{aligned} \quad (5.A.9)$$

Простые, но длительные матричные вычисления показывают, что (5.A.7) равно $\text{Avar}(\tilde{\delta}_{RE}) = D' \text{Avar}(\tilde{\delta}_{RE}) D$, а (5.A.9) = $\widehat{\text{Avar}}(\tilde{\delta}_{RE}) = D' \widehat{\text{Avar}}(\tilde{\delta}_{RE}) D$. В любой конечной выборке (5.A.9) положительно полуопределена. Поэтому статистика Хаусмана гарантированно неотрицательна.

Теперь мы покажем, что матрица (5.A.7) невырождена. В соответствии с (5.1.6), S положительно определена. Следовательно, ранг $\text{Avar}(\hat{q})$ равен рангу

$$B'(Q \otimes \Sigma_{zz}^{-1}) S_{zz} D. \quad (5.A.10)$$

Поскольку столбцы $S^{-1} S_{zz}$ формируют базис для столбцового нуль-пространства B и поскольку $(Q \otimes \Sigma_{zz}^{-1}) S_{zz} D$ имеет полный столбцовый ранг по (5.1.15), по лемме A.5 в [Newey, 1985]¹, ранг (5.A.10) равен:

$$\text{rank}[(\Sigma^{-1} \otimes \Sigma_{zz}^{-1}) S_{zz} : (Q \otimes \Sigma_{zz}^{-1}) S_{zz} D] = L. \quad (5.A.11)$$

где был использован тот факт, что $S = \Sigma \otimes \Sigma_{zz}$.

¹Эта лемма утверждает следующее. Пусть $A = k \times l$ матрица, $B = l \times m$ матрица и $C = l \times n$ матрица. Если столбцы C формируют базис столбцового нуль-пространства матрицы A и $\text{rank}(B) = m$, то $\text{rank}(AB) = \text{rank}(C'B) = m$. Здесь $k = l + MK$, $m = p$ и $n = L$.

Далее, $\mathbf{x}_i$ содержит в себе все уникальные переменные в Z_i . Сделаем перестановку в $\mathbf{x}_i$:

$$\mathbf{x}_i = (f'_1, \dots, f'_{M'}, b'_1)' \quad (5.A.12)$$

Тогда

$$\mathbf{z}'_{im} = \mathbf{x}'_i \mathbf{H}_{im}, \quad \mathbf{H}_{im} \equiv \begin{bmatrix} e_{im} \otimes \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_L - e_{im} \end{bmatrix}, \quad e_{im} = m\text{-й столбец в } \mathbf{I}_{M'}, \quad (5.A.13)$$

так что Σ_{zz} можно записать как

$$\Sigma_{zz} = (\mathbf{I}_M \otimes \Sigma_{zz}) \mathbf{H}, \quad (5.A.14)$$

где $\mathbf{H}' = (\mathbf{H}'_1, \dots, \mathbf{H}'_M)$. Подставив (5.A.14) в (5.A.11), находим, что $\text{rang}(\text{Avar}(\hat{\beta}))$ равен

$$\text{rang}[(\Sigma^{-1} \otimes \mathbf{I}_K) \mathbf{H} : (Q \otimes \mathbf{I}_K) \mathbf{H} \mathbf{D}] = L,$$

где K — строк $(MK \times (L+p))$ матрицы $[(\Sigma^{-1} \otimes \mathbf{I}_K) \mathbf{H} : (Q \otimes \mathbf{I}_K) \mathbf{H} \mathbf{D}]$ имеют вид:

$$\begin{bmatrix} \begin{pmatrix} \sigma^{m1} \\ \vdots \\ \sigma^{mM'} \end{pmatrix} \otimes \mathbf{I}_p & \mathbf{0} & \begin{pmatrix} q_{m1} \\ \vdots \\ q_{mM'} \end{pmatrix} \otimes \mathbf{I}_p \\ \mathbf{0} & (\sum_{i=1}^{M'} \sigma^{mi}) \mathbf{I}_{L-p} & \mathbf{0} \end{bmatrix}. \quad (5.A.15)$$

Следовательно, $\text{rang}[(\Sigma^{-1} \otimes \mathbf{I}_K) \mathbf{H} : (Q \otimes \mathbf{I}_K) \mathbf{H} \mathbf{D}] = L+p$, если $\text{rang}(\Sigma^{-1})$ и $\text{rang}(Q)$ не являются линейно зависимыми, то есть, если Σ^{-1} не пропорциональна Q . Но поскольку $\text{rang}(\Sigma^{-1}) = M \neq \text{rang}(Q) = M-1$, Σ^{-1} не может быть пропорциональна Q .

Набор задач для главы 5

Аналитические упражнения

1. (Оценка фиксированных эффектов как LSDV-оценки) В (5.1.1^o) на с. 358 переопределим α_i как сумму старого α_i и δ_i/γ , так что (5.1.1^o) примет вид:

$$\mathbf{y}_i = \mathbf{F}_i \beta + \mathbf{I}_M \cdot \alpha_i + \boldsymbol{\eta}_i \quad (i = 1, 2, \dots, n). \quad (1)$$

Определим следующие стандартизованные векторы и матрицы:

$$\underset{(nM \times 1)}{\mathbf{y}} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}, \quad \underset{(nM \times p)}{\mathbf{F}} = \begin{bmatrix} \mathbf{F}_1 \\ \vdots \\ \mathbf{F}_n \end{bmatrix}, \quad \underset{(nM \times 1)}{\boldsymbol{\eta}} = \begin{bmatrix} \boldsymbol{\eta}_1 \\ \vdots \\ \boldsymbol{\eta}_n \end{bmatrix}, \quad \underset{(nM \times 1)}{\boldsymbol{\alpha}} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}.$$

$$W_{(Mn \times (n + \# \beta))} = (D : F), \quad \theta_{((n + \# \beta) \times 1)} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}.$$

$$D_{(Mn \times n)} = I_n \otimes \mathbf{1}_M = \begin{bmatrix} \mathbf{1}_M & & & \\ & \mathbf{1}_M & & \\ & & \ddots & \\ & & & \mathbf{1}_M \end{bmatrix}, \quad (2)$$

так что систему (1) из M уравнений для $i = 1, 2, \dots, n$ можно записать как

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \mathbf{1}_M \cdot \alpha_1 \\ \mathbf{1}_M \cdot \alpha_2 \\ \vdots \\ \mathbf{1}_M \cdot \alpha_n \end{bmatrix} + \begin{bmatrix} F_1 \beta \\ F_2 \beta \\ \vdots \\ F_n \beta \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix},$$

или

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \mathbf{1}_M & & & \\ & \mathbf{1}_M & & \\ & & \ddots & \\ & & & \mathbf{1}_M \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{bmatrix} \beta + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix}.$$

или

$$y = D\alpha + F\beta + \eta = W\theta + \eta. \quad (3)$$

OLS-оценка для θ в объединенной выборке объема Mn , (y, W) , равна $(W'W)^{-1}W'y$.

- (а) Покажите, что оценка фиксированных эффектов, заданная в (5.2.4), представляет последние $\# \beta$ элементов этого вектора. **Указание:** Используйте теорему Фриша — Во в регрессии с блочной структурой данных (см. аналитическое упражнение 4(b) в главе 1). На первом шаге проводится регрессия y на D и проводится регрессия каждой столбца F на D . Если мы определим аннулятор для D как

$$M_D = I_n \otimes \mathbf{1}_M = D(D'D)^{-1}D', \quad (4)$$

то векторы остатков можно записать как $M_D y$ и $M_D F$. На втором шаге проводится регрессия $M_D y$ на $M_D F$. Покажите, что оцененный регрессионный коэффициент можно записать как

$$(F'M_D F)^{-1}F'M_D y. \quad (5)$$

Покажите, что $M_D = I_n \otimes Q$ (где Q — аннулятор для $\mathbf{1}_M$, как в тексте), и используйте этот факт, чтобы показать, что (5) действительно является оценкой фиксированных эффектов.

Альтернативно вы можете пройти весь путь обратно к нормальным уравнениям. Если a и b являются OLS-оценками α и β , то (a, b) удовлетворяет системе из $n + \# \beta$ линейных одновременных уравнений:

$$\begin{bmatrix} D'D & D'F \\ F'D & F'F \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} D'y \\ F'y \end{bmatrix}. \quad (6)$$

Возьмите первые n уравнений и решите их относительно a , считая b заданным. Это должно привести к $a = (D'D)^{-1}(D'y - D'Fb)$. Затем подставляйте это выражение в оставшиеся уравнения и решите их относительно b . (Вы можете выполнить то же самое, используя формулу для обращения блочных матриц.)

(b) Покажите:

$$a_i = \bar{y}_i - \left(\frac{1}{n} \sum_{t=1}^n f_{t,i} \right) \tilde{\beta}_1,$$

где a_i является OLS-оценкой для α_i , $f_{t,i}$ является t -й строкой в F_i и $\bar{y}_i = \frac{1}{n} \sum_{t=1}^n (y_{t1} + \dots + y_{tM})$. **Указание:** Используйте нормальные уравнения для α .

(c) Предположим:

(i) $\{y_t, F_t\}$ является i.i.d.,

(ii) $E(\eta_t | F_t) = 0$ (что является более сильным, чем условия ортогональности (5.1 8b)),

(iii) $E(\eta_t \eta_t' | F_t) = \sigma_\eta^2 I_M$ (предположение о сферичности ошибок) и

(iv) W имеет полный столбцовый ранг.

Покажите, что предположения классической модели регрессии выполняются для объединенной регрессии (3). **Указание:** Вы должны показать, что W является строго экзогенной в (3), то есть

$$E(\eta | W) = 0.$$

Также покажите, что $E(\eta \eta' | W) = \sigma_\eta^2 I_{nM}$.

Разработайте для a и b теорию конечных выборок. Например, являются ли они несмещенными? Пусть SSR является суммой квадратов остатков в объединенной регрессии (3). Является ли эта SSR той же самой, что и SSR в регрессии «внутри» (5.2 15)?

2. (GMM-интерпретация FE.) Хорошо известно, что для случая двух уравнений ($M = 2$) оценка фиксированных эффектов является OLS-оценкой в регрессии первых разностей $y_{2t} - y_{1t}$ на $z_{2t} - z_{1t}$. Этот вопрос обобщает доказательство на случай произвольного M .

Мы хотим записать оценку фиксированных эффектов (5.24) в OMM-формате:

$$(S'_{xx} \tilde{W} S_{xx})^{-1} S'_{xx} \tilde{W} y_{xy} \quad (7)$$

для надлежащим образом определенных S_{xx} , $\tilde{W}$ и y_{xy} . Пусть C — такая матрица, что

C является $M \times (M - 1)$ матрицей полного столбцового ранга.

$$\text{удовлетворяющей } C'1_M = 0. \quad (8)$$

Известным примером такой матрицы C является матрица первых $M - 1$ столбцов

$$C_{M \times (M-1)} = \begin{bmatrix} -1 & 0 & 0 & 0 & \dots & 0 \\ 1 & -1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 \\ & & & \ddots & \ddots & \\ 0 & \dots & 0 & 1 & -1 & 0 \\ 0 & \dots & 0 & 0 & 1 & 1 \\ 0 & \dots & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Другим примером является $M \times (M - 1)$ матрица, построенная удалением одного столбца (а, скажем, последнего) из матрицы Q , ангулятора для 1_M .

(a) Проверьте, что эти две матрицы удовлетворяют (8).

(b) Для любого выбранного C проверьте, что вы можете получить следующую систему $M - 1$ преобразованных уравнений, умножив обе части системы M уравнений (5.11') (на с. 358) на C' слева:

$$\tilde{y}_i = \tilde{F}_i \beta + \tilde{\epsilon}_i, \quad (9)$$

где

$$\tilde{y}_i = C' y_i, \quad \tilde{F}_i = C' F_i, \quad \tilde{\epsilon}_i = C' \epsilon_i, \quad (i = 1, \dots, M)$$

(c) (Минимальность.) Проверьте следующее:

Если $\{\hat{y}_i, \hat{Z}_i\}$ удовлетворяет предположениям модели компонент ошибки (которые являются (5.11'), (5.1.2), (5.1.8), (5.1.4), (5.1.5), (5.1.15) и (5.1.6) или (5.1.6') на с. 353), то $\{\hat{y}_i, \hat{F}_i\}$ удовлетворяет предположениям модели случайных эффектов той же самой системы (9) из $M - 1$ уравнений.

(случайная выборка) $\{\hat{y}_i, \hat{F}_i\}$ является итд.

(условия ортогональности) $E(\hat{F}_i \otimes \epsilon_i) = 0$ с ϵ_i = объединенные

$\{\epsilon_{i1}, \dots, \epsilon_{iM}\}$ = уникальные (различные) элементы $\{F_i, b_i\}$.

- (идентифицируемость) $E(\tilde{F}_1 \otimes x_1)$ имеет полный столбцовый ранг,
 (условная homoskedастичность) $E(\tilde{\eta}_1 \tilde{\eta}_1' | x_1) = E(\tilde{\eta}_1 \tilde{\eta}_1')$, последняя матрица невырождена,
 (невырожденность $E(\tilde{g}_1 \tilde{g}_1')$) $E(\tilde{g}_1 \tilde{g}_1')$ невырождена, $\tilde{\eta}_1 \equiv \tilde{\eta}_1 \otimes x_1$.

Указание: Для условий ортогональности заметим, что (5.1.8b) можно записать как $E(\eta_1 \otimes x_1) = 0$ и что $\tilde{\eta}_1 \otimes x_1 = (C' \otimes I_K)(\eta_1 \otimes x_1)$. Условие идентификации эквивалентно (5.1.15) (относительно доказательства см. ответы на вопросы). Чтобы показать, что $E(\tilde{\eta}_1 \tilde{\eta}_1' | x_1)$ не зависит от x_1 , используйте тот факт, что $\tilde{\eta}_1 = C' \eta_1 = C' e_1$, поскольку $e_1 = 1_M \otimes a_1 + \eta_1$. При заданной условной homoskedастичности мы имеем

$$E(\tilde{g}_1 \tilde{g}_1') = E(\tilde{\eta}_1 \tilde{\eta}_1') \otimes E(x_1 x_1').$$

Поэтому последнее условие (невырожденность $E(\tilde{g}_1 \tilde{g}_1')$) является просто условием того, что $E(x_1 x_1')$ невырождена.

- (d) Покажите, что оценку фиксированных эффектов (5.2.4) можно записать в виде GMM-оценки (7) с

$$S_{x\eta} = \frac{1}{n} \sum_{i=1}^n \tilde{F}_i \otimes x_i, \quad S_{\eta\eta} = \frac{1}{n} \sum_{i=1}^n \tilde{\eta}_i \otimes x_i, \quad (10)$$

$$\tilde{W} = (C' C)^{-1} \otimes \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1}.$$

Указание:

Используйте следующий результат из матричной алгебры: Если C удовлетворяет условию (8), то

$$C(C' C)^{-1} C' = Q \quad (11)$$

(аннулятор для 1_K). Элементы f_{ab} являются подмножествами элементов x_i . Проинтегрируйте «исключившиеся» x_i . Поэтому, например,

$$\frac{1}{n} \sum_{i=1}^n f_{ab} x_i \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n x_i f_{cd} = \frac{1}{n} \sum_{i=1}^n f_{ab} f_{cd}.$$

- (e) Пусть $\hat{\Phi}$ $((M-1) \times (M-1))$ является собственной оценкой $\Phi = E(\tilde{\eta}_1 \tilde{\eta}_1')$. Покажите, что эффектная GMM-оценка задается как

$$\hat{\beta} = \left(\frac{1}{n} \sum_{i=1}^n \tilde{F}_i \hat{\Phi}^{-1} \tilde{F}_i' \right)^{-1} \tilde{F}_i' \hat{\Phi}^{-1} \hat{g}_i. \quad (12)$$

Покажите, что

$$\text{Avar}(\hat{\beta}) = \left[E(\hat{F}'_i \hat{\Psi}^{-1} \hat{F}_i) \right]^{-1}. \quad (13)$$

Указание: Примените утверждение 4.7, чтобы представить систему из $M - 1$ уравнений.

- (f) Покажите, что следующая матрица является состоятельной оценкой для Ψ :

$$[(M-1) \times (M-1)] \quad \hat{\Psi} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \hat{F}_i \hat{\beta}_{FE}) (\hat{y}_i - \hat{F}_i \hat{\beta}_{FE})' \quad (14)$$

- (g) (Статистика Саргана.) Выведите статистику Саргана, связанную с эффективной GMM-оценкой. **Указание:** В утверждении 4.7(c) положите

$$\tilde{S} = \hat{\Psi} \odot \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right), \quad g_n(\hat{\beta}) = s_{xy} - S_{xz} \hat{\beta}.$$

- (h) Предположим, что η_i является сферичным и что

$$E(\eta_i \eta_i') = \sigma_\eta^2 I_M. \quad (15)$$

Проверьте, что если $\hat{\beta}_n^2$ является состоятельной оценкой σ_η^2 , то

$$\hat{\Psi} = \hat{\beta}_n^2 C' C \quad (16)$$

является состоятельной для Ψ . Проверьте, что эффективная GMM-оценка (12) с таким выбором $\hat{\Psi}$ численно равна оценке обескоррелированных эффектов.

- (i) (Инвариантность выбора C .) Проверьте: если $M \times (M-1)$ матрица C удовлетворяет (8), то выполняется $B = CA$, где A является невырожденной матрицей $(M-1) \times (M-1)$. Замените всюду в этом вопросе C на B и проверьте, что выбор C не изменяет никакие результаты этого вопроса.

- 3 Пусть SSR — сумма квадратов остатков, определенная в (5.2.15). Для модели компонент ошибки с предположением сферичности $E(\eta_i \eta_i') = \sigma_\eta^2 I_M$ докажите, что

$$\lim_{n \rightarrow \infty} \frac{SSR}{n} = (M-1) \cdot \sigma_\eta^2.$$

Указанне:

$$\begin{aligned}
 SSR &= \sum_{i=1}^n |Q(y_i - F_i \hat{\beta}_{F,D})'| Q(y_i - F_i \hat{\beta}_{F,D})| = \\
 &= \sum_{i=1}^n (y_i - F_i \hat{\beta}_{F,D})' Q(y_i - F_i \hat{\beta}_{F,D}) \\
 &\quad (\text{поскольку } Q \text{ симметричная и идемпотентная}) \\
 &= \sum_{i=1}^n (y_i - F_i \hat{\beta}_{F,D})' C(C'C)^{-1} C'(y_i - F_i \hat{\beta}_{F,D}) = \quad (\text{кз (11)}) \\
 &= \sum_{i=1}^n \tilde{y}_i' (C'C)^{-1} \tilde{y}_i \quad (\tilde{y}_i \equiv C'y_i - C'F_i \hat{\beta}_{F,D}) = \\
 &= \sum_{i=1}^n \text{trace}_1[\tilde{y}_i' (C'C)^{-1} \tilde{y}_i] \\
 &= \sum_{i=1}^n \text{trace}_1[(C'C)^{-1} \tilde{y}_i \tilde{y}_i'] = \\
 &= n \cdot \text{trace}_1[(C'C)^{-1}] \frac{1}{n} \sum_{i=1}^n \tilde{y}_i \tilde{y}_i' \\
 &\quad (\text{поскольку оператор следов линейный}).
 \end{aligned}$$

С обычным регулярным условием $(E(f_{im} f_{ik}'))$ существует и, moreover, для всех m, k , что получается, если $E(x_i x_i')$ существует и конечно; теперь можно стандартным образом доказать, что

$$\frac{1}{n} \sum_{i=1}^n \tilde{y}_i \tilde{y}_i' \xrightarrow{p} E(\tilde{y}_i \tilde{y}_i') \quad \text{при } n \rightarrow \infty,$$

где $\tilde{y}_i = C'y_i - C'F_i \beta$. Покажите, что $E(\tilde{y}_i \tilde{y}_i') = \sigma_u^2 \cdot C'C$.

4. (Независимость возник фиксированных эффектов для динамических моделей) Рассмотрим следующие исходные данные

$$y_{it} = \alpha_i + \rho \cdot y_{i,t-1} + \eta_{it} \quad (\rho = -1, 0, \dots, 1; i = 1, 2, \dots, N),$$

Предположим, что y_{it} есть случайная выборка по $(y_{1t}, y_{2t}, \dots, y_{Nt})$. Мы предполагаем, что $E(\eta_{it} y_{i,t-1}) = 0$ и $E(y_{i,t-1} y_{i,t-1}) = 0$ для всех i . Также предположим, что не существует кросс-секционной корреляции: $E(\eta_{it} \eta_{js}) = 0$ для $i \neq j$ и что второй момент единичный для всех уравнений: $E(\eta_{it}^2) = \sigma_{\eta}^2$.

- (а) Запишите систему M уравнений как (5.1.1^а) на с. 358, надлежащим образом определяя y_t , F_t и b_t .

(b) Покажите $E(\eta_{i,m} \cdot \eta_{j,h}) = 0$, если $m < h$. **Указание:**

$$\eta_{i,m} = \eta_{i,m} + \rho \eta_{i,m-1} + \dots + \rho^{m-1} \eta_{i,1} + \frac{1 - \rho^m}{1 - \rho} \varepsilon_i + \rho^m \eta_{i,0}.$$

(c) Проверьте, что $E(\eta_{i,m} \cdot \eta_i^2) = \rho^{m-1} \sigma_\varepsilon^2$ для $m \geq h$. (Дополнительно.) Покажите, что

$$E(F_i' Q \eta_i) = -\frac{\sigma_\varepsilon^2 (M-1) - M\rho - \rho^M}{M(1-\rho^2)}.$$

(d) Следовательно, если $E(F_i' Q F_i)$ невырождена и если $(M-1) - M\rho + \rho^M \neq 0$, то оценки фиксированных эффектов для этой динамической модели несмещаемы. Какие предположение в утверждении 5.1. врангующие состоятельность оценки фиксированных эффектов, нарушены?

5. (Неисключительно, идентифицирующая «группировка» невырожденность.)

(a) Покажите, что матрица $E(\tilde{F}_i' \tilde{F}_i)$ в (5.2.7) обратима при предположении (5.1.15) с тем, что $E(QF_i \otimes x_i)$ имеет полный столбцовый ранг. **Указание:**

$$E(\tilde{F}_i' \tilde{F}_i) = E(F_i' Q F_i) = \sum_{m=1}^M \sum_{h=1}^M q_{mh} E(f_{im} f'_{ih}),$$

где $q_{mh} = (p, h)$ -й элемент Q . Покажите, что

$$E(f_{im} f'_{ih}) = E(f_{im} x_i') [E(x_i x_i')]^{-1} E(x_i f'_{ih}).$$

(Положите, что элементы f_{im} включены в x_i .) Наконец, покажите, что $x_i f'_{im} = f'_{im} \otimes x_i$, $E(f_i \otimes x_i') = (f_1 \otimes x_1', \dots, f_M \otimes x_M')$ и

$$\begin{aligned} E(\tilde{F}_i' \tilde{F}_i) &= E(F_i \otimes x_i)' [Q \otimes I(x_i x_i')]^{-1} E(F_i \otimes x_i) = \\ &= E(QF_i \otimes x_i) (I_M \otimes [E(x_i x_i')]^{-1}) E(QF_i \otimes x_i). \end{aligned}$$

(b) Используйте аналогичные аргументы, чтобы доказать, что матрица $E(\tilde{F}_i' E(\tilde{\eta}_i \tilde{\eta}_i') \tilde{F}_i)$ в (5.2.8) является невырожденной. **Указание:** Поскольку

$$\tilde{\eta}_i = Q \eta_i = Q \varepsilon_i,$$

мы имеем

$$\tilde{F}_i' E(\tilde{\eta}_i \tilde{\eta}_i') \tilde{F}_i = \tilde{F}_i' Q (E(\varepsilon_i \varepsilon_i')) Q \tilde{F}_i = \tilde{F}_i' E(\varepsilon_i \varepsilon_i') \tilde{F}_i.$$

Также

$$E[\tilde{F}_i' E(\varepsilon_i \varepsilon_i') \tilde{F}_i] = E(QF_i \otimes x_i)' [E(\varepsilon_i \varepsilon_i') \otimes [E(x_i x_i')]^{-1}] E(QF_i \otimes x_i).$$

(с) Покажите, что $E(\tilde{F}_1^* \Phi^{-1} \tilde{F}_1)$ в (13) — аналитическое уравнение 2-й степени невырожденной. **Указание:** Как показано в аналитическом уравнении 2(с), $E(Q F_1 \otimes x_1)$ имеет полный степенной разложения и только тогда, когда $E(C^* F_1 \otimes x_1)$ имеет полный степенной разложение.

6. (Дополнительно, альтернативное условие идентификации.) Рассмотрим модель компонент опенки (5.1.1'') на с. 358. Пусть $p \equiv \# \theta$. Если A_m является матрицей, которая вычитывает f_{1m} из x_1 , ее можно записать как

$$A_m = \begin{bmatrix} e_m \otimes I_p \\ 0 \end{bmatrix},$$

где e_m является m -м столбцом I_p .

(а) Пусть $A = (A_1 \dots A_p)$ является матрицей, полученной стабилизированием $(A_1, \dots, A_p)$. Покажите, что $F_1 = (I_M \otimes x_1') A$.

(б) Покажите, что ранговое условие (5.1.5), объясненное с условием, что $L(x; x')$ является невырожденной (что вытекает из (5.1.5) и (5.1.6)), можно записать как

$$\text{rank}[Q \otimes I_K] A = p \quad (\equiv \# \theta).$$

(Этот результат будет использоваться в следующем упражнении.)
Указание:

$$\begin{aligned} F_1 \otimes x_1 &= [(I_M \otimes x_1') A] \otimes x_1 = \\ &= [(I_M \otimes x_1' \otimes x_1) (A \otimes I)] \\ &= (I_M \otimes x_1' \otimes x_1) A = \\ &= [I_M \otimes (x_1 x_1')] A. \end{aligned}$$

Поэтому

$$\begin{aligned} \tilde{F}_1 \otimes x_1 &= (Q \otimes I_K) (F_1 \otimes x_1) = \\ &= [Q \otimes (x_1 x_1')] A = \\ &= (I_M \otimes x_1 x_1') (Q \otimes I_K) A. \end{aligned}$$

Поэтому $L(\tilde{F}_1 \otimes x_1) = [I_M \otimes E(x_1 x_1')] (Q \otimes I_K) A$.

7. (Дополнительно, проверка невырожденности.) Докажите, что матрица $E(\tilde{F}_1^* \tilde{\eta}_1 \tilde{\eta}_1' \tilde{F}_1)$ в (5.2.23) является невырожденной, выполнив следующие шаги:

- (а) Запишите $\tilde{F}'_i q_i$ как линейную функцию от $g_i \equiv e_i \otimes x_i$. Указание: Пусть A определено, как в предыдущем упражнении.

$$\begin{aligned}\tilde{F}'_i q_i &= A'(I_M \otimes x_i)Qq_i = \\ &= A'(I_M \otimes x_i)Qe_i = \quad (\text{поскольку } e_i = I_M \cdot e_i \otimes 1) = \\ &= A'(I_M \otimes x_i)(Qe_i \otimes 1) = \\ &= A'(Qx_i \otimes x_i) = \\ &= A'(Q \otimes I_N)(e_i \otimes x_i) = \\ &= A'(Q \otimes I_N)g_i.\end{aligned}$$

- (б) Докажите желаемый результат. Указание: $E(g_i g_i')$ положительно определена по 15.1.6).

Эмпирические упражнения

В этом упражнении вы оцените скорость конвергенции, используя несколько различных методов оценивания. Мы используем международную панель, построенную в [Summers and Heston, 1991], которая стала стандартным набором данных для изучения развития стран. Панель Саммерса — Хестона (также называемая Penn World Table) включает основные макроэкономические переменные, измеренные на постоянной основе практически для всех стран мира.

Файл SUM_HES.ASC извлечен из Penn World Table версии 5.6 за 1960–1985 г. данными странам NBER (www.nber.com). Файл организован следующим образом. Для каждой страны информация по стране содержится в нескольких записях (строках). Первая запись содержит идентификационный код страны и две даммы, одна — для коммунистического режима (называемая как *COM* ниже) и другая — для организованных стран — экспортеров нефти (называемая как *OPEC* ниже). Вторая запись содержит год, население (в тысячах), реальный ВВП на душу населения (в долларах США за 1985 год) и норму сбережения (в процентах) для 1960 года. Третья запись содержит то же самое за 1961 год и так далее. Двадцать седьмая запись для страны прекращается в 1985 году. Перечень нескольких записей файла вы видите как

1 0 0			
1960	10800	1723	19.9
1961	11016	1599	21.1
1964	21173	2962	26.4
1965	21548	2986	26.0
2 0 0			
1960	4816	931	3.3
1961	4918	976	2.9

Файл охватывает все 125 стран, для которых доступны данные. (Сопоставление между ID страны и названием страны содержится в `country.d5c`. Страна 1 — это Анжир, страна 2 — Албения и т.д. Заметим, что сопоставление в первом 56 Penn World Table отличается от [Summers and Heston, 1991].) GDP на душу населения — это реальный ВВП страны, пересчитанный в доллары США на основе паритета покупательной способности валюты страны, против 1 января 1985 год (переменная «RGDPCHI» в Penn World Table). Парка сбережений измеряется как отношение реальных инвестиций к реальному ВВП (`inv` в Penn World Table). Дополнительно того, как построены переменные, см. параграфы II и III (в частности, III.C) в [Summers and Heston, 1991].

Первым вопросом, который мы изучаем эмпирически, является условная конвергенция. (На этом этапе было бы полезно было просмотреть работу [Mankiw, Romer, and Weil, 1992].) Пусть s_t является нормой сбережения страны i , а n_t является темпом роста населения страны (который мы будем считать темпом роста трудовых ресурсов). В модели роста Солоу — Сванга с производственной функцией Кобба — Дугласа равновесный уровень выпуска на эффективную единицу труда, обозначенный в основном тексте как q^* , является линейной функцией $\log(n_t) = \log(n_0 + g + \delta)$, где g — темп технологического прогресса, а δ — норма амортизации основного капитала. Тогда, предполагая что $A(0)$, начальный уровень технологии, одинаков для всех стран, уравнение (5.4.10) можно записать как

$$y_{it,t} = \phi_{it} + \rho y_{it,t-1} + \gamma_1 \log(s_t) + \gamma_2 \log(n_t - g + \delta) + \eta_{it}. \quad (1)$$

Из модели следует, что $\gamma_1 = -\gamma_2$, но мы не будем накладывать это условие. Это является стандартной идеей, описанной в табл. IV в [Mankiw et al., 1992]. Как и в [Mankiw et al.], предположим, что $g + \delta$ равно 5% для всех стран, и возьмем s_t как норму сбережения, усредненную за 1960–1985 годы. Мы изберем n_t как среднесловный темп роста населения за тот же период ([Mankiw et al.] используют средний темп роста населения в трудоспособном возрасте). Поскольку наша выборка из 125 стран включает (былише) коммунистические страны и ОПЕС, мы добавляем `COM` (= 1, если коммунистическая, 0 — в противном случае) и `OPEC` (= 1, если страна в ОПЕС, 0 — в противном случае) в уравнение:

$$y_{it,t} = \phi_{it} + \rho y_{it,t-1} + \gamma_1 \log(s_t) + \gamma_2 \log(n_t - g + \delta) + \gamma_3 COM_t + \gamma_4 OPEC_t + \eta_{it}. \quad (2)$$

(а) Вычитая $y_{it,t-1}$ из обеих частей, мы можем переписать (2) как

$$y_{it,t} - y_{it,t-1} = \phi_{it} - (1 - \rho)y_{it,t-1} + \gamma_1 \log(s_t) + \gamma_2 \log(n_t - g + \delta) + \gamma_3 COM_t + \gamma_4 OPEC_t + \eta_{it}. \quad (2')$$

Для этой спецификации положим $M = 1$ (только одно уравнение для оценивания) и возьмем $t_0 = 1960$ и $t_1 = 1985$, так что зависимая переменная $y_{it1} - y_{it0}$ является взвешенным ростом между 1960 и 1985 годами. Постройте график $y_{it1} - y_{it0}$ против y_{it0} для 125 стран, включенных в выборку. Есть ли какая-либо связь между начальным ВВП (y_{it0}) и последующим ростом ($y_{it1} - y_{it0}$)? Предложите в качестве альтернативы регрессорам, входящим в уравнение (2') посредством OLS. Вычислите значение λ (скорость конвергенции), вытекающее из OLS-оценки для ρ . (Она должна быть меньше 1% в год.) Вычислите асимптотическую стандартную ошибку (то есть умноженный на $1/\sqrt{n}$ квадратный корень из асимптотической дисперсии) нашей оценки λ . **Указание:** Из (5.4.3) следует, что $\hat{\lambda} = -\log(\hat{\rho})/25$. Из дельта-метода (лемма 2.5 в параграфе 2.1)

$$\text{Asym}(\hat{\lambda}) = \frac{\text{Asym}(\hat{\rho})}{(25\hat{\rho})^2}$$

получаемую производную $\partial \log(\hat{\rho}) / \partial \rho$ равна $1/\hat{\rho}$. Поэтому стандартная ошибка $\hat{\lambda}$ равна $\text{Asym}(\hat{\lambda})$ равна $\text{Asym}(\hat{\rho})/(25\hat{\rho})^2$. Стандартная ошибка $\hat{\lambda}$ равна квадратному от $1/\sqrt{n}$ квадратному корню из этой величины. Поэтому стандартная ошибка $\hat{\lambda} = (\text{стандартная ошибка } \hat{\rho})/(25\hat{\rho})$.

Можете ли вы подтвердить вывод о том, что «если страна не изменяет своих инвестиций и темпов роста населения, то должна быть устойчивой тенденция более быстрого роста бедных стран по сравнению с богатыми» (Mankiw et al., 1992, p. 428).

- (1) (Оценивание методом фиксированных эффектов.) Подставив $M = 25$, $t_0 = 1960, \dots, t_2 = 1985$, мы можем сформировать систему из 25 уравнений с (5.4.10) в качестве m -го уравнения. Оцените эту систему методом фиксированных эффектов, предполагая, что ошибка сферическая в смысле (5.2.12). Чему равно выведенное значение λ ? (Оно должно быть около 0.4% в год.) (Дополнительно. Вычислите стандартные ошибки без предположения о сферичности ошибок. Как подчеркивается в тексте, оценка фиксированных эффектов не является оптимальной. Мы тем не менее следили применением метода фиксированных эффектов только для получения опыта программирования.)

Совет для пакета TSP: Командой TSP для оценки фиксированных эффектов (с предположением о сферичности ошибок) является PANELE. Она, однако, не принимает организацию данных в SUM FES.ASC. Она требует, чтобы ID-улицы (страны) в год были вынесены в строки, содержащие информацию по группе-году. Файл данных, который PANELE принимает, называется FIXED.ASC. Его первые несколько записей имеют вид:

Ответы на избранные вопросы

Аналитические утверждения

1с. Решение имеет вид:

$$\begin{aligned} E(\eta_i | W) &= E(\eta_i | F) = \quad (\text{поскольку } D \text{ является матрицей констант}) \\ &= E(\eta_i | F_1, \dots, F_n) = \\ &= E(\eta_i | F_1) \\ &(\text{поскольку } \eta_i \text{ не зависит от } (F_1, \dots, F_{i-1}, F_{i+1}, \dots, F_n) \\ &\quad \text{по предположению i.i.d.}) \\ &= 0. \end{aligned}$$

Следовательно, регрессоры являются экзогенными. Также

$$\begin{aligned} E(\eta_i \eta_j' | W) &= E(\eta_i \eta_j' | F) = \\ &= E(\eta_i \eta_j' | F_1) = \\ &= \sigma_{\eta}^2 I_{ii} \quad (\text{по предположению о сферичности ошибок}). \end{aligned}$$

В силу предположения 1 в.с., $E(\eta_i \eta_j' | W) \rightarrow 0$ для $i \neq j$. Поэтому $E(\eta \eta' | W) = \sigma_{\eta}^2 I_{ii}$.

2с. Докажите, что « $E(QF_i \otimes x_i)$ имеет полный столбцовый ранг» $\Leftrightarrow$ « $E(C'F_i \otimes x_i)$ имеет полный столбцовый ранг»:

Пусть $A \equiv E(QF_i \otimes x_i)$ и $B \equiv E(C'F_i \otimes x_i)$. Существует матрица L размера $M \times (M-1)$ полного столбцового ранга, такая что $Q = LC'$ (это фактически $C(C'C)^{-1}$, как следует из (11)). Пусть $D \equiv L \otimes I_K$. Тогда $A = DB$. Поскольку ранг произведения двух матриц не превосходит ранга любой из этих двух матриц,

$$\text{rank}(A) \leq \text{rank}(B). \quad (1)$$

Далее, рассмотрим произведение $D'A = D'DB$.

$$\text{rank}(D'DB) \leq \text{rank}(D'A) \leq \text{rank}(A). \quad (2)$$

Но поскольку L имеет полный столбцовый ранг, $D'D$ является невырожденной. Так как умножение на невырожденную матрицу не изменяет ранг,

$$\text{rank}(D'DB) = \text{rank}(B). \quad (3)$$

Из (1)–(3) следует, что $\text{rank}(A) = \text{rank}(B)$. Поскольку A и B имеют одинаковое число столбцов, из этого следует нужный вывод.

4а.

$$y_1 = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1M} \end{bmatrix}, F_1 = \begin{bmatrix} y_{21} \\ \vdots \\ y_{6,M-1} \end{bmatrix}.$$

Вектора b_i нет.

4д Поскольку $E(\eta_{it} \cdot \eta_{it}) \neq 0$ для $it \geq b$, некоторые из кросс-ортogonalностей не выполняются.

Литература

- Barro, R., and X. Sala-i-Martin. 1995. *Economic Growth*. New York: McGraw-Hill.
- Mankiw, G., D. Romer, and D. Weil. 1992. "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, 107, 407-437.
- Newey, W., 1983. "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29, 229-256.
- Nickell, S., 1981. "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 1417-1428.
- Summers, R., and A. Heston. 1991. "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988," *Quarterly Journal of Economics*, 106, 527-568.

Глава 6. Серийная корреляция

Эта глава обобщает GMM-модель из главы 3, чтобы включить серийную корреляцию в произведение вектора инструментов и ошибки. Для этого нам нужно обобщить центральную предельную теорему из главы 2 на серийно коррелированные процессы. Такое обобщение возможно при определенных условиях, ограничивающих степень серийной корреляции. Соответствующее условие является аналогом для случайных процессов, называемых «*диффузиями*». Линейные процессы важны и сами по себе, и в этой главе их рассмотрению посвящен не менее пяти параграфов. В следующем параграфе обобщенная модель GMM используется для тестирования гипотезы «эффективного рынка» на этот раз вальдически.

Замечание по обозначениям: Поскольку проблемы, рассматриваемые здесь, являются специфическими для временных рядов, в этой главе мы используем индекс «*t*» вместо «*i*».

6.1. Моделирование серийной корреляции: линейные процессы

В параграфе 2.2 были рассмотрены ковариационно-стационарные процессы и их автоковариации. В частности, (скалярный или векторный) процесс **белого шума** (*white noise process*) $\{e_t\}$ является ковариационно-стационарным процессом с нулевым средним и с отсутствием серийной корреляции:

$$E(e_t) = 0 \quad E(e_t^2) = \sigma^2 > 0, \quad E(e_t e_{t-j}) = 0 \quad \text{для } j \neq 0$$

Очень важный класс ковариационно-стационарных процессов, называемых **линейными процессами** (*linear processes*), можно получить, беря скользящее среднее процесса белого шума. Этот параграф излагает основные свойства линейных процессов, методы идентификации и оценивания, называемый **фильтром** (*filter*).

Поскольку текущее значение линейного процесса может зависеть от возможного бесконечного количества прошлых значений процесса белого шума, удобно, чтобы процесс белого шума и линейный процесс, который он порождает, были определены для $t = 0, -1, -2, \dots$ так же как

и для $t = 1, 2, \dots$. Интуитивно ковариационно стационарный процесс, определенный для всех целых t ($t = 0, \pm 1, \pm 2, \dots$), является процессом, начинающимся настолько давно, что его среднее и автоковариация стабилизировались к инвариантным ко времени константам.

MA(q)

Самым простым примером линейных процессов, которые проявляют сериальную корреляцию, является процесс скользящего среднего конечного порядка. Процесс $\{y_t\}$ называется **процессом скользящего среднего q -го порядка** (q -th order moving average process) (MA(q)), если он может быть дан как взвешенное среднее текущих и q самых недавних значений процесса белого шума:

$$y_t = \alpha + \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad \varepsilon_0 = 1. \quad (6.1.1)$$

Очевидно, процесс скользящего среднего является ковариационно стационарным в пределах α . Легко проверить, что автоковариация j -го порядка, $\gamma_j \equiv E[(y_t - \mu)(y_{t-j} - \mu)]$, имеет вид:

$$\gamma_j = (\theta_0 \theta_j + \theta_{j-1} \theta_1 + \dots + \theta_q \theta_{q-j}) \sigma^2 = \sigma^2 \sum_{k=0}^{q-j} \theta_{j+k} \theta_k$$

$$\text{для } j = 0, 1, \dots, q, \quad (6.1.2a)$$

$$\gamma_j = 0 \quad \text{для } j > q. \quad (6.1.2b)$$

где $\mu \equiv E[y_t^2]$. (Эта формула также охватывает γ_{-j} , поскольку $\gamma_j = \gamma_{-j}$.) Например, для $q = 1$ мы имеем:

$$\begin{aligned} \gamma_0 &= (\theta_0^2 + \theta_1^2) \sigma^2 = (1 + \theta_1^2) \sigma^2 \\ &\quad (\text{полагая } j = 0 \text{ и } q = 1 \text{ в (6.1.2a) и замечая, что } \theta_0 = 1), \end{aligned}$$

$$\begin{aligned} \gamma_1 &= \gamma_{-1} = (\theta_1 \theta_0) \sigma^2 = \theta_1 \sigma^2 \\ &\quad (\text{полагая } j = 1 \text{ и } q = 1 \text{ в (6.1.2a) и замечая, что } \theta_0 = 1), \end{aligned}$$

$$\gamma_j = 0 \quad \text{для } j = \pm 2, \pm 3, \dots$$

Полный профиль автоковариаций, $\{\gamma_j\}$, описывается $q+1$ параметрами $(\theta_1, \theta_2, \dots, \theta_q \text{ и } \sigma^2)$, а корреляция $\{\rho_j\}$ (где $\rho_j \equiv \gamma_j/\gamma_0$) — q параметрами $(\theta_1, \theta_2, \dots, \theta_q)$.

MA(∞) как предел в среднем квадратическом

Как иллюстрирует приведенный выше пример, сериальная корреляция в процессах MA(q) полностью пропадает после q лагов. Хотя некоторые временные ряды обладают этим свойством (мы увидим такой пример

в приложении ниже), конечно, желательно, чтобы была возможность моделировать серийную корреляцию, которая являл^ься случайным, т.е. изменяет^ься. Очевидная идея — сделать y_t зависящим от бесконечного прошлого, заменяя сумму конечного количества слагаемых в (6.1.1):

$$\theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q},$$

или бесконечную сумму

$$w_0 \varepsilon_t + w_1 \varepsilon_{t-1} + \dots = \sum_{j=0}^{\infty} w_j \varepsilon_{t-j}, \quad (6.1.3)$$

где $\{w_j\}$ — последовательность действительных чисел.

Для того чтобы эта идея работала, мы должны быть уверены в том, что эта сумма бесконечно большого количества случайных величин определена, а именно что частичная сумма

$$\sum_{j=0}^n w_j \varepsilon_{t-j} \quad (6.1.4)$$

при $n \rightarrow \infty$ сходится (в определенном смысле) к некоторой случайной величине. Одно из необходимых условий, при которых это реализуется, состоит в том, что последовательность действительных чисел $\{w_j\}$ явля^ься абсолютно суммируемой (*absolutely summable*):

$$\sum_{j=0}^{\infty} |w_j| < \infty. \quad (6.1.5)$$

(Это стандартное сокращенное выражение в математическом анализе. Уравнение (6.1.5) следует читать так: частичная сумма $\sum_{j=0}^n |w_j|$ сходится к некоторому действительному числу [то есть к конечному пределу] при $n \rightarrow \infty$, а бесконечная сумма в (6.1.5) по определению обозначает это действительное число.) Чтобы последовательность $\{w_j\}$ была абсолютно суммируемой, необходимо (но, конечно, не достаточно), чтобы $w_j \rightarrow 0$ при $j \rightarrow \infty$. Таким образом, абсолютная суммируемость требует, чтобы влияния прошлых шоков, представленные посредством w_j , в конце концов затухали.

Как заявляется в следующем утверждении, для любого заданного ε частичная сумма (6.1.4) сходится в среднем квадратическом к некоторой случайной величине. Этот предел в среднем квадратическом является единственным в следующем смысле. Если существует другая случайная величина, к которой сходится в среднем квадратическом указанная частичная сумма, то тогда (как вы покажете в аналитическом упражнении 1) эти два предела в среднем квадратическом равны с вероятностью

единица. Мы определяем бесконечную сумму (6.1.3) как единственный предел в среднем квадратическом частичной суммы (6.1.4) и говорим, что эта бесконечная сумма сходится в среднем квадратическом.

Утверждение 6.1 ($MA(\infty)$ с абсолютно суммируемыми коэффициентами): Пусть $\{\varepsilon_t\}$ является белым шумом и $\{\psi_t\}$ является последовательностью действительных чисел, которая абсолютно суммируема. Тогда

(а) Для каждого t

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad (6.1.6)$$

сходится в среднем квадратическом $\{y_t\}$ является ковариационно стационарным. (Процесс $\{y_t\}$ называется **процессом скользящего среднего бесконечного порядка** ($MA(\infty)$) (*infinite-order moving-average process*)).

(б) Математическое ожидание y_t равно μ . Автоковариации $\{\gamma_j\}$ задают соотношение

$$\begin{aligned} \gamma_j &= (\psi_0\psi_0 + \psi_1\psi_1 + \psi_2\psi_2 + \dots)\sigma^2 - \\ &- \sigma^2 \sum_{k=0}^{\infty} \psi_{j+k}\psi_k \quad (j=0, 1, 2, \dots). \end{aligned} \quad (6.1.7)$$

(с) Автоковариации абсолютно суммируемы:

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty. \quad (6.1.8)$$

(д) Если, в дополнение, $\{\psi_t\}$ является $l.d.$, то тогда процесс $\{y_t\}$ является (иногда) стационарным и эргодическим.

Этот результат включает процессы $MA(q)$ в качестве частного случая с $\psi_j = \psi_j$ для $j = 0, 1, \dots, q$ и $\psi_j = 0$ для $j > q$. Очевидно, что такая последовательность абсолютно суммируема. Доказательство частей (а)–(с) является стандартным упражнением. Доказательство (а) будет основано на известном из математического анализа факте о收敛ности Коши (см. контрольный вопрос 1 о последовательности Коши). Часть (б) является очевидной экстраполяцией на бесконечность соответствующих результатов о среднем и автоковариациях процессов MA конечного порядка, но эта экстраполяция предполагает некоторое знание математических ожиданий и суммируемых. Например, чтобы

доказать, что средним $MA(\infty)$ -процессом является μ . Вам придется доказать, что математическое ожидание предела в среднем квадратическом (6.1.3) равно пределу математического ожидания (6.1.4). Тот факт, что эта операция легитимна при сходимости в среднем квадратическом, будет использован для доказательства части (b). Интуитивно часть (a) говорит о том, что если влияние прошлых шоков, представленных ε_t , затухает достаточно быстро, чтобы $\{\psi_t\}$ была абсолютно суммируемой, то тогда и серийная корреляция затухает быстро. Абсолютная суммируемость автоковариаций будет играть ключевую роль в теории ковариационно-стационарных процессов, разрабатываемой в этой главе. Для доказательства (d) см., например: [Hannan, 1970, p. 204, Theorem 3]¹.

Процесс $MA(\infty)$ в (6.1.6) называется **односторонним** (*one-sided*), поскольку скользящее среднее не включает будущее значение ε_t . Это частный случай **линейного процесса** (*linear process*), который определяется как двустороннее скользящее среднее вида:

$$y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j} \quad \text{с} \quad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

Указанное утверждение (b) (все результаты в этом параграфе) можно легко распространить на общие линейные процессы. Мы, однако, фокусируемся на односторонних скользящих средних только потому, что двусторонние скользящие средние редко встречаются в экономике.

Утверждение 6.1 может быть обобщено на случай, когда процесс $\{y_t\}$ является скользящим средним общего ковариационно-стационарного процесса, а не процесса белого шума. В частности, абсолютная суммируемость автоковариаций переживает операцию взятия абсолютно суммируемого бесконечного среднего.

Утверждение 6.2 (Фильтрация ковариационно-стационарного процесса): Пусть $\{x_t\}$ является ковариационно-стационарным процессом, а $\{h_t\}$ — абсолютно суммируемой последовательности действительных чисел. Тогда

(a) Для каждого x бесконечная сумма

$$y_t = \sum_{j=0}^{\infty} h_j x_{t-j} \quad (6.1.9)$$

¹ Абсолютной суммируемости $\{\psi_j\}$ более чем достаточно для выполнения части (a). Как показано в аналитическом утверждении 1 суммируемости с квадратом, которая является более слабой, чем абсолютная суммируемость, достаточно, чтобы гарантировать, что бесконечная сумма также определена как предел в среднем квадратическом. Однако для выполнения других частей этого утверждения необходима абсолютная суммируемость.

сходится в среднем квадратическом. Процесс $\{y_t\}$ является ковариационно стационарным.

(b) Если, кроме того, автоковариации процесса $\{y_t\}$ абсолютно суммируемы, то тогда то же свойство выполняется и для автоковариаций процесса $\{y_t\}$.

За доказательством мы отсылаем, например, к Фуллеру [Fuller, 1996, Theorem 4.3.1], но основная техника здесь та же самая, что и в доказательстве утверждения 6.1. Получение автоковариаций $\{y_t\}$, которые обобщают (6.1.7), является аналитическим упражнением 3.

Фильтры

Операция взятия внешнего среднего (возможно бесконечного количества) последовательных значений процесса, как в (6.1.6) и (6.1.9), называется **фильтрацией** (*filtering*). Ее можно записать в компактном виде, если использовать так называемый **оператор запаздывания** (*lag operator*) L , который определяется соотношением $Lx_t = x_{t-1}$. Мы сейчас введем некоторые понятия, связанные с оператором запаздывания, которые будут полезны в анализе временных рядов¹.

Определение фильтров

Для произвольно заданной последовательности действительных чисел, $(\alpha_0, \alpha_1, \alpha_2, \dots)$, определим **фильтр** (*filter*) как

$$\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \dots \quad (6.1.10)$$

Если мы просто механически, не задумываясь, применим определение оператора запаздывания к процессу $\{x_t\}$, то получим:

$$\begin{aligned} \alpha(L)x_t &= \alpha_0 x_t + \alpha_1 Lx_t + \alpha_2 L^2 x_t + \dots = \\ &= \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots = \\ &= \sum_{j=0}^{\infty} \alpha_j x_{t-j}. \end{aligned} \quad (6.1.11)$$

Это и есть определение $\alpha(L)x_t$. Если $x_t = c$ (константы), то $\alpha(L)c = \alpha(1)c = c \sum_{j=0}^{\infty} \alpha_j$. Если $\alpha_j \neq 0$ для $j = p$ и $\alpha_j = 0$ для $j > p$, фильтр сводится к **лаговому полиному p -й степени** (*p -th degree lag polynomial*):

$$\alpha(L) = \alpha_0 + \alpha_1 L + \dots + \alpha_p L^p.$$

¹ Для более полного обсуждения фильтров см., например, Geweke and Porter, 1997, Sections 6.1 и 7.4.

Когда он применяется к входному процессу, то создает идеальное среднее текущего и p самых недавних значений процесса. Утверждение 6.2 гарантирует нам, что объект $\alpha(L)x_t$, определенный в (6.1.11), является случайной величиной, формирующей ковариационно стационарный процесс, если последовательность $\{\alpha_t\}$ является абсолютно суммируемой и если входной процесс $\{x_t\}$ является ковариационно стационарным. Фильтр с абсолютно суммируемыми коэффициентами будет называться (в этой книге) **абсолютно суммируемым фильтром** (*absolutely summable filter*). Если использовать модный язык теории множеств, то абсолютно суммируемый фильтр является отображением множества ковариационно стационарных процессов в себя.

Произведение фильтров

Пусть $\{\alpha_j\}$ и $\{\beta_j\}$ являются двумя абсолютно суммируемыми последовательностями действительных чисел. Определим последовательность $\{\delta_j\}$ соотношением:

$$\begin{aligned}\alpha_0\beta_0 &= \delta_0, \\ \alpha_0\beta_1 + \alpha_1\beta_0 &= \delta_1, \\ \alpha_0\beta_2 + \alpha_1\beta_1 + \alpha_2\beta_0 &= \delta_2, \\ &\vdots, \\ \alpha_0\beta_j + \alpha_1\beta_{j-1} + \alpha_2\beta_{j-2} + \cdots + \alpha_{j-1}\beta_1 + \alpha_j\beta_0 &= \delta_j, \\ &\vdots.\end{aligned}\quad (6.1.12)$$

Последовательность $\{\delta_j\}$, построенная по этой запутанной формуле, называется **сверткой** (*convolution*) последовательностей $\{\alpha_j\}$ и $\{\beta_j\}$. Пусть $\alpha(L) \equiv \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots$, $\beta(L) \equiv \beta_0 + \beta_1 L + \beta_2 L^2 + \cdots$ и $\delta(L) \equiv \delta_0 + \delta_1 L + \delta_2 L^2 + \cdots$ являются соответствующими фильтрами. Мы определяем **произведение** (*product*) двух функций $\alpha(L)$ и $\beta(L)$ как этот самый фильтр $\delta(L)$ и записываем его как

$$\delta(L) \equiv \alpha(L)\beta(L).$$

Например, для $\alpha(L) = 1 - \alpha_1 L$ и $\beta(L) = 1 + \beta_1 L$ формула свертки приводит к

$$(1 - \alpha_1 L)(1 + \beta_1 L) = 1 + (\beta_1 - \alpha_1)L - \alpha_1\beta_1 L^2.$$

(Если вы знакомы с произведением двух функций обычного вида из математического анализа, вы можете сразу видеть, что это определение для фильтров почти так же аналогично.) Как ясно из определения, фильтры коммутативны (*commutative*):

$$\alpha(L)\beta(L) = \beta(L)\alpha(L).$$

Коммутативность, однако, не будет переноситься на матричные фильтры (см. параграф 5.3 ниже).

Если две последовательности $\{\alpha_i\}$ и $\{\beta_i\}$ абсолютно суммируемы и если процесс $\{x_i\}$ ковариационно стационарный, то утверждение 6.2 гарантирует нам, что процесс $\{\beta(L)x_i\}$ ковариационно стационарный, к тому же $\alpha(L)\beta(L)x_i$ является случайной величиной, равной $\delta(L)x_i$. Таким образом, если $\alpha(L)$ и $\beta(L)$ абсолютно суммируемы и процесс $\{x_i\}$ ковариационно стационарный, то

$$\alpha(L)\beta(L)x_i = \delta(L)x_i \Rightarrow \alpha(L)\beta(L)x_i = \beta(L)\alpha(L)x_i. \quad (6.1.13)$$

Итак, последовательность $\{\beta_i\}$ абсолютно суммируема (см., например: [Fuller, 1996, p. 30]).

Обращение

Особый интерес представляет случай, когда $\delta(L) = 1$ (весьма частый случай фильтра), так что два фильтра $\alpha(L)$ и $\beta(L)$ удовлетворяют соотношению $\alpha(L)\beta(L) = 1$. Мы говорим, что $\beta(L)$ является обращением (инверсом) фильтра $\alpha(L)$ и обозначаем его как $\alpha(L)^{-1}$ или $1/\alpha(L)$. То есть

$$\alpha(L)^{-1} \text{ является фильтром, удовлетворяющим соотношению} \\ \alpha(L)\alpha(L)^{-1} = 1. \quad (6.1.14)$$

При условии, что $\alpha_0 \neq 0$ и $\alpha(L) = \alpha_0 + \alpha_1 L + \dots$, обращение для $\alpha(L)$ может быть определено для произвольной последовательности $\{\beta_i\}$, поскольку множество уравнений (6.1.12) с $\delta_0 = 1$, $\delta_i = 0$ ($i = 1, 2, \dots$) может быть последовательно разрешено относительно $\{\beta_i\}$:

$$\beta_0 = \frac{1}{\alpha_0}, \quad \beta_1 = -\frac{\alpha_1 \beta_0}{\alpha_0} = -\frac{\alpha_1}{\alpha_0^2}, \quad \text{и т.д.}$$

В качестве примера рассмотрим фильтр $1 - L$. Обратный ему фильтр задается как

$$(1 - L)^{-1} = 1 + L + L^2 + L^3 + \dots$$

Последовательность коэффициентов здесь, очевидно, не является абсолютно суммируемой. Этот пример иллюстрирует тот факт, что обратный фильтр может не быть абсолютно суммируемым.

В качестве контрольного вопроса 3 оставляем показать, что

$$\alpha(L)\alpha(L)^{-1} = \alpha(L)^{-1}\alpha(L) \quad (\text{поэтому обращение коммутативно}), \quad (6.1.15a)$$

$$\alpha(L)\beta(L) = \delta(L) \Leftrightarrow \alpha(L) = \alpha(L)^{-1}\delta(L) \Leftrightarrow \alpha(L) = \delta(L)\alpha(L)^{-1}, \quad (6.1.15b)$$

при условии, что $\alpha_1 \neq 0$ и $\beta_0 \neq 0$. Следовательно, например, чтобы решить $\alpha(L)\beta(L) = \delta(L)$ относительно $\beta(L)$, мы умножаем слева на обращение $\alpha(L)$. Мы заметим далее, что коммутативность необходимо исключается для матричных фильтров. Как будет показано в параграфе 6.3, для обращений коммутативность обобщается на матричные фильтры.

Обращение латентных полиномов

В следующем параграфе и приложениях ARMA будет необходимо вычислить обращение латентного полинома p -й степени $\phi(L)$

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p.$$

Такое обращение может быть определено, поскольку $\phi_0 = 1 \neq 0$. Пусть $\psi(L) = \phi(L)^{-1}$. Тогда, по определению,

$$\phi(L)\psi(L) = 1.$$

Формула свертки (6.1.12) предлагает алгоритм решения этого уравнения относительно $\psi(L)$. Полагая в этой формуле $\alpha(L) = \phi(L)$, $\beta(L) = \psi(L)$ и $\delta(L) = 1$, получаем:

$$\begin{aligned} \text{константа: } \psi_0 &= 1, \\ L: \psi_1 &= \phi_1 \psi_0 &= 0, \\ L^2: \psi_2 &= \phi_1 \psi_1 + \phi_2 \psi_0 &= 0, \\ &\dots &\dots \\ L^{p-1}: \psi_{p-1} &= \phi_1 \psi_{p-2} + \phi_2 \psi_{p-3} + \dots + \phi_{p-1} \psi_0 &= 0, \\ L^p: \psi_p &= \phi_1 \psi_{p-1} + \phi_2 \psi_{p-2} + \dots + \phi_{p-1} \psi_1 + \phi_p \psi_0 &= 0, \\ L^{p+1}: \psi_{p+1} &= \phi_1 \psi_p + \phi_2 \psi_{p-1} + \dots + \phi_{p-1} \psi_2 + \phi_p \psi_1 &= 0, \\ L^{p+2}: \psi_{p+2} &= \phi_1 \psi_{p+1} + \phi_2 \psi_p + \dots + \phi_{p-1} \psi_3 + \phi_p \psi_2 &= 0, \\ &\dots &\dots \end{aligned} \quad (6.1.16)$$

Эти уравнения могут быть легко решены последовательно от ψ_0 относительно $\psi_1, \psi_2, \dots$ как

$$\psi_0 = 1, \quad \psi_1 = \phi_1, \quad \psi_2 = \phi_2 + \phi_1^2, \quad \dots$$

Также заметим, что для достаточно большого j (фактически для $j \geq p$) ψ_j подпадает однородному разностному уравнению p -го порядка

$$\psi_j - \phi_1 \psi_{j-1} - \phi_2 \psi_{j-2} - \dots - \phi_{p-1} \psi_{j-p+1} - \phi_p \psi_{j-p} = 0. \quad (6.1.17)$$

Поэтому, вычислив первые p коэффициентов $\psi_0, \psi_1, \dots, \psi_{p-1}$, мы можем использовать это однородное разностное уравнение p -го порядка

для порождения оставшихся коэффициентов ($\psi_j, j \geq p$) с этими p коэффициентами в качестве начальных значений.

Как мы знаем из теории однородных разностных уравнений (кратко описанной в аналитическом выражении 4), последовательность решений $\{\phi_j\}$ уравнения (6.1.17), в конце концов, начинает убывать в геометрической прогрессии, если выполняется то, что известно как **условие стабильности** (*stability condition*). Это условие формулируется следующим образом:

Все корни полиномиального уравнения p -го порядка от z

$$\phi(z) = 0, \quad \text{где} \quad \phi(z) \equiv 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p \quad (6.1.18)$$

больше 1 по абсолютной величине.

Корни полиномиального уравнения $\phi(z) = 0$ могут быть комплексными. Если наше знание комплексными числами ослабло, напомним, что комплексное число z может быть записано как

$$z = a + bi,$$

где a и b — два действительных числа, а $i = \sqrt{-1}$. Поэтому комплексное число может быть представлено как точка на двумерной диаграмме с горизонтальной осью, измеряющей a (действительная часть), и вертикальной осью, измеряющей b (мнимая часть). Его абсолютные значения равно $|z| = \sqrt{a^2 + b^2}$, то есть расстоянию между началом координат и точкой, представляющей z . Мы говорим, что z лежит вне единичного круга, если $|z| > 1$. Поэтому условие стабильности, сформулированное выше, может быть иначе сформулировано как требование того, что корни указанного уравнения лежат вне единичного круга. Поскольку $z = \lambda$ является решением уравнения $\phi(z) = 0$ всякий раз, когда $z = 1/\lambda$ является решением уравнения $z^p - \phi_1 z^{p-1} - \phi_2 z^{p-2} - \dots - \phi_{p-1} z - \phi_p = 0$, условие стабильности может быть эквивалентно сформулировано следующим образом:

Все корни полиномиального уравнения p -го порядка

$$z^p - \phi_1 z^{p-1} - \phi_2 z^{p-2} - \dots - \phi_{p-1} z - \phi_p = 0 \quad (6.1.18')$$

меньше 1 по абсолютному значению (то есть лежат *внутри* единичного круга).

При этом условии, как вам будет предложено доказать (аналитическое выражение 4), существуют два действительных числа, $A > 0$ и $0 < b < 1$, таких что

$$|\psi_j| < Ab^j \quad \text{для всех } j. \quad (6.1.19)$$

При выполнении (6.1.19) абсолютная суммируемость $\{\psi_j\}$ следует из

$$\sum_{j=0}^{\infty} |\psi_j| < \sum_{j=0}^{\infty} Ab^j = \frac{A}{1-b} < \infty.$$

Таким образом, мы доказали

Утверждение 6.3 (абсолютно суммируемые обращения лаговых полиномов): Рассмотрим лаговый полином p -й степени $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$, и положим $\psi(L) \equiv \phi(L)^{-1}$. Если соответствующее полиномиальное уравнение p -й степени $\phi(z) = 0$ удовлетворяет условию стабильности (6.1.18) или (6.1.18'), то последовательность коэффициентов $\{\psi_j\}$ для $\psi(L)$ ограничена по абсолютному значению геометрически убывающей прогрессией (как в (6.1.19)) и, следовательно, абсолютно суммируема.

Для иллюстрации рассмотрим лаговый полином степени единицы, $1 - \phi L$. Корень соответствующего полиномиального уравнения $1 - \phi z = 0$ равен $1/\phi$. Условие стабильности заключается в том, что $|1/\phi| > 1$ или $|\phi| < 1$. Фильтр

$$1 + \phi L + \phi^2 L^2 + \dots \quad (6.1.20)$$

очевидно является обращением рассматриваемого полинома. Последовательность коэффициентов $\{\phi^j\}$ действительно ограничена по абсолютному значению геометрически убывающей последовательностью (например, положим $A = 1,1$ и $b = \phi$).

Контрольные вопросы

1. (Последовательность Коши.) Последовательность действительных чисел, $\{\alpha_n\}$, называют **последовательностью Коши** (Cauchy sequence), если $\alpha_m - \alpha_n \rightarrow 0$ при $m, n \rightarrow \infty$. Важный результат математического анализа заключается в том, что α_n сходится к действительному числу тогда и только тогда, когда эта последовательность является последовательностью Коши. Принимая это как данность, покажите, что $\{\gamma_j\}$ абсолютно суммируема тогда и только тогда, когда $\sum_{j=n+1}^m \gamma_j \rightarrow 0$ при $m, n \rightarrow \infty$. (Вы можете предположить без потери общности, что $m > n$.)
2. (Как фильтрация изменяет автоковариации.) Пусть $\hat{h}(L) = h_0 + h_1 L$ и $y_t = \hat{h}(L)x_t$, где процесс $\{x_t\}$ ковариационно стационарный. Проверьте, что (с автоковариациями процесса $\{x_t\}$, равными $\{\gamma_j^x\}$) автоковариации процесса $\{y_t\}$ задаются соотношением:

$$\gamma_j^y = \sum_{k=0}^1 \sum_{l=0}^1 h_k h_l \gamma_{j-k-l}^x.$$

3. Докажите (6.1.15b). **Указание:** Вот доказательство того, что

$$a(L)\beta(L) = b(L) \Leftrightarrow a(L) = \beta(L)^{-1}b(L);$$

$$a(L)^{-1}b(L) = m(L)^{-1}n(L)\beta(L) = a(L)a(L)^{-1}\beta(L) = \beta(L).$$

4. Удовлетворяет ли $\phi(z) = 1 + \frac{1}{2}z + \frac{1}{8}z^2$ условию стационарности? **Указание:** Корни $\phi(z) = 0$ равны $2(1 \pm \sqrt{3})/5$.

5. (Коэффициенты полиномиальных уравнений.) Рассмотрим полиномиальное уравнение $\phi(z) = 0$, где $\phi(z) = 1 - \alpha_1 z - \alpha_2 z^2$, и пусть (λ_1, λ_2) являются двумя его корнями. Проверьте, что $\phi_1 = 1/\lambda_1 = 1/\lambda_2$, $\phi_2 = 1/(\lambda_1 + \lambda_2)$. **Указание:** Если λ_1 и λ_2 являются двумя корнями для $\phi(z) = 0$, $\phi(z)$ может быть записан как $\phi(z) = (1 - \frac{1}{\lambda_1}z)(1 - \frac{1}{\lambda_2}z)$.

6. (Обращение $1 - \phi(L)$.) Убедитесь, что (6.1.20) является обращением для $1 - \phi(L)$, проверяя формулу свертки (6.1.12) с $\phi_j = 1$, $\phi_j = 0$ для $j \geq 1$.

6.2. Процессы ARMA

После введения понятия фильтров мы можем с легкостью и изяществом изучать класс линейных процессов, называемых процессами ARMA, которые являются параметризацией коэффициентов процесса MA(∞).

AR(1) и его MA(∞)-представление

Процесс авторегрессии первого порядка (AR(1)) (*first-order autoregressive process*) удовлетворяет следующему стохастическому разностному уравнению:

$$y_t = c + \phi y_{t-1} + \varepsilon_t \quad \text{или} \quad y_t - \phi y_{t-1} = c + \varepsilon_t \quad \text{или} \quad (1 - \phi L)y_t = c + \varepsilon_t, \quad (6.2.1)$$

где $\{\varepsilon_t\}$ является белым шумом. Если $\phi \neq 1$, положим $\mu \equiv c/(1 - \phi)$ и перепишем это уравнение как

$$(y_t - \mu) - \phi \cdot (y_{t-1} - \mu) = \varepsilon_t \quad \text{или} \quad (1 - \phi L)(y_t - \mu) = \varepsilon_t. \quad (6.2.1')$$

Как будет видно через мгновение, именно μ , а не c является математическим ожиданием процесса y_t , если ε_t ковариационно стационарный. По этой причине мы будем называть (6.2.1') формой в отклонениях от среднего. Следующее среднее является последовательными значениями $\{y_t\}$, а не $\{\varepsilon_t\}$. Указанное разностное уравнение называется стохастическим из-за наличия случайной величины ε_t .

Мы ищем ковариационно стационарное решение $\{y_t\}$ этого стохастического разностного уравнения. Это решение зависит от того, является ли $|\phi|$ меньшим, равным или большим, чем 1.

Случай 1: $|\phi| < 1$

Решение можно легко получить, используя обращение $(1 - \phi L)^{-1}$, заданное выражением (6.1.20). Так как этот фильтр абсолютно суммируемый, когда $|\phi| < 1$, мы можем применить его к обеим сторонам $\Delta N(1)$ -уравнения (6.1.1'), получая в результате:

$$(1 - \phi L)^{-1}(1 - \phi L)(y_t - \mu) = (1 - \phi L)^{-1}\varepsilon_t.$$

Эта операция легитимна, поскольку благодаря утверждению 6.2 обе стороны этого уравнения определены как пределы в среднем квадратическом. По определению обращения (6.1.14), $(1 - \phi L)(1 - \phi L)^{-1} = 1$, а по коммутативности обращения, $(1 - \phi L)^{-1}(1 - \phi L) = 1$. Следовательно, если $\{y_t\}$ ковариационно стационарный, левая сторона равна $y_t - \mu$ по (6.1.13). Поэтому

$$y_t - \mu = (1 - \phi L)^{-1}\varepsilon_t = (1 + \phi L + \phi^2 L^2 + \dots)\varepsilon_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

или $y_t - \mu = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$. (6.2.2)

То, что мы показали, — это то, что если $\{y_t\}$ является ковариационно стационарным решением стохастического разностного уравнения (6.2.1) или (6.2.1'), то y_t имеет представление в виде сходящегося среднего (*moving-average representation*) как в (6.2.2). Обратим, если процесс y_t имеет представление (6.2.2), то он удовлетворяет указанному разностному уравнению. Это можно легко показать, подставив (6.2.2) в (6.2.1'):

$$\begin{aligned} (1 - \phi L)(y_t - \mu) &= (1 - \phi L)(1 - \phi L)^{-1}\varepsilon_t = \\ &= \varepsilon_t \quad (\text{поскольку } (1 - \phi L)(1 - \phi L)^{-1} = 1 \text{ по определению обращения}). \end{aligned}$$

Таким образом, процесс $\{y_t\}$, заданный в (6.2.2), является *единственным* ковариационно стационарным решением стохастического дифференциального уравнения, если $|\phi| < 1$. Математическим ожиданием этого процесса, согласно утверждению 6.1(б), является μ .

Условие $|\phi| < 1$, называемое условием стабильности (6.1.18), в сочетании с полиномиальным уравнением первой степени $1 - \phi z = 0$ называется **условием стационарности** (*stationarity condition*) в текущем контексте процессов авторегрессии. Интуитивно этот результат говорит о том, что соответствующий процесс, если он начинается очень давно, стабилизируется, при условии, что $|\phi| < 1$, так что влияние прошлого геометрически затухает с течением времени.

Случай 2: $|\phi_1| > 1$

Сдвигая время вперед на один период (то есть заменяя « t » на « $t+1$ »), умножая обе стороны на ϕ^{-1} и приравняв перестановку, стохастическое разностное уравнение (6.2.1') можно записать в виде:

$$y_t - \mu = \phi^{-1} (y_{t+1} - \mu) - \phi^{-1} \varepsilon_{t+1}. \quad (6.2.3)$$

Применяя того же рассуждения, что и выше, но на этот раз с перемещением в обратном направлении во времени, доказывает, что единственным ковариационно-стационарным решением является здесь

$$\mu = \mu - \sum_{j=1}^{\infty} \phi^{-j} \varepsilon_{t+j}. \quad (6.2.4)$$

То есть текущее значение y является скалярным средним будущих значений ε . Эта бесконечная сумма определена, поскольку последовательность $\{\phi^{-j}\}$ абсолютно суммируема, если $|\phi| > 1$.

Случай 3: $\phi = 1$

Рассматриваемое стохастическое разностное уравнение не имеет ковариационно-стационарного решения. Например, если $\phi = 1$, это стохастическое разностное уравнение принимает вид:

$$\begin{aligned} y_t &= c + y_{t-1} + \varepsilon_t, \\ &= c + (c + y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t. \quad \text{Поскольку } y_{t-1} = c + y_{t-2} + \varepsilon_{t-1} = \\ &= c + (c + (c + y_{t-3} + \varepsilon_{t-2}) + \varepsilon_{t-1}) + \varepsilon_t, \quad \text{и т.д.} \end{aligned}$$

Подставляя эти же рассуждения для подстановки j раз, мы получаем:

$$y_t = y_{t-j} + (c \cdot j) + (\varepsilon_t + \varepsilon_{t-1} + \dots + \varepsilon_{t-j+1}).$$

(Если $\{\varepsilon_t\}$ является независимым белым шумом, этот процесс является случайным блужданием со сносом c (*random walk with drift c*)). Предположим, вопреки нашему заявлению, что этот процесс ковариационно-стационарный. Тогда дисперсия $y_t - y_{t-j}$ равна $2(\gamma_0 - \gamma_j)$. Дисперсия правой части равна $\sigma^2 \cdot j$. Поэтому

$$\begin{aligned} 2 \cdot (\gamma_0 - \gamma_j) &= \sigma^2 \cdot j \quad \text{или} \\ \rho_j &= 1 - \frac{\sigma^2}{2\gamma_0} \cdot j < -1 \quad \text{для достаточно больших } j \\ & \quad (\text{вспомните: } \rho_j = \gamma_j / \gamma_0). \end{aligned}$$

Это является противоречием, поскольку коэффициенты автокорреляции не могут быть больше единицы по абсолютному значению. Поэтому

процесс $\{y_t\}$ не может быть ковариационно стационарным. Процессы с $\phi = 1$ являются примерами «процессов с единичным корнем» и будут изучаться в главе 9.

Решение (6.1.4) для случая $|\phi| > 1$, в котором текущее значение y_t связано с будущими значениями порождающие их $\{z_t\}$, непригодно для экономики, потому что не наделяет экономических агентов рациональным предвидением. В дальнейшем, если не указано противное, мы будем использовать термин «AR(1)-процесс» для обозначения ковариационно стационарного решения АП(1)-уравнения (6.2.1) или (6.2.1') при выполнении условия единичной стационарности.

Аппроксимация AR(1)

Автокорреляции $\{\gamma_j\}$ для AR(1) (для случая $|\phi| < 1$) могут быть вычислены двумя способами. Первый способ заключается в представлении AR(1) в виде MA(∞), как в (6.2.2), и использовании (6.1.7). Поскольку для AR(1) $\gamma_2 = \phi^2$, мы имеем

$$\begin{aligned} \gamma_j &= (\phi^2 + \phi^{2j} - \phi^{2j+1} + \dots) \sigma^2 = \frac{\phi^{2j} \sigma^2}{1 - \phi^2} \\ \gamma_j / \gamma_0 &(\equiv \gamma_j / \gamma_0) = \phi^{2j} \quad (j = 0, 1, \dots). \end{aligned} \quad (6.2.5)$$

Поэтому коррелограмма процесса AR(1) очень проста, она убывает со j с постоянной геометрической скоростью.

Второй способ основан на использовании так называемых **уравнений Юла — Уокера** (*Yule — Walker equations*); см. аналитическое уравнение (6.1.6) и численное уравнение Юла — Уокера при численном автокорреляции.

AR(p) и его MA(∞)-представление

Все результаты, только что полученные для AR(1), могут быть обобщены на АП(p), процесс авторегрессии **p-го порядка** (*p-th order autoregressive process*), который удовлетворяет стохастическому разностному уравнению p-го порядка:

$$\begin{aligned} y_t &= c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t, \quad \text{или} \\ y_t &= c_1 y_{t-1} + \dots + c_p y_{t-p} + c + \varepsilon_t, \quad \text{или} \\ \phi(L)y_t &= c + \varepsilon_t \quad c = \phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p. \end{aligned} \quad (6.2.6)$$

где $\{\varepsilon_t\}$ — белый шум и $\phi_p \neq 0$. Очевидно, $\phi(1) = 1 - \phi_1 - \dots - \phi_p$. Если $\phi(1) \neq 0$, положим $\mu \equiv c/(1 - \phi_1 - \dots - \phi_p) = c/\phi(1)$. Как показано ниже, μ является математическим ожиданием процесса y_t , если y_t является ковариационно стационарным. Выразим теперь c AR(p)-уравнение (6.2.6)

можно эквивалентно записать в форме отклонений от среднего:

$$(y_t - \mu) = \phi \cdot (y_{t-1} - \mu) + \phi' \cdot \dots + \phi_n \cdot (y_{t-n} - \mu) + \varepsilon_t \\ \text{или } \phi(L)(y_t - \mu) = \varepsilon_t. \quad (6.2.6')$$

Следующее утверждение является обобщением на процесс $AR(p)$ результата, полученного ранее для процессов $AR(1)$.

Утверждение 6.4 ($AR(p)$ как $MA(\infty)$ с абсолютно суммируемыми коэффициентами): *Предположим, что полином p -й степени $\phi(z)$ удовлетворяет условию стационарности (стабильности) (6.1.16) или (6.1.18'). Тогда*

- (а) Единственное каузально стационарное решение стохастического разностного уравнения в отрезках (6.2.6) или (6.2.6') имеет $MA(\infty)$ -представление

$$y_t = \mu + \phi(L)\varepsilon_t, \quad \phi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \dots, \quad (6.2.7)$$

где $\phi(L) = \phi(L)^{-1}$. Последовательность коэффициентов $\{\psi_j\}$ ограничена геометрически убывающей последовательностью r , следовательно, абсолютно суммируема.

- (б) Математическое ожидание μ указанного процесса задается как

$$\mu = \phi(L)^{-1}c_1, \quad \text{где } c_1 \text{ — некоторая константа в (6.2.6)}. \quad (6.2.8)$$

- (в) $\{\psi_j\}$ ограничена по абсолютному значению последовательностью, которая геометрически убывает по j . Следовательно, абсолютно суммируема.

Если не указано противное, мы будем использовать термин «процесс $AR(p)$ » для единственного стационарного каузального решения (6.2.7) $AR(p)$ -уравнения (6.2.6), удовлетворяющего условию стационарности. Абсолютная суммируемость обращения $\phi(L)^{-1}$ в части (а) этого утверждения непосредственно следует из утверждения 6.3. При наличии абсолютной суммируемости вторая часть (а) может быть доказана с использованием той же аргументации, которую мы использовали для $AR(1)$: просто умножьте обе стороны (6.2.6') на $\phi(L)^{-1}$ и используйте (6.2.7).

$$\phi(L)^{-1}\phi(L) = \phi(L)\phi(L)^{-1} = 1$$

Часть (б) непосредственно следует из утверждения 6.1(б). Часть (в) можно легко доказать, объединяя (а) и утверждение 6.1(б). В аналитическом упражнении 7 мы будем предложено дать альтернативное доказательство

этого утверждения без использования фильтров, но с использованием инструмента под названием **сопровождающая форма** (*companion form*).

Как и для АН(1), автокорреляции могут быть получены двумя способами. Первый способ использует представление МА(∞). Коэффициенты $\{\phi_i\}$ в представлении МА являются коэффициентами в обращении лаг-ового полинома $\phi(L)$, в алгоритм для вычисления таких коэффициентов уже был представлен ранее (см. (5.1.15)). При заданных $\{\psi_i\}$ используйте (5.1.7) для вычисления автоковариаций $\{\gamma_i\}$. Второй метод использует уравнения Юла — Уокера.

ARMA(p, q)

Процесс ARMA(p, q) объединяет АН(p) и МА(q).

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \\ + \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q},$$

или

$$\theta(L)y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} = \\ = c + \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q},$$

или

$$\phi(L)y_t = c + \theta(L)\varepsilon_t, \quad \phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p \\ \theta(L) = \theta_0 + \theta_1 L + \dots + \theta_q L^q, \quad (6.2.9)$$

где $\{\varepsilon_t\}$ — белый шум. Если $\phi(L) \neq 0$, положим $\mu = c/\phi(1)$. Форма в отклонениях от среднего имеет вид

$$(y_t - \mu) = \phi_1 (y_{t-1} - \mu) + \dots + \phi_p (y_{t-p} - \mu) = \\ = \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

или

$$\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t \quad (6.2.9')$$

Это все еще стохастическое разностное уравнение p -го порядка, но с серийно коррелированными порождающим процессом $\{\theta(L)\varepsilon_t\}$ вместо процесса белого шума $\{\varepsilon_t\}$. Утверждение 6.4 легко обобщается на следующий.

Утверждение 6.5 (ARMA(p, q) как МА(∞) с абсолютно суммируемыми коэффициентами): Предположим, что p -й степени $\phi(L)$ удовлетворяет условиям стационарности (6.1.18) или (6.1.18'). Тогда

(a) Единственное ковариационно-стационарное решение стохастического разностного уравнения p -го порядка (6.2.9) или (6.2.9') имеет

MA(∞) представление

$$y_t = \mu + \psi(L)\varepsilon_t, \quad \psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \dots, \quad (6.2.10)$$

где $\psi(L) \equiv \phi(L)^{-1}\theta(L)$. Последовательность коэффициентов $\{\psi_j\}$ ограничена геометрически убывающей последовательностью λ , следовательно, абсолютно суммируема.

(b) Математическое ожидание μ указанного процесса задается как

$$\mu = \phi(1)^{-1}c, \quad \text{где } c \text{ является константой в (6.2.9)}. \quad (6.2.11)$$

(c) $\{\psi_j\}$ ограничена по абсолютному значению последовательностью, которая геометрически убывает по j . Следовательно, автоковариансы абсолютно суммируемы.

Если не указывать противное, мы будем использовать термин «процесс ARMA» для единственного решения (6.2.10) ARMA уравнения (6.2.9) или (6.2.8'), которые удовлетворяют условиям стационарности. Представление MA(∞) можно легко получить, умножив обе стороны (6.2.8') на обращение $\phi(L)^{-1}$ и замечая, что $\phi(L)^{-1}\phi(L) = \phi(L)\phi(L)^{-1} = 1$. Для оставшейся части (a) и части (b) единственной неотрецируемая часть доказательства заключается в том, чтобы показать, что $\{\psi_j\}$ ограничена по абсолютному значению геометрически убывающей последовательностью. В случае AR(p) мы использовали алгоритм, основанный на множестве уравнений (6.1.16), чтобы решить $\phi(L)\psi(L) = \theta(L)$ относительно $\psi(L)$. Этот алгоритм может быть легко обобщен. На этот раз $\psi(L) \equiv \phi(L)^{-1}\theta(L)$ или

$$\phi(L)\psi(L) = \theta(L).$$

Чтобы решить это относительно $\psi(L)$, применим формулу обратной (6.1.12) с $\phi(L) = \phi(L)$, $\beta(L) = \psi(L)$ и $\delta(L) = \theta(L)$ (вместо (1). Результатирующее множество уравнений то же самое, что и (6.1.16), за исключением того, что правая часть уравнения для константы равна θ_0 , а не 1, правая часть уравнения для L равна θ_1 , а не 0, правая часть уравнения для L^2 равна θ_2 , и так далее, вплоть до уравнения для L^p , правая часть которого равна θ_p . При этом часть оставшихся уравнений равна 0. Как и в случае процесса AR(p), эти уравнения могут быть легко разрешены последовательно относительно $\{\psi_0, \psi_1, \psi_2, \dots\}$, при $\theta_0 \neq 0$ имеем:

$$\psi_0 = \theta_0^{-1}, \quad \psi_1 = -\phi_1 + \phi_0\psi_1, \quad \psi_2 = \theta_2 - \phi_2 - \phi_1\psi_1 + \phi_0^2 \text{ и т.д.}$$

И здесь, как в случае AR(p), λ достаточно близка к 1 (близка к 1 для $j \geq \max\{p, q\} + 1$, так мы проверите в контрольном вопросе 51, $\{\psi_j\}$ следует тому же правилу геометрического разложения, что и процесс (6.1.7). Поскольку $\{\psi_j\}$ ограничена по абсолютному значению геометрически убывающей последовательностью λ , следовательно, абсолютно суммируема.

Снова существуют два способа получить выражение для автоковариационных функций $\{\phi_j\}$ в терминах $\{\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2\}$. Вы можете вычислить коэффициенты $\{\phi_j\}$ в МА-представлении как только что описано, а затем использовать (6.1.7). Другой метод основан на уравнениях Юла-Уокера.

ARMA(p, q) с общими корнями

Предположим, что в уравнении ARMA(p, q) (6.2.9') $\phi(L)$ удовлетворяет условию стационарности, как в утверждении 6.5, но предположим, что один из корней $\phi(z) = 0$ является также и корнем $\theta(z) = 0$. Если этот корень равен λ , полиномы $\phi(z)$ и $\theta(z)$ имеют факторизацию:

$$\phi(z) = \alpha(z)\phi^*(z), \quad \theta(z) = \alpha(z)\theta^*(z) \quad \text{с } \alpha(z) = 1 - \lambda^{-1}z.$$

Обращение для $\phi(L)$ равно $\phi^*(L)^{-1}\alpha(L)^{-1}$. Поэтому $\phi(L)$ в утверждении 6.5 можно записать как

$$\phi(L) = \phi^*(L)^{-1}\theta(L) = \phi^*(L)^{-1}\alpha(L)^{-1}\alpha(L)\theta^*(L) = \phi^*(L)^{-1}\theta^*(L). \quad (6.2.12)$$

и это показывает, что уравнение ARMA(p, q) (6.2.9') и более простое уравнение ARMA($p-1, q-1$)

$$\phi^*(L)(y_t - \mu) = \theta^*(L)\varepsilon_t$$

имеют в качестве единственного ковариационно стационарного решения один и тот же процесс¹. Из соображений экономности, уравнения ARMA с общими корнями редко используются для параметризации ковариационно стационарных процессов.

Обратимость

Для уравнения ARMA(p, q) (6.2.9) или $\phi(z) = 0$ и $\theta(z) = 0$ не имеют общих корней и если $\theta(z)$ удовлетворяет условию стабильности (6.1.18) или (6.1.18'), то ARMA процесс называется **обратимым** (*invertible*), а условие стабильности для $\theta(z)$ называется **условием обратимости** (*invertibility condition*)². Поскольку $\theta(L)^{-1}$ абсолютно суммируема, если $\theta(z)$

¹Поскольку мы определили фильтры для последовательностей действительных чисел, обсуждение в тексте предполагает, что корень λ вещественный. Если λ комплексный, то должен существовать другой общий корень, который является комплексно сопряженным для λ . Пусть $\lambda_1 = a - bi$ и $\lambda_2 = a + bi$ являются такой парой комплексных корней. Тогда, если мы определим

$$\alpha(z) = 1 - (\lambda_1^{-1} + \lambda_2^{-1})z + \frac{1}{(\lambda_1^{-1} - \lambda_2^{-1})}z^2 = 1 - \frac{2a}{(a^2 + b^2)}z - \frac{1}{(a^2 + b^2)}z^2,$$

приведенное α факти обсуждение корней верно.

²Легко спутать обратимость и существование обращения. Поскольку $\theta_0 = 1 \neq 0$, $\theta(L)$ имеет обращение без условия обратимости. Обращение, однако, может не быть абсолютно суммируемым, если не выполнено условие обратимости.

удовлетворяет условию обратимости (стабильности), мы можем умножить обе стороны ARMA-уравнения (6.2.9) на $\delta(L)^{-1}$ и использовать соотношение $\delta(L)^{-1}\phi(L)^{-1} = \phi(L)^{-1}$, чтобы получить

$$\delta(L)^{-1}\phi(L)y_t = \frac{e}{\theta(L)} + \varepsilon_t. \quad (6.2.13)$$

Таким образом, рассматриваемый процесс имеет представление в виде процесса авторегрессии бесконечной памяти (AR(∞)). Это получение не требует выполнения условия стационарности (того, что $\phi(z)$ удовлетворяет условию стабильности). Если оба $\phi(L)$ и $\theta(L)$ удовлетворяют условию стабильности, то процесс ARMA(p, q) имеет и MA(∞), и AR(∞) представления.

Производящая функция автоковариаций и спектр

Векторной формой компактной записи информации о последовательности автоковариаций $\{\gamma_j\}$ ковариационного стационарного процесса $\{y_t\}$ является **производящая функция автоковариаций** (*autocovariance-generating function*):

$$\begin{aligned} g_Y(z) &\equiv \sum_{j=-\infty}^{\infty} \gamma_j z^j = \\ &= \gamma_0 + \sum_{j=1}^{\infty} \gamma_j (z^j + z^{-j}) \quad (\text{так как } \gamma_{-j} = \gamma_j). \end{aligned} \quad (6.2.14)$$

Аргументом этой функции $g_Y(z)$ является комплексный скаляр. Поскольку суммирование включает бесконечно много слагаемых, возникает вопрос о существовании такой функции. Известный результат из математического анализа гласит о том, что указывая бесконечная сумма сходящаяся по крайней мере для $|z| = 1$ (единичная окружность), если $\{\gamma_j\}$ абсолютно суммируема. Например, для $z = 1$

$$g_Y(1) = \sum_{j=-\infty}^{\infty} \gamma_j = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j. \quad (6.2.15)$$

Это выражение очевидно не определено вследствие сходимости бесконечной суммы.

Любое комплексное число на единичной окружности может быть представлено как

$$z = \exp(i\theta) = \exp(i\omega\lambda) = e^{i\omega\lambda}, \quad (6.2.16)$$

¹ Факт: Если $\{\gamma_j\}$ абсолютно суммируема, то для суммируемых (то есть чл. н. с. сумма $\{\gamma_j\}$ сходится к конечному пределу)

где $i = \sqrt{-1}$ и ω является взятым со знаком «минус» углом z радианов, который z образует с действительной осью. Если производящая функция автокорреляций вычислена для точки z в радианах $z \in [0, 2\pi]$, результирующая функция от ω ,

$$\gamma_r(\omega) = \frac{1}{2\pi} g_Y(\cos(\omega) - i \sin(\omega)) = \frac{1}{2\pi} g_Y(e^{-i\omega}) \quad (6.2.17)$$

называется (энергетическим) спектром (*(power) spectrum*), или спектральной плотностью, функцией спектральной плотности (*spectral density function*)-процесса $\{y_t\}$, а ω называется частотой (*frequency*).

Можно сказать, что в случае абсолютно суммируемых автокорреляций все эти автокорреляции могут быть вычислены на основе спектра (отсюда и термин «производящая функция автокорреляций»). Поэтому существует взаимнооднозначное отображение между последовательностью $\{\gamma_t\}$ и $g_Y(z)$ или $g_Y(\omega)$. Таким образом, подход во временной области (*time domain approach*), который концентрируется на последовательности $\{\gamma_t\}$ и представляет именно то, что мы делали до сих пор, и подход в частотной области (*frequency domain approach*), который основан на интерпретации спектра, эквивалентны, хотя некоторые результаты могут быть более четко сформулированы по одному подходу, нежели при другом. В этой книге подход в частотной области не изучается далее того, что уже сказано.

Если $\{y_t\}$ — белый шум, то очевидно, что $g_Y(z)$ постоянна и равна дисперсии. Для MA(1) $g_Y(z)$ получается простой подстановкой (6.1.2) для $q = 1$ в общее выражение (6.2.14).

$$\begin{aligned} g_Y(z) &= \gamma_0 z^{-1} + \gamma_0 + \gamma_1 z \\ &= (\theta_1 \sigma^2) z^{-1} + (1 + \theta_1^2) \sigma^2 + (\theta_1 \sigma^2) z = \\ &= \sigma^2 \cdot [\theta_1 z^{-1} + (1 + \theta_1^2) + \theta_1 z] = \\ &= \sigma^2 \cdot (1 - \theta_1 z)(1 - \theta_1 z^{-1}). \end{aligned}$$

Поэтому

$$g_Y(z) = \sigma^2 \theta(z) \theta(z^{-1}), \quad (6.2.18)$$

где $\theta(z) = 1 + \theta_1 z$. Это последнее выражение обобщается на случай MA(q), где $\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q$. Вы можете проверить, что путем последовательного умножения в этом выражении для $g_Y(z)$ в случае MA(q), приводя подобные по степеням z и убавляя коэффициент z при z^2 и z^{-2} . Это выражение также подходит и для процесса MA(∞), если его коэффициенты абсолютно суммируемы, как нам гарантирует следующий результат.

Утверждение 6.8 (производящая функция автокорреляций для про- фильрованного процесса): Пусть $\{z_t\}$ является белым шумом и пусть $\psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \dots$ является абсолютно суммируемым фильтром (то есть последовательность $\{\psi_t\}$ абсолютно суммируема). Тогда

производящая функция автоковариаций MA(∞)-процесса $\{y_t\}$, где $y_t := \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, равна:

$$g_Y(z) = \sigma^2 \psi(z) \psi(z^{-1}). \quad (6.2.19)$$

В более общем случае, пусть $\{x_t\}$ является ковариационно стационарным процессом с абсолютно суммируемыми автоковариациями и $g_X(z)$ является производящей функцией автоковариаций для $\{x_t\}$. Производящая функция автоковариаций профильтрованного ряда $\{y_t\}$, где $y_t = h(L)x_t$ (h абсолютно суммируемой $\{h_j\}$) равна:

$$g_Y(z) = h(z) g_X(z) h(z^{-1}). \quad (6.2.20)$$

Мы в будущем докажем этот результат; см., например, теорему 4.3.1 в [Fullev, 1996].

Поскольку процессы AR(p) и ARMA(p, q) имеют представление MA(∞), их производящие функции автоковариаций могут быть легко получены из (6.2.19). Так как $\psi(L) := 1/\phi(L)$ для AR(p) и $\psi(L) := \theta(L)/\phi(L)$ для ARMA(p, q), то ихеее

$$\text{AR}(p): g_Y(z) = \sigma^2 \frac{1}{\phi(z) \phi(z^{-1})}, \quad (6.2.21)$$

$$\text{ARMA}(p, q): g_Y(z) = \sigma^2 \frac{\theta(z) \theta(z^{-1})}{\phi(z) \phi(z^{-1})}. \quad (6.2.22)$$

Контрольные вопросы

1. Если $\{y_t\}$ является ковариационно стационарным решением AR(1)-уравнения (6.2.1) с $|\alpha| < 1$, то что представляет собой

$$\tilde{E}^*(y_t, 1, y_{t-1})$$

(проекции методом наименьших квадратов)? Указание: $E(y_t - \varepsilon_t) = 0$.

Зависят ли коэффициенты этой проекции от t ? Верно ли, что $\tilde{E}^*(y_t, 1, y_{t-1}) = E(y_t | y_{t-1})$? Указание: $E(y_t | y_{t-1}) = 0$, но означает ли это, что $E(\varepsilon_t | y_{t-1}) = 0$?

Что представляет собой $\tilde{E}^*(y_t, 1, y_{t-1}, y_{t-2})$? Предположим теперь, что $\{y_t\}$ — ковариационно стационарное решение уравнения AR(1), тогда $|\alpha| < 1$. Верно ли, что $E(y_t | y_{t-1}) = 0$? (Омнег: Нет!) Верно ли, что $\tilde{E}^*(y_t, 1, y_{t-1}) = \alpha + \varepsilon_{t-1}$, как было в случае $|\alpha| < 1$?

2. (AR(1) с $\alpha = -1$) Для простоты предположим, что $\varepsilon = 0$ в AR(1)-уравнении (6.2.1). Покажите, что AR(1)-уравнение с $\alpha = -1$ не имеет ковариационно стационарного решения. Указание: Тот же самый год последовательной подстановка, который мы использовали для случая $\alpha = 1$, приводит к соотношению:

$$y_t = (-1)^j y_{t-j} - \varepsilon_t - \varepsilon_{t-1} + \dots + (-1)^{j-1} \varepsilon_{t-j+1}.$$

Из этого соотношения получаем

$$(-1)^j \rho_j = 1 - \frac{\sigma^2}{2\pi} \cdot j.$$

3. (Об утверждении 6.4.) Откуда мы знаем, что $\phi(1) \neq 0$ в (6.2.8)? Укажите: *используя условия стационарности.*

Докажите часть (б) утверждения 6.4. *Указание: Возьмите математическое ожидание от обеих сторон (6.2.6) и используйте ковариационную стационарность процесса $\{x_t\}$.*

4. (Об утверждении 6.5.) Проверьте, что $y_t = \rho + \phi(L)^{-1}\theta(L)\varepsilon_t$ является решением уравнения ARMA(p, q).

5. ($\{\varepsilon_t\}$ для ARMA(p, q).) В случае AR(p) мы использовали уравнения (6.1.12) для вычисления MA-коэффициентов $\{\varepsilon_t\}$. Запишите соответствующие уравнения для ARMA(3, 1). С какого лага j стартует последовательность $\{\varepsilon_t\}$, следующая соотношению

$$\varepsilon_t - \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} - \phi_3 \varepsilon_{t-3} = 0,$$

что является (6.1.17) для $p = 3$? Следуйте по же схеме для ARMA(1, 2). С какого лага j стартует последовательность $\{\varepsilon_t\}$, следующая соотношению (6.1.17) для $p = 12$? (Ответ: $j = 3$.)

6. (Как фильтрация изменяет спектр.) Пусть фильтр $h(L)$ абсолютно суммируем. По утверждению 5.2, если процесс $\{x_t\}$ является ковариационно стационарным с абсолютно суммируемыми автоковариациями, то это выполняется и для процесса $\{y_t\}$, где $y_t = h(L)x_t$. Используя (6.2.20), проверьте, что

$$g_Y(\omega) = h(\sigma^{-1/2})g_X(\omega)h(\sigma^{1/2}).$$

7. (Спектр является вещественнозначным.) Покажите, что спектр является вещественнозначным. *Указание: Подставьте (6.2.16) в (6.2.14). Используйте следующие факты из комплексного анализа: $\cos(\omega) = \text{real}(\omega)^2 = \cos(\omega) = -i \sin(\omega)$, $\sin(-\omega) = -\sin(\omega)$.*

6.3. Векторные процессы

Все понятия, введенные до сих пор, легко обобщить на векторные процессы.

Векторный процесс белого шума (*vector white noise process*) $\{\varepsilon_t\}$ является совместно ковариационно стационарным процессом, удовлетворяющим соотношениям

$$\begin{aligned} E(\varepsilon_t) &= 0, \quad E(\varepsilon_t \varepsilon'_s) = \Omega \text{ (положительно определена)} \\ &\text{и } E(\varepsilon_t \varepsilon'_{t-j}) = 0 \text{ для } j \neq 0, \end{aligned} \quad (6.3.1)$$

Поскольку Δt не ограничена диагональным видом, между элементами ε_t могут быть одновременные корреляции. Полная корреляция между элементами ε_t исключается, поскольку от матрицы Ω требуется положительная определенность.

Векторный МА(∞)-процесс (*vector MA(∞) process*) является одной из векторных цепочек (6.1 б):

$$z_t = \mu + \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}, \quad \varepsilon \sim \Psi_0 = I. \quad (6.3.2)$$

где $\{\Psi_j\}$ — заданная матрица. Последовательность матриц коэффициентов называется абсолютно суммируемой, если каждый элемент ψ_{ij} абсолютно суммируем. То есть, если ψ_{ij} является (i, j) -й элементом Ψ_j

$$\{\Psi_j\} \text{ — абсолютно суммируема} \Leftrightarrow \sum_{j=0}^{\infty} \|\Psi_j\| < \infty \text{ для всех } (i, j). \quad (6.3.3)$$

При таком определении утверждение 6.1 очевидным образом обобщается на многомерный случай. В частности, если $\Gamma_j := E[(y_t - \mu)(y_{t-j} - \mu)'] =$ — автоковариационная матрица y с порядка j , то выражение для автоковариаций в части (б) утверждения 6.1 принимает вид:

$$\Gamma_j = \sum_{i=0}^{\infty} \Psi_{j+i} \Omega \Psi_i' \quad (j = 0, 1, 2, \dots). \quad (6.3.4)$$

(Эта формула распространяется также и на $j = -1, -2, \dots$, поскольку $\Gamma_{-j} = \Gamma_j'$.)

Многомерный фильтр (*multivariate filter*) может быть записан как:

$$H(L) = H_0 + H_1 L + H_2 L^2 + \dots$$

где $\{H_j\}$ — последовательность (необязательно квадратных) матриц. Поэтому, если H_{ij} является (i, j) -й элементом H_j , (k, l) -й элемент $H(L)$ равен:

$$H_{kl} + H_{kl} L + H_{kl} L^2 + \dots$$

Многомерная версия утверждения 6.2 выглядит с $y_t = H(L)x_t$.

Произведение фильтров

Пусть $A(L)$ и $B(L)$ — два фильтра, где A_j имеет размер $m \times r$, а B_j имеет размер $r \times n$, так что можно определить произведение матриц $A_j B_k$. Произведение двух фильтров, $D(L) = A(L)B(L)$, является оператором

размера $n \times n$, последовательность матриц коэффициентов которого $\{D_j\}$ задается m -мерной версией формулы свертки (6.1.12):

$$\begin{aligned} A_0 B_0 &= D_0, \\ A_0 B_1 + A_1 B_0 &= D_1, \\ A_0 B_2 + A_1 B_1 + A_2 B_0 &= D_2, \\ &\dots \\ A_0 B_j + A_1 B_{j-1} + A_2 B_{j-2} + \dots + A_{j-1} B_1 + A_j B_0 &= D_j, \\ &\dots \end{aligned} \quad (6.3.5)$$

Обращение

Пусть $A(L)$ и $B(L)$ — две функции, матрицы коэффициентов которых квадратны. $B(L)$ называется **обратным** (*inverse*) для $A(L)$ и обозначается $A(L)^{-1}$, если

$$A(L) B(L) = I \quad (6.3.6)$$

Для любой произвольной последовательности квадратных матриц $\{A_j\}$ обращение для $A(L)$ существует, если A_0 невырождена. Легко показать, что обращения коммутируют (см. вспомогательный лемму 2), поэтому

$$A(L) A(L)^{-1} = A(L)^{-1} A(L).$$

Абсолютно суммируемые обратимые элементы кольца

Леговый полином p -й степени (*p -th degree lag matrix polynomial*) имеет вид:

$$\Phi(L) = I - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p, \quad (6.3.7)$$

где $\{\Phi_i\}$ является последовательностью $n \times n$ матриц с $\Phi_p \neq 0$. Как следует из теории однородных разностных уравнений, условие стабильности для m -мерного случая заключается в следующем.

Все корни детерминантного уравнения

$$|I - \Phi_1 z - \dots - \Phi_p z^p| = 0 \quad (6.3.8)$$

больше единицы по абсолютной величине (то есть лежат вне единичного круга).

Для удобства пишем

Все корни детерминантного уравнения

$$Iz^p - \Phi_1 z^{p-1} - \dots - \Phi_p = 0 \quad (6.3.8')$$

меньше единицы по абсолютному значению (то есть лежат внутри единичного круга).

В качестве примера рассмотрим лаговый полином первого порядка $\Phi(L) = I - \Phi_1 L$, где $\Phi_1 = (\phi_{ij})$ является 2×2 -матрицей. Уравнение (6.3.8) может быть записано как

$$\begin{bmatrix} 1 - \phi_{11}z & -\phi_{12}z \\ -\phi_{21}z & 1 - \phi_{22}z \end{bmatrix} \cdot \begin{bmatrix} z \\ 1 \end{bmatrix} = (\phi_{11} + \phi_{22})z + (\phi_{11}\phi_{22} - \phi_{12}\phi_{21})z^2 = 0.$$

С обобщенным таким образом условием стабильности утверждение 6.3 обобщается следующим образом. Пусть $\Phi(L) = \Phi(L)^{-1}$. Каждая компонента последовательности матриц коэффициентов $\{\Psi_j\}$ будет ограничена по абсолютному значению геометрически убывающей последовательностью.

Многомерным аналогом процесса AR(p) является процесс **векторной авторегрессии** p -го порядка (**VAR(p)**) (*vector autoregressive process of p -th order*). Он является единственным ковариационно стационарным решением при условии стационарности (6.3.6) для следующего векторного стохастического разностного уравнения:

$$y_t - \Phi_1 y_{t-1} - \dots - \Phi_p y_{t-p} = \varepsilon_t + e_t \quad \text{или} \quad \Phi(L)y_t = \varepsilon_t, \\ \text{где } \Phi(L) = I_p - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p \quad \text{и} \quad \mu = \Phi(1)^{-1}e. \quad (6.3.9)$$

Здесь $\Phi_p \neq 0$. Например, двучленный процесс VAR(1) может быть записан в явном виде как система двух уравнений с общей регрессоркой

$$\begin{aligned} y_{1t} &= \alpha_1 + \phi_{11}y_{1,t-1} + \phi_{12}y_{2,t-1} + \varepsilon_{1t}, \\ y_{2t} &= \alpha_2 + \phi_{21}y_{1,t-1} + \phi_{22}y_{2,t-1} + \varepsilon_{2t}, \quad \text{где } \Phi_1 = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}. \end{aligned}$$

В более общем случае p -членный VAR(p) можно записать как множество из p уравнений с $p+1$ общими регрессорами. Утверждение 6.4 непосредственно обобщается на m -мерный случай. В частности, каждый элемент Γ_j ограничен по абсолютному значению геометрически убывающей последовательностью.

Векторные авторегрессии являются очень популярным инструментом для анализа динамических взаимосвязей между элементарными макропеременными. Подробный анализ можно найти в [Hamilton, 1994, Chapter 11].

Многомерным аналогом ARMA(p, q) служит **векторный ARMA(p, q)** (**VARMA(p, q)**) (*vector ARMA(p, q)*). Он является единственным ковариационно стационарным решением при условии стационарности для сле-

дующего векторного стохастического разностного уравнения.

$$\begin{aligned} y_t = & \varepsilon_t + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + \varepsilon_t + \\ & + \Theta_1 \varepsilon_{t-1} + \Theta_2 \varepsilon_{t-2} + \dots + \Theta_q \varepsilon_{t-q} \quad (6.3.9) \\ & \Phi(L)(y_t - \mu) = \Theta(L)\varepsilon_t, \end{aligned}$$

где

$$\begin{aligned} \Phi(L) = I - \Phi_1 L - \dots - \Phi_p L^p, \quad \Theta(L) = I + \Theta_1 L + \dots + \Theta_q L^q, \\ \mu = \Phi(1)^{-1} \varepsilon. \end{aligned} \quad (6.3.10)$$

и при этом $\Phi(L)_p \neq 0$ и $\Theta_q \neq 0$. Утверждение 6.5 обобщается очевидным способом

Применимая функция спектограммы

Многомерная версия производящей функции автоковариаций для векторного ковариационно-стационарного процесса $\{y_t\}$ имеет вид

$$G_Y(z) = \sum_{j=-\infty}^{\infty} \Gamma_j z^j = \Gamma_0 + \sum_{j=1}^{\infty} (\Gamma_j z^j + \Gamma_j' z^{-j}) \quad (\text{поскольку } \Gamma_{-j} = \Gamma_j'), \quad (6.3.11)$$

Спектр векторного процесса $\{y_t\}$ определяется как

$$S_Y(\omega) = \frac{1}{2\pi} G_Y(e^{-i\omega}) \quad (6.3.12)$$

Утверждение 6.6 легко обобщается. В частности, если $H(L)$ является абсолютно суммируемым фильтром $n \times k$, а $G_X(z)$ является производящей функцией автоковариаций n -мерного ковариационно-стационарного процесса $\{x_t\}$ с абсолютно суммируемой автоковариационной производящей функцией автоковариаций k для $y_t = H(L)x_t$, задается как

$$G_Y(z) = H(z) G_X(z) H(z^{-1})', \quad (6.3.13)$$

В качестве частных случаев этого соотношения мы имеем

$$\text{векторный белый шум: } G_Y(z) = \Omega, \quad (6.3.14)$$

$$\text{VAR}(x_t): G_Y(z) = \Psi(z) \Omega \Psi(z^{-1})', \quad (6.3.15)$$

$$\text{VAR}(y_t): G_Y(z) = |\Phi(z)|^{-1} \Omega |\Phi(z^{-1})|^{-1'}, \quad (6.3.16)$$

$$\text{VARMA}(p, q): G_Y(z) = |\Phi(z)|^{-1} [\Theta(z)] \Omega [\Theta(z^{-1})' |\Phi(z^{-1})|^{-1}]. \quad (6.3.17)$$

Напомним, если $y_t = \varepsilon_t + \Theta_1 \varepsilon_{t-1}$,

$$G_Y(z) = (I + \Theta(z) \Omega (I + \Theta(z^{-1})'$$

Контрольные вопросы

1. Проверьте (6.34) для авторегрессии $MA(1)$.
2. (Коммутативность обращения.) Проверьте коммутативность обращения, то есть « $A(L)B(L) = I$ » $\Leftrightarrow$ « $B(L)A(L) = I$ » для двух квадратных фильтров размерности $n \in \mathbb{N}$, $A_n \neq 0$, $B_n \neq 0$. Указание: Положите $D_n = I$, $D_j = 0$ ($j \geq 1$) в системе (6.35) и решите ее относительно B . Покажите, что этот $B(L)$ удовлетворяет соотношению $B(L)A(L) = I$.
3. (Отсутствие коммутативности для многомерных фильтров.) Предложите пример: для $A(L)$ и $B(L)$ — 2×2 , но $A(L)B(L) \neq B(L)A(L)$. Указание: Как выбрать $A(L) = I + A_1L$ и $B(L) = I + B_1L$? Найдите две квадратные матрицы A_1 и B_1 , такие что $A_1B_1 \neq B_1A_1$.
4. Проверьте многомерную версию (6.1.10), а именно:

$$\begin{aligned} \bullet A(L)B(L) = D(L) &\Leftrightarrow \bullet B(L) = A(L)^{-1}D(L) \Leftrightarrow \\ &\Leftrightarrow \bullet A(L) = D(L)B(L)^{-1}, \end{aligned}$$

при условии невырожденности A_n и B_n . Поэтому, чтобы решить $A(L)B(L) = D(L)$, используйте обе стороны равенств на $A(L)^{-1}$.

5. Покажите, что $[A(L)B(L)]^{-1} = B(L)^{-1}A(L)^{-1}$ при условии невырожденности A_n и B_n . Указание: По определению,

$$A(L)B(L)[A(L)B(L)]^{-1} = I.$$

6.4. Оценивание авторегрессий

Авторегрессионные процессы (авторегрессии) популярны в эконометрике, не только потому что они имеют естественную интерпретацию, но также потому, что они легко оцениваются. Этот параграф рассматривает оценивание авторегрессий. Оценивание ARMA-процессов будет затронуто в конце параграфа.

Оценивание AR(1)

Вспомним, что $AR(1)$ является $MA(\infty)$ с абсолютно суммируемыми коэффициентами. Поэтому, если $\{z_t\}$ является независимым белым шумом (последовательностью i.i.d. случайных величин с нулевым средним и конечной дисперсией), он является строго стационарным и эргодическим (утверждение 6.1(d)). Полагая $x_t = (1, y_t, y_t^2, \dots)$ и $\beta = (1, \rho, \rho^2, \dots)$, $AR(1)$ -уравнение (6.2.1) можно записать как уравнение регрессии

$$y_t = x_t'\beta + z_t. \quad (6.4.1)$$

Мы предполагаем, что выборка состоит из $(y_0, y_1, y_2, \dots, y_n)$, включая y_0 , так что (6.4.1) для $t = 1$ (которое имеет y_0 в правой части) может быть включено в оценивание. Поэтому период выборки с $t = 1$ по n

Мы теперь покажем, что все условия утверждения 2.5 об асимптотических свойствах OLS-оценки для β при наличии условной гомоскедастичности здесь удовлетворяются. Во первых, очевидно, что линейность (предположение 2.1) удовлетворяется. Во вторых, как только что было замечено, процесс $\{y_t, x_t\}$ совместно стационарный и эргодический (предположение 2.2). В-третьих, поскольку y_{t-1} (независимый регрессор в x_t) является функцией от $\{y_{t-1}, x_{t-1}, \dots\}$, он не зависит от ошибки ε_t по предположению 1.1.б. для $\{e_t\}$. Таким образом,

$$E(\varepsilon_t^2 | x_t) = \sigma^2, \quad (6.4.2)$$

то есть ошибка условно гомоскедастична (предположение 2.7). В-четвертых, если $g_t = x_t' \varepsilon_t = (x_t, y_{t-1}, \varepsilon_t)'$, то $\{g_t\}$ является мартингал-разностью, как требуется в предположении 2.5. Поскольку

первый элемент $E(g_t | g_{t-1}, g_{t-2}, \dots) =$

$$\begin{aligned} & E(\varepsilon_t | x_{t-1}, y_{t-2}, \dots) \\ &= E(\varepsilon_t | x_{t-1}, x_{t-2}, \dots, x_{t-2}, y_{t-1}, y_{t-2}, \dots) = \\ &= 0 \quad (\text{поскольку } \{\varepsilon_t\} \text{ является i.i.d. и } y_{t-1} \text{ является функцией от} \\ & \quad (x_{t-1}, x_{t-2}, \dots)) \end{aligned}$$

и

второй элемент $E(g_t | g_{t-1}, g_{t-2}, \dots) =$

$$\begin{aligned} &= E(\varepsilon_t | y_{t-1}, g_{t-2}, \dots) = \\ &= E[E(\varepsilon_t | y_{t-1}, g_{t-2}, \dots, g_{t-3}, \dots) | g_{t-2}, \dots] \\ & \quad (\text{по закону композиции математических ожиданий}) \\ &= E[g_{t-1} | E(\varepsilon_t | y_{t-1}, g_{t-2}, \dots) | g_{t-2}, \dots] = \\ &= 0 \quad (\text{так как } E(\varepsilon_t | y_{t-1}, x_{t-1}, x_{t-2}, \dots, g_{t-2}, y_{t-1}, y_{t-2}, \dots) = 0). \end{aligned} \quad (6.4.3)$$

В частности, $E(\varepsilon_t | y_{t-1}) = 0$, поэтому ε_t ортогонален ошибке (предположение 2.3). Наконец, ранговое условие в матрице коварианс $E(x_t x_t')$ (предположение 2.4) удовлетворяется, поскольку определитель для

$$E(x_t x_t') = \begin{bmatrix} 1 & 0 \\ 0 & \gamma_0 + \sigma^2 \end{bmatrix} \quad (6.4.4)$$

равен $\gamma_0 > 0$. Это также означает, что матрица $E(y_t y_t')$ невырождена, поскольку γ_0 — условной гомоскедастичности $E(y_t y_t') = \sigma^2 E(x_t x_t')$. Поэтому удовлетворяется условие невырожденности в предположении 2.5.

Таким образом, AR(1) с процессом независимого белого шума удовлетворяет всем предположениям утверждения 2.5, так что части (a)–(c)

этого утверждения остаются здесь верными с $(c, \phi)'$ в качестве β , $(1, y_{t-1})'$ в качестве x_t и σ^2 в качестве дисперсии ошибок. В частности, OLS-оценка дисперсии ошибок

$$s^2 = \frac{1}{n-2} \sum_{t=1}^n e_t^2, \quad e_t = y_t - \hat{c} - \hat{\phi} y_{t-1}, \quad \text{где } (\hat{c}, \hat{\phi}) - \text{OLS-оценки,}$$

является состоятельной для σ^2 .

Оценивание AR(p)

Аналогично AR(1), авторегрессионные коэффициенты $(\phi_1, \dots, \phi_p)$ можно оценить состоятельно, оценивая регрессию y_t на константу и лагированные y_t ($y_{t-1}, \dots, y_{t-p}$). Мы предполагаем, что выборка состоит из $(y_{-p+1}, y_{-p+2}, \dots, y_0, y_1, \dots, y_n)$, исключая p лагированных значений, предшествующих y_1 , так что уравнение AR(p) может быть записано начиная с $t = 1$ и оценивание производится по периоду с $t = 1$ до n .

Утверждение 6.7 (оценивание AR-коэффициентов): Пусть $\{y_t\}$ является AR(p)-процессом, следующим (6.2.6) с условием стационарности (6.1.18). Предположим далее, что $\{e_t\}$ является независимым белым шумом. Тогда OLS-оценка для $(c, \phi_1, \phi_2, \dots, \phi_p)$, полученная путем регрессии y_t на константу и p лагированных значений y_t для периода выборки $t = 1, 2, \dots, n$, является состоятельной и асимптотически нормальной. Если $\beta = (c, \phi_1, \phi_2, \dots, \phi_p)'$, $x_t = (1, y_{t-1}, \dots, y_{t-p})'$, и $\hat{\beta}$ — OLS-оценка для β , то мы имеем:

$$\text{Avar}(\hat{\beta}) = \sigma^2 E(x_t x_t')^{-1},$$

что состоятельно оценивается как

$$\widehat{\text{Avar}}(\hat{\beta}) = s^2 \cdot \left(\frac{1}{n} \sum_{t=1}^n x_t x_t' \right)^{-1},$$

где s^2 — OLS-оценка дисперсии ошибок, заданная как

$$s^2 = \frac{1}{n-p-1} \sum_{t=1}^n (y_t - \hat{c} - \hat{\phi}_1 y_{t-1} - \dots - \hat{\phi}_p y_{t-p})^2. \quad (6.4.5)$$

Доказательство заключается в применении утверждения 2.5, что может быть сделано аналогично тому, как мы это делали для AR(1). Трудность состоит в том, чтобы показать, что матрица $E(x_t x_t')$ невырождена (предположение 2.4), что эквивалентно требованию того, чтобы автоковариационная матрица $\text{Var}(y_t, \dots, y_{t-p+1})$ размерности $p \times p$ являлась

невырожденной (см. контрастный возраст 1[6] и же). Можно показать, что автокорреляционная матрица ковариационного стационарного процесса является невырожденной для любого p , если $\gamma_0 > 0$ и $\gamma_j \rightarrow 0$ при $j \rightarrow \infty$. Поэтому предположение 2.4 также выполняется.

Мы не рассматривали оценивание $AR(p)$ методом максимального правдоподобия, но для тех случаев, кто интересуется связью с ML, OLS-оценка численно совпадает с оценкой условного гауссовского ML и асимптотически эквивалентна оценке безусловного гауссовского ML. См. параграф 8.7.

Выбор глубины запаздываний

Все эти приятные результаты об оценивании авторегрессий предполагают, что порядок (глубина запаздываний) авторегрессии p известен. Как мы должны действовать, если порядок p неизвестен? Прежде всего ясно, что утверждение 6.7, которое основывается на авторегрессиях p -го порядка, применимо к авторегрессиям более низкого порядка, скажем, $r < p$. То есть, даже если $\phi_1 \neq 0$, но $\phi_{r+1} = \phi_{r+2} = \dots = \phi_p = 0$, это утверждение по-прежнему применимо, до тех пор пока $(\phi_1, \phi_2, \dots, \phi_r)$ удовлетворяет условию стационарности. В частности, OLS-оценки для $(\phi_1, \phi_2, \dots, \phi_r)$ будут сходиться к истинным нулевым значениям. Чтобы представить то же доказательство иначе, предположим, что истинный порядок авторегрессии равен p (так что $\phi_p \neq 0$), и предположим, что мы знаем о p то, что он меньше или равен некоторому известному целому числу $p_{\max}$. Полагая r и только что r и используя доказательство равенств p и p равным $p_{\max}$, мы видим, что утверждение 6.7 применимо к авторегрессиям с $p_{\max}$ лагами. Если $\{y_t\}$ является стационарным $AR(p)$ -процессом и если мы оцениваем регрессию y_t на константу и $p_{\max}$ лагов от y_t , то OLS-оценки коэффициентов будут состоятельными и асимптотически нормальными.

Естественно думать, что истинная глубина запаздываний может быть оценена по указанной OLS-оценке для $(\phi_1, \dots, \phi_{p_{\max}})$. Мы рассмотрим для класса правил для определения глубины запаздываний. Первое — это

Последовательное t -критерий от общего к частному: Начните с авторегрессии с $p_{\max}$ лагами. Если последний лаг значим на некотором предварительно специфицированном уровне значимости (скажем, 10%), то закончите процедуру и поместите глубину запаздываний равной $p_{\max}$. В противном случае удалите последний лаг, переоцените авторегрессию с меньшим на единицу количеством лагов и повторите тот же самый тест. Если такой процесс продолжается, то так пор пока в авторегрессию остается

¹ См., например, Proposition 5.1.1 в [Box-Jenkins and Pierce, 1991]

только один лаг, и если этот лаг является незначимым, то тогда продолжите глубину запаздываний равной 0 (отсутствие лагов).

Это правило обладает следующими свойствами, которые будут показаны через мгновение. Поскольку t -тест является состоятельным, глубина запаздываний, выбранная таким образом, никогда не будет меньше, чем p (истинная глубина запаздываний) на больших выборках. Однако вероятность избыточности (того, что глубина запаздываний выбрана большей, чем истинная глубина запаздываний p) отлична от нуля даже на больших выборках. То есть если $\hat{p}$ — глубина запаздываний, выбранная указанным последовательным t -правилом, то

$$\lim_{n \rightarrow \infty} \text{Prob}(\hat{p} < p) = 0 \quad \text{и} \quad \lim_{n \rightarrow \infty} \text{Prob}(\hat{p} > p) > 0. \quad (5.4.6)$$

Чтобы пояснить это, предположим, что $p = 2$ и $p_{\max} = 3$. Последовательное t -правило начинается с трех лагов в авторегрессии:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \varepsilon_t.$$

Мы проверяем нулевую гипотезу о том, что $\phi_3 = 0$, на 10%-ном уровне значимости. С вероятностью 10% мы отвергаем (верную) нулевую гипотезу и полагаем глубину запаздываний равной 3 или принимаем нулевую гипотезу с вероятностью 90% на больших выборках. В случае принятия нулевой гипотезы мы полагаем $\phi_3 = 0$, оцениваем авторегрессию с двумя лагами и тестируем гипотезу о том, что $\phi_2 = 0$. Поскольку в действительности $\phi_2 \neq 0$ (так как $p = 2$), абсолютное значение t -ста-дистика, основанной на OLS-оценке для ϕ_2 , может быть очень велико на больших выборках, так что мы никогда не примем (неверную) нулевую гипотезу о том, что второй лаг нулевой. Следовательно, в этом примере $\text{Prob}(\hat{p} = 2) = 90\%$ и $\text{Prob}(\hat{p} = 3) = 10\%$ на больших выборках.

Есть две незначительные вариации в выборе периода выборки с данными $(y_1, \dots, y_n)$. Одна всегда использует одинаковый период $t = p_{\max} - 1, p_{\max} - 2, \dots, n$. Другая использует период выборки «эластичным», расширяя его при все меньшем и меньшем количестве включенных лагов. То есть, когда оцениваемая авторегрессия имеет j лагов, периодом выборки является $t = j + 1, j + 2, \dots, n$.

Второе правило для выбора глубины запаздываний использует как **информационный критерий Акаике (AIC)** (*Akaike information criterion*) или **информационный критерий Шварца (SIC)** (*Schwartz information criterion*), также известный как **байесовский информационный критерий (BIC)** (*Bayesian information criterion*). Эта процедура полагает глубину запаздываний равной значению j , которое минимизирует

$$\log(\text{SSR}_j/n) + (j-1)C(n)/n \quad (5.4.7)$$

по $j = 0, 1, 2, \dots, p_{\max}$, где SSR_j является суммой квадратов остатков для авторегрессии с j лагами. В этой формуле « $j + 1$ » входит в качестве числа параметров (коэффициентов при константе и j лагов). Для AIC $C(n) = 2$, в то время как для BIC $C(n) = \log(n)$. Несколько комментариев об этом правиле, основанном на информационных критериях:

- Поскольку количество параметров возрастает вместе с j , термин, слогаемое целой функции (6.4.7) уменьшается, так как подгонка уравнения становится лучше, но второй слагаемый увеличивается. Информационный критерий обеспечивает баланс между лучшей подгонкой и сложностью модели.
- Без верхней границы $p_{\max}$ может быть выбрано сколь угодно большое значение j (очевидно, информационный критерий достигает минимума $-\infty$ при $j = 1 - n$, поскольку $SSR_j = 0$ для $j = n - 1$). В текущем контексте, где истинный DGP является авторегрессией конечного порядка, разумной верхней границей является максимально возможная глубина запаздываний.
- И здесь существуют два способа устанавливать период выборки. Вы можете использовать фиксированный период выборки $t = p_{\max} + 1, p_{\max} + 2, \dots, n$, в этом случае SSR_j равна сумме $n - p_{\max}$ квадратов остатков. Или вы можете использовать «эластичный» период выборки, в этом случае SSR_j равна сумме $n - j$ квадратов остатков. В любом случае есть еще один вариант, который заключается в замене n в информационном критерии (6.4.7) на фактически: 2 объем выборки. Поэтому вы заменяете n на $n - p_{\max}$, если используется фиксированная выборка. На основе эмпирических и некоторых теоретических соображений [Ng and Perron, 2000] рекомендуют использовать фиксированный, а не «эластичный» период выборки, с заменой n на $n - p_{\max}$ в информационном критерии.

Пусть $\hat{\rho}_{AIC} \neq \hat{\rho}_{BIC}$ являются количествами запаздываний, выбранными посредством AIC и BIC соответственно. Как и глубина запаздываний, выбранная посредством t -правила от общего к частному, они есть функции от данных и, следовательно, являются случайными величинами. Для них можно показать следующее (см., например: [Lütkepohl, 1993, Section 4.3]):

- (i) $\hat{\rho}_{AIC} \leq \hat{\rho}_{BIC}$, если объем выборки не очень мал. (Если используется фиксированная выборка $t = p_{\max} + 1, \dots, n$ и $m = p_{\max}$ заменяет n в (6.4.7), это верно для любой конечной выборки с $n \geq p_{\max} + 8$.) Это является алгебраическим результатом и выполняется для любой выборки y .

(2) Предположим, что $\{y_t\}$ — стационарный $AR(p)$ и ε_t — независимый белый шум с конечным четвертым моментом. Тогда

$$\lim_{n \rightarrow \infty} \hat{\rho}_{AIC} = \rho, \quad \lim_{n \rightarrow \infty} \text{Prob}(\hat{\rho}_{AIC} < \rho) = 0, \quad \lim_{n \rightarrow \infty} \text{Prob}(\hat{\rho}_{AIC} > \rho) > 0. \quad (6.4.8)$$

То есть, точно так же, как и последовательное t -правило от общего к частному, правило AIC имеет положительную вероятность завышения значения ρ . В противоположность им правило BIC является состоятельным. Кроме того, [Наппа и Дефет 1988, Theorem 5.4.1(c) and Remark 1] показали, что состоятельность $\hat{\rho}_{AIC}$ выполняется, когда верхняя граница ρ_{max} делится возрастающей со скоростью $\log(n)$ (то есть когда она падает быстрее, чем $[\log(n)]$ [целой частью $\leq \log(n)$] для любого заданного $\varepsilon > 0$). Это означает, что вам не нужно знать верхнюю границу для состоятельного оценивания порядка авторегрессии посредством BIC.

Оценивание VAR

Оценивание коэффициентов VAR столь же просто. Если y_t является M -мерным вектором, VAR(p) (6.3.9) является системой из M уравнений и может быть записана в виде:

$$y_{tm} = x_t' \delta_{tm} + \varepsilon_{tm}, \quad (m = 1, 2, \dots, M). \quad (6.4.9)$$

где y_{tm} является m -м элементом в y_t , δ_{tm} — m -м элементом в δ_t , и

$$x_t = \begin{bmatrix} 1 \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{bmatrix}, \quad \delta_{tm} = \begin{bmatrix} c_m \\ \phi'_{1m} \\ \phi'_{2m} \\ \vdots \\ \phi'_{pm} \end{bmatrix},$$

c_m — m -й элемент вектора c , ϕ'_{jtm} — m -я строка матрицы Φ_j .

Эти M уравнений имеют одно и то же множество регрессоров. Легко проверить, что x_t ортогональны ошибке в каждом уравнении и что ошибки условно гомосkedастичны (доказательство очень похоже на доказательство для $AR(p)$). Таким образом, указанная система из M уравнений с i.i.d. $\{\varepsilon_t\}$ удовлетворяет условиям многомерной регрессионной модели из параграфа 4.5. Поэтому эффективное GMM-оценивание осуществляется очень просто: достаточно применить OLS поочередно к каждому уравнению системы. Выражения для асимптотической дисперсии и ее состоятельной оценки могут быть получены из утверждения 4.6 с $x_{itm} = x_t$ (так что (4.5.13') на с. 309 становится равным $\frac{1}{n} \sum_t x_t x_t'$). Если δ является $(Mp - 1)M$ -мерным штабелированным вектором, созданным

из $(\delta_1, \dots, \delta_M)$, и если $\tilde{\delta}$ является оценкой для δ , получаемой последовательным применением OLS к отдельным уравнениям, то мы имеем из (4.5.17):

$$\widehat{\text{Var}}(\tilde{\delta}) = \tilde{\Omega} * \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1}, \quad (6.4.10)$$

где

$$\tilde{\Omega} \equiv \frac{1}{n - Mp - 1} \sum_{i=1}^n \tilde{\varepsilon}_i \tilde{\varepsilon}_i', \quad \tilde{\varepsilon}_i = y_i - \tilde{\varepsilon} - \tilde{\Phi}_1 y_{i-1} - \dots - \tilde{\Phi}_p y_{i-p}, \quad (6.4.11)$$

Если мы не знаем глубину запаздываний p , оцениваем ее по данным. В принципе, основанном на информационном критерии, глубина запаздываний, которую мы выбираем, равна значению k , которое минимизирует

$$\log \left(\left| \frac{1}{n} \sum_{i=1}^n \tilde{\varepsilon}_i \tilde{\varepsilon}_i' \right| \right) + (kMp' + Mp') \cdot \frac{C(p')}{n}, \quad (6.4.12)$$

по $p' = 0, 1, \dots, p_{\max}$. Здесь, как и в одномерном случае, $C(p') = 2$ для AIC и $\log(n)$ для BIC. Точно такие же результаты, перечисленные ранее как (1) и (2) для одномерного случая, выполняются и для многомерного случая (см., например: [Lütkepohl, 1993, Section 4.3]).

Оценивание ARMA(p, q)

Положим

$$y_t \equiv z_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}, \\ z_t = (1, y_{t-1}, y_{t-2}, \dots, y_{t-p})', \quad \delta \equiv (\varphi_1, \varphi_2, \dots, \varphi_p)'$$

одномерное уравнение ARMA(p, q) (6.2.9) можно записать в виде:

$$z_t = \varphi_t' \delta + u_t, \quad (6.4.13)$$

Компоненты скалярного сдвига в ARMA создают две проблемы. Во-первых, оценивая, что ошибки u_t серияльно коррелированы (фактически они образуют процесс $MA(q)$). Во-вторых, поскольку лагированные зависимые переменные, включенные в z_t , коррелированы с самим z_t , включенным в ошибки, регрессоры z_t не являются ортогональными ошибке u_t . То есть серияльная корреляция при наличии лагированных зависимых переменных приводит к смещенному оцениванию. Со второй проблемой можно было бы справиться с помощью подбора лагированных зависимых переменных $\{y_{t-1}, y_{t-2}, \dots\}$ в качестве инструментов. Что касается первой проблемы, метод, который будет разработан позднее в этой главе, предоставляет состоятельные оценки вектора коэффициентов при

наличии сериальной корреляции. При задании достоверной оценки δ , МА-параметры (θ -ы) могут быть оценены по остаткам. Эта процедура, однако, не является эффективной¹. Мы не будем рассматривать такую эффективную оценивающую процедуру ARMA(p, q) с гауссовскими ошибками. Интересующиеся читатели могут обратиться, например, к [Fuller, 1996, Chapter 8] или [Pamilton, 1994, Chapter 5].

Контрольные вопросы

1. (Август для AR-коэффициентов)

(а) Для AR(1) покажите, что

$$\text{Avar}(\hat{\phi}) = \frac{\sigma^2}{n} (1 - \sigma^2),$$

где $\hat{\phi}$ — OLS-оценка для ϕ . Указание: Это равно умноженному на σ^2 (2.2)-му элементу обращения (6.4.4).

(б) Для AR(p) пусть $\hat{\phi}$ является OLS-оценкой для $(\phi_1, \phi_2, \dots, \phi_p)'$. Покажите, что

$$\text{Avar}(\hat{\phi}) = \sigma^2 V^{-1},$$

где V является автоковариационной матрицей размера $p \times p$:

$$V = \text{Cov}(\hat{y}_1, \dots, \hat{y}_{p+1}) = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \dots & \gamma_{p-2} \\ \vdots & \gamma_1 & \gamma_2 & \dots & \vdots \\ \gamma_{p-2} & \dots & \gamma_1 & \gamma_0 & \gamma_1 \\ \gamma_{p-1} & \dots & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix}.$$

Указание: Если x_i является тем же, что и в утверждении 6.7, то

$$E(x_i x_j') = \begin{bmatrix} 1 & \mu 1' \\ \mu 1 & \mu^2 11' \end{bmatrix},$$

где $1 = p \times 1$ вектор из единиц. Покажите, что обращение этой матрицы задается как

$$A = \begin{bmatrix} \sigma & c' \\ c & V^{-1} \end{bmatrix}, \text{ где } \sigma = 1 - \mu^2 1' V^{-1} 1, \text{ а } c = -\mu V^{-1} 1.$$

Это также показывает, что матрица $E(x_i x_j')$ является невырожденной тогда и только тогда, когда это выполняется для V . (Если V невырожденная, то $E(x_i x_j')$ невырожденная из-за ее обратимости. Если V вырожденная, то существует p -мерный вектор $d \neq 0$, такой что $Vd = 0$ и $E(x_i x_j')f = 0$, где $f' = (-\mu 1'd, d')$.)

¹Эта модель приводит к бесконечно большому количеству условий ортогональности. $E(u_t y_{t-j}) = 0$ для всех $j \leq p - q - 1$. Процедура GMM может использовать только конечное количество условий ортогональности.

2. (Уменьшение коэффициентов VAR(1)) Рассмотрим VAR(1) без константы $y_t = A y_{t-1} + \varepsilon_t$. Проверьте, что верно для A в одномерной регрессии при регрессии на выборки $\{y_1, y_2, \dots, y_n\}$ (так что верная выборка — $t = 2, 3, \dots, n$) равна:

$$\hat{A} = \left(\sum_{t=2}^n y_t y_{t-1}' \right) \left(\sum_{t=2}^n y_{t-1} y_{t-1}' \right)^{-1}.$$

Указание: Постройте автофакторизацию и покажите, что равен n -й строке A .

6.5. Асимптотика для выборочных средних сериально коррелированных процессов

Этот параграф изучает асимптотические свойства выборочного среднего

$$\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t$$

для сериально коррелированных процессов. (Следует иметь в виду, что $\bar{y}$ зависит от n , хотя обозначение не указывает на это явно.) Для состоятельности выборочного среднего $\bar{y}$ уже есть достаточное условие в функциональной теореме главы 4. Первое утверждение в этом параграфе предоставляет другое достаточное условие в форме предельных нормальных стационарных процессов. Мы также предлагаем две CLT для сериально коррелированных процессов, одну — для линейных процессов, а другую — для эргодических стационарных процессов. Глава 2 включает CLT для эргодических стационарных процессов, но она не учитывает сериальную корреляцию, поскольку предполагается, что процесс является мартингал разностью. CLT для эргодических стационарных процессов этим параграфом обобщается, добавляя сериальную корреляцию.

LLN для ковариационно стационарных процессов

Утверждение 6.8 (LLN для ковариационно стационарных процессов с затухающими автоковариациями): Пусть $\{y_t\}$ является ковариационно стационарным процессом с ковариационными ожиданиями μ и γ_j являются автоковариациями процесса $\{y_t\}$. Тогда

(a) $\bar{y} \rightarrow_{\text{пл.в.}} \mu$ при $n \rightarrow \infty$, если $\lim_{n \rightarrow \infty} \gamma_j = 0$.

(б)

$$\lim_{n \rightarrow \infty} \text{Var}(\sqrt{n} \bar{y}) = \sum_{j=-\infty}^{\infty} \gamma_j < \infty, \quad \text{если } \{\gamma_j\} \text{ суммируема}^1. \quad (6.5.1)$$

¹Anderson (1971, Lemma 3.3.1, p. 460) Аббревиатура суммируемости $\{\gamma_j\}$ не является проблемой

Поскольку из сходимости в среднем квадратическом вытекает сходимость по вероятности, часть (а) показывает, что достаточно очень слабых условий на ковариационно-стационарный процесс, чтобы выборочные средние были состоятельными для μ . Для доказательства части (а) по слабому закону больших чисел Чебышева (см. параграф 2.1), достаточно показать, что $\lim_{n \rightarrow \infty} \text{Var}(\bar{y}) = 0$, что довольно легко (аналитическое выражение (8)), после того как получено выражение для $\text{Var}(\bar{y})$. В следующем параграфе с опущением серьезной корреляции в $\{y_t\}$, вычисления упрощаются для $\bar{y}$ является более провозвешен, поскольку должны быть приняты во внимание ковариации между двумя различными слагаемыми:

$$\begin{aligned} n \cdot \text{Var}(\bar{y}) &= \text{Var}(\sqrt{n} \bar{y}) = \\ &= \frac{1}{n} [\text{Cov}(y_1, y_1) + \dots + y_n] + \dots + \text{Cov}(y_n, y_1 + \dots + y_n) = \\ &= \frac{1}{n} [(y_0 - y_1) + \dots + (y_{n-1} - y_n)] + (y_1 + y_2 + \dots + y_{n-1}) + \\ &\quad + \dots + (y_{n-1} + y_n - 1 + \dots + y_1 - y_n) = \\ &= \frac{1}{n} [ny_0 - 2(1 - 1)y_1 + \dots - 2(1 - 1)y_{n-1} + \dots + 2y_n] = \\ &= y_0 - 2 \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \gamma_j. \end{aligned} \quad (6.5.2)$$

(Если так просто получить это, положите n , скажем, 3 для проверки.) Для каждого фиксированного $j \geq 1$ коэффициент при γ_j в (6.5.2) сходится к 1 для $n \rightarrow \infty$, так что все аппроксимации в конечном счете входят в сумму с коэффициентом, равным единице. Это не является достаточным для того, чтобы получить желаемый результат (6.5.1). Доказательство части (ii) является аналитическим выражением (10).

Для ковариационно-стационарного процесса $\{y_t\}$ мы определяем *длительную дисперсию* (long-run variance) как предел $\text{Var}(\sqrt{n} \bar{y})$ при $n \rightarrow \infty$ (если он существует). Поэтому часть (b) утверждения говорит о том, что длительная дисперсия равна $\sum_{j=-\infty}^{\infty} \gamma_j$, что, в свою очередь, равно значению производящей функции автоковариаций $g(\lambda)$ для $\lambda = 1$ (см. (6.2.15)) или умноженному на 2π значению спектра на нулевой частоте (см. (6.2.17)).

Для центральных предельные теоремы

Мы представим две CLT, которые включают гравитацию корреляции. Первая является обобщением CLT Линдберга — Лёва.

Утверждение 6.9 (CLT для $MA(\infty)$): Пусть $y_t = a + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, где $\{\varepsilon_t\}$ — независимый белый шум, и $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Тогда

$$\sqrt{n}(\bar{y} - \mu) \rightarrow N\left(0, \sum_{j=-\infty}^{\infty} \gamma_j\right). \quad (6.5.3)$$

Мы не будем доказывать этот результат (для доказательства см.: [Andersson, 1971, Theorem 7.7.8] или [Brockwell and Davis, 1991, Theorem 7.1.2]), но мы не должны упускать тот факт, что $\text{Var}(\bar{y})$ (дисперсия предельного распределения $\sqrt{n}(\bar{y} - \mu)$) равна для временной дисперсии $\sum_{j=-\infty}^{\infty} \gamma_j$. Мы знаем из леммы 2.1, что « $x_n \rightarrow_d x$ », « $\text{Var}(x_n) \rightarrow c < \infty$ », « $E(|x_n|^{2+\delta}) < \infty$ » $\Rightarrow$ « $\text{Var } x = c$ ». Положим здесь $x_n = \sqrt{n}(\bar{y} - \mu)$. По (6.5.1), $\text{Var}(e_n)$ сходится к конечному пределу $\sum \gamma_j$. Поэтому дисперсия предельного распределения $\sqrt{n}(\bar{y} - \mu)$ равна $\sum \gamma_j$.

По утверждению 6.1(d), процесс в утверждении 6.9 является (стационарным и) эргодическим. Вспомните, что эргодичность ограничивает серийную корреляцию требованием того, что две случайные величины, отстоявшие достаточно далеко друг от друга в последовательности, являются почти независимыми. Но эргодическая стационарность сама по себе не является достаточной для гарантирования асимптотической нормальности $\bar{y}$. Возникает естественный вопрос: существуют ли дополнительные ограничения на эргодичность, если не считать линейные процессы, как в утверждении 6.9, которые обеспечивают асимптотическую нормальность? Ограничение, которое становится все более популярным — это то, что в этой книге называется **условием Гурдина** (*Gordin's condition*) для эргодических стационарных процессов. У него есть три части. Первые две части следующие:

- (a) $E(y^2) < \infty$. (Это является ограничением на [строго] стационарные процессы, поскольку, строго говоря, стационарные процессы могут не иметь конечных вторых моментов.)
- (b) $E(y(y_{-j}, y_{-2}, \dots)) \xrightarrow{j \rightarrow \infty} 0$. (Так как процесс $\{y_t\}$ стационарный, это условие эквивалентно тому, что $E(y(y_{-j}, y_{-j-1}, \dots)) \xrightarrow{j \rightarrow \infty} 0$. Поскольку $E(y(y_{-j}, y_{-j-1}, \dots))$ является случайной величиной, эта сходимость не может быть обычной сходимостью для действительных чисел.)

Из этого условия вытекает, что безусловное среднее равно нулю¹, $E(y_t) = 0$. Поскольку процесс $\{y_t\}$ имеет математическое ожидание, он должен приближаться к нулю, поскольку на отсутствие информации, то есть к безусловному математическому ожиданию. Для информации см. стрелку

¹См. также доказательства см.: Lemma 5.14 в [White, 1984].

части положим $I_t \equiv (y_t, y_{t-1}, y_{t-2}, \dots)$ и запишем y_t как

$$\begin{aligned} y_t &= y_t - [E(y_t|I_{t-1}) - E(y_t|I_{t-1})] - [E(y_t|I_{t-2}) - E(y_t|I_{t-2})] - \dots - \\ &\quad - [E(y_t|I_{t-j}) - E(y_t|I_{t-j})] = \\ &= [y_t - E(y_t|I_{t-1})] + [E(y_t|I_{t-1}) - E(y_t|I_{t-2})] + \dots + \\ &\quad + [E(y_t|I_{t-j-1}) - E(y_t|I_{t-j})] + E(y_t|I_{t-j}) = \\ &= (r_{t1} + r_{t2} + \dots + r_{t,j-1}) + E(y_t|y_{t-j}, y_{t-j-1}, \dots), \end{aligned}$$

где

$$r_{tk} \equiv E(y_t|I_{t-k}) - E(y_t|I_{t-k-1}).$$

Это r_{tk} является изменением (срессмотром) ожиданий относительно y_t при увеличении информационного множества с I_{t-k-1} до I_{t-k} . По (b), $y_t = (r_{t1} + r_{t2} + \dots + r_{t,j-1})$ сходится в среднем квадратическом к нулю при $j \rightarrow \infty$ для каждого t . Поэтому y_t можно записать как

$$y_t = \sum_{j=1}^{\infty} r_{tj}.$$

Эта сумма называется **телескопической суммой** (*telescoping sum*). Третьей частью условия Гордина является

$$c) \sum_{i=0}^{\infty} [E(r_{i0}^2)]^{1/2} < \infty.$$

Телескопическая сумма указывает, как «шоки», представленные $(r_{t0}, \dots)$, влияют на текущее значение y_t . Условие (с), грубо говоря, показывает, что шоки, которые произошли давно, не оказывают непропорционально большого влияния. В таком виде это условие ограничивает степень сериальной автокорреляции в $\{y_t\}$.

Для иллюстрации указанного условия возьмем пример AR(1) с нулевым средним и с независимыми ошибками:

$$y_t = \phi y_{t-1} + \varepsilon_t, \quad |\phi| < 1,$$

$$\{\varepsilon_t\} = \text{независимый белый шум с } \sigma^2 = \text{Var}(\varepsilon_t).$$

Очевидно, что условие Гордина (a) удовлетворяется. Условие (b) также удовлетворяется, поскольку

$$E(y_t|y_{t-j}, y_{t-j-1}, \dots) = \phi^j y_{t-j} \quad (6.5.4)$$

и $E[(\phi^j y_{t-j})^2] \rightarrow 0$ при $j \rightarrow \infty$. Пересмотр ожиданий r_{tj} может быть записан как

$$r_{tj} = \phi^j y_{t-j} - \phi^{j+1} y_{t-j-1} = \phi^j (y_{t-j} - \phi y_{t-j-1}) = \phi^j \varepsilon_{t-j}. \quad (6.5.5)$$

Потоку телекоммуникаций сужа для $AP(1)$ и является $MA(\infty)$ -представлением. Условие (с) удовлетворяется, поскольку $|E\epsilon_{1j}^2|^{1/2} = |\omega|^{1/2}$ и

$$\sum_{j=0}^{\infty} |\omega|^{1/2} = \frac{1}{1 - |\omega|^{1/2}} < \infty.$$

Исходная CLT такова

Утверждение 6.10 (CLT Гордина для эргодических стационарных процессов с нулевым средним¹): *Предположим, что процесс $\{y_t\}$ стационарный и эргодический, и предположим, что выполняется условие Гордина. Тогда $E(y_t) = 0$, автоковариации $\{\gamma_j\}$ абсолютно суммируемы и*

$$\sqrt{n}\bar{y} \rightarrow N\left(0, \sum_{j=-\infty}^{\infty} \gamma_j\right).$$

Поскольку условие Гордина удовлетворяется, если $\{y_t\}$ является мартигал-разностью, которая не проявляет серийной корреляции, утверждение 6.10 является обобщением CLT для гауссовской стационарной мартигал-разности (2.4) главы 2.

Многомерное обобщение

Обобщение приведенных выше результатов на многомерный случай произвольной зависимости. Пусть

$$\bar{y} = \frac{1}{n} \sum_{t=1}^n \mathbf{y}_t$$

является выборочным средним векторного процесса $\{\mathbf{y}_t\}$

■ Многомерная версия утверждения 6.8 заключается в том, что

- $\bar{y} \rightarrow_{P} \mu$, если каждый диагональный элемент в Γ_j стремится к нулю при $j \rightarrow \infty$, и
- предел $\lim_{n \rightarrow \infty} \text{Var}(\sqrt{n}\bar{y})$ (который является долговременной ковариационной матрицей (long run covariance matrix) для $\{\mathbf{y}_t\}$) равен $\sum_{j=-\infty}^{\infty} \Gamma_j$, если $\{\Gamma_j\}$ является суммируемой (то есть если каждая компонента Γ_j является суммируемой).

¹ Этот результат, относящийся к [Gordin, 1968], был популяризирован в [Billingsley and Pittman, 1966; White, 1984]. Утверждение об абсолютной суммируемости автоковариаций отменяется в версии 1984 [Majumdar, 1985].

- Таким образом, автоковариационная матрица $\{y_t\}$ может быть записана как

$$G_Y(1) = 2\pi\alpha_Y(0) \sum_{j=-\infty}^{\infty} \Gamma_j = \Gamma_0 + \sum_{j=1}^{\infty} (\Gamma_j + \Gamma_j'), \quad (6.5.6)$$

где $G_Y(z)$ (производящая функция автоковариаций) и $\alpha_Y(\omega)$ (спектр) для векторного процесса $\{y_t\}$ определяются в параграфе 6.3.

- Многомерная версия утверждения 6.4 заключается в том, что

$$\sqrt{n}(\bar{y} - \mu) \xrightarrow{d} N\left(0, \sum_{j=-\infty}^{\infty} \Gamma_j\right), \quad (6.5.7)$$

если $\{y_t\}$ является векторным МД(∞) с абсолютными суммируемыми коэффициентами, а $\{x_{jt}\}$ являются векторным независимым белым шумом.

- Многомерное обобщение условия Гордина на эргодические стационарные процессы очевидно¹:

Условие Гордина на эргодические стационарные процессы

(a) $E(y_t y_t')$ существует и конечно.

(b) $E(y_t | y_{t-j}, y_{t-j-1}, \dots) \rightarrow_{a.s.} 0$ для $j \rightarrow \infty$.

(c) $\sum_{j=0}^{\infty} \|E(y_t' y_{t-j})\|_2 < \infty$, где

$$\tau_0 \equiv \|y_t, y_{t-j}, y_{t-j-1}, \dots\| = E(y_t | y_{t-j}, y_{t-j-1}, \dots).$$

- Многомерная версия утверждения 6.10 (последняя) предположим, что условие Гордина выполняется для векторного эргодического стационарного процесса $\{y_t\}$. Тогда $E(y_t) = 0$, $\{\Gamma_j\}$ является абсолютно суммируемой и

$$\sqrt{n} \bar{y} \xrightarrow{d} N\left(0, \sum_{j=-\infty}^{\infty} \Gamma_j\right). \quad (6.5.8)$$

Контрольные вопросы

1. Покажите, что $\text{Var}(\sqrt{n} \bar{y}) = \Gamma' \text{Var}(y_1, y_2, \dots, y_{n+1}) \Gamma / n$, где $\text{Var}(y_1, y_2, \dots, y_{n+1}) = \Gamma$ — автоковариационная матрица размера $n \times n$ ковариаций в стационарном процессе $\{y_t\}$.

¹ Сформулировано как Assumption 3.5 в [Hansen, 1992].

2. $\Lambda_{\text{MA}}(\bar{y})$ для $\text{MA}(x_t)$. Проверьте, что $\Lambda_{\text{MA}}(\bar{y})$ для процесса $\text{MA}(x_t)$ в утверждении 5.9 равна

$$\sigma^2 \cdot \left(\sum_{j=0}^{\infty} \delta_j^2 \right)^2$$

Указание: Она должна быть равна $\rho_Y(1)$ для $\text{MA}(x_t)$.

3. (Асимптотика $\bar{y}$ для $\text{AR}(1)$) Рассмотрим $\text{AR}(1)$ -процесс $y_t = \rho + \phi y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, $\rho < 1$. Верно ли, что $\bar{y} \rightarrow \rho$? Добавьте предположение, что $\{\varepsilon_t\}$ является процессом независимого белого шума. Вычислите $\Lambda_{\text{MA}}(\bar{y})$. [Ответ: $(1 - \sigma^2/(1 - \rho^2))\sigma^2$.]

4. (Долгосрочная ковариационная матрица профилей временного ряда.) Пусть $H(I) \equiv H_1 - H_2(I)$, $n \leq |I|$, — ковариационно-стационарный процесс с абсолютно суммируемыми автоковариациями. Проверьте, что ковариационная матрица для $y_n = H(I)y_t$ задается как

$$(H_n + H_1)G_n(I)(H_n + H_1)'$$

Указание: Используйте (6.3.1b) и (6.5.1).

5. (Долгосрочная дисперсия процесса «MA с длинным хвостом».) Какова долгосрочная дисперсия процесса $y_t = \varepsilon_t + \varepsilon_{t-1} + \dots + \varepsilon_{t-p}$, где ε_t является белым шумом? [Ответ: равна нулю.]

6.6. Включение сериальной корреляции в GMM

Модель и асимптотические результаты

Сейчас у нас есть все инструменты, необходимые для включения сериальной корреляции в модель GMM с одним уравнением главы 3. Вспомним, что $\bar{y}_t$ из главы 3 (с индексом «t», зачехленным на «~») является K -мерным вектором, определенным как $\bar{y}_t = \bar{y}_t$, при условии K -мерного вектора инструментов $\bar{z}_t$ и скалярной ошибки ε_t . По предположению 3.3 (условия ортогональности), среднее $\bar{y}_t$ равно нулю. Мы предположили в предположении 3.5, что $\{\bar{y}_t\}$ является мартингальностью. Поскольку при этом предположении не допускается сериальная корреляция, матрица $\bar{S}_n$ определена как асимптотическая дисперсия $\bar{y}_t = \left(\sum_{t=1}^n \bar{y}_t \bar{y}_t' \right)$, была просто дисперсией $\bar{y}_t$. Теперь мы допускаем сериальную корреляцию в $\{\bar{y}_t\}$, ослабляем предположение 3.3 следующим образом:

Предположение 3.6' (условие Гордина, ограничивающее степень сериальной корреляции): $\{\bar{y}_t\}$ удовлетворяет условию Гордина. Его долгосрочная ковариационная матрица невырождена.

Тогда по многомерной версии утверждения 6.10, $\sqrt{n}\bar{y}$ сходится к нормальному распределению. Дисперсия этого предельного распределения

(и, конечно $\text{Avar}(\hat{\theta})$), обозначенная как S , равна долговременной ковариационной матрице.

$$S = \sum_{j=-\infty}^{\infty} \Gamma_j = \Gamma_0 + \sum_{j=1}^{\infty} (\Gamma_j + \Gamma_j'), \quad (6.6.1)$$

где Γ_j — автоковариационная матрица j -го лагуна

$$\Gamma_j = E(g_t g_{t-j}') \quad (j = 0, \pm 1, \pm 2, \dots). \quad (6.6.2)$$

(Так как $E(g_t) = 0$ по условию ортогональности, здесь нет необходимости считать среднее из g_t .)

То есть $N(\text{Avar}(\hat{\theta}))$ равна Γ_0 при предположении 3.5 и равна $\sum_{j=-\infty}^{\infty} \Gamma_j$ при предположении 3.5'. Это *существенное* различие между моделями GMM с наличием и отсутствием зеркальной корреляции. Соответственно переищем все результаты, полученные в параграфах 3.5–3.7 при условии, что S задается сейчас соотношением (6.6.1) и $\hat{S}$, некоторая состоятельная оценка для S , переопределяется, чтобы принять во внимание зеркальную корреляцию. Более конкретно:

- GMM-оценки $\hat{\theta}(\hat{W})$ остаются состоятельными и асимптотически нормальными. Ее асимптотическая дисперсия *состоятельно* оценивается посредством (3.5.2) при условии, что $\hat{S}$ является состоятельной оценкой для S , определенной в (6.6.1). Об этой оценочной асимптотической дисперсии иногда говорят как о *состоятельной к гетеросkedастичности и автокорреляции (НАС)*.
- GMM-оценка достигает минимальной дисперсии, при $\lim_{n \rightarrow \infty} \hat{W} = S^{-1}$, где S задается как в (6.6.1).
- Двухшаговая процедура, описанная в параграфе 3.5, все еще дает эффективную GMM-оценку при условии, что $\hat{S}$, введенная на шаге 1, является состоятельной для S .
- С такой, надлежащим образом определенной $\hat{S}$, выражения для статистик t , W , J , C и LP в параграфах 3.5–3.7 остаются в силе, эти статистики сохраняют те же асимптотические распределения и при наличии зеркальной корреляции.

Оценки N , когда автоковариации обращаются в нуль после конечного количества лагов

Все эти результаты подразумевают, что у нас есть состоятельная оценка, $\hat{S}$, долговременной ковариационной матрицы S . Получение такой оценки требует некоторого интеллектуального усилия, особенно когда автоковариации не обращаются в нуль после конечного количества лагов.

Мы начинаем наш поиск с состоятельного оценивания отдельных автоковариаций. Естественной оценкой является

$$\tilde{\Gamma}_j = \frac{1}{n} \sum_{i=j+1}^n \hat{y}_i \hat{y}_{i-j} \quad (j = 0, 1, \dots, n-1), \quad (6.6.3)$$

где

$$\hat{y}_i \equiv x_i - \bar{x}, \quad \bar{x} = \bar{y} - x'\bar{\beta}, \quad \bar{\beta} \text{ состоятельна для } \beta.$$

Если бы у нас были действительные значения y_i вместо оцененных значений $\hat{y}_i$ в этой формуле, то $\tilde{\Gamma}_j$ была бы состоятельной для Γ_j при предположениях 3.1 и 3.2, обеспечивающих эргодическую стационарность для $\{y_t\}$. Но мы должны использовать для вычисления автоковариаций не истинный, а оцененный ряд $\{\hat{y}_t\}$. По этой причине нам нужно подходящее условие на четвертый момент (о формулировании которого мы не будем беспокоиться) для оцененных автоковариаций, чтобы они были состоятельными. Доказательство состоятельности здесь не дается, поскольку оно очень близко доказательству утверждения 3.4.

При заданных оцененных автоковариациях следует рассмотреть две возможности вычисления $\tilde{S}$. Если мы *априори* знаем, что $\Gamma_j = 0$ для $j > q$, где q известно и конечно, то ясно, что $\tilde{S}$ может быть состоятельно оценена как

$$\tilde{S} = \tilde{\Gamma}_0 + \sum_{j=1}^q (\tilde{\Gamma}_j + \tilde{\Gamma}_j') = \sum_{j=-q}^q \tilde{\Gamma}_j \quad (\text{напомним: } \tilde{\Gamma}_{-j} = \tilde{\Gamma}_j'). \quad (6.6.4)$$

Более сложным является другой случай, когда мы не знаем значений q (которое может быть, а может и не быть конечным). Как мы можем оценить долговременную дисперсионную матрицу, которая включает возможно бесконечно много параметров? Есть два подхода.

Использование ядер для оценивания $\tilde{S}$

Недавно были предложены несколько процедур для оценивания долговременной дисперсионной матрицы¹ $\tilde{S}$. Так называемые **ядерные оценки** (*kernel-based estimators*) (или «непараметрические») могут быть представлены как взвешенное среднее оцененных автоковариаций:

$$\hat{\tilde{S}} = \sum_{j=-n+1}^{n-1} k\left(\frac{j}{q(n)}\right) \cdot \tilde{\Gamma}_j. \quad (6.6.5)$$

¹Для того чтобы результаты, сформулированные в этом параграфе и в следующем, выполнялись, нам необходимо наложить ряд технических условий (в дополнение к условию Гордана выше), которые еще более суживают характер серийной корреляции. Относительно формулировки этих условий и более детального обсуждения темы, рассматриваемой в этом параграфе и в следующем, см.: [Deo Haan and Levin, 1996b].

Здесь функция $k(\cdot)$, которая задает наш автоковарианс, называется **ядром** (*kernel*), а $q(n)$ называется **шириной окна** (*bandwidth*). Ширина окна может зависеть от объема выборки n . Оценка (6.6.4) является частным случаем ядерной оценки с $q(n) = q$ и

$$k(x) = \begin{cases} 1 & \text{для } |x| \leq 1, \\ 0 & \text{для } |x| > 1. \end{cases} \quad (6.6.6)$$

Это ядро называется **усеченным ядром** (*truncated kernel*). В данном контексте неизвестного q мы можем использовать усеченное ядро с шириной окна $q(n)$, которая вырастает вместе с объемом выборки. Если $q(n)$ увеличивается до бесконечности, то вычисления $\hat{S}$ может быть включено все больше и больше (и в конечном итоге все) автоковариаций. Однако на конечных выборках не гарантируется положительная полуопределенность этой усеченной ядерной оценки. Это является проблемой, поскольку тогда оцененная асимптотическая дисперсия GMM-оценки может не быть положительно полуопределенной.

[Newey and West, 1987] заметили, что ядерная оценка может быть сделана неотрицательно определенной на конечных выборках, если ее ядро является **ядром Бартлетта** (*Bartlett kernel*).

$$k(x) = \begin{cases} 1 - |x| & \text{для } |x| \leq 1, \\ 0 & \text{для } |x| > 1. \end{cases} \quad (6.6.7)$$

Ядерная оценка для S , основанная на ядре Бартлетта, называется (в эконометрике) **оценкой Ньюн — Веста** (*Newey — West estimator*). Например, для $q(n) = 3$ ядерная оценка включает автоковариации до двух (и трех) лагов:

$$\hat{S} = \hat{\Gamma}_0 + \left(\frac{2}{3}\right)(\hat{\Gamma}_1 + \hat{\Gamma}_1') + \left(\frac{1}{3}\right)(\hat{\Gamma}_2 + \hat{\Gamma}_2'). \quad (6.6.8)$$

В своем наиболее полном исследовании ядерного оценивания для одновременной дисперсии [Andrews, 1991] проанализировал положительно полуопределенные ядра, а именно класс ядер (включая ядро Бартлетта), которые приводят к положительной полуопределенности $\hat{S}$. Он показал, что при некоторых соответствующих условиях регулярности, таких как выпуклость, с которой $\hat{S}$ сходится (в среднем квадратическом) к S , зависит от ядра $\propto q(n)$. Самая высокая возможная скорость равна $n^{2/5}$ (то есть последовательность $n^{2/5} \cdot (\hat{S} - S)$ остается стохастически ограниченной)¹.

¹ Последовательность случайных величин $\{y_n\}$ называется **стохастически ограниченной** (*stochastically bounded*), что записывается как $y_n = O_p(1)$, если для любого $\epsilon > 0$ существует $M > 0$, такое что $\text{Prob}(|y_n| > M) < \epsilon$ для всех n . Если последовательность сходится по вероятности к некоторой случайной величине, то эта последовательность стохастически ограничена.

что происходит, когда $q(n)$ растет со скоростью $n^{1/2}$ (то есть $q(n)/n^{1/2}$ остается ограниченным), а ядро принадлежит определенному подмножеству положительно полуопределенных ядер. Ядро Бартлетта не входит в это подмножество. Для ядра Бартлетта самая высокая возможная скорость равна $n^{1/3}$, что достигается при росте $q(n)$ со скоростью $n^{1/3}$. Следовательно, по крайней мере на асимптотических основаниях, положительно полуопределенные ядра из этого подмножества должны быть предпочтительнее ядра Бартлетта. Пример таких ядер — **квадратичное спектральное (QS)** (*quadratic spectral*) ядро:

$$k(x) = \frac{25}{12\pi^2 x^2} \left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right). \quad (6.6.9)$$

Поскольку в QS-ядре $k(x) \neq 0$ для $|x| > 1$, все оцененные автоковариации $\hat{\Gamma}_j$ ($j = 0, 1, \dots, n-1$) входят в вычисление $\hat{S}$, даже если $q(n) < n-1$.

Надо отметить, что знания скорости роста недостаточно для определения значения ширины окна для конкретных данных конечной длины. [Andrews, 1991] рассмотрел также формулу для определения ширины окна на основе данных, которая не только растет с оптимальной скоростью, но и минимизирует некоторый соответствующий критерий (асимптотическую среднеквадратическую ошибку). Но даже в этой формуле некоторые значения параметров должны быть предложены пользователем, так что эта формула на самом деле не предоставляет автоматический выбор ширины окна на основе данных.

VARHAC

Другая процедура для оценивания S , называемая систематической к гетероскедастичности и автокорреляции VAR (**VARHAC**) оценкой, была предложена в [Det Nian and Levin, 1996a]. Идея заключается в том, чтобы подогнать VAR конечного порядка к K -мерному ряду $\{\hat{g}_t\}$, а затем инвертировать соответствующую ковариационную матрицу, подразумеваемую оценкой VAR. Это вовсе не означает, что $\hat{g}_t$ предполагается VAR конечного порядка. Фактически в этой процедуре (а также и в процедуре, использующей ядро) $\hat{g}_t$ даже не обязательно должен быть линейным процессом, не говоря уже об авторегрессии конечного порядка. Процедура VARHAC состоит из двух шагов.

Шаг 1: Выбор глубины запаздываний для каждого уравнения VAR.

В этом шаге для каждого уравнения VAR вы определяете глубину запаздываний на основе Байесовского информационного критерия (BIC). Пусть $\hat{g}_{kt}$ — k -й элемент K -мерного вектора $\hat{g}_t$. Тогда k -е уравнение VAR может быть записано в виде:

$$\hat{g}_{kt} = \omega_{11}^{(k)} \hat{g}_{k,t-1} + \omega_{12}^{(k)} \hat{g}_{k,t-2} + \dots + \omega_{1p}^{(k)} \hat{g}_{k,t-p} +$$

$$\begin{aligned}
& + \phi_{21}^{(k)} \hat{y}_{2,t-1} + \phi_{22}^{(k)} \hat{y}_{2,t-2} + \dots + \phi_{2p}^{(k)} \hat{y}_{2,t-p} \\
& \dots \\
& \phi_{K1}^{(k)} \hat{y}_{K,t-1} + \phi_{K2}^{(k)} \hat{y}_{K,t-2} + \dots + \phi_{Kp}^{(k)} \hat{y}_{K,t-p} + \epsilon_{kt} = \\
& = \phi_1^{(k)} \hat{y}_{1,t-1} + \dots + \phi_p^{(k)} \hat{y}_{1,t-p} + \epsilon_{kt}, \quad (6.6.10)
\end{aligned}$$

где

$$\phi_j^{(k)} = (\phi_{1j}^{(k)}, \dots, \phi_{Kj}^{(k)}), \quad (j = 1, 2, \dots, p).$$

При выборе глубины запаздываний по BIC должна быть специфицирована максимальная глубина запаздываний $p_{\max}$. В параграфе 6.4, когда BIC вводился в контексте оценивания истинного порядка (глубины запаздываний) авторегрессии конечного порядка, $p_{\max}$ предполагалось некоторому целому числу, в котором было известно, что оно больше или равно истинной глубине запаздываний. В текущем контексте, где g_t не обязательно является VAR конечного порядка, [Den Haan and Levin, 1996a, Theorem 2(c)] показывают, что оценка VARHAC является состоятельной для K , когда $p_{\max}$ растет со скоростью $n^{1/3}$, и рекомендуют положить $p_{\max}$ равным $\lfloor n^{1/3} \rfloor$ (целая часть $n^{1/3}$). Чтобы подчеркнуть ее зависимость от объема выборки, мы обозначаем максимальную глубину запаздываний как $p_{\max}(n)$. Пусть $SSR_p^{(k)}$ является суммой квадратов остатков (SSR) при OLS-оценивании (6.6.10) (k -е уравнение в VAR) для $t = p_{\max}(n) + 1, p_{\max}(n) + 2, \dots, n$. Для $p = 0$ $SSR_p^{(k)} = \sum_{t=p_{\max}(n)+1}^n (\hat{y}_{kt})^2$, SSR в регрессии $\hat{y}_{kt}$ при отсутствии регрессоров. Поскольку в k -м уравнении имеется $p(k)$ коэффициентов, критерий BIC равен:

$$\log(SSR_p^{(k)}/n) + p(k) \cdot \log(n)/n. \quad (6.6.11)$$

(в этом выражении n можно было бы заменить на фактический объем выборки $n - p_{\max}$ с отсутствием асимптотических последствий.) Пусть $p(k)$ равно значению p , которое минимизирует информационный критерий по $p = 0, 1, \dots, p_{\max}(n)$ для k -го уравнения VAR. Также пусть $\hat{\phi}_j^{(k)}$ ($j = 1, 2, \dots, p(k)$) является OLS-оценкой VAR коэффициентов $\phi_j^{(k)}$ ($j = 1, 2, \dots, p(k)$), когда (6.6.10) оценивается посредством OLS для $p = p(k)$.

Шаг 2: Вычисление подразумеваемой долгосрочной дисперсии.

Пусть

$$P = \max\{p(1), p(2), \dots, p(K)\}.$$

¹Тем что период выборки фиксируется на протяжении всего процесса выбора глубины запаздываний.

наибольшая глубина запаздываний в K уравнениях. По конструкции, $P \leq p_{\max}(n)$. Модель дает оцененную VAR, которая может быть записана как

$$\begin{aligned} \hat{g}_t &= \hat{\Phi}_1 \hat{g}_{t-1} + \dots + \hat{\Phi}_P \hat{g}_{t-P} + \hat{\varepsilon}_t \\ (K \times 1), \quad (K \times K), (K \times 1), \quad & \dots, \quad (K \times K), (K \times 1), \quad (K \times 1) \\ (t = p_{\max}(n) + 1, p_{\max}(n) + 2, \dots, n). \end{aligned} \quad (6.6.12)$$

Поскольку глубина запаздываний отличается для разных уравнений, некоторые строки в $\hat{\Phi}_p$ ($p = 1, 2, \dots, P$) могут быть нулевыми. Строки $\hat{\Phi}_p$ связаны с $\hat{\phi}_p^{(k)}$ ($p = 1, 2, \dots, P(k)$), полученными в главе 1, следующим образом:

$$k\text{-я строка в } \hat{\Phi}_p = \begin{cases} \hat{\phi}_p^{(k)} & \text{for } 1 \leq p \leq P(k), \\ 0 & \text{for } P(k) < p \leq P. \end{cases} \quad (6.6.13)$$

Пусть $\hat{\Omega}$ является ковариационной матрицей остатков VAR:

$$\hat{\Omega}_{(K \times K)} \equiv \frac{1}{n} \sum_{t=p_{\max}(n)+1}^n \hat{\varepsilon}_t \hat{\varepsilon}_t'. \quad (6.6.14)$$

Вспомогите, что одновременно ковариационная матрица VAR задается соотношением (6.3.16) для $z = 1$. Оценка для $\hat{S}$, подразумеваемая оцененной VAR (6.6.12), равна тогда

$$\hat{S} = \left[\mathbf{I}_K - \sum_{p=1}^P \hat{\Phi}_p' \right]^{-1} \hat{\Omega} \left[\mathbf{I}_K - \sum_{p=1}^P \hat{\Phi}_p \right]^{-1}. \quad (6.6.15)$$

Эта VARHAC-оценка явно определена по построению. [Det Nag and Levin, 1996a] показала, что (для подходящих условий регулярности) эта оценка сходится к S с быстрой скоростью, чем любая положительно определенная ядерная оценка, для почти любой автокорреляционной структуры процесса g_t (который обязательно является VAR конечного порядка). В частности, если g_t является VARMA конечного порядка, то порядок запаздываний, выбранный по BIC, дает параболическую скорость, в VARHAC-оценке сходится со скоростью, сколь угодно близкой к $n^{1/2}$, что быстрее, чем квадратичная скорость $n^{2/3}$, достигаемая положительно полуопределенной ядерной оценкой.

Контрольные вопросы

1. Что является эффективной GMM-оценкой, когда $x_t = x_t^*$?
2. (Подобные интегрированные сериальной корреляции.) Рассмотрим эффективную GMM-оценку, которая предполагает, что $\{g_t\}$ сериально коррелированы. Будет ли эта оценка состоятельной, если $\{g_t\}$ на самом деле сериально некоррелирована? Будет ли она эффективной?

6.7. Оценивание при условной гомоскедастичности (дополнительно)

Мы видели в параграфе 3.8, что при условной гомоскедастичности GMM-оценка редуцируется к 2SLS-оценке. Этот параграф показывает, как 2SLS можно обобщить для включения сериальной корреляции. При этом станет ясно, что GLS-оценка не является GMM-оценкой с сериальной корреляцией. Последняя половина этого параграфа включает соотношение между GLS и GMM.

Ядерное оценивание S при условной гомоскедастичности

Соотношение между сериальной корреляцией в $g_t \equiv x_t' \cdot \varepsilon_t$ и сериальной корреляцией в ошибках ε_t становится более ясным при наличии условной гомоскедастичности. Пусть

$$\omega_j \equiv E(\varepsilon_t \varepsilon_{t-j}) \quad \text{для всех } j. \quad (6.7.1)$$

Поскольку $\{\varepsilon_t\}$ является стационарным по предположениям 3.1 и 3.2, ω_j не зависит от t . Если $E(\varepsilon_t) = 0$ (что имеет место, когда x_t включает константу), то ω_j равна автоковариации $\{\varepsilon_t\}$ j -го порядка. Предположим, что ошибки условно гомоскедастичны:

условная гомоскедастичность:

$$E(\varepsilon_t \varepsilon_{t-j} | x_t, x_{t-j}) = \omega_j \quad (j = 0, +1, +2, \dots). \quad (6.7.2)$$

При этом дополнительном условии к Γ_j можно применить обычные аргументы, использующие закон полных математических ожиданий:

$$\begin{aligned} \Gamma_j &= E(g_t g_{t-j}') \quad (\text{поскольку } E(g_t) = 0) = \\ &= E(x_t \varepsilon_{t-j} x_{t-j}' \varepsilon_t) = \\ &= E[(E(x_t \varepsilon_{t-j} | x_t, x_{t-j}) x_{t-j}' \varepsilon_t)] \\ &\quad (\text{по закону полных математических ожиданий}) \\ &= \omega_j E(x_t x_{t-j}'). \end{aligned} \quad (6.7.3)$$

Таким образом, если $E(x_t x'_{t-j})$ не равно нулю, $\{x_t\}$ серийно коррелированы тогда и только тогда, когда $\{\varepsilon_t\}$ серийно коррелированы.

Для оценивания Γ_j мы используем тот факт, что Γ_j является произведением двух вторых моментов, $E(x_t x'_{t-j})$ и $E(\varepsilon_t \varepsilon'_{t-j})$. Естественной оценкой для $E(\varepsilon_t \varepsilon'_{t-j})$ является $\frac{1}{n} \sum_{t=j+1}^n x_t x'_{t-j}$. Легко показать, что ω_j состоятельно оценивается посредством $\frac{1}{n} \sum_{t=j+1}^n \tilde{\varepsilon}_t \tilde{\varepsilon}'_{t-j}$, где $\tilde{\varepsilon}_t$ является остатком для некоторой состоятельной оценки (доказательство почти то же самое, как в случае с отсутствием серийной корреляции, см. утверждение 3.2). Таким образом, естественной оценкой для Γ_j при условной гомоскедастичности является

$$\hat{\Gamma}_j = \left(\frac{1}{n} \sum_{t=j+1}^n \tilde{\varepsilon}_t \tilde{\varepsilon}'_{t-j} \right) \left(\frac{1}{n} \sum_{t=j+1}^n x_t x'_{t-j} \right). \quad (6.7.4)$$

При наличии этих оцененных автоковариаций ядерное оценивание $\hat{S}$ производится точно так же, как и в случае отсутствия условной гомоскедастичности, описанном в предыдущем параграфе.

Представление оцененной долговременной дисперсии через матрицу данных

С целью сопоставления GMM-оценки с серийной корреляцией, но с условной гомоскедастичностью и обычных оценок полезно представление $\hat{S}$ через матрицу данных. Определяя $n \times n$ матрицу $\hat{\Omega}$ соответствующим образом, мы можем записать $\hat{S}$ как

$$\hat{S} = \frac{1}{n} X' \hat{\Omega} X, \quad (6.7.5)$$

где X является матрицей данных размера $n \times K$, t -я строка которой равна x'_t . Матрица $\hat{\Omega}$ выглядит так же, как автоковариационная матрица $\{\varepsilon_t\}$

$$\hat{\Omega}_{(n \times n)} = \begin{bmatrix} \hat{\omega}_0 & \hat{\omega}_1 & \hat{\omega}_2 & \cdots & \hat{\omega}_{n-1} \\ \hat{\omega}_1 & \hat{\omega}_0 & \hat{\omega}_1 & \cdots & \hat{\omega}_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\omega}_{n-2} & \cdots & \hat{\omega}_1 & \hat{\omega}_0 & \hat{\omega}_1 \\ \hat{\omega}_{n-1} & \cdots & \hat{\omega}_2 & \hat{\omega}_1 & \hat{\omega}_0 \end{bmatrix}. \quad (6.7.6)$$

Если мы *априорно* знаем, что $\text{Cov}(\varepsilon_t, \varepsilon'_{t-j}) = 0$ (так что $\Gamma_j = 0$) для $j > q$, то специфицируем элементы этой $\hat{\Omega}$ как

$$\hat{\omega}_j = \begin{cases} \frac{1}{n} \sum_{t=j+1}^n \tilde{\varepsilon}_t \tilde{\varepsilon}'_{t-j}, & \text{если } 0 \leq j \leq q, \\ 0, & \text{если } j > q. \end{cases} \quad (6.7.7)$$

Тогда $\hat{S}$ в (6.6.4) может быть записана как (6.7.5). Если мы не знаем, что существует такое q , то используем ядерную оценку для S . Для ядра Бартлетта положим

$$\hat{\omega}_j = \begin{cases} \frac{1}{n} \left[\frac{j}{q(n)} \right] \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i \hat{\varepsilon}_{i-j}, & \text{если } 0 \leq j \leq q(n) - 1, \\ 0, & \text{если } j > q(n) - 1. \end{cases} \quad (6.7.6)$$

Тогда состоятельная оценка (6.6.3) с (6.6.4) равна (6.7.5).

Теперь легко показать следующее обобщение 2SLS оценки с целью охвата серийной корреляции.

- Эффективная GMM оценка при условной гомоскедастичности может быть записана как

$$\hat{\delta}(\hat{S}^{-1}) = [Z'X(X'\hat{\Omega}X)^{-1}X'Z]^{-1}Z'X(X'\hat{\Omega}X)^{-1}X'y, \quad (6.7.9)$$

где Z и y — матрицы данных для регрессоров и зависимой переменной. Это обобщает выражение (3.8.3') на с. 259.

- Состоятельная оценка $\text{Avar}(\hat{\delta}(\hat{S}^{-1}))$ редуцируется к

$$\begin{aligned} \text{Avar}(\hat{\delta}(\hat{S}^{-1})) &= \left[\frac{1}{n} Z'X \left(\frac{1}{n} X'\hat{\Omega}X \right)^{-1} \frac{1}{n} X'Z \right]^{-1} \\ &= n \cdot \left[Z'X(X'\hat{\Omega}X)^{-1}X'Z \right]^{-1}. \end{aligned} \quad (6.7.10)$$

Поэтому квадратные корни из диагональных элементов

$$\left[Z'X(X'\hat{\Omega}X)^{-1}X'Z \right]$$

являются стандартными ошибками. Это обобщает выражение (3.8.5') на с. 259.

Связь с GLS

Процедура для устранения серийной корреляции в GMM-оценивании, которую мы описали, отличается от GLS-процедуры. Чтобы увидеть это отличие более ясно, рассмотрим случай, когда $\varepsilon_t = \varepsilon_t$. Если имеется L регрессоров, то $X'Z$ и $X'ZX$ в (6.7.9) являются матрицами размера $L \times L$, поэтому эффективной GMM-оценкой, которая использует условия ортогональности $E(z_t \cdot \varepsilon_t) = 0$, является OLS-оценка $\hat{\delta}_{OLS} = (Z'Z)^{-1}Z'y$. Состоятельная оценка ее асимптотической дисперсии получается, если положить $X = Z$ в (6.7.10):

$$\text{Avar}(\hat{\delta}_{OLS}) = n \cdot (Z'Z)^{-1}(Z'\hat{\Omega}Z)(Z'Z)^{-1}. \quad (6.7.11)$$

Это может показаться странным, поскольку мы знаем из параграфа 1.6, что GLS-оценка является более эффективной, чем OLS, когда ошибки сериально коррелированы. Этот результат относится к конечным выборкам, но при условии, что у нас есть состоятельная оценка $\hat{\Omega}$ для Ω . мы можем взять формулу GLS и оценить δ как

$$\hat{\delta}_{GLS} = (Z' \hat{\Omega}^{-1} Z)^{-1} Z' \hat{\Omega}^{-1} y.$$

Не является ли эта оценка асимптотически более эффективной, чем OLS, в том смысле, что ее асимптотическая дисперсия меньше? Проблема с GLS заключается в том, что состоятельность не гарантируется, если регрессоры просто предопределенные, как показано ниже. Как мы подчеркивали в главе 2, во многих моделях временных рядов регрессоры не являются строго экзогенными. Отсюда следует, что *GLS не должен использоваться для коррекции сериальной корреляции в ошибках для моделей, не имеющих строгой экзогенности*. В частности, для случая $x_t = x_t$ корректная процедура для коррекции сериальной корреляции заключается в том, чтобы оставить точечную оценку неизменной и в то же время включить сериальную корреляцию в оценку асимптотической дисперсии.

То, что GLS-оценка в общем случае является несостоятельной, можно увидеть следующим образом. Чтобы сосредоточиться на решаемой проблеме, предположим, что истинная автоковариационная матрица пропорциональна некоторой известной $n \times n$ матрице V :

$$\text{Var}(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) = \sigma^2 \cdot V.$$

Как показано в параграфе 1.6, GLS-оценка коэффициента δ в регрессионной модели $y = Z\delta + \varepsilon$ может быть вычислена как OLS-оценка в преобразованной модели $Cy = CZ\delta + C\varepsilon$, где C является «квадратным корнем» из V^{-1} , так что $C'C = V^{-1}$. Но рассмотрим на i -е наблюдение преобразованной модели, которое может быть записано как

$$\tilde{y}_i = \tilde{x}_i' \delta + \tilde{\varepsilon}_i, \quad (6.7.12)$$

где

$$\tilde{y}_i = c_{i1}y_1 + \dots + c_{in}y_n,$$

$$\tilde{x}_i = c_{i1}x_1 + \dots + c_{in}x_n,$$

$$\tilde{\varepsilon}_i = c_{i1}\varepsilon_1 + \dots + c_{in}\varepsilon_n,$$

$(c_{i1}, \dots, c_{in})$ — i -я строка в C .

Для состоятельности GLS-оценки необходимо, чтобы $E(\tilde{x}_i' \tilde{\varepsilon}_i) = 0$. Если

часть этого условия может быть записана как

$$E(\hat{z}_k - \hat{z}_k') = (a) = 0, \\ (a) = \sum_{k=1}^n c_k^1 E(z_k - z_k'), \quad (b) = \sum_{k=2}^n c_k + c_0 E(z_1 - z_1').$$

По условию ортогональности $E(z_k - z_k') = 0$, (a) равно нулю. Что касается (b), то поскольку регрессоры просто предопределены и не предполагаются строго экзогенными, то нет гарантии того, что $E(z_k - z_k')$ равно нулю для $k \neq 1$. Поэтому (b) не равно нулю, если c не принимают определенных значений, для того чтобы компенсировать ненулевые слагаемые.

Существует один важный частный случай, в котором GLS-оценка состоятельна, — это случай, когда ошибки образуют процесс авторегрессии конечного порядка. Чтобы взять самый простой случай, предположим, что $\{z_t\}$ является AR(1): $z_t = \phi z_{t-1} + \eta_t$. Его автоковариационная матрица равна

$$\text{Var}(z_1, z_2, \dots, z_n) = \sigma^2 \cdot V, \quad V = \frac{1}{1 - \phi^2} \begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^{n-1} \\ \phi & 1 & \phi & \dots & \phi^{n-2} \\ & & \ddots & \ddots & \\ \phi^{n-2} & \dots & \phi & 1 & \phi \\ \phi^{n-1} & \dots & \phi^2 & \phi & 1 \end{bmatrix}.$$

Легко проверить, что квадратный корень из $V^{-1} \cdot C$ (удлиненной матрицы $C \cdot C = V^{-1}$), имеет вид:

$$C = \begin{bmatrix} \sqrt{1 - \phi^2} & 0 & 0 & \dots & 0 & 0 \\ -\phi & 1 & 0 & \dots & 0 & 0 \\ 0 & \phi & 1 & \dots & 0 & 0 \\ & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\phi & 1 \end{bmatrix}. \quad (6.7.13)$$

Поэтому h преобразованное уравнение имеет вид:

$$y_t - \phi y_{t-1} = (z_t - \phi z_{t-1})\beta + \eta_t \quad (6.7.14)$$

GLS-оценка для β , которая и является OLS-оценкой в регрессии $y_t - \phi y_{t-1}$ на $z_t - \phi z_{t-1}$, является состоятельной, если η_t ортогональна к z_t и z_{t-1} . Однако эти условия ортогональности вытекают из условий ортогональности для непробранной модели, которые заключаются в том, что $E(z_t - z_t') = 0$.

Контрольные вопросы

1. Проверьте (6.7.5) для $n = 3$, $q = 1$.
2. (Статистика Саргана.) Получите обобщение статистики Саргана (формула (3.8.10')) на с. 259) на случай сериально коррелированных ошибок.
3. Предположим, что $\mathbf{z}_t$ является собственным подмножеством $\mathbf{x}_t$. Мы видели в параграфе 3.8, что оценка 2SLS редуцируется к OLS-оценке. Верно ли это также и здесь? То есть редуцируется ли (6.7.9) к $(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$? [Ответ: Нет.]
4. Предоставленно $\hat{\mathbf{S}}$ через матрицу данных без условий гомоскедастичности.) Определите $\hat{\mathbf{Q}}$ подходящим образом так, чтобы (6.7.5) было равно ядерной оценке Бартлетта для $\hat{\mathbf{S}}$ без условной гомоскедастичности, а именно (6.6.5) с (6.6.7). Укажите: Запишите $\hat{\mathbf{Q}}$ или (6.7.6). Если глубина запаздывающей q известна, то $\hat{Q}_{ij} = \hat{Q}_{ji}$ для $i - j \leq q$ и равно 0 для $|i - j| > q$, где $\hat{Q}_{ij}$ является (i, j) -элементом в $\hat{\mathbf{Q}}$. Как вы определите $\hat{Q}_{ij}$ для ядерной оценки Бартлетта?
5. Рассмотрим простую регрессионную модель: $y_t = \beta z_t + \varepsilon_t$, где $\{\varepsilon_t\}$ является АБ(1), $\varepsilon_t = \phi\varepsilon_{t-1} + \eta_t$ с известным ϕ . Предположим, что $E(\varepsilon_t \varepsilon_s) = 0$, $E(\varepsilon_t \eta_s) = 0$, $1/(1 - \phi^2) > 0$ и условную гомоскедастичность. Также, для упрощения, пусть $E(z_t) = 0$. Покажите, что

$$\text{Avar}(\hat{\beta}_{OLS}) = \frac{\sigma_\eta^2}{\sigma_z^2} \frac{1 - 2\sum_{j=1}^{\infty} \phi^j \rho_j}{1 - \phi^2}, \quad \text{Avar}(\hat{\beta}_{GLS}) = \frac{\sigma_\eta^2}{\sigma_z^2} \frac{1}{(1 - \phi^2) - 2\phi\rho_1},$$

где $\sigma_z^2 \equiv \text{Var}(z_t)$, $\sigma_\eta^2 \equiv \text{Var}(\eta_t)$, $\rho_j = \text{Cov}(z_t, z_{t-j})$. Найдите такую конфигурацию для $\{\phi, \{\rho_j\}\}$, чтобы $\text{Avar}(\hat{\beta}_{GLS}) > \text{Avar}(\hat{\beta}_{OLS})$. Объясните, почему это соответствует тому факту, что $\hat{\beta}_{GLS}$ является эффективной ГММ-оценкой.

6.8. Приложение: Форвардные обменные курсы как оптимальные предикторы

На валютных рынках процентная разница между форвардным и спотовым обменными курсами называется **форвардной премией** (*forward premium*). Соотношение между форвардной премией и ожидаемым темпом укрепления/ослабления валюты в течение срока действия форвардного контракта является предметом большого количества эмпирических исследований. Самая простая гипотеза об этом соотношении заключается в том, что эти два значения одинаковы. В данном параграфе мы проверим эту гипотезу, приводя те же самые недельные данные, которые были использованы в [Baksett and Hodrick, 1993].

Данные охватывают период с 1973 по 1989 год и включают следующие переменные:

S_t — спотовый обменный курс в единицах иностранной валюты за доллар (например, 125 иен за доллар) в пятницу t -й недели,

F_t — 30-дневный форвардный обменный курс в пятницу t -й недели, а именно цена доллара в единицах иностранной валюты, поставляемый через 30 дней после пятницы t -й недели.

S_{30} — спотовая ставка на дату поставки 30-дневного форвардного контракта, созданного в пятницу t -й недели.

Важной особенностью этих данных является то, что срок погашения контракта охватывает несколько интервалов выборки: датой поставки является дата после пятницы $t + 4$ -й недели, но до пятницы $t + 5$ -й недели.

Гипотеза эффективности рынка

Чтобы проверить предположения, лежащие в основе этой гипотезы, рассмотрим три стратегии инвестирования долларовых средств на 30 дней в момент t . Первая заключается в инвестировании в инструмент внутреннего денежного рынка (например, казначейские векселя США), срок погашения которого составляет 30 дней. Пусть i_t является нормой доходности за 30-дневный период. Вторая инвестиционная стратегия заключается в конвертации долларов в единицы иностранной валюты, покупке 30-дневных инструментов денежного рынка, номинированных в иностранной валюте, процентная ставка которой обозначается как i_t^* , и конвертации обратно в доллары в момент погашения. Норма доходности этой инвестиционной стратегии равна:

$$S_t \cdot (1 + i_t^*) / S_{30}.$$

Эта стратегия включает риск изменения курса валюты, поскольку в момент инвестирования, t , вы не знаете значения S_{30} , которое будет являться спотовой ставкой через 30 дней. Третья инвестиционная стратегия отличается от второй тем, что она хеджирует против риска изменения курса путем покупки долларового форварда по цене F_t в тот же самый момент, когда долларовые средства конвертируются в единицы иностранной валюты. Эта хеджированная иностранная инвестиция обеспечивает гарантированную доходность, равную:

$$S_t \cdot (1 + i_t^*) / F_t.$$

Поскольку первая и третья стратегии не включают риска, арбитраж требует, чтобы нормы доходности были одинаковыми:

$$1 + i_t = S_t \cdot (1 + i_t^*) / F_t.$$

Взяв логарифмы от обеих частей и используя аппроксимацию $\log(1 + r) \approx r$ д, мы получаем уравнение **покрытого процентного паритета** (covered interest parity):

$$i_t = r_t + f_t - i_t^*, \quad \text{или} \quad f_t - r_t = i_t^* - i_t. \quad (6.8.1)$$

где строчные буквы для S и F обозначают натуральные логарифмы. То есть форвардная премия, $f_t = s_t$, равна дифференциалу процентных ставок. Это соотношение выполняется почти точно в реальных данных. Фактически люди обычно вычисляют форвардную премию как дифференциал процентных ставок.

Норма доходности во второй стратегии не определена, но, если инвесторы нейтральны к риску, ее ожидаемая доходность равна гарантированной норме доходности. Кроме того, если их ожидания рациональны, ожидаемая доходность является условным математическим ожиданием, основанным на доступной информации в валюте, так что

$$E_t S_t \cdot (1 + i_t^*) / S30_t(I_t) = S_t \cdot (1 + i_t^*) / F_t, \quad (6.8.2)$$

где $E(\cdot|I_t)$ является оператором математического ожидания, а I_t является информационным множеством по состоянию на дату t . Поскольку S_t , F_t и i_t^* известны в дату t , это равенство редуцируется к

$$E(F_t / S30_t | I_t) = 1.$$

И опять, по логарифмической аппроксимации, $F_t / S30_t \approx 1 + f_t - s30_t$. Следовательно,

$$E(s30_t | I_t) = f_t, \quad \text{или} \quad E(\varepsilon_t | I_t) = 0, \quad \text{где } \varepsilon_t = s30_t - f_t. \quad (6.8.3)$$

Это гипотеза эффективности рынка. Как ясно из вывода, она предполагает нейтральность к риску и рациональные ожидания. Эту гипотезу иногда описывают как то, что форвардная ставка является оптимальным прогнозом будущей спотовой ставки, поскольку условное математическое ожидание является оптимальным прогнозом в том смысле, что оно минимизирует среднеквадратичную ошибку (см. утверждение 2.7).

Тестирование равенства нулю безусловного среднего

Беря безусловное математическое ожидание от обеих сторон (6.8.3), мы получаем безусловное соотношение:

$$E(s30_t) = E(f_t), \quad \text{или} \quad E(\varepsilon_t) = 0. \quad (6.8.4)$$

То есть прогноз должен быть в среднем правильным. Рис. 6.1 изображает ошибку прогноза ε_t для обменного курса фунта/доллар. Она похожа на стационарный χ ряд, и мы находим из предположения, что это так¹.

¹Используя короткий периодный интервал, получаем $\varepsilon_t = (s30_t - s_t) - (f_t^* - f_t)$. Если мы верим, что в 30-дневная скорость накопления $s30_t = s_t$ и дифференциал процентных ставок $f_t^* = f_t$, то получим стационарный χ ряд. Мы должны верить в то, что ошибка прогноза является стационарным

Поскольку ошибка прогноза наблюдаема, мы можем проверить гипотезу о том, что безусловное среднее равно нулю. В этом заключается отличие от утверждения, посвященного работе Фама из главы 2, где ошибка прогноза (относительно будущей инфляции) наблюдаема только с точностью до константы.

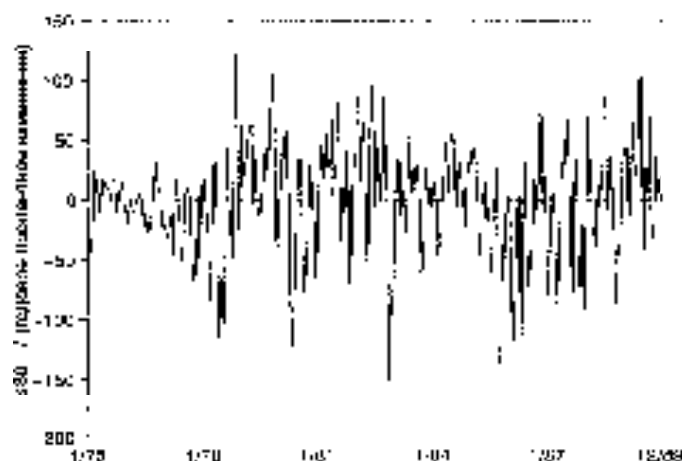


Рис. 6.1. Ошибка прогноза, йена/доллар

Для проверки гипотезы эффективности рынка нам нужно получить асимптотическое распределение выборочного среднего $\bar{e}$ при этой гипотезе. Эта задача в техническом отношении сложная, поскольку последовательность $\{e_t\}$ сериально коррелирована. Чтобы увидеть, почему ошибка прогноза имеет сериальную корреляцию, заметим, что I_t , включая $\{e_1, \dots, e_{t-4}, \dots\}$, не включает $\{e_t, \dots, e_{t+4}\}$, поскольку e_t , который зависит от x_t^* , не может наблюдаться до момента $t+4$. Поэтому

$$E(e_t | e_1, \dots, e_{t-4}, \dots) = 0. \quad (6.8.5)$$

Из этого следует, что $\text{Cov}(e_t, e_{t-j}) = 0$ для $j \geq 4$, но необязательно для $j < 4$. В утверждении, посвященном работе Фама, я рассуждал, что реальная процентная ставка (которая равна прогнозу инфляции плюс константа) была сериально некоррелирована, поскольку срок погашения процентной ставки и интервалы выборок были одинаковыми. Здесь же ошибка прогноза имеет сериальную корреляцию, поскольку срок погашения фьючерсного контракта (30 дней) длиннее, чем интервал выборки (календарный). Оцененная коррелограмма (образованная из более чем 2000 наблюдений) для стандартных ошибок на рис. 6.2¹. Стандартная ошибка вычисляется в предположении, что ошибка прогноза является $i.i.d.$. Поскольку она равна $1/\sqrt{N}$, как

¹Хотя теория подразумевает, что последовательное среднее равно нулю, я не тестировал это утверждение при вычислении выборочных автокорреляций.

в табл. 2.1. Поведение автокорреляции в целом соответствует гипотезе: она остается значительно выше двух стандартных ошибок для первых четырех лагов, а впоследствии, в основном, лежит внутри доверия.

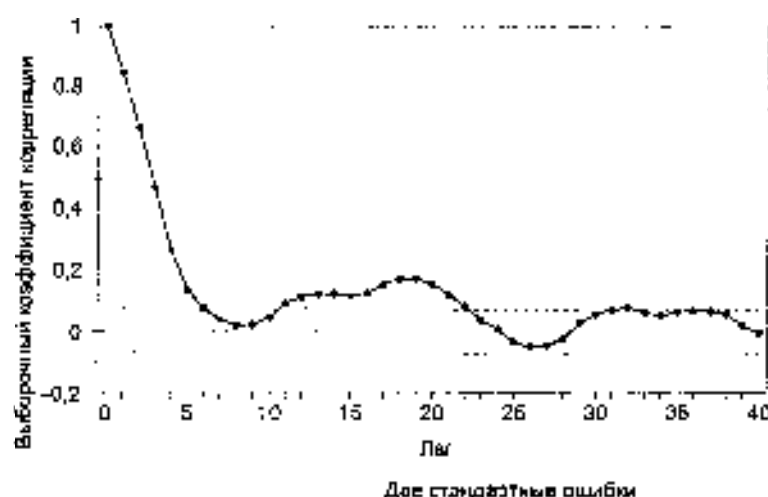


Рис. 6.2. Коррелограмма $\Delta 30$ f , желт./доллар

При нулевой гипотезе об эффективности рынка, поскольку серийная корреляция исчезает после конечного лага, легко видеть, что $\{\varepsilon_t\}$ удовлетворяет существенной части условия Гордина, ограничивающей серийную корреляцию. По закону повторных математических ожиданий, из (6.8.5) следует $E(\varepsilon_t | \varepsilon_{t-j}, \varepsilon_{t-j-1}, \dots) = 0$ для $j \geq 5$, из чего немедленно вытекает часть (b) условия Гордина. Поскольку пересмотр ожиданий, r_{tj} , равен нулю для $j \geq 5$, часть (с) выполняется, при условии, что r_{tj} ($0 \leq j < 4$) имеет конечные четвертые моменты. Тогда из утверждения 6.10 следует, что

$$\frac{\sqrt{n} \bar{\varepsilon}}{\sqrt{\gamma_0 + 2 \sum_{j=1}^4 \gamma_j}} = \frac{\bar{\varepsilon}}{\sqrt{(\gamma_0 + 2 \sum_{j=1}^4 \gamma_j)/n}} \xrightarrow{d} N(0, 1). \quad (6.8.6)$$

Кроме того, знаменатель достоверно оценивается путем замены γ_j на ее выборочный аналог $\hat{\gamma}_j$. Полученное отношение

$$\frac{\bar{\varepsilon}}{\sqrt{(\hat{\gamma}_0 + 2 \sum_{j=1}^4 \hat{\gamma}_j)/n}} \quad (6.8.7)$$

также является асимптотически нормальным. Знаменатель этого выражения будет называться стандартной ошибкой выборочности среднего.

В табл. 6.1 приведены результаты тестирования следствия из гипотезы, состоящего в том, что безусловное среднее ошибки прогноза равно

ную, для трех курсов иностранных валют по отношению к доллару. Эти три валюты — немецкая марка, британский фунт и японская иена. В первых двух столбцах таблицы приведены среднее и стандартное отклонение фактической скорости изменения спотовой ставки, $\Delta 30 - s$, и форвардной премии, $f_1 - s_1$. (Поскольку берется спотовая ставка по отношению к доллару, положительная скорость изменения означает, что доллар укрепляется.) Месyачные скорости изменения и форвардная премия были умножены на 1200, чтобы выразить их в годовых процентах. Гипотеза эффективности рынка заключается в том, что форвардная премия является ожидаемой скоростью изменения спотовой ставки. Очевидно, что фактическая скорость изменения намного (более отрицательна, чем форвардная премия. Третий столбец сообщает выборочное среднее для $\Delta 1$, которое равно разности между первыми двумя столбцами для выборочного среднего. Числа в скобках под выборочными средними представляют стандартные ошибки, вычисленные, как описывалось выше. Ни в одном случае нет значимого свидетельства против нулевой гипотезы о нулевом среднем.

Таблица 6.1

Средние скорости изменения, форвардных премий и разности между ними:
1975–1989 годы

Обменный курс	Средние и стандартные отклонения		Разность (ожидаемая скорость изменения)
	$\Delta 30 - s$ (фактическая скорость изменения)	$f - s$ (ожидаемая скорость изменения)	
иена/доллар	4.98 (41.6)	-3.74 (3.65)	1.25 (42.4)
			стандартная ошибка = 3.56
Европейская марка/доллар	-1.78 (40.6)	-3.91 (2.17)	2.13 (41.1)
			стандартная ошибка = 3.25
фунт стерлингов/доллар	3.59 (39.2)	2.16 (3.49)	1.43 (39.9)
			стандартная ошибка = 3.29

Примечание: $\Delta 30 - s$ — скорость изменения в течение 30 дней и $f - s$ — форвардная премия, выраженные в годовых процентных изменениях. Стандартные отклонения показаны в скобках. Стандартные ошибки — значения знаменателя в (6.7). Данные недельные с 1975 по 1989 г. Объем выборки — 776.

Регрессионные тесты

Безусловный тест, однако, не используется по причине эффективности рынка, которое заключается в том, что ошибки прогноза ортогональны обуславливающей информации t_i . Вероятно, наиболее важной переменной

ной в информационном множестве I_t для будущих спотовых ставок является форвардная ставка f_t . Мы можем тестировать, является ли ошибка прогноза некоррелированной с форвардной ставкой, регрессируя ε_t на координату f_t . Это эквивалентно оцениванию следующей регрессии:

$$\varepsilon(t) = \beta_0 + \beta_1 f_t + \varepsilon_t \quad (6.8.8)$$

и тестированию равенств $\beta_0 = 0$ и $\beta_1 = 1$. Поскольку ошибка является здесь ошибкой прогноза, ортогональной всему, что известно на дату t , включая f_t , регрессоры гарантированно будут ортогональными ошибкам. Поэтому может показаться, что для получения корректных статистических выводов мы можем оценить (6.8.8) посредством OLS и произвести коррекцию на серийную корреляцию в ошибках, как указано в параграфе 6.6.



Рис. 6.8. Спотовая ставка йена/доллар, январь 1975 г. — декабрь 1989 г.

Однако, как показывает рис. 6.8, спотовый обменный курс подпадает как процесс с возрастающей дисперсией. Как мы убедимся в главе 9, мы не можем отвергнуть гипотезу о том, что этот процесс имеет единичный корень. Форвардная ставка обладает таким же свойством. (Ее график не показан здесь, поскольку он неотличим от графика спотовой ставки.) Таким образом, предположение 3.2, которое требует, чтобы регрессор $\{f_t\}$ и зависимая переменная $\{\varepsilon(t)\}$ были стационарными, не выполняется. Проблема, в действительности, является более сложной. Мы видели на рис. 6.1, что разность между $\varepsilon(t)$ и f_t ведет себя как стационарный ряд. На языке вводимом далее в главе 10, $\varepsilon(t)$ и f_t «коинтегрированы» с коинтегрирующим вектором $(-1, 1)$, в этом случае OLS-оценка для β_1

в (6.8.8) быстро сходится к 1. Это можно увидеть на рис. 5.4, где показана зависимость $\hat{\alpha}_T$ от f_0 . Если получить оцененную линию регрессии, то ее угловой коэффициент равен 0.99. С точки зрения тестирования эффективности рынка, проблема состоит в том, что OLS-оценка сходится к 1 независимо от того, верна эта гипотеза или нет. Следовательно, у теста нет мощности.

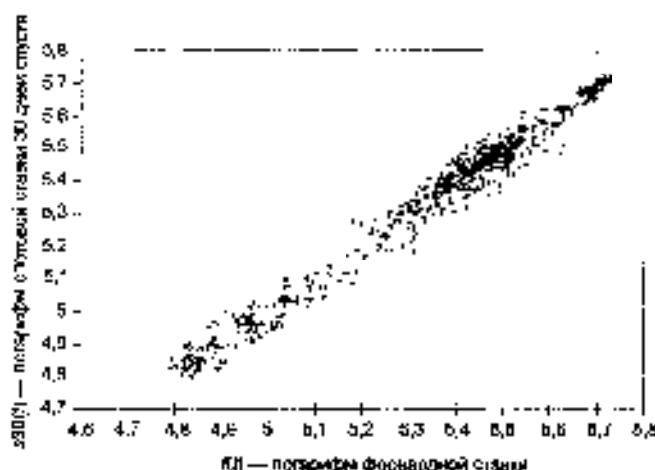


Рис. 6.4. График зависимости $\hat{\alpha}_T$ от f_0 , пенд/доллар

Эта проблема может быть решена путем оценивания другой регрессии:

$$\hat{\alpha}_T t_i - \alpha_T = \beta_0 + \beta_1 (f_T - \alpha_T) + \varepsilon_i. \quad (6.8.9)$$

Судя по графикам $\hat{\alpha}_T t_i - \alpha_T$ и $f_T - \alpha_T$ (не показаны), разумно предположить, что эти переменные являются стационарными, что соответствует предположению 3.2 (эргодическая стационарность). То, что при эффективности рынка это уравнение удовлетворяет другим предположениям СММ, можно проверить следующим образом. При $\beta_0 = 0$ и $\beta_1 = 1$ ошибка ε_i равна ошибке прогноза $\hat{\alpha}_T t_i - f_T$, и поэтому условия ортогональности $E(\varepsilon_i)$ и $E[(f_T - \alpha_T) \cdot \varepsilon_i]$ (предположения 3.3 с $x = \alpha_T$) выполняются. Условие идентификации заключается в том, что второй момент для П. $f_T - \alpha_T$ является невырожденным, что верно тогда и только тогда, когда $\text{Var}(f_T - \alpha_T) > 0$. Это очевидно, выполняется, если бы популяционная дисперсия была равна нулю, мы не наблюдали бы никакой вариации в $f_T - \alpha_T$. Это отличается от предположения 3.5', которое требует, чтобы и

$$g_1 \equiv x_2 \cdot x_1' = \begin{bmatrix} 1 & 1 \\ f_T - \alpha_T & \alpha_T \end{bmatrix}$$

было не единственным свойством серийной корреляции. Как уже было замечено, поскольку I_T включает $\{e_t, e_{t-1}, \dots\}$, ошибка прогноза удовлетворяет (6.8.5). В контрольном вопросе 3 нам будет предложено показать, что $\{g_t\}$ наследует то же самое свойство, а именно:

$$E(g_t | g_{t-2}, g_{t-1}, \dots) = 0. \quad (6.8.10)$$

Поэтому, как и в безусловном тесте, условие Гордона выполняется, если соответствующие угловые моменты конечны. Поскольку $E(g_t g'_{t-j}) = 0$ для $j \geq 3$, долговременную ковариационную матрицу S можно оценить, как в (6.6.4), с $q = 4$. (В эмпирическом упражнении вас попросят оценить $\hat{S}$ с помощью ядра Бартлетта, а также посредством VARHAC.)

Подводя итог, эффективной GMM-оценкой для $\beta = (\beta_0, \beta_1)'$ в (6.8.9) является OLS-оценка, и состоятельной оценкой ее асимптотической ковариационной матрицы является

$$\text{Avar}(\hat{\beta}_{OLS}) = S_{xx}^{-1} \hat{S} S_{xx}^{-1}, \quad (6.8.11)$$

где $\hat{S}$ задается в (6.6.4) с $q = 4$. В табл. 6.2 сообщаются результаты регрессионного теста для трех валют. Соответствующий график для обменного курса иена/доллар показан на рис. 6.5. Заметим, что для всех трех валют угловой коэффициент не только значимо отличается от 1,

Таблица 6.2

Регрессионные тесты на эффективность рынка: 1975–1988 годы

$s30_t - s_t = \beta_0 + \beta_1 (f_t - s_t) + \varepsilon_t$				
Валюта	Регрессионные коэффициенты		R^2	Статистика Вальда для $H_0: \beta_0 = 0, \beta_1 = 1$
	Константа	Формальная премия		
иена/доллар	-12.9 (4.01)	-2.10 (0.788)	0.034	18.6 ($p = 0.009\%$)
немецкая марка / доллар	-13.6 (5.72)	-3.01 (1.37)	0.026	8.7 ($p = 1.312\%$)
фунт стерлингов / доллар	7.95 (3.54)	3.02 (0.85)	0.033	12.9 ($p = 0.156\%$)

Примечание. Стандартные ошибки приведены в скобках. Они являются робастными к гетероскедастичности и допускают серийные корреляции до четырех лагов и вычислены по формуле (6.8.11). Объем выборки равен 776. Наши результаты несколько отличаются от результатов [Beckett and Hendrick, 1993], поскольку эти авторы использовали для оценивания S -матрицы на основе ядра Бартлетта с $q(n) = 4$.

но и является отрицательным. Вытекающая из эффективности рынка нулевая гипотеза о том, что $\beta_0 = 0$ и $\beta_1 = 1$, может быть окончательно отвергнута. Это отражается в слишком большом количестве наблюдений в северо-западном квадранте на рис. 6.5; слишком часто ожидаемое ослабление доллара (отрицательные $f_t - s_t$) связано с фактическим укреплением доллара (положительные $s_{30,t} - s_t$).

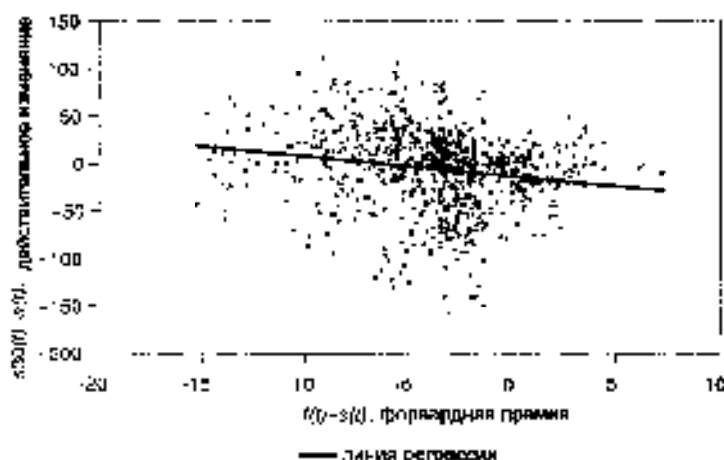


Рис. 6.5. График зависимости $s_{30} - s$ от $f - s$, пенн./доллар

Контрольные вопросы

1. (Стандартная ошибка безусловного теста.) Положите $\varepsilon_t = s_{30,t} - f_t$ и распишите регрессию ε_t на константу. Для учета сериальной корреляции используйте (6.6.4) с $q = 4$ для $\hat{S}$. Проверьте, что t -значение для гипотезы о том, что константа равна нулю, дается в (6.8.7).
2. (Отсутствие идентификации.) Для простоты предположим, что f_t^* является случайным форвардным обменным курсом, так что $s_{30,t}$ в (6.8.9) можно заменить на s_{t+2} . Предположим, что спотовый обменный курс является случайным блужданием и $I_t = \{x_1, x_2, \dots\}$. Каким должен быть f_t при нулевой гипотезе? Какое из требуемых предположений нарушается? (Ответ: Отсутствие идентификации.) Предположим вместо этого, что s_t является $AB(1)$ -процессом, удовлетворяющим $s_t = \alpha + \phi s_{t-1} + \eta_t$, где $\{\eta_t\}$ является i.i.d. и $|\alpha| < 1$. Проверьте, что $f_t = (1 - \alpha)s_t + \phi^2 x_t$. Указание: $s_{t+2} = \alpha(1 - \alpha)s_t + \phi^2 x_t + \eta_{t+2} + \phi \eta_{t+1}$. Проверьте, что $\varepsilon_t = \eta_{t+2} + \phi \eta_{t+1}$.
3. Докажите (6.8.10). Указание: $E(\varepsilon_t | I_t) = 0$. Проверьте, что I_t включает $s_t - f_t$ и $\{g_{t-2}, g_{t-3}, \dots\}$. Используйте закон повторных математических ожиданий.

Набор задач для главы 6

Аналитические упражнения

1. (Пределы в среднем квадратическом и их единственность.) В главе 2 мы определили сходимость в среднем квадратическом последовательности случайных величин $\{z_n\}$ к случайной величине z . В этом и в других вопросах мы будем иметь дело с последовательностями случайных величин с конечными вторыми моментами. Для таких исчислительных целей определение сходимости в среднем квадратическом таково:

Определение. Пусть $\{z_n\}$ является последовательностью случайных величин с $E\{z_n^2\} < \infty$. Мы говорим, что $\{z_n\}$ сходится в среднем квадратическом к случайной величине z с $E\{z^2\} < \infty$, записывая это как $z_n \rightarrow_{\text{в.к.}} z$, если $E\{z_n - z\}^2 \rightarrow 0$ при $n \rightarrow \infty$.

- (а) Покажите, что если $z_n \rightarrow_{\text{в.к.}} z$ и $z_n \rightarrow_{\text{в.к.}} z'$, то $E\{z - z'\}^2 = 0$. Укажите: $z = z' = (z - z_n) + (z_n - z')$. Определим $\|z\| = \sqrt{E\{z^2\}}$. Тогда

$$\begin{aligned} \|z - z'\|^2 &= E\{z - z_n\}^2 + 2E\{(z - z_n)(z_n - z')\} + E\{z_n - z'\}^2 \leq \\ &\leq \|z - z_n\|^2 + 2\sqrt{E\{(z - z_n)^2\}}\sqrt{E\{z_n - z'\}^2} + \|z_n - z'\|^2 \leq \\ &\text{по неравенству Коши—Шварца } E\{xy\} \leq \sqrt{E\{x^2\}}\sqrt{E\{y^2\}} = \\ &= \|z - z_n\| + \|z_n - z'\|. \end{aligned}$$

Поэтому $\|z - z'\| \leq \|z - z_n\| + \|z_n - z'\|$. Это называется неравенством треугольника.

- (б) Покажите, что $\text{Prob}\{z = z'\} = 1$. **Указание:** Используйте неравенство Чебышева:

$$\text{Prob}\{|z - z'| > \varepsilon\} \leq \frac{1}{\varepsilon^2} \|z - z'\|^2.$$

2. (Доказательство теоремы 5.1.) Хорошо известный результат о сходимости в среднем квадратическом (см., например, [Brockwell] and Davis, 1961, § 2.1 или § 2.1, записывается так, что

- (а) $\{z_n\}$ сходится в среднем квадратическом тогда и только тогда, когда $E\{z_n - z_0\}^2 \rightarrow 0$ при $n, 0 \rightarrow \infty$.

- (б) Если $z_n \rightarrow_{\text{в.к.}} z$ и $z_0 \rightarrow_{\text{в.к.}} z$, то

$$\lim_{n \rightarrow \infty} E\{z_n\} = E\{z\} \quad \text{и} \quad \lim_{n, 0 \rightarrow \infty} E\{z_n z_0\} = E\{z^2\}.$$

Отвечая, считайте этот результат доказанным.

- (а) Докажите, что при выполнении предположений утверждения 6.1 (6.1.6) сходится в среднем квадратическом при $n \rightarrow \infty$. **Указание:** Пусть

$$y_{n,n} = n^{-1} \sum_{j=0}^n \psi_j \varepsilon_{n-j}.$$

Что должно быть доказано, так это то, что $\{y_{n,n}\}$ сходится в среднем квадратическом при $n \rightarrow \infty$ для каждого t . Учитывая (i) выше, достаточно доказать, что (предполагая $m \geq n$ без потери общности)

$$E \left[\sum_{j=n+1}^m \psi_j \varepsilon_{m-j} \right]^2 = \sigma^2 \sum_{j=n+1}^m \psi_j^2 \rightarrow 0 \text{ при } m, n \rightarrow \infty.$$

Используйте тот факт, что абсолютно суммируемая последовательность является **квадратично суммируемой** (*quadrat summable*), то есть

$$\sum_{j=0}^{\infty} |\psi_j| < \infty \Rightarrow \sum_{j=0}^{\infty} \psi_j^2 < \infty,$$

и тот факт, что последовательность действительных чисел $\{\alpha_n\}$ сходится (к конечному пределу) тогда и только тогда, когда α_n является последовательностью Коши:

$$\alpha_n \rightarrow \alpha \Leftrightarrow |\alpha_m - \alpha_n| \rightarrow 0 \text{ при } m, n \rightarrow \infty.$$

Положите $\alpha_n = \sum_{j=0}^n \psi_j^2$.

- (b) Докажите, что $E(\mu_t) = \mu$. **Указание:** Пусть $y_{n,n} = n^{-1} \sum_{j=0}^n \psi_j \varepsilon_{n-j}$, как в (a). В (a) было показано, что $y_{n,n} \rightarrow_{\text{в.к.}} \mu$.
 (c) (Доказательство части (b) утверждения 6.1.) Покажите, что

$$E[(y_n - \mu)(y_{n-j} - \mu)] = \lim_{n \rightarrow \infty} E[(y_{n,n} - \mu)(y_{n-j,n} - \mu)].$$

Показали ли мы, что $\{y_t\}$ является ковариационно стационарным? [Ответ: Да.]

- (d) (Дополнительно.) Докажите часть (c) утверждения 6.1, приняв как доказанные следующие факты из математического анализа.

- (i) Если $\{a_j\}$ абсолютно суммируема, то $\{a_j\}$ суммируема (то есть $-\infty < \sum_{j=0}^{\infty} a_j < \infty$) и

$$\left| \sum_{j=0}^{\infty} a_j \right| \leq \sum_{j=0}^{\infty} |a_j|.$$

- ii) Рассмотрим двойную сумму с двумя индексами $\{a_{jk}\}$ ($j, k = 0, 1, 2, \dots$). Предположим, что $\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} |a_{jk}| < \infty$ для каждого n , и пусть $s_n \equiv \sum_{j=0}^n \sum_{k=0}^n |a_{jk}|$. Предположим, что $\{s_n\}$ абсолютно суммируема. Тогда

$$\left| \sum_{j=0}^{\infty} \left(\sum_{k=0}^{\infty} a_{jk} \right) \right| < \infty \quad \text{и} \quad \sum_{j=0}^{\infty} \left(\sum_{k=0}^{\infty} a_{jk} \right) = \sum_{k=0}^{\infty} \left(\sum_{j=0}^{\infty} a_{jk} \right) < \infty.$$

Указание: Получите

$$\sum_{j=0}^N |s_j| \leq \sigma^2 \sum_{j=0}^N \left(\sum_{k=0}^N |a_{j-k}| \cdot |a_k| \right) < \infty.$$

3. (Автоковариация $h(L)x_t$) Мы хотим получить автоковариации для $\{y_t\}$ в утверждении 6.2. Как и в утверждении 6.2, будем предполагать, что $\{x_t\}$ является абсолютно стационарным. Сначала рассмотрим взвешенное среднее конечного количества слагаемых:

$$y_{n,t} = h_0 x_t + h_1 x_{t-1} + \dots + h_n x_{t-n}.$$

- (a) Покажите, что для этого $\{y_{n,t}\}$

$$\gamma_j = \sum_{k=0}^n \sum_{l=0}^n h_k h_l \gamma_{j-k+l}^x,$$

где γ_j^x является автоковариацией $\{x_t\}$ j -го порядка.

Проверьте, что это равносильно к (6.1.2), если $\{x_t\}$ является белым шумом. **Указание:** Если вы найдете это сложным, начинайте сначала с $n = 1$, а затем увеличивайте n , чтобы установить структуру.

- (b) Рассмотрим теперь взвешенное среднее возможно бесконечного числа слагаемых:

$$y_t = h_0 x_t + h_1 x_{t-1} + \dots$$

Примем утверждение 6.2(a) как доказанное, покажите, что

$$\gamma_j = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} h_k h_l \gamma_{j-k+l}^x, \quad \text{если } \{h_j\} \text{ абсолютно суммируема.}$$

Указание: $y_{n,t} \rightarrow y_t$ по при $n \rightarrow \infty$ по утверждению 6.2. Продолжите, как в 2(c).

4. (Однородные линейные разностные уравнения.) Рассмотрим однородное линейное разностное уравнение p -го порядка:

$$y_j - \phi_1 y_{j-1} - \phi_2 y_{j-2} - \dots - \phi_p y_{j-p} = 0 \quad (j = p, p+1, \dots), \quad (1)$$

где $\{\phi_1, \phi_2, \dots, \phi_p\}$ являются вещественными константами с $\phi_p \neq 0$ с начальным условием в виде заданных значений $\{y_0, y_1, \dots, y_{p-1}\}$. Соответствующее полиномиальное уравнение имеет вид:

$$1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^{p-1} - \phi_p z^p = 0. \quad (2)$$

Пусть λ_k ($k=1, 2, \dots, K$) являются различными корнями этого уравнения (они могут быть комплексными) и пусть r_k является кратностью λ_k . Таким образом, $r_1 + r_2 + \dots + r_K = p$. Указанное полиномиальное уравнение можно записать как

$$\prod_{k=1}^K (1 - \lambda_k^{-1} z)^{r_k} = 0. \quad (3)$$

В качестве примера рассмотрим полиномиальное уравнение второго порядка

$$1 - z + \frac{1}{4}z^2 = 0. \quad (4)$$

Оно имеет корень 2 кратности 2, поскольку

$$1 - z + \frac{1}{4}z^2 = \left(1 - \frac{1}{2}z\right)^2.$$

Поэтому в этом примере $K=1$, $r_1=2$ и $\lambda_1=2$.

Как вы проверите ниже, решение однородного уравнения p -го порядка можно записать в виде:

$$y_j = \sum_{k=1}^K \sum_{r=0}^{r_k-1} c_{kr} \cdot (j)^r \lambda_k^{-j}. \quad (5)$$

Коэффициенты c_{kr} (их ровно p , поскольку $r_1 + r_2 + \dots + r_K = p$) можно определить из начальных условий, которые специфицируют значения для $\{y_0, y_1, \dots, y_{p-1}\}$. В примере выше, (уравнение (4)),

$$y_j = c_{10} \cdot 2^{-j} + c_{11} \cdot j \cdot 2^{-j}. \quad (6)$$

В важном частном случае, когда все корни различны, $K=p$ и $r_k=1$ для всех k , и (5) принимает вид:

$$y_j = \sum_{k=1}^p c_{1k} \lambda_k^{-j}. \quad (7)$$

(а) Для четного случая $p = 2$ и различных корней, (7) имеет вид:

$$y_j = c_{10}\lambda_1^{-j} + c_{20}\lambda_2^{-j}. \quad (8)$$

Проверьте, что это является решением однородного разностного уравнения (1) с $p = 2$ для любого значения (c_{10}, c_{20}) . Чтобы определить (c_{10}, c_{20}) , запишите (8) для $j = 0$ и $j = 1$ и решите это относительно значений c как функцию от $(y_0, y_1, \lambda_1, \lambda_2)$.

(б) Чтобы научиться (вспомнить), как работать с несколькими корнями, рассмотрим пример (4) выше. Проверьте, что (6) является решением однородного разностного уравнения второго порядка

$$y_j - y_{j-1} + \frac{1}{3}y_{j-2} = 0.$$

(в) (Дополнительно.) Докажите следующую лемму:

Лемма. Пусть $0 \leq \xi < 1$ и пусть n является неотрицательным целым числом. Тогда существуют для действительных числа A и b , такие что $\xi < b < 1$ и $(j)^n \xi^j < Ab^j$ для $j = 0, 1, 2, \dots$.

Указание: Выберите любое b , такое, что $\xi < b < 1$. Поскольку при достаточно j , b^j в конце концов принимает значения, большие, чем $(j)^n \xi^j$, существует некоторое j , назовем его J , такое, что $(j)^n \xi^j < b^j$ для всех $j \geq J$.

(г) Предположим, что выполняется условие стабильности для линейного однородного разностного уравнения p -го порядка. Докажите, что существуют две числа, A и b , такие что $0 < b < 1$ и

$$|y_j| < Ab^j \quad \text{для } j = 0, 1, 2, \dots$$

Указание:

$$\begin{aligned} |y_j| &= \left| \sum_{k=1}^K \sum_{n=0}^{r_k-1} c_{kn} \cdot (j)^n \lambda_k^{-j} \right| \leq \\ &\leq \sum_{k=1}^K \sum_{n=0}^{r_k-1} |c_{kn}| \cdot j^n \cdot |\lambda_k|^{-j} p \leq \\ &\leq c \sum_{k=1}^K \sum_{n=0}^{r_k-1} j^n |\lambda_k|^{-j} p, \end{aligned}$$

где $c = \max\{|c_{kn}|\}$. Поскольку $|\lambda_k|^{-1} < 1$ по условию стабильности, мы можем применить лемму с $\xi = |\lambda_k|^{-1}$ и утверждать, что

$$(j)^n |\lambda_k|^{-j} p < A_k (b_k)^j \quad \text{для всех } j$$

для некоторого $A_k > 0$ и $|A_k^{-1}| < b_k < 1$. Положим $A^* = \max\{A_k\}$ и $b = \max\{b_k\}$, так что $|A_k^{-1}b_k| \leq A^*b$ для всех k . Поэтому

$$|y_j| \leq c \sum_{i=1}^K \sum_{n=0}^{j-1} A^n b^n = crA^*b^j.$$

(Напомним, что $p = c_1 + \dots + c_K$.) Тогда определим A как crA^* . Вы можете найти такой способ доказательства, используя неравенства сложения. Если это так, покажите $p \cdot K$ и $(c_1, \dots, c_K)$ некоторым конкретным числом (например, $p = 3$, $K = 2$, $c_1 = 1$, $c_2 = 2$).

5. (Уравнения Юля - Уокера.) В основном тексте дековарияции $AV(1)$ -процесса мы получили на основании представления $MA(x)$. Здесь мы рассмотрим уравнения Юля - Уокера.

(а) Из (6.2.1') вы с. 405 получите, что

$$y_j - \phi y_{j-1} = E_{j+1}(y_j - \mu) \quad (j = 0, 1, 2, \dots).$$

(б) Легко. Используя MA -представление, покажите, что

$$E_{j+1}(y_{j-1} - \mu) = \begin{cases} \sigma^2 & (j = 0); \\ 0 & (j = 1, 2, \dots). \end{cases}$$

(с) Выведите уравнения Юля - Уокера:

$$y_0 - \phi y_{-1} = \sigma^2, \quad (9)$$

$$y_j - \phi y_{j-1} = 0 \quad (j = 1, 2, \dots), \quad (10)$$

(г) Вычислите $(y_0, y_1, \dots)$ из (9) и (10). **Указание:** Сначала положите $j = 1$ в (10) для получения $y_1 = \phi y_0$. Это в (9) можно решить относительно (y_0, y_1) .

6. ($AV(1)$ без стационарных фильтров.) Мы хотим доказать без помощи утверждения 6.2, что $\{y_t\}$, следующий $AV(1)$ -уравнению (6.2.1) с условием стационарности $|\phi| < 1$, должен иметь $MA(x)$ -представление, если он является ковариационно стационарным.

(а) Пусть $x_t \equiv y_t - \mu$, так что $E(x_t) = 0$. Последовательной подстановкой из (6.2.1') (см с. 405) получите

$$x_t = \tau_{t,2} + \phi^n x_{t-n},$$

где

$$\tau_{t,n} = \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots + \phi^{n-1} \varepsilon_{t-n+1}$$

(б) Покажите, что $x_{t,n} \rightarrow_{\text{пл.в.}} x_t$.

Указание: Надо доказать, что $E[(x_{t,n} - x_t)^2] \rightarrow 0$ при $n \rightarrow \infty$. Поскольку $(x_{t,n} - x_t)^2 = \phi^{2n} x_{t-n}^2$ и $\{x_t\}$ предполагается ковариационно стационарным, $E[x_{t-n}^2] < \infty$.

(в) (Тривиально.) Вспомогат. что бесконечная сумма

$$\sum_{i=0}^{\infty} \phi^2 \varepsilon_{t-i}$$

определяется как предел в среднем квадратическом частичной суммы $x_{t,n}$, т.е. доказать, что y_t можно записать как

$$y_t = \mu + \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}.$$

7. (См. сопровождающая форма.) Мы хотим обобщить доказательство, только не построенное в вопросе 6, на $AR(p)$. Без потери общности, мы можем предположить, что $\mu = 0$ (или переопределять y_t как отклонение от среднего для y_t). Определим

$$\xi_{t+1} = \begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \quad \pi_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$F = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

Рассмотрим следующие векторное стохастическое разностное уравнение первого порядка:

$$\xi_t = F\xi_{t-1} + \pi_t. \quad (11)$$

В этой системе p уравнений, называемой сопровождающей формой (companion form), первое уравнение точно такое же, как и $AR(p)$ -уравнение (5.2 б'), т.е. (10), а остальные даны в зависимости от последующих для ϕ_{-j} ($j = 1, 2, \dots, p-1$).

(а) Предположительной подсказкой покажите, что

$$\xi_t = \xi_{t,n} + (F)^n \xi_{t-n},$$

$$\xi_{t,n} \equiv y_t - Fy_{t-1} + F^2 y_{t-2} + \dots + F^{n-1} y_{t-n+1}. \quad (12)$$

Собственные значения (eigenvalues), или характеристические корни (characteristic roots) $p \times p$ -матрицы F являются корнями детерминантного уравнения

$$|F - \lambda I_p| = 0. \quad (13)$$

Можно показать, что левая часть этого уравнения равна $\lambda^p - \phi_1 \lambda^{p-1} + \dots + \phi_{p-1} \lambda - \phi_p$. Поэтому это детерминантное уравнение преобразуется под обозначениями уравнение p -го порядка

$$\lambda^p - \phi_1 \lambda^{p-1} + \dots + \phi_{p-1} \lambda - \phi_p = 0. \quad (14)$$

(b) Проверьте это для $p = 2$.

В оставшейся части этого вопроса предположим, что все корни полиномиального уравнения p -го порядка (которые, как было только что показано, являются собственными значениями F) различны. Пусть $\{\lambda_1, \lambda_2, \dots, \lambda_p\}$ — эти различные собственные значения. Из матричной алгебры известно, что существует невырожденная $p \times p$ -матрица T , такая что

$$P = T A T^{-1}, \quad A_{(p \times p)} = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_{p-1} & 0 \\ 0 & \dots & 0 & 0 & \lambda_p \end{bmatrix}. \quad (15)$$

(Этот факт не зависит от особой формы матрицы P ; все, что нужно для (15), — это то, что значения λ являются различными собственными значениями матрицы F .) Из (15) следует, что

$$(F)^n = T(A)^n T^{-1}.$$

(c) (Очень легко.) Докажите это для $n = 3$.

(d) Покажите, что $\xi_{t+n} \rightarrow_{n \rightarrow \infty} \xi$ при $n \rightarrow \infty$, если выполняется условие стационарности относительно $AR(p)$ -уравнения и если $\{\rho_i\}$ являются каноническими стационарными. Указание: Матрица стационарности требует, чтобы все корни в (13) были меньше 1 по абсолютной величине. В силу (12), то, что нужно показать — это то, что $(F^n) \xi_{t-n} \rightarrow_{n \rightarrow \infty} 0$. Вспомните, что $z_n \rightarrow_{n \rightarrow \infty} \alpha$, если сходимость в среднем квадратическом происходит экспоненциально.

(e) Покажите, что y_t имеет $MA(\infty)$ -представление

$$y_t = \phi_0 \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \dots$$

Как ψ_2 связан с характеристическими корнями? **Указание:** Из (ii) немедленно следует, что

$$\xi_1 = \psi_2 - F\psi_{-1} = (F')^2\psi_{-2} + \dots.$$

Верхнее уравнение этого выражения является MA(∞)-представлением для ψ_1 .

(Если ψ_2 все вероятные значения рекуррентна, то вместо (15) там можно использовать так называемое Жорданово разложение для F' .)

8. (AR(1), который начинался в момент $t = 0$.) В предыдущем тексте MA(∞)-представление (6.2.2) AR(1)-процесса, следующего $y_t = c + \phi y_{t-1} + \varepsilon_t$, подразумевает, что процесс определен для $t = -1, -2, \dots$, так же как и для $t = 0, 1, 2, \dots$. Вместе этого рассмотрим AR(1)-процесс, который начинается в $t = 0$ с начальным значением y_0 . Предположим, что $c \neq 0$ и что y_0 не коррелировано с последующим процессом белого шума $(\varepsilon_1, \varepsilon_2, \dots)$.

(a) Запишите дисперсию для y_t как функцию от σ^2 ($\equiv \text{Var}(\varepsilon_t)$, $\text{Var}(y_0)$) и ϕ . **Указание:** Пошаговая рекурсия дает:

$$y_t = (1 + \phi + \phi^2 + \dots + \phi^{t-1})c + \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \dots + \phi^{t-1}\varepsilon_1 + \phi^t y_0.$$

Пусть μ и $\{\gamma_j\}$ являются средним и автоковариационными функциями стационарного AR(1)-процесса, так что $\mu = c/(1 - \phi)$, а $\{\gamma_j\}$ заданы в (6.2.5). Покажите, что

$$\lim_{t \rightarrow \infty} E(y_t) = \mu, \quad \lim_{t \rightarrow \infty} \text{Var}(y_t) = \gamma_0, \quad \text{и} \quad \lim_{t \rightarrow \infty} \text{Cov}(y_t, y_{t-j}) = \gamma_j.$$

(b) Покажите, что $\{y_t\}$ является ковариационно стационарным, если $E(y_0) = \mu$ и $\text{Var}(y_0) = \gamma_0$.

9. (Доказательство утверждения 6.5(2).) Мы хотим доказать, что утверждение 6.5(a). По неравенству Чебышева достаточно показать, что

$$\lim_{n \rightarrow \infty} \text{Var}(\bar{y}) = 0.$$

Из (6.5.2):

$$\begin{aligned} \text{Var}(\bar{y}) &= \frac{\sigma^2}{n} - \frac{2}{n} \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \gamma_j = \\ &= \frac{\gamma_0}{n} - \frac{2}{n} \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \gamma_j \\ &\quad \left(\text{покажем, что } 1 - \frac{j}{n} \rightarrow 0 \text{ для } j = n\right) \\ &\leq \frac{\gamma_0}{n} - \frac{2}{n} \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \gamma_j \leq 0. \end{aligned}$$

$$\leq \frac{M}{n} + \frac{2}{n} \sum_{j=1}^n |\gamma_j|.$$

(а) (Дополнительно.) Докажите следующий результат:

$$\lim_{j \rightarrow \infty} a_j = 0 \Rightarrow \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n a_j = 0.$$

Указание: $\{a_j\}$ сходится к 0. Поэтому (i) $|a_j| < M$ для всех j и (ii) для любого заданного $\varepsilon > 0$ существует положительное целое N , такое, что $|a_j| < \varepsilon/2$ для всех $j > N$. Поэтому

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n |a_j| &= \frac{1}{n} \sum_{j=1}^N |a_j| + \frac{1}{n} \sum_{j=N+1}^n |a_j| < \frac{1}{n} \sum_{j=1}^N M + \frac{1}{n} \sum_{j=N+1}^n \frac{\varepsilon}{2} = \\ &= \frac{NM}{n} + \frac{(n-N)\varepsilon}{2n} < \frac{NM}{n} + \frac{\varepsilon}{2}. \end{aligned}$$

Выбирая n достаточно большим, NM/n можно сделать меньше, чем $\varepsilon/2$.

(b) Принимая результат в (а) как доказанный, покажите, что $\text{Var}(\bar{y}) \rightarrow 0$ при $n \rightarrow \infty$.

10. (Доказательство утверждения 6.8(ii).)

(а) (Дополнительно.) Докажите следующий результат:

$$\sum_{j=1}^{\infty} a_j < \infty \Rightarrow \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n j a_j = 0.$$

Указание:

$$\begin{aligned} \sum_{j=1}^n j a_j &= a_1 + 2a_2 + 3a_3 + \dots + n a_n = (a_1 + a_2 + \dots + a_n) + \\ &+ (a_2 + a_3 + \dots + a_n) + \dots + (a_{n-1} + a_n) + a_n = \\ &= \sum_{j=1}^n \sum_{k=j}^n a_k \leq \sum_{j=1}^n \left| \sum_{k=j}^n a_k \right| = \sum_{j=1}^n \sum_{k=j}^n |a_k| = \sum_{j=1}^n \left| \sum_{k=j}^n a_k \right|. \end{aligned}$$

Поскольку $\{a_j\}$ суммируема, (i) $|\sum_{k=j}^n a_k| < M$ для всех j, n и (ii) существует положительное целое N , такое, что $|\sum_{k=j}^n a_k| < \varepsilon/2$ для любого заданного $\varepsilon > 0$ для всех $j, n > N$ (это следует из того, что $\{\sum_{j=1}^k a_j\}$ является последовательностью Коши).

(b) Принимая этот результат как доказанный, покажите утверждение 6.8(b).

Эмпирические упражнения

1. Прежде чем отвечать, прочитайте по крайней мере п. 115–119 (за исключением последних двух параграфов на п. 119) в [Bekeirt and Hadrick, 1988]. Файлы данных DM.ASC (для немецкой марки), POLIND.ASC (для японской иены) и YEN.ASC (японская иена) содержат еженедельные данные по следующим пунктам.

Столбец 1: дата наблюдения (например, «19850104» обозначает 4 января 1985 г.).

Столбец 2: цена предложения доллара в единицах иностранной валюты на спотовом рынке в пятницу текущей недели (S_t).

Столбец 3: цена предложения доллара в единицах иностранной валюты на 30-дневном форвардном рынке в пятницу текущей недели (F_t).

Столбец 4: цена предложения доллара в единицах иностранной валюты на спотовом рынке на дату поставки из текущий форвардный контракт ($S30_t$).

Период выборки — с первой недели 1975 г. по последнюю неделю 1989 г. Объем выборки равен 775. Как и в основном тексте, определим $e_t \equiv \log(S_t)$, $f_t \equiv \log(F_t)$, $s30_t \equiv \log(S30_t)$. Если f_t является информацией, доступной в пятницу недели t , то она включает в себя $\{s_{t-4}, s_{t-3}, \dots, f_{t-4}, f_{t-3}, \dots, s30_{t-4}, s30_{t-3}, \dots\}$. Заметим, что $s30_t$ не наблюдается до момента после пятницы недели $t+4$. Определим $e_t \equiv s30_t - f_t$.

Выберите предпочтительную вами валюту для ответа на следующие вопросы:

- (a) (Работа в библиотеке/интернете.) Для вашего выбора иностранной валюты идентифицируйте неделю, когда абсолютное значение форвардной премии наибольшее. Для этой недели найдите некоторую меру внутренней одномесячной процентной ставки (например, одномесячной ставки безрискового ссуда) для США и немецкого валюты для примерки того, что дифференциал процентной ставки соответствует форвардному премия.
- (г) (Коррелограмма $\{e_t\}$.) Изобразите выборочную коррелограмму для e_t с 40 лагами. Исчезает ли автокорреляция после 4 лагов? (Важно помнить, вычитать ли выборочное среднее при вычислении выборочных корреляций. Теория говорит нам, что безусловное среднее равно нулю, что вы, возможно, захотите использовать в расчетах. При вычислении коррелограммы

для обменного курса иена/доллар, показанной на рис. 6.2, выборочное среднее было вычтено.)

- (c) (Является ли логарифм спотовой ставки случайным блужданием?) Изобразите выборочную коррелограмму $\hat{\gamma}_{k+1} - \hat{\gamma}_k$ с 40 лагами. Для этих 40 автокорреляций используйте статистику Льюнга — Бокса для тестирования на сериальную корреляцию. Можете ли вы отвергнуть гипотезу о том, что $\{x_t\}$ является случайным блужданием со сносом?

- (d) (Безусловный тест.) Проведите безусловный тест. Можете ли вы повторить результаты в табл. 6.1 для выбранной валюты?

- (e) (Дополнительно, регрессионный тест с усеченным ядром.) Проведите регрессионный тест. Можете ли вы повторить результаты в табл. 6.2 для выбранной валюты?

Советы для пакета RATS: Используйте LINREG с опцией ROBUSTERRORS. Для включения автокорреляций до четвертой включительно положите LAGS = 4 в LINREG. Опция DAMP = 0 указывает LINREG использовать усеченное ядро.

Советы для пакета TSP: В процедуре GMM в TSP есть опция KERNEL, но выбор ограничен ядром Бартлетта и ядром Парзена. Вы можете использовать матричные команды в TSP чтобы выполнить оценивание с усеченным ядром.

- (f) (Ядро Бартлетта.) Для проведения регрессионного теста используйте оценку для S на основе ядра Бартлетта. [Newey and West, 1994] предлагают процедуру автоматического выбора ширины окна на основе данных. Прикиньте на глаз, что глубина запаздываний автокорреляций, определенная этой процедурой, равна 13 для иены/доллар (поэтому графическая оценка S используются автоковариация до двенадцатого порядка), 9 для DM/доллар и 17 для фунта/доллар.

Советы для пакета RATS: Опция ROBUSTERRORS, DAMP = 1 указывает LINREG использовать ядро Бартлетта. Для иены/доллар, евро/доллар, положите LAGS = 12.

Советы для пакета TSP: Положите hat, kernel = bartlett в процедуре GMM. Положите kMA = 12 для иены/доллар.

Стандартная ошибка для коэффициента при $f - s$ для иены/доллар должна быть равной 0.6516.

- (g) (Дополнительно, VARHAC.) В регрессионном тесте используйте оценку VARHAC для S . Положите ρ_{max} равным 100% числа $n^{1/3}$. На шаге 1 VARHAC вы можете оценить двумерную VAR и для каждого из двух уравнений VAR можете выбрать глубину

зависимый y , используя BIC на фиксированной выборке $y = y_{\min} + 1, \dots, T$. Информационный критерий может быть взят в виде (6.6.11) или, альтернативно,

$$\log \left(N S T_{y, K}^2 / (n - p_{\max}) \right) + pK \cdot \log(n - p_{\max}) / (n - p_{\max}). \quad (15)$$

Только для того, чтобы иметь четкое представление о том, как вы выбрали p и каков ответ, используйте последнюю целевую функцию для выбора p посредством BIC. Для случая нсн/доллар выбранный глубина запаздываний должна быть равна 4 для первого уравнения VAR и 6 для второго уравнения. Для нсн/доллар стандартная ошибка для коэффициента при f — n должна быть равной 0,8027.

2. (Продолжение упражнения по работе Фамы в лабе 2.) Теперь мы знаем, как использовать трехмесячную ставку по казначейским векселям для тестирования эффективности рынка казначейских векселей США, используя месячные данные. Пусть

π_t — темп инфляции (в процентах, годовой темп), по трехмесячному месяцу, заканчивающемуся в месяц $t - 1$.

R_t — трехмесячная ставка по казначейским векселям (в процентах, годовая ставка) на начало месяца t .

Рассмотрим уравнения Фамы для проверки эффективности рынка:

$$\pi_{t+j} = \beta_0 + \beta_1 R_t + \varepsilon_t.$$

- (a) (Тривиально.) Почему зависимой переменной является π_{t+j} , а не, скажем, π_t ?
- (b) Покажите, что для гипотезы эффективности рынка $\beta_1 = 1$ и автоковариации $\{\gamma_j\}$ исчезают после второго лага (то есть $\gamma_j = 0$ для $j > 2$).
- (c) Пусть $\pi_{t-1} \equiv R_t - \pi_{t+j}$ является ex post процентной ставкой по трехмесячным казначейским векселям. Используйте *PAGE* и *MISHKIN.ASC* как меру трехмесячного темпа инфляции для вычисления ex post годовой процентной ставки. (Датировка переменных в *MISHKIN.ASC* такая, что лабильное процентной ставки за январь использует данные по казначейской ставке, истекающей в декабре, и январское лабильное для трехмесячного темпа инфляции состоит из данных CPI с декабря

по мере. Поэтому в регрессии можно сопоставлять PAB и TAB (треугольники и ставку по казначейским векселям) для одного и того же момента.) Изобразите выборочную коррелограмму для периода выборки 1/53–7/71. Что можно предположить о качестве эффективности рынка по этой коррелограмме?

- (д) Проверьте эффективность рынка, оценивая регрессию Филла для периода выборки 1/53–7/71. (Объем выборки должен быть равен 223, не выбрасывайте две трети наблюдений, как это сделал Филла в своих табл. 6–5.) Используйте (6.6.4) с $q = 2$ или 8.

Ответы на избранные вопросы

Аналитические упражнения

- 2с. Поскольку $\{c_j\}$ абсолютно суммируема, $c_j \rightarrow 0$ при $j \rightarrow \infty$. По этому для любого j существует $A > 0$, такое что $|c_{j+k}| \leq A$ для всех j, k . Следовательно, $|\psi_{j+k} \cdot c_k| \leq A|c_k|$. Поскольку $\{c_k\}$ (и, следовательно, $\{Ac_k\}$) абсолютно суммируема, это верно и для $\{\psi_{j+k} \cdot c_k\}$ ($k = 0, 1, 2, \dots$) для любого заданного j . Таким образом, в силу (1),

$$|\gamma_j| = \sigma^2 \sum_{k=0}^{\infty} c_{j+k} c_k \leq \sigma^2 \sum_{k=0}^{\infty} |\psi_{j+k} c_k| = \sigma^2 \sum_{k=0}^{\infty} |\psi_{j+k}| |c_k| < \infty.$$

Теперь положим в (1) вместо $\psi_j = c_j - c_{j-1}$. Тогда

$$\sum_{j=0}^{\infty} |c_k| = \sum_{j=0}^{\infty} |c_k - c_{j-k}| \leq |c_k| \sum_{j=0}^{\infty} |c_j| < \infty.$$

Пусть

$$M \equiv \sum_{j=0}^{\infty} |c_j| < \infty, \quad c_k \equiv |c_k| \sum_{j=0}^{\infty} |c_{j-k}|.$$

Тогда $\{c_k\}$ суммируема, поскольку $|c_k| \leq |c_k| \cdot M$ и $\{c_k\}$ абсолютно суммируема. Следовательно, из (1),

$$\sum_{j=0}^{\infty} \left(\sum_{k=0}^{\infty} c_{j+k} |c_k| \right) < \infty.$$

Эта двойная сумма, умноженная на σ^2 , больше или равна $\sum_{j=0}^{\infty} |\gamma_j|$. Поэтому $\{\gamma_j\}$ абсолютно суммируема.

7а. δ_j^2 является элементом $\{1, 1\}$ в $T(\Lambda)^2 T^{n-1}$.

8а.

$$\begin{aligned} E(y_t) &= \frac{1 - \phi^t}{1 - \phi} c + \phi^t E(y_0) = \\ &= (1 - \phi^t) \mu + \phi^t E(y_0) \\ &\quad (\text{поскольку } \mu = c/(1 - \phi)) \\ &= \mu + \phi^t \cdot [E(y_0) - \mu]. \\ \text{Var}(y_t) &= (1 + \phi^2 + \phi^4 + \dots + \phi^{2(t-2)}) \sigma^2 + \phi^{2t} \text{Var}(y_0) \\ &\quad (\text{поскольку } \text{Cov}(y_t, y_t) = 0) \\ &= \frac{1 - \phi^{2t}}{1 - \phi^2} \sigma^2 + \phi^{2t} \text{Var}(y_0) = \\ &= (1 + \phi^{2t}) \gamma_0 + \phi^{2t} \text{Var}(y_0) \\ &\quad (\text{поскольку } \gamma_0 \equiv \sigma^2/(1 - \phi^2)), \\ &= \gamma_0 + \phi^{2t} \cdot [\text{Var}(y_0) - \gamma_0]. \\ \text{Cov}(y_t, y_{t-j}) &= (\phi^j - \phi^{j+2} + \dots - \phi^{2t-2-j}) \sigma^2 + \phi^{2t-j} \text{Var}(y_0) = \\ &= \phi^j \cdot (1 + \phi^2 + \dots + \phi^{2t-2j-2}) \sigma^2 + \phi^{2t-j} \text{Var}(y_0) = \\ &= \phi^j \cdot \frac{1 - \phi^{2(t-j)}}{1 - \phi^2} \sigma^2 + \phi^{2t-j} \text{Var}(y_0) = \\ &= \phi^j \cdot (1 - \phi^{2(t-j)}) \gamma_0 + \phi^{2t-j} \text{Var}(y_0) = \\ &= \gamma_0 \phi^j + \phi^{2t-j} \cdot [\text{Var}(y_0) - \gamma_0] = \\ &= \gamma_0 + \phi^{2t-j} \cdot [\text{Var}(y_0) - \gamma_0]. \end{aligned}$$

Заключительные приложения

- 1а. Статистика Льюнга—Бокса с 40 лагами равна 84.0 с p -значением 0.00000%. Поэтому гипотеза о том, что спотовая ставка является случайным блужданием со сносом, может быть отвергнута. См. рис. 6.6 для коррелограммы кена/доллар.
- 1б. Для кены/доллар стандартная ошибка равна 3,72 для канализации и 0,68 для коэффициента при $f = \lambda$. Статистика Вальда для гипотезы о том, что $\beta_0 = 0$ и $\beta_1 = 1$, равна 21.85 ($p = 0.0027$).
- 2а. Коррелограмма для трехкратной ех-рост реальной ставки показана на рис. 6.7. Теория предсказывает, что сезонная корреляция исчезает после второго лага. Коррелограмма в основном соответствует этому предсказанию.

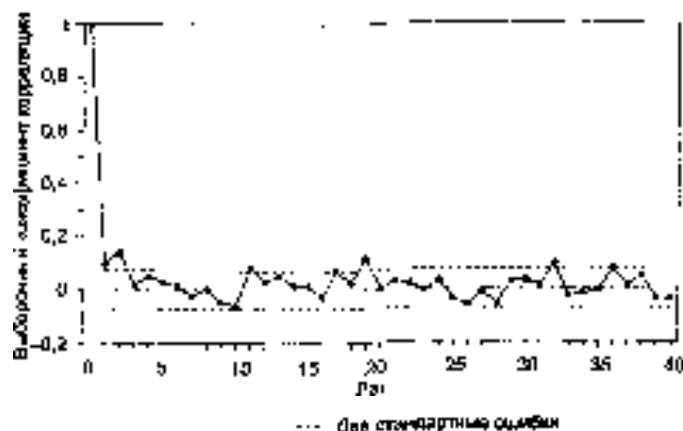
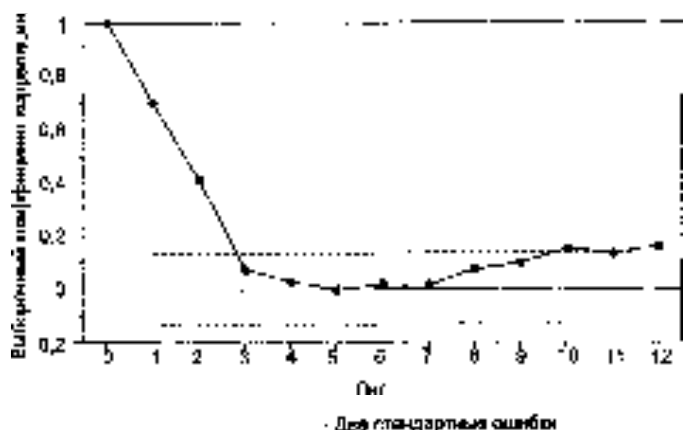
Рис. 6.6. Correlogramma $y_{t+1} - \hat{y}_t$ логар./логар

Рис. 6.7. Correlogramma трехмесячной ex rate реальной ставки

Литература

- Andrews, D., 1991, "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.
- Bekaert, G., and R. Hodrick, 1993, "On Biases in the Measurement of Foreign Exchange Risk Premiums," *Journal of International Money and Finance*, 12, 115-136.
- Brockwell, P., and R. Davis, 1991, *Time Series: Theory and Methods* (2nd ed.), Berlin: Springer-Verlag.
- Don Hanh, W., and A. Levin, 1996a, "Inferences from Parametric and Non-Parametric Covariance Matrix Estimation Procedures," Technical Working Paper Series 195, NBER.

- , 1990b, "A Practitioner's Guide to Robust Covariance Matrix Estimation," Technical Working Paper Series 197, NBER.
- Fueller, W., 1996, *Introduction to Statistical Time Series* (2d ed.), New York: Wiley.
- Gordin, M. I., 1969, "The Central Limit Theorem for Stationary Processes," *Soviet Math. Dokl.*, 10, 1174–1176.
- Gourieroux, C., and A. Monfort, 1997, *Time Series and Dynamic Models*, Cambridge: Cambridge University Press.
- Hannan, J., 1991, *Time Series Analysis*, Princeton: Princeton University Press.
- Hannan, E. J., 1970, *Multiple Time Series*, New York: Wiley.
- Hannan, E. J., and M. Deistler, 1998, *The Statistical Theory of Linear Systems*, New York: Wiley.
- Hansen, L., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.
- Lütkepohl, H., 1993, *Introduction to Multiple Time Series Analysis* (2d ed.), New York: Springer – Verlag.
- Newey, W., and K. West, 1997, "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 65, 703–708.
- , 1994, "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.
- Ng, S., and P. Perron, 2001, "A Note on Model Selection in Time Series Analysis," mimeo., Boston: Boston College.
- White, H., 1981, *Asymptotic Theory for Econometricians*, New York: Academic.

Глава 7. Экстремальная оценка

В предыдущих главах для оценивания различных рассматриваемых нами моделей мы выбирали обобщенный метод моментов (GMM). Мы также кратко описали метод максимального правдоподобия (ML). В этой главе мы расширим наш арсенал методов оценивания, исследуя класс оценок, называемых «экстремальными оценками», который включает в себя метод наименьших квадратов (линейный и нелинейный), обобщенный метод моментов (линейный и нелинейный) и метод максимального правдоподобия в качестве частных случаев. Как было подчеркнуто [Агетсиза, 1985] и другими, этот унифицированный подход полезен тем, что он выявляет общую структуру, лежащую в основе этих внешне различных методов оценивания.

Параграф 7.1 вводит в рассмотрение экстремальные оценки. Два следующих параграфа посвящены асимптотическим свойствам экстремальных оценок. Параграф 7.4 определяет триаду тестовых статистик — статистики Вальды, множителей Лагранжа и отношения правдоподобия — для экстремальных оценок в общем случае и показывает, как эти статистики конкретизируются в частных случаях. Параграф 7.5 посвящен краткому обсуждению вычислительных аспектов экстремальных оценок.

Эта глава может показаться чрезвычайно сложной, так как в ней сочетается интенсивное использование асимптотической теории с математическим анализом. Основная идея, однако, довольно проста. Если вы можете читать вузовские учебники по микро- и макроэкономике, не испытывая особых трудностей, то будете способны понять большую часть текста при повторном (если не при первом) его прочтении.

Если вас не интересует асимптотика экстремальных оценок *сама по себе*, то вы собираетесь изучить главу 8, вы можете не читать эту главу целиком. Просто прочитайте параграф 7.1, а затем постарайтесь понять, о чем идет речь в теоремах 7.3, 7.6, 7.8, 7.9 и 7.11.

Примечание, касающееся обозначений: В отличие от предыдущих глав мы будем обозначать истинное значение векторного

параметра вектором с близким (подстрочным) индексом 0. Таким образом, θ_0 — это истинные значения параметра, тогда как θ обозначает гипотетическое значение параметра. Такие обозначения стандартны в литературе для «избранных», посвященной нелинейному оцениванию. Кроме того, мы будем использовать « i » вместо « i » для обозначения номера наблюдения.

7.1. Экстремальные оценки

Оценка $\hat{\theta}$ называется **экстремальной** (*extremal*), если существует такая целевая функция $Q_n(\theta)$, что

$$\hat{\theta} \text{ максимизирует } Q_n(\theta) \text{ при условии, что } \theta \in \Theta \subset \mathbb{R}^p, \quad (7.1.1)$$

где Θ , называемое **пространством параметров** (*parameter space*), множество всех возможных значений параметра. В этой книге мы ограничиваем наше внимание случаем, когда Θ является подмножеством конечномерного Евклидова пространства $\mathbb{R}^p$. Целевая функция $Q_n(\theta)$ зависит не только от θ , но также от выборки или от наблюдений $(w_1, \dots, w_n)$, где w_i — i -е наблюдение, а n — длина выборки. В наших обозначениях эта зависимость целевой функции от выборки размера n указывает нижний индекс n . Метод максимального правдоподобия (ML) и линейный и нелинейный обобщенный метод моментов (GMM) являются частными случаями экстремальных оценок. В этом параграфе представлены эти и другие экстремальные оценки и показывается, как они связаны друг с другом.

«Измеримость» $\hat{\theta}$

Строго говоря, задача максимизации (7.1.1) обязательно имеет решение. Но вспомним следующий факт из математического анализа:

Пусть отображение $h: \mathbb{R}^p \rightarrow \mathbb{R}$ непрерывно и $A \subset \mathbb{R}^p$ — компакт (замкнутое и ограниченное множество). Тогда h имеет точку максимума на множестве A . То есть существует такой $x^* \in A$, что $h(x^*) \geq h(x)$ для любого $x \in A$.

Таким образом, если $Q_n(\theta)$ непрерывна по θ для любых данных $(w_1, w_2, \dots, w_n)$ и Θ — компакт, то тогда существует θ — решение задачи (7.1.1) для любой заданной выборки $(w_1, w_2, \dots, w_n)$. В случае когда решение неединственно, мы выберем одно из них. Таким образом, $\hat{\theta}$, однозначно определенная для любого набора данных $(w_1, w_2, \dots, w_n)$, есть функция от данных. Однако, строго говоря, того, что $\hat{\theta}$ является функцией от вектора случайных величин $(w_1, w_2, \dots, w_n)$, недостаточно для того, чтобы $\hat{\theta}$ была определена как случайная величина: необходимо,

чтобы $\hat{\theta}$ была «измеримой» функцией от $(w_1, w_2, \dots, w_n)$. (Если функция непрерывна, то она измерима.) Следующая лемма показывает, что $\hat{\theta}$ измерима, если таковой является функция $Q_n(\theta)$.

Лемма 7.1 (существование экстремальных оценок): (Предположим, что (i) пространство параметров Θ — компактное подмножество $\mathbb{R}^p$, (ii) $Q_n(\theta)$ непрерывна по θ для любых данных $(w_1, \dots, w_n)$ и (iii) $Q_n(\theta)$ — измеримая функция от данных для всех θ из Θ . Тогда существует измеримая функция $\hat{\theta}$, зависящая от наблюдений, которая является решением (7.1.1).

По поводу доказательства см., например [Gouriéroux and Monfort, 1995, Property 24.1]. Во всех примерах, приведенных в этой главе, $Q_n(\theta)$ измерима.

В большинстве прикладных задач мы не знаем верхнюю или нижнюю границу для истинного вектора параметров. Даже если мы знаем эти граничные значения, они не обязательно включены в пространство параметров, так что оно не будет замкнутым (см. пример CES далее). Таким образом, решение предположения об ограниченности и замкнутости Θ — это то, чего мы хотим избежать. Для некоторых асимптотических результатов в данной главе мы будем явственно предполагать компактность некоторыми другими условиями, которые выполняются во многих прикладных задачах.

Два класса экстремальных оценок

В следующих двух параграфах мы получим асимптотические результаты для экстремальных оценок, а затем конкретизируем эти результаты для следующих классов двух экстремальных оценок.

1. **M-оценки.** Экстремальная оценка является **M-оценкой** (*M-estimator*), если целевая функция является выборочным средним:

$$Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n m(w_i; \theta), \quad (7.1.2)$$

где m — функция $(w_i; \theta)$, принимающая действительные значения. Два изучаемых нами примера, которые дают M-оценки, — это метод максимального правдоподобия (ML) (*maximum likelihood*) и нелинейный метод наименьших квадратов (NLS) (*nonlinear least squares*).

2. **GMM.** Экстремальная оценка является **GMM-оценкой** (*GMM estimator*), если целевая функция может быть записана в виде

$$Q_n(\theta) = -\frac{1}{2}g_n(\theta)' \tilde{W} g_n(\theta) \quad \text{с} \quad g_n(\theta) \equiv \frac{1}{n} \sum_{i=1}^n g(w_i; \theta), \quad (7.1.3)$$

где $\tilde{W}$ — симметричная положительно определенная матрица размера $K \times K$, определяющая расстояние до $g_n(\theta)$ от нуля. Она может зависеть от данных. Максимизация целевой функции равносильна минимизации расстояния $g_n(\theta)' \tilde{W} g_n(\theta)$, что соответствует нашему определению GMM в главе 3. Как вскоре станет понятно, деление расстояния на два упрощает выражение для производной целевой функции. Мы вводили GMM-оценивание для линейных моделей, в которых $g(w_i; \theta)$ линейна по θ . Ниже мы покажем, что GMM-оценивание может быть распространено и на нелинейные модели.

Существуют также экстремальные оценки, которые не попадают ни в один из этих двух классов. Самый известный пример таких методов оценивания — **классические оценки минимального расстояния** (*classical minimum distance estimators*). Целевая функция в этом случае может быть представлена как $-g_n(\theta)' \tilde{W} g_n(\theta)$, однако $g_n(\theta)$ необязательно является выборочным средним. Теорема о состоятельности в следующем параграфе достаточно обща, чтобы охватить и такого рода оценки.

В оставшейся части этого параграфа мы представим примеры M-оценок и GMM-оценок. Каждый пример является комбинацией *модели* (множества процессов, порождающих данные, {DGP}) и *метода оценивания*.

Метод максимального правдоподобия (ML)

Основной пример M-оценки — это метод максимального правдоподобия в случае, когда $\{w_i\}$ — i.i.d. (независимые, одинаково распределенные случайные величины). Моделью здесь является множество i.i.d.-последовательностей $\{w_i\}$, где функция плотности вероятности w_i принадлежит семейству плотностей, описываемых конечномерным вектором θ : $f(w_i; \theta)$, $\theta \in \Theta$. (Поскольку $\{w_i\}$ одинаково распределены, функциональная форма $f(\cdot; \cdot)$ не зависит от i .) Функциональная форма $f(\cdot; \cdot)$ известна. Модель **параметрическая** (*parametric*), так как вектор параметров θ конечномерный. При истинном значении параметра θ_0 , функция плотности вероятности истинного DGP (того DGP, который на самом деле породил данные) есть $f(w_i; \theta_0)$ ¹. Мы говорим, что модель **правильно специфицирована** (*correctly specified*), если $\theta_0 \in \Theta$.

¹ Если бы мы следовали обозначениям, введенным в предыдущих параграфах, то истинное значение параметра записывали бы как θ , а гипотетическое значение как $\tilde{\theta}$. В этом главе мы используем θ_0 для обозначения истинного значения и θ — для обозначения гипотетического значения вектора параметров.

Так как w_i распределены независимо, совместная плотность вероятности данных $(w_1, w_2, \dots, w_n)$ равна:

$$f(w_1, w_2, \dots, w_n; \theta_0) = \prod_{i=1}^n f(w_i; \theta_0). \quad (7.1.4)$$

Поскольку распределение данных специфицировано плотностью, естественным методом оценивания является метод максимального правдоподобия, который производится следующим образом. После замены вектора истинных параметров θ_0 на вектор гипотетических значений θ эта функция плотности, рассматриваемая как функция от θ , называется **функцией правдоподобия** (*likelihood function*). Оценкой параметра θ_0 по методу максимального правдоподобия (ML) является значение θ , максимизирующее функцию правдоподобия. Так как логарифмирование — монотонное преобразование, максимизация функции правдоподобия эквивалентна максимизации **логарифма функции правдоподобия** (*log likelihood function*)¹:

$$\ln f(w_1, w_2, \dots, w_n; \theta) = \ln \left[\prod_{i=1}^n f(w_i; \theta) \right] = \sum_{i=1}^n \ln f(w_i; \theta). \quad (7.1.5)$$

Оценка ML для θ_0 , таким образом, является M-оценкой с

$$\rho(w_i; \theta) = \ln f(w_i; \theta), \quad \text{то есть} \quad Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n \ln f(w_i; \theta). \quad (7.1.6)$$

Прежде чем иллюстрировать ML примером, полезно сделать три замечания:

- (Автокоррелированные наблюдения.) Среднее значение логарифма функции правдоподобия не будет иметь такую простую форму, если $\{w_i\}$ будут автокоррелированы. Примеры ML для автокоррелированных наблюдений будут изучены в следующей главе.
- (Альтернативные оценки.) Метод максимального правдоподобия — не единственный способ оценить вектор параметров. Например, пусть $\mu(\theta)$ — математическое ожидание w_i , соответствующее функции плотности вероятности $f(w_i; \theta)$. Это каноничная функция от θ .

¹Для θ , таких, что $f(w_i; \theta) = 0$, операция логарифмирования не определена. Это не является проблемой, потому что такое значение θ никогда не максимизирует функцию правдоподобия (заметим, что $f(w_i; \theta) \geq 0$). Для таких значений θ мы можем приписать $\ln f(w_i; \theta)$ какое угодно малое значение таким образом, что значение максимизирующей функцию правдоподобия, будет также максимизировать и логарифм функции правдоподобия. См. White, 1984, p. 15, для более подробного рассмотрения этого вопроса.

По построению, должно выполняться следующее условие равенства нулю математического ожидания:

$$E[w_t - \mu(\theta_0)] = 0. \quad (7.1.7)$$

Вектор параметров θ_0 может быть оценен посредством GMM с $g(w_t; \theta) = w_t - \mu(\theta)$ в качестве целевой функции GMM (7.1.3).

- (Эффективность ML.) Как будет показано в двух следующих параграфах, при выполнении соответствующих условий ML состоятельна и асимптотически нормальна. Достаточно распространенной была вера в то, что оценка ML эффективна (то есть обладает минимальной асимптотической дисперсией) в классе всех состоятельных и асимптотически нормальных оценок. С помощью контрпримера (описанного, например, в работе [Amemiya, 1985, Example 4.2.4]) было показано, что эта вера является ложной. Тем не менее оценка ML является эффективной в достаточно обширных классах асимптотически нормальных оценок. Одним из таких общих классов являются GMM-оценки. В параграфе 7.3 мы отметим, что асимптотическая дисперсия любой оценки θ_0 , полученной GMM, не ниже, чем у оценки ML.

Пример 7.1 (оценивание математического ожидания нормального распределения): Пусть данные $(w_1, \dots, w_n)$ — последовательность скалярных i.i.d. полученных величин и распределение w_t задается $N(\mu, \sigma^2)$. Таким образом, $\theta = (\mu, \sigma^2)'$

$$f(w_t; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(w_t - \mu)^2}{2\sigma^2} \right].$$

Среднее значение логарифма функции правдоподобия для данных $(w_1, \dots, w_n)$

$$\frac{1}{n} \sum_{t=1}^n \ln f(w_t; \mu, \sigma^2) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma^2) - \frac{1}{n} \sum_{t=1}^n \left[\frac{(w_t - \mu)^2}{2\sigma^2} \right]. \quad (7.1.8)$$

Оценка ML для (μ_0, σ_0^2) — это экстремальная оценка, где в качестве $Q_n(\theta)$ выступает это среднее значение логарифма функции правдоподобия. Предположим, что дисперсия предполагается положительной, но никак не ограничена на μ_0 *a priori* не задается; так что пространство параметров Θ — это $\mathbb{R} \times \mathbb{R}_+$, где $\mathbb{R}_+$ — множество положительных действительных чисел. Нетрудно проверить, что ML-оценка для μ_0 — это выборочное среднее w_t , GMM-оценка для μ_0 , основанная на условии равенства нулю математического ожидания $E(w_t - \mu_0) = 0$, также является выборочным средним w_t .

Метод углового максимального правдоподобия

В большинстве приложений вектор y_i разбивается на две группы — y_1 и x_1 , и интерес для исследователя представляет изучение того, как x_1 влияет на условное распределение y_1 при заданном x_1 . Мы называли переменную y_1 зависимой переменной, а x_1 — регрессорами. Никакие результаты в этой главе не зависят от того, является ли y_1 скаляром; y_1 может быть вектором, и все равно все результаты останутся верными, если вместо скаляра « y_1 » написать вектор « y_1 ». Тем не менее мы будем использовать скалярное обозначение просто потому, что во всех примерах, рассматриваемых в этой главе, будет все-таки одна зависимая переменная.

Пусть $f(y_1|x_1;\theta_0)$ — условная функция плотности y_1 при заданном x_1 , а $f(x_1;\psi_0)$ — безусловная (частная) функция плотности x_1 . Тогда

$$f(y_1, x_1; \theta_0, \psi_0) = f(y_1|x_1; \theta_0)f(x_1; \psi_0) \quad (7.1.9)$$

будет совместной функцией плотности вероятности $(y_1, x_1)'$. Пока что предположим, что θ_0 и ψ_0 не связаны функциональной зависимостью (см. следующий абзац для примера функциональной связи). Среднее значение логарифма функции правдоподобия для данных $(y_1, x_1, \dots, y_n, x_n)$ равно:

$$\frac{1}{n} \sum_{i=1}^n \ln f(y_i; \theta, \psi) = \frac{1}{n} \sum_{i=1}^n \ln f(y_i|x_i; \theta) + \frac{1}{n} \sum_{i=1}^n \ln f(x_i; \psi). \quad (7.1.10)$$

Первый элемент, стоящий справа от знака равенства, — это среднее значение логарифма условной функции правдоподобия (*log conditional likelihood*). **Оценка условного ML** (*conditional ML estimator*) для θ_0 максимизирует этот первый элемент, игнорируя, таким образом, второй элемент. Это M-оценка с

$$\psi(y_i; \theta) = \ln f(y_i, x_i; \theta) \quad \text{то есть} \quad Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n \ln f(y_i|x_i; \theta). \quad (7.1.11)$$

Второй элемент, стоящий справа от знака равенства в (7.1.10), — среднее значение логарифма частной функции правдоподобия. Оно не зависит от θ , так что оценка условного ML для θ_0 численно совпадает с (совместной или полной) оценкой ML, которая максимизирует среднее значение совместной функции правдоподобия (7.1.10).

Предполагая теперь, что θ_0 и ψ_0 связаны друг с другом функциональной зависимостью. Например, векторы θ_0 и ψ_0 можно разбить как $\theta_0 = (\alpha_0', \beta_0')$, $\psi_0 = (\beta_0', \gamma_0')$. В этом случае оценки полного (безусловного) ML и условного ML перестанут быть численно совпадающими.

Контингентно понятно, что условному МЛ достаточно информации, которая могла бы быть получена из частной функции правдоподобия. На самом деле можно показать, что оценка условного МЛ для θ_0 эквивалентна, чек оценка полного МЛ, полученная из задачи максимизации логарифма совместной функции правдоподобия (7.1.10) (см., например: [Coulter and Moffitt, 1995, Section 7.3.2]). В пользу этого предположения исследования для потери эффективности невозможны, так как мы не знаем, а не хотим специфицировать параметрическую форму $f(x_i; \theta)$ частного распределения.

Вот для примера условного МЛ (другие примеры будут в следующей главе).

Пример 7.2 (модель линейной регрессии с нормальными ошибками): В главе 2 мы рассматривали модель линейной регрессии с условной гомоскедастичностью. Предположим далее, что ошибка имеет нормальное распределение, так что

$$\{y_i, x_i\} \text{ i.i.d.}, \quad y_i = x_i' \beta_0 + \varepsilon_i, \quad \varepsilon_i | x_i \sim N(0, \sigma_0^2) \quad (7.1.12)$$

Условная функция правдоподобия g_i при условии x_i , $f(y_i | x_i; \theta)$, — это функция плотности для $N(x_i' \beta, \sigma^2)$. Таким образом, для $\theta = (\beta', \sigma^2)'$ и $\omega_i = (y_i, x_i)$ функция g_i будет выглядеть как

$$\begin{aligned} g_i(\omega_i; \theta) &= \frac{1}{\sigma} f(y_i | x_i; \beta, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(y_i - x_i' \beta)^2}{2\sigma^2} \right] = \\ &= \frac{1}{\sqrt{2\pi}} \ln(\pi) - \ln(\sigma) - \frac{1}{2} \left(\frac{y_i - x_i' \beta}{\sigma} \right)^2 \end{aligned} \quad (7.1.13)$$

Процессом параметров θ будет $\mathbb{R}^K \times \mathbb{R}_{++}$, где K — размерность β , а $\mathbb{R}_{++}$ — множество положительных действительных чисел, что отражает *априорное* предположение на дисперсию ошибки: $\sigma_0^2 > 0$. Как было показано в главе 1, оценкой МЛ для β_0 является оценка МНК, а оценкой МЛ для σ_0^2 будет сумма квадратов остатков, деленная на размер выборки n .

Пример 7.3 (модель пробит): В модели *пробит* (*probit model*) скалярная зависящая переменная y_i бинарна: $y_i \in \{0, 1\}$. Например, $y_i = 1$, если женщина решит выйти на рынок труда, и $y_i = 0$ в противном случае; а x_i может включать заработок мужа. Условное распределение y_i при заданном векторе x_i рассматривается как

$$\begin{cases} P(y_i = 1 | x_i; \theta_0) = \Phi(x_i' \theta_0), \\ P(y_i = 0 | x_i; \theta_0) = 1 - \Phi(x_i' \theta_0), \end{cases} \quad (7.1.14)$$

где $\Phi(\cdot)$ — функция распределения стандартного нормального распределения. Более компактно это можно записать как

$$f(y; x; \theta_0) = \Phi(x'\theta_0)^y (1 - \Phi(x'\theta_0))^{1-y}.$$

Если не заданы никакие *априорные* ограничения на θ_0 , множество параметров Θ представляет собой $\mathbb{R}^p$. Оценка ML для θ_0 будет ML-оценкой с функцией цели вида:

$$m(x; \theta) = \ln f(y|x, \theta) = y \ln \Phi(x'\theta) + (1 - y) \ln[1 - \Phi(x'\theta)]. \quad (7.1.15)$$

Инвариантность оценок ML

Оценки безусловного и условного ML обладают некоторыми важными свойствами. Одно из них, которое выполняется в конечных выборках, — это **инвариантность** (*invariance*). Чтобы описать свойство инвариантности для экстремальных оценок в общем случае, рассмотрим **репараметризацию** (*reparameterizing*) модели с помощью отображения или функции $\lambda = \tau(\theta)$, определенной на Θ . Пусть Λ будет множеством значений отображения:

$$\Lambda \equiv \tau(\Theta) = \{\lambda; \lambda = \tau(\theta) \text{ для некоторого } \theta \in \Theta\}. \quad (7.1.16)$$

По определению, для любого λ из Λ существует хотя бы один вектор θ , такой, что $\lambda = \tau(\theta)$. Отображение $\tau: \Theta \rightarrow \Lambda$ называется репараметризацией, если оно однозначное, то есть если для каждого λ из Λ существует ровно один θ , такой, что $\lambda = \tau(\theta)$. Этот единственный θ является, таким образом, функцией от λ . Это отображение Λ на Θ называется **обратным** (*inverse*) к τ и обозначается как τ^{-1} . Мы будем говорить, что экстремальная оценка $\hat{\theta}$ **инвариантна** (*invariant*) к репараметризации τ , если экстремальной оценкой для репараметризованной модели будет $\tau(\hat{\theta})$. Пусть $Q_n(\lambda)$ — целевая функция, соответствующая репараметризованной модели. Экстремальная оценка будет инвариантной, если и только если

$$\hat{Q}_n(\lambda) = Q_n(\tau^{-1}(\lambda)) \quad \text{для всех } \lambda \in \Lambda. \quad (7.1.17)$$

Чтобы доказать это, положим $\tilde{\lambda} = \tau(\tilde{\theta})$. Для любого $\lambda \in \Lambda$ мы имеем $\hat{Q}_n(\lambda) = Q_n(\tau^{-1}(\lambda)) \leq Q_n(\tilde{\theta})$, поскольку $\tilde{\theta}$ максимизирует $Q_n(\theta)$ по множеству Θ , а $\tau^{-1}(\lambda)$ принадлежит Θ . Но $Q_n(\tilde{\theta}) = Q_n(\tau^{-1}(\tilde{\lambda})) = \hat{Q}_n(\tilde{\lambda})$. Так что $\hat{Q}_n(\lambda) \leq \hat{Q}_n(\tilde{\lambda})$ для всех $\lambda \in \Lambda$.

Метод максимального правдоподобия (условный и безусловный) инвариантен к репараметризации, поскольку функция правдоподобия после репараметризации принимает вид $\tilde{f}(y; \lambda) = f(y; \tau^{-1}(\lambda))$, что дает оценку

функцию, которая удовлетворяет (7.1.17). Существует ли коинвариантная оценка? Ответ: да, может: в вопросе 5 вас попросят проверить, что оценка ПММ не является инвариантной.

Нелинейный метод наименьших квадратов (NLS)

В NLS, как и в условном ML, вектор наблюдений y_t разбивается на две группы: y_t и x_t ($y_t = (y_t, x_t')'$). Моделью в NLS является набор случайных процессов $\{y_t, x_t\}$, таких, что условное математическое ожидание $E(y_t|x_t)$, являющееся функцией от x_t , принадлежит семейству функций $\varphi(x_t; \theta)$, $\theta \in \Theta$. Функциональная форма $\varphi(\cdot; \cdot)$ известна. Если θ_0 — истинное значение вектора параметров, то $E(y_t|x_t) = \varphi(x_t; \theta_0)$ для истинного процесса порождения данных $\{y_t, x_t\}$. Если мы определим $\varepsilon_t = y_t - E(y_t|x_t)$, то правильно специфицированная модель может быть записана как

$$y_t = \varphi(x_t; \theta_0) + \varepsilon_t, \quad E(\varepsilon_t|x_t) = 0, \quad \theta_0 \in \Theta. \quad (7.1.18)$$

Наиболее часто использующийся метод для оценивания θ_0 — это метод **наименьших квадратов** (*least squares*), который минимизирует сумму квадратов остатков. Таким образом, оценивание NLS, которое является методом наименьших квадратов, примененным к вышеописанной модели, является М-оценкой с

$$\rho(y_t; \theta) = -[y_t - \varphi(x_t; \theta)]^2, \quad \text{то есть} \quad Q_n(\theta) = -\frac{1}{n} \sum_{t=1}^n [y_t - \varphi(x_t; \theta)]^2. \quad (7.1.19)$$

Здесь максимизация $Q_n(\theta)$ идентична минимизации суммы квадратов остатков.

Пример 7.4 (производственная функция CES с аддитивными ошибками): В качестве иллюстрации рассмотрим производственную функцию CES¹

$$Q_t = A \cdot [\delta_0 K_t^\rho + (1 - \delta_0) L_t^\rho]^{-1/\rho} + \varepsilon_t, \quad (7.1.20)$$

где Q_t — выпуск в период времени t , K_t — затраты капитала, L_t — затраты труда, а ε_t — аддитивный к производственной функции шок с $E(\varepsilon_t|K_t, L_t) = 0$. Это модель условного математического ожидания (7.1.18) с $y_t = Q_t$, $x_t = (K_t, L_t)'$, $\theta_0 = (A, \delta_0, \rho)'$, $\varphi(x_t; \theta) = A \cdot [\delta K_t^\rho + (1 - \delta) L_t^\rho]^{-1/\rho}$. Стандартные свойства производственных функций (монотонность и вогнутость) выполняются, если $A > 0$, $0 < \rho < 1$, $1 < \rho$, и это определит пространство параметров Θ .

¹ CES — Constant elasticity of substitution — производственная функция с постоянной эластичностью замещения. — Прим. науч. ред. перевод.

Линейный и нелинейный GMM

В главе 3 мы применяли GMM для оценивания линейных уравнений. Оцениваемым линейным уравнением было

$$y_t = x_t' \theta_0 + \varepsilon_t. \quad (7.1.21)$$

Вектором инструментов был x_t , а условиями ортогональности (равенства нулю математического ожидания) были уравнения:

$$E[x_t \cdot (y_t - x_t' \theta_0)] = 0. \quad (7.1.22)$$

Правильно специфицированной моделью здесь является набор эргодических стационарных процессов $w_t = (y_t, x_t', x_t')$, таких, что эти условия равенства нулю математических ожиданий выполняются для θ_0 , принадлежащего Θ . Оценкой линейного GMM для θ_0 является GMM-оценка с функцией g в целевой функции GMM (7.1.13), заданной следующим образом:

$$g(w_t; \theta) = x_t \cdot (y_t - x_t' \theta) = x_t \cdot y_t - x_t x_t' \theta. \quad (7.1.23)$$

GMM может быть легко применен и к нелинейным уравнениям. Предположим, что оцениваемое уравнение нелинейно, а также что y_t входит в это уравнение неявно:

$$a(y_t, z_t; \theta_0) = \varepsilon_t. \quad (7.1.24)$$

Просто возьмем в качестве функции g следующую: $g(w_t; \theta) = x_t \cdot a(y_t, z_t; \theta)$. Оценка такого вида называется **обобщенной оценкой нелинейного метода инструментальных переменных** (*generalized nonlinear instrumental variables estimator*)¹. Это все еще особый вид GMM, потому что функция g здесь может быть записана как произведение вектора инструментов и ошибки. Рассматриваемые в данной главе свойства GMM-оценок не опираются на эту особую структуру.

В качестве примера нелинейного GMM рассмотрим нелинейное уравнение Эйлера [Hansen and Singleton, 1982].

Пример 7.5 (нелинейное уравнение Эйлера для потребления): Уравнение Эйлера для оптимизационной задачи домохозяйства, являющееся стандартным для макроэкономики, имеет вид:

$$E \left[R_{t+1} \frac{\lambda_t u'(c_{t+1})}{u'(c_t)} \mid \mathcal{F}_t \right] = 1, \quad (7.1.25)$$

где R_{t+1} — задовая ех пост норма отдачи (1 плюс норма отдачи) на рассматриваемом активе, c_t — потребление, λ_t — дисконт-фактор, $u'(c_t)$ —

¹ Эта оценка более общая, нежели оценка для линейного метода инструментальных переменных, потому что здесь не предполагается условная homoskedasticность.

предельная полезность, а I_t — информация, доступная в момент времени t . Возьмем в качестве функции полезности $u(c) = c^{1-\alpha_0}/(1-\alpha_0)$. Тогда $u'(c) = c^{-\alpha_0}$, и уравнение Эйлера приобретает вид:

$$E\left[a\left(\frac{c_{t+1}}{c_t}, R_{t+1}; \alpha_0, \beta_0\right) \mid I_t\right] = 0, \\ \text{где } a\left(\frac{c_{t+1}}{c_t}, R_{t+1}; \alpha_0, \beta_0\right) \equiv R_{t+1} \cdot \beta_0 \cdot \left(\frac{c_{t+1}}{c_t}\right)^{-\alpha_0} - 1. \quad (7.1.26)$$

Если $\mathbf{x}_t$ — вектор переменных, значения которых известны в момент времени t , то $\mathbf{x}_t \in I_t$, и, в соответствии с (7.1.26),

$$E\left[\mathbf{x}_t \cdot a\left(\frac{c_{t+1}}{c_t}, R_{t+1}; \alpha_0, \beta_0\right) \mid I_t\right] = 0. \quad (7.1.27)$$

Применяя оператор безусловного математического ожидания к обеим сторонам и используя закон полных математических ожиданий ($E[E(x|I_t)] = E(x)$), мы получим условия ортогональности (равенства нулю математического ожидания) $E[g(\mathbf{w}_t; \boldsymbol{\theta}_0)] = 0$, где

$$g(\mathbf{w}_t; \boldsymbol{\theta}_0) = \mathbf{x}_t \cdot a\left(\frac{c_{t+1}}{c_t}, R_{t+1}; \alpha_0, \beta_0\right) \quad \text{с} \quad \mathbf{w}_t = \begin{bmatrix} \mathbf{x}_t \\ c_{t+1}/c_t \\ R_{t+1} \end{bmatrix}, \quad \boldsymbol{\theta}_0 = \begin{bmatrix} \beta_0 \\ \alpha_0 \end{bmatrix}. \quad (7.1.28)$$

Контрольные вопросы

1. (Оценивание модели пробит посредством NLS.) Проверьте, что для модели пробит $E(y_t | \mathbf{x}_t) = \Phi(\mathbf{x}_t' \boldsymbol{\theta}_0)$. Рассмотрите применение метода наименьших квадратов к этой модели. Специфицируйте функцию m для целевой функции М-оценки (7.1.2).
2. (Оценивание модели пробит посредством GMM.) Проверьте, что для модели пробит выполняются условия ортогональности $E(\mathbf{x}_t \cdot (y_t - \Phi(\mathbf{x}_t' \boldsymbol{\theta}_0))) = 0$. Специфицируйте функцию g для целевой функции оценки GMM (7.1.13).
3. (Оценивание модели GMM посредством ML?) Предположим, что величины $\{y_t, z_t, \mathbf{x}_t\}$ i.i.d. в правильно специфицированной линейной модели, описанной в последнем подпараграфе. Сможете ли вы оценить $\boldsymbol{\theta}_0$ с помощью ML? [Ответ: Нет. Потому что эта модель не специфицирует параметрическую форму функции плотности $\mathbf{w}_t$.]
4. (Отличие $E(\varepsilon_t | \mathbf{x}_t) = 0$ от $E(\varepsilon_t \cdot \mathbf{x}_t) = 0$ в NLS.) Рассмотрите квадратичную модель NLS, где $\varphi(\mathbf{x}_t; \boldsymbol{\theta}) = \theta_0 + \theta_1 x_t + \theta_2 x_t^2$. Предположим, что $E(\varepsilon_t | \mathbf{x}_t) = 0$, но $E(\varepsilon_t x_t^2) \neq 0$. Выполняется ли условие $E(\varepsilon_t \cdot \mathbf{x}_t) = 0$? [Ответ: Нет. Если $E(\varepsilon_t | \mathbf{x}_t) = 0$, то $E(\varepsilon_t h(\mathbf{x}_t)) = 0$ для любой (измеримой) функции h .] Будет ли оценка NLS для $\boldsymbol{\theta}_0$ состоятельной? [Ответ: Нет.]

3. (Отсутствие свойства инвариантности для GMM.) Рассмотрим линейную модель $y_i = \theta_0 x_i + \varepsilon_i$, где θ_0 и ε_i — скаляры. Условия ортогональности: $E[y_i \cdot (y_i - \theta_0 x_i)] = 0$. Запишите целевую функцию GMM $Q_n(\theta)$. Предположите, что $\Theta = \mathbf{R}_+$ (то есть $\theta_0 \geq 0$), и рассмотрите репараметризацию $\lambda = 1/\theta$. Используя эту репараметризацию, данное уравнение можно переписать как $x_i = \lambda_0 y_i - \lambda_0 \varepsilon_i$, а условия ортогональности можно переписать как $E[x_i \cdot (x_i - \lambda_0 y_i)] = 0$. Пусть $\tilde{Q}_n(\lambda)$ будет целевой функцией GMM, ассоциирующей с этим набором условий ортогональности. Будут ли в данном случае выполняться равенства $Q_n(\theta) = \tilde{Q}_n(1/\theta)$ или $\tilde{Q}_n(\lambda) = Q_n(1/\lambda)$, если $\tilde{W}$ одинакова в обеих целевых функциях? (Ответ: Нет!)

7.2. Состоятельность

В этом параграфе мы сначала представим набор достаточных условий, при выполнении которых экстремальная оценка состоятельна. Эти условия будут потом адаптированы для NLS, ML и GMM.

Две теоремы о состоятельности экстремальных оценок

Целевая функция $Q_n(\cdot)$ — случайная функция, потому что для каждого значения θ ее значение $Q_n(\theta)$, будучи зависящим от данных $(y_1, \dots, y_n)$, является случайной величиной. Основополагающая идея состоятельности экстремальной оценки состоит в том, что если $Q_n(\theta)$ сходится по вероятности к $Q_0(\theta)$ и истинный параметр θ_0 решает «предельную задачу» максимизации предельной функции $Q_0(\theta)$, то пределом максимизирующего значения $\hat{\theta}$ должен быть вектор θ_0 . Достаточными условиями того, что максимизирующее значение предельной функции будет пределом максимизирующего значения, являются требование к сходимости по вероятности быть «равномерной» (мы проясним смысл этого очень скоро) и требование к пространству параметров Θ быть компактным множеством. Как уже говорилось, пространство параметров не является компактом в большинстве прикладных задач. Поэтому мы приведем альтернативный набор условий состоятельности, которые не требуют, чтобы Θ было компактом.

Чтобы представить первую теорему о состоятельности экстремальных оценок, надо уточнить, что же мы имеем в виду, когда говорим, что сходимость «равномерная». Если вы уже знакомы из математического анализа с понятием равномерной сходимости последовательности функций, поймите, что определение, которое вы сейчас увидите, — это расширенное определение на последовательность случайных функций, $Q_n(\cdot)$ ($n = 1, 2, \dots$). Точечная сходимость по вероятности (pointwise convergence in probability) $Q_n(\cdot)$ к некоторой неслучайной функции $Q_0(\cdot)$ просто означает, что $\lim_{n \rightarrow \infty} Q_n(\theta) = Q_0(\theta)$ для всех θ , а именно что последовательность случайных величин $|Q_n(\theta) - Q_0(\theta)|$ ($n = 1, 2, \dots$)

сходится по вероятности к 0 для любого θ . **Равномерная сходимость по вероятности** (*uniform convergence in probability*) сильнее. Сходимость должна происходить «равномерно» на пространстве параметров Θ в следующем смысле:

$$\sup_{\theta \in \Theta} |Q_n(\theta) - Q_0(\theta)| \xrightarrow{P} 0 \quad \text{при } n \rightarrow \infty. \quad (7.2.1)$$

Как и в случае сходимости по вероятности, можно просто расширить определение равномерной сходимости по вероятности на последовательности *векторных* случайных функций или *матричных* случайных функций (рассматривая матрицы как векторы, элементы которых были переставлены), требуя равномерной сходимости каждого элемента: последовательность векторных случайных функций $\{h_n(\cdot)\}$ равномерно сходится по вероятности к неслучайной функции $h_0(\cdot)$, если каждый элемент сходится равномерно. Эта поэлементная сходимость эквивалентна сходимости по норме:

$$\sup_{\theta \in \Theta} \|h_n(\theta) - h_0(\theta)\| \xrightarrow{P} 0 \quad \text{при } n \rightarrow \infty, \quad (7.2.2)$$

где $\|\cdot\|$ — обычная Евклидова норма¹.

В следующем утверждении нашей первой теоремы о состоятельности условия, обеспечивающие однозначную определенность $\hat{\theta}$, обозначены как (i)–(iii), а те условия, которые обеспечивают состоятельность, — это (a) и (b).

Утверждение 7.1 (состоятельность для компактного пространства параметров): Пусть (i) Θ — компактное подмножество $\mathbb{R}^p$, (ii) $Q_n(\theta)$ непрерывна по θ для любого набора данных $\{y_1, \dots, y_n\}$ и (iii) $Q_n(\theta)$ — измеримая функция от данных для любых значений θ из Θ (так что по лемме 7.1 экстремальная оценка $\hat{\theta}$, определенная в (7.1.1), — это корректно определенная случайная величина). Если существует функция $Q_0(\theta)$, такая что

- (a) (Идентифицируемость.) $Q_0(\theta)$ достигает своего максимального значения на Θ исключительно в $\theta_0 \in \Theta$,
 (b) (Равномерная сходимость.) $Q_n(\cdot)$ сходится равномерно по вероятности к $Q_0(\cdot)$.

то тогда $\hat{\theta} \rightarrow_p \theta_0$.

¹ Евклидова норма *р-нормы* в векторе $x = (x_1, \dots, x_p)'$, обозначаемая как $\|x\|$, определяется как $\sqrt{x_1^2 + \dots + x_p^2}$.

Мы не будем доказывать этот результат по поводу доказательства смотрите, например: [Amemiya, 1985, Theorem 4.1.1]. Для оценок, таких как оценки ML, NLS и GMM, мы позднее сформулируем условие идентификации (а) в более простом виде, чтобы его можно было проще интерпретировать интуитивно.

Теорема, которая не требует компактности Θ , выглядит следующим образом:

Утверждение 7.2 (состоятельность без требования компактности):

Пусть (i) истинный вектор параметров θ_0 — является внутренним элементом выпуклого пространства параметров Θ ($\subset \mathbb{R}^n$), (ii) $Q_n(\theta)$ возмуща¹ на пространстве параметров для любого набора данных $\{w_1, \dots, w_n\}$ и (iii) $Q_n(\theta)$ — измеримая функция от данных для любого θ из Θ . Пусть $\hat{\theta}$ — экстремальная оценка, определенная в (7.1.1) (подождите немного, чтобы узнать о ее существовании). Если существует функция $Q_0(\theta)$, такая, что

(a) (Идентифицируемость.) $Q_n(\theta)$ достигает своего максимального значения на Θ исключительно в $\theta_0 \in \Theta$,

(b) (Точечная сходимость.) $Q_n(\theta)$ сходится по вероятности к $Q_0(\theta)$ для всех $\theta \in \Theta$,

то тогда, при $n \rightarrow \infty$, $\hat{\theta}$ существует с вероятностью, приближающейся к 1, и $\hat{\theta} \rightarrow_p \theta_0$.

По поводу доказательства см.: [Newey, McFadden 1994, pp. 2133–2144]. В большинстве прикладных задач Θ — открытое выпуклое множество (как в примерах 7.1–7.4), так что условие (i) выполняется. Требуется только точечная сходимость $Q_n(\theta)$ к $Q_0(\theta)$, что проще проверять в прикладных задачах. Однако на это приходится платить требованием выпуклости $Q_n(\theta)$ для любого набора данных (при выпуклости этой функции равномерная сходимость будет следовать из точечной сходимости), но во многих прикладных задачах это условие выполняется. Ниже мы проверим, что условие выпуклости выполняется для ML-оценивания модели линейной регрессии (после репараметризации), ML-оценивания модели пробит и линейного GMM.

¹ Скалярная функция $b(x)$ называется **выпуклой**, если $\alpha b(x_1) + (1 - \alpha)b(x_2) \leq b(\alpha x_1 + (1 - \alpha)x_2)$ для всех x_1, x_2 , $\theta \leq \alpha \leq 1$. Так, линейная функция выпукла. Некоторые авторы используют термин «глобальная выпуклость», чтобы отличать этот одномерный случай от концепции «сложной выпуклости», которая требует, чтобы гессиан, вычисленный для рассматриваемого вектора параметров, был неотрицательно определен. Если функция выпукла, она непрерывна. Так что условие (ii) в этом утверждении сильнее, чем условие (ii) в предыдущей теореме.

Состоятельность M -оценок

Для M -оценок целевая функция задается соотношением (7.1.3). Если $\{w_t\}$ эргодический и стационарный, то из эргодичности вытекает те же самая сходимость $Q_n(\theta); Q_n(\theta)$ почти по вероятности и для любого значения $\theta \in \Theta$ к $Q(\theta)$, заданной

$$Q(\theta) = E[m(w_t, \theta)]. \quad (7.2.3)$$

(естественно, при условии, что $E[m(w_t, \theta)]$ существует и конечно). Чтобы использовать первый полученный результат о состоятельности (утверждение 7.1) для M -оценок, так необходимо показать, что сходимость $\frac{1}{n} \sum_{t=1}^n m(w_t, \cdot)$ к $E[m(w_t, \theta)]$ является равномерной. Стандартный способ доказать равномерную сходимость по непрерывности, который использует тип функ., что выражение является выборочным средним, — это **равномерный закон больших чисел** (*uniform law of large numbers*). Версия этого закона для эргодических стационарных процессов (приведена как Lemma 2.4 в [Newey, McFadden, 1994] и как Theorem A.2.2 в [White, 1994]) выглядит следующим образом.

Лемма 7.2 (равномерный закон больших чисел): Пусть $\{w_t\}$ — эргодический стационарный процесс. Предположим, что (i) множество Θ компактно, (ii) $m(w_t, \theta)$ непрерывна по θ для всех w_t , и (iii) $m(w_t, \theta)$ измерима по w_t для всех θ из Θ . Предположим также, что выполняется

условие доминирования (*dominance condition*): существует функция $d(w_t)$ (иногда называемая *доминирующей функцией*), такая, что $|m(w_t, \theta)| \leq d(w_t)$ для всех $\theta \in \Theta$ и $E[d(w_t)] < \infty$.

Тогда $\frac{1}{n} \sum_{t=1}^n m(w_t, \cdot)$ сходится равномерно по вероятности к $E[m(w_t, \cdot)]$ на Θ . Кроме того, $E[m(w_t, \theta)]$ — непрерывная функция от θ .

Так как, по построению $|m(w_t, \theta)| \leq \sup_{\theta \in \Theta} |m(w_t, \theta)|$ для всех $\theta \in \Theta$, мы можем использовать $\sup_{\theta \in \Theta} |m(w_t, \theta)|$ для $d(w_t)$, так что предыдущее выше условие доминирования может быть записано просто как $E[\sup_{\theta \in \Theta} |m(w_t, \theta)|] < \infty$. Равномерный закон больших чисел может быть применен к векторным случайным функциям.

Пусть $\{w_t\}$ — эргодический стационарный процесс. Предположим, что (i) пространство параметров Θ компактно, (ii) $h(w_t, \theta)$ непрерывна по θ для всех w_t , и (iii) $h(w_t, \theta)$ измерима по w_t для всех θ из Θ . Предположим также, что выполняется

условие доминирования (*dominance condition*):

$$E[\sup_{\theta \in \Theta} |h(w_t, \theta)|] < \infty.$$

Тогда $E[h(w_t, \theta)]$ — непрерывная функция от θ и $\frac{1}{n} \sum_{t=1}^n h(w_t, \cdot)$ сходится равномерно по вероятности к $E[h(w_t, \cdot)]$ на Θ .

Комбинируем эту лемму с первой теоремой состоятельности экстремальных оценок и устанавливаем, что $Q_n(\theta)$ непрерывна, если $m(\omega; \theta)$ непрерывна по θ , и что $Q_n(\theta) \rightarrow$ измеримая функция данных, если $m(\omega; \theta)$ измерим по ω , и получаем

Утверждение 7.3 (состоятельность M-оценок с компактным пространством параметров): Пусть $\{\omega_i\}$ — эргодический стационарный процесс. Предположим, что (i) пространство параметров Θ — компактное подмножество $\mathbb{R}^d$, (ii) $m(\omega; \theta)$ непрерывна по θ для почти ω , и (iii) $m(\omega; \theta)$ измерима по ω для всех $\theta \in \Theta$. Пусть $\tilde{\theta} = M$ -оценка, определенная в (7.1.1) и (7.1.2). Предположим также, что

(a) (Идентификация.) $E[m(\omega; \theta)]$ принимает максимальное значение на множестве Θ исключительно в $\theta_0 \in \Theta$,

(b) (Ограниченность.) $E[|m(\omega; \theta)|] < \infty$.

Тогда $\tilde{\theta} \rightarrow_p \theta_0$.

Однако в формулировку о состоятельности экстремальных оценок для M-оценок добавляется даже проще. Цельная функция $Q_n(\theta)$ вогнута, если $m(\omega; \theta)$ является вогнутой по θ^1 . Как сказано выше, точечная сходимость $Q_n(\theta)$ к $Q_0(\theta) (= E[m(\omega; \theta)])$ обеспечивается эргодической теоремой, если $E[m(\omega; \theta)]$ существует и ограничен для всех θ . Таким образом,

Утверждение 7.4 (состоятельность M-оценок без требования компактности): Пусть $\{\omega_i\}$ — эргодический стационарный случайный процесс. Предположим, что (i) истинный вектор параметров θ_0 является внутренним элементом выпуклого пространства параметров Θ ($\mathbb{R}^d \supset \Theta$), (ii) $m(\omega; \theta)$ непрерывна на пространстве параметров для всех ω , и (iii) $m(\omega; \theta)$ измерима по ω для всех θ из Θ . Пусть $\tilde{\theta} = M$ -оценка, определенная в (7.1.1) и (7.1.2). Предположим также, что

(a) (Идентификация.) $E[m(\omega; \theta)]$ принимает максимальное значение на множестве Θ исключительно в $\theta_0 \in \Theta$,

(b) $E[|m(\omega; \theta)|] < \infty$ (то есть $E[m(\omega; \theta)]$ существует и конечно) для всех $\theta \in \Theta$.

Тогда при $n \rightarrow \infty$ $\tilde{\theta}$ существует с вероятностью, приближающейся к 1 и $\tilde{\theta} \rightarrow_p \theta_0$.

¹ Это свойство математического ожидания, если $f(x)$ и $g(x)$ вогнутые функции, то точечная функция будет и их сумма $f(x) + g(x)$.

Следующий пример показывает, что оценка условного МЛ для модели пробит удовлетворяет условию вогнутости.

Пример 7.6 (вогнутость функции правдоподобия для модели пробит): Функция η для модели пробит задается в (7.1.15), где $\Phi(\cdot)$ — функция распределения вероятности для стандартного нормального распределения. Так как нормальное распределение симметрично, $\Phi(-x) = 1 - \Phi(x)$ и функция η может быть переписана как

$$\eta(x; \theta) = x_1' \ln \Phi(x_1' \theta) + (1 - x_1') \ln \Phi(-x_1' \theta). \quad (7.2.4)$$

Вогнутость η следует из того, что $x_1' \ln \Phi(x_1' \theta)$ и $x_1' \ln \Phi(-x_1' \theta)$ вогнуты по θ . Так как $x_1' \theta$ линейно по θ , достаточно показать, что $\ln \Phi(x)$ вогнута. Первая производная $\ln \Phi(x)$ равна

$$\frac{d \ln \Phi(x)}{dx} = \frac{\phi(x)}{\Phi(x)}, \quad (7.2.5)$$

где $\phi(x) (= \Phi'(x))$ — функция плотности для $N(0, 1)$. Известно, что $\phi(x)/\Phi(x)$ — убывающая функция. Таким образом, $\ln \Phi(x)$ — (строго) вогнутая функция.

Вогнутость после репараметризации

Утверждение 7.4 может быть легко расширено на оценки, основанные на функциях, которые будут вогнутыми после репараметризации. Предположим, что все условия утверждения, кроме вогнутости, выполнены и что существует непрерывное взаимно однозначное отображение $\tau(\theta): \Theta \rightarrow A \equiv \tau(\Theta)$, такое, что

$$\tilde{\eta}(\omega; \lambda) \equiv \eta(\omega; \tau^{-1}(\lambda))$$

вогнута по $\lambda \in A \equiv \tau(\Theta)$ — выпуклое множество. Пусть

$$\tilde{Q}_n(\lambda) \equiv \frac{1}{n} \sum_{i=1}^n \tilde{\eta}(\omega_i; \lambda) =$$

целая функция после данной репараметризации. Если что все наши предположения утверждения 7.4 выполнены для $\lambda \in A$ и $\tilde{\eta}(\omega; \lambda) \equiv \eta(\omega; \tau^{-1}(\lambda))$ — внутренняя точка A , $E[\tilde{\eta}(\omega; \lambda)]$ принимает максимизирующее значение исключительно в λ_0 и $E[\tilde{\eta}(\omega; \lambda)] < \infty$ для всех $\lambda \in A$. Тогда для достаточно большого n существует МЛ-оценка $\tilde{\lambda}$, такая, что $\|\tilde{\lambda}_n - \lambda_0\| \rightarrow 0$. Так как $\tilde{Q}_n(\lambda)$ удовлетворяет (7.1.17) для взаимно однозначного отображения τ , $\theta = \tau^{-1}(\tilde{\lambda})$ максимизирует исходную целевую

функцию $Q_n(\theta)$. Оценка $\hat{\theta}$ состоятельна, потому что

$$\begin{aligned}\text{plim}_{n \rightarrow \infty} \hat{\theta} &= \text{plim}_{n \rightarrow \infty} \tau^{-1}(\bar{\lambda}) = \\ &= \tau^{-1}(\text{plim}_{n \rightarrow \infty} \bar{\lambda}) \quad (\text{так как } \tau^{-1} \text{ непрерывно}) \quad (7.2.6) \\ &= \tau^{-1}(\lambda_0) = \theta_0.\end{aligned}$$

Пример 7.7 (Вогнутость функции правдоподобия для линейной регрессии): В условном ML-оценивании линейной модели регрессии $m(w_t; \theta)$ задана в (7.1.13), где $\theta = (\beta', \sigma^2)'$ и $w_t = (y_t, x_t')'$. Эта функция необязательно вогнута¹, но можно рассмотреть репараметризацию: $\gamma = 1/\sigma$ и $\delta = \beta/\sigma$. Это непрерывное взаимно однозначное отображение между $\Theta = \mathbb{R}^k \times \mathbb{R}_{++}$ и $\Lambda = \mathbb{R}^k \times \mathbb{R}_{++}$. Новое пространство параметров Λ выпукло. После репараметризации, в которой $\lambda = (\delta', \gamma)'$, функция m приобретает вид:

$$\tilde{m}(w_t; \lambda) = -\frac{1}{2} \ln(2\pi) + \ln(\gamma) - \frac{1}{2}(\gamma y_t - x_t' \delta)^2. \quad (7.2.7)$$

Так как $\ln(\gamma)$ вогнут по γ , он вогнут по λ . Так как $v \equiv \gamma y_t - x_t' \delta$ линейна по λ и $-\frac{1}{2}v^2$ вогнута, $-\frac{1}{2}(\gamma y_t - x_t' \delta)^2$ вогнута по λ . Так что функция $\tilde{m}$, приведенная выше, которая является суммой двух вогнутых функций (плюс константа), вогнута по λ для всех w_t . Таким образом, учитывая то, что все условия утверждения 7.4, кроме вогнутости, выполняются для θ , Θ и $\log m(w_t; \theta)$, оценка условного ML $\hat{\theta}$ будет состоятельной.

Идентификация в NLS и ML

Условия идентификации (а) (п том, что $E[m(w_t; \theta)]$ должна принимать максимальное значение исключительно в θ_0) в теоремах о состоятельности М-оценок могут быть сформулированы для NLS и ML более простым образом.

NLS

Рассмотрим сначала NLS, где $m(w_t; \theta) = -(y_t - \varphi(x_t; \theta))^2$ и $E(y_t | x_t) = \varphi(x_t; \theta_0)$. Мы показывали в параграфе 2.9, что условное математическое ожидание минимизирует среднеквадратическую ошибку, то есть для лю-

¹ Например, вторая частная производная по σ^2 может быть для θ как отрицательной, так и положительной в зависимости от значения $w_t = (y_t, x_t')$. Вспомним, что необходимым условием вогнутости для дважды дифференцируемой функции $b(x_1, \dots, x_r)$ является положительность $\partial^2 b / \partial x_j \partial x_j$ для всех $j = 1, 2, \dots, r$.

бой (измеримой) функции $h(x_t)$

$$\begin{aligned} \text{среднеквадратическая ошибка} &= E[y_t - h(x_t)]^2 = \\ &= E[y_t - E(y_t|x_t)]^2 + \\ &= E[\{F(y_t|x_t) - h(x_t)\}^2] \geq \\ &\geq E[y_t - E(y_t|x_t)]^2. \end{aligned} \quad (7.2.8)$$

Здесь мы всего лишь воспроизвели (2.9.4) в текущих обозначениях. Более того, условное математическое ожидание является по существу единственным решением задачи минимизации в следующем смысле.

Будучи функциями от x_t , $h(x_t)$ и $E(y_t|x_t)$, являются случайными величинами. Для двух случайных величин X и Y пусть $X \neq Y$ означает $\text{Prob}(X \neq Y) > 0$ (то есть $\text{Prob}(X = Y) < 1$), а $X = Y$ означает $\text{Prob}(X \neq Y) = 0$. (Если X и Y отличаются друг от друга только для тех событий, вероятность которых равна нулю, нам не надо беспокоиться об этом отличии.) Таким образом, $h(x_t) \neq E(y_t|x_t)$ означает, что $\text{Prob}(h(x_t) \neq E(y_t|x_t)) > 0$. Если $h(x_t) \neq E(y_t|x_t)$ в этом смысле, то $\{E(y_t|x_t) - h(x_t)\}^2 > 0$ с положительной вероятностью. Следовательно, $E[\{E(y_t|x_t) - h(x_t)\}^2]$ в (7.2.8) положительное $\neq$ среднеквадратическая ошибка строго больше, чем $E[\{y_t - E(y_t|x_t)\}^2]$.

Прости задав функцию $h(x_t) = \varphi(x_t; \theta)$ и заметив, что $\varphi(x_t; \theta_0) = E(y_t|x_t)$, мы можем сделать вывод, что

$$\begin{cases} E[y_t - \varphi(x_t; \theta)]^2 > E[y_t - \varphi(x_t; \theta_0)]^2, & \text{если } \varphi(x_t; \theta) \neq \varphi(x_t; \theta_0), \\ E[y_t - \varphi(x_t; \theta)]^2 = E[y_t - \varphi(x_t; \theta_0)]^2, & \text{если } \varphi(x_t; \theta) = \varphi(x_t; \theta_0). \end{cases} \quad (7.2.9)$$

Таким образом, условие идентификации — что $E[y_t - \varphi(x_t; \theta)] = E[y_t - \varphi(x_t; \theta)]^2$ максимизируется (то есть $E[y_t - \varphi(x_t; \theta)]^2$ минимизируется) исключительно в θ_0 — выполняется, если

«идентификация условного математического ожидания»

(conditional mean identification)

$$\varphi(x_t; \theta) \neq \varphi(x_t; \theta_0) \quad \text{для всех } \theta \neq \theta_0. \quad (7.2.10)$$

ML

В случае ML роль, которую играло выражение (7.2.8) для NLS, будет играть **информационное неравенство Кульбака — Лейблера** (Kullback — Leibler information inequality). Здесь мы исследуем идентификацию для условного ML (делать то же самое для безусловного или совместного ML проще). Гипотетическая функция условной плотности $f(y_t|x_t; \theta)$, будучи функцией от (y_t, x_t) , является случайной величиной

для любого значения θ . Математическое ожидание случайной величины $\ln f(y_i | x_i; \theta)$ может быть записано как¹

$$E[\ln f(y_i | x_i; \theta)] = \int \ln f(y_i | x_i; \theta) f(y_i, x_i; \theta_0, \psi_0) dy_i dx_i. \quad (7.2.11)$$

Обратите особое внимание на то, что математическое ожидание здесь берется по отношению к истинной совместной функции плотности $f(y_i, x_i; \theta_0, \psi_0)$. Информационное неравенство Куна-Лейблера (адаптированное к условным плотностям) утверждает, что

$$\begin{cases} E[\ln f(y_i | x_i; \theta)] < E[\ln f(y_i | x_i; \theta_0)], & \text{если } f(y_i | x_i; \theta) \neq f(y_i | x_i; \theta_0), \\ E[\ln f(y_i | x_i; \theta)] = E[\ln f(y_i | x_i; \theta_0)], & \text{если } f(y_i | x_i; \theta) = f(y_i | x_i; \theta_0). \end{cases} \quad (7.2.12)$$

(Конечно, при условии того, что $E[\ln f(y_i | x_i; \theta)]$ существует и что они конечные) (По поводу вывода этого неравенства см. аналитическое упрощение 1.) Отсюда немедленно следует, что условие идентификации (что $E[\ln f(y_i | x_i; \theta)]$ максимизируется исключительно в θ_0) выполнено, если

«идентификация условной функции плотности»

(conditional density identification):

$$f(y_i | x_i; \theta) \neq f(y_i | x_i; \theta_0) \quad \text{для всех } \theta \neq \theta_0. \quad (7.2.13)$$

Мы можем превратить теоремы о состоятельности М-оценок (утверждения 7.3 и 7.4) в теоремы о состоятельности оценок условного МЛ с $w_i = \{y_i, x_i\}'$ и $m(w_i; \theta) = \ln f(y_i | x_i; \theta)$. Заменив условия идентификации для М-оценок на условия идентификации условной функции плотности, мы получим два следующих самодостаточных утверждения:

Утверждение 7.5 (Состоятельность оценки условного МЛ с компактным пространством параметров): Пусть $\{y_i, x_i\}$ — эргодический стационарный случайный процесс с условной функцией плотности $f(y_i | x_i; \theta_0)$ и пусть $\hat{\theta}$ — оценка условного МЛ, которая максимизирует среднее значение логарифма условной функции правдоподобия (выведенной в предположении, что $\{y_i, x_i\}$ i.i.d.):

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \ln f(y_i | x_i; \theta).$$

¹ Если $f(y_i | x_i; \theta) = 0$, то тогда нельзя определить логарифм этой функции. Если это происходит, важно указать, что $f(y_i | x_i; \theta)$ положительна для всех y_i, x_i и θ . См. [Wilde, 1964, p. 9] относительно доказательства, которое избегает такой упрямой процедуры.

Предположим, что модель правильно специфицирована, так что θ_0 принадлежит Θ . Предположим, что (i) пространство параметров Θ — компактное подмножество $\mathbb{R}^p$, (ii) $f(y_i|x_i;\theta)$ непрерывна по θ для всех (y_i, x_i) и (iii) $f(y_i|x_i;\theta)$ измерима по (y_i, x_i) для всех $\theta \in \Theta$ (так что по лемме 7.1 $\hat{\theta}$ определена как случайная величина). Предположим также, что

(a) (Идентификация.) $\text{Prob}\{f(y_i|x_i;\theta) \neq f(y_i|x_i;\theta_0)\} > 0$ для всех $\theta \neq \theta_0$ в Θ .

(b) (Доминирование.) $E[\sup_{\theta \in \Theta} \ln f(y_i|x_i;\theta)] < \infty$ (заметьте: математическое ожидание берется по y_i и по x_i).

Тогда $\hat{\theta} \rightarrow_p \theta_0$.

Утверждение 7.6 (состоятельность оценки условного ML, без компактности): Пусть $\{y_i, x_i\}$ эргодический стационарный с условной функцией плотности $f(y_i|x_i;\theta_0)$ и пусть $\hat{\theta}$ — оценка условного ML, которая максимизирует среднее значение логарифма условной функции правдоподобия (помноженной в произвольных, что $\{y_i, x_i\}$ i.i.d.):

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \ln f(y_i|x_i;\theta).$$

Предположим, что модель правильно специфицирована, так что θ_0 принадлежит Θ . Предположим, что (i) вектор истинных параметров θ_0 является внутренним элементом выпуклого множества параметров Θ ($\subseteq \mathbb{R}^p$), (ii) $\ln f(y_i|x_i;\theta)$ вогнута по θ для всех (y_i, x_i) и (iii) $\ln f(y_i|x_i;\theta)$ измерима по (y_i, x_i) для всех θ из Θ . (Для достаточно большого n $\hat{\theta}$ однозначно определена, см. несколько строк ниже.) Предположим, также, что

(a) (Идентификация.) $\text{Prob}\{f(y_i|x_i;\theta) \neq f(y_i|x_i;\theta_0)\} > 0$ для всех $\theta \neq \theta_0$ из Θ .

(b) $E[\ln f(y_i|x_i;\theta)] < \infty$ (то есть $E[\ln f(y_i|x_i;\theta)]$ существует и конечно) для всех $\theta \in \Theta$ (заметьте: математическое ожидание берется по y_i и по x_i).

Тогда, для $n \rightarrow \infty$, $\hat{\theta}$ существует с вероятностью, стремящейся к 1, и $\hat{\theta} \rightarrow_p \theta_0$.

Следует сделать две замечания.

- Информационные критерии Кюльбака — Лейблера (7.2.1^о) выполняются также для безусловных функций плотности (доказательство для безусловных плотностей тривиально). Таким образом, утверждения 7.5 и 7.6 верны и для безусловного ML: надо просто заменить $f(y|x; \theta)$ на $f(x; \theta)$.
- Важной особенностью этих двух теорем о состоятельности оценки ML является то, что, несмотря на то что логарифм функции правдоподобия выводится при предположении о независимости и одинаковом распределении $\{y_i, x_i\}$, оценки будут состоятельными даже в случае, когда $\{y_i, x_i\}$ не i.i.d., а просто является стационарным и эргодическим процессом. Это так, потому что оценка ML является M-оценкой. Оценка ML, которая максимизирует функцию правдоподобия, отличную от функции правдоподобия модели, называется **оценкой квази-ML (quasi-ML estimator)**, или **оценкой QML (QML estimator)**. Другими словами, оценка QML максимизирует функцию правдоподобия в неверно специфицированной модели. Таким образом, оценки метода «максимального правдоподобия» в утверждениях 7.5 и 7.6 будут оценками QML, если $\{y_i, x_i\}$ стационарный и эргодический, но не i.i.d. Эти теоремы говорят о том, что данные конкретные оценки QML состоятельны.

В частных случаях условие идентификации для условного ML может быть записано в еще более простом виде.

Пример 7.5 (идентификация в модели линейной регрессии): В случае линейной регрессии с нормальными ошибками логарифм условной функции правдоподобия для наблюдений i равен:

$$\ln f(y_i, x_i; \theta) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (y_i - x_i'\beta)^2.$$

Так как логарифм — строго монотонная функция, идентификация условной плотности эквивалентна условию о том, что

$$\ln f(y_i, x_i; \theta) \neq \ln f(y_i, x_i; \theta_0)$$

с положительной вероятностью, если $\theta \neq \theta_0$. Это условие выполняется, если $E(x_i x_i')$ невырождена (т. е. следовательно, положительно определена). Чтобы понять, почему это так, обозначим $\theta = (\beta' \sigma^2)'$ и $\theta_0 = (\beta_0' \sigma_0^2)'$. Ясно, что $\ln f(y_i, x_i; \theta) \neq \ln f(y_i, x_i; \theta_0)$ с положительной вероятностью, если $\sigma^2 \neq \sigma_0^2$. Так что рассмотрим случай, когда $\sigma^2 = \sigma_0^2$, но $\beta \neq \beta_0$. Тогда

$$E[(x_i'\beta - x_i'\beta_0)^2] = E[(x_i'\beta - \beta_0)^2] = (\beta - \beta_0)' E(x_i x_i') (\beta - \beta_0) > 0$$

Таким образом, $x_i'\beta / x_i'\beta_0$ с положительной вероятностью, если $x_i'\beta = x_i'\beta_0$ с вероятностью 1 (то есть если $x_i'\beta / x_i'\beta_0$ с нулевой вероятностью), то тогда $E\{x_i'\beta - x_i'\beta_0\}^2$ будет равняться нулю. Значит, $y_i = x_i'\beta / y_i = x_i'\beta_0$ с положительной вероятностью, из чего следует, что условие идентификации выполнено. Из того же условия относительно $E(x_i x_i')$ следует также и условие (b) в утверждении 7.6, потому что

$$\begin{aligned} E(y_i - x_i'\beta)^2 &= E[(x_i'(\beta_0 - \beta))]^2 \text{ (поскольку } y_i = x_i'\beta_0 \text{ i.s.t.)} \\ &= F(x_i^2) + (\beta_0 - \beta)' F(x_i x_i') (\beta_0 - \beta), \\ &\text{так как } F(x_i' x_i) = 0. \end{aligned}$$

Последнее слагаемое будет конечным, если конечно $E(x_i x_i')$. Из этих результатов и отмеченного ранее факта, что логарифм функции правдоподобия возрастает после репараметризации, следует, что оценки условного ML-модели линейной регрессии с нормальными ошибками (выведенная из предположения, что $\{y_i, x_i\}$ i.i.d.) состоятельны, если $\{y_i, x_i\}$ эргодический и стационарный и если $E(x_i x_i')$ невырождена.

Пример 7.9 (идентификация в моделях пробит): Тот же результат, описанный выше, выведен в следующем примере для линейной модели, также работающей в случае кривых пробит, для которой условная функция правдоподобия для наблюдений i равна

$$f(y_i, x_i; \theta) = \Phi(x_i'\theta)^\gamma \Phi(-x_i'\theta)^{1-\gamma},$$

Та же самая логарифмизация, которую использовалась в примере 7.8, приводит к тому, что условная логарифмичность $F(x_i x_i')$, $x_i'\theta \neq x_i'\theta_0$ с положительной вероятностью, если $\theta \neq \theta_0$. Но так как $\Phi(\cdot)$ строго монотонна, из $x_i'\theta \neq x_i'\theta_0$ с положительной вероятностью следует, что $\Phi(x_i'\theta) \neq \Phi(x_i'\theta_0)$ и $\Phi(x_i'\theta) / \Phi(x_i'\theta_0)$ с положительной вероятностью. Значит, условие идентификации (a) в утверждении 7.6 выполняется, если $E(x_i x_i')$ невырождена.

Из невырожденности $E(x_i x_i')$ следует также и то, что условие (b) утверждения 7.6 выполняется для модели пробит. Очень просто показать, что

$$E\{\Phi(x_i'\theta)\} \leq E\{\Phi(x_i'\theta_0)\} + |x_i'(\theta - \theta_0)| \text{ для всех } \theta. \quad (7.2.14)$$

Комбинируя это ограничение на $E\{\Phi(x_i'\theta)\}$ с тем фактом, что Φ и $1 - \Phi$ по абсолютной величине меньше или равны единице, легко показать, что $E\{|x_i'(\theta - \theta_0)|\} < \infty$, если $E(x_i x_i')$ существует и конечно (условие вы-

получается из за невырожденности $E(x_i x_i')$ ¹. Мы, таким образом, делаем вывод, что ошибка модели пробит ML (логарифм функции правдоподобия для которой выводился в предположении i.i.d. $\{y_i, x_i\}$) состоятельна, если (y_i, x_i) стационарный и эргодический, и $E(x_i x_i')$ невырождена.

Состоятельность GMM

Перейдем теперь к GMM и адаптируем утверждение 7.1 к GMM, проверяя выполнение условий этого утверждения при выполнении набора соответствующих достаточных условий. Целевая функция GMM задается (7.1.13). Ясно, что непрерывность $Q_n(\theta)$ по θ будет обеспечена, если $g(w_i; \theta)$ непрерывна по θ для всех w_i . Если $\{w_i\}$ стационарный и эргодический, то $g_n(\theta) \equiv \sum_{i=1}^n g(w_i; \theta) \rightarrow_p E[g(w_i; \theta)]$. Следовательно, предельная функция $Q_0(\theta)$ имеет вид:

$$Q_0(\theta) = -\frac{1}{2} E[g(w_i; \theta)]' W E[g(w_i; \theta)]. \quad (7.2.15)$$

Эта функция положительно-определена, если $W (\equiv \text{plim } \tilde{W})$ положительно определена. Функция достигает максимального значения, равного нулю, в точке θ_0 , потому что $E[g(w_i; \theta)] = 0$ по условиям ортогональности. Таким образом, условие идентификации (что $Q_0(\theta)$ принимает максимальное значение исключительно в θ_0) утверждения 7.1 будет выполняться, если $E[g(w_i; \theta)] \neq 0$ для $\theta \neq \theta_0$. Учитывая равномерную сходимость $Q_n(\theta)$ к $Q_0(\theta)$, несложно показать, что это условие будет выполнено, если $g_n(\cdot)$ сходится равномерно к $E[g(w_i; \cdot)]$ (см.: [Newey and McFadden, 1994, p. 2132]. Многомерная версия равномерного закона больших чисел, приведенная прямо под леммой 7.2, в свою очередь, дает достаточное условие равномерной сходимости. Так что мы доказали

Утверждение 7.7 (состоятельность GMM с компактным пространством параметров): Пусть $\{w_i\}$ — стационарный и эргодический и пусть θ

¹Только для заинтересованных читателей здесь прикратно доказательство:

$$\begin{aligned} \ln f(w; \theta) &\leq \ln \Phi(z; \theta) = (1 - \rho) \ln \Phi(-x; \theta) \leq \\ &\leq \ln \Phi(x; \theta) + |\ln \Phi(-x; \theta)| \text{ (так как } \rho \leq 1 \text{ и } 1 - \rho \leq 1) \leq \\ &\leq 2 \times |\ln \Phi(0) - \ln \Phi(-x; \theta)| \text{ (то (7.2.14))} \leq \\ &\leq 2 \times |\ln \Phi(0) + |\ln \cdot| \cdot |\theta| + |x|^2 \cdot |\theta|^2|. \end{aligned}$$

Последнее неравенство следует из неравенства Коши — Буняковского: $|\ln \Phi(x; \theta)| \leq |\ln \cdot| \cdot |\theta|$ для двух любых векторов x и y произвольной размерности. Кроме того, $|\ln \Phi(x; \theta)|$ доказывается суммой $K|x_j^2|$ по j , конечно, по тому что из невырожденности $K|x_j^2|$ следует что $K|x_j^2| \leq \infty$ для всех j . Кроме $K|x_j^2| \leq \infty$, $\ln \Phi(x; \theta) \leq \infty$, $\ln \Phi(x; \theta) \leq \infty$.

СММ-оценка, определенная как

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \left[\frac{1}{n} \sum_{i=1}^n g(w_i; \theta) \right]' \tilde{W} \left[\frac{1}{n} \sum_{i=1}^n g(w_i; \theta) \right],$$

где симметричная матрица $\tilde{W}$ по предположению сходится по вероятности к некоторой симметричной положительно определенной матрице W . Предположим, что модель специфицирована правильно, в том смысле, что $E[g(w_i; \theta_0)] = 0$ (то есть выполняются условия ортогональности) для $\theta_0 \in \Theta$. Предположим, что (i) пространство параметров Θ — компактное подмножество $\mathbb{R}^k$, (ii) $g(w_i; \theta)$ непрерывна по θ для всех w_i и (iii) $g(w_i; \theta)$ измерима по w_i для всех θ из Θ (так что $\hat{\theta}$ — корректно определенная случайная величина по лемме 7.1). Предположим также, что

(a) (Идентификация.) $E[g(w_i; \theta)] \neq 0$ для всех $\theta \neq \theta_0$ из Θ .

(b) (Доминирование.) $E[\sup_{\theta \in \Theta} |g(w_i; \theta)|] < \infty$.

Тогда $\hat{\theta} \rightarrow_p \theta_0$.

Если $\mathcal{Q}_n(\theta)$ вогнута, то требования компактности Θ , непрерывности $g(w_i; \theta)$ и условие доминирования могут быть заменены на условие, что $E[g(w_i; \theta)]$ существует и конечно для всех θ . Мы не будем формулировать этот результат в качестве теоремы, потому что существует очень мало прикладных задач, в которых $\mathcal{Q}_n(\theta)$ вогнута. Например, целевые функции СММ для уравнения Эйлера в примере 7.5 не являются вогнутой. Пусть ли не единственным известным случаем является тот, в котором $g(w_i; \theta)$ линейна по θ , по этому важному случаю был довольно подробно рассмотрен в главе 3. Когда $g(w_i; \theta)$ нелинейна по θ , специфицировать простейшие условия для условий (a) (идентификация) и (b) (доминирование) в утверждении 7.7 в общем случае довольно сложно. Так что в большинстве прикладных задач выполнение этих условий просто предполагается.

Контрольные вопросы

1. (Необходимость идентификации функции плотности.) Мы показали, для условного МЛ, что идентификация условной функции плотности (7.2.13) достаточно для выполнения условия (a) (т.е. $E[h(x_1, x_2; \theta)]$ принимает наибольшее значение исключительно в θ_0). То, что это является и необходимым условием, ясно из информационного неравенства Куляшова — Лейблера. Не используя это неравенство, покажите необходимость идентификации условной плотности. Указание: Покажите, что если бы (7.1.13) не выполнялось, то (a) в этом случае не выполнялось бы. Предположите, что существует $\theta_1 \neq \theta_2$, такое, что $\theta_1 \in \Theta$ и $f(x_1, x_2; \theta_1) = f(x_1, x_2; \theta_2)$. Тогда $E[h(x_1, x_2; \theta_1)] = E[h(x_1, x_2; \theta_2)]$.

2. (Необходимость идентификации условного математического ожидания) Не используя (7.2.8), покажите, что условие идентификации условного математического ожидания является как необходимым, так и достаточным для идентификации в NLS.
3. (Идентификация в NLS.) Предположим, что функция φ в NLS линейна: $\varphi(\mathbf{x}_1; \theta) = \mathbf{x}_1' \theta$. Проверьте, что условие идентификации выполняется, если $E(\mathbf{x}_1 \mathbf{x}_1')$ невырождена.
4. (Идентификация в NLS-регрессии коффи гробят.) В модели регрессии $E(y_1; \mathbf{x}_1) = \Phi(\mathbf{x}_1' \theta_0)$. Пусть θ_0 оценивается NLS. Проверьте, что условие идентификации условного математического ожидания выполняется, если $E(y_1 \mathbf{x}_1)$ невырождена.
5. (Квази ML.) Предположим, что в утверждении 7.5 $\{y_1, \mathbf{x}_1\}$ i.i.d. так что функции правдоподобия можно себе специфицировать правильно. Однако предположим, что истинное значение вектора параметра θ_0 не включено в предполагаемый пространство параметров Θ . Так что в этом смысле оценка $\hat{\theta}$ будет пылаться оценкой квази-ML. Оутекин условие (а), предположите вместо этого, что существует единственное значение θ^* из Θ , которое максимизирует $E[f(y_1; \mathbf{x}_1; \theta)]$. Покажите, что оценка квази-ML $\hat{\theta}$ будет состоятельной оценкой для θ^* . Указание: Замените θ_0 в утверждении 7.3 на θ^* . В утверждениях 7.1–7.4 θ_0 не обязан быть вектором «истинных» значений параметров.
6. (Идентификация в линейной GMM.) Рассмотрим модель линейного GMM из главы 3. Проверьте, что ранговое условие идентификации, выведенное в главе 3 эквивалентно условию идентификации в утверждении 7.7.
7. (Вогнутость целевой функции в случае линейного GMM.) Проверьте, что целевая функция GMM в линейном GMM вогнута θ . Указание: Гессиан (матрица вторых производных) функции $Q_n(\theta)$ равен $S_{xx}' W S_{xx}$, где $S_{xx} \equiv \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i'$. Двежды непрерывно дифференцируемая функция вогнута, если и только если ее гессиан отрицательно полуопределен во всех точках.
8. (Идентификация GMM при вырожденной W .) Если бы мы требовали, что W должна быть хотя бы положительно полуопределена, а не обязательно строго положительно определена, то каким бы было условие идентификации в утверждении 7.7? Указание: Предположите, что W положительно полуопределена. Покажите, что если $E[g(\mathbf{y}_1; \mathbf{x}_1; \theta_0)] = 0$ и $WE[g(\mathbf{y}_1; \mathbf{x}_1; \theta)] \neq 0$ для $\theta \neq \theta_0$, то $Q_n(\theta) = -\frac{1}{2} E[g(\mathbf{y}_1; \mathbf{x}_1; \theta)]' W E[g(\mathbf{y}_1; \mathbf{x}_1; \theta)]$ имеет единственную точку максимума в θ_0 .

7.3. Асимптотическая нормальность

Базовым инструментом для вывода асимптотического распределения экстремальной оценки является теорема о среднем значении из математического анализа. Как она используется, зависит от того, является ли рассматриваемая оценка М-оценкой или оценкой GMM. Таким образом

доказательство асимптотической нормальности будет приведено отдельно для M -оценок (включая оценку ML как частного случая M -оценок) и для оценок СММ. За этим последуют краткие замечания об относительной эффективности СММ и ML. Подводя итоги, мы докажем в конце параграфа, что разложение в ряд Тейлора для отклонения выборочной оценки от истинного параметра имеет общую структуру для M -оценок и СММ.

Асимптотическая нормальность M -оценок

Приведенная в (7.1.2) целевая функция для M -оценки имеет вид:

$$Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n m(w_i; \theta). \quad (7.3.1)$$

Удобно обозначить градиент (вектор первых производных) и гессиан (матрицу вторых производных) функции m как

$$J(w_i; \theta) = \frac{\partial m(w_i; \theta)}{\partial \theta}, \quad (7.3.2)$$

$$H(w_i; \theta) = \frac{\partial^2 m(w_i; \theta)}{\partial \theta^2} = \frac{\partial^2 m(w_i; \theta)}{\partial \theta \partial \theta'}, \quad (7.3.3)$$

По аналогии с ML $J(w_i; \theta)$ будет называться **вектором вклада для наблюдения i** (*score vector for observation i*). Этот $J(w_i; \theta)$ не следует путать с вкладом в привычном понимании, который является градиентом целевой функции $Q_n(\theta)$. Функцию вклада во втором случае мы будем обозначать далее в этой главе как $J_n(\theta)$. То же самое относится к гессиану: $H(w_i; \theta)$ будет рассматриваться как **гессиан для наблюдения i** (*Hessian for observation i*), а гессиан $Q_n(\theta)$ будет обозначаться $H_n(\theta)$.

Целью этого параграфа является асимптотическая нормальность $\hat{\theta}$, описанная ниже в (7.3.10). В процессе вывода этого свойства мы сделаем несколько предположений, которые будут собраны ниже в утверждении 7.8. Предположим, что $m(w_i; \theta)$ дифференцируема по θ и что $\bar{\theta}$ — внутренняя точка Θ . Тогда $\bar{\theta}$, будучи внутренним решением задачи максимизации $Q_n(\theta)$, удовлетворяет условиям первого порядка

$$\begin{aligned} \mathbf{0}_{(p \times 1)} &= \frac{\partial Q_n(\bar{\theta})}{\partial \theta} = \\ &= \frac{1}{n} \sum_{i=1}^n J(w_i; \bar{\theta}) \quad (\text{по (7.3.1) и (7.3.2)}) \end{aligned} \quad (7.3.4)$$

Сейчас мы будем использовать следующий результат из математического анализа:

Теорема о среднем значении: Пусть $h: \mathbb{R}^n \rightarrow \mathbb{R}^m$ непрерывно дифференцируемо. Тогда $h(x)$ можно представить в виде **разложения в среднем значении** (mean value expansion),

$$h(x) = h(x_0) + \frac{\partial h(x_0)}{\partial x} (x - x_0),$$

где α — среднее значение, лежащее между x и x_0 ¹.

Положив $y = x$, $x = \tilde{\theta}$, $x_0 = \theta_0$ и $h(\cdot) = \frac{\partial Q(\cdot)}{\partial \theta}$ в теореме о среднем значении, мы получим следующее разложение,

$$\begin{aligned} \frac{\partial Q(\tilde{\theta})}{\partial \theta} &= \frac{\partial Q(\cdot)}{\partial \theta}(\theta_0) + \frac{\partial^2 Q(\cdot)}{\partial \theta \partial \theta}(\bar{\theta}) (\tilde{\theta} - \theta_0) = \\ &= \frac{1}{n} \sum_{i=1}^n s(w_i; \theta_0) + \left[\frac{1}{n} \sum_{i=1}^n H(w_i; \bar{\theta}) \right] (\tilde{\theta} - \theta_0) \end{aligned} \quad (7.3.5)$$

(по (7.3.1)–(7.3.3)).

где $\bar{\theta}$ — среднее значение, лежащее между $\tilde{\theta}$ и θ_0 . Требования нормальности о среднем значении относительно непрерывной дифференцируемости будет выполняться, если $s(w_i; \theta)$ всегда непрерывно дифференцируемо по θ . Объединив это уравнение с вышеуказанным условием первого порядка, мы получим:

$$0 = \frac{1}{n} \sum_{i=1}^n s(w_i; \theta_0) + \left[\frac{1}{n} \sum_{i=1}^n H(w_i; \bar{\theta}) \right] (\tilde{\theta} - \theta_0). \quad (7.3.6)$$

Предполагая, что $\frac{1}{n} \sum_{i=1}^n H(w_i; \bar{\theta})$ невырождена, это уравнение может быть решено относительно $(\tilde{\theta} - \theta_0)$, что даст,

$$\sqrt{n}(\tilde{\theta} - \theta_0) = \left[\frac{1}{n} \sum_{i=1}^n H(w_i; \bar{\theta}) \right]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n s(w_i; \theta_0). \quad (7.3.7)$$

Это выражение для (умноженный на $\sqrt{n}$) ошибки оценки будет называться **разложением в среднем значении ошибки оценки** (mean value expansion for the scoring error). Заметим, что вектор вклада $s(w_i; \theta)$ рассчитывается при истинном значении θ_0 .

¹Теорема о среднем значении применяется только к скалярным значениям h , так что на самом деле x заменяется ст. значения и элементу векторного уравнения. Это упрощение не влияет на приведенный в тексте анализ.

Теперь, поскольку $\bar{\theta}$ лежит между θ_0 и $\hat{\theta}$, $\bar{\theta}$ — состоятельная оценка вектора θ_0 , если таковой является $\hat{\theta}$. Если $\{w_t\}$ — эргодический стационарный процесс, естественно будет предположить, что

$$\frac{1}{n} \sum_{t=1}^n H(w_t; \theta) \xrightarrow{p} E[H(w_t; \theta_0)]. \quad (7.3.8)$$

Стационарности и эргодичности w_t и состоятельности $\bar{\theta}$ самих по себе недостаточно, чтобы обеспечить это; необходимы предположить выполнение некоторых технических условий. Одним из таких технических условий является равномерная сходимость $\frac{1}{n} H(w_t; \cdot)$ к $E[H(w_t; \cdot)]$ в окрестности θ_0 ¹. Но, по теореме о равномерной сходимости, равномерная сходимость $\frac{1}{n} H(w_t; \cdot)$ будет выполняться, если для гессиана выполняется следующее условие доминирования: для некоторой окрестности N точки θ_0 $E[\sup_{\theta \in N} |H(w_t; \theta)|] < \infty$ (это условие (4) ниже)². Это достаточное условие для (7.3.8): если возможно напрямую проверить (7.3.8)), то тогда нет необходимости проверять условие доминирования для асимптотической нормальности.

Наконец, если выполняется (7.3.8) и если

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n s(w_t; \theta_0) \xrightarrow{d} N(0, \Sigma), \quad (7.3.9)$$

то, по теореме Слайского (см. лемму 2.4(c)), мы имеем:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, (E[H(w_t; \theta_0)])^{-1} \Sigma (E[H(w_t; \theta_0)])^{-1}\right). \quad (7.3.10)$$

Собирая предположения, которые мы сделали ранее, мы получаем:

Утверждение 7.8 (асимптотическая нормальность М-оценок):

Предположим, что выполняются предположения утверждения 7.3 или утверждения 7.4, так что $\{w_t\}$ — эргодический стационарный процесс

¹Это условие не приводилось ниже результата (см., например, Theorem 4.1.5 в [Аллен-Сул, 1983]).

Предположим, что последовательность суммарных функций $h_n(\theta)$ сходится равномерно по θ к некоторой функции $h_0(\theta)$ в открытой окрестности θ_0 . Тогда $\lim_{n \rightarrow \infty} h_n(\hat{\theta}) = h_0(\theta_0)$, если $\lim \hat{\theta} = \theta_0$, и $h_0(\theta)$ непрерывна в θ_0 .

Пусть, кроме того, выполняется $\frac{1}{n} H(w_t; \cdot)$, а роль $h_n(\theta) = E[H(w_t; \theta)]$. Непрерывность $E[H(w_t; \theta)]$ обеспечивается теоремой о равномерной сходимости. Равномерная сходимость необходима только в окрестности θ_0 , а не во всем пространстве параметров, потому что $\hat{\theta}$ близко к θ_0 для достаточно больших n .

²Матрицу можно рассматривать как вектор, элементы которого были переставлены. Так что Евклидова норма матрицы, такая как $\|H(w_t; \theta)\|$, есть квадратный корень из суммы квадратов всех ее элементов.

я M -оценка $\hat{\theta}$, определенная в (7.1.1) и (7.1.2), состоятельна. Предположим также, что

- (1) θ_0 — внутренний элемент Θ ,
- (2) $\eta(\omega; \theta)$ дважды непрерывно дифференцируема по θ для любого ω ,
- (3) $\frac{1}{n} \sum_{i=1}^n \eta(\omega_i; \theta_0) \rightarrow_p N(0, \Sigma)$, Σ положительно определенная матрица, где $\eta(\omega; \theta)$ определена в (7.3.2),
- (4) (условие локального доминирования по Гессе) для некоторой окрестности N' точки θ_0

$$E \left[\sup_{\theta \in N'} |H(\omega; \theta)| \right] < \infty,$$

так что для любой состоятельной оценки $\hat{\theta} \xrightarrow{p} \theta_0$, $\frac{1}{n} \sum_{i=1}^n H(\omega_i; \hat{\theta}) \rightarrow_p E[H(\omega; \theta_0)]$, где $H(\omega; \theta)$ определена в (7.3.3).

- (5) $E[H(\omega; \theta_0)]$ невырождена.

Тогда $\hat{\theta}$ асимптотически нормальна:

$$\text{Asym}(\hat{\theta}) = (E[H(\omega; \theta_0)])^{-1} \Sigma (E[H(\omega; \theta_0)])^{-1}.$$

(Это Теорема 4.1.3 в [Амиссива, 1995], примененная к M -оценкам.) Необходимо сделать два замечания:

- На предположений, которые мы сделали в процессе вывода асимптотической нормальности, следующие не перечислены в утверждении: (i) θ_0 — внутренняя точка и (ii) $\frac{1}{n} \sum_{i=1}^n H(\omega_i; \hat{\theta})$ невырождена. Интуитивно понятно, что эти условия выполняются, потому что $\hat{\theta}$ сходится по вероятности к внутренней точке θ_0 , а $\frac{1}{n} \sum_{i=1}^n H(\omega_i; \hat{\theta})$ сходится по вероятности к невырожденной матрице $E[H(\omega; \theta_0)]$. См. [Newey and McFadden, 1994, p. 2152] для строгих доказательств. Технические моменты такого рода будут кифирированы в оставшейся части главы.
- Если ω стационарный и эргодический, то таким же образом вектор выклада $\eta(\omega; \theta_0)$ и матрица Σ — дисперсионная ковариационная матрица $\{\eta(\omega; \theta_0)\}^T$. Достаточным условием для (3) является условие Тюринга, приведенное в параграфе 6.5. Так что условие (3) в утверждении может быть заменено условием Тюринга на $\{\eta(\omega; \theta_0)\}$. Оно будет выполняться, если, например, ω i.i.d. и $E[\eta(\omega; \theta_0)] = 0$. Предположение о том, что Σ положительно

Достаточно — ковариационная матрица была введена в параграфе 7.4

определены, на самом деле не нужно для вывода утверждения, но мы можем также его сделать, потому что практически во всех практических задачах оно выполняется (или предполагается) и также потому, что оно потребуется для проверки гипотез далее в этой главе.

Состоятельная оценка асимптотической ковариационной матрицы

Чтобы использовать асимптотические результаты для проверки гипотез, нам необходима состоятельная оценка для

$$\text{Avar}(\tilde{\theta}) = (E[H(w_i; \theta_0)])^{-1} \Sigma(E[H(w_i; \theta_0)])^{-1}.$$

Так как $\tilde{\theta} \rightarrow_p \theta_0$, из условия (4) утверждения 7.8 следует, что

$$\frac{1}{n} \sum_{i=1}^n H(w_i; \tilde{\theta}) \xrightarrow_p E[H(w_i; \theta_0)].$$

Таким образом, три условия того, что нам доступна состоятельная оценка $\tilde{\Sigma}$ матрицы Σ ,

$$\text{Avar}(\tilde{\theta}) = \left\{ \frac{1}{n} \sum_{i=1}^n H(w_i; \tilde{\theta}) \right\}^{-1} \tilde{\Sigma} \left\{ \frac{1}{n} \sum_{i=1}^n H(w_i; \tilde{\theta}) \right\}^{-1} \quad (7.3.11)$$

является состоятельной оценкой асимптотической ковариационной матрицы. Чтобы получить $\tilde{\Sigma}$, можно применить предложенные в параграфе 6.6 методы, такие, как VARLAC, к числовым рядам $\{e(w_i; \tilde{\theta})\}$ при условии выполнения некоторых лежащих в их технических условий.

Асимптотическая нормальность условного ML

Мы применим сейчас этот результат об асимптотической нормальности к ML для случая i.i.d. данных. В ML, где функцией m является логарифм (условной) функции правдоподобия, условия первого порядка (7.3.4) называются **уравнениями правдоподобия** (*likelihood equations*). То, что функцией m является логарифм (условной) функции правдоподобия, приводит к упрощению асимптотической ковариационной матрицы $\tilde{\theta}$. Мы докажем это для условного ML, сделав то же самое для безусловного и совместного ML, позже.

Рассмотрим также два равенства в условии (3) утверждения 7.8. Второе равенство называется **равенством для информационных матриц** (*information matrix equality*) (не путать с информационным неравенством Кушмана — Лейблера). Оно называется так, потому что матрица $E[\delta(w_i; \theta_0; \delta(w_i; \theta_0)']$ — это **информационная матрица для наблюдения i** (*information matrix for observation i*). Вывод этих двух равенств

при выполнении указанных технических условий, рассмотренных выше, для переформулировки дифференцирования и интегрирования, остается в качестве аналитического упражнения 2. Здесь мы просто отметим применимость этих двух лемм к утверждению 7.8. Если $\omega_1 \in \Omega$, то из ЦПТ в формулировке Линдберга — Лева и первого равенства $E[s(\omega_1; \theta_0)] = 0$ следует, что

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n s(\omega_i; \theta_0) \rightarrow_d N(0, \Sigma), \quad \text{где } \Sigma = E[s(\omega_1; \theta_0)s(\omega_1; \theta_0)']. \quad (7.3.12)$$

Из этого и из равенства для дифференциальных матриц следует, что $\text{Asymp}(\tilde{\theta})$ в утверждении 7.8 упрощается двумя способами как

$$\text{Asymp}(\tilde{\theta}) = -\{E[H(\omega_1; \theta_0)]\}^{-1} \cdot \{E[s(\omega_1; \theta_0)s(\omega_1; \theta_0)']\}^{-1}. \quad (7.3.13)$$

Важно отметить, мы доказали

Утверждение 7.9 (асимптотическая нормальность условного ML): Пусть $\omega_1 \equiv (y_1, x_1')$ i.i.d. Предположим, что выполняются предпосылки утверждения 7.5 или утверждения 7.6, так что $\tilde{\theta} \rightarrow_p \theta_0$. Предположим также, что

- (1) θ_0 — внутренняя точка Θ ,
- (2) $f(y|x; \theta)$ дважды непрерывно дифференцируема по θ для всех (y, x) ,
- (3) $E[s(\omega_1; \theta_0)] = 0$ и $E[H(\omega_1; \theta_0)] = E[s(\omega_1; \theta_0)s(\omega_1; \theta_0)']$, где функции s и H определены в (7.3.2) и (7.3.3),
- (4) (условие локального нормирования на границе) для некоторой окрестности N точки θ_0

$$E[\sup_{\theta \in N} \|H(\omega_1; \theta)\|] < \infty,$$

так что для любой последовательной оценки $\tilde{\theta}$,

$$\frac{1}{n} \sum_{i=1}^n H(\omega_i; \tilde{\theta}) \rightarrow_p E[H(\omega_1; \theta_0)].$$

- (5) $E[H(\omega_1; \theta_0)]$ невырождена

тогда $\tilde{\theta}$ асимптотически нормальна с $\text{Asymp}(\tilde{\theta})$, заданной в (7.3.13).

Для замечания относительно этого утверждения:

- Условие (3) формулируется здесь как предположение, потому что его вывод требует перестановочности дифференцирования и интегрирования (см. аналитические упражнения 2). Техническое условие, при котором такая перестановочность возможна, доступен из математического анализа (см., например, Lemma 3.6 в [Newey and McFadden, 1994]). Так что условие (3) может быть заменено таким техническим условием на $f(y_i|x_i;\theta)$. Мы не делаем этого здесь, потому что в большинстве прикладных задач условие (3) может быть проверено ланямую.
- Из обсуждения, проведенного выше, должно быть понятно, как можно применить эти утверждения к безусловному ML: просто замените $f(y_i|x_i;\theta)$ на $f(w_i;\theta)$.

Чтобы использовать этот асимптотический результат для проверки гипотез, нам нужна состоятельная оценка $\text{Avar}(\hat{\theta})$. Учитывая, что она может быть записана как $-\{E[H(w_i;\theta_0)]\}^{-1}$, естественной оценкой является

$$\text{первая оценка } \text{Avar}(\hat{\theta}) = -\left\{\frac{1}{n} \sum_{i=1}^n H(w_i;\hat{\theta})\right\}^{-1}. \quad (7.3.14)$$

Поскольку $\hat{\theta} \rightarrow_p \theta_0$, эта оценка состоятельна по условию (4) утверждения 7.9. Другая оценка, основанная на соотношении

$$\text{Avar}(\hat{\theta}) = \{E[s(w_i;\theta_0)s(w_i;\theta_0)']\}^{-1}, \quad \text{это}$$

$$\text{вторая оценка } \text{Avar}(\hat{\theta}) = \left\{\frac{1}{n} \sum_{i=1}^n s(w_i;\hat{\theta})s(w_i;\hat{\theta})'\right\}^{-1}. \quad (7.3.15)$$

Чтобы обеспечить состоятельность этой оценки, нам необходимо сделать некоторое предположение в дополнение к гипотезе утверждения 7.9 (см., например, Theorem 4.4 в [Newey and McFadden, 1994]). Мы не будем приводить его здесь, потому что во многих прикладных задачах состоятельность этой второй оценки может быть проверена напрямую.

Не существует никаких неопровержимых доводов для предпочтения одной оценки $\text{Avar}(\hat{\theta})$ другой. С одной стороны, вторую оценку, которая требует только первых производных логарифма функции правдоподобия, вычислить проще, чем первую. Это может быть важным доводом, когда невозможно вычислить производные функции плотности аналитически и приходится использовать различные численные методы. С другой стороны, по крайней мере в некоторых случаях, первая оценка дает лучшую аппроксимацию асимптотического распределения (см., например: Davidson и McKinnon, 1993).

Два примера

Чтобы лучше понять результаты, касающиеся асимптотической нормальности условного МЛ, давайте рассмотрим, как эти результаты могут быть применены к известным примерам. Мы рассмотрим два из них.

Пример 7.10 (асимптотическая нормальность в модели линейной регрессии). Запишем логарифм условной функции плотности для наблюдения i для модели линейной регрессии с нормальными ошибками:

$$\ln f(y_i; x_i; \beta, \sigma^2) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma^2) - \frac{(y_i - x_i' \beta)^2}{2\sigma^2}. \quad (7.3.16)$$

Пространство $\Theta = \mathbb{R}^K \times \mathbb{R}_+$ (где K — размерность Θ), и условие (1) утверждения 7.9 выполняются. Условие (2) выполняется очевидным образом. Чтобы проверить условие (3), проведем рутинные расчеты и получим, что¹

$$\begin{aligned} a(w_i; \theta) &= \begin{bmatrix} \frac{1}{\sigma^2} x_i \cdot \tilde{\varepsilon}_i \\ -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^2} \tilde{\varepsilon}_i^2 \end{bmatrix}, \quad H(w_i; \theta) = \begin{bmatrix} -\frac{1}{\sigma^2} x_i x_i' & -\frac{1}{\sigma^2} x_i \cdot \tilde{\varepsilon}_i \\ -\frac{1}{\sigma^2} x_i' \cdot \tilde{\varepsilon}_i & \frac{1}{\sigma^2} - \frac{1}{\sigma^2} \tilde{\varepsilon}_i^2 \end{bmatrix}, \\ a(w_i; \theta) a(w_i; \theta)' &= \begin{bmatrix} \frac{1}{\sigma^2} x_i x_i' \tilde{\varepsilon}_i^2 & -\frac{1}{2\sigma^2} x_i \cdot \tilde{\varepsilon}_i + \frac{1}{2\sigma^2} x_i \cdot \tilde{\varepsilon}_i^3 \\ \frac{1}{2\sigma^2} x_i' \cdot \tilde{\varepsilon}_i + \frac{1}{2\sigma^2} x_i' \cdot \tilde{\varepsilon}_i^3 & \frac{1}{\sigma^2} - \frac{1}{\sigma^2} \tilde{\varepsilon}_i^2 + \frac{1}{\sigma^2} \tilde{\varepsilon}_i^4 \end{bmatrix}, \end{aligned} \quad (7.3.17)$$

где $w_i = (y_i, x_i')$, $\theta = (\beta', \sigma^2)'$, а $\tilde{\varepsilon}_i \equiv y_i - x_i' \beta$ (который следует отличать от $\varepsilon_i \equiv y_i - x_i' \beta_0$). Так что для $\theta = \theta_0$ $\tilde{\varepsilon}_i$ в данной записи может быть заменен на ε_i . В модели линейной регрессии $E(\varepsilon_i | x_i) = 0$. Также, так как $\varepsilon_i \sim N(0, \sigma^2)$, мы имеем $E(\varepsilon_i^2) = 0$, $E(\varepsilon_i^2) = 2\sigma^2$. Используя эти соотношения, легко проверить (3). В частности,

$$-E[H(w_i; \theta_0)] = E[a(w_i; \theta_0) a(w_i; \theta_0)'] = \begin{bmatrix} -\frac{1}{\sigma_0^2} E(x_i x_i') & 0 \\ 0 & \frac{1}{2\sigma_0^2} \end{bmatrix}. \quad (7.3.18)$$

Если $E(x_i x_i')$ невырождена, то $E[H(w_i; \theta_0)]$ невырождена и (5) выполняется. Для рассмотрения условия (4) пусть $\tilde{\varepsilon}_i \equiv y_i - x_i' \tilde{\beta}$ для некоторой произвольной оценки $\tilde{\beta}$ к β^0 . Условие (7.3.8) в этом примере выполняется как

$$\begin{aligned} & \begin{bmatrix} -\frac{1}{\sigma_0^2} \frac{1}{n} \sum_{i=1}^n x_i x_i' & -\frac{1}{\sigma_0^2 \sigma_0^2} \frac{1}{n} \sum_{i=1}^n x_i \cdot \tilde{\varepsilon}_i \\ -\frac{1}{\sigma_0^2 \sigma_0^2} \frac{1}{n} \sum_{i=1}^n x_i' \cdot \tilde{\varepsilon}_i & \frac{1}{2\sigma_0^2 \sigma_0^2} - \frac{1}{\sigma_0^2 \sigma_0^2} \frac{1}{n} \sum_{i=1}^n \tilde{\varepsilon}_i^2 \end{bmatrix} \xrightarrow{P} \\ & \xrightarrow{P} \begin{bmatrix} -\frac{1}{\sigma_0^2} E(x_i x_i') & 0 \\ 0 & \frac{1}{2\sigma_0^2} \end{bmatrix}. \end{aligned} \quad (7.3.19)$$

¹Для параметра σ^2 дифференцирование производится по логарифму параметру σ^2 , а не по σ .

Показать это просто, учитывая, что $\varepsilon_i = y_i - x_i'(\tilde{\beta} - \beta_0)$. Мы можем сделать вывод, что все условия утверждения 7.9 выполняются, если $E(x_i x_i')$ невырождена.

Пример 7.11 (асимптотическая нормальность ML-оценки модели пробит): Запишем логарифм условной функции правдоподобия для модели пробит.

$$\ln f(y_i|x_i; \theta) = y_i \ln \Phi(x_i'\theta) + (1 - y_i) \ln [1 - \Phi(x_i'\theta)]. \quad (7.3.20)$$

Так как $\Theta = \mathbb{R}^p$, условие (1) утверждения 7.9 выполняется. Условие (2) очевидным образом выполняется. Для проверки условия (3) проведем скучные вычисления, которые дадут нам следующее:

$$s(w_i; \theta) = \frac{(y_i - \Phi(x_i'\theta))x_i(x_i'\theta)}{[\Phi(x_i'\theta)(1 - \Phi(x_i'\theta))]^{1/2}}. \quad (7.3.21)$$

$$H(w_i; \theta) = \left\{ - \frac{y_i - \Phi(x_i'\theta)}{[\Phi(x_i'\theta)(1 - \Phi(x_i'\theta))]^{3/2}} [\phi(x_i'\theta)]^2 + \left[\frac{y_i}{\Phi(x_i'\theta)} \frac{\Phi(x_i'\theta)}{[1 - \Phi(x_i'\theta)]} \right] \phi'(x_i'\theta) \right\} x_i x_i'. \quad (7.3.22)$$

(Чтобы выписать (7.3.22), используйте тот факт, что $y_i^2 = y_i$.) Здесь $\Phi(\cdot)$ — это (кумулятивная) функция распределения вероятности для $N(0, 1)$, а $\phi(\cdot) \equiv \Phi'(\cdot)$ — ее функция плотности. Тогда легко доказать условный вариант (3):

$$\begin{aligned} E[s(w_i; \theta_0)|x_i] &= 0 \quad \text{и} \\ E[H(w_i; \theta_0) | x_i] &= E[s(w_i; \theta_0) s(w_i; \theta_0)' | x_i]. \end{aligned} \quad (7.3.23)$$

Так что (3) выполняется по закону полных математических ожиданий. В частности, для модели пробит просто показать, что

$$E[H(w_i; \theta_0)] = E[s(w_i; \theta_0) s(w_i; \theta_0)'] = E[\lambda(x_i'\theta_0) \lambda(-x_i'\theta_0) x_i x_i'], \quad (7.3.24)$$

где

$$\lambda(x) \equiv \frac{\phi(x)}{1 - \Phi(x)} \quad (7.3.25)$$

называют **обратным отношением Миллса** (*inverse Mill's ratio*), или **мерой риска** (*hazard*) для $N(0, 1)$. Можно показать, что элементы скобок в (7.3.22) лежат между 0 и 2 (см. [Newey and McFadden, 1994, p. 2147]). Поэтому

$$\|H(w_i; \theta)\| \leq 2 \|x_i x_i'\|. \quad (7.3.26)$$

Евклидова норма $\|\mathbf{z}_1 \mathbf{z}_2'\|$ есть квадратный корень из суммы квадратов элементов $\mathbf{z}_1 \mathbf{z}_2'$. Поэтому математическое ожидание $\|\mathbf{z}_1 \mathbf{z}_2'\|^2$ и, следовательно, математическое ожидание $\|\mathbf{z}_1 \mathbf{z}_2'\|$ конечны, если $E(\mathbf{z}_1 \mathbf{z}_1')$ (соответствует $\mathbf{z}$) конечно. Тогда условие локального доминирования на гессиан (условие (4)) выполняется, если $E(\mathbf{z}_1 \mathbf{z}_1')$ невырождена. Можно показать, что условие (5) выполняется, если $E(\mathbf{z}_1 \mathbf{z}_1')$ невырождена (см. справку на р. 2:47 в [Newey and McFadden, 1994]). Мы снова можем сделать вывод о том, что условие утверждения 7.9 выполняется, если $E(\mathbf{z}_1 \mathbf{z}_1')$ невырождена.

Асимптотическая нормальность GMM

Рассмотрим теперь GMM. Целевая функция GMM выглядит следующим образом:

$$Q_n(\theta) = -\frac{1}{2} g_n(\theta)' \tilde{W} g_n(\theta), \quad \text{где } g_n(\theta) = \frac{1}{n} \sum_{i=1}^n g(\mathbf{w}_i; \theta). \quad (7.3.27)$$

Как и в случае M-оценок, мы применим теорему о среднем значении к условиям первого порядка задачи максимизации, но теорема будет применена к $g_n(\theta)$, а не к ее исходной аппроксимации. Таким образом, в отличие от случая M-оценок, целевая функция должна быть непрерывно дифференцируемой только один раз, а не дважды. Это отражает тот факт, что выборочное среднее входит в целевую функцию GMM по-другому.

В предположении, что $\hat{\theta}$ — внутренняя точка, условия первого порядка максимизации целевой функции таковы:

$$\mathbf{0}_{(p \times 1)} = \frac{\partial Q_n(\hat{\theta})}{\partial \theta} = -G_n(\hat{\theta})' \tilde{W} g_n(\hat{\theta}), \quad (7.3.28)$$

где $G_n(\theta)$ — матрица Якоби для $g_n(\theta)$:

$$G_n(\theta) = \frac{\partial g_n(\theta)}{\partial \theta'}. \quad (7.3.29)$$

Теперь применим теорему о среднем значении к $g_n(\theta)$, а не к $g(\mathbf{w}_i; \theta)$, как в случае M-оценок, чтобы получить разложение в среднем значении:

$$g_n(\hat{\theta}) = g_n(\theta_0) + G_n(\bar{\theta})(\hat{\theta} - \theta_0), \quad (7.3.30)$$

Подставляя это в запись уравнения первого порядка, мы получаем:

$$\mathbf{0}_{(p \times 1)} = -G_n(\hat{\theta})' \tilde{W} g_n(\hat{\theta}) = -G_n(\hat{\theta})' \tilde{W} G_n(\bar{\theta})(\hat{\theta} - \theta_0), \quad (7.3.31)$$

Решая это уравнение относительно $(\hat{\theta} - \theta_0)$ и учитывая, $= \frac{1}{n} \sum_{t=1}^n g(w_t, \theta)$, мы получаем:

$$\sqrt{n}(\hat{\theta} - \theta_0) = - \begin{bmatrix} G_n(\hat{\theta})' & \widehat{W} & G_n(\bar{\theta}) \end{bmatrix}^{-1} \times \\ \times G_n(\hat{\theta})' & \widehat{W} & \frac{1}{\sqrt{n}} \sum_{t=1}^n g(w_t; \theta_0). \\ (p \times K) & (K \times K) & (K \times p) \quad (K \times 1)$$

В этом выражении якобиан $G_n(\theta)$ функции $g_n(\theta)$ вычисляется в разных точках — $\hat{\theta}$ и $\bar{\theta}$. Так как $g_n(\theta) = \frac{1}{n} \sum_{t=1}^n g(w_t, \theta)$, якобиан любой оценки $\hat{\theta}$ может быть записан как

$$G_n(\hat{\theta}) = \frac{1}{n} \sum_{t=1}^n \frac{\partial g(w_t; \hat{\theta})}{\partial \theta'}. \quad (7.3.3)$$

Если w_t — стационарный и эргодический процесс и $\hat{\theta}$ состоятельна, естественно предположить, что это выражение сходится к $E[\partial g(w_t; \theta_0)/\partial \theta']$. Так же, как в случае М-оценок, это предположение будет верным, если $\partial g(w_t; \theta_0)/\partial \theta'$ удовлетворяет соответствующим условиям доминирования, сформулированным в условии (4) ниже. Теперь должно быть понятно, что эквивалентом утверждения 7.8 для GMM будет

Утверждение 7.10 (асимптотическая нормальность GMM): Предположим, что выполняются предположения утверждения 7.7, так что $\{w_t\}$ стационарный и эргодический процесс, $\widehat{W} (K \times K)$, сходится по вероятности к симметричной положительно определенной матрице W , и p -мерная GMM-оценка $\hat{\theta}$ состоятельна. Предположим также, что

- (1) θ_0 — внутренняя точка Θ .
- (2) $g(w_t; \theta)$ непрерывно дифференцируема по θ для любых w_t , $(K \times 1)$.
- (3) $\frac{1}{\sqrt{n}} \sum_{t=1}^n g(w_t; \theta_0) \rightarrow_d N(0, S_{\frac{1}{\sqrt{n}}})$, S положительно определенная.
- (4) (Условие локального доминирования $\frac{\partial g(w_t; \theta)}{\partial \theta'}$) для некоторой окрестности N' точки θ_0 .

$$\sup_{\theta \in N'} \left\| \frac{\partial g(w_t; \theta)}{\partial \theta'} \right\| < \infty,$$

так что для любой состоятельной оценки $\hat{\theta}$,

$$\frac{1}{n} \sum_{t=1}^n \frac{\partial g(w_t; \hat{\theta})}{\partial \theta'} \xrightarrow{p} E \left[\frac{\partial g(w_t; \theta_0)}{\partial \theta'} \right],$$

(б) $E \left[\frac{\partial g(\omega_i; \theta_0)}{\partial \theta} \right]_{(K \times p)}$ — матрица полного столбцового ранга.

Тогда будут верны следующие результаты:

(а) (Асимптотическая нормальность.) $\hat{\theta}$ асимптотически нормальна с

$$\text{Avar}(\hat{\theta}) = (G'WG)^{-1}G'WSWG(G'WG)^{-1}, \quad (7.3.33)$$

где $G_{(K \times p)} \triangleq E \left[\frac{\partial g(\omega_i; \theta_0)}{\partial \theta} \right]$.

(б) (Состоятельная оценка асимптотической ковариационной матрицы.) Предположим, что имеется состоятельная оценка $\hat{S}$ для S . Асимптотическая ковариационная матрица состоятельна и оценивается как

$$\widehat{\text{Avar}}(\hat{\theta}) = (\hat{G}'\hat{W}\hat{G})^{-1}\hat{G}'\hat{W}\hat{S}\hat{W}\hat{G}(\hat{G}'\hat{W}\hat{G})^{-1}, \quad (7.3.34)$$

где $\hat{G}_{(K \times p)} \triangleq G_n(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n \frac{\partial g(\omega_i; \hat{\theta})}{\partial \theta}$.

Предположение о том, что S положительно определена, на самом деле не нужно для доказательства утверждения, однако мы можем его сделать, потому что практически во всех прикладных задачах это предположение выполняется (или предполагается), а также это нам будет необходимо позднее, при обсуждении проверки гипотез в данной главе.

Полезно отметить аналогию между линейным и нелинейным GMM, сравнив это утверждение с утверждением 3.1. В случае линейного GMM, $g_n(\theta)$ задается как

$$g_n(\theta) = \left(\frac{1}{n} \sum_{i=1}^n x_i \cdot y_i \right) - \left(\frac{1}{n} \sum_{i=1}^n x_i z_i' \right) \theta. \quad (7.3.35)$$

Так что $\hat{G} \equiv G_n(\hat{\theta})$ в утверждении 7.10 сводится к $(-\frac{1}{n} \sum x_i z_i')$, а G сводится к $-E(x_i z_i')$. Суть утверждения 7.10 состоит в том, что асимптотическое распределение нелинейной оценки GMM может быть получено путем линейной аппроксимации $g_n(\theta)$ в окрестности вектора истинных значений параметра.

Аналогия с линейным GMM наблюдается и далее.

- (Эффективность нелинейного GMM.) Используя текущие обозначения, запишем матричное неравенство (3.5.11) из параграфа 3.5.

$$(G'WG) - G'WSWG(G'WG)^{-1} \geq (G'S^{-1}G)^{-1}, \quad (7.3.36)$$

которое утверждает, что нижней границей для $\text{Avar}(\hat{\theta})$ является $(G'WG)^{-1}$. Нижняя граница достигается, если $\hat{W}$ удовлетворяет условию эффективности: $\text{plim}_{n \rightarrow \infty} \hat{W} = S^{-1}$, то есть $\hat{W} = \hat{S}^{-1}$ для нелинейной состоятельной оценки S .

- (*J*-статистика для проверки сверхидентифицируемости ограничений)
Предположим, что условия утверждения 7.10 выполнены и что $\tilde{W} = \tilde{S}^{-1}$, так что $\tilde{W}$ удовлетворяет условной эффективности: $\tilde{W} = S^{-1}$. Тогда минимизированное расхождение $J = g_n(\hat{\theta})' \tilde{S}^{-1} g_n(\hat{\theta})$ ($= -2n(\hat{Q}_n(\hat{\theta}))$) имеет асимптотическое распределение χ^2 -квадрат с $K - p$ степенями свободы. Доказательство этого результата точно такое же, как и в линейном случае, см.: [Newey and McFadden, 1994, Section 9.5] для доказательства.
- (Оценивание S) $S \sim$ это дисперсионная ковариационная матрица $\{g(w_i; \theta_0)\}$. При выполнении некоторых технических условий методы, представленные в параграфе 6.6, — такие, как VARHAC, — могут применяться к оценочным рядам $\{g(w_i; \hat{\theta})\}$ для получения $\tilde{S}$. В частности если w_i i.i.d., то тогда $S = E[g(w_i; \theta_0) \times g(w_i; \theta_0)']$ и в роли $\tilde{S}$ может выступать второй выборочный момент $g(w_i; \hat{\theta})$.

Сопоставление ГММ с МЛ

Вопрос, на который мы только что ответили, таков: если принимать условия ортогональности как заданные, то каков оптимальный выбор взвешивающей матрицы W ? Связанный с предыдущим, но все же отдельный вопрос: каков оптимальный выбор условий ортогональности¹? Этот вопрос также имеет однозначный ответ при условии, что параметрическая форма функции плотности для данных $(w_1, \dots, w_n)$ известна. Здесь мы сконцентрируемся на случае i.i.d., считая, что плотность вероятности w_i задана функцией $f(w_i; \theta)$. Пусть $\hat{\theta}$ — оценка ГММ, связанная с условиями ортогональности $E[g(w_i; \theta_0)] = 0$. Ее асимптотическая ковариационная матрица $\text{Avar}(\hat{\theta})$ задается в утверждении 7.10. Несложно показать, что

$$\text{Avar}(\hat{\theta}) \geq E[s(w_i; \theta_0) s(w_i; \theta_0)']^{-1}, \quad \text{где } s(w_i; \theta_0) = \frac{\partial \ln f(w_i; \theta)}{\partial \theta} \quad (7.3.37)$$

То есть матрица, обратная к информационной матрице $E(ss')$, — нижняя граница для асимптотической ковариационной матрицы оценок ГММ. Это матричное неравенство выполняется при выполнении предположений утверждения 7.10 и некоторых технических условий на $f(w_i; \theta)$.

¹Еще один вопрос об эффективности касается оптимального выбора инструментов в обобщенном нелинейном методе инструментальных переменных (частный случай ГММ, в котором функция g — произведение вектора инструментов на вектор ошибок для оцениваемого уравнения). По поводу обоснования оптимальных инструментов см.: [Newey, McFadden, 1994, Section 5.4]

которые допускают перестановочность дифференцирования и интегрирования. См.: [Newey and McFadden 1994, Theorem 5.1] для формулировки этих условий и доказательства. Тогда, исходя из этого результата, асимптотическая эффективность оценок ML выше, чем у GMM-оценок, потому что указанная нижняя граница $[E(\dot{y}_t \dot{y}_t')]^{-1}$ — это асимптотическая ковариационная матрица оценки ML. Превосходство ML, однако, не очень удивительно, учитывая то, что ML использует информацию относительно параметрической формы $f(w_t; \theta)$ функции плотности вероятности, в то время как GMM не делает этого.

Как бы покажете в контрольном вопросе 4 ниже, GMM достигает этой нижней границы, когда функция g и условия ортогональности являются функцией вклада для наблюдения t :

$$g(w_t; \theta) = s(w_t; \theta) = \frac{\partial \ln f(w_t; \theta)}{\partial \theta}. \quad (7.3.38)$$

Таким образом, оценка GMM с оптимальными условиями ортогональности асимптотически эквивалентна ML. В действительности они численно эквивалентны, что можно показать следующим образом. Поскольку K (количество условий ортогональности) равно p (количеству параметров), когда g выбрано оптимально, как в (7.3.38), оценка GMM $\hat{\theta}$ должна удовлетворять условию $g_n(\hat{\theta}) = 0$, что, в соответствии с (7.3.38), может быть записано как

$$\frac{1}{n} \sum_{t=1}^n s(w_t; \hat{\theta}) = 0. \quad (7.3.39)$$

Это не что иное, как уравнения правдоподобия (то есть условия первого порядка для ML). Они имеют хотя бы одно решение, потому что оценка ML удовлетворяет им. Если у них ровно одно решение, то $\hat{\theta}$ является оценкой ML, равно как и оценкой GMM. Если у них более одного решения, то нам нужно выбрать одно из них в качестве GMM-оценки. Так как одним из решений является оценка ML, мы можем просто выбрать ее в качестве оценки GMM.

Выражение для ошибки оценки в едином формате

Чтобы подготовиться к изучению вопросов следующего параграфа, полезно привести несколько отличающееся доказательство асимптотической нормальности. Мы использовали теорему о среднем значении, чтобы вывести разложение в среднем значении для ошибки оценки, которое представлено в (7.3.7) для M-оценок и в (7.3.32) для GMM. В этом параграфе мы покажем ниже, что они могут быть записаны в одинаковом для обеих оценок виде (7.3.43).

Рассмотрим сначала M -оценку. Учитывая, что

$$\frac{\partial Q_n(\theta_0)}{\partial \theta} = \frac{1}{n} \sum_{i=1}^n s(w_i; \theta_0),$$

разложение в среднем значении (7.3.7) может быть записано как

$$\sqrt{n}(\hat{\theta} - \theta_0) = \left[\frac{1}{n} \sum_{i=1}^n H(w_i; \bar{\theta}) \right]^{-1} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta}. \quad (7.3.40)$$

По условию (4) утверждения 7.8, $\frac{1}{n} \sum_{i=1}^n H(w_i; \bar{\theta})$ сходится по вероятности к некоторой симметричной матрице Ψ размера $p \times p$, заданной как

$$\Psi_{(p \times p)} = E[H(w_i; \theta_0)]. \quad (7.3.41)$$

Перепишем (7.3.40) в виде:

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta_0) &= -\Psi^{-1} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \\ &= \left\{ \left[\frac{1}{n} \sum_{i=1}^n H(w_i; \bar{\theta}) \right]^{-1} - \Psi^{-1} \right\} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \end{aligned} \quad (7.3.42)$$

По построению, выражение, заключенное в скобки, сходится по вероятности к нулю. По условию (3) утверждения 7.8,

$$\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} (= \frac{1}{\sqrt{n}} \sum_{i=1}^n s(w_i; \theta_0))$$

сходится к случайной величине. Так что последняя составляющая, которая является произведением выражения, заключенного в скобки, на $\sqrt{n} \times \frac{\partial Q_n(\theta_0)}{\partial \theta}$, сходится по вероятности к нулю, по лемме 2.4(b). Этот факт можно компактно записать в виде **разложения Тейлора** (*Taylor expansion*):¹

$$\sqrt{n}(\hat{\theta} - \theta_0) = \underbrace{-\Psi^{-1}}_{(p \times p)} \underbrace{\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta}}_{(p \times 1)} + \underbrace{o_p}_{(p \times 1)}, \quad (7.3.43)$$

где слагаемое o_p обозначает «некоторую случайную величину, которая сходится к нулю по вероятности»². Точное выражение для слагаемого o_p будет зависеть от контекста. Здесь оно равно последней составляющей

¹Для этого представление более подходит термин «разложение Тейлора», а не «разложение среднего значения», потому что частная, которая перемножается с $\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta}$, вычисляется в θ_0 , а не в $\bar{\theta}$, так это делается в разложении в среднем значении.

²Обозначение o_p было введено в параграфе 2.1.

в (7.3.42). Как вы увидите, все, что важно для явного анализа, — это чтобы слагаемое o_p исчезло (чтобы оно сходилось к нулю по вероятности), а не его выражение. Так как разность между $\sqrt{n}(\hat{\theta} - \theta_0)$ и $-\Psi^{-1} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta}$ исчезает, асимптотическое распределение $\sqrt{n}(\hat{\theta} - \theta_0)$ точно такое же, как асимптотическое распределение $-\Psi^{-1} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta}$ по лемме 2.4(a). Тогда из этого следует, что $\sqrt{n}(\hat{\theta} - \theta_0)$ сходится к нормальному распределению с нулевым математическим ожиданием и асимптотической ковариационной матрицей, заданной соотношением

$$\text{Avar}(\hat{\theta}) = \Psi^{-1} \Sigma \Psi^{-1} \quad \text{где} \quad \Sigma \equiv \text{Avar} \left(\frac{\partial Q_n(\theta_0)}{\partial \theta} \right). \quad (7.3.44)$$

Для M-оценок $\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{i=1}^n s(w_i; \theta)$ и Σ является долгосрочный ковариационной матрицей $s(w_i; \theta)$. Если положить $\Psi = E[H(w_i; \theta)]$, то мы получим выражение для $\text{Avar} \hat{\theta}$ из утверждения 7.8.

Теперь обратимся к GMM. Выражение для $\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta}$ выглядит следующим образом:

$$\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} = -[G_n(\theta_0)]' \hat{W} \frac{1}{\sqrt{n}} \sum_{i=1}^n g(w_i; \theta_0). \quad (7.3.45)$$

(Вспомним, что $G_n(\theta) \equiv \frac{\partial g_n(\theta)}{\partial \theta}$.) Так как $G_n(\theta_0) = \frac{1}{n} \sum \frac{\partial g(w_i; \theta_0)}{\partial \theta} \rightarrow_p G \equiv E\left[\frac{\partial g(w_i; \theta_0)}{\partial \theta}\right]$, $\hat{W} \rightarrow_p W$, и $\frac{1}{\sqrt{n}} \sum_{i=1}^n g(w_i; \theta_0) \rightarrow_d N(0, S)$ при выполнении условий утверждения 7.10, $\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta}$ будет сходиться к нормальному распределению с нулевым математическим ожиданием и

$$\Sigma \equiv \text{Avar} \left(\frac{\partial Q_n(\theta_0)}{\partial \theta} \right) = \begin{matrix} G' & W & S & W & G \\ (p \times p) & (p \times K) & (K \times K) & (K \times p) & (p \times p) \end{matrix} \quad (7.3.46)$$

Теперь перепишем разложение в среднем значении (7.3.32) для ошибки GMM-оценки как

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta_0) &= B^{-1}e, \quad B \equiv G_n(\hat{\theta})' \hat{W} G_n(\bar{\theta}), \\ e &\equiv -G_n(\hat{\theta})' \hat{W} \frac{1}{\sqrt{n}} \sum_{i=1}^n g(w_i; \theta_0). \end{aligned} \quad (7.3.47)$$

Далее, в контрольном вопросе 5, будет предложено показать, что это можно записать как разложение Тейлора (7.3.43) с симметричной матрицей Ψ , заданной соотношением:

$$\Psi \equiv - \begin{matrix} G' & W & G \\ (p \times p) & (p \times K) & (K \times p) \end{matrix} \quad (7.3.48)$$

Подставляя (7.3.48) и (7.3.46) в (7.3.44), мы получим асимптотическую ковариационную матрицу GMM-оценки, описанную в утверждении 7.10. Табл. 7.1 подытоживает анализ, проведенный в этом подпараграфе, показывая, как стоит проводить подстановку, чтобы сделать общие формулы применимыми для M-оценок и GMM.

Таблица 7.1

Разложение Тейлора для отклонения
выборочной оценки от истинного значения параметра

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\Psi^{-1} \sqrt{n} \frac{\partial \psi_n(\theta_0)}{\partial \theta} + o_p(1) \sqrt{n} \frac{\partial \psi_n(\theta_0)}{\partial \theta} \rightarrow N(0, \Sigma), \text{ Var}(\hat{\theta}) = \Psi^{-1} \Sigma \Psi^{-1}$$

Элементы для замены	M-оценки	GMM
$\psi_n(\theta)$	$\frac{1}{n} \sum_{i=1}^n \psi(w_i; \theta)$	$\frac{1}{2} g_n(\theta)' \bar{W} g_n(\theta)$
$\sqrt{n} \frac{\partial \psi_n(\theta_0)}{\partial \theta}$	$\frac{1}{n} \sum_{i=1}^n \psi(w_i; \theta_0)$	$-[G_n(\theta_0)' \bar{W}] \frac{1}{n} \sum_{i=1}^n g(w_i; \theta_0)$
Ψ	$E[\psi(w_i; \theta_0)]$	$-G'WG$
Σ	долготренинный ковариационная матрица для $\psi(w_i; \theta_0)$	$G'WSWG$

Примечание. Оператором определены $\psi(w_i; \theta)$ и $H(w_i; \theta)$ см. (7.3.2) и (7.3.3). Кроме того

$$g_n(\theta) = \frac{1}{n} \sum_{i=1}^n g(w_i; \theta) \text{ и } G = E \left[\frac{\partial g(w_i; \theta_0)}{\partial \theta} \right].$$

Между прочим, для анализа, проводимого в следующем параграфе, будет полезно заметить, что $\Sigma = -\Psi$ для ML с i.i.d. наблюдениями и для эффективного GMM (GMM с $W = S^{-1}$). Для ML $\Sigma = -E[\psi(w_i; \theta_0)\psi(w_i; \theta_0)']$, что равняется матрице Ψ в (7.3.41), взятой со знаком минус. Благодаря желательности информации матрицы. Для эффективного GMM, задав $W = S^{-1}$ в (7.3.46), мы получим

$$\Sigma = G'S^{-1}G, \quad (7.3.49)$$

что есть Ψ в (7.3.48) со знаком минус.

Контрольные вопросы

1. (Вклад и гессиан целевой функции) Для M-оценок и, в частности, ML мы определили вклад наблюдения i , $\psi(w_i; \theta)$ как градиент (вектор первых производных) и гессиан для наблюдения i , $H(w_i; \theta)$, как гессиан функции $\psi(w_i; \theta)$. Пусть вклад (без добавления слов «для наблюдения i ») — это градиент целевой функции $\psi_n(\theta)$, обозначаемый как $\psi_n(\theta)$. Пусть гессиан (без добавления слов «для наблюдения i ») — это матрица гессиан

функции $Q_n(\theta)$, обозначая их как $H_n(\theta)$. При выполнении условий утверждения 7.9 покажите, что

$$E[s_n(\theta_0)] = 0, \quad -\frac{1}{n} E[H_n(\theta_0)] = E[s_n(\theta_0)s_n'(\theta_0)].$$

Указание: $\{s_n(w_i; \theta_0)\}_{i=1}^n$

2. (Равенство для условных информационных матриц в модели линейной регрессии.) Для линейной регрессии в примере 7.10 проверьте, что

$$E[s(w_i; \theta_0) | x_i] = 0 \quad \text{и} \quad -E[s(w_i; \theta_0) | x_i] = E[s(w_i; \theta_0)s(w_i; \theta_0)' | x_i].$$

3. (Анализ модели линейной регрессии.) Для модели линейной регрессии из примера 7.10 пусть $(\hat{\beta}, \hat{\sigma}^2)$ — МЛ-оценка (β, σ^2) . Запишите первую оценку $\text{Avar}(\hat{\beta})$, определенную в (7.3.14). Верно ли равенство:

$$\widehat{\text{Avar}}(\hat{\beta}) = \hat{\sigma}^2 \left(\frac{1}{n} \sum x_i x_i' \right)^{-1}.$$

(Ответ: Да.)

4. (GMM с оптимальными условиями ортогональности.) Предположим, что соответствующие технические условия на гипотетическую функцию плотности $f(w; \theta)$ выполняются, так что равенство для информационных матриц будет верным. Используя утверждение 7.10, покажите, что, когда функция g с моментных условиях задается как в (7.3.38), асимптотическая ковариационная матрица оценки GMM равняется матрице, обратной к информационной матрице для наблюдения i (являющейся нижней границей в (7.3.37)). Указание: Так как K равняется $p \times p$ -квадратная матрица и выражение для $\text{Avar}(\hat{\theta})$ в утверждении 7.10 редуцируется к $\text{Avar}(\hat{\theta}) = G^{-1} S G^{-1}$. Из выбора g следует, что $G = E[H(w; \theta_0)]$ и что $S = E[s(w; \theta_0)s(w; \theta_0)']$.

5. (Разложение Тейлора для ошибки GMM-оценки.) Покажите, что (7.3.47) может быть записано как разложение в ряд Тейлора (7.8.43) при выполнении условий утверждения 7.10, выполняя следующие этапы:

(а) Покажите, что $B^{-1} = \Psi^{-1} + Y_n$, где $Y_n \rightarrow_p 0$.

(б) Покажите, что $a = \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} = \Psi_n$, где $\Psi_n \rightarrow_p 0$. Указание:

$$Q_n(\hat{\theta}) - Q_n(\theta_0) = (G_n(\hat{\theta}) - G) + (G - G_n(\theta_0)) \geq_p 0 \quad \text{и}$$

$$\bar{W} \frac{1}{\sqrt{n}} \sum_{i=1}^n g(w_i; \theta_0)$$

сходится по распределению к случайной величине.

(в) Покажите, что $B^{-1}a = -\Psi^{-1} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} = x_n$, где

$$x_n = - \left[Y_n \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} + \Psi^{-1} y_n + Y_n y_n \right].$$

(г) Покажите, что $x_n \rightarrow_p 0$.

6. Почему расстояние в GMM делится на 2? Проверьте, что если бы минус $g_n(\theta)' \hat{S}^{-1} g_n(\theta)$ не делилось на 2 в определении целевой функции эффективного GMM, то мы бы не имели $\Sigma = -\Psi$.

7.4. Проверка гипотез

Как хорошо известно для ML, существует триада статистик — статистика Вальда, статистика множителей Лагранжа (LM) и статистика отношения правдоподобий (LR), которые могут быть использованы для проверки нулевой гипотезы. Эти три статистики асимптотически эквивалентны, в том смысле, что у них одинаковое асимптотическое распределение (χ^2). В этом параграфе мы докажем, что асимптотическая эквивалентность этой триады может быть распространена на GMM, выстраивая доказательство, применимое как к ML, так и к GMM. Краеугольным камнем этого доказательства будет общий формат ошибки оценки, который был получен в конце предыдущего параграфа. Читая этот параграф впервые, вы, возможно, захотите прочесть следующий подпараграф, а потом перейти к последнему подпараграфу.

Нулевая гипотеза

Мы уже рассматривали в контексте модели линейной регрессии проблему тестирования набора l возможно нелинейных ограничений (см. параграф 2.4). Кратко повторим: пусть θ_0 — p -мерный параметр модели. Нулевая гипотеза может быть выражена как

$$H_0: \alpha(\theta_0) = 0. \quad (7.4.1)$$

($l \times 1$)

Мы предполагаем, что $\alpha(\cdot)$ непрерывно дифференцируема. Также обозначим

$$A(\theta) \equiv \frac{\partial \alpha(\theta)}{\partial \theta} = \quad (7.4.2)$$

($l \times p$)

якобиан $\alpha(\theta)$. Мы предполагаем, что

$$A_0 \equiv A(\theta_0) \text{ имеет полный столбцовый ранг.} \quad (7.4.3)$$

($l \times l$)

Это гарантирует, что указанные l ограничений не дублируют друг друга. Это условие на ранг верно как $l \leq p$.

Пусть $\hat{\theta}$ — рассматриваемая экстремальная оценка. Это либо ML-оценка, либо GMM-оценка. Она решает задачу безусловной оптимизации (7.1.1). Статистика Вальда для проверки нулевой гипотезы использует оценку

без ограничений. LM-статистика, напротив, использует *оценку с ограничениями* (*constrained estimator*), обозначаемую как $\hat{\theta}$, которая является решением задачи

$$\max_{\theta \in \Omega} Q_n(\theta) \text{ при ограничениях } a(\theta) = 0. \quad (7.4.4)$$

Как было показано в параграфе 7.2, истинное значение параметра θ_0 является решением «предельной задачи безусловной оптимизации», где Q_n в задаче безусловной оптимизации (7.4.1) заменяется предельной функцией Q_0 . Она также является решением «предельной задачи условной оптимизации», где Q_n в выпуклостной задаче условной оптимизации заменяется предельной функцией Q_0 , потому что θ_0 удовлетворяет ограничениям. Оказывается, что равномерная сходимость по вероятности $Q_n(\cdot)$ к $Q_0(\cdot)$ обеспечивает то, что предел оценки с ограничениями $\hat{\theta}$ является решением предельной задачи условной оптимизации, что есть θ_0 . По поводу доказательства см.: [Newey and McFadden, 1994, Theorem 9.1] для GMM и [Gallant and White, 1988, Lemma 7.3(h)] для общего случая экстремальных оценок. Также возможно показать, что если все условия, обеспечивающие асимптотическую нормальность для оценки модели без ограничений, выполняются, то тогда $\hat{\theta}$ будет также асимптотически нормальной при выполнении нулевой гипотезы. Для доказательства этого свойства для GMM-оценок см.: [Newey and McFadden, 1994, Theorem 9.2]; для ML, см., например: [Amemiya, 1985, Section 4.5].

Рабочие предположки

Доказательство того, что статистики Вальды, LM и LR имеют асимптотическое распределение $\chi^2(r)$, базируется на выполнении следующих условий:

(A) $\sqrt{n}(\hat{\theta} - \theta_0)$ допускает разложение в ряд Тейлора (7.3.43), которое имеет вид:

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\Psi^{-1} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} + o_p. \quad (7.4.5)$$

(B) $\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \rightarrow_d N(0, \Sigma)$, Σ положительно определена.

(C) $\sqrt{n}(\hat{\theta} - \theta_0)$ сходится по распределению к некоторой случайной величине.

(D) $\Sigma = -\Psi$.

¹Если нам интересно доказательство, то оно представляет собой всего лишь вариант разложения в ряд Тейлора из рассуждений предыдущего параграфа, посвященного LM-статистике, который базируется на разложении Тейлора.

Мы показали в предыдущем параграфе, что первые два условия выполняются для условного ML при выполнении условий утверждения 7.9. для безусловного ML — при выполнении этих условий, надлежащим образом модифицированных, а для GMM — при выполнении условий утверждения 7.10. Как только что было замечено, при выполнении тех же условий, оценка модели с параметрами θ состоятельна и асимптотически нормальна, что обеспечивает выполнение условия (C). Таким образом, условия (A)–(C) вытекают из условий, которые обеспечивают асимптотическую нормальность экстремальной функции.

Статистики Вайлда и LM, если их определять должным образом, так, чтобы они не зависели от выполнения условия (D), асимптотически распределены по закону $\chi^2(r)$. По сути условие (D) выполняется, можно упростить выражения для этих статистик. Более того, без этого условия статистика LR не будет распределена асимптотически по $\chi^2(r)$. Как было отмечено в конце предыдущего параграфа, условие (D) выполняется для ML и для эффективного GMM с $W = S^{-1}$. Таким образом, та часть анализа, которая базируется на условии (D), применима к неэффективному GMM.

Степанецкий Валерий

В этой книге мы несколько раз получали статистику Вальда, используя «дедуктивный» из леммы 2.5. Здесь мы приведем несколько отличающуюся процедуру вычисления этой статистики, основанную на разложении Тейлора. Применяя теорему о среднем значении к $\sigma^2(\hat{\theta})$ в окрестности θ_0 и учитывая, что $\sigma^2(\theta_0) = 0$ при выполнении нулевой гипотезы, мы получим следующее приближенное выражение для $\sqrt{n}\sigma(\hat{\theta})$:

$$\sqrt{n}n(\hat{\theta}) = A(\tilde{\theta}); \quad \sqrt{n}(\hat{\theta} - \theta_0) \quad (7.4.6)$$

Полученное разложение Тейлора через разложение в средние значения более прозрачно, чем то, которое было проведено в предыдущем параграфе. Так как $\bar{\theta}_j$ лежит между θ_j и $\bar{\theta}$, состоятельность и так как $A(\cdot)$ непрерывна, $A(\bar{\theta})$ сходится по вероятности к $A_0 (= A(\theta_0))$. Второе слагаемое, на который умножается $A(\bar{\theta})$, $\sqrt{n}(\bar{\theta} - \theta_j)$, сходится к некоторой случайной величине. Так как $(A(\bar{\theta}) - A_0)\sqrt{n}(\bar{\theta} - \theta_j)$ стремится (сходится по вероятности к нулю), по лемме 2.4(b). Обозначая это слагаемое как α_j , перепишем разложение в среднем значениях в виде разложения Тейлора:

$$\frac{\sqrt{n}m(\hat{\theta})}{\sqrt{g(\hat{\theta})}} = \frac{A_0}{\sqrt{g(\hat{\theta})}} \frac{\sqrt{n}(\hat{\theta} - \theta_0)}{(\hat{\theta} - \theta_0)} + o_p(1) \quad (7A.7)$$

(Учитите, что такая аргументация была бы неверна, если бы $\sqrt{a}(\vec{\theta} - \theta_0)$ не сходилось к некоторой случайной величине.)

Подставляя разложение Тейлора ((7.4.5) в это равенство, мы получаем уравнение, связывающее $\sqrt{n}\alpha(\hat{\theta})$ и $\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta}$:

$$\sqrt{n}\alpha(\hat{\theta}) = - \underset{(r \times r)}{A_0} \underset{(r \times p)}{\Psi^{-1}} \underset{(p \times p)}{\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta}} + o_p. \quad (7.4.8)$$

(Здесь o_p — это умноженное на A_0 o_p из (7.4.7) плюс o_p из (7.4.7), но оно все равно сходится к нулю по вероятности.) Из этого выражения и асимптотической нормальности $\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta}$ (условие (B)) сразу же следует, что $\sqrt{n}\alpha(\hat{\theta})$ сходится к нормальному распределению с нулевым математическим ожиданием и

$$\begin{aligned} \text{Var}(\alpha(\hat{\theta})) &= A_0 \Psi^{-1} \Sigma \Psi^{-1} A_0' = \\ &= A_0 \underset{(r \times p)}{\Sigma^{-1}} \underset{(p \times p)}{A_0'} \quad (\text{если выполняется условие (D)}). \end{aligned} \quad (7.4.9)$$

Поскольку матрица A_0 размера $r \times p$ — матрица полного строкового ранга и Σ — положительно определенная матрица (по условию (B)), эта матрица размера $r \times r$ положительно определена. Таким образом, соответствующая квадратичная форма

$$n\alpha(\hat{\theta})' [A_0 \Sigma^{-1} A_0']^{-1} \alpha(\hat{\theta}) \quad (7.4.10)$$

имеет асимптотическое распределение $\chi^2(r)$ при нулевой гипотезе. Как было замечено выше, $A(\hat{\theta}) \rightarrow_p A_0$. Так что если имеется состоятельная оценка $\hat{\Sigma}$ для Σ , то $A(\hat{\theta})\hat{\Sigma}^{-1}A(\hat{\theta})'$ — состоятельная оценка для $A_0\Sigma^{-1}A_0'$, из чего, по лемме 2.4(б), следует, что статистика Вальда определена как

$$W \equiv n\alpha(\hat{\theta})' [A(\hat{\theta})\hat{\Sigma}^{-1}A(\hat{\theta})']^{-1} \alpha(\hat{\theta}), \quad (7.4.11)$$

также имеет асимптотическое $\chi^2(r)$ распределение при нулевой гипотезе.

Как получить состоятельную оценку $\hat{\Sigma}$ для $\Sigma(-\Psi)$, зависит от того, является ли экстремальная оценка ML-оценкой или оценкой эффективности GMM. Для ML она задается эффективной оценкой $-\Psi = -E[H(w_i; \theta_0)]$:

$$-\frac{\partial^2 Q_n(\hat{\theta})}{\partial \theta \partial \theta'} = -\frac{1}{n} \sum_{i=1}^n H(w_i; \hat{\theta}) \quad (7.4.12)$$

или состоятельной оценкой $\Sigma = E[s(w_i; \theta_0)s(w_i; \theta_0)']$:

$$\frac{1}{n} \sum_{i=1}^n s(w_i; \hat{\theta}) s(w_i; \hat{\theta})'. \quad (7.4.13)$$

Для эффективного GMM, $\Sigma = G'G^{-1}G$ (см. (7.3.49)), что можно достаточно оценить как

$$\hat{\Sigma} = \hat{G}' \hat{S}^{-1} \hat{G} \quad \text{с} \quad \hat{G}_{(K \times n)} \equiv G_n(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n \frac{\partial g(w_i; \hat{\theta})}{\partial \theta}. \quad (7.4.14)$$

Как уже отмечалось, $\hat{\theta}$ — оценка одновременной ковариационной матрицы, полученная из оценочного ряда $\{g(w_i; \hat{\theta})\}$.

Статистика множителей Лагранжа (LM-статистика)

Пусть γ_0 ($r < 1$) — множитель Лагранжа для функции ограничения (7.4.4). Оценка модели с ограничениями $\hat{\theta}$ удовлетворяет условиям первого порядка:

$$\sqrt{n} \frac{\partial Q_n(\hat{\theta})}{\partial \theta} + A(\hat{\theta})' \sqrt{n} \gamma_0 = 0_{(r \times 1)}, \quad (7.4.15)$$

$$\sqrt{n} a(\hat{\theta}) = 0_{(r \times r)}. \quad (7.4.16)$$

Мы хотим получить разложение Тейлора для этих условий первого порядка. По условию (C), $\sqrt{n}(\hat{\theta} - \theta_0)$ сходится к некоторой случайной величине. Тогда $\sqrt{n} a(\hat{\theta})$ допускает разложение Тейлора:

$$\sqrt{n} a(\hat{\theta}) = A_0 \sqrt{n}(\hat{\theta} - \theta_0) + o_p. \quad (7.4.17)$$

В контрольном вопросе 2 будет предложено показать, что следующее разложение Тейлора верно как для ML, так и для GMM.

$$\sqrt{n} \frac{\partial Q_n(\hat{\theta})}{\partial \theta} = \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} + \Psi \sqrt{n}(\hat{\theta} - \theta_0) + o_p. \quad (7.4.18)$$

Следствием этого равенства является то, что $\sqrt{n} \frac{\partial Q_n(\hat{\theta})}{\partial \theta}$ сходится к некоторой случайной величине. Тогда $\sqrt{n} \gamma_0$, удовлетворяющее (7.4.15), сходится к случайной величине. Следовательно, поскольку $A(\hat{\theta}) \rightarrow_p A_0$, мы получаем:

$$A(\hat{\theta})' \sqrt{n} \gamma_0 = A_0' \sqrt{n} \gamma_0 + (A(\hat{\theta}) - A_0)' \sqrt{n} \gamma_0 = A_0' \sqrt{n} \gamma_0 + o_p. \quad (7.4.19)$$

Подставляя эти три уравнения в условия первого порядка, мы получаем

$$\begin{bmatrix} \Psi & A_0' \\ (r \times r) & (r \times r) \\ A_0 & 0 \\ (r \times r) & (r \times r) \end{bmatrix} \begin{bmatrix} \sqrt{n}(\hat{\theta} - \theta_0) \\ (r \times 1) \\ \sqrt{n} \gamma_0 \\ (r \times r) \end{bmatrix} = \begin{bmatrix} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \\ (r \times 1) \\ 0 \\ (r \times r) \end{bmatrix} + o_p. \quad (7.4.20)$$

Используя формулу обращения для блочных матриц, можно решить это уравнение и получить в итоге:

$$\sqrt{n} \gamma_n = -(A_0 \Psi^{-1} A_0')^{-1} A_0 \Psi^{-1} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} + o_p. \quad (7.4.21)$$

$$\sqrt{n}(\hat{\theta} - \theta_0) = -[\Psi^{-1} - \Psi^{-1} A_0' (A_0 \Psi^{-1} A_0')^{-1} A_0 \Psi^{-1}] \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} + o_p. \quad (7.4.22)$$

Дальнейший вывод LM-статистики такой же, как и для статистики Вальда. Согласно (7.4.21) $\sqrt{n} \gamma_n$ сходится к нормальному распределению с нулевым математическим ожиданием и асимптотической ковариационной матрицей, заданной соотношением:

$$\begin{aligned} \text{Var}(\gamma_n) &= (A_0 \Psi^{-1} A_0')^{-1} A_0 \Psi^{-1} \Sigma \Psi^{-1} A_0' (A_0 \Psi^{-1} A_0')^{-1} = \\ &= (A_0 \Sigma^{-1} A_0')^{-1} \quad (\text{ср. выполняется условие (D)}). \end{aligned} \quad (7.4.23)$$

Так как эта матрица размера $r \times r$ положительно определена, соответствующая ей квадратичная форма

$$n \gamma_n' (A_0 \Sigma^{-1} A_0')^{-1} \gamma_n \quad (7.4.24)$$

при выполнении нулевой гипотезы имеет асимптотическое распределение $\chi^2(r)$. Если существует состоятельная оценка матрицы Σ , обозначаемая как $\tilde{\Sigma}$, то тогда $A(\tilde{\theta}) \tilde{\Sigma}^{-1} A(\tilde{\theta})'$ будет состоятельной оценкой для $A_0 \Sigma^{-1} A_0'$. Так что LM-статистика, определенная первой строкой выражения, приведенного ниже, при выполнении нулевой гипотезы имеет асимптотическое распределение $\chi^2(r)$. Эту статистику можно привести к более красивому виду, используя условия первого порядка (7.4.15):

$$\begin{aligned} LM &\equiv n \gamma_n' \underbrace{\left[A(\tilde{\theta}) \tilde{\Sigma}^{-1} A(\tilde{\theta})' \right]}_{(r \times r)} \gamma_n = \\ &= n (A(\tilde{\theta})' \gamma_n)' \tilde{\Sigma}^{-1} (A(\tilde{\theta})' \gamma_n) = \\ &= n \left(\frac{\partial Q_n(\tilde{\theta})}{\partial \theta} \right)'_{(r \times p)} \tilde{\Sigma}^{-1} \left(\frac{\partial Q_n(\tilde{\theta})}{\partial \theta} \right)_{(p \times 1)} \\ &\quad (\text{так как } A(\tilde{\theta})' \gamma_n = - \frac{\partial Q_n(\tilde{\theta})}{\partial \theta} \text{ по (7.4.15)}). \end{aligned} \quad (7.4.25)$$

Последнее выражение для LM-статистики объясняет, почему ее так же называют **тестовой статистикой вклада** (*score test statistic*). Она представляет собой расстояние от нуля до вклада, рассчитанного для оценки с ограничениями $\tilde{\theta}$. Однако вы не должны обманываться красотой этой статистики: несмотря на то что это квадратичная форма,

соответствующая p -мерному вектору, количество степеней свободы ее асимптотического распределения χ^2 равно p , а не p_1 , что ясно из приведенного процесса вывода этой статистики.

Для ML $\tilde{\Sigma}$ задается посредством (7.4.12) или (7.4.13), а численных $\tilde{\theta}$. Для GMM достоверной оценкой Σ будет:

$$\tilde{\Sigma} = \tilde{G}' \tilde{S}^{-1} \tilde{G} \quad \text{с} \quad \tilde{G}_{(K \times p)} = G_{\theta}(\tilde{\theta}) = \frac{1}{n} \sum_{i=1}^n \frac{\partial g(w_i; \tilde{\theta})}{\partial \theta}. \quad (7.4.26)$$

Здесь $\tilde{S}$ — оценка долговременной ковариационной матрицы, построенной для оценочных рядов $\{g(w_i; \tilde{\theta})\}$.

Статистика отношения правдоподобий (LR-статистика)

LR-статистика определяется как умноженная на 2 n разность $Q_n(\tilde{\theta}) - Q_n(\tilde{\theta}_0)$. (Для эффективного GMM называть это статистикой отношения правдоподобия не совсем корректно, потому что Q_n для GMM — это не функция правдоподобия, но мы все же будем использовать этот термин как для ML, так и для GMM.) Для ML и GMM несложно показать (см. контрольные вопросы 6 и 7), что

$$LR = 2n \cdot [Q_n(\tilde{\theta}) - Q_n(\tilde{\theta}_0)] = -n(\tilde{\theta} - \tilde{\theta}_0)' \Psi(\tilde{\theta} - \tilde{\theta}_0) + o_p. \quad (7.4.27)$$

Справедливость этого выражения не зависит от условия (D).

Из этого выражения можно вывести асимптотическое распределение LR-статистики. Вычитая (7.4.22) из (7.4.5), мы получаем:

$$\sqrt{n}(\tilde{\theta} - \tilde{\theta}_0) = -\Psi^{-1} A_0' (A_0 \Psi^{-1} A_0')^{-1} A_0 \Psi^{-1} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} + o_p. \quad (7.4.28)$$

Подставляя это в (7.4.27), можно записать LR-статистику как

$$LR = \left(\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \right)' \Psi^{-1} A_0' (A_0 \Psi^{-1} A_0')^{-1} A_0 \Psi^{-1} \left(\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \right) + o_p. \quad (7.4.29)$$

Теперь используем предположение (D). LR-статистика преобразуется в

$$LR = \left(\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \right)' \Sigma^{-1} A_0' [A_0 \Sigma^{-1} A_0']^{-1} A_0 \Sigma^{-1} \left(\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \right) + o_p. \quad (7.4.30)$$

Это выражение распределится асимптотически по закону $\chi^2(p)$, потому что ковариационная матрица предельного распределения величины $A_0 \Sigma^{-1} \left(\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \right)$ — положительно определенная матрица 4-го блока.

Это завершает доказательство асимптотической эквивалентности триады статистик. Но из этого доказательства можно было бы показать, что

выполняется нечто более сильное, чем просто асимптотическая эквивалентность (тот факт, что три статистики имеют одинаковое асимптотическое распределение). Мы показали и доказываем, что разность между статистикой Вальда и (7.4.16) есть $o_p(1)$. Но, по (7.4.8), разность между вторым выражением и квадратичной формой в выражении для LR-статистики (7.4.30) — это $o_p(1)$, и следовательно когда $\Sigma = -\Psi$. То есть численный разрыв между статистикой Вальда и LR-статистикой будет $o_p(1)$, если $\Sigma = -\Psi$. То же самое верно и для LM-статистики. Разностью между ней и (7.4.24) снова является $o_p(1)$, когда $\Sigma = -\Psi$.

Итоговые выводы для триады статистик

Чтобы подытожить довольно длинное обсуждение в этом параграфе, приведем

Утверждение 7.11 (триада статистик): Пусть $\tilde{\theta}$ — экстремальная оценка, определенная в (7.1.1). Рассмотрим нулевую гипотезу (7.4.1), где $\alpha(\cdot)$ непрерывно дифференцируема (так что якобиан $A(\theta) \equiv \frac{\partial \alpha(\theta)}{\partial \theta}$ непрерывен) и выполняется ранговое условие (7.4.3). Пусть $\tilde{\theta}$ — экстремальная оценка модели с ограничениями, определенная в (7.4.4). Предположим:

- выполнение предположений утверждения 7.9, если экстремальная оценка — это оценка условного ML;
- выполнение предположений, соответствующим образом модифицированных, как показано прямо под утверждением 7.9, в случае оценки безусловного ML;
- выполнение предположений утверждения 7.10 с $W = S^{-1}$ и наличие состоятельной оценки $\tilde{S}$ для S , построенной из $\tilde{\theta}$, и $\tilde{S}$ построенной из $\tilde{\theta}$, в случае эффективного GMM.

Определим статистики Вальда, LM, и LR, как это сделано в табл. 7.2. Тогда

- экстремальная оценка с ограничениями $\tilde{\theta}$ состоятельна и асимптотически нормальна;
- при выполнении нулевой гипотезы все три статистики сходятся по распределению к $\chi^2(r)$, где r — количество ограничений в нулевой гипотезе;
- более того, при выполнении нулевой гипотезы численное различие между указанными тремя статистиками сходится к нулю по распределению.

Таблица 7.1

Триада статистик

Статистики Биванде: $n\alpha(\tilde{\theta})' [A(\tilde{\theta}) \tilde{\Sigma}^{-1} A(\tilde{\theta})]^{-1} \alpha(\tilde{\theta})$
 $\alpha(\tilde{\theta}) = (g(\tilde{\theta}), \frac{\partial g(\tilde{\theta})}{\partial \theta})'$

$$\text{LM-статистика: } n \left(\frac{\frac{\partial Q_n(\tilde{\theta})}{\partial \theta}}{\sqrt{n}} \right)' \tilde{\Sigma}^{-1} \left(\frac{\frac{\partial Q_n(\tilde{\theta})}{\partial \theta}}{\sqrt{n}} \right)$$

$$\text{LR-статистика: } 2n \cdot (Q_n(\tilde{\theta}) - Q_n(\hat{\theta}))$$

Элементы для подстановки	Условный ML	Эффективный GMM
$Q_n(\theta)$	$\frac{1}{n} \sum_{i=1}^n \log f(y_i x_i; \theta)$	$-\frac{1}{2} y_n(S)' \tilde{S}^{-1} y_n(S)$
$\tilde{\Sigma}$	$-\frac{\partial^2 Q_n(\tilde{\theta})}{\partial \theta \partial \theta'}$ или $\frac{1}{n} \sum_{i=1}^n v(w_i; \tilde{\theta}) v(w_i; \tilde{\theta})'$	$\tilde{G}' \tilde{S}^{-1} \tilde{G}$, $\tilde{G} = G_n(\tilde{\theta})$
$\tilde{\Sigma}$	заменить $\tilde{\theta}$ на $\hat{\theta}$ в выражении выше	$\tilde{G}' \hat{S}^{-1} \tilde{G}$, $\tilde{G} = G_n(\tilde{\theta})$

Примечание: По закону определения $v(w_i; \theta)$ и $H(w_i; \theta)$ см. (7.3.2) и (7.3.3). Кроме того,

$$y_n(S) = \frac{1}{n} \sum_{i=1}^n g(w_i; \theta), \quad G_n(\theta) = \frac{\partial g_n(\theta)}{\partial \theta'}, \quad \text{и } G = E \left[\frac{\partial g(w_i; \theta_i)}{\partial \theta'} \right].$$

В $Q_n(\tilde{\theta})$ и $Q_n(\hat{\theta})$ тем GMM используется одна и та же матрица, стоящая здесь посредстве

Контрольные вопросы

1. Проверьте, что LM-статистика для ML может быть записана как

$$n \left(\frac{1}{n} \sum_{i=1}^n v(w_i; \tilde{\theta}) \right)' \left\{ \frac{1}{n} \sum_{i=1}^n v(w_i; \tilde{\theta}) v(w_i; \tilde{\theta})' \right\}^{-1} \left(\frac{1}{n} \sum_{i=1}^n v(w_i; \tilde{\theta}) \right).$$

2. Выведите (7.4.18). **Указание:** Для ML замените $\tilde{\theta}$ на $\tilde{\theta}$ в (7.3.5). Для GMM выведите

$$\begin{aligned} \sqrt{n} \frac{\partial Q_n(\tilde{\theta})}{\partial \theta} &= \\ &= -G_n(\tilde{\theta})' \tilde{W} \frac{1}{\sqrt{n}} \sum_{i=1}^n g(w_i; \theta_i) - G_n(\tilde{\theta})' \tilde{W} G_n(\tilde{\theta}) \sqrt{n}(\tilde{\theta} - \theta_0), \end{aligned}$$

где $\tilde{\theta}$ лежит между θ_0 и $\tilde{\theta}$. Вывод этого должен быть таким же простым, как и вывод (7.3.11).

3. Объясните, почему LR-статистика не будет асимптотически распределена по закону χ^2 , если не выполняется условие $\Sigma = -\Psi$.
4. Рассмотрите триаду статистик, рассчитанных так, как указано в табл. 7.2, для ML, но предположите, что данные w_i автокоррелированы. Какая из трех оценок все еще будет распределена асимптотически по закону χ^2 ? [Ответ: Никакая.]

5. Причисливая, что условие $\Sigma = -\Psi$ не выполняется, мы предположим, что у нас есть состоятельная оценка $\hat{\Psi}$ для матрицы Ψ , также как и состоятельная оценка $\hat{\Sigma}$ для Σ . Предложите LM-статистику, которая имеет асимптотически нормальное распределение $\chi^2(r)$. Укажите: Используйте члены разложения в (7.4.23), которое отвечает верным и без выполнения $\Sigma = -\Psi$. Ответ так:

$$n \left(\frac{\partial Q_n(\hat{\theta})}{\partial \theta} \right)' \hat{\Psi}^{-1} A(\hat{\theta})' [A(\hat{\theta}) \hat{\Psi}^{-1} \hat{\Sigma} \hat{\Psi}^{-1} A(\hat{\theta})']^{-1} A(\hat{\theta}) \hat{\Psi}^{-1} \left(\frac{\partial Q_n(\hat{\theta})}{\partial \theta} \right). \quad (7.4.31)$$

6. (По желанию доказательства (7.4.27).) Докажите (7.4.27) для M-оценок. Укажите: Воспользуйтесь разложением в среднем значения которого порядка в окрестности $\hat{\theta}$

$$Q_n(\hat{\theta}) = Q_n(\bar{\theta}) + \frac{\partial Q_n(\bar{\theta})}{\partial \theta'} (\hat{\theta} - \bar{\theta}) + \frac{1}{2} (\hat{\theta} - \bar{\theta})' \frac{\partial^2 Q_n(\bar{\theta})}{\partial \theta \partial \theta'} (\hat{\theta} - \bar{\theta}). \quad (7.4.32)$$

где $\bar{\theta}$ лежит между $\hat{\theta}$ и $\hat{\theta}$. По условию первого порядка $\frac{\partial Q_n(\bar{\theta})}{\partial \theta} = 0$. Также $\frac{\partial^2 Q_n(\bar{\theta})}{\partial \theta \partial \theta'} \rightarrow E[H(m; \theta_n)]$.

7. (По желанию доказательства (7.4.27) для GMM.) Докажите (7.4.27) для GMM. Укажите: Выведите разложение Тейлора из разложения в среднем значения $g_n(\hat{\theta})$ в окрестности $\hat{\theta}$

$$g_n(\bar{\theta}) = g_n(\hat{\theta}) + G_n(\bar{\theta})(\bar{\theta} - \hat{\theta}) + o_p. \quad (7.4.33)$$

Подставьте это выражение в $Q_n(\hat{\theta}) = -\frac{1}{2} g_n(\hat{\theta})' \bar{W} g_n(\hat{\theta})$. После этого используйте условия первого порядка $E[g_n(\hat{\theta})' \bar{W} g_n(\hat{\theta})] = 0$.

7.6. Численная оптимизация

До этого момента нас не интересовали вычислительные аспекты асимптотической оценки. В ML-оценках модели линейной регрессии, например, целевая функция $Q_n(\theta)$ квадратична по θ , так что существует явная формула для оптимального значения параметра — оценки ML. В линейном GMM целевая функция будет также линейной, и оценка линейного GMM может быть выписана в явном виде. В большинстве других случаев, однако, нет таких явных формул и необходимо применить какой-нибудь численный алгоритм, чтобы найти максимум. Существует несколько алгоритмов, но в этом параграфе будут рассмотрены только два самых важных. Больше деталей можно найти, например, в: [Judge et al., 1985, Appendix B]. Первый алгоритм используется для расчета M-оценок, в то время как второй используется для GMM.

Алгоритм Ньютона — Раффани

Сначала мы рассмотрим M -оценку, целевой функцией $Q_n(\theta)$ которой является (7.1.2). Так как $Q_n(\theta)$ дважды непрерывно дифференцируема на M -оценке, для нее существует разложение Тейлора второго порядка

$$Q_n(\theta) \approx Q_n(\tilde{\theta}_j) + s_n(\tilde{\theta}_j)(\theta - \tilde{\theta}_j) + \frac{1}{2}(\theta - \tilde{\theta}_j)' H_n(\tilde{\theta}_j)(\theta - \tilde{\theta}_j), \quad (7.5.1)$$

где $\tilde{\theta}_j$ — оценка на j -м шаге итерационной процедуры, которая вскоре будет достигнута, а s_n и H_n — градиент и гессиан целевой функции:

$$s_n(\theta) \equiv \frac{\partial Q_n(\theta)}{\partial \theta}, \quad H_n(\theta) \equiv \frac{\partial^2 Q_n(\theta)}{\partial \theta \partial \theta'}. \quad (7.5.2)$$

Оценка $(j+1)$ -го шага $\tilde{\theta}_{j+1}$ максимизирует квадратичную функцию в правой части (7.5.1). Она задается формулой

$$\tilde{\theta}_{j+1} = \tilde{\theta}_j - [H_n(\tilde{\theta}_j)]^{-1} s_n(\tilde{\theta}_j). \quad (7.5.3)$$

Эта итерационная процедура называется **алгоритмом Ньютона — Раффани** (*Newton — Raftery algorithm*). Если целевая функция вогнута, то этот алгоритм часто быстро сходится к глобальному максимуму.

Для оценок ML, когда глобальный максимум $\hat{\theta}$ достигается таким образом, оценка $\Delta \ln \ell(\hat{\theta})$ получается как $-H_n(\hat{\theta})^{-1}$ (эквивалентно $H_n(\hat{\theta}) \equiv \frac{1}{n} \sum_{i=1}^n H(\theta_i; \hat{\theta})$).

Алгоритм Гарриса — Ньютона

Теперь обратимся к GMM, у которого целевой функцией является $Q_n(\theta) = -\frac{1}{2} g_n'(\theta) \hat{V} g_n(\theta)$ (см. (7.1.13)). Как и при методе асимптотически оптимального разложения, мы будем работать с линейной целевой функцией $g_n(\theta)$. Разложение Тейлора первого порядка для $g_n(\theta)$ около $\tilde{\theta}_j$ имеет следующий вид

$$\begin{aligned} g_n(\theta) &\approx g_n(\tilde{\theta}_j) + G_n(\tilde{\theta}_j)(\theta - \tilde{\theta}_j) \\ &= [g_n(\tilde{\theta}_j) + G_n(\tilde{\theta}_j)\tilde{\theta}_j] - [-G_n(\tilde{\theta}_j)]\tilde{\theta}_j \\ &= v_j - G_j\theta. \end{aligned} \quad (7.5.4)$$

где

$$v_j = g_n(\tilde{\theta}_j) + G_n(\tilde{\theta}_j)\tilde{\theta}_j, \quad G_j \equiv -G_n(\tilde{\theta}_j). \quad (7.5.5)$$

(Вспомните, что $G_n(\theta) \equiv \frac{\partial g_n(\theta)}{\partial \theta}$.) Если бы функция $g_n(\theta)$ в выражении для целевой функции GMM была этой линейной функцией $v_j -$

$G_j \theta$, т.е. тогда целевая функция была бы квадратичной по θ и значение, максимизирующее целевую функцию (или минимизирующее расстояние GMM), было бы линейной оценкой GMM:

$$\hat{\theta}_{j+1} = (G_j' \tilde{W} G_j)^{-1} G_j' \tilde{W} y_j. \quad (7.5.6)$$

Это оценка $(j+1)$ -го шага в алгоритме Гаусса — Ньютона (*Gauss — Newton algorithm*)¹. В отличие от алгоритма Гьютона — Рафсона, здесь нет необходимости вычислять вторые производные.

**Запись алгоритмов Ньютона — Рафсона
и Гаусса — Ньютона в общем формате**

В алгоритмах Гаусса — Ньютона и Ньютона — Рафсона имеются и общие черты. Подставляя (7.5.5) в (7.5.6) и проводя перестановки, мы получаем:

$$\hat{\theta}_{j+1} = \hat{\theta}_j - [-G_n(\hat{\theta}_j)' \tilde{W} G_n(\hat{\theta}_j)]^{-1} [-G_n(\hat{\theta}_j)' \tilde{W} g_n(\hat{\theta}_j)]. \quad (7.5.7)$$

Выражение во вторых скобках есть не что иное, как градиент, вычисленный в $\hat{\theta}_j$ для целевой функции GMM $Q_n(\theta) = -\frac{1}{2} g_n(\theta)' \tilde{W} g_n(\theta)$. Роль, которую играет гессиан в целевой функции в (7.5.3), здесь играет выражение в первых скобках, которое является оценкой $-G'WG$ в табл. 7.1, вычисленной в $\hat{\theta}_j$. Таким образом, аналогия между M-оценками и оценками GMM, описанная в табл. 7.1, наблюдается также и здесь. Более того, если g_n линейна, то эквивалент гессиана (выражение в первых скобках в (7.5.7)) и есть гессиан целевой функции GMM. Так что аналогия точная: алгоритм Гаусса — Ньютона совпадает с алгоритмом Ньютона — Рафсона.

Уравнения, нелинейные только по параметрам

В алгоритме Гаусса — Ньютона (7.5.7), тогда, когда $g_n(\theta) (\equiv \frac{1}{n} \sum g(w_i; \theta))$ нелинейны по θ , в общем случае необходимо вычислять средние по наблюдениям, чтобы вычислить градиент и эквивалент гессиана на каждой итерации. (Это также верно в общем случае для алгоритма Ньютона — Рафсона, если целевая функция не квадратичная по θ .) Для больших массивов данных это операция, требующая больших затрат времени работы центрального процессора. Есть, однако, один случай, когда усреднение по наблюдениям не требуется. Это верно для класса обобщенных нелинейных моделей инструментальных переменных (частный случай нелинейного GMM, в котором $g(y, x, \theta)$ можно представить

¹ Алгоритм Гаусса — Ньютона был изначально разработан для нелинейного метода наименьших квадратов. По поводу того, как алгоритм Гаусса — Ньютона применяется к M-оценкам, контрольный вопрос 2. Представленный здесь алгоритм является, таким образом, адаптацией для GMM.

как $\sigma(y_i, z_i; \theta; x_i)$. Предположим, что выражение для $\sigma(y_i, z_i; \theta)$ имеет следующий вид:

$$\sigma(y_i, z_i; \theta) = \sigma_0(y_i, z_i) + \alpha_1(y_i, z_i)' \alpha(\theta), \quad (7.5.8)$$

$(1 \times q); \quad (q \times 1)$

где $\sigma_0(\cdot)$ и $\alpha(\cdot)$ — известные функции от (y_i, z_i) (но не функции от θ). Функция такого вида называется **нелинейной только по параметрам** (*nonlinear in parameters only*), или **линейной по переменным** (*linear in variables*). Можно легко показать, что

$$\begin{aligned} g_n(\theta) &= \frac{1}{n} \sum_{i=1}^n g(y_i, z_i; \theta) = \frac{1}{n} \sum_{i=1}^n \sigma(y_i, z_i; \theta) \cdot x_i = \\ &= \left(\frac{1}{n} \sum_{i=1}^n \sigma_0(y_i, z_i) \cdot x_i \right) + \left(\frac{1}{n} \sum_{i=1}^n x_i \alpha_1(y_i, z_i)' \right) \alpha(\theta), \end{aligned} \quad (7.5.9)$$

$(K \times 1) \quad (K \times q)$

$$G_n(\theta) = \frac{\partial g_n(\theta)}{\partial \theta'} = \left(\frac{1}{n} \sum_{i=1}^n x_i \alpha_1(y_i, z_i)' \right) \frac{\partial \alpha(\theta)}{\partial \theta'}. \quad (7.5.10)$$

$(K \times q) \quad (q \times q)$

Из этих уравнений становится понятно, что выборочные средние нужно рассчитывать только один раз до начала итераций.

Контрольные вопросы

1. (Возрастает ли Q_n во время итераций?) Положим $\theta = \hat{\theta}_{j-1}$ в (7.5.1) и перепишем это выражение в виде:

$$\begin{aligned} Q_n(\hat{\theta}_{j+1}) - Q_n(\hat{\theta}_j) &\cong x_n(\hat{\theta}_j)' (\hat{\theta}_{j+1} - \hat{\theta}_j) + \\ &- \frac{1}{2} (\hat{\theta}_{j+1} - \hat{\theta}_j)' H_n(\bar{\theta}_j) (\hat{\theta}_{j+1} - \hat{\theta}_j). \end{aligned}$$

Рассмотрим алгоритм Ньютона — Рафсона (7.5.3). Предположим, что $\hat{\theta}_{j+1}$ достаточно близка к $\hat{\theta}_j$, чтобы знак $Q_n(\hat{\theta}_{j+1}) - Q_n(\hat{\theta}_j)$ был таким же, как знак правой части записанного уравнения. Покажите, что $Q_n(\hat{\theta}_{j+1}) \geq Q_n(\hat{\theta}_j)$, если $Q_n(\theta)$ вогнута. **Указание:** Используя (7.5.11), можно переписать правую часть уравнения как

$$\frac{1}{2} (\hat{\theta}_{j+1} - \hat{\theta}_j)' H_n(\bar{\theta}_j) (\hat{\theta}_{j+1} - \hat{\theta}_j).$$

2. (Алгоритм Гаусса — Ньютона для NLS.) В NLS (недифференциальном методе наименьших квадратов) целевая функция задается соотношением (7.1.19).

Пусть $\hat{\theta}_j$ — оценка на j -м шаге. Рассмотрим линейную аппроксимацию

$$\begin{aligned}\varphi(x_i; \theta) &\approx \varphi(x_i; \hat{\theta}_j) + \left(\frac{\partial \varphi(x_i; \hat{\theta}_j)}{\partial \theta} \right)'_{(n \times 1)} (\theta - \hat{\theta}_j) = \\ &= \left[\varphi(x_i; \hat{\theta}_j) + \left(\frac{\partial \varphi(x_i; \hat{\theta}_j)}{\partial \theta} \right)' \hat{\theta}_j \right] - \left(\frac{\partial \varphi(x_i; \hat{\theta}_j)}{\partial \theta} \right)' \theta. \quad (7.5.11)\end{aligned}$$

Замените $\varphi(x_i; \theta)$ в функции l на эту линейную по θ функцию. Покажите, что минимум l этой линейной задачи достигается при

$$\begin{aligned}\hat{\theta}_{j+1} = \hat{\theta}_j + \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{\partial \varphi(x_i; \hat{\theta}_j)}{\partial \theta} \right) \left(\frac{\partial \varphi(x_i; \hat{\theta}_j)}{\partial \theta} \right)' \right]^{-1} \times \\ \times \left\{ \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial \varphi(x_i; \hat{\theta}_j)}{\partial \theta} \right)' [y_i - \varphi(x_i; \hat{\theta}_j)] \right\}. \quad (7.5.12)\end{aligned}$$

Набор задач для главы 7

Аналитические упражнения

(Информационное неравенство Кульбака — Лейблера.)

Пусть $f(y|x; \theta)$ — параметрическое семейство гипотетических функций условной плотности вероятности, с истинной функцией плотности является $f(y|x; \theta_0)$. Предположим, что $E[\ln f(y|x; \theta)]$ существует и конечно для всех θ . Информационное неравенство Кульбака — Лейблера утверждает, что

$$\begin{aligned}\text{Prob}[f(y|x; \theta) \neq f(y|x; \theta_0)] > 0 \Rightarrow \\ \Rightarrow E[\ln f(y|x; \theta)] < E[\ln f(y|x; \theta_0)].\end{aligned}$$

Докажите это, предприняв следующие шаги. В доказательстве, исключительно ради простоты, предположите, что $f(y|x; \theta) > 0$ для всех $(y|x)$ и θ , так что можно логарифмировать $f(y|x; \theta)$.

(a) Предположите, что $\text{Prob}[f(y|x; \theta) \neq f(y|x; \theta_0)] > 0$. Пусть $w = (y, x)'$. Определим

$$a(w) = f(y|x; \theta) / f(y|x; \theta_0)$$

Проверьте, что $a(w) \neq 1$ с положительной вероятностью, так что $a(w)$ — случайная величина, не являющаяся константой.

- (b) Строгая версия **неравенства Йенсена (Jensen's inequality)** утверждает, что

если $c(x)$ — строго выпуклая функция, а x — случайная величина, отличная от постоянной, то $E[c(x)] < c(E(x))$.

Используя этот факт, покажите, что

$$E[\ln a(w)] < \ln \{E[a(w)]\}.$$

Указание: $\ln(x)$ — строго вогнутая функция.

- (c) Покажите, что $E[a(w)] = 1$. **Указание:** условное математическое ожидание $a(w)$ равно 1, потому что

$$\begin{aligned} E[a(w)|x] &= \int a(w)f(y|x;\theta_0)dy = \\ &= \int \frac{f(y|x;\theta)}{f(y|x;\theta_0)}f(y|x;\theta_0)dy = \\ &= \int f(y|x;\theta)dy = \\ &= 1. \end{aligned}$$

Последнее равенство выполняется, потому что $f(y|x;\theta)$ — функция плотности вероятности для фиксированных x и θ . Заметьте, что условное математическое ожидание берется по отношению к истинной условной функции плотности $f(y|x;\theta_0)$.

- (d) Наконец, покажите исходный результат.

2. (Равенство для информационных матриц) Для ML вектор оценок s иessian для наблюдения (y, x) могут быть записаны ($s \equiv (s, x')$) как

$$\begin{aligned} s(w; \theta) &= \frac{\partial \ln f(y|x;\theta)}{\partial \theta}, \\ H(w; \theta) &= \frac{\partial s(w; \theta)}{\partial \theta'} = \frac{\partial^2 \ln f(y|x;\theta)}{\partial \theta \partial \theta'}. \end{aligned}$$

Как обычно, для простоты предположите, что $f(y|x;\theta) > 0$ для всех (y, x) и θ , так что можно логарифмировать функцию плотности.

- (a) Предполагая, что интегрирование (то есть вычисление математических ожиданий) и дифференцирование можно менять местами, покажите, что $E[s(w; \theta_0)] = 0$. **Указание:** Так как

$f(y|x; \theta)$ — гистетическая функция плотности. Интеграл от нее равен единице:

$$\int f(y|x; \theta) dy = 1.$$

Продифференцируйте обе стороны этого уравнения по θ , а затем поменяйте порядок дифференцирования и интегрирования, чтобы получить тождество:

$$\int \eta(\theta; \theta) f(y|x; \theta) dy = \mathbf{0}_{(p \times 1)}.$$

Положите $\theta = \theta_0$. Получите $E[\eta(\theta; \theta_0)|x] = \mathbf{0}$.

(b) Докажите равенство для информационных матриц:

$$-E[H(\theta; \theta_0)] = E[\eta(\theta; \theta_0)\eta(\theta; \theta_0)'].$$

Указание: Дифференцируя обе стороны тождества в предыдущем указании и предполагая, что дифференцирование и интегрирование можно поменять местами, мы получаем:

$$\int \frac{\partial}{\partial \theta'} \underbrace{[s(\theta; \theta) f(y|x; \theta)]}_{(p \times 1)} dy = \mathbf{0}_{(p \times p)}.$$

Покажите, что подынтегральное выражение может быть записано в виде:

$$\frac{\partial}{\partial \theta'} [s(\theta; \theta) f(y|x; \theta)] = H(\theta; \theta) f(y|x; \theta) + s(\theta; \theta) s(\theta; \theta)' f(y|x; \theta).$$

3. (Группа статистик для модели линейной регрессии.) Рассмотрим модель линейной регрессии с нормальными ошибками, в которой условная плотность для наблюдения i задается соотношением:

$$f(y_i|x_i; \beta, \sigma^2) = \frac{1}{\sigma} \ln(2\pi) - \frac{1}{2\sigma^2} \ln(\sigma^2) - \frac{(y_i - x_i'\beta)^2}{2\sigma^2}.$$

Пусть $(\tilde{\beta}, \tilde{\sigma}^2)'$ — МЛ-оценка параметра $\theta = (\beta', \sigma^2)'$ для модели без ограничений, а $(\tilde{\beta}, \tilde{\sigma}^2)'$ — МЛ-оценка для модели с ограничениями $H\beta = c$, где H — матрица размера $r \times K$, состоящая из известных констант. Предположим, что $\theta \in \mathbb{R}^K \times \mathbb{R}_+$ и что $E(x_i x_i')$ невырождена. Пусть также

$$\tilde{\Sigma} = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_i x_i' & \mathbf{0} \\ \mathbf{0}' & \frac{1}{n\tilde{\sigma}^2} \end{bmatrix}, \quad \tilde{\Sigma} = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_i x_i' & \mathbf{0} \\ \mathbf{0}' & \frac{1}{n\tilde{\sigma}^2} \end{bmatrix}.$$

- (а) Проверьте, что $\hat{\beta}$ минимизирует сумму квадратов остатков, то есть является оценкой МНК. Проверьте, что $\tilde{\beta}$ минимизирует сумму квадратов остатков при условии выполнения ограничения $R\tilde{\beta} = c$, то есть является МНК-оценкой для модели с ограничениями.

- (б) Пусть $Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n \ln f(y_i; x_i; \beta, \sigma^2)$. Покажите, что

$$Q_n(\hat{\theta}) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} - \frac{1}{2} \ln \left(\frac{SSR_0}{n} \right),$$

$$Q_n(\tilde{\theta}) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} - \frac{1}{2} \ln \left(\frac{SSR_R}{n} \right),$$

где $SSR_0 \equiv \sum (y_i - x_i' \hat{\beta})^2$ — сумма квадратов остатков в модели без ограничений, а $SSR_R \equiv \sum (y_i - x_i' \tilde{\beta})^2$ — сумма квадратов остатков в модели с ограничениями. **Указание:** Покажите, что $\hat{\sigma}^2 = SSR_0/n$ и $\tilde{\sigma}^2 = SSR_R/n$.

- (в) Проверьте, что оценка $\hat{\Sigma}$, хотя и не равняется $-\frac{1}{n} H(w_0; \hat{\theta})$, является состоятельной оценкой для $-E[H(w_0; \theta_0)]$. Проверьте, что $\tilde{\Sigma}$, хотя и не равняется $-\frac{1}{n} H(w_0; \tilde{\theta})$, является состоятельной оценкой для $-E[H(w_0; \theta_0)]$. **Указание:** Но обсуждая в параграфе 7.8 вытекает, что $\tilde{\theta}$ состоятельна. Как отмечалось в параграфе 7.4, $\tilde{\theta}$ также состоятельна при выполнении нулевой гипотезы.

- (д) Покажите, используя $\hat{\Sigma}$ и $\tilde{\Sigma}$, формулы для которых приведены в табл. 7.2, что статистики Вайта, LM и LR могут быть записаны как

$$W = n \cdot \frac{(R\hat{\beta} - c)' [R(X'X)^{-1}R']^{-1} (R\hat{\beta} - c)}{SSR_0},$$

$$LM = n \cdot \frac{(y - X\tilde{\beta})' P(y - X\tilde{\beta})}{SSR_R},$$

$$LR = n \cdot \left\{ \ln \left(\frac{SSR_R}{n} \right) - \ln \left(\frac{SSR_0}{n} \right) \right\},$$

где $g(c \times 1)$ и $X(n \times K) =$ — вектор наблюдений и матрица, для наблюдений y_i и x_i , а $P = X(X'X)^{-1}X'$.

Указание: $A(\theta)(r \times (K+1))$ в табл. 7.2 равно:

$$\begin{pmatrix} R & : & 0 \\ 0 & : & I_{r \times r} \end{pmatrix}.$$

- (в) Покажите, что эти три статистики можно записать также следующим образом:

$$\begin{aligned} W &= n \cdot \frac{SSR_R - SSR_U}{SSR_U}, \\ LM &= n \cdot \frac{SSR_R - SSR_U}{SSR_R}, \\ LR &= n \cdot \ln \left(\frac{SSR_R}{SSR_U} \right). \end{aligned}$$

Указание: Как мы показали в аналитическом упражнении к главе 1,

$$\begin{aligned} SSR_R - SSR_U &= (\tilde{\beta} - \hat{\beta})'(X'X)(\tilde{\beta} - \hat{\beta}) = \\ &= (R\tilde{\beta} - c)'[R(X'X)^{-1}R']^{-1}(R\tilde{\beta} - c) = \\ &= (y - X\tilde{\beta})'P(y - X\tilde{\beta}). \end{aligned}$$

- (п) Покажите, что $W \geq LR \geq LM$. (Эти неравенства не всегда выполняются в моделях нелинейной регрессии.)

Ответы на избранные вопросы

- 2а. Так как $f(y|x;\theta)$ — гипотетическая функция плотности вероятности, интеграл от нее равен единице:

$$\int f(y|x;\theta)dy = 1. \quad (1)$$

Это тождество, верное для любого значения $\theta \in \Theta$. Дифференцируя обе части этого тождества по θ , мы получим:

$$\frac{\partial}{\partial \theta} \int f(y|x;\theta)dy = \mathbf{0}_{(p \times 1)}. \quad (2)$$

Если операции дифференцирования и интегрирования можно менять местами, то

$$\frac{\partial}{\partial \theta} \int f(y|x;\theta)dy = \int \frac{\partial}{\partial \theta} f(y|x;\theta)dy. \quad (3)$$

По определению вклада, $s(w;\theta)f(y|x;\theta) = \frac{\partial}{\partial \theta} f(y|x;\theta)$. Подставляя это в (3), получаем:

$$\int s(w;\theta)f(y|x;\theta)dy = \mathbf{0}_{(p \times 1)}. \quad (4)$$

Это верно для любого значения $\theta \in \Theta$, в частности для θ_0 .
 Полагая $\theta = \theta_0$, мы получим:

$$\int \psi(w; \theta_0) f(y|x; \theta_0) dy = E[\psi(w; \theta_0)'x] = 0. \quad (5)$$

Затем, используя закон повторных математических ожиданий, мы получаем следующий результат.

Литература

- Arellano, T., 1985, *Advanced Econometrics*, Cambridge: Harvard University Press.
- Daytonson, R., and J. MacKinnon, 1983, *Estimation and Inference in Econometrics*, New York: Ox and University Press.
- Gallant, R., and H. White, 1988, *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, New York: Basic Blackwell.
- Gunteroux, C., and A. Monfort, 1994, *Statistics and Econometric Models*, New York: Cambridge University Press.
- Hansen, L. E., and K. Singleton, 1982, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1209–1286.
- Judge, G., W. Griffiths, R. Hill, D. Lutkepohl, and T. Lee, 1983, *The Theory and Practice of Econometrics* (2d ed.), New York: Wiley.
- Newey, W., and D. McFadden, 1994, "Large Sample Estimation and Hypothesis Testing" Chapter 36 in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Volume IV, New York: North-Holland.
- White, H., 1994, *Estimation, Inference, and Specification Analysis*, New York: Cambridge University Press.

Глава 8. Примеры применения метода максимального правдоподобия

Метод максимального правдоподобия (МП) был изложен в предыдущей главе как частный случай экстремальных оценок. Вследствие его важности в эконометрике в этой главе рассматривается применение МП к некоторым известным моделям. Мы уже упоминали о некоторых из этих моделей в главах 3 и 4 в контексте оценивания обобщенным методом моментов.

При выводе асимптотических свойств МП-оценок моделей, рассматриваемых здесь, будут использованы общие результаты для МП, полученные в главе 7. Перед чтением этой главы нам следует просмотреть параграфы 7.1, предположения 7.5, 7.6, 7.8, 7.9 и 7.11.

Замечания по обозначениям: Мы сохраним обозначения, введенные в предыдущей главе: истинное значение параметра обозначается индексом 0. Так что θ — гипотетическое значение параметра, θ_0 — истинное значение параметра.

8.1. Модели с качественным откликом

Во многих ситуациях в экономике и в других науках зависимая переменная представляет собой категоризованную переменную. Например, человек может ходить или не ходить в состав рабочей силы; пассажир выбирает тот или иной вид транспорта; пациент может умереть или выжить и так далее. Регрессионные модели, в которых зависимая переменная принимает дискретные значения, называются **моделями с качественным откликом** (*qualitative response*) — (QR). Модель с качественным откликом описывается параметризованной функцией плотности $f(y, x; \theta)$, представляющей семейство дискретных распределений вероятностей y при условии x , индексированных параметром θ .

QR-модель называется **моделью с бинарным выбором** (*binary response model*), если зависимая переменная может принимать только два значения (их можно считать равными 0 и 1 без потери общности), и называется **моделью с множественным выбором** (*multinomial response model*), если зависимая переменная принимает более двух значений.

В этом подразделе рассматриваются только модели с бинарным выбором. По поводу более подробного обсуждения моделей с бинарными и множественными откликами см. [Алленгуэ, 1985, Chapter 9].

Наиболее популярной моделью с бинарным откликом является пробит-модель, которую уже была представлена в примере 7.3. Другой популярной бинарной моделью является логит-модель (*logit model*):

$$\begin{cases} f(y = 1 | x; \theta_0) = A(x'\theta_0), \\ f(y = 0 | x; \theta_0) = 1 - A(x'\theta_0), \end{cases} \quad (8.1.1)$$

где A — функция логистического распределения

$$A(v) \equiv \frac{\exp(v)}{1 + \exp(v)}. \quad (8.1.2)$$

При этом функция плотности $f(y|x; \theta)$ может быть записана как

$$f(y|x; \theta_0) = A(x'\theta_0)^y \times [1 - A(x'\theta_0)]^{1-y}. \quad (8.1.3)$$

Взяв логарифм от обеих частей этого выражения и заменив истинное значение параметра θ_0 его гипотетическим значением θ , мы получаем логарифмическое правдоподобие для наблюдения i :

$$\log f(y_i | x_i; \theta) = y_i \log A(x_i'\theta) + (1 - y_i) \log [1 - A(x_i'\theta)] \quad (8.1.4)$$

Целевая функция логит-модели $Q_n(\theta)$ равна умноженному на $1/n$ логарифмическому правдоподобию выборки $(y_1, x_1, y_2, x_2, \dots, y_n, x_n)$. В предположении, что $\{y_i, x_i\} = i.i.d.$, логарифмическое правдоподобие выборки представляет собой сумму из n логарифмов правдоподобия для наблюдения i . Следовательно,

$$Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n \{y_i \log A(x_i'\theta) + (1 - y_i) \log [1 - A(x_i'\theta)]\}. \quad (8.1.5)$$

Ранее мы показали, что для пробит-модели функция правдоподобия выпукла (пример 7.6) и что ML-оценки состоятельны (пример 7.9) и асимптотически нормальны (пример 7.11), а предположим, что $F(x, x')$ — невырожденная матрица. Вы вскоре увидите, что вывод этих свойств для логит-модели проще (Однако при первом применении вы, возможно, захотите пропустить обсуждение состоятельности и асимптотической нормальности оценок.)

Вклад и гессин для наблюдений i

Функция логистического распределения объясняет следующим удобным свойством:

$$A'(v) = A(v)(1 - A(v)), \quad A''(v) = -2A(v)[1 - A(v)]. \quad (8.1.6)$$

Используя его, несложно вывести (см. контрольный вопрос 1(а)) следующие выражения для функции вклада и гессиана¹ для наблюдения i :

$$j(w_i; \theta) \left(= \frac{\partial \log f(y_i | x_i; \theta)}{\partial \theta} \right) = [y_i - A(x_i' \theta)] x_i, \quad (8.1.7)$$

$$H(w_i; \theta) \left(= \frac{\partial^2 j(w_i; \theta)}{\partial \theta^2} \right) = -A(x_i' \theta)[1 - A(x_i' \theta)] x_i x_i'. \quad (8.1.8)$$

где $w_i = (y_i, x_i')$.

Состоятельность

Поскольку $x_i x_i'$ — положительно полуопределенная матрица, из выражения для $H(w_i; \theta)$ немедленно вытекает отрицательная полуопределенность гессиана, и поэтому $Q_n(\theta)$ вогнута. Следовательно, подходящей теоремой о состоятельности является утверждение 7.6. Мы покажем здесь, что условия этого утверждения выполняются при невырожденности $E(x_i x_i')$. Условие (а) этого утверждения представляет собой идентификацию (условной) функции плотности: $f(y_i | x_i; \theta) \neq f(y_i | x_i; \theta_0)$ с положительной вероятностью для $\theta \neq \theta_0$ ². Поскольку взятие логарифма является строго монотонным преобразованием, это условие эквивалентно

$$\log f(y_i; x_i; \theta) \neq \log f(y_i | x_i; \theta_0) \\ \text{с положительной вероятностью для } \theta \neq \theta_0. \quad (8.1.9)$$

где θ_0 — истинное значение параметра, и функция плотности f дана в (8.1.3). Проверка этого условия аналогична случаю лоббит-модели, рассмотренному в примере 7.8, и производится следующим образом. При обсуждении примера 7.8 было установлено, что

$$E(x_i x_i') \text{ невырождена} \rightarrow x_i' \theta \neq x_i' \theta_0 \\ \text{с положительной вероятностью для } \theta \neq \theta_0. \quad (8.1.10)$$

Поскольку логистическая функция распределения $A(\cdot)$ — строго монотонная функция от $\cdot$, мы имеем $A(x_i' \theta) \neq A(x_i' \theta_0)$, когда $x_i' \theta \neq x_i' \theta_0$. Таким образом, из (а) следует из невырожденности $E(x_i x_i')$. Переходя к условию (б) (покажем, что $E[\log f(y_i; x_i; \theta)] < \infty$ для всех θ), легко показать, что

$$|(\log A(\cdot))_1| \leq \log A(0) + v. \quad (8.1.11)$$

¹ Напомним, что гессианов и их элементы являются вторыми производными. — Прим. ред. переводс.

² Другими словами, пусть $X \equiv f(y_i; x_i; \theta)$ и $Y \equiv f(y_i; x_i; \theta_0)$. X и Y — случайные величины, так как они являются функциями от $w_i = (y_i, x_i')$. Условие может быть переформулировано как $\text{Prob}(X \neq Y) > 0$ для $\theta \neq \theta_0$.

Условие (1) можно проверять аналогично проверке такого же условия для пробит-в примере 7.9. Таким образом, как и в пробит-модели, невырожденность $E(x_i x_i')$ обеспечивает состоятельность оценок МЛ в логит-модели.

Асимптотическая нормальность

Чтобы доказать асимптотическую нормальность при том же условии, мы проверим выполнение условий утверждения 7.9. Условие (1) этого утверждения выполняется, если в качестве пространства параметров Θ берется R^k . Условие (2) выполняется очевидным образом. Условие (3) легко можно проверить следующим образом. Так как $E[y_i | x_i] = A(x_i' \theta)$, то из (8.1.7) следует $E[s(w_i; \theta_0) | x_i] = 0$. А значит, $E[s(w_i; \theta_0)] = 0$, по закону полных математических ожиданий². Читателю остается в качестве контрольного вопроса (6) вывести равенство для условных информационных матриц

$$E[s(w_i; \theta_0) s(w_i; \theta_0)' | x_i] = -E[H(w_i; \theta_0) | x_i]. \quad (8.1.12)$$

Следовательно, (3) выполняется по закону полных математических ожиданий³. Условие локального доминирования (условие (4)) для логит-модели проверить легко, так как поскольку $|A(x_i' \theta)| = |1 - A(x_i' \theta)| \leq 1$, то $|H(w_i; \theta)| \leq |x_i x_i'|$ для всех θ . Математическое ожидание $\|x_i x_i'\|^2$ (суммы квадратов элементов $\tilde{x}_i \tilde{x}_i'$) конечно, если $E(x_i x_i')$ — невырожденная матрица. Поэтому корректно $E(|x_i x_i'|)$. Что касается условия (5), то аргумент в скобке 29 Ньюи и Мак-Фаддена [Newey and McFadden, 1994]), использовавшийся для проверки того же условия в пробит-модели, также может быть использован и для логит-модели⁴.

Таким образом, мы приходим к выводу, что если $\{y_i, x_i\} \sim \text{i.i.d.}$ и $E(x_i x_i')$ — невырожденная матрица, то тогда в логит-модели МЛ-оценка $\hat{\theta}$ истинного значения параметра θ_0 является состоятельной и асимптотически нормальной с асимптотической ковариационной матрицей, задаваемой выражением:

$$\text{Avar}(\hat{\theta}) = \left[\frac{1}{n} \sum_{i=1}^n H(w_i; \hat{\theta}) \right]^{-1}, \quad (8.1.13)$$

где $w_i = (y_i, x_i')$, и $H(w_i; \theta)$ задается в (8.1.8)).

² Мы приведем доказательство только для заинтересованных читателей. Для любого $\varepsilon > 0$, существует C , такое, что $|A(c)| - A(c)| \geq C > 0$ для $|c| \leq \varepsilon$. Поэтому

$$E[|A(x' \theta_0) - A(x' \hat{\theta})| x x'] \geq E[|1 - A(x' \hat{\theta})| - A(x' \hat{\theta})| x x'] \geq \varepsilon E[|1 - A(x' \hat{\theta})| x x']$$

и все вышеописанное — матричные неравенства. $|c| \leq \varepsilon$ — индикаторная функция. Последний член правой части определен (и не отрицателен) и не отрицателен для достаточно больших n вследствие невырожденности $E(x x')$.

Контрольные вопросы

1. (Функция оценок и гессиан для наблюдения i для функции плотности общего вида.) Заменяя вероятностную функцию распределения $A(x_i; \theta_0)$ в (6.1.1) на произвольную функцию распределения $f(x_i; \theta_1)$,

(а) Проверьте, что функции оценок и гессиан для наблюдения i задаются следующими выражениями:

$$a(x_i; \theta) = \frac{\partial}{\partial \theta} \log f(x_i; \theta) = \frac{f'(x_i; \theta)}{f(x_i; \theta)},$$

$$H(x_i; \theta) = - \left[\frac{f''(x_i; \theta)}{f(x_i; \theta)} - \left(\frac{f'(x_i; \theta)}{f(x_i; \theta)} \right)^2 \right] = - \left[\frac{f''(x_i; \theta)}{f(x_i; \theta)} - \left(\frac{f'(x_i; \theta)}{f(x_i; \theta)} \right)^2 \right],$$

где $F = F(x; \theta_0)$, $f = f(x; \theta)$, $f' = f'(x; \theta)$, $f'' = f''(x; \theta)$ — функция плотности.

(б) Проверьте равенство для элементов информационных матриц (8.1.12) для функции распределения F общего вида. Указание: Покажите, что математическое ожидание вектора матрицы информации для гессиана в предыдущем параграфе равно нулю.

2. Проверьте (6.1.7) и (8.1.8).

3. (Логит-модели для бинарных коррелированных наблюдений.) Предположим, что последовательность $\{y_i, x_i\}$ является стационарной и эргодической, но небылительной iid. Достаточно ли ML-оценка для логит-модели, описанная в тексте? [Ответ: Да. Так как утверждение 7.6, применимое в данном случае, не требует условия независимости. Однако выражение для $\text{Asym}(\hat{\theta})$ нельзя получить как обратную со знаком «минус» матрицу, обратную к выборочному гессиану.]

8.2. Усеченные регрессионные модели

Модель с ограниченной зависимой переменной (LDV) (limited dependent variable model) представляет собой регрессионную модель, в которой одна зависимая переменная y_i ограничена каким-либо образом, как наблюдений для которых y_i не удовлетворяет некоторому предопределенному критерию, исключаются из выборки LDV-модель первого типа называется **цензурированной моделью регрессии (censored regression model)**, а модель второго типа — **усеченной моделью регрессии (truncated regression model)**. В этом параграфе представлены усеченные регрессионные модели; цензурированные модели регрессии рассматриваются в следующем параграфе.

Модель

Пусть $\{y_i, x_i\}$ iid и

$$y_i = x_i' \beta_0 + \varepsilon_i, \quad \varepsilon_i | x_i \sim N(0, \sigma_0^2). \quad (8.2.1)$$

Эта модель была бы в точности стандартной линейной регрессионной моделью с нормально распределенными ошибками, если бы не следующая особенность: в выборку включены только те наблюдения, для которых y_i удовлетворяет некоторому заранее установленному критерию. Мы рассмотрим только простейшее правило усечения: $y_i > c$, где c — известная константа. Это правило часто называют «усечением снизу». После рассмотрения этого частного случая несложно обобщить его на более общие правила усечения.

Усеченные распределения

Для проведения моделирования нам необходимы два результата из теории вероятностей.

Функция плотности усеченной случайной величины: Если непрерывная случайная величина y имеет функцию плотности $f(y)$ и c — некоторая константа, то функция плотности после усечения для $y > c$ определена на интервале (c, ∞) и задается следующим образом:

$$f(y|y > c) = \frac{f(y)}{\text{Prob}\{y > c\}} \quad (8.2.2)$$

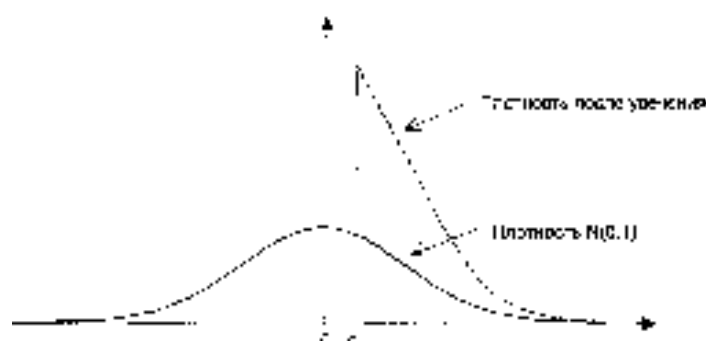


Рис. 8.1. Выпуклые усечения

На рис. 8.1 показано, как усечение влияет на представление для $N(0,1)$. Сплошной линией изображена в точности вероятность $\Phi(0,1)$. Поскольку эта площадь известна, площадь под этой кривой равна единице. Пунктирная кривая изображает усеченную функцию плотности на интервале (c, ∞) . Она находится выше сплошной кривой, так что площадь под пунктирной кривой также равна единице.

Далее приведен второй результат, который будет полезен в дальнейшем.

Моменты усеченного нормального распределения:

Если $y \sim N(\mu_0, \sigma_0^2)$ и c — константа, то математическое ожидание и дисперсия усеченного распределения равны соответственно

$$E(y|y > c) = \mu_0 + \sigma_0 \lambda(v), \quad (8.2.3)$$

$$\text{Var}(y|y > c) = \sigma_0^2 \{1 - \lambda(v)[\lambda(v) - v]\}, \quad (8.2.4)$$

где $v = (c - \mu_0)/\sigma_0$ и $\lambda(v) \equiv \frac{\phi(v)}{1 - \Phi(v)}$.

Эта формула для математического ожидания ясно показывает, что выборочное среднее выборки из усеченного распределения не состоятельно в отношении μ_0 . Это пример **смещения, обусловленного селективностью выборки** (*sample selection bias*). Это смещение равно $\sigma_0 \lambda(v)$. Функция $\lambda(v)$ называется **обратным отношением Миллса** (*inverse Mill's ratio*), или **функцией риска** (*hazard function*). Функция $\lambda(v)$ выпукла и имеет асимптоты v при $v \rightarrow \infty$ и нуль при $v \rightarrow -\infty$. Поэтому ее производная $\lambda'(v)$ лежит в пределах от 0 до 1. Легко показать, что

$$\lambda'(v) = \lambda(v)[\lambda(v) - v]. \quad (8.2.5)$$

Используя (8.2.3) и (8.2.4), мы можем продемонстрировать, как усечение изменяет форму регрессии y_t на x_t . Из (8.2.1) следует, что распределением y_t при условии x_t до усечения выборки является $N(x_t'\beta_0, \sigma_0^2)$. Наблюдение t присутствует в выборке, если и только если $y_t > c$. Поэтому

$$E(y_t|x_t, t \text{ в выборке}) = x_t'\beta_0 + \sigma_0 \lambda\left(\frac{c - x_t'\beta_0}{\sigma_0}\right), \quad (8.2.6)$$

$$\begin{aligned} \text{Var}(y_t|x_t, t \text{ в выборке}) &= \\ &= \sigma_0^2 \left\{ 1 - \lambda\left(\frac{c - x_t'\beta_0}{\sigma_0}\right) \left[\lambda\left(\frac{c - x_t'\beta_0}{\sigma_0}\right) - \frac{c - x_t'\beta_0}{\sigma_0} \right] \right\}. \end{aligned} \quad (8.2.7)$$

Эта формула показывает, что OLS-оценка коэффициента при x_t из линейной регрессии y_t на x_t не состоятельна, так как дополнительная составляющая $\sigma_0 \lambda\left(\frac{c - x_t'\beta_0}{\sigma_0}\right)$ (которая коррелирована с x_t) будет включена в ошибку линейной регрессии. Поскольку функциональная форма функций риска $\lambda(v)$ известна (при предположении о нормальности ε_t), мы можем избежать смещения, обусловленного селективностью выборки, применяя нелинейный метод наименьших квадратов для оценивания параметров (β_0, σ_0^2) , однако ML оценка более предпочтительна, так как метод максимального правдоподобия является асимптотически более эффективным.

Заметим попутно, что проблема смещения, обусловленного селективностью выборки, не возникает, если отбор основан на значениях регрессоров, а не на значениях зависимой переменной. Предположим,

что наблюдение t включается в выборку, если $\phi(x_t) \geq 0$. Тогда

$$\begin{aligned} E(y_t | x_t, t \text{ в выборке}) &= E(x_t' \beta_1 + \varepsilon_t | x_t, \phi(x_t) \geq 0) = \\ &= x_t' \beta_1 + E(\varepsilon_t | x_t, \phi(x_t) \geq 0) = \\ &= x_t' \beta_1. \end{aligned} \quad (8.2.8)$$

Последнее равенство выполнено, так как $E(\varepsilon_t | x_t, \phi(x_t) \geq 0) = E(\varepsilon_t | x_t) = 0$, если $\phi(x_t) \geq 0$.

Функция правдоподобия

При нахождении функции правдоподобия используется уравнение (8.2.2). Как мы только что заметили, распределение $y_t | x_t$ до усечения — это нормальное распределение $N(x_t' \beta_0, \sigma_0^2)$ с функцией плотности

$$\frac{1}{\sqrt{2\pi}\sigma_0} \exp \left\{ -\frac{1}{2} \left(\frac{y_t - x_t' \beta_0}{\sigma_0} \right)^2 \right\} = \frac{1}{\sigma_0} \varphi \left(\frac{y_t - x_t' \beta_0}{\sigma_0} \right), \quad (8.2.9)$$

где $\varphi(\cdot)$ — функция плотности стандартного нормального распределения $N(0, 1)$. Вероятность того, что наблюдение t представлено в выборке, равна

$$\begin{aligned} \text{Prob}(y_t > c | x_t) &= 1 - \text{Prob}(y_t \leq c | x_t) = \\ &= 1 - \text{Prob} \left(\frac{y_t - x_t' \beta_0}{\sigma_0} \leq \frac{c - x_t' \beta_0}{\sigma_0} \middle| x_t \right) = \\ &= 1 - \Phi \left(\frac{c - x_t' \beta_0}{\sigma_0} \right) \quad \text{поскольку } \varphi = \frac{d}{du} \Phi \quad x_t \sim N(0, 1). \end{aligned} \quad (8.2.10)$$

Поэтому из (8.2.2) следует, что функция плотности после усечения величин, распределенных независимо $(\varepsilon, \varepsilon')$, равна

$$f(y | x_t, t \text{ в выборке}; \beta, \sigma^2) = \frac{1}{\sigma} \varphi \left(\frac{y - x_t' \beta}{\sigma} \right) \cdot \frac{1}{1 - \Phi \left(\frac{c - x_t' \beta}{\sigma} \right)}. \quad (8.2.11)$$

Эта функция плотности (условия по отношению к x_t) для наблюдения t *присуща выборке* и *выборке*. Прогнозируемая величина (β_0, σ_0^2) на их гипотетические значения (β, σ^2) , следовательно, дифференциальное условие правдоподобия для наблюдений t ,

$$\begin{aligned} \log f(y_t | x_t; \beta, \sigma^2) &= \\ &= \left\{ -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma^2) - \frac{1}{2} \left(\frac{y_t - x_t' \beta}{\sigma} \right)^2 \right\} - \\ &- \log \left[1 - \Phi \left(\frac{c - x_t' \beta}{\sigma} \right) \right]. \end{aligned} \quad (8.2.12)$$

Если бы не последняя составляющая, то выражение численно бы аналогической функцией правдоподобия (при условии x_i) для обычной линейной регрессионной модели. Однако последняя составляющая в этом выражении необходима, так как значение зависимой переменной для наблюдения i , прошедшее тест $y_i > c$ и, следовательно, отсутствующее в выборке, всегда больше c .

Репараметризация функции правдоподобия

Чтобы упростить анализ, логарифмическую функцию правдоподобия можно репараметризовать. Для этого введем обозначения:

$$\delta = \beta/\sigma, \quad \gamma = 1/\sigma. \quad (8.2.13)$$

(Мы уже использовали такую репараметризацию в примере 7.7 для линейной регрессионной модели.) Это взаимно однозначное соответствие между $\{\beta, \sigma^2\}$ и $\{\delta, \gamma\}$ обретает соответствие дается формулами: $\beta = \delta/\gamma$, $\sigma^2 = 1/\gamma^2$. После репараметризации логарифмическое правдоподобие принимает вид:

$$\begin{aligned} \log \hat{L}(\beta, \sigma; \mathbf{X}) &= \\ &= \left[-\frac{1}{2} \log(2\pi) - \log(\gamma) - \right. \\ &\quad \left. \frac{1}{2} \gamma^2 (y_i - x_i' \delta)^2 \right] - \log[1 + \Phi(\gamma c - x_i' \delta)] \end{aligned} \quad (8.2.14)$$

Целевая функция ML представляет собой умноженное на $1/n$ логарифмическое правдоподобие выборки. Для случайной выборки логарифмическое правдоподобие выборки равно сумме из n логарифмических правдоподобий для наблюдений i . Таким образом, целевая функция в ML-оценке репараметризованной ускоренной регрессионной модели, обозначенная как $Q_n(\delta, \gamma)$, представляет собой среднее по i выражений (8.2.14). Оценка максимального правдоподобия $(\hat{\delta}, \hat{\gamma})$ истинных параметров (δ_0, γ_0) — это такие значения (δ, γ) , которые максимизируют эту целевую функцию.

Проверка состоятельности и асимптотической нормальности

Следующая задача состоит в выводе теоремы и проверке выполнения условий состоятельности и асимптотической нормальности для репараметризованного логарифмического правдоподобия (8.2.14). (При первом прочтении вам, возможно, захочется пропустить эту часть.)

Вклад и гессиян для наблюдения i

Утомительные вычисления дают:

$$s(\omega_i; \delta, \gamma) = \left[\frac{(\gamma y_i - x_i' \delta) x_i}{\frac{1}{\gamma} - (\gamma y_i - x_i' \delta) y_i} \right] - \lambda_i v_i \begin{bmatrix} -x_i \\ c \end{bmatrix}, \quad (8.2.15)$$

$$\begin{aligned} H(\omega_i; \delta, \gamma) = & - \begin{bmatrix} x_i x_i' & y_i x_i' \\ -y_i x_i' & \frac{1}{\gamma^2} + y_i^2 \end{bmatrix} + \\ & + \lambda_i(v_i)[\lambda_i(v_i) \cdot v_i] \begin{bmatrix} x_i x_i' & -c x_i \\ -c x_i & c^2 \end{bmatrix}, \end{aligned} \quad (8.2.16)$$

где K — количество регрессоров, $\omega_i = (y_i, x_i')$ и $v_i = \gamma c - x_i' \delta \left(= \frac{c - x_i' \delta}{\sigma} \right)$. Гессиян отрицательно полуопределен на *всей области определения* даже после репараметризации. Следовательно, целевая функция $\tilde{Q}_n(\delta, \gamma)$ не является *застутой*¹. Поэтому в данном случае подходящей теоремой о состоятельности является утверждение 7.6 (а не утверждение 7.5), тогда как подходящей теоремой об асимптотической нормальности остается утверждение 7.9.

Состоятельность

Условие (а) утверждения 7.5 есть идентифицируемость (условный) функции плотности:

$$\tilde{f}(y_i | x_i; \delta, \gamma) \neq \tilde{f}(y_i | x_i; \delta_0, \gamma_0)$$

с положительной вероятностью для $(\delta, \gamma) \neq (\delta_0, \gamma_0)$.

где (δ_0, γ_0) — истинные значения параметров, а репараметризованный логарифм функции плотности $\tilde{f}(y_i | x_i; \delta, \gamma)$ представлен в (8.2.14). Очевидно, что при заданных (y_i, x_i) значение $\tilde{f}(y_i | x_i; \delta, \gamma)$ отличается для различных значений γ . Поэтому идентификация сводится к условию: $x_i' \delta \neq x_i' \delta_0$ с положительной вероятностью для $\delta \neq \delta_0$. Но (8.1.10) указывает, что это условие выполнено, если матрица $E(x_i x_i')$ невырождена. Переходя к условию (б) утверждения (условие доминирования для

¹Одновременно можно сказать, что целевая функция $Q_n(\theta)$ отрицательно полуопределена в любой точке, представляющей решение уравнений правдоподобия (см. [Юлис, 1989]). Так как это исключает наличие глобальных точек и локальных максимумов, ситуация, когда существует несколько локальных максимумов, невозможна. Поэтому при условии, что локальный максимум $\tilde{Q}_n(\delta, \gamma)$ не является ни нейтральной точкой (не равенства параметров) условия первого порядка являются необходимыми и достаточными для существования *глобального* максимума. Если некоторый алгоритм (такой как алгоритм Ньютона — Рафсона) сойдется, найденное решение является *глобальным* максимумом.

$\hat{f}(y_i; x_i; \hat{\delta}, \hat{\gamma})$. Легко показать, что это условие выполнено при предположении о невырожденности $E(x_i x_i')$. Поэтому ML-оценка $(\hat{\delta}, \hat{\gamma})$ истинных значений параметров (δ, γ) является состоятельной при предположении о невырожденности $E(x_i x_i')$.

Асимптотическая нормальность

Используя (8.2.6), (8.2.7), (8.2.13), с помощью довольно громоздких вычислений можно показать, что $E[y|x] = 0$. Чрезвычайно громоздкие вычисления приводят к равенству для условной информационной матрицы: $E[yy'|x_i] = -E[H|x_i]$. Поэтому условие (3) утверждения 7.9 выполняется. Мы не будем обсуждать проверку условий (4) и (5) этого утверждения².

Итак, можно сделать следующий вывод: если $\{y_i, x_i\}$ — последовательность независимых и одинаково распределенных случайных величин и $E(x_i x_i')$ — невырожденная матрица, то тогда ML-оценка параметров $(\hat{\delta}, \hat{\gamma})$ усеченной регрессионной модели является состоятельной и асимптотически нормально распределенной с асимптотической ковариационной матрицей, состоятельно оцениваемой посредством

$$\text{Var}(\hat{\delta}, \hat{\gamma}) = - \left[\frac{1}{n} \sum_{i=1}^n H(w_i; \hat{\delta}, \hat{\gamma}) \right]^{-1}, \quad (8.2.17)$$

где $w_i = (y_i, x_i')$ и $H(w_i; \theta)$ определяется выражением (8.2.16).

Восстановление исходных параметров

Вследствие инвариантности оценок ML (см. параграф 7.1), ML-оценка для (β_1, σ_1^2) может быть посредством обратного отображения: $\hat{\beta} = \hat{\delta}/\hat{\gamma}$ и $\hat{\sigma}^2 = 1/\hat{\gamma}^2$. Если ML-оценка репараметризованной модели $(\hat{\delta}, \hat{\gamma})$ состоятельна, то такой же будет и ML-оценка $(\hat{\beta}, \hat{\sigma}^2)$, восстановленная таким

² Только для заинтересованных читателей: необходимо показать, что

$$E \sup_{\theta \in \Theta} |\log \hat{f}(y|x; \theta)| < \infty,$$

где $\theta = (\delta', \gamma')$ и Θ — компакт. Расширение аргументации примера 7.8 показывает, что существует доминирующая функция для выражения в скобках в уравнении (8.2.14). Чтобы показать существование доминирующей функции для $|\log \hat{f}(y|x; \theta)|$, можно использовать неравенство (7.2.14).

³ Похоже, что на основании опубликованных работ, проверяющих выполнение условий (4) и (5). Для случая, когда $\{x_i\}$ — последовательность фиксированных значений. Сакри [Sarg, 1992] доказал асимптотическую нормальность в предположении, что достаточно и $\lim_{n \rightarrow \infty} \sum_{i=1}^n x_i x_i' = H$ — невырожденная матрица. Его доказательство дано для более общего случая наличия сериальной корреляции наблюдений. Представляется разумным предположить, что для случая стохастических объясняющих переменных x_i (как здесь) достаточным условием состоятельности и асимптотической нормальности является невырожденность матрицы $E(x_i x_i')$.

образом. Асимптотическая ковариационная матрица $(\hat{\beta}, \hat{\sigma}^2)$ может быть получена тем же методом (см. лемма 2.5) как $\hat{J} \text{Var}(\hat{\beta}, \hat{\sigma}^2) \hat{J}'$, где $\hat{J}$ — оценка матрицы Якоби для обратного отображения ($\theta = \beta/\sigma, \eta = 1/\sigma^2$), вычисленная в точке $(\hat{\beta}, \hat{\sigma}^2)$:

$$\hat{J} = \begin{bmatrix} \frac{1}{\hat{\sigma}^2} I & -\frac{\hat{\beta}}{\hat{\sigma}^3} \\ 0 & -\frac{2}{\hat{\sigma}^3} \end{bmatrix}. \quad (8.2.18)$$

Контрольные вопросы

1. (Второй момент β_1) Из (8.2.6) и (8.2.7) выведите:

$$\begin{aligned} E(y_i^2 | x_i; \beta, \sigma^2) \text{ присутствует в выборке} &= \\ &= \frac{1}{\sigma^2} \left[1 + \lambda(\tau_1)\tau_1 + (x_i'\beta)^2 + \lambda(\tau_1)\tau_1^2 \delta \right], \end{aligned}$$

где $\tau_1 = \tau(x_i) = x_i'\beta$.

2. (Оценка усеченной регрессионной модели посредством GMM) Усеченная модель регрессии также может быть оценена с помощью GMM. Рассмотрим $K+1$ условий ортогональности. Первые K из них выглядят следующим образом:

$$g_1(w_i; \beta, \gamma) = \left(y_i - \frac{1}{\gamma} x_i'\beta - \frac{\lambda(\tau_1)}{\gamma} \right) \cdot x_i, \quad (8.2.19)$$

где $w_i = (y_i, x_i')$ и $\tau_1 = \tau(x_i) = x_i'\beta$ ($K+1$ -е условие ортогональности имеет следующий вид:

$$g_2(w_i; \beta, \gamma) = y_i^2 - \frac{1}{\gamma^2} \left[1 + \lambda(\tau_1)\tau_1 + (x_i'\beta)^2 + \lambda(\tau_1)\tau_1^2 \delta \right]. \quad (8.2.20)$$

- (а) Проверьте, что $E[g_1(w_i; \beta_0, \gamma_0)] = 0$ и $E[g_2(w_i; \beta_0, \gamma_0)] = 0$. Указание: Используйте закон итерных математических ожиданий. τ_1 является функцией от x_i .

- (б) Покажите, что пара (β^*, γ^*) удовлетворяет уравнениям правдоподобия тогда и только тогда, когда она удовлетворяет указанным $K+1$ условиям ортогональности. Указание: Пусть $g_2(w_i; \beta, \gamma) = (K+1)$ -й член $g(w_i; \beta, \gamma)$ в (8.2.15). Тогда

$$g_2(w_i; \beta, \gamma) + \gamma g_1(w_i; \beta, \gamma) = (x_i'\beta) \left(y_i - \frac{1}{\gamma} x_i'\beta - \frac{\lambda(\tau_1)}{\gamma} \right).$$

8.3. Цензурированные регрессионные (тобит) модели

Цензурированная модель регрессии называется также **моделью тобит** (*Tobit model*) в честь Дж. Тобина [Tobin, 1958], который был первым.

кто эвас поллитик, к кураторам в эконотку. Простейшая тобит модель может быть записана следующим образом:

$$y_i = x_i' \beta_0 + \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (8.3.1)$$

$$y_i = \begin{cases} y_i^* & \text{если } y_i^* > c, \\ c & \text{если } y_i^* \leq c. \end{cases} \quad (8.3.2)$$

Здесь ε_i распределены нормально с параметрами $N(0, \sigma_0^2)$, и $\{y_i, x_i\}$ ($i = 1, 2, \dots, n$) — i. d. Паровое значение c известно. В отличие от усеченной модели регрессии, здесь усечения нет: наблюдения, для которых значение зависимой переменной y_i^* не удовлетворяет критерию $y_i^* > c$, присутствующим в выборке. Основное отличие тобит-модели от обычной линейной регрессионной модели состоит в том, что зависимая переменная цензурирована (то есть ограничена каким-то диапазоном значений), что приводит к необходимости различать наблюдаемую зависимую переменную y_i и ненаблюдаемую (латентную) переменную y_i^* ; c — цензурированное значение y_i^* . Модель можно переписать в компактном виде.

$$y_i = \max\{x_i' \beta_0, c\}, \quad i = 1, \dots, n. \quad (8.3.3)$$

Функция правдоподобия для тобит-модели

Для likelihood, не затронутых цензурированием Тобина [Tobin, 1958], т.е. для так называемых *unlimited observations*, функция плотности y_i (уточним: относительная y_i) записывается как

$$\frac{1}{\sigma_0} \phi\left(\frac{y_i - x_i' \beta_0}{\sigma_0}\right). \quad (8.3.4)$$

Эта функция плотности отличается от likelihood для усеченной модели (8.2.11), поскольку в случае тобит-модели выборка не усечена. Для тех наблюдений в которых значение зависимой переменной изменяется из-за цензурирования (**limit observations**) все, что мы можем знать, — это то, что $y_i^* \leq c$, вероятность чего равна:

$$\begin{aligned} \text{Prob}(y_i^* \leq c | x_i) &= \text{Prob}\left(\frac{y_i^* - x_i' \beta_0}{\sigma_0} \leq \frac{c - x_i' \beta_0}{\sigma_0} \mid x_i\right) = \\ &= \Phi\left(\frac{c - x_i' \beta_0}{\sigma_0}\right) \quad (\text{так как } \frac{y_i^* - x_i' \beta_0}{\sigma_0} | x_i \sim N(0, 1)). \end{aligned} \quad (8.3.5)$$

Поэтому функция плотности y_i , определенная на интервале $[c, \infty)$, описывается уравнением (8.3.4) для $y_i > c$, а в точке $y_i = c$ сосредоточена

вероятностная масса $\Phi\left(\frac{c - x_i'\beta}{\sigma}\right)$. Эта функция плотности может быть описана как

$$\frac{1}{\sigma} \phi\left(\frac{y_i - x_i'\beta}{\sigma}\right) \cdot (1 - D_i) + \left[\Phi\left(\frac{c - x_i'\beta}{\sigma}\right)\right]^{D_i}, \quad (8.3.6)$$

где дамми-переменная D_i является функцией от y_i и определяется как

$$D_i = \begin{cases} 0, & \text{если } y_i > c \text{ (то есть } y_i^* > c), \\ 1, & \text{если } y_i = c \text{ (то есть } y_i^* \leq c). \end{cases} \quad (8.3.7)$$

Переходя к логарифмам и заменяя (y_i, σ_i^2) на их гипотетические значения (β, σ^2) , мы получаем логарифмическое условное правдоподобие для наблюдений i :

$$\log f(y_i | x_i; \beta, \sigma^2) = (1 - D_i) \log \left[\frac{1}{\sigma} \phi\left(\frac{y_i - x_i'\beta}{\sigma}\right) \right] + D_i \log \Phi\left(\frac{c - x_i'\beta}{\sigma}\right). \quad (8.3.8)$$

Таким образом, если выборка случайна, то среднее логарифмическое правдоподобие для табличного метода имеет вид:

$$\begin{aligned} Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n \left\{ (1 - D_i) \log \left[\frac{1}{\sigma} \phi\left(\frac{y_i - x_i'\beta}{\sigma}\right) \right] + D_i \log \Phi\left(\frac{c - x_i'\beta}{\sigma}\right) \right\} = \\ = \frac{1}{n} \sum_{y_i > c} \left\{ \log \left[\frac{1}{\sigma} \phi\left(\frac{y_i - x_i'\beta}{\sigma}\right) \right] \right\} + \frac{1}{n} \sum_{y_i \leq c} \left\{ \log \Phi\left(\frac{c - x_i'\beta}{\sigma}\right) \right\}. \end{aligned} \quad (8.3.9)$$

Регуляризация

Таким же, как и в случае усеченной модели регрессии, анализ становится проще, если сразу же параметризацию (8.2.13) регуляризовывать. Записанное логарифмическое правдоподобие имеет вид

$$\log \hat{f}(y_i | x_i; \delta, \gamma) = (1 - D_i) \left\{ -\frac{1}{2} \log(2\pi) - \frac{1}{2} (y_i - x_i'\delta)^2 \right\} + D_i \log \Phi(\gamma - x_i'\delta). \quad (8.3.10)$$

Мы уже видели, что выражение в фигурных скобках является вогнутой по (δ, γ) (см. пример 7.7). Мы также видели, что $\log \Phi(\cdot)$ является вогнутой функцией (см. пример 7.6), из чего следует, что $\log \Phi(\gamma - x_i'\delta)$ вогнута по (δ, γ) . То есть регуляризованное логарифмическое правдоподобие для регрессионной модели является вогнутой функцией, так как представляет собой неотрицательное взвешенное среднее двух вогнутых

функций (это было обнаружено Олсеном [Olson, 1978]). Как и в случае с пробит и логит-моделями, подходящей теоремой о состоятельности оценок для тобит-модели является утверждение 7.6.

Чтобы проверить выполнение условий, обеспечивающих состоятельность и асимптотическую нормальность репараметризованного логарифмического правдоподобия, остается только получить выражения для градиента и гессаана.

Вклад и гессаан для наблюдений i

Используя тот факт, что $\lambda(-v) = \frac{\phi(-v)}{\Phi(-v)}$ и (8.2.5), можно показать, что

$$s(w_i; \delta, \gamma) = (1 - I_i) \cdot \left[\frac{(\gamma y_i - x_i' \delta) x_i}{\gamma - (\gamma y_i - x_i' \delta) y_i} \right] + I_i \cdot \lambda(-v_i) \left[\frac{-x_i}{v_i} \right], \quad (8.3.11)$$

$$\begin{aligned} H(w_i; \delta, \gamma) = & -(1 - I_i) \cdot \left[\frac{x_i x_i'}{-y_i x_i'} \frac{1}{\gamma} + \frac{y_i x_i}{y_i^2} \right] - \\ & - I_i \cdot \lambda(-v_i) [\lambda(-v_i) + v_i] \begin{bmatrix} x_i x_i' & -x_i v_i \\ -x_i v_i & v_i^2 \end{bmatrix}. \end{aligned} \quad (8.3.12)$$

где $v_i \equiv \gamma y_i - x_i' \delta (= \frac{c - x_i' \beta}{\sigma})$.

Состоятельность и асимптотическая нормальность

Для тобит-модели мы не будем проверять выполнение условий, обеспечивающих состоятельность и асимптотическую нормальность оценок. Для случая, когда $\{w_i\}$ является последовательностью фиксированных констант, Амеция [Amemiya, 1973] доказал состоятельность и асимптотическую нормальность ML-оценок в тобит-модели в предположении, что x_i ограничены и что матрица $\lim \frac{1}{n} \sum_{i=1}^n x_i x_i'$ невырождена. Для случая, когда w_i стохастические (как здесь), разумным кажется предположение о том, что достаточным условием для состоятельности и асимптотической нормальности оценок является невырожденность матрицы $E(x_i x_i')$.

Восстановление исходных параметров

Так же, как и в случае усеченной модели регрессии, ML-оценки параметров (β_0, σ_0^2) может быть получена как $\hat{\beta} = \bar{\delta} / \hat{\gamma}$ и $\hat{\sigma}^2 = 1 / \hat{\gamma}^2$. Асимптотическая ковариационная матрица $(\hat{\beta}, \hat{\sigma}^2)$ может быть получена дельта-методом как $\hat{J} \text{Avar}(\bar{\delta}, \hat{\gamma}) \hat{J}'$, где $\hat{J}$ обозначает то же, что и в уравнении (8.2.18), и $\text{Avar}(\bar{\delta}, \hat{\gamma})$ равна $-\left[\frac{1}{n} \sum_{i=1}^n H(w_i; \bar{\delta}, \hat{\gamma}) \right]^{-1}$.

Контрольные вопросы

1. (Сравнение цензурированной и усеченной регрессий.) Одной из возможностей указать различия между усечением и цензурированием является вычисление $E(y_i | x_i)$ для цензурированной модели регрессии. Покажите, что

$$E(y_i | x_i) = \left[1 - \Phi \left(\frac{c - x_i' \beta_0}{\sigma_0} \right) \right] \times \left[x_i' \beta_0 + \sigma_0 \lambda \left(\frac{c - x_i' \beta_0}{\sigma_0} \right) \right] + \Phi \left(\frac{c - x_i' \beta_0}{\sigma_0} \right) c,$$

Указание: При $y_i > c$ распределение y_i (условное относительно x_i) представляет собой не что иное, как усеченное распределение, которое упоминалось в предыдущем параграфе, с математическим ожиданием (8.2.4). Вероятность того, что $y_i > c$, дана в (8.2.10).

2. Покажите, что гессиан в (8.3.12) отрицательно полуопределен. **Указание:**

$$\begin{bmatrix} x_i x_i' & -y_i x_i' \\ -y_i x_i' & y_i^2 + y_i^2 \end{bmatrix} = \begin{bmatrix} x_i \\ -y_i \end{bmatrix} \begin{bmatrix} x_i' & y_i \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{y_i} \end{bmatrix},$$

$$\begin{bmatrix} x_i x_i' & c x_i' \\ -c x_i' & c^2 \end{bmatrix} = \begin{bmatrix} x_i \\ -c \end{bmatrix} \begin{bmatrix} x_i' & c \end{bmatrix}.$$

Кроме того, $\lambda(r) \geq r$ для любого r (то есть $\lambda(-1) + r \geq 0$ для всех $-r$).

3. (Тобит при наличии серийной корреляции наблюдений.) Пусть $\{y_i, x_i\}$ — эргодический стационарный, но обязательно i.i.d. Является ли тобит ML-оценка, описанная в тексте, состоятельной? (Ответ: Да.)

8.4. Многомерные регрессии

В параграфе 1.3 и в примере 7.2 главы 7 мы уже отмечали, что OLS-оценка линейной регрессионной модели численно совпадает с ML-оценкой той же модели при нормально распределенных ошибках. Мы вывели GMM для моделей из одного уравнения с заданными регрессорами в главе 3 и для случая нескольких уравнений в главе 4, но мы не указывали, каковы ML-оценки для этих моделей. В этом параграфе мы рассмотрим свойства ML-оценки модели многомерной регрессии, которая является простейшим частным случаем модели, состоящей из нескольких уравнений.

Переформулирование модели многомерной регрессии

Модель многомерной регрессии описывается предположениями 4.1–4.5 и 4.7, где $x_{itm} = x_{itm} = x_i$ для всех $m = 1, 2, \dots, M$. Изменяя обозначения, введенные в главе 4, систему из M уравнений можно записать как

$$y_{itm} = x_{itm}' \beta_{0im} + e_{itm} \quad (im = 1, 2, \dots, M; t = 1, 2, \dots, n), \quad (8.4.1)$$

(1 × K) (K × 1)

Здесь мы продолжаем придерживаться соглашения, согласно которому истинные значения параметра снабжаются нижним индексом «0». Если определить

$$\underset{(M \times 1)}{y_t} = \begin{bmatrix} y_{t1} \\ y_{t2} \\ \vdots \\ y_{tM} \end{bmatrix}, \quad \underset{(M \times 1)}{v_t} = \begin{bmatrix} v_{t1} \\ v_{t2} \\ \vdots \\ v_{tM} \end{bmatrix}, \quad \underset{(K \times M)}{\Pi_0} = [\pi_{01} \quad \pi_{02} \quad \cdots \quad \pi_{0M}],$$
(8.4.2)

то эта система M уравнений может быть записана более компактно:

$$y_t = \Pi_0' x_t + v_t \quad (t = 1, 2, \dots, n). \quad (8.4.3)$$

Эта модель имеет следующие особенности:

- $\{y_t, x_t\}$ стационарный и эргодический (предположение 4.2).
- Обычно регрессоры x_t являются предопределенными; $E(x_t \cdot v_{t+m}) = 0$ для любого m (предположение 4.3).
- Расположение условия идентификации (предположение 4.4) сводится к тому, что $E(x_t x_t')$ — невырожденная матрица.
- Вектор остатков условно гомосkedастичен: $E(v_t v_t' | x_t) = \Omega_0$, и Ω_0 положительно определена (предположение 4.7).

СММ-оценка π_{0m} совпадает с OLS-оценкой:

$$\hat{\pi}_{0m} = \left(\frac{1}{n} \sum_{t=1}^n x_t x_t' \right)^{-1} \left(\frac{1}{n} \sum_{t=1}^n x_t \cdot y_{tm} \right). \quad (8.4.4)$$

Поэтому оценка СММ (OLS) для Π_0 может быть найдена как

$$\underset{(M \times M)}{\hat{\Pi}} = \left(\underset{(K \times K)}{\frac{1}{n} \sum_{t=1}^n x_t x_t'} \right)^{-1} \left(\underset{(K \times M)}{\frac{1}{n} \sum_{t=1}^n x_t y_t'} \right). \quad (8.4.5)$$

Функция правдоподобия

Мы еще недостаточно специфицировали модель, чтобы было возможным найти ее ML-оценку. Введем дополнительные предположения о том, что 1) $v_t | x_t \sim N(0, \Omega_0)$ и 2) $\{y_t, x_t\}$ — последовательность независимых одинаково распределенных случайных величин (эти предположения усиливают предположения 4.2 и 4.5).

Многомерная регрессионная модель с предположением о нормальности (1) означает, что $y_i, x_i \sim N(\Pi'x_i, \Omega_i)$. Функция плотности этого многомерного нормального распределения имеет вид:

$$(2\pi)^{-1/2} |\Omega_i|^{-1/2} \exp \left[-\frac{1}{2} (y_i - \Pi'x_i)' \Omega_i^{-1} (y_i - \Pi'x_i) \right]. \quad (8.4.6)$$

Заменяя истинные значения параметров (Π_0, Ω_0) их гипотетическими значениями (Π, Ω) и беря логарифм, мы получаем логарифмическое условное правдоподобие для наблюдения i :

$$\log f(y_i|x_i; \Pi, \Omega) = -\frac{M}{2} \log(2\pi) - \frac{1}{2} \log(|\Omega|^{-1}) - \\ - \frac{1}{2} (y_i - \Pi'x_i)' \Omega^{-1} (y_i - \Pi'x_i), \quad (8.4.7)$$

где был использован следующий факт: $\log(|\Omega|^{-1}) = \log(|\Omega|^{-1})$.

Для случайной выборки объема n на $\frac{1}{n}$ логарифмическое условное правдоподобие выборки равно сумме из n логарифмических условных правдоподобий для наблюдений i . То есть целевая функция $Q_n(\Pi, \Omega)$ та же, что и в (8.4.7), но с суммированием по i и получается как

$$Q_n(\Pi, \Omega) = -\frac{M}{2} \log(2\pi) - \frac{1}{2} \log(|\Omega|^{-1}) - \\ - \frac{1}{2n} \sum_{i=1}^n (y_i - \Pi'x_i)' \Omega^{-1} (y_i - \Pi'x_i). \quad (8.4.8)$$

Доказательство того, что последняя составляющая может быть переписана как

$$\frac{1}{2n} \sum_{i=1}^n (y_i - \Pi'x_i)' \Omega^{-1} (y_i - \Pi'x_i) = \frac{1}{2} \text{trace}[\Omega^{-1} \tilde{\Omega}(\Pi)], \quad (8.4.9)$$

где $\tilde{\Omega}(\Pi)$ определяется как

$$\tilde{\Omega}(\Pi) \equiv \frac{1}{n} \sum_{i=1}^n (y_i - \Pi'x_i)(y_i - \Pi'x_i)', \quad (8.4.10)$$

оставшаяся в качестве контрольного вопроса 1. Тогда целевая функция может быть переписана как

$$Q_n(\Pi, \Omega) = -\frac{M}{2} \log(2\pi) + \frac{1}{2} \log(|\Omega|^{-1}) - \frac{1}{2} \text{trace}[\Omega^{-1} \tilde{\Omega}(\Pi)]. \quad (8.4.11)$$

ML-оценки для (Π_0, Ω_0) — это такие значения (Π, Ω) , которые максимизируют целевую функцию. Пространство параметров для (Π, Ω)

определяется как

$\{(\mathbf{P}, \Omega) : \Omega \text{ — симметричная и положительно определенная матрица}\}.$

Максимизация функции правдоподобия

Сейчас мы покажем, что решение указанной задачи максимизации численно совпадает с GMM-оценкой. То есть ML-оценка $\mathbf{P}_0$ — это GMM/OLS-оценка $\hat{\mathbf{P}}$ в (8.4.5), а ML-оценка Ω_0 дается соотношением:

$$\hat{\Omega} = \frac{1}{n} \sum_{i=1}^n \hat{e}_i \hat{e}_i' = \hat{\Omega}(\hat{\mathbf{P}}), \quad (8.4.12)$$

где $\hat{e}_i = y_i - \hat{\mathbf{P}}'x_i$.

При предположениях многомерной регрессии, упомянутых выше, $\hat{\Omega}(\mathbf{P})$ положительно определена с вероятностью, равной единице, для любой заданной $\mathbf{P}$ при достаточно больших n (доказательство этого утверждения остается читателю как аналитическое упражнение 2). Поэтому мы можем предположить, что $\hat{\Omega}(\mathbf{P})$ положительно определена, а не просто положительно полуопределена. Доказательство того, что GMM-оценка численно совпадает с оценкой максимального правдоподобия, основано на следующей двухшаговой процедуре максимизации целевой функции.

Шаг 1: Первый шаг состоит в максимизации $Q_n(\mathbf{P}, \Omega)$ по Ω в предположении, что $\mathbf{P}$ задан. При этом полезно воспользоваться следующим результатом из матричной алгебры:

Неравенство, включающее в себя след и определитель матрицы¹: Пусть $\mathbf{A}$ и $\mathbf{B}$ — симметричные и положительно определенные матрицы одинакового размера. Тогда функция

$$f(\mathbf{A}) = \log |\mathbf{A}| - \text{tr}(\mathbf{A}\mathbf{B})$$

достигает единственного максимума в точке $\mathbf{A} = \mathbf{B}^{-1}$.

Этот результат с $\mathbf{A} = \Omega^{-1}$ и $\mathbf{B} = \hat{\Omega}(\mathbf{P})$ немедленно приводит к выводу о том, что целевая функция (8.4.11) достигает своего единственного максимума в точке $\Omega = \hat{\Omega}(\mathbf{P})$ при заданной $\mathbf{P}$. Подстановка этой Ω в (8.4.11) дает нам (умноженную на $\frac{1}{n}$)

¹ Этот результат вытекает из Леммы 6A [Jørgensen, 1993]. Единственность точки максимума достигается благодаря линейности оператора следа и строгой вогнутости $\log|\mathbf{A}|$, отмеченной в [Magnus and Neudecker, 1988, с. 232].

концентрированную логарифмическую функцию правдоподобия (*concentrated log likelihood function*) (концентрированную по отношению к Ω):

$$\begin{aligned} Q_n(\Pi) &= Q_n(\Pi, \hat{\Omega}(\Pi)) = \\ &= -\frac{M}{2} \log(2\pi) - \frac{1}{2} \log(|\hat{\Omega}(\Pi)|^{-1}) - \frac{1}{2} \text{trase}[\hat{\Omega}(\Pi)^{-1} \hat{\Omega}(\Pi)] = \\ &= -\frac{M}{2} \log(2\pi) - \frac{1}{2} \log(|\hat{\Omega}(\Pi)|) - \frac{M}{2} \end{aligned} \quad (8.4.13)$$

Шаг 2: Внимательно глядя на концентрированную логарифмическую функцию правдоподобия $Q_n(\Pi)$, можно заметить, что ML-оценка Π_0 должна доставлять минимум

$$|\hat{\Omega}(\Pi)| = \left| \frac{1}{n} \sum_{t=1}^n (y_t - \Pi' x_t)(y_t - \Pi' x_t)' \right|. \quad (8.4.14)$$

Доказательство того факта, что OLS-оценка $\hat{\Pi}$ из (8.4.5) доставляет минимум этому выражению, оставлено читателю в качестве аналитического упражнения 1. Наконец, подстановка $\Pi = \hat{\Pi}$ в (8.4.10) дает (8.4.12).

Состоятельность и асимптотическая нормальность

В главе 4 было показано, что без предположения о нормальности остатков $\{e_t\}$ (распределении нормально) или предположения о том, что $\{y_t, x_t\}$ — i.i.d., GMM/OLS-оценка $\hat{\Pi}$ из (8.4.5) является состоятельной и асимптотически нормальной и что $\hat{\Omega}$ из (8.4.12) состоятельна. Из этого тривиально следует, что ML-оценка Π_0 также является состоятельной и асимптотически нормальной и что ML-оценка Ω_0 состоятельна. Также следует отметить, что ML-оценка Π_0 , основанная на функции правдоподобия, которая предполагает нормальность, является состоятельной и асимптотически нормальной, даже если ошибки не распределены по нормальному закону.

Мы не будем заниматься асимптотической нормальностью ML-оценки $\Omega_0, \hat{\Omega}(\hat{\Pi})$. Показать асимптотическую нормальность ML-оценки можно, например, проверив условие соответствующей теоремы об асимптотической нормальности из предыдущей главы.

Контрольные вопросы

1. Докажите (8.4.9). **Указание:** Воспользуйтесь следующими свойствами следов: следа, 1) $\text{trase}(x) = x$, если $x = \text{diag}(x)$, 2) $\text{trase}(A - B) = \text{trase}(A) - \text{trase}(B)$ и 3) $\text{trase}(AB) = \text{trase}(BA)$ при условии, что

матричные произведения AB и BA определены. Ответ:

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n (y_t - \Pi'x_t)' \Omega^{-1} (y_t - \Pi'x_t) &= \\ &= \text{trace} \left[\frac{1}{n} \sum_{t=1}^n (y_t - \Pi'x_t)' \Omega^{-1} (y_t - \Pi'x_t) \right] \\ &= \frac{1}{n} \sum_{t=1}^n \text{trace} [(y_t - \Pi'x_t)' \Omega^{-1} (y_t - \Pi'x_t)] \\ &= \frac{1}{n} \sum_{t=1}^n \text{trace} [\Omega^{-1} (y_t - \Pi'x_t) (y_t - \Pi'x_t)'] \\ &= \text{trace} \left[\frac{1}{n} \sum_{t=1}^n \Omega^{-1} (y_t - \Pi'x_t) (y_t - \Pi'x_t)' \right] \\ &= \text{trace} \left[\Omega^{-1} \frac{1}{n} \sum_{t=1}^n (y_t - \Pi'x_t) (y_t - \Pi'x_t)' \right]. \end{aligned}$$

8.5. FIML

Как было показано в главе 4, многомерная модель регрессии является частным случаем модели кажущихся несвязанными регрессий (SUR), которая, в свою очередь, является частным случаем системы одновременных уравнений, для которой применим трехэтапный метод наименьших квадратов (3SLS). Двухэтапный OLS (2SLS) совпадает с GMM, когда каждое уравнение системы оценивается отдельно, в то время как 3SLS совпадает с GMM, когда оценивается вся система в целом. В этом параграфе представим ML-анализ для 2SLS и 3SLS.

Система одновременных уравнений с общими инструментами: новые обозначения

Напомним, что эта модель представляет собой **модель из M уравнений** *M-equation system*, описываемую предположениями 4.1–4.5 и 4.7, где $x_{tm} = x_t$ для всех $m = 1, 2, \dots, M$ (то есть набор инструментов одинаков для всех уравнений). Сохраняя договоренность снабжать истинные значения параметров нижним индексом 0, мы можем записать M уравнений системы следующим образом:

$$y_{tm} = z'_{tm} \beta_{tm0} + \varepsilon_{tm} \quad (m = 1, 2, \dots, M; t = 1, 2, \dots, n). \quad (8.5.1)$$

(1 × L_m) (1 × K)

Отметим следующие особенности этой модели:

- В отличие от модели многомерной регрессии, в каждом из m уравнений регрессоры z_{tm} могут не быть ортогональны ошибкам, однако имеется K доступных предопределенных переменных x_t для

которых условие ортогональности выполняется: $E(x_i \cdot \varepsilon_{im}) = 0$ для всех m (это предположение 4.3). Определим вектор ε_i размерности $M \times 1$ как

$$\varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iM} \end{bmatrix}. \quad (8.5.2)$$

Условие ортогональности можно выразить как

$$E(x_i \varepsilon_i') = 0 \quad (8.5.3)$$

Эти K предопределенных переменных используются как общий набор инструментов для GMM-оценки.

- Переменные, входящие в систему, но не включенные в x_i (кроме ошибок), называются **эндогенными переменными** (*endogenous variables*).
- Второе условие идентификации (предположение 4.4) состоит в том, что матрица $E(x_i x_{im}')$ размерности $K \times L_m$ имеет полный столбцовый ранг для всех m . Уравнения называются **сверхидентифицированными** (*overidentified*), если условие ранга выполнено и $L_m < K$.
- Вектор ошибок условно гомосkedастичен: $E(\varepsilon_i \varepsilon_i' | x_i) = \Sigma_0$ и Σ_0 положительно определена (предположение 4.7).
- $E(x_i x_i')$ — невырожденная матрица (следствие условной гомосkedастичности и требования невырожденности в предположении 4.5).

GMM-оценки системы (1) здесь являются 3SLS-оценкой, приведенной в (4.5.12) и поэтому как оценивание каждого уравнения в отдельности с помощью GMM представляет собой 2SLS-уравнений соотношения (3.6.3).

Следующие примеры кратко иллюстрируют эти понятия.

Пример 8.1 (модель рыночного равновесия): Расширенная версия примера Уоркинга, рассмотренная в параграфе 3.1, представляет собой систему из двух уравнений:

$$q_1 = \alpha_1 p_1 + \beta_1 + \beta_2 m_1 + \varepsilon_{11} \quad (\text{спрос}), \quad (8.5.4)$$

$$q_1 = \alpha_2 p_1 + \beta_1 + \beta_2 m_1 + \varepsilon_{12} \quad (\text{предложение}). \quad (8.5.5)$$

В этой системе q_1 — количество кофе, p_1 — цена кофе. Переменная q_1 фигурирует как в уравнении спроса, так и в уравнении предложения. Переменная m_1 представляет собой количество осадков.

которых можно на предложенные коэф. Эта система может быть представлена в общей форме введением следующих обозначений:

$$y_{1i} = q_i, \quad y_{2i} = q_i, \quad z_{1i} = \begin{bmatrix} \beta_1 \\ 1 \\ \alpha_i \end{bmatrix}, \quad z_{2i} = \begin{bmatrix} \beta_2 \\ 1 \\ \alpha_i \end{bmatrix},$$

$$\delta_{1i} = \begin{bmatrix} \gamma_{11} \\ \beta_{11} \\ \beta_{12} \end{bmatrix}, \quad \delta_{2i} = \begin{bmatrix} \gamma_{21} \\ \beta_{21} \\ \beta_{22} \end{bmatrix}.$$

(Заметим, что y_{1i} и y_{2i} являются одной и той же переменной.) Если $E(\varepsilon_{1j}) = 0$, $E(\varepsilon_{2j}) = 0$ и $E(\alpha_j \varepsilon_{1j}) = 0$ для $j = 1, 2$, то тогда $(1, \alpha_i, \alpha_i)$ могут быть включены в общий набор инструментов, так что

$$z_i = \begin{bmatrix} 1 \\ \alpha_i \\ \alpha_i \end{bmatrix}.$$

Остальные переменные (кроме ошибок), входящие в систему (то есть в данном примере β_1 и α_1) являются эндогенными переменными. При условии, что ранг матрицы условий идентификации бесконечен, каждое уравнение идентично (ровно) одному. Поскольку количество переменных в правой части уравнений равно количеству инструментов,

Пример 8.2 (уравнение заработной платы): В главе 3 мы оценивали структурное уравнение заработной платы. Дополним это уравнение еще одним, объясняющим результат теста *Knowledge of the World of Work* (KW/W).

$$LW_i = \gamma_{11}\delta_i + \beta_{11} + \beta_{12}Q_i + \varepsilon_{1i}, \quad (8.5.6)$$

$$KW/W_i = \gamma_{21}\delta_i + \beta_{21} + \beta_{22}Q_i, \quad (8.5.7)$$

где δ_i — количество лет обучения индивида i . Предположим, что $E(\varepsilon_{1j}) = 0$ и $E(\varepsilon_{2j}) = 0$ для $j = 1, 2$. Пусть ε_{1i} как распределение является переменной MED (облажилище матерн), которая не фигурирует ни в одном из уравнений, но которая является предсказательной, в том смысле, что $E(MED_i \varepsilon_{1j}) = 0$ для $j = 1, 2$. То есть общий набор инструментов есть $z_i = (1, Q_i, MED_i)'$. Остальные переменные (кроме ошибок), входящие в систему, являются эндогенными. В этой системе из двух уравнений имеется три эндогенные переменные: LW , S и KW/W . При условии выполнения рыночного условия идентификации первое уравнение точно идентифицировано, поскольку количество переменных в правой части уравнения равно количеству инструментов. Второе уравнение не идентифицируемо.

Полная система одновременных уравнений

ML-аналог 3SLS называется **оценкой максимального правдоподобия с полной информацией** (*full-information maximum likelihood*) — **FIML**. Перед применением FIML на модель односторонних уравнений необходимо наложить несколько дополнительных условий до введения предположений о нормальном распределении и независимости и одинаковости распределенности. Эти дополнительные требования состоят в том, чтобы модель из M уравнений была «полной» системой одновременных уравнений. Полнота требует выполнения двух условий:

1. Количество эндогенных переменных в системе должно равняться числу уравнений. Это условие подразумевает, что если y_t — вектор длины M , содержащий эти M эндогенных переменных, то $(y_{t1}, \dots, y_{tm}, z_{t1}, \dots, z_{tn})$ — это вектор длины $(m+n)$, который позволяет записать систему из M уравнений (8.5.1) в виде:

$$\Gamma_1 y_t + B_0 z_t = \varepsilon_t \quad (t=1, 2, \dots, n), \quad (8.5.8)$$

$(M \times M) \quad (1 \times 1) \quad (M \times n) \quad (1 \times 1) \quad (n \times 1)$

где m -е уравнение имеет вид: $y_{tm} = z_{tm}\delta_{tm} + \varepsilon_{tm}$. (Эти трижды конструировано для примера 8.1 выше.) Такая запись называется **структурной формой**, а (Γ_1, B_0) называются **параметрами структурной формы** (*structural form parameters*). Как будет показано ниже, каждый из структурных параметров является функцией от $(\delta_1, \dots, \delta_{m+n})$.

Система уравнений из примера 8.1 удовлетворяет этому требованию. Система уравнений из примера 8.2 не является полной системой, поскольку количество эндогенных переменных превышает количество наблюдений. С помощью FIML, невозможно оценить полную систему (за исключением случаев, когда мы дополним систему соответствующими уравнениями, см. приложение FIML, далее). Напротив, полную систему можно оценить с помощью 3SLS (или с помощью GMM для системы уравнений, если не вводить предположение об условии гомосkedастичности), если система удовлетворяет ряду другим условиям идентифицируемости.

2. Квадратная матрица Γ_0 невырождена. Из этого следует, что структурная форма может быть решена относительно эндогенной переменной y_t следующим образом:

$$y_t = -\Gamma_1^{-1} B_0 z_t + \Gamma_1^{-1} \varepsilon_t \equiv \Pi_0' z_t + \eta_t. \quad (8.5.9)$$

302

$$\Pi_0^* \equiv - \Gamma_0^{-1} B_0, \quad (8.5.10)$$

$$(M \times K) \quad (M \times M) \quad (M \times K)$$

$$v_i \equiv \Gamma_0^{-1} \varepsilon_i, \quad (8.5.11)$$

$$(M \times 1) \quad (M \times M) \quad (M \times 1)$$

Выражение (8.5.9) называется **приведенной формой** (*reduced-form representation*) структурной формы. Элементы Π_0 называются **коэффициентами приведенной формы**. В каждом уравнении приведенной формы, регрессоры представляют собой единичный вектор перекрестных z_t . Из (8.5.3) и (8.5.11) следует, что $E(z_t v_t) = 0$, то есть все регрессоры являются предопределенными. Следовательно, система уравнений в приведенной форме является многомерной регрессионной моделью.

Вычисления между (Γ_0, B_0) и δ_0

Пусть δ_0 — штабелированный вектор, включающий в себя все коэффициенты системы из M уравнений (8.5.1):

$$\delta_0 \equiv \begin{bmatrix} \delta_{01} \\ \delta_{02} \\ \vdots \\ \delta_{0M} \end{bmatrix}, \quad (8.5.12)$$

$$(M \times 1) \quad (M \times 1)$$

Для понимания механизма FIML важно посмотреть, как матрица коэффициентов (Γ_0, B_0) зависит от δ_0 . В качестве иллюстрации рассмотрим пример 5.1. В этом примере вектор δ_0 определяется как

$$\delta_0 \equiv \begin{bmatrix} \gamma_{11} \\ \gamma_{12} \\ \gamma_{21} \\ \gamma_{22} \end{bmatrix}, \quad (8.5.13)$$

$$(4 \times 1)$$

Два эндогенных переменных в примере — это (p_t, q_t) . Порядок включения этих переменных не важен. Для определенности пусть p_t первой входит в вектор y_t : $y_t = (p_t, q_t)'$. Тогда параметры структурной формы могут быть записаны как

$$\Gamma_0 \equiv \begin{bmatrix} -\gamma_{11} & 1 \\ -\gamma_{21} & 1 \end{bmatrix}, \quad B_0 \equiv \begin{bmatrix} -\gamma_{12} & -\gamma_{12} & 0 \\ \gamma_{22} & 0 & \gamma_{22} \end{bmatrix}, \quad (8.5.14)$$

$$(2 \times 2) \quad (2 \times 3)$$

Мы приведем этот пример для иллюстрации трех аспектов. Во-первых, каждая строка матрицы Γ_0 содержит единицу, что отражает тот факт, что коэффициент при зависимой переменной в каждом из M уравнений (8.5.1) равен единице. В этом смысле Γ_0 уже нормализована. Во-вторых, некоторые элементы Γ_0 и B_0 равны нулю, что отражает тот факт, что некоторые из эндогенных или предопределенных переменных не входят в отдельные уравнения системы. Эта особенность матрицы коэффициентов структурной формы называется **исключающими ограничениями** (*exclusion restrictions*) (В примере 8.1 ни один из элементов Γ_0 не равен нулю, так как обе эндогенные переменные входят в оба уравнения системы.) В-третьих, в системе нет ограничений, затрагивающих сразу несколько уравнений, поэтому каждый элемент δ_0 понимается в (Γ_0, B_0) как бы один раз.

Функция правдоподобия для FIML

Чтобы полная система одновременных уравнений могла оцениваться посредством FIML, мы предполагаем, что 1) вектор структурных ошибок ε_t имеет совместное нормальное распределение, условное относительно x_t (то есть $\varepsilon_t|x_t \sim N(0, \Sigma_0)$) и 2) $\{y_t, x_t\}$ — последовательность независимых одинаково распределенных случайных величин, а не просто ортогональные стационарные мартигал-разности (это предположение — более жесткий вариант предположений 4.2 и 4.5).¹

Далее предположения о нормальном распределении структурных ошибок ε_t (1) и (8.5.11) мы записываем $\varepsilon_t|x_t \sim N(0, \Gamma_0^{-1}\Sigma_0(\Gamma_0^{-1})')$. Связывая это с приведенной формой $y_t = \Pi_0'x_t + v_t$ в (8.5.10), получаем:

$$y_t|x_t \sim N(-\Gamma_0^{-1}B_0x_t, \Gamma_0^{-1}\Sigma_0(\Gamma_0^{-1})'). \quad (8.5.16)$$

Почему логарифмическое правдоподобие для наблюдений t принимает вид:

$$\begin{aligned} \log f(y_t, x_t; \delta, \Sigma) &= -\frac{M}{2} \log(2\pi) - \frac{1}{2} \log |\Gamma^{-1}\Sigma(\Gamma^{-1})'| - \\ &\quad - \frac{1}{2} [y_t - \Gamma^{-1}Bx_t]' [\Gamma^{-1}\Sigma(\Gamma^{-1})']^{-1} [y_t - \Gamma^{-1}Bx_t]. \end{aligned} \quad (8.5.16')$$

Это правдоподобие является функцией от (δ, Σ) , поскольку, как было показано в (8.5.14), коэффициенты структурной формы (Γ, B) являются функциями от δ . Теперь

$$\begin{aligned} [y_t - \Gamma^{-1}Bx_t]' [\Gamma^{-1}\Sigma(\Gamma^{-1})']^{-1} [y_t - \Gamma^{-1}Bx_t] &= \\ = [y_t - \Gamma^{-1}Bx_t]' [\Gamma'\Sigma^{-1}\Gamma] [y_t - \Gamma^{-1}Bx_t] &= \\ = (y_t - Bx_t)' \Sigma^{-1} [\Gamma y_t + Bx_t] \end{aligned} \quad (8.5.17)$$

и

$$\Gamma^{-1} \Sigma (\Gamma^{-1})' = \Gamma^{-1} [\Sigma + (\Gamma^{-1})']^{-1} \Sigma (\Gamma^{-1})'. \quad (8.5.18)$$

Подставляя эти выражения в (8.5.16), получаем:

$$\begin{aligned} \log f(y_i | x_i; \delta, \Sigma) = & -\frac{M}{2} \log(2\pi) - \frac{1}{2} \log |\Gamma^{-2}| - \frac{1}{2} \log |\Sigma| - \\ & - \frac{1}{2} [\Gamma y_i + B x_i]' \Sigma^{-1} [\Gamma y_i + B x_i]. \end{aligned} \quad (8.5.19)$$

Усредняя по i , получаем целевую функцию FIML для случайной выборки

$$\begin{aligned} Q_n(\delta, \Sigma) = & -\frac{M}{2} \log(2\pi) - \frac{1}{2} \log |\Gamma^{-2}| - \frac{1}{2} \log |\Sigma| - \\ & - \frac{1}{2n} \sum_{i=1}^n [\Gamma y_i + B x_i]' \Sigma^{-1} [\Gamma y_i + B x_i]. \end{aligned} \quad (8.5.20)$$

Оценка FIML истинных параметров $(\delta_0, \Sigma_0) = \arg(\delta, \Sigma)$, доставляющая максимум целевой функции,

Концентрированная функция правдоподобия для FIML

Как и в случае с максимизацией целевой функции для многомерной регрессионной модели, мы можем решить вопрос в два этапа. Сначала можно найти максимум $Q_n(\delta, \Sigma)$ по Σ , считая δ заданным. В результате получаем:

$$\tilde{\Sigma}(\delta) = \frac{1}{n} \sum_{i=1}^n (\Gamma y_i + B x_i)(\Gamma y_i + B x_i)'. \quad (8.5.21)$$

Элемент (i, k) может быть записан как

$$\tilde{\sigma}_{ik} = \frac{1}{n} \sum_{i=1}^n (y_i - x_{i0}' \delta_0)(y_i - x_{ik}' \delta_k). \quad (8.5.22)$$

Подставляя (8.5.2) обратно в целевую функцию (8.5.20), умноженную на $1/n$, концентрированную функцию правдоподобия для FIML можно записать в виде:

$$\begin{aligned} Q_n(\delta) &= Q_n(\delta, \tilde{\Sigma}(\delta)) \\ &= -\frac{M}{2} \log(2\pi) - \frac{M}{2} + \frac{1}{2} \log |\Gamma^{-2}| - \frac{1}{2} \log [\tilde{\Sigma}(\delta)] = \\ &= -\frac{M}{2} \log(2\pi) - \frac{M}{2} + \frac{1}{2} \log \left| \sum_{i=1}^n (y_i - \Gamma^{-1} B x_i)(y_i + \Gamma^{-1} B x_i)' \right|. \end{aligned} \quad (8.5.23)$$

Второй шаг состоит в максимизации этой концентрированной функции правдоподобия по δ . FIML-оценка $\hat{\delta}$ истинного значения параметра δ_0 — это такое значение δ , которое доставляет максимум этой функции¹. FIML-оценка $\Sigma_{II} = \text{cov } \hat{\Sigma}(\hat{\delta})$.

Тестирование сверхидентифицирующих ограничений

Сравним теперь концентрированную функцию правдоподобия для FIML $Q_n(\delta)$ из (8.5.23) с концентрированной функцией правдоподобия для модели многомерной регрессии $Q_n(\Pi)$ из (8.4.13). Концентрированную функцию правдоподобия для FIML $Q_n(\delta)$ можно получить из $Q_n(\Pi)$ путем введения определенных ограничений. Эти ограничения следуют из следующей гипотезы:

$$U_0 = \Pi_0' - \Gamma_0^{-1}B; \text{ или } \Gamma_0\Pi_0' - B_0 = 0. \quad (8.5.24)$$

Иными словами, FIML-оценка $\hat{\delta}_0$ — это оценка многомерной регрессии с ограничениями. Поэтому истинность нулевой гипотезы можно проверить с помощью принципа отношения правдоподобия. Поскольку OLS-оценка $\hat{\Pi}$ из (8.4.5) максимизирует концентрированную функцию правдоподобия для многомерной регрессии $Q_n(\Pi)$ из (8.4.13), а FIML-оценка $\hat{\delta}$ максимизирует концентрированную функцию правдоподобия для FIML $Q_n(\delta)$ из (8.5.23), статистика теста отношения правдоподобия принимает следующий вид:

$$\begin{aligned} LR &= 2n \cdot (Q_n^*(\hat{\Pi}) - Q_n^*(\hat{\delta})) = \\ &= n \cdot \{\log[\hat{\Pi}(-(\hat{\Gamma}^{-1} - \hat{B})')] - \log \hat{\Pi}(\hat{\Pi})\}, \end{aligned} \quad (8.5.25)$$

где $\hat{\Pi}(\Pi)$ определена в (8.4.10) и $(\hat{\Gamma}, \hat{B})$ — значение (Γ, B) при $\hat{\delta}$. Так как размерность δ равна $\sum_{m=1}^M L_m$, количество ограничений, накладываемых нулевой гипотезой H_0 на Π , равно:

$$KIM = \sum_{m=1}^M L_m = \sum_{m=1}^M (K - I_{m1}), \quad (8.5.26)$$

что равно общему числу сверхидентифицирующих ограничений в системе. Поэтому тест отношения правдоподобия, основанный на LR-статистике из (8.5.25), называется **тестом на сверхидентифицирующие**

¹То есть FIML-оценка $\hat{\delta}$ является экстремальной оценкой, которая максимизирует

$$\frac{1}{2} \log \frac{1}{n} \sum_{i=1}^n (y_i - \Gamma^{-1} B x_i)(y_i + \Gamma^{-1} B x_i)'$$

Некоторые свойства этой экстремальной оценки рассматриваются в дополнительном аналитическом утверждении 3.

ограничения (*test of overidentifying restrictions*). Он является тестом на спецификацию модели, так как тестируемые ограничения — это условия, предполагаемое в модели.

Свойства FIML-оценок

Для FIML-оценок не существует аналитического решения. Значит, для проверки состоятельности и асимптотической нормальности нам будет необходимо проверять условия соответствующих теорем из предыдущего параграфа. Далее мы приведем без доказательства основные свойства FIML-оценок.

Идентифицируемость

Формулирование условий идентифицируемости в различных формах для полной системы одновременных уравнений — одна из важнейших тем большинства учебников. Здесь же мы предлагаем в качестве основного аналитического выражения 3 указать, что условие идентифицируемости FIML, как экстремальной оценки равносильно ранговому условию идентифицируемости (то есть что матрица $E(\mathbf{z}_i \mathbf{z}_i')$ имеет полный столбцовый ранг для всех $m = 1, 2, \dots, M$).

Асимптотические свойства

Хотя целевая функция FIML (8.5.20) выводилась в предположении нормального распределения ошибок $\varepsilon_i | \mathbf{z}_i$, FIML-оценка $\hat{\delta}_0$ является состоятельной и асимптотически нормально распределенной и без этого предположения. Доказательство состоятельности приведено, например, в [Аптекуа, 1985, pp. 232–233]. Элегантное доказательство асимптотической нормальности распределения $\hat{\delta}$ принадлежит Хаусману [Хаусман, 1975], который показывает, что итерационный метод инструментальных переменных может рассматриваться как алгоритм для решения условий первого порядка максимизации $Q_n(\delta, \Sigma)^1$. Это доказательство показывает, что FIML-оценка $\hat{\delta}_0$ асимптотически эквивалентна 3SLS-оценке². Поэтому ее асимптотическая дисперсия дается выражением (4.5.14) и состоятельной оценкой асимптотической дисперсии является (4.5.17).

¹Мы не будем приводить условия первого порядка (уравнения правдоподобия) для FIML-оценок, так как их вывод требует громоздких матричных обозначений. Детали можно найти в [Аптекуа, 1985, pp. 233–234] или [Davieson and MacKinnon, 1994, pp. 654–660]. Для SUR-модели уравнения приводимой итерационной процедуры описать намного проще, см. следующий параграф.

²Эта асимптотическая эквивалентность не сохраняется при переходе к системе нелинейных уравнений. Нелинейный FIML является более эффективной, чем нелинейный 3SLS. См. [Аптекуа, 1985, Section 8.2].

Идентифицируемость

Так как FIML — это ML-оценка, она обладает свойством эффективности, которое обсуждалось в параграфе 7.1. Чтобы проиллюстрировать это свойство, вернемся к структуре 8.1. Предположим, мы хотим репараметризовать уравнение предложения следующим образом:

$$p_t = \beta_{21}y_t - \bar{\beta}_2 - \bar{\beta}_{22}x_t + \varepsilon_t \quad (\text{предложение}). \quad (8.5.27)$$

Параметры старого уравнения, связывающего предложение $(\gamma_{21}, \beta_{21}, \beta_{22})$, и параметры нового уравнения предложения $(\bar{\beta}_2, \bar{\beta}_{21}, \bar{\beta}_{22})$ связаны взаимным однозначным соответствием

$$\bar{\beta}_{21} = 1/\gamma_{21}, \bar{\beta}_2 = -\beta_{21}/\gamma_{21}, \bar{\beta}_{22} = -\beta_{22}/\gamma_{21}. \quad (8.5.28)$$

Если $(\hat{\gamma}_{21}, \hat{\beta}_{21}, \hat{\beta}_{22})$ — это FIML-оценка параметров $(\gamma_{21}, \beta_{21}, \beta_{22})$ для системы уравнений, состоящей из уравнения спроса (8.5.4) и старого уравнения предложения (8.5.5), то тогда FIML-оценка, основанная на системе уравнений из (8.5.4) и нового уравнения предложения, численно совпадает со значениями вышеупомянутого отображения в точке $(\hat{\gamma}_{21}, \hat{\beta}_{21}, \hat{\beta}_{22})$. 2SLS-оценка (и 2SLS по той же причине) не обладает таким свойством. Вышеизложенная дискуссия о FIML-оценках может быть подытожена следующим образом.

Утверждение 8.1 (асимптотические свойства FIML): Рассмотрим систему из M уравнений (8.5.1). Пусть δ_0 — латентный вектор, включающий все коэффициенты этой системы. Пусть выполнены следующие предположения:

- выполнено ранговое условие идентификации: матрица $E(x_i x_{im}')$ имеет полный столбцовый ранг для всех $m = 1, 2, \dots, M$,
- $E(x_i x_i')$ — невырожденная матрица,
- указанная система из M уравнений может быть представлена в форме полной системы одновременных уравнений (8.5.8) с невырожденной матрицей Γ_0 ,
- $x_i | x_i \sim N(0, \Sigma_0)$, Σ_0 положительно определена,
- $(y_i, x_i) = (y_i, x_i)$,
- пространство параметров для (δ_0, Σ_0) является компактом, причем вектор постоянных параметров (δ_0, Σ_0) является его внутренней точкой.

Тогда

- (a) FIML-оценка $(\hat{\delta}, \hat{\Sigma})$, достигавшая максимум (8.5.20), состоятельна и асимптотически нормальна.

(5) асимптотическая дисперсия $\hat{\delta}$ равна (4.5.15), а постоянная базиса асимптотической дисперсии задается (4.5.17), где $\partial^{mb} = (m, b)$ -й элемент $\hat{\Sigma}^{-1}$.

(6) статистика отклонения правдоподобия (8.5.25) для тестирования сверхидентифицирующих ограничений асимптотически распределена как χ^2 с $KM = \sum_{m=1}^M L_m$ степенями свободы.

Кроме того, эти асимптотические свойства $\hat{\delta}$ сохраняются, даже если распределение $\varepsilon_i|x_i$ не является нормальным.

ML-оценивание SUR-модели

Модель SUR является частным случаем FIML, где x_{im} является подвектором x_i . Следовательно, это модель из M уравнений (8.5.1) с предопределенными регрессорами x_{im} . В отличие от многомерной модели регрессии, набор регрессоров может быть разным для разных уравнений системы. Легко доказать (см. контрольный вопрос 1), что линейные из двух зависимых переменных не могут быть одной и той же переменной. То есть $y_i = (M \times 1)$ в структурной форме (8.5.8) просто содержит M зависимых переменных. Кроме того, поскольку x_{im} не включает в себя эндогенных переменных, Γ_1 в структурной форме представляет собой единичную матрицу. Таким образом, целевая функция FIML (8.5.20) принимает вид:

$$Q_N(\delta, \Sigma) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma| - \frac{1}{2N} \sum_{i=1}^N [\Gamma'y_i + Bx_i]'\Sigma^{-1}[\Gamma'y_i + Bx_i], \quad (8.5.29)$$

где Γ — единичная матрица. (Мы продолжаем писать Γ , чтобы упростить сравнение с целевой функцией LIML, которая будет рассмотрена в следующем параграфе.) Как и в FIML, Σ , которая максимизирует целевую функцию при заданном δ , — это $\hat{\Sigma}(\delta)$, выражение для которой приведено в (8.5.21).

Для SUR-модели вклад по отношению к δ может быть записан довольно просто. Каждая строка матрицы $\Gamma y_i + Bx_i$ представляет собой $y_{im} + x'_{im}\delta$. Определим y размера $(MN \times 1)$ и Z размера $(MN \times \sum_{m=1}^M L_m)$, как в первом аналитическом упражнении к главе 4. Тогда

$$\sum_{i=1}^N [\Gamma y_i + Bx_i]'\Sigma^{-1}[\Gamma y_i + Bx_i] = (y - Z\delta)'(\Sigma^{-1} \otimes I_N)(y - Z\delta). \quad (8.5.30)$$

Поэтому

$$\frac{\partial Q_N(\delta, \Sigma)}{\partial \delta} = -\frac{1}{N} Z'(\Sigma^{-1} \otimes I_N)(y - Z\delta). \quad (8.5.31)$$

(Проплести то же самое для более общего случая FIML не так просто, так как, хотя (8.5.30) выписано для FIML также, вклад $\frac{\partial^2 Q_n}{\partial \delta^2}$ представляет собой более сложное выражение из-за наличия члена $\log |\Gamma^2|$ в FIML функции правдоподобия. Иными словами, если $|\Gamma|$ не зависит от δ , то тогда $\frac{\partial^2 Q_n}{\partial \delta^2}$ для FIML совпадает с (8.5.31). Приравняв это к нулю и решив относительно δ , получаем:

$$\hat{\delta}(\Sigma) = [Z'(\Sigma^{-1} \times I_n)Z]^{-1} Z'(\Sigma^{-1} \times I_n)y. \quad (8.5.32)$$

Под условием, что $\hat{\Sigma}$ — состоятельная оценка Σ_0 , $\hat{\delta}(\hat{\Sigma})$ представляет собой SUR-оценку δ_0 , введенную в главе 4.

Эти две функции $(\hat{\delta}(\Sigma), \hat{\Sigma}(\delta))$ определяют отображение из пространства параметров для (δ, Σ) в эквив. решение условий первого порядка для максимизации целевой функции $Q_n(\delta, \Sigma)$ является неподвижной точкой этого отображения. Эта неподвижная точка может быть вычислена с помощью **итерационного SUR** (*iterative SUR*). Пусть $(\hat{\delta}^{(j)}, \hat{\Sigma}^{(j)})$ — оценка (δ_0, Σ_0) на шаге j . Оценка на следующем шаге вычисляется по формуле:

$$\hat{\delta}^{(j+1)} = \hat{\delta}(\hat{\Sigma}^{(j)}), \quad \hat{\Sigma}^{(j+1)} = \hat{\Sigma}(\hat{\delta}^{(j+1)}). \quad (8.5.33)$$

Первая часть этой итерации состоит в получении SUR-оценки при заданной $\hat{\Sigma}^{(j)}$, в то время как вторая часть обновляет полученную ранее оценку Σ_0 с использованием текущей оценки δ_0 . Если этот процесс сойдется, то его результат — решение условий первого порядка.

Контрольные вопросы

1. (Вывод $f(y|x_i)$ из $f(x_i, y)$.) В этом упражнении мы хотим вывести функцию правдоподобия для наблюдения i для FIML-оценки модели. При этом мы будем использовать следующий результат из теории вероятностей:

Формула замены переменной: Пусть ε — непрерывный M -мерный случайный вектор с функцией плотности $f_\varepsilon(\varepsilon)$ и $g: \mathbb{R}^M \rightarrow \mathbb{R}^M$ — взаимно однозначное дифференцируемое отображение на открытом множестве S , при этом обратное отображение обозначается как $g^{-1}(\cdot)$. Пусть $M \times M$ матрица $J(\varepsilon)$ — матрица Якоби $\frac{\partial g_\varepsilon}{\partial \varepsilon}$ ($M \times M$ матрица частных производных, в которой элемент (j, i) равен $\frac{\partial g_{j,\varepsilon}}{\partial \varepsilon_i}$). Пусть $|J(\varepsilon)|$ не равен нулю в любой точке S . Тогда функция плотности $y = g(\varepsilon)$ задается как

$$f(y) = \frac{f_\varepsilon(g^{-1}(y))}{\det |J(g^{-1}(y))|}.$$

(Заметьте, что $|J(g^{-1}(y))|$ — это определитель, который может как быть, так и не быть положительным.) Используя этот факт, выведите функцию плотности $f(y|x_i)$ из $f(x_i, y)$. Также проверьте, что соответствующий

логарифм функция правдоподобия задается (8.5.16). **Указание:** Обратное отображение $g^{-1}(y_t)$ представляет собой левую часть (8.5.8). Следовательно, $J(\varepsilon_t) = J_0^{-1}$. Кроме того, $\log(\det(J)) = \frac{1}{2} \log(\det J^{-2})$.

2. (Инструменты за пределами системы.) Рассмотрим полную систему одновременных уравнений (8.5.8) с $\Gamma_0 \neq 0$. Пусть одна из K предопределенных переменных, скажем x_{TK} , не входит ни в одно из M уравнений. Пусть $\tilde{K}^*(y; x_T)$ — проекция ОН S_y на π . Покажите, что $\tilde{E}^*(y_{TK}; x_T) = \tilde{E}^*(y_{TK}; \tilde{x}_T)$, где $\tilde{x}_T = (x_{T1}, \dots, x_{TK-1})'$ (так что $x_T = (\tilde{x}_T', x_{TK})'$). (Как будет показано в аналитическом упражнении 4, включение x_{TK} в набор инструментов в дополнение к $\tilde{x}_T$ не влияет на асимптотическую дисперсию FIML-оценки δ_{TK} ; если при этом свойства в малых выборках, вероятно, ухудшатся.)
3. Вспомните, что в примере 8.1 y_{11} и y_{12} являются одной и той же переменной. В SUR модели такого произойти не может: никакие два элемента из $(y_{11}, y_{12}, \dots, y_{KM})$ не являются одной и той же переменной, если выполнены предположения модели. Докажите это. **Указание:** (Доказательство от противного.) Предположите, что y_{11} и y_{12} являются одной и той же переменной. Тогда

$$\varepsilon_{11} - \varepsilon_{12} = z_{11}'\delta_{0,1} - z_{12}'\delta_{0,1}.$$

Так как z_{11} и z_{12} — подвекторы x_T , левая часть может быть переписана как $w_1'\alpha$ для некоторого K -мерного вектора α . Используйте условие ортогональности $E(w_1\varepsilon_1') = 0$ и вырожденность $E(x_T x_T')$, чтобы показать, что $\alpha = 0$.

8.6. LIML

Определение LIML

Преимуществом метода FIML является то, что он позволяет использовать всю информацию, содержащуюся в полной системе одновременных уравнений. С другой стороны, это же является его слабым местом, поскольку оценка становится несостоятельной, если какая-либо часть системы неправильно специфицирована (этим же недостатком страдают все системные или совместные методы оценивания). Если вы уверены, что интересующее вас уравнение специфицировано правильно, а относительно прочих уравнений системы такой уверенности нет, возможно, будет разумнее применить методы оценивания отдельного уравнения (такие, как 2SLS). Остальная часть этого параграфа посвящена выводу ML-оценки, которая называется **оценкой максимального правдоподобия с ограниченной информацией** (*limited information maximum likelihood*), (LIML) и которая представляет собой ML аналог 2SLS-оценки.

Пусть нас интересует l -е уравнение системы из M уравнений. J_{lm} регрессоров z_{lm} являются либо эндогенными, либо предопределенными. Пусть $\tilde{y}_l$ — вектор, состоящий из M_m эндогенных переменных, а $\tilde{x}_l$ —

вектор, состоящий из M_m предопределенных параметров, где $M_m + M_{mm} + 1 = I_{mm} + \delta_{mm}$ разбивается аналогичным образом как $\delta_{mm} = (\delta'_{0m}, \delta'_{1m})'$. Тогда это уравнение может быть записано как

$$M_m + \sum_{i=1}^{M_m} \delta'_{im} \delta_{im} + C_m = \sum_{i=1}^{M_m} \hat{y}_i' \pi_{mi} + \sum_{i=1}^{M_m} \hat{x}_i' \delta_{im} + \varepsilon_m. \quad (8.6.1)$$

Очевидно, что это уравнение, рассматриваемое в эквивалент, представляет собой неслучайную систему из-за существования M_m включенных эндогенных переменных $\hat{y}_i$. Однако если это уравнение дополнить M_m уравнениями в приведенной форме для $\hat{y}_i$ из (8.5.9), то система из этих $1 + M_m$ уравнений будет полной системой одновременных уравнений. Чтобы объяснить эту идею более подробно, рассмотрим Π_0 — соответствующую матрицу коэффициентов приведенной формы системы уравнений. Возьмем из Π_0 соответствующие M_m столбцов и запишем уравнения в приведенной форме для M_m включенных в уравнения эндогенных переменных:

$$\hat{y}_i = \frac{\hat{\Pi}_0'}{(1 - \hat{\Pi}_0' K)(K \times 1) + M_m \times 1} x_i - \bar{y}_i. \quad (8.6.2)$$

Соединяя (8.6.1) и (8.6.2), получим систему из $1 + M_m$ уравнений:

$$\bar{\Gamma}_0 \begin{bmatrix} \bar{y}_0 \\ \bar{y}_1 \end{bmatrix} + \bar{M}_0 \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \bar{B}_0 \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \end{bmatrix}, \quad (8.6.3)$$

где

$$\bar{y}_0 = \begin{bmatrix} \delta_{0m} \\ \hat{y}_1 \end{bmatrix}_{(M_m+1) \times 1}, \quad \bar{y}_1 = \begin{bmatrix} \varepsilon_{1m} \\ \hat{y}_1 \end{bmatrix}_{(M_m+1) \times 1},$$

$$\bar{\Gamma}_0 = \begin{bmatrix} 0 & -\gamma'_{0m} \\ 0 & (1 - \hat{\Pi}_0') \end{bmatrix}_{(M_m+1) \times (M_m+1)}, \quad \bar{M}_0 = \begin{bmatrix} -\beta'_{0m} & 0' \\ -\hat{\Pi}_0' & 0' \end{bmatrix}_{(M_m+1) \times (M_m+1)}. \quad (8.6.4)$$

Здесь мы без труда могли бы показать, что включенные в уравнение предопределенные переменные $\hat{y}_i$ являются функциями K_m элементов x_i . Система уравнений (8.6.3) является полной системой $1 + M_m$ одновременных уравнений, поскольку

$$\bar{\Gamma}_0' = -I = 0. \quad (8.6.5)$$

В качестве примера рассмотрим график 8.3. Включенная эндогенная переменная уравнения заработной платы — это S_1 . Добавим к уравнению заработной платы уравнение в приведенной форме для S_1 регрессию S_1 на константу, $I(0)$ и $M \in D_1$. Матрица $\bar{\Gamma}_0$ для системы из двух уравнений имеет вид

$$\bar{\Gamma}_0 = \begin{bmatrix} 1 & 0 \\ 0 & \gamma_1' \end{bmatrix}.$$

Гипотетическое значение $\bar{\Gamma}$ для $\bar{\Gamma}_1$ имеет такую же структуру, как и $\bar{\Gamma}_0$ в (8.5.4). Поэтому $|\bar{\Gamma}| = 1$. Заменив в функции FIML (8.5.20) (δ, B) на $(\gamma_{0m}, \beta_{0m}, \bar{\Pi})$, Γ на $\bar{\Gamma}$ и Σ на $\bar{\Sigma}$ и используя условие $|\bar{\Gamma}| = 1$, мы получаем целевую функцию LIML:

$$Q_0(\gamma_{0m}, \beta_{0m}, \bar{\Pi}, \bar{\Sigma}) = -\frac{M_m}{2} \log(2\pi) - \frac{1}{2} \log|\bar{\Sigma}| - \\ - \frac{1}{2\pi} \sum_{i=1}^n (\bar{\Gamma} \bar{y}_i - Bx_i)' (\bar{\Sigma})^{-1} (\bar{\Gamma} \bar{y}_i - Bx_i), \quad (8.6.5)$$

где $\bar{\Sigma}$ — ковариационная матрица $\bar{\epsilon}_1$ размерности $(1 + M_m) \times (1 + M_m)$. Пусть $(\hat{\gamma}_{0m}, \hat{\beta}_{0m}, \hat{\Pi}, \hat{\Sigma}) \sim \text{FIML-оценка}$ этой системы. **LIML-оценка** (LIML estimator) для $\delta_{0m} = (\gamma'_{0m}, \beta'_{0m})'$ есть $(\hat{\gamma}'_{0m}, \hat{\beta}'_{0m})'$. Представленный способ вычисления LIML-оценки принадлежит Пагану [Pagan, 1979] и отличается от способа, предложенного Андерсоном и Рубином [Anderson and Rubin, 1949]. Поскольку LIML-оценивание является частным случаем FIML-оценивания, предложение 8.1 гарантирует состоятельность и асимптотическую нормальность LIML-оценки, даже в случае, когда не выполняются предположения о нормальности ошибок.

Вычисление LIML

Для случая SUR-модели мы показали, что ML-оценка может быть вычислена с помощью итерационной SUR-процедуры. Если внимательно взглянуть на доказательства, то можно заметить, что оно не зависит от того, является ли матрица Γ в целевой функции SUR (8.5.29) единичной матрицей. Единственным требованием является равенство определителя матрицы Γ константе. Поэтому та же аргументация может быть применена к целевой функции LIML (8.6.5). То есть мы можем получить LIML-оценку параметра δ_{0m} , применяя итерационную SUR-процедуру к системе (8.6.5), состоящей из $(1 + M_m)$ уравнений. В частности, если Σ_1 — известная матрица, то LIML-оценка является SUR-оценкой. Наш вывод о том, что SUR может быть использован для оценки коэффициентов при включенных эндогенных переменных, является удивительным. Вовсе нет. Причина этого в следующем: ошибка оценки $\hat{\delta}_{0m} - \delta_{0m}$ зависит не только от корреляции между включенными эндогенными переменными $\bar{y}_i$ и ошибкой ϵ_{0m} , но и от корреляции между $\bar{y}_i$ и $\bar{\epsilon}_i$. Эти две корреляции взаимно уничтожают друг друга. Этот факт проверяется в следующем разделе на простом примере.

Однако эта итерационная SUR-процедура обычно не используется для вычисления LIML-оценки, так как в случае LIML решение может быть найдено аналитически. Система одновременных уравнений (8.6.5) обладает двумя отличительными чертами. Первая — это особая структура $\bar{\Gamma}$, что мы уже отмечали. Вторая особенность состоит в том, что

на строки матрицы $\bar{D}$ не являются исключительные ограничения, соответствующие вычислительным переменным $\hat{y}_i$. Благодаря этим особенностям LIML-оценка $\hat{\delta}_{\text{LIML}} = (\hat{\beta}_{\text{LIML}}, \hat{\gamma}_{\text{LIML}})'$ и статистика максимального правдоподобия (8.5.25) могут быть вычислены в явном виде.

Чтобы записать в явном виде LIML-оценку и статистику отклонения правдоподобия, как необходимо ввести новые обозначения. Пусть $Z_m, \bar{Y}, \bar{X}, X$ и y_m — матрицы и вектор данных, которые соотносятся с $z_m, \hat{y}, \hat{x}, x$ и y_m соответственно следующим образом:

$$\begin{aligned} Z_m &\equiv \begin{bmatrix} z_{1m} \\ \vdots \\ z_{nm} \end{bmatrix}, \quad \bar{Y} \equiv \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{bmatrix}, \quad \bar{X} \equiv \begin{bmatrix} \hat{x}_1' \\ \vdots \\ \hat{x}_n' \end{bmatrix}, \\ X &\equiv \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}, \quad y_m \equiv \begin{bmatrix} y_{1m} \\ \vdots \\ y_{nm} \end{bmatrix}, \end{aligned} \quad (8.6.7)$$

и пусть M и $\bar{M}$ — матрицы, основанные на X и $\bar{X}$ соответственно:

$$M \equiv I_n - X(X'X)^{-1}X', \quad \bar{M} \equiv I_n - \bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}'. \quad (8.6.8)$$

Тогда LIML-оценка $\hat{\delta}_{\text{LIML}}$ может быть записана как

$$\hat{\delta}_{\text{LIML}} \equiv [Z_m'(I_n - kM)Z_m]^{-1} Z_m'(I_n - kM)y_m, \quad (8.6.9)$$

где k — наименьший характеристический корень

$$\bar{W}W^{-1}, \quad (8.6.10)$$

где $\bar{W}$ и W^{-1} — матрицы размерности $(1 - M_m) \times (1 - M_m)$, определяемые как

$$\bar{W} = [y_m - \bar{Y}]' \bar{M} [y_m - \bar{Y}], \quad W = [y_m - \bar{Y}]' M [y_m - \bar{Y}]. \quad (8.6.11)$$

Вывод можно найти, например, у Дэвидсона и Мак-Кенны (Davidson and MacKinnon, 1993, pp. 645–649). Статистика отклонения правдоподобия (8.5.25) для сверхидентифицирующей модели (8.5.25) может быть записана в следующем виде:

$$LN = -n \log K, \quad (8.6.12)$$

Так как в дополнительных M_m уравнениях (8.6.2) не содержится сверхлинейных ограничений, предельное число уравнений является единственно возможным сверхидентифицирующим ограничением, количество которых равно $K - L_m$. Поэтому из утверждения 8.1 следует, что эти статистики для проверки сверхидентифицирующих ограничений

асимптотически распределена как $\chi^2(K - L_m)$. Когда уравнение идентифицировано точно ($K = L_m$), LR статистики должна быть равна нулю. Кроме того, может быть показано, что в этом случае $k = 1$ (подробности см., например, в Дэвидсоне и Мак-Киннэне [Davidson and MacKinnon, 1993, p. 647]).

Если мы необязательно требуем, чтобы k было именно таким, как только что было определено, то оценка (8.6.9) называется **оценкой k -класса** (*k-class estimator*). Таким образом, LIML является оценкой k -класса с k , определенным выше. Изучение формулы для 2SLS-оценки в терминах матриц данных (см. (3.8.3')) в с. 259) немедленно показывает, что 2SLS-оценка является оценкой k -класса с $k = 1$, а OLS оценка — оценка k -класса с $k = 0$. Следовательно, LIML и 2SLS численно совпадают, когда уравнение точно идентифицировано (так что $k = 1$).

Мы уже показали, что LIML-оценка является состоятельной и асимптотически нормальной. Кроме того, используя явную формулу (8.6.9), можно показать, что асимптотическое распределение LIML-оценки δ_{lim} совпадает с асимптотическим распределением оценки 2SLS (см.: [Amemiya, 1985, Section 7.3.4]). Поэтому формула для асимптотической дисперсии LIML дается в (3.8.4), а ее состоятельная оценка — в (3.8.5).

Сравнение LIML и 2SLS

Поскольку LIML и 2SLS имеют одинаковое асимптотическое распределение, выбор в пользу одного из этих методов нельзя сделать на основании их асимптотических свойств. В конечных выборках LIML обладает свойством инвариантности, тогда как 2SLS таким свойством не обладает. Кроме того, из обширной литературы, посвященной конечным выборкам (см., например, Джадж и др. [Judge et al., 1985, Section 15.4]), можно заключить, что LIML предпочтительнее 2SLS. Ниже перечислены основные выводы из этой литературы. Однако в них предполагается, что предопределенные переменные x_i — фиксированные константы.

- Пусть ошибки нормально распределены. p -й момент 2SLS-оценки существует, если и только если $p < K - L_m + 1$, где K — количество предопределенных переменных и L_m — количество регрессоров в том уравнении. Таким образом, если уравнение идентифицировано точно (то есть если $K = L_m$), то у 2SLS-оценки даже не существует математического ожидания. Это утверждение справедливо, даже если ошибки имеют нормальное распределение, у которого все моменты конечны.
- У LIML-оценки нет конечных моментов, даже при предположении о нормальности ошибок.

- Однако большинство симулиций методов Монте-Карло (см., например, [Anderson, Kanitono, and Sawa, 1982]) демонстрирует, что LIML сходится к нормальному распределению быстрее, чем 2SLS.

Контрольные вопросы

1. Почему SUR может быть использован для оценивания уравнений в структурной форме? Чтобы понять, почему SUR может использоваться для оценивания коэффициентов при экзогенных эндогенных переменных, рассмотрим следующий простой пример:

$$\begin{cases} y_{1t} = \gamma_0 y_{2t} + \varepsilon_{1t} \\ y_{2t} = \beta_0 y_{1t} + v_t \end{cases}$$

где y_{it} — предопределенная переменная. Для простоты предположим, что

$$\Sigma_{\varepsilon} \equiv E \left(\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} & v_t \end{bmatrix} \right)$$

известна. Пусть $(\hat{\gamma}, \hat{\beta})$ — SUR-оценка истинных значений параметров (γ_0, β_0) .

- (a) Покажите, что

$$\begin{bmatrix} \hat{\gamma} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \gamma_0 \\ \beta_0 \end{bmatrix} + \begin{bmatrix} \sigma^{11} \frac{1}{n} \sum y_{2t}^2 & \sigma^{12} \frac{1}{n} \sum y_{2t} v_t \\ \sigma^{21} \frac{1}{n} \sum y_{1t} y_{2t} & \sigma^{22} \frac{1}{n} \sum v_t^2 \end{bmatrix}^{-1} \times \begin{bmatrix} \sigma^{11} \frac{1}{n} \sum y_{2t} \varepsilon_{1t} + \sigma^{21} \frac{1}{n} \sum y_{2t} v_t \\ \sigma^{21} \frac{1}{n} \sum y_{1t} \varepsilon_{1t} + \sigma^{22} \frac{1}{n} \sum y_{1t} v_t \end{bmatrix},$$

где σ^{ij} — элемент матрицы $\Sigma_{\varepsilon}^{-1}$ с номером (i, j) . Укажите: формулы для SUR-оценок $\hat{\gamma}$ и $\hat{\beta}$ (4.5.13), (4.5.13') и (4.5.14') на с. 309.

- (b) Покажите, что

$$\sigma^{11} \frac{1}{n} \sum_{t=1}^n y_{2t}^2 \rightarrow \sigma^{11} \frac{1}{n} \sum_{t=1}^n y_{2t}^2$$

сходящееся по вероятности к нулю. Укажите:

$$E(y_{2t} \varepsilon_{1t}) = E(\varepsilon_{1t} v_t) = \sigma_{12}, E(y_{2t} v_t) = E(v_t^2) = \sigma_{22}.$$

8.7. Сериально коррелированные наблюдения

При выводе асимптотической дисперсии ML-оценок мы до сих пор предполагали независимость и одинаковую распределенность наблюдений. Этот параграф посвящен ML-оцениванию в случае, когда наблюдения сериально коррелированы

Два вопроса

Когда между наблюдениями имеется сериальная корреляция, возникают два важных вопроса. Первый: является ли МЛ-оценка, минимизирующая предположения о независимости и одинаковой распределенности наблюдений, состоятельной и асимптотически нормальной, если наблюдения сериально коррелированы? Предположим, что вектор наблюдений $\mathbf{w}_i$ — стационарный и эргодический, но обязательно *i.i.d.* Рассмотрим МЛ-оценку, которая максимизирует

$$Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log f(\mathbf{w}_i; \theta), \quad (8.7.1)$$

где $\log f(\mathbf{w}_i; \theta)$ — логарифмическое правдоподобие для наблюдения i . Если бы наблюдения были независимыми и одинаково распределенными, то эта целевая функция представляла бы собой умноженное на $1/n$ логарифмическое правдоподобие выборки $(\mathbf{w}_1, \dots, \mathbf{w}_n)$. Даже при том, что логарифмическая функция правдоподобия для наблюдения i , $\log f(\mathbf{w}_i; \theta)$, специфицирована правильно, эта целевая функция специфицирована неправильно, поскольку ошибочно предполагает, что $\{\mathbf{w}_i\}$ — *i.i.d.* Так как целевая функция специфицирована неправильно, МЛ-оценка, которая максимизирует эту целевую функцию, представляет собой квази-МЛ-оценку. Является ли эта квази-МЛ-оценка, базирующаяся на неверной предпосылке об отсутствии сериальной корреляции, состоятельной и асимптотически нормальной?

Мы в действительности уже дали ответ на этот первый вопрос в предыдущей главе. Из утверждений 7.5 и 7.6 следует, что квази-МЛ-оценка является состоятельной, так как она является частным случаем М-оценки с $\psi(\mathbf{w}_i; \theta) = \log f(\mathbf{w}_i; \theta)$, равной логарифмическому правдоподобию для наблюдения i . Для установления асимптотической нормальности этой оценки мы можем воспользоваться утверждением 7.5, в котором перечислены условия, при которых М-оценка асимптотически нормально распределена. Простое воспроизведение заключения утверждения 7.6, получаем, что выражение для асимптотической дисперсии задается как

$$\text{Avar}(\hat{\theta}) = (E[H(\mathbf{w}_i; \theta_0)])^{-1} \Sigma(E[H(\mathbf{w}_i; \theta_0)])^{-1}, \quad (8.7.2)$$

где θ_0 — истинное значение параметров θ , $H(\mathbf{w}_i; \theta)$ — первая для наблюдения i и Σ — дисперсия или дисперсия Бласса для наблюдения i , вычисленная в точке θ_0 , $\psi(\mathbf{w}_i; \theta_0)$. Состоятельная оценка этой асимптотической дисперсии может быть найдена как

$$\widehat{\text{Avar}}(\hat{\theta}) = \left\{ \frac{1}{n} \sum_{i=1}^n H(\mathbf{w}_i; \hat{\theta}) \right\}^{-1} \hat{\Sigma} \left\{ \frac{1}{n} \sum_{i=1}^n H(\mathbf{w}_i; \hat{\theta}) \right\}^{-1}, \quad (8.7.3)$$

где $\hat{\Sigma}$ — состоятельная оценка Σ . Любой «непараметрический» метод оценивания, обсуждавшийся в главе 6 (например, VARHAC), может быть использован для расчета $\hat{\Sigma}$ по оцененной последовательности $\{s(w_t, \hat{\theta})\}$. Кроме того, нет необходимости в параметрической спецификации сериальной корреляции в $\{s(w_t, \theta_0)\}$.

Второй вопрос: как выглядит функция правдоподобия, которая учитывает наличие сериальной корреляции, и каковы свойства «истинной» ML-оценки, которая максимизирует эту целевую функцию? Этот вопрос может быть адресован как условному, так и безусловному методу максимального правдоподобия. Говоря об условном ML, мы разбиваем вектор наблюдений w_t на две группы: зависимая переменная y_t и регрессоры x_t и исследуем условное распределение $(y_1, \dots, y_n)$ при условии $(x_1, \dots, x_n)$. Параметризация этого условного распределения довольно проста, когда $\{w_t\}$ независимы и одинаково распределены, поскольку это распределение может быть записано как произведение по всем t условных распределений y_t при условии x_t . На самом деле случаи независимых и одинаково распределенных величин — это то, что мы рассматривали в предыдущих главах. В случае, когда $\{w_t\}$ сериально коррелированы, исследователь не знает динамический взаимосвязи между y_t и x_t достаточно хорошо, чтобы записать условное распределение¹. В оставшейся части этого параграфа мы будем рассматривать только примеры безусловного ML-оценивания с сериально коррелированными наблюдениями.

Безусловный ML для зависимых наблюдений

Мы начинаем с простейшего одномерного случая. Пусть $(y_1, y_2, \dots, y_n)$ — имеющаяся в распоряжении выборка. (Для удобства обозначений мы предполагаем, что в выборку включается наблюдение для $t = 0$.) Когда между наблюдениями существует сериальная корреляция, функция плотности выборки больше не может быть записана как произведение индивидуальных функций плотности. Однако она может быть записана в терминах функций условной плотности. Совместная функция плотности (y_1, y_2) может быть записана следующим образом:

$$f(y_1, y_2) = f(y_1) f(y_2|y_1). \quad (8.7.4)$$

Аналогично совместная функция плотности для (y_0, y_1, y_2) :

$$f(y_0, y_1, y_2) = f(y_2|y_1, y_0) f(y_1, y_0). \quad (8.7.5)$$

¹Обсуждение условного ML в случаях, когда полная спецификация динамической взаимосвязи невозможно или нежелательна, см. у Вулдриджа [Wooldridge, 1994, Section 3]. Обсуждение возможных ограничений, накладываемых на динамические взаимосвязи между переменными (такими как «границность по Грейнджеру» и «слабая экзогенность») см., например, в работе Давидсона и МакКиннона [Davidson and MacKinnon, 1993, Section 18.2].

Подстановка (8.7.4) в (8.7.5), получаем:

$$f(y_0, y_1, y_2) = f(y_2|y_1, y_0)f(y_1|y_0)f(y_0). \quad (8.7.6)$$

Повторяя такую последовательную подстановку, в итоге получим:

$$f(y_0, y_1, \dots, y_n) = \prod_{i=1}^n f(y_i|y_{i-1}, \dots, y_1, y_0)f(y_0). \quad (8.7.7)$$

Если условная функция плотности $f(y_i|y_{i-1}, \dots, y_1, y_0)$ и безусловная функция плотности $f(y_0)$ могут быть параметризованы с помощью конечного вектора параметра θ , то тогда умножение на $1/n$ логарифмическое правдоподобие выборки $(y_0, y_1, \dots, y_n)$ может быть записано как

$$\frac{1}{n} \sum_{i=1}^n \log f(y_i|y_{i-1}, \dots, y_1, y_0; \theta) + \frac{1}{n} \log f(y_0; \theta). \quad (8.7.8)$$

Оценка максимального правдоподобия $\hat{\theta}$ истинного вектора параметров θ_0 — это такие значения θ , которые максимизируют эту логарифмическую функцию правдоподобия. Такую оценку мы будем называть **точной ML-оценкой** (*exact ML estimator*).

ML-оценка процесса AR(1)

В параграфе 6.2 мы довольно подробно рассмотрели процесс авторегрессии первого порядка AR(1). Если мы предположим, что ошибки независимы и нормальны, мы получим гауссовский AR(1)-процесс (*Gaussian AR(1) process*):

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t, \quad (8.7.9)$$

где $\varepsilon_t \sim i.i.d. N(0, \sigma_\varepsilon^2)$. Мы предположим, что условие стационарности процесса $|\alpha_1| < 1$ выполнено. Поэтому $\{y_t\}$ — эргодический стационарный процесс, в соответствии с утверждением 6.1. Из параграфа 6.2 мы знаем, что безусловная математическая expectation y равна $\alpha_0/(1 - \alpha_1)$, а безусловная дисперсия равна $\sigma_\varepsilon^2/(1 - \alpha_1^2)$. Кроме того, поскольку y может быть представлен как взвешенное среднее $(\varepsilon_1, \varepsilon_2, \dots)$, его распределение нормально при предположении о нормальности ε_t . Поэтому

$$y_0 \sim N\left(\frac{\alpha_0}{1 - \alpha_1}, \frac{\sigma_\varepsilon^2}{1 - \alpha_1^2}\right). \quad (8.7.10)$$

Кроме того, из (8.7.9) и предположения о нормальности ошибок несложно получить сразу, что

$$y_0(y_1, \dots, y_n) \sim N(y_0 | \alpha_0, \sigma_\varepsilon^2). \quad (8.7.11)$$

Далее, используя формулу (8.7.7), совместную функцию плотности $(y_0, y_1, \dots, y_n)$ для гипотетического значения параметров $\theta \equiv (c, \phi, \sigma^2)$ можно записать как

$$f(y_0, y_1, \dots, y_n; \theta) = \prod_{i=1}^n \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (y_i - c - \phi y_{i-1})^2 \right] \right\} \times \\ \times \frac{1}{\sqrt{2\pi\frac{\sigma^2}{1-\phi^2}}} \exp \left[-\frac{(y_0 - \frac{c}{1-\phi})^2}{2\frac{\sigma^2}{1-\phi^2}} \right]. \quad (8.7.12)$$

Прологарифмировав и разделив обе части выражения на n , получим целевую функцию для точной ML-оценки

$$\hat{Q}_n(\theta) = \frac{1}{n} \sum_{i=1}^n \left\{ -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (y_i - c - \phi y_{i-1})^2 \right\} + \\ + \frac{1}{n} \left\{ -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \left(\frac{\sigma^2}{1-\phi^2} \right) - \left[\frac{(y_0 - \frac{c}{1-\phi})^2}{2\frac{\sigma^2}{1-\phi^2}} \right] \right\}. \quad (8.7.13)$$

Точная ML-оценка $\hat{\theta}$ параметров AR(1)-процесса — это такое значение θ , которое максимизирует эту целевую функцию. В предположении, что решение внутреннее, эта оценка является решением условий первого порядка (уравнений правдоподобия), полученных путем приравнивания $\frac{\partial \hat{Q}_n}{\partial \theta} = 0$ к нулю. Они представляют собой систему нелинейных уравнений по θ , так что поиск $\hat{\theta}$ требует применения итерационных алгоритмов, таких как рассмотренные в параграфе 7.5.

Условное ML-оценивание AR(1)-процессов

Истинным недостатком уравнений правдоподобия является логарифм правдоподобия $\log$. Как альтернативу, мы можем эту составляющую и найдем максимум следующей функции:

$$Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n \left\{ -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (y_i - c - \phi y_{i-1})^2 \right\}. \quad (8.7.14)$$

Используя соотношение $f(y_0, y_1, \dots, y_n) = f(y_1, \dots, y_n | y_0) f(y_0)$ и (8.7.12), мы видим, что эта целевая функция равна умноженной на $1/n$ логарифмическому правдоподобию, условному относительно первого наблюдения y_0 .

$$f(y_1, y_2, \dots, y_n; \text{усл } \theta) = \prod_{i=1}^n \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (y_i - c - \phi y_{i-1})^2 \right] \right\}, \quad (8.7.15)$$

Пусть $\hat{\theta}$ — значение параметра θ , которое максимизирует условное логарифмическое правдоподобие (8.7.14) (при условии y_0). Эта ML-оценка будет поэтому называться **y_0 -условной ML-оценкой** (*y_0 -conditional ML estimator*). То, что здесь оценка ищется при условии y_0 , концептуально отличает эту оценку от ML-оценок, рассмотренных в предыдущих параграфах, где оценка выводилась при условии x_1 .

Прежде чем анализировать связь между двумя этими условными ML-оценками (точнее ML-оценка $\hat{\theta}$ и y_0 -условная ML-оценка $\hat{\theta}$), мы выведем два асимптотических свойства y_0 -условной ML-оценки. Первый вывод основан на поиске выражения для оценки в явном виде. Вспомним из параграфа 1.5 и примера 7.2, что целевая функция ML линейной регрессионной модели $y_t = x_t' \beta_0 + \varepsilon_t$ для независимых одинаково распределенных наблюдений представляется как

$$\frac{1}{n} \sum_{t=1}^n \left\{ -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (y_t - x_t' \beta)^2 \right\}. \quad (8.7.16)$$

В параграфе 1.5 было показано (и это очень легко показать), что ML-оценка β_0 совпадает с OLS-оценкой, а ML-оценка σ_0^2 равна сумме квадратов остатков, деленной на размер выборки. Теперь, при $x_t = (1, y_{t-1})'$ и $\beta_0 = (\alpha_0, \phi_0)'$, целевая функция линейной регрессионной модели упрощается до $Q_n(\theta)$ (8.7.14) — условного логарифмического правдоподобия для AR(1)-процесса. Поэтому алгебраически условная ML-оценка $(\hat{\alpha}, \hat{\phi})$ для (α_0, ϕ_0) численно совпадает с OLS-оценкой коэффициентов регрессии y_t на y_{t-1} , и условная ML-оценка $\hat{\sigma}^2$ для σ_0^2 равна сумме квадратов остатков, деленной на n . Что касается асимптотических свойств оценки, то мы показали в параграфе 6.4, что $(\hat{\alpha}, \hat{\phi})$ — состоятельная и асимптотически нормальная оценка с асимптотической дисперсией, указанный в утверждении 6.7 для $\hat{\mu} = \mu$, и что $\hat{\sigma}^2$ — состоятельная оценка σ_0^2 . Поэтому для асимптотической нормальности условной ML-оценки $(\hat{\alpha}, \hat{\phi})$, которая максимизирует y_0 -условную функцию правдоподобия для нормально распределенных ошибок, предположение о нормальном распределении не требуется.

Второй вывод менее прямой, но, несмотря на это, более instructive. Как только что было замечено, целевая функция $Q_n(\theta)$ совпадает с (8.7.16) при $x_t = (1, y_{t-1})'$ и $\beta_0 = (\alpha_0, \phi_0)'$. Поскольку $\{y_t, x_t\}$ только стационарны и эргодичны, но необязательно независимы и одинаково распределены, условная ML-оценка параметров $(\hat{\alpha}, \hat{\phi}, \hat{\sigma}^2)$ может рассматриваться как квази-ML-оценка, рассмотренная в утверждении 7.6. В примере 7.8 было проверено, что эта целевая функция удовлетворяет условиям утверждения 7.5, если $E(x_t x_t')$ — невырожденная матрица ($x_t = (1, y_{t-1})'$). В параграфе 6.4 мы показали, что это условие невырожденности выполнено для процесса AR(1), если $\sigma_0^2 > 0$. Таким обра-

зом, оценка условного ML состоятельна. Для доказательства асимптотической нормальности рассмотрим условную ML-оценку как M-оценку с $w_t = (y_t, y_{t-1})'$ и

$$m(w_t; \theta) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (y_t - \mu - \phi y_{t-1})^2. \quad (8.7.17)$$

Следовательно, подходящей теоремой здесь будет утверждение 7.8. Объясним как $s(w_t; \theta)$ — вклад наблюдения t , и $H(w_t; \theta)$ — гессиан для наблюдения t , связанный с этой m функцией, и применим во внимание условия (1)–(5) утверждения 7.8. Из обсуждения в примере 7.10 следует, что все условия, кроме (3), выполнены, если $E(x_t x_t')$ — невырожденная матрица (где $x_t = (1, y_{t-1})'$). Как мы только что отметили, это условие выполнено при $\sigma_y^2 > 0$. Нам осталось рассмотреть только выполнение условия (3): $\frac{1}{n} \sum_{t=1}^n s(w_t; \theta_0) \rightarrow_d N(0, \Sigma)$. В модели AR(1) $f(y_t | y_{t-1}, \dots, y_1, y_0) = f(y_t | y_{t-1})$. Как вы сможете показать в контрольном вопросе 1, это специфическое свойство совместной функции плотности указывает на то, что последовательность $\{s(w_t; \theta_0)\}$ является мартингал-разностью. Так как $s(w_t; \theta_0)$ является функцией от $w_t = (y_t, y_{t-1})'$ и $\{y_t\}$ стационарна и эргодична, последовательность $\{s(w_t; \theta_0)\}$ также стационарна и эргодична. Используя CLT для эргодических стационарных мартингал-разностей из главы 2, можно заключить, что условие (3) выполняется с $\Sigma = E[s(w_t; \theta_0) s(w_t; \theta_0)']$.

$$\Sigma = E[s(w_t; \theta_0) s(w_t; \theta_0)']. \quad (8.7.18)$$

Наконец, легко показать, что для наблюдений t выполняется равенство для информационных матриц $-E[H(w_t; \theta_0)] = E[s(w_t; \theta_0) s(w_t; \theta_0)']$ (доказательство было приложено в примере 7.10). Подставляя это выражение в (8.7.18) и (8.7.2), мы заключаем, что асимптотическая дисперсия условной ML-оценки $\hat{\theta}$ истинного значения параметра θ_0 дается выражением:

$$\text{Avar}(\hat{\theta}) = -(E[H(w_t; \theta_0)])^{-1} = -(E[s(w_t; \theta_0) s(w_t; \theta_0)'])^{-1}. \quad (8.7.19)$$

То есть обычные выводы для правильно специфицированной ML-оценки для независимых и одинаково распределенных наблюдений применимы и для условной ML-оценки, которая максимизирует правильно специфицированную функцию правдоподобия (8.7.14) для зависимых наблюдений¹.

Разность между (8.7.13) и (8.7.14) представляет логарифмическое правдоподобие $\ln$, деленное на n . Если размер выборки n достаточно велик, то первое наблюдение y_0 вносит пренебрежимо малый вклад

¹ Это доказательство можно переписать и в более общей форме, как авторегрессионные процессы. См.: Lemma 5.2 в [Wooldridge, 1994].

в функцию правдоподобия выборки. Точная ML-оценка и условная ML-оценка имеют одинаковое асимптотическое распределение при условии, что $|\phi_0| < 1$ (см., например, [Fuller, 1996, Sections 8.1 и 8.4]). По этой причине в большинстве случаев параметры авторегрессии оцениваются OLS (или условным ML), а не точным ML.

Условное ML-оценивание процессов $AR(p)$ и $VAR(p)$

Гауссовский $AR(p)$ -процесс (*Gaussian $AR(p)$ process*) может быть записан как

$$y_t = c_0 + \phi_{01}y_{t-1} + \phi_{02}y_{t-2} + \dots + \phi_{0p}y_{t-p} + \varepsilon_t, \quad (8.7.20)$$

где $\varepsilon_t \sim i.i.d.N(0, \sigma_\varepsilon^2)$. Условное ML-оценивание параметров процесса $AR(p)$ абсолютно аналогично случаю $AR(1)$. Если имеется выборка $(y_{-p+1}, y_{-p+2}, \dots, y_0, y_1, \dots, y_n)$, то тогда умноженное на $1/n$ логарифмическое правдоподобие этой выборки при условии $(y_{-p+1}, y_{-p+2}, \dots, y_0)$ задается соотношением (8.7.16), где

$$\mathbf{y}_t = (y_{t-1}, \dots, y_{t-p})' \text{ и } \boldsymbol{\beta}_0 = (c_0, \phi_{01}, \phi_{02}, \dots, \phi_{0p})'.$$

Поэтому условная ML-оценка (условная относительно $(y_{-p+1}, \dots, y_0)$) параметров $(c_0, \phi_{01}, \phi_{02}, \dots, \phi_{0p})$ численно совпадает с OLS-оценкой коэффициентов регрессии y_t на константу и p лаговых значений y_t . Условная ML-оценка σ_ε^2 равна сумме квадратов остатков, деленной на размер выборки. Из утверждения 6.7 следует, что условная ML-оценка коэффициентов является состоятельной и асимптотически нормальной, если выполняются условия стационарности (то есть когда корни характеристического полинома $1 - \phi_{01}z - \dots - \phi_{0p}z^p = 0$ превосходят 1 по абсолютному значению). Как и в случае процесса $AR(1)$, для выполнения свойств состоятельности и асимптотической нормальности не требуется предположения о нормальном распределении ошибок. Точная ML-оценка и условная ML-оценка имеют одинаковое асимптотическое распределение, если условие стационарности выполнено.

Эти результаты можно обобщить на случай векторного процесса. **Гауссовская $VAR(p)$** (*Gaussian $VAR(p)$*) (авторегрессия порядка p) может быть записана как

$$y_t = c_0 + \Phi_{01}y_{t-1} + \dots + \Phi_{0p}y_{t-p} + \varepsilon_t, \quad (8.7.21)$$

(47 > 1)

где $\varepsilon_t \sim N(0, \Omega_{11})$. Целевая функция в условном ML-оценивании равна

умноженному на $1/n$ условному логарифмическому правдоподобию:

$$Q_n(\theta) = -\frac{M}{2} \log(2\pi) + \frac{1}{2} \log(|\Omega^{-1}|) - \\ - \frac{1}{2n} \sum_{i=1}^n \{(y_i - \Pi'x_i)' \Omega^{-1} (y_i - \Pi'x_i)\}, \quad (8.7.22)$$

где

$$\Pi' \equiv [c \quad \Phi_1 \quad \Phi_2 \quad \dots \quad \Phi_p], \quad x_i' \equiv \begin{bmatrix} 1 \\ y_{i-1} \\ y_{i-2} \\ \vdots \\ y_{i-p} \end{bmatrix}. \quad (8.7.23)$$

Это совпадает с логарифмической функцией для МЛ-оценивания кластерной регрессионной модели, рассмотренной в параграфе 8.4. Поэтому условная МЛ-оценка параметров $(\hat{c}_n, \hat{\Pi}_n)$ численно совпадает с оценкой, получаемой применением OLS отдельно к каждому уравнению. Она является состоятельной и асимптотически нормальной, если матрицы коэффициентов $x_i x_i'$ образуют условную стационарность: все корни характеристического уравнения $|I_p - \Phi_1 z - \dots - \Phi_p z^p| = 0$ превосходят 1 по абсолютному значению. Этот результат же требует нормальности распределения вектора ε_i .

Контрольные вопросы

1. (Вклад является мартигал-разностью.) Рассмотрим гауссовский $AR(1)$ -процесс (8.7.9). Пусть $s(w_t; \theta)$ — функция вклада наблюдения t (градиент (8.7.17)), где $w = (x_t, w_{t-1})'$ и $\theta = (c, \omega, \sigma^2)'$.

(a) Покажите, что

$$s(w_t; \theta) = \begin{bmatrix} \frac{1}{\sigma^2} x_t \cdot (y_t - x_t' \beta) \\ \frac{1}{2\sigma^2} \cdot \frac{1}{2\sigma^2} (y_t - x_t' \beta)^2 \end{bmatrix},$$

где $\beta = (c, \omega)'$ и $x_t = (1, y_{t-1})'$.

(b) Покажите, что

$$E[s(w_t; \theta) | y_{t-1}, y_{t-2}, \dots] = 0.$$

Указание: При $\theta = \theta_0$, $y_t - x_t' \beta = \varepsilon_t$.

- (c) Покажите, что $\{s(w_t; \theta_0)\}$ является martingal-разностью. Укажите, что $s(w_t; \theta_0)$ является (нормальной) функцией от $(y_{t-1}, y_{t-2})'$. Следовательно, $s(w_{t-j}; \theta_0)$ ($j \geq 1$) можно вычислить на основе $(y_{t-1}, y_{t-2}, \dots, 1)$.
- (d) Можно ли перенести результат пункта (c) на случай гауссовского $AR(p)$ -процесса? [Ответ: Да.]

Набор задач для главы 8

Аналитические упражнения

1. Покажите, что

$$\frac{1}{n} \sum_{i=1}^n (y_i - \Pi' x_i)(y_i - \Pi' x_i)' \geq \left| \frac{1}{n} \sum_{i=1}^n \tilde{v}_i \tilde{v}_i' \right|, \quad (1)$$

где $\tilde{v}_i = y_i - \tilde{\Pi}' x_i$ и $\tilde{\Pi}$ определена в (8.4.5). (Так как $\tilde{v}_i$ не зависит от Π , по построению, это неравенство показывает, что левая часть достигает минимума при $\Pi = \tilde{\Pi}$.) **Указание:** Используйте стратегию «добавить и вычитать», чтобы вывести соотношение:

$$y_i - \Pi' x_i = \tilde{v}_i - (\tilde{\Pi} - \Pi)' x_i.$$

Тогда

$$\begin{aligned} \sum_{i=1}^n (y_i - \Pi' x_i)(y_i - \Pi' x_i)' &= \\ &= \sum_{i=1}^n \tilde{v}_i \tilde{v}_i' + (\tilde{\Pi} - \Pi)' \sum_{i=1}^n x_i \tilde{v}_i' + \sum_{i=1}^n \tilde{v}_i x_i' (\tilde{\Pi} - \Pi) + \\ &+ (\tilde{\Pi} - \Pi)' \left(\sum_{i=1}^n x_i x_i' \right) (\tilde{\Pi} - \Pi) = \\ &= \sum_{i=1}^n \tilde{v}_i \tilde{v}_i' + (\tilde{\Pi} - \Pi)' \left(\sum_{i=1}^n x_i x_i' \right) (\tilde{\Pi} - \Pi) \\ &\quad \text{(так как } \sum_{i=1}^n x_i \tilde{v}_i' = 0). \end{aligned}$$

Далее воспользуйтесь следующим результатом из матричной алгебры:

Пусть A и B — положительно полуопределенные матрицы одинакового размера. Тогда $|A + B| \geq |A|$.

2. ($\hat{\Omega}; \Pi$) положительно определена.) В условиях многомерной регрессионной модели покажите, что $\hat{\Omega}(\Pi)$, определенная в (8.4.10), является положительно определенной для любой заданной Π . **Указание:** Вычтите соотношение $y_i - \Pi' x_i = v_i + (\Pi_0 - \Pi)' x_i$. Так как $\{y_i, x_i\}$ — последовательность независимых одинаково распределенных величин, $\hat{\Omega}(\Pi)$ сходится почти наверное к $E[(y_i - \Pi' x_i)(y_i - \Pi' x_i)']$. Покажите, что

$$E[(y_i - \Pi' x_i)(y_i - \Pi' x_i)'] = \Omega_0 + (\Pi_0 - \Pi)' E(x_i x_i') (\Pi_0 - \Pi).$$

Затем используйте матричное неравенство, упомянутое в предыдущем упражнении. Ω_0 положительно определена по предположению.

- 3 (Дополнительное, идентификация δ в FIML.) Как было замечено в главе 4, с. 566 параграфа 8.5, FIML-оценка $\hat{\delta}_n$ является экстремальной оценкой с целевой функцией

$$Q_n(\delta) = -[\tilde{L}(\delta)], \quad (2)$$

где

$$\tilde{L}(\delta) \equiv \frac{1}{n} \sum_{i=1}^n \{y_i + \Gamma' A x_i (y_i - \Gamma' A x_i)\}'. \quad (3)$$

Пусть

$$Q_0(\delta) \equiv \lim_{n \rightarrow \infty} Q_n(\delta). \quad (4)$$

Чтобы доказать состоятельность этой экстремальной оценки, необходимо проверить два условия утверждения 7.1 из предыдущей главы. Воспроизведем эти условия, используя текущие обозначения:

идентифицируемость: $Q_0(\delta)$ достигает единственного максимума на компактном пространстве параметров при $\delta = \delta_0$.

равномерная сходимость: $Q_n(\cdot)$ равномерно сходится по вероятности к $Q_0(\cdot)$.

В этом упражнении мы доказываем, что ранговое условие идентифицируемости $\text{rk}(\alpha - \alpha'_{\min})$ имеет полный столбчатый ранг для всех m) эквивалентно условию идентифицируемости указанной выше экстремальной оценки. В процессе доказательства все три предложения относительно FIML-оценок, которые мы поставили в начале, состоять из ранее известных.

Чтобы начать обсуждение, давайте переформулируем обозначения. Воспроизведем m -е уравнение структурной формы из текста:

$$y_{it} = \mathbf{z}_{it}' \gamma + \alpha_{it} + \varepsilon_{it}, \quad (5)$$

$$\mathbf{z}_{it} = (\mathbf{z}_{it1}, \mathbf{z}_{it2}, \mathbf{z}_{it3})'$$

Регрессоры $\mathbf{z}_{it}$ состоят из M_m эндогенных переменных и K_m предопределенных, где $M_m + K_m = L_m$. Обозначим

$\mathbf{B}_m$: матрица, выбирающая M_m эндогенных переменных из $\mathbf{z}_{it}$, $\mathbf{B}_m' \in M_m$

$\mathbf{C}_m$: матрица, выбирающая K_m предопределенных переменных из $\mathbf{z}_{it}$, $\mathbf{C}_m' \in K_m$

Для администрции, S_1 и C_1 для примера 8.1 в $y = (y_1, y_2)'$ и $x_1 = (1, x_1, x_2)'$ определяются как

$$S_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, C_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Таким образом, регрессоры m -го уравнения могут быть представлены как

$$z'_m = \left(y'_1 S'_m + x'_2 C'_m \right).$$

(а) Покажите, что

$$E(z_m z'_m) = E(y_1 x'_1) \underbrace{\begin{bmatrix} \Pi_0 & S'_m & C'_m \\ (N \times M) & (M \times M) & (K \times K) \end{bmatrix}}_{(K \times M)}. \quad (6)$$

Указание: Используйте приведенную форму (8.3.9) и тот факт, что $U(x, y) = 0$. Так как умножение на невырожденную матрицу (здесь $E(x_1 x'_1)$) не меняет ранга, из уравнения (6) следует, что ранг $E(z_m z'_m)$ идентичен рангу $E(x_1 x'_1)$. Таким образом, ранг $\begin{bmatrix} \Pi_0 S'_m & C'_m \end{bmatrix}$ равен L_m .

(б) Покажите, что

$$\lim_{n \rightarrow \infty} \hat{\Omega}(\delta) = \Gamma_0^{-1} \Sigma_0 (\Gamma_0^{-1})' = \left[\Pi_0' + \Gamma^{-1} B' E(x_1 x'_1) \right] \Pi_0 + \Gamma^{-1} B'. \quad (7)$$

где $\Pi_0' \equiv -\Gamma_0^{-1} B_0$. **Указание:** Так как (y_1, x_1) — i.i.d., указанный предел по вероятности равен $E[(y_1 + \Gamma^{-1} B x_1)(y_1 + \Gamma^{-1} B x_1)']$. Кроме того,

$$y_1 + \Gamma^{-1} B x_1 = e_1 + (\Pi_0' - \Gamma^{-1} B') x_1.$$

(в) Покажите, что $\lim_{n \rightarrow \infty} |\hat{\Omega}(\delta)|$ достигает минимума, только если $\Gamma \Pi_0' + B = 0$. **Указание:** Так как $E(x_1 x'_1)$ положительно определенно, второе слагаемое в правой части уравнения (7) равно нулю, только если $\Pi_0' + B'(\Gamma^{-1})' = 0$. Также используйте результат из предыдущей главы, упомянутый в объяснении (а) выше.

Потому условие $\Gamma \Pi_0' + B = 0$ необходимо для существования минимума $|\hat{\Omega}(\delta)|$ достигается, единственного максимума из пространства также не существует, даже если $(\Gamma, B) = (\Gamma_0, B_0)$ — единственное решение $\Gamma \Pi_0' + B = 0$, рассматриваемой как система однородных уравнений для (Γ, B) .

- (ii) Пусть α'_m размера $(1 \times (M + K))$ — m -я строка матрицы $[T : B]$ и κ'_m — вектор размера $(1 \times (M + K))$, элементы которого, соответствующие y_m , равны единице, а все другие равны нулю. Для иллюстрации α' для примера 8.1, где $y_k = (y_1, y_2)'$ и $x_1 = (1, 0, 0, 0)'$ равен $(0, 1, 0, 0, 0)$. Проверьте для примера 8.1 общий результат о том, что

$$\alpha'_m = \kappa'_m + \delta'_m \begin{bmatrix} S_{22} & 0 \\ (S_{21} \times M) & (S_{11} \times K) \\ 0 & C_m \\ (K_m \times M) & (C_m \times K) \end{bmatrix}, \quad (8)$$

- (iii) Перепишите $\Pi \Pi'_0 + B = 0$ как

$$[T : B] \begin{bmatrix} \Pi_0 \\ I_K \end{bmatrix} = 0. \quad (9)$$

Покажите, что m -я строка $[T \Pi'_0 + B] = 0$ может быть записана в виде

$$\delta'_m \begin{bmatrix} S_{22} \Pi_0 \\ C_m \end{bmatrix} = \pi'_m, \quad (10)$$

где

$$\pi_m \equiv [\Pi_0 : I_K] e_m. \quad (11)$$

- (iv) Проверьте, что $\delta_m = \delta_{0m}$ является одним из решений (10). Покажите, что необходимым и достаточным условием того, чтобы $\delta_m = \delta_{0m}$ было единственным решением (10), является ранговое условие идентификации уравнения m . **Указание:** Вспомните следующий факт из линейной алгебры: Пусть $Ax = y$ имеет решение x_0 . Необходимым и достаточным условием того, чтобы это решение было единственным, является полный столбцовый ранг матрицы A .
- (v) Объясните, почему мы можем сделать вывод о том, что единственное решение $\Pi \Pi'_0 + B = 0$ является $(\Gamma, H) = (\Gamma_0, H_0)$. **Указание:** Каждый элемент δ может встречаться в Γ или B только один раз.

4. Дополнительные инструменты, которые не появляются в системе. Векторизовать полную систему одностепенных уравнений, обсуждавшуюся в параграфе 8.5. Пусть

$$x_k = \begin{bmatrix} \tilde{P}_k \\ (x' - 1) \times 1 \\ \tilde{P}_k \end{bmatrix}$$

Предположим, что x_{ijk} не входит ни в одно из M структурных уравнений системы. Покажите, что исключение x_{ijk} из набора инструментов не влияет ни на вычислимость рангового условия идентификации, ни на асимптотическую дисперсию FIML-оценок параметров δ_{ij} . Указание: Определите матрицы S_m и C_m как в выражении (6), так что $x_{ijk}^* = (y_i S_m' : x_{ijk}^* C_m')$ и используйте (6). FIML-оценка асимптотически эквивалентна 3SLS, следовательно, ее асимптотическая дисперсия задается в (4,5,15). Последняя строка C_m представляет собой вектор из нулей. Пусть Π_C (размера $((K-1) \times M)$) — матрица коэффициентов приведенной формы, когда x_{ijk} исключены из системы. Тогда

$$\Pi_C = \begin{bmatrix} \tilde{\Pi}_0 \\ (K-1) \times 0 \\ 0' \\ 1 \times M \end{bmatrix}$$

Литература

- Amemiya, T., 1974, "Regression Analysis: When the Dependent Variable is Truncated Normal", *Econometrica*, 41, 987–1016.
- , 1985, *Advanced Econometrics*, Cambridge: Harvard University Press.
- Anderson, T. W., and H. Rubin, 1949, "Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations", *Annals of Mathematical Statistics*, 20, 46–65.
- Anderson, T. W., A. Kunitomo, and I. Sawa, 1982, "Evaluation of the Distribution Function of the Least Information Maximum Likelihood Estimator", *Econometrica*, 50, 1009–1027.
- Davidson, R., and J. MacKinnon, 1980, *Estimation and Inference in Econometrics*, New York: Oxford University Press.
- Fuhrer, W., 1996, *Introduction to Statistical Time Series* (2d ed.), New York: Wiley.
- Hausman, J., 1975, "An Instrumental Variable Approach to Full Information Estimators for Linear and Certain Nonlinear Econometric Models", *Econometrica*, 43, 727–738.
- Johansen, S., 1995, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, New York: Oxford University Press.
- Judge, G., W. Griffiths, R. Hill, H. Lutkepohl, and T. Lee, 1985, *The Theory and Practice of Econometrics* (2d ed.), New York: Wiley.
- Magnus, J., and H. Neudecker, 1985, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, New York: Wiley.
- Newey, W., and D. McFadden, 1994, "Large Sample Estimation and Hypothesis Testing", Chapter 30 in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Volume IV, New York: North-Holland.
- Olsen, E. J., 1975, "Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model", *Econometrica*, 43, 1211–1215.
- Orma, C., 1989, "On the Uniqueness of the Maximum Likelihood Estimator in Truncated Regression Models", *Econometric Reviews*, 8, 217–222.

- Pagan, A., 1979, "Some Consequences of Viewing LIML As an Iterated Aitken Estimator", *Economics Letters*, 3, 369–372.
- Sarna, S. K., 1992, "Asymptotic Properties of a Quasi-Maximum Likelihood Estimator in Truncated Regression Model with Serial Correlation", *Econometric Reviews*, 11, 253–260.
- Tobin, J., 1958, "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, 26, 24–36.
- Woolbridge, J., 1994, "Estimation and Inference for Dependent Processes", Chapter 45 in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Volume IV, New York: North-Holland.

Глава 9. Эконометрика единичного корня

До настоящего момента наш анализ был ограничен стационарными процессами. В параграфе 9.1 вводится два класса процессов с трендом: процессы, стационарные относительно тренда, и процессы с единичным корнем. Процесс с единичным корнем называется также стационарным в разностях или интегрированным порядка 1 ($I(1)$), поскольку его первая разность является стационарным, или $I(0)$ -процессом. Технические средства для получения правильного распределения статистик, включающих $I(0)$ - и $I(1)$ -процессы, собраны в параграфе 9.2. Используя эти средства, мы получим некоторые тесты на единичный корень при нулевой гипотезе о том, что рассматриваемый процесс является $I(1)$ -процессом. Эти тесты не только являются популярными, но и обладают лучшими свойствами на конечных выборках по сравнению со многими другими существующими тестами. В параграфе 9.3 мы рассматриваем тесты Дickey — Фуллера, которые были разработаны для тестирования нулевой гипотезы о том, что процесс является случайным блужданием, т.е. стационарным $I(1)$ -процессом, первая разность которого сепарально стационарна. Параграф 9.4 показывает, что эти тесты можно обобщить таким образом, чтобы охватить $I(1)$ -процессы с сепарально коррелированными первыми разностями. Эти и другие тесты на единичный корень кратко сравниваются в параграфе 9.5. В прикладном параграфе 9.6 этой главы тесты на единичный корень используются для проверки паритета покупательной способности (PPP), фундаментального утверждения в международной экономике о том, что номинальные курсы подпадают под национальные уровни цен.

9.1. Моделирование трендов

Нет недостатка в примерах таких временных рядов, которые можно было бы разумно считать как тренды и циклические. На рис. 9.1 показано положение тарифов реально и ВВП США. Тарифы ВВП очевидно имеют

возрастающий тренд, в том смысле, что его среднее, вместо того чтобы быть константой, как у стационарных процессов, неуклонно возрастает с увеличением времени. Тренд среднего называется **детерминированным трендом** (*deterministic trend*) или **временным трендом** (*time trend*). В случае логарифма ВВП США детерминированный тренд выглядит линейным. Другой вид тренда можно увидеть на рис. 6.3 на с. 454, где показано поведение логарифма обменного курса иена/доллар. Логарифм этого обменного курса не имеет тренда среднего, но кажется, что каждое его изменение оказывает перманентное воздействие на будущие значения ряда, так что наилучшим предиктором его будущих значений является его текущее значение. Процесс с этим свойством, который не принадлежит к стационарным процессам, называется **стохастическим трендом** (*stochastic trend*). Если вы вспомните определение мартингала, то оно соответствует этому описанию стохастического тренда. Стохастические тренды и мартингалы являются синонимами.

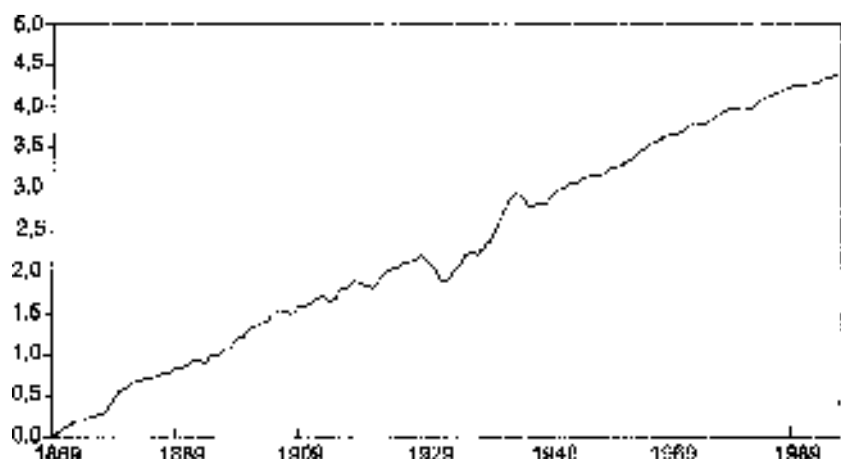


Рис. 9.1. Логарифм реального ВВП США, значение для 1869 года приравнено 0

Основная предпосылка этой главы и следующей заключается в том, что экономический временной ряд может быть представлен в виде суммы линейного временного тренда, стохастического тренда и стационарного процесса¹.

¹Замечание о семантике. Процессы с трендами часто называются «нестационарными процессами». Мы не будем использовать этот термин в оставшейся части этой главы, поскольку процесс может быть нестационарным и без наличия тренда. Например, пусть ε_t является IID-процессом с единичной дисперсией и пусть ϕ_t принимает значение 1 для нечетных t и 2 для четных t . Тогда процесс (u_t) описанный как $u_t = \phi_t \varepsilon_t$, не является стационарным, поскольку его дисперсия зависит от t . Тем не менее этот процесс не может быть разумно описан как процесс с трендом.

Интегрированные процессы

Случайное блуждание является примером класса трендовых процессов, известных как **интегрированные процессы** (*integrated processes*). Что бы дать точное определение интегрированным процессам, мы рассмотрим сначала $I(0)$ -процессы. **$I(0)$ -процесс** (*$I(0)$ process*) является (строго) стационарным процессом, **долговременная дисперсия** (*long-run variance*) которого (определяемая в обсуждении, следующем за утверждением 6 в параграфе 6.5) конечна и положительна. (Через игнорирование мы объясним, почему долговременная дисперсия должна быть положительной.) Следуя, например, [Hamilton 1994, гл. 435], мы допускаем, что $I(0)$ -процессы могут иметь ненулевые средние (некоторые авторы, например [Stock, 1994], требуют, чтобы $I(0)$ -процессы имели нулевое среднее). Следовательно, $I(0)$ -процесс может быть задан как

$$x_t \equiv y_t, \quad (9.1.1)$$

где $\{y_t\}$ является стационарным процессом с нулевым средним и положительной долговременной дисперсией.

Определение интегрированных процессов вытекает из определения $I(0)$ процессов. Пусть Δ является оператором разности, так что для последовательности $\{x_t\}$

$$\begin{aligned} \Delta x_t &= (1 - L)x_t = x_t - x_{t-1}, \\ \Delta^2 x_t &= (1 - L)^2 x_t = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}), \text{ и т.д.} \end{aligned} \quad (9.1.2)$$

Определение 9.1 ($I(d)$ -процессы): Процесс называется интегрированным порядка d ($\equiv I(d)$), ($d = 1, 2, \dots$), если d -я разность, $\Delta^d x_t$, является $I(0)$ процессом. В частности, процесс $\{x_t\}$ является интегрированным порядка 1 ($\equiv I(1)$), если его первая разность, Δx_t , является $I(0)$ -процессом.

Причина требования того, чтобы долговременная дисперсия $I(0)$ -процесса была положительной, заключается в том, чтобы исключить следующие типы информации о процессе: Рассмотрим процесс y_t , определяемый как

$$y_t \equiv x_t - x_{t-1}, \quad (9.1.3)$$

где $\{x_t\}$ является произвольным блужданием. Как мы увидим в следующем вопросе 3 в параграфе 6.5, долговременная дисперсия $\{y_t\}$ равна нулю. Если бы указанного требования к долговременной дисперсии не было, то $\{y_t\}$ был бы $I(0)$ -процессом. Но тогда, так как $\Delta y_t = x_t$, процесс $\{x_t\}$ был бы $I(1)$ -процессом. Если процесс $\{y_t\}$ понимается в виде первой разности $I(0)$ -процесса

то он называется $I(-1)$ -процессом. Долгосрочная дисперсия $I(-1)$ -процесса растет быстро (доказательство: см. ниже), поэтому контрольный вариант $I(1)$.

В оставшейся части этой главы интегрированные процессы, с которыми мы имеем дело, имеют порядок 1. Ниже приведено несколько замечаний относительно $I(1)$ -процессов.

- (Когда он стационарен?) Как это же о для случайных блужданий, дисперсия $I(1)$ -процесса аналогично возрастает по времени. Таким образом, если бы процесс начался в бесконечном прошлом, дисперсия была бы бесконечной. Чтобы сосредоточиться на $I(1)$ -процессах с конечной дисперсией, мы предполагаем, что процесс начался с непустым прошлым, т. е. без отрицательности, мы можем предположить, что начальной датой является $t = 0$. Поскольку $I(0)$ -процесс можно записать как (9.1.1) и, следовательно, по определению $I(1)$ -процесса, $\{x_t\}$ удовлетворяет соотношению $\Delta x_t = \delta + u_t$, мы можем записать x_t в уровнях как

$$x_t = x_0 + \delta \cdot t + (u_1 + u_2 + \dots + u_t), \quad (9.1.4)$$

где $\{u_t\} \sim i.i.d.$ с нулевым средним. Поэтому спецификация процесса $\{x_t\}$ в уровнях должна включать предположение о начальном значении. Если не указано обратное, мы всегда предполагаем, что $E(x_0^2) < \infty$, так что x_0 может быть случайным.

- (Среднее процесса $I(0)$ является трендом в $I(1)$.) Как ясно из (9.1.4) $I(1)$ -процесс может иметь линейный тренд, $\delta \neq 0$. Это является следствием допущения того, что $I(0)$ -процесс имеет ненулевой тренд δ . Если $\delta = 0$, то $I(1)$ -процесс не имеет тренда и называется **$I(1)$ -процессом без сноса** (*driftless $I(1)$ process*), в то время как если $\delta \neq 0$, процесс называется **$I(1)$ -процессом со сносом** (*$I(1)$ process with drift*). Очевидно, что $I(1)$ -процесс со сносом может быть задан как сумма линейного тренда и $I(1)$ -процесса без сноса.
- (Два других названия $I(1)$ -процессов.) У $I(1)$ -процессов есть два других названия. Он называется **процессом, стационарным в разностях** (*difference-stationary process*), поскольку его первая разность стационарна. Он также называется **процессом с единичным корнем** (*unit-root process*). Чтобы понять, почему он так называется, рассмотрим модель

$$(1 - \rho L)y_t = \delta + u_t, \quad (9.1.5)$$

где $\{u_t\} \sim i.i.d.$ с нулевым средним. Это авторегрессионная модель с возможно serially коррелированными ошибками, представлен-

ными посредством π_t . Если авторегрессионный корень ρ равен единице, то первая разность y_t является $I(0)$ -процессом, так что $\{y_t\}$ является $I(1)$ -процессом.

Почему важно знать, является ли рассматриваемый процесс $I(1)$ -процессом

В оставшейся части этой главы мы будем иметь дело с различием **процессов, стационарных относительно тренда** (*trend-stationary processes*), которые можно записать как сумму линейного временного тренда (если таковой имеется) и стационарного процесса с одной стороны и процессов, стационарных в разностях (то есть $I(1)$ -процессов с или без сдвига), с другой. Как будет показано в утверждении 9.1, $I(1)$ -процесс можно записать как сумму линейного тренда (если таковой имеется), стационарного процесса и стохастического тренда. Таким образом, различия между указанными двумя классами процессов состоят в наличии стохастического тренда. Имеются по крайней мере две причины, почему такое различие является важным.

1. Первая: это имеет большое значение для прогнозирования. Чтобы показать это, рассмотрим следующий простой $AR(1)$ -процесс с трендом:

$$\begin{aligned} y_t &= \alpha + \delta \cdot t + \varepsilon_t, \\ \varepsilon_t &= \rho \varepsilon_{t-1} + z_t, \end{aligned} \quad (9.1.6)$$

где $\{z_t\}$ является независимым белым шумом. Прогноз на s шагов вперед y_{t+s} условный относительно $(y_t, y_{t-1}, \dots)$, можно записать как

$$\begin{aligned} E(y_{t+s} | y_t, y_{t-1}, \dots) &= \\ &= E[\alpha + \delta \cdot (t+s) | y_t, y_{t-1}, \dots] + E(z_{t+s} | y_t, y_{t-1}, \dots) = \\ &= \alpha + \delta \cdot (t+s) - E(z_{t+s} | y_t, y_{t-1}, \dots). \end{aligned} \quad (9.1.7)$$

И $(y_t, y_{t-1}, \dots)$, и $(z_t, z_{t-1}, \dots)$ содержат одну и ту же информацию, поскольку между ними существует взаимно однозначное соответствие. Поэтому последнее слагаемое можно записать как

$$E(z_{t+s} | y_t, y_{t-1}, \dots) = E(z_{t+s}, z_t, z_{t-1}, \dots) = \rho^s z_t, \quad (9.1.8)$$

поскольку $z_{t+s} = z_{t+s} + \rho z_{t+s-1} + \dots + \rho^{s-1} z_{t+1} + \rho^s z_t$. Подставляя (9.1.8) в (9.1.7), мы получаем следующее выражение для прогноза на s шагов вперед:

$$\begin{aligned} E(y_{t+s} | y_t, y_{t-1}, \dots) &= \alpha + \delta \cdot (t+s) + \rho^s z_t = \\ &= \alpha + \delta \cdot (t+s) + \rho^s \cdot (y_t - \alpha - \delta \cdot t). \end{aligned} \quad (9.1.9)$$

Следует рассмотреть два случая:

Случай $\rho < 1$ Если $\rho < 1$, то $\{z_t\}$ является стационарным AR(1)-процессом и, таким образом, является I(0)-процессом с нулевым средним. Поэтому $\{y_t\}$ является процессом, стационарным относительно тренда. В этом случае, поскольку $E(z_t^2) < \infty$, мы имеем:

$$E[(\rho^s z_t)^2] = \rho^{2s} E(z_t^2) \rightarrow 0 \text{ при } s \rightarrow \infty.$$

Следовательно, прогноз на s периодов вперед $E(y_{t+s} | y_t, y_{t-1}, \dots)$ при $s \rightarrow \infty$ сходится к среднему квадратическому и линейному временному тренду $\alpha + \delta \cdot (t + s)$. Более точно:

$$E\{[E(y_{t+s} | y_t, y_{t-1}, \dots) - \alpha - \delta \cdot (t + s)]^2\} \rightarrow 0 \text{ как } s \rightarrow \infty. \quad (9.1.10)$$

То есть текущее и прошлые значения y не влияют на прогноз, если горизонт прогноза s достаточно далекий. В частности, если $\delta = 0$, то

$$E(y_{t+s} | y_t, y_{t-1}, \dots) \xrightarrow{s \rightarrow \infty} E(y_t) \text{ при } s \rightarrow \infty. \quad (9.1.11)$$

То есть долговременный прогноз является безусловное среднее. Это свойство, называемое **возвращением к среднему** (*mean reversion*), выполняется для любого линейного стационарного процесса, а не только для стационарного AR(1)-процесса¹. По этой причине линейный стационарный процесс иногда называют **переходящей компонентой** (*transitory component*).

Случай $\rho = 1$ Предположим теперь, что $\rho = 1$. Тогда $\{z_t\}$ является случайным блужданием без сноса (частный случай стохастического тренда), и y_t можно записать в виде:

$$y_t = (\alpha + z_0) + \delta \cdot t + \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t. \quad (9.1.12)$$

показывающем, что $\{y_t\}$ является случайным блужданием со сносом с начальным значением $z_0 + \alpha$ (частный случай I(1) или стационарного в разностях процесса). Полагая $\rho = 1$ в (9.1.9), мы получаем:

$$E(y_{t+s} | y_t, y_{t-1}, \dots) = \delta \cdot s + y_t. \quad (9.1.13)$$

То есть случайное блуждание со сносом δ , как ожидается, растет с постоянной скоростью δ за период, каким бы ни было текущее значение y_t . В отличие от случая стационарности относительно тренда текущее значение оказывает здесь перманентное влияние

¹Дополнительно см., например, в: [Hamilton, 1994, гл. 439].

на прогноз для всех горизонтов прогноза. Это происходит из-за наличия стохастического тренда z_t . Стохастический тренд называют также **перманентной компонентой** (*permanent component*).

- Вторая причина для различения стационарных относительно тренда и стационарных и разностных процессов возникает в контексте статистических выводов. Предположим, например, что вы заинтересованы в проверке некоторой гипотезы относительно параметра β в уравнении регрессии.

$$y_t = \alpha + \beta z_t + \varepsilon_t \quad (9.1.14)$$

где ε_t является стационарной ошибкой. С одной стороны, если регрессор z_t стационарен относительно тренда, то тогда, как было показано в параграфе 2.12, формула для OLS-оценки параметра β будет иметь асимптотически стандартное нормальное распределение. С другой стороны, если z_t стационарен *в разности*, то есть, если он имеет стохастический тренд, то это существенно имеет важное различное предельное распределение, так что делать статистические выводы относительно β обычным способом нельзя. Регрессия в рассматриваемом случае называется «интегрирующей регрессией». Статистические выводы относительно интегрирующей регрессии будут излагаться в следующей главе.

Что следует брать в качестве нулевой гипотезы, $I(0)$ или $I(1)$?

Для различения стационарных относительно тренда и стационарных и разностных процессов, или $I(1)$ -процессов, в литературе обычной практикой является проверка нулевой гипотезы о том, что рассматриваемый процесс является $I(1)$ -процессом, а не проверка нулевой гипотезы о стационарности относительно тренда. Предположительно у этого есть две причины. Одна из них состоит в том, что исследователь убежден, что в представляющем интерес уравнении переменные являются $I(1)$, и хочет контролировать ошибку 1 рода относительно нулевой гипотезы $I(1)$. Однако это в конечном счете не является убедительным, поскольку обычно исследователь не обладает строгим априорным убеждением относительно того, какими являются переменные, $I(0)$ или $I(1)$. Краткое изложение проблемы, которую это создает для статистических выводов, имеется в [Watson, 1994, pp. 2867–2870]. Другая причина состоит просто в том, что литература по тестам для нулевой гипотезы стационарности относительно тренда против альтернативы $I(1)$ все еще недостаточно разработана. Есть несколько тестов для проверки нулевой гипотезы о том, что процесс является $I(0)$ или, в более общем случае, стационарным относительно тренда; см. статьи [Tsayaka, 1990], [Kwiatkowski, et al., 1992]

и [Leybourne and McCabe, 1994] а также обзоры [Stock, 1994, Section 4] и [Maddala and Kim, 1998, Section 4.5]. На момент написания этой книги нет единственного теста для проверки $I(0)$ -гипотезы, использующегося большинством исследователей, который имел бы хорошие свойства в конечных выборках. Поэтому в этой книге мы сосредоточимся на тестах для проверки нулевой гипотезы $I(1)$.

Другие подходы к моделированию трендов

В дополнение к стохастическим и временным трендам имеются два других популярных подхода к моделированию трендов: дробная интегрируемость и тренды с изломом с неизвестной датой излома. Они не будут рассматриваться в этой книге. В качестве обзора первого подхода см.: [Raiffa, 1996]. Количество публикаций о единицах в тренде быстро возрастает. Обзоры можно найти в [Stock, 1994, Section 5] и [Maddala and Kim, 1998, Chapter 13].

Контрольные вопросы

1. (Диффузионная дисперсия разностей $I(0)$ -процессов.) Пусть $\{u_t\}$ есть $I(0)$ с $\text{Var}(u_t) < \infty$. Проверьте, что процесс $\{\Delta u_t\}$ является ковариационно стационарным и что долговременная дисперсия $\{\Delta u_t\}$ равна нулю.

Указание: Диффузионная дисперсия $\{u_t\}$ (где $u_t = \Delta u_t$) определяется как предел для $\text{Var}(\sqrt{T}u)$, при $t \rightarrow \infty$.

$$\sqrt{T}u = (u_1 + u_2 + \dots + u_T)/\sqrt{T} = (u_1 + \dots + u_n)/\sqrt{T}.$$

9.2. Инструментарий для эконометрики единичных корней

Основным инструментом эконометрики единичных корней является то, что называют **функциональной центральной предельной теоремой** (*functional central limit theorem, FCLT*). Для того чтобы FCLT была применима, мы должны сузить класс $I(1)$ -процессов, накладывая ограничения на соответствующие $I(0)$ -процессы. Решающие о том, какие ограничения следует наложить, то есть на каком классе $I(0)$ процессов сосредоточиться, изменяется от автора к автору в зависимости от используемой версии FCLT. В данной работе, следуя сложившейся в литературе практике, мы фокусируемся на линейных $I(0)$ -процессах. Вслед за определением линейных $I(0)$ -процессов в этом параграфе вводятся некоторые понятия и результаты, которые будут неоднократно использоваться в оставшейся части данной главы.

Линейные I(0)-процессы

Наше определение линейных I(0)-процессов заключается в следующем:

Определение 9.2 (линейные I(0)-процессы): Линейный I(0)-процесс можно записать как константу плюс линейный процесс с нулевым средним $\{\varepsilon_t\}$, такой что

$$u_t = \varepsilon_t(t)\varepsilon_t, \quad \varepsilon(t) = \varepsilon_0 + \varepsilon_1 t + \varepsilon_2 t^2 + \dots \quad \text{при } t = 0, \pm 1, \pm 2, \dots \quad (9.2.1)$$

$$\{\varepsilon_t\} \text{ является независимым белым шумом} \\ (\text{i.i.d. со средним } 0 \text{ и } E(\varepsilon_t^2) = \sigma^2 > 0). \quad (9.2.2)$$

$$\sum_{j=0}^{\infty} j|\varepsilon_j| < \infty. \quad (9.2.3a)$$

$$\varepsilon(1) \neq 0 \quad (9.2.3b)$$

Предполагается, что процесс инноваций $\{\varepsilon_t\}$ является белым шумом i.i.d. Это требование введено только для упрощения изложения, и его можно существенно ослабить¹. В остальной части этой главы мы называем линейные I(0)-процессы просто I(0)-процессами, без уточнения «линейные». Также, если не указано иное, $\{\varepsilon_t\}$ будет независимым процессом белого шума, а $\{u_t\}$ — I(0)-процессом с нулевым средним.

Условие (9.2.3a) (иногда называемое условием **I-суммируемости** (*summability*)) является более сильным, чем более известное условие абсолютной суммируемости. Это более сильное предположение сделает более легким доказательство результатов для больших выборок в следующем параграфе с помощью разложения Беттериджа — Нельсона, которое будет введено в ближайшее время. Поскольку последовательности $\{\varepsilon_t\}$ абсолютно суммируемы, линейный процесс $\{u_t\}$ является строго стационарным и эргодическим в силу утверждения 6.1(ii). Автоковариация u_t j -го порядка будет обозначаться как γ_j . Поскольку $E(u_t) = 0$, $\gamma_j = E(u_t u_{t-j})$. По утверждению 6.1(c), автоковариации абсолютно суммируемы.

Чтобы понять условие (9.2.3a), вспомним, что, по утверждению 6.8(b), долговременная дисперсия (которую мы обозначаем в этой главе λ^2) равна значению производящей функции автоковариаций при $z = 1$. По утверждению 6.6 она дается формулой:

$$\lambda^2 \equiv \text{долговременная дисперсия } \{u_t\} = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j = \sigma^2 [\varepsilon(1)]^2. \quad (9.2.4)$$

¹ Например, процесс инноваций может быть стационарной илточчал разности $\nabla \varepsilon_t$, например, Пеллери 198 в [Галлах, 1996, с. 50].

Условие (9.2.3b) гарантирует положительность дисперсионной матрицы:

Аппроксимация $I(1)$ -процесса случайным блужданием

Пусть $\{\xi_t\}$ является $I(1)$ -процессом, так что $\Delta \xi_t = \delta + \eta_t$, где $\eta_t \equiv \xi_t - \xi_{t-1}$ является $I(0)$ -процессом с нулевым средним, удовлетворяющим (9.2.1)–(9.2.3) с $E(\xi_0^2) < \infty$. Используя следующую тождество (его легко проверить и контролировать вкратце 1))

$$\eta(L) = \eta(1) + \Delta \eta(L), \quad \Delta = 1 - L,$$

$$\eta(L) = \sum_{j=0}^{\infty} \alpha_j L^j, \quad \alpha_j = (\eta_{j+1} + \eta_{j+2} + \dots + \eta_{j+q})/q, \quad j = 0, 1, 2, \dots, \quad (9.2.5)$$

мы можем записать η_t как

$$\eta_t = \eta(1)z_t + \Delta \eta(L)z_t = \eta + \eta(L)z_t. \quad (9.2.6)$$

Можно доказать (см. аналитическое выражение 7)) что $\eta(L)$ является абсолютно суммируемым. Поэтому, по утверждению 6.1(a) $\{\eta_t\}$ — ковариационно-стационарный процесс с нулевым средним (фактически он является процессом стационарным по утверждению 6.1(d)). Подставляя (9.2.6) в (9.1.4), мы получаем то, что известно в эконометрике как **разложение Бевриджа – Нельсона (Beveridge – Nelson decomposition)**:

$$\begin{aligned} \xi_t &= \delta \cdot t + \sum_{i=1}^p \lambda_i(1) \cdot x_i + \eta_0 - \eta_{t-1} + \xi_0 = \\ &= \delta \cdot t + \lambda(1) \sum_{i=1}^p x_i + \eta - (\xi_t - \eta_t) \\ &\quad \left(\text{поскольку } \sum_{i=1}^p (\eta_i - \eta_{i-1}) = \eta - \eta_0 \right). \end{aligned} \quad (9.2.7)$$

Таким образом, любой линейный $I(1)$ -процесс можно записать как сумму линейного тренда ($\delta \cdot t$), случайного блуждания без сноса, или стохастического тренда ($\lambda(1)(x_1 + x_2 + \dots + x_t)$), стационарного процесса (η) и начального значения ($\xi_0 - \eta_0$). Формулируя несколько отвлеченным образом, имеем

Утверждение 9.1 (разложение Бевриджа – Нельсона):

Пусть $\{\eta_t\}$ является $I(0)$ -процессом с нулевым средним, удовлетворяющим (9.2.1)–(9.2.3). Тогда

$$\eta_1 + \eta_2 + \dots + \eta_t = \lambda(1)(x_1 + x_2 + \dots + x_t) + \eta - \eta_0.$$

где $\eta_t = \alpha(L)\varepsilon_t$, $\alpha_t = (\alpha_{t-1} + \alpha_{t+2} + \dots)$. Процесс $\{\eta_t\}$ является эргодически стационарным с нулевым средним¹.

На мгновение положим в (9.2.7) $\delta = 0$, так что $\{\xi_t\}$ является $I(1)$ -процессом без сноса. Важным следствием указанного разложения является то, что любой $I(1)$ -процесс без сноса доминируется стохастическим трендом $\omega(1) \sum_{s=1}^t \varepsilon_s$ в следующем смысле. Разделим обе части разложения (9.2.7) (с $\delta = 0$) на $\sqrt{t}$ и получим:

$$\frac{\xi_t}{\sqrt{t}} = \omega(1) \cdot \frac{1}{\sqrt{t}} \sum_{s=1}^t \varepsilon_s + \left[\frac{\xi_0}{\sqrt{t}} + \frac{\eta_t}{\sqrt{t}} - \frac{\eta_1}{\sqrt{t}} \right]. \quad (9.2.8)$$

Поскольку, по предположению, $E(\xi_0^2) < \infty$, $E[(\xi_0/\sqrt{t})^2] \rightarrow 0$ при $t \rightarrow \infty$, поэтому $\xi_0/\sqrt{t}$ сходится в среднем квадратическом (и, следовательно, по вероятности) к нулю. То же самое применимо к $\eta_t/\sqrt{t}$ и $\eta_1/\sqrt{t}$. Поэтому слагаемые в квадратных скобках в правой части (9.2.8) можно игнорировать асимптотически. В отличие от этого первое слагаемое имеет предельное распределение $N(0, \sigma^2 \cdot [\omega(1)]^2)$ по СЛТ Ляндеберга—Цевя из параграфа 2.1. В этом смысле стохастический тренд растет со скоростью $\sqrt{t}$. Используя (9.2.4), стохастический тренд можно записать как

$$\omega(1)(\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t) = \lambda \cdot \left(\frac{\varepsilon_1}{\sigma} + \frac{\varepsilon_2}{\sigma} + \dots + \frac{\varepsilon_t}{\sigma} \right), \quad (9.2.9)$$

что показывает: изменения в стохастическом тренде в ξ_t имеют дисперсию λ^2 , долговременную дисперсию $\{\Delta \xi_t\}$.

Теперь предположим, что $\delta \neq 0$, так что $\{\xi_t\}$ является $I(1)$ -процессом со сносом. Как ясно из деления обеих сторон (9.2.7) на t , а не на $\sqrt{t}$ и применения той же самой аргументации, которая была только что приведена для случая $\delta = 0$, стохастический тренд, так же как и стационарную компоненту, можно игнорировать асимптотически, с ξ_t/t , сходящимся по вероятности к δ . В этом смысле временной тренд доминирует $I(1)$ -процесс на больших выборках.

Связь с моделями ARMA

В параграфе 6.2 мы определили стационарный ARMA-процесс с нулевым средним $\{\eta_t\}$ как единственный ковариационно стационарный процесс, удовлетворяющий стационарному ARMA(p, q) уравнению:

$$\phi(L)a_t = \theta(L)\varepsilon_t, \quad (9.2.10)$$

¹Опосредственно более удобно указать, при каком $\eta_1 = \dots = 0$ разлагается, как показано здесь, и который не требует, чтобы $I(0)$ -процесс $\{\varepsilon_t\}$ был линейным, см.: Theorem 5.4 в [Hall and Heyde, 1980]. В этой теореме последовательность $\{\varepsilon_t\}$ является стационарной мартингал разностью, поэтому стохастический тренд является мартингалом.

где $\phi(L)$ удовлетворяет условию стационарности. Если $\theta(1) \neq 0$, то процесс $\text{ARIMA}(p, q)$ является $I(0)$ -процессом с нулевым средним. Чтобы увидеть это, заметим, что, в соответствии с утверждением 6.5(a), $\phi(L)$ можно записать как

$$\phi(L) = \phi(L) + \theta(L), \quad (9.2.11)$$

Поскольку $\phi(1) \neq 0$ по условию стационарности и $\theta(1) \neq 0$ по предположению, условие (9.2.3b) (а так же $\psi(1) \neq 0$) выполняется. По утверждению 6.4(a) последовательность коэффициентов $\{\psi_j\}$ ограничена по абсолютному значению геометрически убывающей последовательностью, что гарантирует выполнение (9.2.3a).

При задании только что определенном стационарном $\text{ARIMA}(p, q)$ процессе $\{x_t\}$ с нулевым средним, соответствующий ему $I(1)$ -процесс $\{\xi_t\}$, определенный соотношением $\Delta \xi_t = x_t$, называется **интегрированным процессом авторегрессии — скользящего среднего** ($\text{ARIMA}(p, 1, q)$) (*integrated moving autore*). Подставляя соотношение $x_t = - (1 - L)\xi_t$ в (9.2.10), мы получаем:

$$\phi^*(L)\xi_t = \theta(L)\xi_t, \quad (9.2.12)$$

где $\phi^*(L) = \phi(L)(1 - L)$, $\phi(L)$ стационарный и $\theta(1) \neq 0$. Один из корней соответствующего авторегрессионного полиномиального уравнения $\phi^*(z) = 0$ равен единице, а все другие корни лежат за пределами единичного круга. В более общем случае класс $I(d)$ -процессов можно представить как удовлетворяющий (9.2.12), где $\phi^*(L)$ определяется уже как

$$\phi^*(L) = \phi(L)(1 - L)^d. \quad (9.2.13)$$

Поэтому $\phi^*(z) = 0$ имеет теперь единичный корень кратности d (то есть d единичных корней), а все другие корни лежат вне единичного круга. Процессы $I(d)$ называют интегрированными процессами авторегрессии — скользящего среднего ($\text{ARIMA}(p, d, q)$). Первый параметр (p) равен количеству авторегрессионных лагов (с учетом единичных корней), второй параметр (d) относится к порядку интегрированности, а третий параметр (q) равен количеству лагов в скользящем среднем. Поскольку $\phi(L)$ удовлетворяет условию стационарности, взятие d -х разностей $\text{ARIMA}(p, d, q)$ процесса приводит к стационарному $\text{ARIMA}(p, q)$ -процессу с нулевым средним. Поэтому $\text{ARIMA}(p, d, q)$ -процесс называется $I(d)$.

Винеровский процесс

В двух следующих параграфах будут представлены разнообразные тесты на единичный корень. Промежные распределения их тестовых статистик будут описаны в терминах **винеровского процесса** (*Wiener process*).

(известного также как **процесс броуновского движения** (*Brownian motion process*)). Некоторые из вас, возможно, уже знакомы с этим процессом по финансовым моделям с непрерывным временем, но чтобы освежить вашу память, приведем его определение.

Определение 9.3 (стандартный винеровский процесс): Стандартный винеровский процесс (броуновское движение) $W(t)$ — это случайный процесс с непрерывным временем, satisfying с каждым моментом $t \in [0, 1]$ скалярную случайную величину $W(t)$, такую, что

(1) $W(0) = 0$,

(2) для любых моментов $0 \leq t_1 < t_2 < \dots < t_k < 1$ значения

$$W(t_2) - W(t_1), W(t_3) - W(t_2), \dots, W(t_k) - W(t_{k-1})$$

являются независимыми нормально распределенными случайными величинами $W(t_k) - W(t_{k-1}) \sim N(0, (t_k - t_{k-1}))$ (так что, в частности, $W(1) \sim N(0, 1)$),

(3) с вероятностью 1 реализации $W(t)$ непрерывны по t .

Грубейшими, **функциональная центральная предельная теорема** (*functional central limit theorem*) (FCLT) называемая также **принципом инвариантности** (*invariance principle*), утверждает, что винеровский процесс является пределом случайного блуждания без сноса в непрерывном времени. Представьте, что вы генерируете реализацию стандартного случайного блуждания (каждое из которых имеет единичную дисперсию) длины T — изменяете масштаб графика по вертикали, деля все значения реализации на $\sqrt{T}$, а затем сжимаете этот нормализованный график по горизонтали так, чтобы он соответствовал единичному интервалу $[0, 1]$. На панели (a) рис. 9.2 это сделано для размера выборки $T = 10$. Размер выборки увеличивается до 100 на панели (b), а затем — до 1000 — на панели (c). Как показано на рисунке, график становится все более и более похожим на непрерывный. FCLT гарантирует вам, что существует предельный процесс при $T \rightarrow \infty$ и что предельный процесс является винеровским процессом. Свойство (2) определения 9.3 является математической формулировкой того, что последовательность числовых изменений винеровского процесса является i.i.d. Аналог в непрерывном времени случайного блуждания без сноса, каждое из которых имеет дисперсию σ^2 , а следовательно дисперсию можно записать как $\sigma^2 W(t)$.

Чтобы провести аналогию между случайным блужданием без сноса и винеровским процессом чуть дальше, рассмотрим «центрированное стандартное случайное блуждание», которое строится из стандартных

случайного блуждания без сноса путем вычитания выборочного среднего. То есть пусть $\{\xi_t\}$ является случайным блужданием без сноса

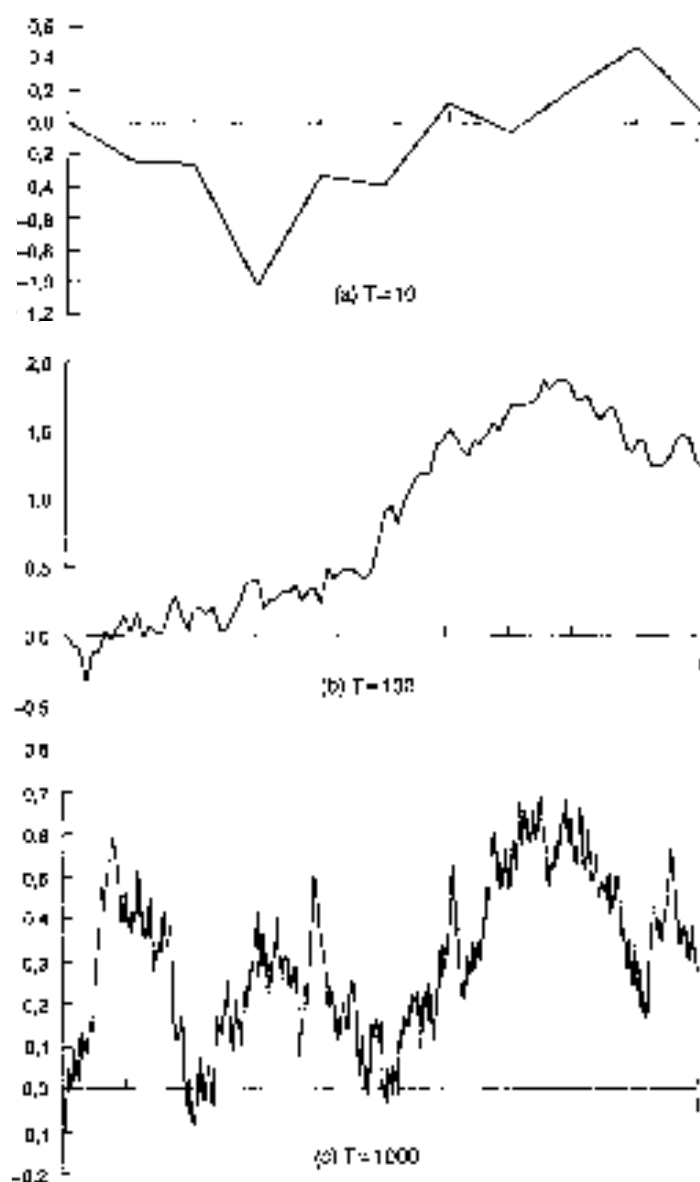


Рис. 9.2. Иллюстрация функциональной центральной предельной теоремы

с $\text{Var}(\Delta \xi_t) = 1$ по определению:

$$\xi_t^D = \xi_t - \frac{\xi_1 + \xi_2 + \dots + \xi_{T-1}}{T} \quad (t = 0, 1, \dots, T-1). \quad (9.2.14)$$

Аналогом в непрерывном времени центрированного стандартного случайного блуждания является **центрированный стандартный винеровский процесс** (*detrended standard Wiener process*), определяемый как

$$W^D(r) \equiv W(r) - \int_0^1 W(s) ds. \quad (9.2.15)$$

(Мы определяли центрированные ряды для $t = 0, 1, \dots, T-1$, чтобы соответствовать соглашениям в датировке утверждений 9.2 выше. Если бы центрированный ряд был определен для $t = 1, 2, \dots, T$, то его аналогом в непрерывном времени тоже был (9.2.15).)

Мы можем также создать из стандартного случайного блуждания $\{\xi_t\}$ **детрендеризованный ряд**:

$$\xi_t^D \equiv \xi_t - \bar{\xi} \quad (t = 0, 1, \dots, T-1), \quad (9.2.16)$$

где $\bar{\xi} \equiv \bar{\xi}^D$ сходится OLS-оценке константы и коэффициента линейного тренда в регрессии ξ_t на $(1, t)$ ($t = 0, 1, \dots, T-1$). Аналогом в непрерывном времени детрендеризованного случайного блуждания является **детрендеризованный стандартный винеровский процесс** (*detrended standard Wiener process*), определяемый как

$$\begin{aligned} W^D(r) &\equiv W(r) - a - b \cdot r, \\ a &\equiv \int_0^1 (1-s) W(s) ds, \\ b &\equiv \int_0^1 (s - \frac{1}{2}) W(s) ds. \end{aligned} \quad (9.2.17)$$

Коэффициенты a и b являются предельными случайными величинами при $T \rightarrow \infty$ для $\bar{\xi}$ и $\bar{\xi}^D$ соответственно (вывод см., например, в [Phillips and Dunsmuir, 1986]). Следовательно, если линейный временной тренд подгоняется посредством OLS к случайному блужданию без смещения, оцененный линейный временной тренд $\bar{\xi}$ сходится по распределению к случайной величине a (также на больших выборках значения b сходятся к b), будет случайным, равносильно случаю $\bar{\xi}$. Этот феномен известен как **случайная ложная детрендеризация** (*spurious detrending*).

Полезная лемма

Тесты на единичный корень используют статистики, включающие $I(0)$ - и $I(1)$ -процессы. Какимые результаты, собранные в следующем утверждении, будут использоваться для получения предельных распределений статистик тестов на единичный корень в параграфах 9.3–9.4.

Утверждение 9.2 (предельные распределения статистик, включающих $I(0)$ - и $I(1)$ -переменные): Пусть $\{\xi_t\}$ является $I(1)$ -процессом без сноса, так что $\Delta\xi_t$ является $I(0)$ -процессом с конечным средним, удовлетворяющим (9.2.1)–(9.2.3), и $E(\xi_t^2) < \infty$. Пусть

$$\lambda^2 \equiv \text{долгосрочная дисперсия } \{\Delta\xi_t\} \text{ и } \gamma \equiv \text{Var}(\Delta\xi_1).$$

Тогда

$$(a) \quad \frac{1}{T^2} \sum_{t=1}^T (\xi_{t-1})^2 \rightarrow_p \lambda^2 \cdot \int_0^1 W(r)^2 dr,$$

$$(b) \quad \frac{1}{T} \sum_{t=1}^T \Delta\xi_t \xi_{t-1} \rightarrow_p \frac{\lambda^2}{2} [W(1)^2 - W^2(0)] = -\frac{\gamma}{2}.$$

Пусть $\{\xi_t^D\}$ является центрированным рядом, полученным из $\{\xi_t\}$ по формуле (9.2.4). Тогда

$$(c) \quad \frac{1}{T^2} \sum_{t=1}^T (\xi_t^D)^2 \rightarrow_p \lambda^2 \cdot \int_0^1 [W^D(r)]^2 dr,$$

$$(d) \quad \frac{1}{T} \sum_{t=1}^T \Delta\xi_t \xi_{t-1}^D \rightarrow_p \frac{\lambda^2}{2} [W^D(1)^2 - W^D(0)^2] = -\frac{\gamma}{2}.$$

Пусть $\{\xi_t^D\}$ является детрендированным рядом, полученным из $\{\xi_t\}$ по формуле (9.2.6). Тогда

$$(e) \quad \frac{1}{T^2} \sum_{t=1}^T (\xi_{t-1}^D)^2 \rightarrow_p \lambda^2 \cdot \int_0^1 [W^D(r)]^2 dr,$$

$$(f) \quad \frac{1}{T} \sum_{t=1}^T \Delta\xi_t \xi_{t-1}^D \rightarrow_p \frac{\lambda^2}{2} [W^D(1)^2 - W^D(0)^2] = -\frac{\gamma}{2}.$$

Сходимость здесь совместная. То есть векторы, составленный из статистик, указанных в (a)–(f), сходится к случайному вектору, элементы которого являются соответствующими случайными величинами, также указанными в (a)–(f).

Чтобы (a), например, читается следующим образом: последовательность случайных величин $\{(1/T^2) \sum_{t=1}^T (\xi_{t-1})^2\}$, индексированная параметром T , сходится по распределению к случайной величине $\lambda^2 \int W^2 dr$. Все предельные случайные величины в этом утверждении записываются в терминах стандартного винеровского процесса. Замечим, что W^D — тот же самый винеровский процесс $W(\cdot)$ (см. (9.2.4) и (9.2.6)) и что центрированный и детрендированный винеровские процессы в (c)–(f) получены из винеровского процесса, показанного в (a) и (b). Поэтому предельные случайные величины в (a)–(f) могут быть коррелированными.

Чтобы получить определенное понимание этого результата, временно предположим, что $\{\xi_t\}$ является случайным блужданием без сноса

с $\text{Var}(\Delta\xi_t) = \sigma^2$ и фиксированным ξ_1 . Учитывая, что винеровский процесс является его аналогом в непрерывном времени, рассмотрим, что « $\sum_{i=1}^T (\xi_i - 1)^2$ », соответствующим образом нормализованная на некоторую степень T , становится « $\sigma^2 \int_0^1 W^2$ ». Возможно, что вы не могли добиться нормализовать ее посредством T' . Один из способов понять, почему именно квадрат T дает правильную нормализацию, — это вспомнить, что $E[(\xi_t)^2] = \text{Var}(\xi_t) = \sigma^2 \cdot t$. Таким образом,

$$E\left[\sum_{i=1}^T (\xi_i - 1)^2\right] = \sigma^2 \sum_{i=1}^T (i - 1) = \sigma^2 \cdot \frac{T(T-1)}{2}. \quad (9.2.18)$$

Так что среднее $\sum_{i=1}^T (\xi_i - 1)^2$ растет со скоростью T' . Чтобы получить случайную величину, которая имела бы правильное распределение, придется разделить $\sum_{i=1}^T (\xi_i - 1)^2$ на T' .

Прежде, нежели понять, что $\{\xi_t\}$ является общим $I(1)$ -процессом без сноса, полезно заметить, что его значение является процессом $I(0)$ с нулевым средним, удовлетворяющим (9.2.1)–(9.2.3). Тогда случайную корреляцию $\Delta\xi_t$ можно приписать во движении кутек значения σ^2 на λ^2 (долгое время для дисперсии $\{\Delta\xi_t\}$). Эти свойства — два одних следствия (9.2.2) разложения Бессежа — Феллера, reminding, что $I(1)$ -процесс без сноса — это блуждание, которое подвергается случайным блужданиям, имеющим среднего квадрат дисперсию λ^2 . Иначе говоря, приведенное распределение для $\sum_{i=1}^T (\xi_i - 1)^2 / T^2$ было бы тем же самым, если бы $\{\xi_t\}$, вместо того чтобы являться общим $I(1)$ -процессом без сноса с коррелированными первыми разностями, был случайным блужданием без сноса, изменения которого имеют дисперсию λ^2 .

По поводу доказательства утверждения 9.2 см., например, (Stock, 1994, pp. 2751–2753). Это, скорее, математические приложения FCLT и теоремы, называемой теоремой о непрерывном отображении. Поскольку $W(\cdot) \sim N(0, 1)$, предельная случайная величина в (b) принимает вид:

$$\left(\frac{\lambda^2}{2}\right) X = \frac{\lambda_0}{2}, \quad (9.2.19)$$

где $X \sim \chi^2(1)$ (хи-квадрат с одной степенью свободы). Доказательство (b) можно выполнять без упомянутого аппарата FCLT и теоремы о непрерывном отображении, и это оставим в качестве аналитического упражнения 1.

Контрциклические вопросы

1. Проверка разложения Бессежа — Феллера для простого случая: Проверьте (9.2.5) для $\phi(f) = 1 + \alpha \cdot f$. (Подставьте то же самое для $\phi(f) = (1 + \alpha f)^2$ при $\alpha < 1$).

2. (CLT для $I(0)$.) Утверждение 6.9 касается асимптотики выборочного среднего линейного процесса. Проверьте, что вблиз. расположенный непосредственно к утверждению 9.1, является доказательством утверждения 6.9 при допущении предположении о 1-суммируемости.

3. (Является ли стационарная компонента в разложении Бекериджа — Нельсона $I(0)$ процессом?) Рассмотрите процесс $I(0)$ с нулевым средним:

$$y_t = \alpha y_{t-1} + 2\alpha y_{t-2} + \varepsilon_{t-2}.$$

Покажите, что долгосрочная дисперсия процесса $\{y_t\}$ в (9.2.6) равна нулю.

4. (Названные значения и центрированные величины.) Пусть $\{\xi_t\}$ является $I(1)$ -процессом без сноса, так что ξ_t можно записать как $\xi_t = \xi_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_p t^p$. Влияет ли значение ξ_0 на значение ξ_t^* , где $\{\xi_t^*\}$ является центрированным рядом, полученным по формуле (9.2.14)? [Ответ: Нет.]

5. (Линейные тренды и детрендированные величины.) Пусть $\{\xi_t\}$ — $I(1)$ -процесс со сном, так что его можно записать как сумму линейного временного тренда $\alpha + \delta t$ и $I(1)$ процесса без сноса. Влияют ли значения α и δ на значение ξ_t^* , где $\{\xi_t^*\}$ является центрированным рядом, полученным по формуле (9.2.16)? [Ответ: Нет.]

9.3. Тесты Дики — Фуллера

В этом параграфе и в следующем мы представим несколько тестов для проверки нулевой гипотезы о том, что процесс порождения данных (DGP) является $I(1)$ -процессом. В этом параграфе $I(1)$ -процесс является случайным блужданием с или без сноса и тесты основаны на OLS-оценках соответствующих AR(1) уравнений. Эти тесты были разработаны в [Dickey, 1976] и [Fuller, 1978] и называются **тестами Дики — Фуллера** (*Dickey — Fuller tests*), или **DF-тестами** (*DF tests*). Мы принимаем соглашение о том, что выборка содержит $T - 1$ наблюдений $\{y_1, y_2, \dots, y_T\}$, так что AR(1)-уравнение можно оценить для $t = 1, 2, \dots, T$

AR(1)-модель

Тестовые статистики Дики — Фуллера получаются из оцениваемой модели авторегрессии первого порядка:

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\} \text{ — независимый белый шум.} \quad (9.3.1)$$

(Случай с константой будет изучен через мгновение.) Эта модель исключает как $I(1)$, так $I(0)$ процессы: если $\rho = 1$, то есть если соответствующее полиномиальное уравнение $1 - \rho z = 0$ имеет единичный корень, то $\Delta y_t = \varepsilon_t$ и $\{y_t\}$ является случайным блужданием без тренда, а если $|\rho| < 1$, процесс является стационарным AR(1) с нулевым средним. Таким образом, гипотезу о том, что данные порождаются случайным

Блужданием без тренда можно сформулировать как нулевую гипотезу о том, что $\alpha = 1$. Если мы принимаем ту позицию, что DGP или $I(0)$, или $I(1)$, то $-1 < \alpha \leq 1$. При условии этого одностороннего ограничения на ρ альтернативу $I(0)$ можно выразить как $\alpha < 1$. Поэтому для проверки нулевой гипотезы $\rho = 1$ против альтернативы $I(0)$ будут использоваться односторонние тесты.

Соответствующие тесты будут основаны на $\hat{\rho}$, OLS-оценке для α в AR(1)-уравнении (9.3.1). При нулевой гипотезе $I(1)$, $\rho = 1$, ошибка оценки равна $\hat{\rho} - 1$ и задается выражением:

$$\hat{\rho} - 1 \equiv \frac{\sum_{t=1}^T y_t y_{t-1}}{\sum_{t=1}^T (y_{t-1})^2} - 1 = \frac{\sum_{t=1}^T \Delta y_t y_{t-1}}{\sum_{t=1}^T (y_{t-1})^2}. \quad (9.3.2)$$

Как будет проверено через мгновенные числители имеет предельное распределение, если разделить это на T , а знаменатель имеет предельное распределение, если разделить его на T^2 . Поэтому рассмотрим ошибку оценки, увеличенную в T раз:

$$T \cdot (\hat{\rho} - 1) = T \cdot \frac{\sum_{t=1}^T \Delta y_t y_{t-1}}{\sum_{t=1}^T (y_{t-1})^2} = \frac{\frac{1}{T} \sum_{t=1}^T \Delta y_t y_{t-1}}{\frac{1}{T^2} \sum_{t=1}^T (y_{t-1})^2}. \quad (9.3.3)$$

Получение предельного распределения при нулевой гипотезе $I(1)$

Получить предельное распределение $T \cdot (\hat{\rho} - 1)$ при нулевой гипотезе $I(1)$ (случайное блуждание без тренда) очень просто. Поскольку $\{y_t\}$ является $I(1)$ -процессом без сноса, применимы части (а) и (б) утверждения 9.2 с $\xi_t = y_t$. Кроме того, поскольку Δy_t является независимым белым шумом, λ^2 (долговременная дисперсия $\{\Delta y_t\}$) равна γ_0 (дисперсия Δy_t). Таким образом,

$$\omega_{1T} \equiv \frac{1}{T} \sum_{t=1}^T \Delta y_t y_{t-1} \xrightarrow{d} \frac{\gamma_0}{2} W(1)^2 - \frac{\gamma_0}{2} \equiv \omega_1. \quad (9.3.4)$$

$$\omega_{2T} \equiv \frac{1}{T^2} \sum_{t=1}^T (y_{t-1})^2 \xrightarrow{d} \gamma_0 \int_0^1 W(r)^2 dr = \omega_2. \quad (9.3.5)$$

Как отмечалось в указанном утверждении, сходимость является совместной, так что $\omega_T \equiv (\omega_{1T}, \omega_{2T})'$ для ω_T сходится к 2×1 случайному вектору $\omega \equiv (\omega_1, \omega_2)'$. Поскольку (9.3.3) является непрерывной функцией от ω_T (или равно ω_{1T}/ω_{2T}), из леммы 2.3(b) следует, что

$$\begin{aligned} T \cdot (\hat{\rho} - 1) &= \frac{\frac{1}{T} \sum_{t=1}^T \Delta y_t y_{t-1}}{\frac{1}{T^2} \sum_{t=1}^T (y_{t-1})^2} = \frac{\omega_{1T}}{\omega_{2T}} \xrightarrow{d} \frac{\omega_1}{\omega_2} = \frac{\frac{\gamma_0}{2} W(1)^2 - \frac{\gamma_0}{2}}{\gamma_0 \int_0^1 W(r)^2 dr} = \frac{\frac{1}{2}(W(1)^2 - 1)}{\int_0^1 W(r)^2 dr} \equiv D_{DF}^{\rho}. \end{aligned} \quad (9.3.6)$$

Тестовая статистика $T \cdot (\hat{\rho} - 1)$ называется **ρ -статистикой Дики — Фуллера** (*Dickey — Fuller, или DF, ρ statistic*). Стоит отметить несколько моментов:

- Невырожденное предельное распределение имеет $T \cdot (\hat{\rho} - 1)$, а не обычное $\sqrt{T} \cdot (\hat{\rho} - 1)$. Оценка $\hat{\rho}$ называется **суперсостоятельной** (*superconsistent*), поскольку она сходится к истинному значению, равному единице, с более высокой скоростью (T).
- Нулевая гипотеза не специфицирует значения $\text{Var}(\varepsilon_t)$ и y_0 (начальное значение) в DGP, однако они не влияют на предельное распределение, распределение случайной величины DF_ρ . То есть предельное распределение не включает эти **мешающие параметры** (*nuisance parameters*). Поэтому мы можем использовать $T \cdot (\hat{\rho} - 1)$ в качестве тестовой статистики. Этот тест на случайное блуждание без сноса называется **ρ -тестом Дики — Фуллера** (*Dickey — Fuller ρ test*).
- Заметим, что и числитель, и знаменатель включают *один и тот же винеровский процесс*, так что они коррелированы. Поскольку $W(1)$ имеет стандартное нормальное распределение, числитель в выражении для DF_ρ можно было бы записать как $(\chi^2(1) - 1)/2$. Но при этом стало бы неясным сделанное здесь замечание.

Обычный t -тест для нулевой гипотезы $\rho = 1$ также имеет предельное распределение. Это всего лишь вопрос простой алгебры — показать, что t -значение $\rho = 1$ можно записать как

$$t = \frac{\hat{\rho} - 1}{s / \sqrt{\sum_{t=1}^T (y_{t-1})^2}} = \frac{\frac{1}{T} \sum_{t=1}^T \Delta y_t y_{t-1}}{s \cdot \sqrt{\frac{1}{T^2} \sum_{t=1}^T (y_{t-1})^2}}, \quad (9.3.7)$$

где s является стандартной ошибкой регрессии:

$$s \equiv \sqrt{\frac{1}{T-1} \sum_{t=1}^T (y_t - \hat{\rho} y_{t-1})^2}. \quad (9.3.8)$$

Легко доказать (см. контрольный вопрос 3), что s^2 является состоятельной оценкой для $\sigma_\varepsilon^2 (\equiv \text{Var}(\Delta y_t))$. Из этого, (9.3.4) и (9.3.5), тогда немедленно следует, что

$$t \xrightarrow{d} \frac{\frac{1}{2}(W(1)^2 - 1)}{\sqrt{\int_0^1 W(r)^2 dr}} \equiv \tau_{DF_\rho}. \quad (9.3.9)$$

Это предельное распределение также не зависит от мешающих параметров. Поэтому t -значение можно использовать для тестирования, однако

при этом нельзя использовать критические значения стандартного нормального распределения; критические значения следует брать из (приведенных ниже) таблиц для DF_t . Этот тест называется **DF t -тестом** (*DF t test*).

Суммируем изложенное.

Утверждение 9.3 (тесты Дики — Фуллера на случайное блуждание без сноса, случай без константы): Предположим, что $\{y_t\}$ является случайным блужданием без сноса (так что $\{\Delta y_t\}$ является независимым белым шумом) с $E(y_{1t}^2) < \infty$. Рассмотрим регрессию y_t на y_{t-1} (без константы) для $t = 1, 2, \dots, T$. Тогда

$$DF_{\rho}\text{-статистика: } T \cdot (\hat{\rho} - 1) \xrightarrow{d} DF_{\rho},$$

$$DF_t\text{-статистика: } t \xrightarrow{d} DF_t,$$

где $\hat{\rho}$ является OLS-оценкой коэффициента при y_{t-1} , t является t -значением для гипотезы о том, что коэффициент при y_{t-1} равен 1, а DF_{ρ} и DF_t являются случайными величинами, определенными в (9.3.6) и (9.3.9).

Таким образом, у нас есть две тестовые статистики, $T \cdot (\hat{\rho} - 1)$ и t -значение, для одной и той же нулевой гипотезы $I(1)$. Сравнение (и/или объединение) этих DF-тестов будет обсуждаться в следующем параграфе.

В части (а) табл. 9.1 приведены критические значения для распределения на конечных выборках случайной величины $T \cdot (\hat{\rho} - 1)$ для различных объемов выборки. Сплошная кривая на рис. 9.3 изображает распределение $\hat{\rho}$ на конечных выборках для $T = 100$, показывающее, что OLS-оценка для ρ смещена вниз (то есть среднее $\hat{\rho}$ меньше 1). Все это получено методом Монте-Карло в предположении, что распределение ε_t является гауссовским с $\mu_0 = 0$. По утверждению 9.3, при возрастании объема выборки T эти критические значения сходятся к критическим значениям для распределения DF_{ρ} , приведенным в строке для $T = \infty$. (Поскольку для выполнения асимптотических результатов в утверждении 9.3 μ_0 обязательно должно быть нулевым, а распределение ε обязательно должно быть гауссовским, мы должны получить те же самые предельные критические значения, если критические значения для конечных выборок вычислять при различных предположениях о начальном значении μ_0 и распределении ε_t .) По критическим значениям для $T = \infty$ мы можем видеть, что распределение DF_{ρ} также смещено влево с 5-процентным левосторонним критическим значением, равным -8.1 (так что $\text{Prob}(DF_{\rho} < -8.1) = 5\%$), и 5-процентным правосторонним критическим значением, равным 1.28. Смещение в $\hat{\rho}$ на конечных выборках отражается в скошенности предельного распределения $T \cdot (\hat{\rho} - 1)$.

Таблица 9.1

Критические значения для теста ρ Дики – Фуллера

Объем выборки (T)	Вероятность того, что статистика меньше, чем указанное число						
	0,01	0,025	0,05	0,10	0,90	0,95	0,99
Панель (a): $T \cdot (\hat{\rho} - 1)$							
25	-11,6	-9,3	-7,3	-5,3	1,01	1,41	2,28
50	-12,8	-9,9	-7,7	-5,5	0,97	1,34	2,16
100	-13,3	-10,2	-7,9	-5,6	0,95	1,31	2,09
350	-13,6	-10,4	-8,0	-5,7	0,94	1,29	2,05
500	-13,7	-10,4	-8,0	-5,7	0,93	1,28	2,04
∞	-13,8	-10,5	-8,1	-5,7	0,93	1,28	2,03
Панель (b): $T \cdot (\hat{\rho}^2 - 1)$							
25	-17,2	-14,6	-12,5	-10,2	-0,76	0,00	1,39
50	-18,9	-15,7	-13,3	-10,7	-0,81	-0,07	1,22
100	-19,8	-16,3	-13,7	-11,0	-0,83	-0,11	1,14
250	-20,3	-16,7	-13,9	-11,1	-0,84	-0,13	1,08
500	-20,5	-16,8	-14,0	-11,2	-0,85	-0,14	1,07
∞	-20,7	-16,9	-14,1	-11,3	-0,85	-0,14	1,06
Панель (c): $T \cdot (\hat{\rho}^3 - 1)$							
25	-22,5	-20,0	-17,9	-15,6	-3,65	-2,51	-0,46
50	-25,8	-22,4	-19,7	-16,8	-3,71	-2,60	-0,67
100	-27,4	-23,7	-20,6	-17,5	-3,74	-2,63	-0,76
250	-28,5	-24,4	-21,3	-17,9	-3,76	-2,65	-0,83
500	-28,9	-24,7	-21,5	-18,1	-3,76	-2,66	-0,86
∞	-29,4	-25,0	-21,7	-18,3	-3,77	-2,67	-0,88

Источник: [Fuller, 1976 Table 2.5.], исправленный в [Fuller, 1996, Table 10.A.1].

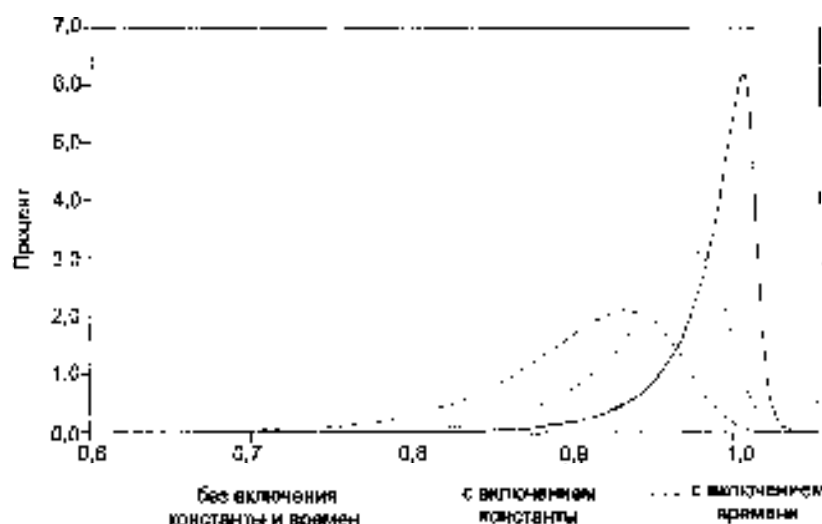


Рис. 9.3. Распределение на конечных выборках OLS-оценок для AR(1)-коэффициента, $T = 100$

Критические значения для DF_i (предельного распределения i -значения) находятся в строке для $T \rightarrow \infty$ в табл. 9.2, панель (а). Как и DF_μ , оно скошено влево. В этой таблице также приведены критические значения для конечных объемов выборки. И опять, они предполагают, что $y_t = \Pi + \varepsilon_t$ имеет нормальное распределение.

Таблица 9.2

Критические значения для теста τ Дики — Фуллера

Объем выборки (T)	Вероятность того, что статистика меньше, чем указанное число							
	0,01	0,025	0,05	0,10	0,50	0,95	0,975	0,99
Панель (а): τ								
25	-2,85	-2,26	-1,95	-1,60	0,92	1,36	1,70	2,15
50	-2,62	-2,25	-1,95	-1,61	0,91	1,31	1,56	2,06
100	-2,60	-2,24	-1,95	-1,61	0,90	1,29	1,61	2,04
250	-2,58	-2,24	-1,95	-1,62	0,89	1,28	1,63	2,02
500	-2,58	-2,23	-1,95	-1,62	0,89	1,28	1,62	2,01
∞	-2,58	-2,23	-1,95	-1,62	0,89	1,28	1,62	2,01
Панель (b): τ^*								
25	-3,75	-3,33	-2,99	-2,64	-0,37	0,00	0,34	0,71
50	-3,59	-3,23	-2,93	-2,60	-0,41	-0,04	0,29	0,66
100	-3,50	-3,17	-2,90	-2,58	-0,42	-0,06	0,26	0,63
250	-3,45	-3,14	-2,88	-2,58	-0,42	-0,07	0,24	0,62
500	-3,44	-3,13	-2,87	-2,57	-0,44	-0,07	0,24	0,61
∞	-3,42	-3,12	-2,86	-2,57	-0,44	-0,08	0,23	0,60
Панель (с): τ^b								
25	-4,38	-3,95	-3,60	-3,24	-1,14	-0,81	-0,50	-0,15
50	-4,16	-3,80	-3,50	-3,16	-1,19	-0,87	-0,58	-0,24
100	-4,05	-3,73	-3,45	-3,15	-1,22	-0,90	-0,62	-0,28
250	-3,98	-3,69	-3,42	-3,13	-1,23	-0,92	-0,64	-0,31
500	-3,97	-3,67	-3,42	-3,13	-1,24	-0,93	-0,65	-0,32
∞	-3,96	-3,67	-3,41	-3,13	-1,25	-0,94	-0,66	-0,32

Источник: [Fuller, 1976, Table 8.5.2], исправленный в [Fuller, 1996, Table 10.A.2].

Включение постоянной составляющей

Недостатком тестов, основанных на $AR(1)$ -уравнении без постоянной составляющей, является отсутствие инвариантности к прибавлению к ряду константы. Если тест проводится для ряда в логарифмах (как в примере 9.1 ниже), изменение единицы измерения ряда приводит к прибавлению к ряду в логарифмах некоторой константы, что влияет на значение тестовой статистики. Чтобы сделать тестовую статистику инвариантной

к прибавлению константы, мы обобщаем модель, на основе которой получается тестовая статистика. Рассмотрим модель

$$y_t = \alpha + z_t, \quad z_t = \rho z_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\} - \text{независимый белый шум.} \quad (9.3.10)$$

Процесс $\{y_t\}$ получается прибавлением постоянной α к процессу, удовлетворяющему (9.3.1). При нулевой гипотезе $\rho = 1$ $\{z_t\}$ является случайным блужданием без сноса. Поскольку y_t можно записать в виде:

$$y_t = \alpha + z_0 + \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t.$$

$\{y_t\}$ также является случайным блужданием без сноса с начальным значением $y_0 \equiv \alpha + z_0$. При альтернативе $\rho < 1$ $\{y_t\}$ является стационарным AR(1) со средним α . Таким образом, класс I(0)-процессов, охватываемый моделью (9.3.10), шире, чем класс процессов, охватываемых моделью (9.3.1).

Исключая $\{z_t\}$ из (9.3.10), мы получаем AR1-уравнение с константой:

$$y_t = \alpha^* + \rho y_{t-1} + \varepsilon_t, \quad (9.3.11)$$

где

$$\alpha^* = (1 - \rho)\alpha. \quad (9.3.12)$$

Поскольку $\alpha^* = 0$ при $\rho = 1$, нулевая гипотеза I(1) (о том, что DGP является случайным блужданием без сноса) является совместной гипотезой о том, что $\rho = 1$ и $\alpha^* = 0$ в терминах коэффициентов регрессии (9.3.1). Без ограничения $\alpha^* = 0$ $\{y_t\}$ было бы случайным блужданием со сномом. Мы разработаем тесты на случайное блуждание со сномом чуть позже. До этого мы продолжаем брать в качестве нулевой гипотезу случайного блуждания без сноса.

Пусть $\hat{\rho}^n$ является здесь OLS-оценкой для ρ в (9.3.11), а t^{ρ} — t -значение для нулевой гипотезы о том, что $\rho = 1$. Должно быть ясно, что α не будет влиять на значение $\hat{\rho}^n$ или на ее OLS стандартной ошибки: прибавление одной и той же постоянной к y_t для всех t просто изменяет оцененную постоянную составляющую. Следовательно, распределение **DF ρ^{μ} -статистики** (DF ρ^{μ} statistic), $T \cdot (\hat{\rho}^n - 1)$ и **DF t^{μ} -статистики** (DF t^{μ} statistic), t^{ρ} , как на конечных выборках, так и на больших выборках, не будет зависеть от α вне зависимости от значения ρ .

Как вам будет предложено доказать в аналитическом упражнении 2, $T \cdot (\hat{\rho}^n - 1)$ сходится по распределению к случайной величине, представляемой как

$$DF_{\rho}^{\mu} \equiv \frac{\frac{1}{2}([W^{\mu}(1)]^2 - [W^{\mu}(0)]^2 - 1)}{\int_0^1 [W^{\mu}(r)]^2 dr}, \quad (9.3.13)$$

а t^{ρ} сходится по распределению к

$$DF_t^{\mu} \equiv \frac{\frac{1}{2}([W^{\mu}(1)]^2 - [W^{\mu}(0)]^2 - 1)}{\sqrt{\int_0^1 [W^{\mu}(r)]^2 dr}}. \quad (9.3.14)$$

Здесь $W^{\mu}(\cdot)$ является центрированным стандартным винеровским процессом, введенным в параграфе 9.2. Эти предельные распределения не зависят от мешающих параметров, таких как α . Основная идея доказательства заключается в следующем. Сначала получаем центрированный ряд, вычитая выборочное среднее из $\{y_t\}$ (или, равносильно, получаем OLS-остатки от регрессии $\{y_t\}$ на константу). Во втором шаге используем теорему Фриша — Во (приведенную в аналогичном упражнении 4 в главе 1), чтобы записать $\hat{\rho}' \cdot (\hat{\rho}^{\mu} - 1)$ и t -значение в виде формул, включающих центрированные ряды. Наконец, используем утверждение 9.2(c) и (d). Подведем итог.

Утверждение 9.4 (тесты Дики — Фуллера о случайном блуждании без сноса, случай с постоянной составляющей): Предположим, что $\{y_t\}$ является случайным блужданием без сноса (так что $\{\Delta y_t\}$ является независимым белым шумом) с $E(y_0^2) < \infty$. Рассмотрим регрессию y_t на $(1, y_{t-1})$ для $t = 1, 2, \dots, T$. Тогда

$$DF_{\rho'} \text{ — статистика: } T \cdot (\hat{\rho}' - 1) \xrightarrow{d} DF_{\rho'}^{\mu},$$

$$DF_{t^{\mu}} \text{ — статистика: } t^{\mu} \xrightarrow{d} DF_{t^{\mu}}^{\mu},$$

где $\hat{\rho}'$ — OLS-оценка коэффициента при y_{t-1} , t^{μ} — t -значение для гипотезы о том, что коэффициент при y_{t-1} равен 1, а $DF_{\rho'}^{\mu}$ и $DF_{t^{\mu}}^{\mu}$ — две случайные величины, определенные в (9.3.13) и (9.3.14).

Критические значения для $DF_{\rho'}^{\mu}$ приведены в табл. 9.1, панель (a), а для $DF_{t^{\mu}}^{\mu}$ — в табл. 9.2, панель (b), в строке для $T = \infty$. Распределение сложнее всего сильнее, чем в случае без постоянной составляющей. В таблицах также есть критические значения для выборок конечного объема, основанные на предположении о том, что y_t имеет нормальное распределение. Распределение $\hat{\rho}'$ на конечных выборках для $T = 100$ изображено на рис. 9.3. Оно показывает более сильное смещение для $\hat{\rho}'$, чем для $\hat{\rho}$ (OLS-оценка без постоянной составляющей). В отличие от случая без включения постоянной составляющей критические значения на конечных выборках не зависят от y_0 . Это происходит потому, что, когда $\rho = 1$, изменение начального значения y_0 просто прибавляет одну и ту же постоянную к y_t для всех t , так что значения $DF_{\rho'}^{\mu}$ и t^{μ} -статистика инвариантны к y_0 .

Мы начали этот подраздел, сказав, что нулевая гипотеза $I(1)$ является совместной гипотезой о том, что $\rho = 1$ и $\mu^* = 0$. Тем не менее упомянутые тестовые статистики используются для проверки гипотезы о том, что $\rho = 1$. Для тестов на единичный корень это практика в профессии; тестовые статистики для совместной гипотезы используются редко.

Пример 9.1 (являются ли обменные курсы случайным блужданием?): Пусть y_t является спотовым обменным курсом иена/доллар. Чтобы сделать тесты инвариантными к выбору единиц (например, иена/доллар против иена/другая валюта), мы подбираем AR(1)-модель с постоянной оказывающейся, так что применимо утверждение 9.4. Для недельных данных по обменному курсу, использованных в эмпирическом приложении к главе 6 и изображенных на рис. 6.3 на с. 454, указанные AR(1)-уравнения, оцененные посредством OLS, имеют вид:

$$y_t = 0.762 + 0.9983376 y_{t-1}, \quad R^2 = 0.997, \quad SE/R = 2.824. \\ (0.13\%) \quad (0.0010285)$$

В скобках приведены стандартные ошибки. Из-за наличия лагированной переменной y_{t-1} фактический период выборки начинается со второй недели 1975 года, а не с первой. Поэтому $T = 777$ и

$$T \cdot (\hat{\rho} - 1) \rightarrow 777 \cdot (0.9983376 - 1) \rightarrow -1.20, \\ \hat{\rho} = \frac{0.9983376 - 1}{0.0010285} \rightarrow -0.86.$$

Поскольку альтернативной является гипотеза о том, что $\rho < 1$, следует использовать односторонний тест. (Если вы полагаете, что процесс может быть взрывным AR(1) с $\rho > 1$, то вы должны применить двухсторонний тест.) Из табл. 9.1, панель (b) для DF_{ρ}^n (асимптотическое) критическое значение, которое дает 5% нижнему хвосту, равно -14.1 . То есть

$$P(DF_{\rho}^n < -14.1) = 0.05.$$

Значение -1.20 статистики теста значительно выше этого критического значения, поэтому нулевая гипотеза без сомнения не отвергается на 5%-ном уровне значимости. Возвращаясь к DF t тесту, из табл. 9.2, панель (b), находим, что 5-процентное критическое значение равно -2.86 . t значение -0.86 больше его, и поэтому нулевая гипотеза, без сомнения, не отвергается.

Включение временного тренда

Как было отмечено в параграфе 9.1, во многих экономических временных рядах проявляются временные тренды. Чтобы сделать DF-тест применимым к временным рядам с трендом, мы еще раз обобщаем модель, добавляя на этот раз линейный временной тренд:

$$y_t = \alpha + \delta \cdot t + z_t, \quad z_t = \rho z_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\} \text{ — независимый белый шум.} \\ (9.3.15)$$

Здесь δ — угловой коэффициент линейного тренда, который может быть равным, а может быть и не равным нулю. При нулевой гипотезе $\rho = 1$ $\{z_t\}$ является случайным блужданием без сноса, и y_t можно записать как

$$y_t = \alpha + \delta \cdot t + z_0 + \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t = \\ = z_0 + \delta \cdot t + (\varepsilon_1 + \dots + \varepsilon_t) \quad (9.3.16)$$

с $z_0 = \alpha + z_0$. Поэтому $\{y_t\}$ — случайное блуждание со сном, если $\delta \neq 0$, и без сноса, если $\delta = 0$. При альтернативе $\rho < 1$ процесс, являющийся суммой линейного тренда и стационарного АП(1)-процесса с нулевым средним, — стационарный относительно тренда.

Устраняя $\{z_t\}$ из (9.3.15), мы получаем:

$$y_t = \alpha^* + \delta^* \cdot t + \rho y_{t-1} + \varepsilon_t, \quad (9.3.17)$$

где

$$\alpha^* = (1 - \rho)\alpha + \rho\delta, \quad \delta^* = (1 - \rho)\delta. \quad (9.3.18)$$

Поскольку $\delta^* = 0$, когда $\rho = 1$, нулевая гипотеза $\{I\}$ (и том, что DGP является случайным блужданием с или без сноса) является совместной гипотезой о том, что $\rho = 1$ и $\delta^* = 0$ в терминах регрессионных коэффициентов. На практике, однако, тесты на единичный корень обычно фокусируются на единственном ограничении $\rho = 1$.

Как в модели AR(1) с постоянной составляющей, формулировку утверждения 9.3 можно легко модифицировать для AR(1)-модели с трендом. Пусть $\hat{\rho}$ является OLS-оценкой для ρ в (9.3.17), а t^* является t -значением для нулевой гипотезы $\rho = 1$. Как вам будет предложено доказать в аналитическом упражнении 2, $T \cdot (\hat{\rho} - 1)$ сходится по распределению к

$$DF_{\rho} = \frac{\frac{1}{2}([W^{\tau}(1)]^2 - [W^{\tau}(0)]^2 - 1)}{\int_0^1 [W^{\tau}(r)]^2 dr}, \quad (9.3.19)$$

а t^* сходится по распределению к

$$DF_t^* = \frac{\frac{1}{2}([W^{\tau}(1)]^2 - [W^{\tau}(0)]^2 - 1)}{\sqrt{\int_0^1 [W^{\tau}(r)]^2 dr}}. \quad (9.3.20)$$

Здесь $W^{\tau}(\cdot)$ является детрендованным стандартным винеровским процессом, введенным в параграфе 9.2. Основная идея доказательства — использовать теорему Фрунзе — Во, чтобы записать $T \cdot (\hat{\rho} - 1)$ и t^* с помощью формул, включающих детрендованные ряды, и затем применить утверждение 9.2(e) и (j). Подытожим сказанное.

Утверждение 9.5 (тесты Дики — Фуллера о случайном блуждании с или без сноса): Предположим, что $\{y_t\}$ является случайным блужданием с или без сноса с $E(y_0^2) < \infty$. Рассмотрим регрессию y_t на $(1, t, y_{t-1})$ для $t = 1, 2, \dots, T$. Тогда

$$DF_{\rho}^T \text{ — статистика: } T \cdot (\hat{\rho}^T - 1) \rightarrow DF_{\rho}^T, \quad (9.3.21)$$

$$DF_{t^*}^T \text{ — статистика: } t^* \rightarrow DF_{t^*}^T. \quad (9.3.22)$$

где $\hat{\rho}^T$ — OLS-оценка коэффициента при y_{t-1} , t^* — t -значение для гипотезы о том, что коэффициент при y_{t-1} равен 1, а DF_{ρ}^T и $DF_{t^*}^T$ — две случайные величины, определенные в (9.3.19) и (9.3.20).

Критические значения для DF_{ρ}^T приведены в табл. 9.1, панель (с), а для $DF_{t^*}^T$ — в табл. 9.2, панель (с). В таблицах также приведены критические значения для выборок конечного объема, основанные на предположении о том, что ε_t имеет нормальное распределение (начальное значение y_0 , так же как и параметры тренда α и δ , не нужно специфицировать для вычисления распределений на конечных выборках, поскольку они не влияют на численные значения тестовых статистик при нулевой гипотезе $\rho = 1$). Для обеих DF-статистик ρ^T - и t^* -распределение даже более сильно скошено влево, чем для случая с константой. Для всех объемов выборок, отраженных в таблицах, 99%-ые критические значения отрицательны, что означает, что более чем в 99% случаев $\hat{\rho}^T$ меньше единицы, когда фактически истинное значение ρ равно единице! Это демонстрируется штрихпунктирной кривой на рис. 9.3, которая представляет график распределения в конечных выборках $\hat{\rho}^T$ для $T = 100$.

Дадим два комментария к утверждению 9.5:

- (Численная инвариантность) Тестовые статистики $\hat{\rho}^T$ и t^* инвариантны к (α, δ) вне зависимости от значения ρ ; прибавление $\alpha + \delta \cdot t$ к временному ряду $\{y_t\}$ просто изменяет OLS-оценки α^T и δ^T , но не $\hat{\rho}^T$ и ее t -значение. Следовательно, распределения $T \cdot (\hat{\rho}^T - 1)$ и t^* в конечных выборках, так же как и на больших, не будут зависеть от (α, δ) вне зависимости от значения ρ .
- (Нужно ли включать в регрессию время?) Поскольку тестовые статистики инвариантны к значению δ в DGP, утверждение 9.5 применимо, даже когда $\delta = 0$. То есть статистики $\hat{\rho}^T$ и t^* можно использовать для проверки нулевой гипотезы о случайном блуждании без тренда. Однако если мы уверены в том, что процесс под нулевой гипотезой не имеет тренда, то следует использовать DГ-тесты $\hat{\rho}^D$ и t^D в утверждении 9.4 с регрессией, не включающей время, поскольку мощность в конечных выборках против стационарной

альтернативы в общем случае больше с $\hat{\rho}^D$ и $\hat{\rho}^F$, а не с $\hat{\rho}^T$ и $\hat{\rho}^L$ (вам попросят проверить это в упражнении 1 на метод Монте-Карло). А если вы думаете, что процесс может иметь тренд, то должны использовать $\hat{\rho}^T$ и $\hat{\rho}^L$, с компонентой времени в регрессии. Если бы не включали время в регрессию и используете критические значения из табл. 9.1, панель (h), и табл. 9.2, панель (h), в то время как DGP действительно имеет тренд, то тест не будет иметь правильного размера на больших выборках (то есть вероятность отвергнуть нулевую гипотезу, когда она верна, не будет приближаться к предопределенно специфицированному уровню значимости (номинальному размеру) при неограниченном увеличении объема выборки). Это происходит просто потому, что, когда DGP имеет тренд, предельные распределения $\hat{\rho}^D$ и $\hat{\rho}^F$ не есть DF_ρ^D или DF_ρ^F .

Контрольные вопросы

1. (Уверенность в тесте.) Пусть $\hat{\rho}$ является OLS-оценкой коэффициента, как в (9.3.2). Покажите, что $T^{-1/2} \cdot (\hat{\rho} - 1) \rightarrow_p 0$ для $0 < \rho < 1$, если $\rho = 1$.

2. (Предельное распределение для общего I(1)-процесса без наклона.)

В AR(1) модели (9.3.1) докажите предположение о том, что ошибки являются независимыми белыми шумом, значения ε_t на a_t и предположения, что $\{\varepsilon_t\}$ является общим I(0)-процессом с нулевым средним, удовлетворяющим (9.2.1)–(9.2.3). Следовательно, рассматриваемый процесс при нулевой гипотезе $\rho = 1$ является общим I(1) процессом без тренда. Покажите, что $T \cdot (\hat{\rho} - 1)$ сходится по распределению к

$$\frac{\frac{1}{2}(W(1)^2 - \frac{2}{3})}{\int_0^1 W(r)^2 dr},$$

где $W^2 =$ — непрерывная дисперсия $\{\Delta y_t\}$ и $\gamma_0 \equiv \text{Var}(\Delta y_t)$.

3. (Состоятельность s^2 .) Обычная стандартная ошибка OLS-регрессии для AR(1) уравнения x_t определяется в (9.3.6). Если $\rho < 1$, то мы знаем из параграфа 6.4, что s^2 является состоятельной оценкой для $\text{Var}(\varepsilon_t)$. Покажите, что при нулевой гипотезе $\rho = 1$ s^2 также является состоятельной для $\gamma_0 \equiv \text{Var}(\Delta y_t)$. Укажите, почему s^2 не является состоятельной для γ_0 .

$$\begin{aligned} s^2 &\equiv \frac{1}{T-1} \sum_{t=1}^T (\Delta y_t - (\hat{\rho} - 1)y_{t-1})^2 = \\ &= \frac{1}{T-1} \sum_{t=1}^T (\Delta y_t)^2 - \frac{2}{T-1} \cdot [T \cdot (\hat{\rho} - 1)] \cdot \frac{1}{T} \sum_{t=1}^T \Delta y_t y_{t-1} = \\ &= \frac{1}{T-1} \cdot [T \cdot (\hat{\rho} - 1)^2] \cdot \frac{1}{T^2} \sum_{t=1}^T (y_{t-1})^2. \end{aligned}$$

Также $E[(\Delta y_t)^2] = \gamma_0$. Используйте результат, который мы использовали постоянно: если $x_T \rightarrow_p 0$ и $y_T \rightarrow_p y$, то $x_T \cdot y_T \rightarrow_p 0$.

4. (Для теста на случайное блуждание.) Рассмотрим две регрессии. Одна является регрессией Δy_t на y_{t-1} , а другая — регрессией Δy_t на yt_{-1} . Чтобы протестировать нулевую гипотезу о том, что $\{y_t\}$ является случайным блужданием без дрейфа, какое распределение следует использовать? Значения из первой регрессии или t -значения из второй регрессии?

5. (Состоятельность DF-тестов.) Напомним, что тест называется **состоятельным** (*consistent*) против некоторой множества альтернатив, если вероятность отвержения нулевой гипотезы стремится к единице при возрастании объема выборки. Пусть $T \rightarrow (\hat{\rho} - 1)$ и предположим, как в § 8.3.

(а) Предположим, что $\{y_t\}$ является $I(0)$ -процессом с нулевым средним и что его коэффициент автокорреляции первого порядка меньше 1 (так что $E(y_t^2) > 2E(y_{t-1}y_t)$). Покажите, что

$$\lim_{T \rightarrow \infty} T \cdot (\hat{\rho} - 1) = -\infty.$$

Таким образом, DF- ρ -тест является состоятельным против общих $I(0)$ -альтернатив. **Указание:** Покажите, что если $\{y_t\}$ является $I(0)$ -процессом, то

$$\lim_{T \rightarrow \infty} \hat{\rho} = \frac{E(y_t y_{t-1})}{E(y_t^2)}.$$

(б) Покажите, что DF- t -тест является состоятельным против общих $I(0)$ -альтернатив с нулевым средним. **Указание:** Когда $\{y_t\}$ является $I(0)$ -процессом с нулевым средним, $\hat{\rho}$ сходится по вероятности к некоторому положительному числу.

6. (Влияние y_t на тестовые статистики.) Проверьте, что, когда $\rho = 1$, значение $\hat{\rho}$ не принимает значения $\hat{\rho}^*$ для n конечных выборок. Остается ли это верным, когда $\rho < 1$? [Ответ: Нет. Поскольку влияние замены y_t на $\hat{\rho}$ зависит от t , зависит ли мощность на конечных выборках против альтернативы $\rho < 1$ от y_0 ?] [Ответ: Да.]

7. (Статистика Саргана — Бхаргавы [Sargan and Bhargava, 1983].) Использование тестовой статистики, основанной на OLS оценке коэффициента ρ — не самый строгий способ получения теста на единичный корень. Для выборки $(y_1, y_2, \dots, y_T)$ статистика Саргана — Бхаргавы (*Sargan-Bhargava statistic*) определяется как

$$SB = \frac{T \sum_{t=1}^T e(y_t)^2}{\frac{1}{T} \sum_{t=1}^T (\Delta y_t)^2}.$$

(а) Проверьте, что она является обратной к статистике Дарбина — Уотсона.

(б) Покажите, что если $\{y_t\}$ является случайным блужданием без дрейфа, то

$$SB \xrightarrow{p} \int_0^1 [W(r)]^2 dr.$$

Указание: $\sum_{t=1}^T y_t^2 = \sum_{t=1}^T y_{t-1}^2 + y_T^2$. Также $y_T^2/T^2 \rightarrow 0$.

1c) Проверьте, что при $I(0)$ -альтернативе $E[(\Delta y_t)^2] \neq 0$ $N \rightarrow \infty$. (Потому этот тест отвергает нулевую гипотезу $I(1)$ при ложных альтернативах $N(0)$.)

9.4. Расширенные тесты Дики — Фуллера

Процесс $I(1)$ при нулевой гипотезе в предыдущем параграфе был случайным блужданием, которое не допускает стационарной корреляции в первых разностях. Тесты этого параграфа характеризуют стационарную корреляцию путем добавления в авторегрессионное уравнение лагированных первых разностей. Они исходят из [Dickey and Fuller, 1979] и стали известны как **расширенные тесты Дики — Фуллера (ADF) (augmented Dickey — Fuller tests)**.

Расширенная авторегрессия

Нулевая гипотеза $I(1)$ в самом простом ADF-тесте состоит в том, что $\{y_t\}$ является $ARIMA(p, 1, 0)$ -процессом, то есть $\{\Delta y_t\}$ является стационарным $AR(p)$ -процессом с нулевым средним:

$$\Delta y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \dots + \zeta_p \Delta y_{t-p} + \varepsilon_t, \quad E(\varepsilon_t^2) \equiv \sigma^2, \quad (9.4.1)$$

где $\{\varepsilon_t\}$ является независимым белым шумом, а все корни соответствующего характеристического уравнения, $1 - \zeta_1 z - \zeta_2 z^2 - \dots - \zeta_p z^p$, лежат за пределами единичного круга. (Позднее мы будем допускать, что $\{\Delta y_t\}$ является стационарным $ARMA(p, q)$ -процессом. Таким образом, процесс в первых разностях $\{\Delta y_t\}$ является $I(0)$ -процессом, поскольку его можно записать как $\Delta y_t = \omega(I)\varepsilon_t$, где

$$\omega(I) = (1 - \zeta_1 I - \zeta_2 I^2 - \dots - \zeta_p I^p)^{-1}. \quad (9.4.2)$$

Модель, использованная для получения тестовых статистик, включает этот процесс в качестве частного случая:

$$y_t = \rho y_{t-1} + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \dots + \zeta_p \Delta y_{t-p} + \varepsilon_t. \quad (9.4.3)$$

Действительно, когда $\rho = 1$, (9.4.3) редуцируется к (9.4.1).

Уравнение (9.4.3) можно также записать как $AR(p+1)$:

$$y_t = \phi_0 y_{t-1} + \phi_1 y_{t-2} + \dots + \phi_{p+1} y_{t-p-1} + \varepsilon_t, \quad (9.4.4)$$

где значения ϕ_j связаны с $(\rho, \zeta_1, \dots, \zeta_p)$ соотношениями

$$\begin{aligned} \rho &\equiv \phi_0 + \phi_1 + \dots + \phi_{p+1}, \\ \zeta_j &\equiv -(\phi_{j+1} + \phi_{j+2} + \dots + \phi_{p+1}) \quad (j = 1, 2, \dots, p). \end{aligned} \quad (9.4.5)$$

Обратно, любую авторегрессию $(p+1)$ -го порядка можно записать в виде (9.4.3). Эта форма (9.4.3) была изначально предложена в [Filer, 1976, p. 374]. Она называется **расширенной авторегрессией** (*augmented autoregression*), поскольку ее можно получить путем добавления лагированных приращений в правую часть AR(1) уравнения в уровнях (9.3.1).

Легко показать (контрольный вопрос 2), что полиномиальное уравнение, связанное с (9.4.4), имеет корни внутри единичного круга, если $\rho > 1$. Таким образом, если мы принимаем условие, что DGP является или I(1)-, или I(0) процессом, то ρ не может быть больше 1, и I(0)-альтернатива состоит в том, что $\rho < 1$.

Предельное распределение OLS-оценки

Создав модель, мы обратимся теперь к задаче получения предельного распределения OLS-оценки для ρ в расширенной авторегрессии (9.4.3) той же самой модели I(1). Частично ту же самую оценку ρ можно получить из AR($p+1$) уравнения (9.4.4), вычисляя сумму OLS-оценок $(\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_{p+1})$ (подумайте, что $\rho = \hat{\rho}_1 + \dots + \hat{\rho}_{p+1}$). Однако для целей получения предельного распределения расширенной регрессии более удобна преобразованная модель I(1) с $\rho = 1$ регрессором из (9.4.4) и инновациями I(1)-процессом без дрейфа, поскольку маркирует тот факт, который используется далее в (9.4.3) ясно, что ρ из данных комбинации являются I(0) с нулевыми средними.

Для упрощения вычислений рассмотрим частный случай $\rho = 1$ (одно лагированное приращение в y_t для AR(2)), так что расширенная авторегрессия имеет вид:

$$y_t = \rho y_{t-1} + \zeta_t \Delta y_{t-1} + \varepsilon_t. \quad (9.4.6)$$

Мы предполагаем, что выборка есть $(y_1, y_2, \dots, y_T)$, так что расширенную авторегрессию можно оценить для $t = 1, 2, \dots, T$ (или альтернативно выборка есть $(y_1, y_2, \dots, y_T)$, но суммирование в дальнейшем происходит с $t = 1$ по $t = T$, а не с $t = 1$ по $t = T$). Запишем это в виде:

$$y_t = x_t' \beta + \varepsilon_t$$

с

$$x_t = \begin{bmatrix} y_{t-1} \\ \Delta y_{t-1} \end{bmatrix}, \quad \beta = \begin{bmatrix} \rho \\ \zeta \end{bmatrix}. \quad (9.4.7)$$

Если $b = (\hat{\rho}, \hat{\zeta})'$ является OLS-оценкой коэффициентов $\beta \equiv (\rho, \zeta)'$ расширенной авторегрессии, то оценка ошибки задается следующей:

$$b - \beta = \left(\sum_{t=1}^T x_t x_t' \right)^{-1} \sum_{t=1}^T x_t \varepsilon_t. \quad (9.4.8)$$

Отдельные компоненты в (9.4.8) задаются посредством соотношений:

$$X'X \equiv \sum_{i=1}^T \mathbf{x}_i \mathbf{x}_i' = \begin{bmatrix} \sum_{i=1}^{T-1} (y_{i-1})^2 & \sum_{i=1}^{T-1} y_{i-1} \Delta y_{i-1} \\ \sum_{i=1}^{T-1} \Delta y_{i-1} y_{i-1} & \sum_{i=1}^{T-1} (\Delta y_{i-1})^2 \end{bmatrix},$$

$$\sum_{i=1}^T \mathbf{x}_i \cdot \varepsilon_i = \begin{bmatrix} \sum_{i=1}^{T-1} y_{i-1} \varepsilon_i \\ \sum_{i=1}^{T-1} \Delta y_{i-1} \varepsilon_i \end{bmatrix}. \quad (9.4.9)$$

Мы можем применить универсальный же символ трюка, который использовали в регрессии на время в параграфе 2.12. Для некоторой невырожденной матрицы Υ_T , которая специфицируется ниже, перепишем (9.4.8) в виде:

$$\Upsilon_T \cdot (b - \beta) = \Upsilon_T \begin{bmatrix} \hat{\beta} - \rho \\ \hat{\zeta}_1 - \zeta_1 \end{bmatrix} = A_T^{-1} e_T. \quad (9.4.10)$$

где

$$A_T = \Upsilon_T^{-1} \left[\sum_{i=1}^T \mathbf{x}_i \mathbf{x}_i' \right] \Upsilon_T^{-1}, \quad e_T = \Upsilon_T^{-1} \sum_{i=1}^T \mathbf{x}_i \cdot \varepsilon_i. \quad (9.4.11)$$

Мы введем новую матрицу Υ_T , чтобы $\Upsilon_T \cdot (b - \beta)$ имело невырожденное предельное распределение при нулевой гипотезе $H(1)$. Это достигается, если выбрать:

$$\Upsilon_T = \begin{bmatrix} T & 0 \\ 0 & \sqrt{T} \end{bmatrix}. \quad (9.4.12)$$

При таком выборе скалирующей матрицы Υ_T из (9.4.9) и (9.4.11) следует, что

$$A_T = \begin{bmatrix} \frac{1}{T} \sum_{i=1}^{T-1} (y_{i-1})^2 & \frac{1}{\sqrt{T}} \sum_{i=1}^{T-1} y_{i-1} \Delta y_{i-1} \\ \frac{1}{\sqrt{T}} \sum_{i=1}^{T-1} \Delta y_{i-1} y_{i-1} & \frac{1}{T} \sum_{i=1}^{T-1} (\Delta y_{i-1})^2 \end{bmatrix},$$

$$e_T = \begin{bmatrix} \frac{1}{\sqrt{T}} \sum_{i=1}^{T-1} y_{i-1} \varepsilon_i \\ \frac{1}{\sqrt{T}} \sum_{i=1}^{T-1} \Delta y_{i-1} \varepsilon_i \end{bmatrix}. \quad (9.4.13)$$

Рассмотрим теперь A_T и e_T и найдем их предельные распределения при нулевой гипотезе $\rho = 1$.

(1.1)-й элемент A_T . Так как $\{y_t\}$ является $I(1)$ -процессом без сноса, этот элемент сходится по распределению к $\lambda^2 \cdot \int W^2 dt$ по утверждению 9.2(а), где $\lambda^2 \equiv \lim$ и где $\{W_t\}$ — процесс дисперсии $\{\Delta y_t\}$.

(2.2)-й элемент A_T . Он сходится по вероятности (а, следовательно, по распределению) к $-\frac{1}{2} \cdot \text{Var}(\Delta y_t)$, поскольку $\{\Delta y_t\}$, будучи стационарным $AR(1)$, является эргодическим стационарным процессом с нулевым средним.

Внедиагональные элементы A_T . Они равны произведению $1/\sqrt{T}$ на

$$\frac{1}{T} \sum_{i=1}^T \Delta y_{t-1} \varepsilon_{t-1}, \quad (9.4.14)$$

что является средним произведением $I(0)$ -переменной с нулевым средним на $I(1)$ -переменную без дрейфа. По предположению 9.2(b), подобное произведение $(1/T) \sum_{i=1}^T \Delta y_{t-1} \varepsilon_{t-1}$ сходится по распределению к некоторой случайной величине. Легко показать (контрольный вопрос 2), что (9.4.14) также сходится к случайной величине. Поэтому умноженное на $1/\sqrt{T}$ (9.4.14), являясь суммой внедиагональных элементов A_T , сходится по вероятности (в силу закона больших чисел по распределению) к 0.

Следовательно, A_T , являющаяся образом перенормированная матрица $X'X$, является асимптотически диагональной:

$$A_T \rightarrow \begin{bmatrix} \lambda^2 \cdot \int_0^1 W(r)^2 dr & 0 \\ 0 & \sigma_\varepsilon^2 \end{bmatrix},$$

так что

$$(A_T)^{-1} \rightarrow \begin{bmatrix} (\lambda^2 \cdot \int_0^1 W(r)^2 dr)^{-1} & 0 \\ 0 & \sigma_\varepsilon^{-2} \end{bmatrix}. \quad (9.4.15)$$

Мы переходим теперь к элементам c_T .

Первый элемент c_T . Используя разложение Босердажа — Нельсона, можно показать (аналитическое упражнение 4), что

$$\frac{1}{T} \sum_{i=1}^T y_{t-1} \varepsilon_t \rightarrow_d c_1 = \frac{1}{2} \sigma_\varepsilon^2 \cdot \phi(1) \cdot [F(1)^2 - 1], \quad (9.4.16)$$

для любого ИУР-процесса с нулевым средним

$$\Delta y_t = \phi(L) \varepsilon_t. \quad (9.4.17)$$

В данном случае при нулевой гипотезе $\Delta y_t = \zeta_1 \Delta y_{t-1} + \varepsilon_t$, так что $\phi(L) = (1 - \zeta_1 L)^{-1}$ и

$$\phi(F) = \frac{1}{1 - \zeta_1}. \quad (9.4.18)$$

Второй элемент c_T . Легко показать (контрольный вопрос 4), что $\{\Delta y_{t-1} \varepsilon_t\}$ является стационарной и квадратической мартигал-разностью, поэтому

$$\text{второй элемент } c_2 = \frac{1}{\sqrt{T}} \sum_{i=1}^T \Delta y_{t-1} \varepsilon_t \rightarrow_d c_2 \sim N(0, \gamma_0 \cdot \sigma^2), \quad (9.4.19)$$

где $\gamma_0 \equiv \text{Var}(\Delta y_1)$ и $\sigma^2 \equiv \text{Var}(\varepsilon_t)$.

Подставляя (9.4.15)–(9.4.19) в (9.4.10) и заменяя, что истинное значение ρ равно единице при нулевой гипотезе $I(1)$, мы получаем, что для OLS оценки расширенной авторегрессии¹

$$T \cdot (\hat{\rho} - 1) = \left[\frac{T \cdot (\hat{\rho} - 1)}{\sqrt{T}(\hat{\zeta}_1 - \zeta_1)} \right] \rightarrow \begin{bmatrix} (\lambda^2 \cdot \int_0^1 W(r)^2 dr)^{-1} & 0 \\ 0 & \gamma_0^{-1} \end{bmatrix} \begin{bmatrix} \zeta_{c1} \\ \zeta_{c2} \end{bmatrix} = \\ = \begin{bmatrix} (\lambda^2 \cdot \int_0^1 W(r)^2 dr)^{-1} \cdot \zeta_{c1} \\ \gamma_0^{-1} \cdot \zeta_{c2} \end{bmatrix}.$$

Таким образом,

$$T \cdot (\hat{\rho} - 1) \rightarrow \frac{\sigma^2 \psi(1)}{\lambda^2} \cdot \frac{\frac{1}{2} |W(1)|^2 - 1}{\int_0^1 W(r)^2 dr} \quad \text{или} \\ \frac{\lambda^2}{\sigma^2 \psi(1)} \cdot T \cdot (\hat{\rho} - 1) \rightarrow DF_{\rho}. \quad (9.4.20)$$

$$\sqrt{T} \cdot (\hat{\zeta}_1 - \zeta_1) \rightarrow N\left(0, \frac{\sigma^2}{\gamma_0}\right), \quad (9.4.21)$$

где DF_{ρ} — что иное, как DF ρ -расширение, определенное в (9.3.6). Заметим, что оцененный коэффициент при $I(0)$ -переменной с нулевым средним (Δy_{T-1}) имеет обычную $\sqrt{T}$ асимптотику, в то время как оцененный коэффициент при $I(1)$ -переменной без сноса (y_{T-1}) имеет нестандартное предельное распределение.

Получение тестовых статистик

В предельное распределение статистики $\frac{\lambda^2}{\sigma^2 \psi(1)} \cdot T \cdot (\hat{\rho} - 1)$ в (9.4.20) мешающие параметры не входят, но эта статистика сама по себе включает мешающие параметры через $\frac{\lambda^2}{\sigma^2 \psi(1)}$. По формуле (9.2.4) для долговременной дисперсии λ^2 и (Δy_t) , это отношение равно $\pi(1)$, которое, в силу (9.4.18), оказывается равным $\frac{1}{1-\zeta_1}$ при нулевой гипотезе. Вернемся теперь к расширенной авторегрессии (9.4.6). Поскольку $\hat{\zeta}_1$, OLS-оценка для ζ_1 , состоятельна в силу (9.4.21), состоятельная оценка для $\frac{1}{1-\zeta_1}$ равна $(1 - \hat{\zeta}_1)^{-1}$. Тогда из (9.4.20) вытекает, что

$$\frac{T \cdot (\hat{\rho} - 1)}{1 - \hat{\zeta}_1} \rightarrow DF_{\rho}. \quad (9.4.22)$$

¹Здесь мы используем лемму 2.3(б) для утверждения о том, что $A_T^{-1} \sigma_T$ сходится по распределению к $A^{-1} \sigma$, где A и σ являются предельными случайными величинами для A_T и σ_T соответственно. Заметим, что в отличие от ситуаций, которые рассматриваются в стационарном случае, сходимость A_T к A есть сходимость по распределению, но не по вероятности. Мы доказали в учебнике, что A_T сходится по распределению к A , а σ_T к σ , но мы не показывали, что $A_T \sigma_T$ сходится по распределению совместно к $A \sigma$, что нужно для леммы 2.3(б). Однако, как видно из Theorem 2.2 in Billings and Woyl (1983) сходимость является совместной.

То есть требуемая коррекция для $T \cdot (\hat{\rho} - 1)$ выполняется через односторонний коэффициент t -приращения Δy_t в разнице между авторегрессией. После проведения такой коррекции предельное распределение статистики не зависит от значения параметров, контролирующей серийную корреляцию $\{\Delta z_t\}$, таких как λ^2 и σ_ε^2 .

OLS t -значение для гипотезы $\rho = 1$ также имеет предельное распределение. Его можно записать как

$$\begin{aligned} t &= \frac{\hat{\rho} - 1}{\sigma \cdot \sqrt{(1, 1) \text{ элемент матрицы } \left(\sum_{t=1}^T x_t x_t' \right)^{-1}}} = \\ &= \frac{T \cdot (\hat{\rho} - 1)}{\sigma \cdot \sqrt{(1, 1) \text{ элемент матрицы } (A_T')^{-1}}}. \end{aligned} \quad (9.4.23)$$

где σ является обычной стандартной ошибкой OLS-регрессии. Из (9.4.1b), (9.4.20), состоятельности λ^2 для σ^2 и утверждений 9.2(a) и (b) для $\varepsilon_t = y_t - \rho y_{t-1}$, легко следует, что

$$t \xrightarrow{d} DF_T, \quad (9.4.24)$$

где DF_T является DF T -распределением, определенным в (9.3.9). Следовательно, для t -значения на самом деле нет необходимости производить коррекцию того факта, что лагированные Δy_t были исключены в разностную авторегрессию.

Все эти результаты для $\rho = 1$ можно легко обобщить:

Утверждение 9.6 (расширенный тест Дики — Фуллера на единичный корень без константы): Предположим, что $\{y_t\} \sim \text{ARIMA}(p, 1, 0)$ -процесс, так что $\{\Delta y_t\}$ является стационарным $\text{AR}(p)$ -процессом, следующим (9.4.1). Рассмотрим оценивание расширенной авторегрессии (9.4.3), и пусть $\{\hat{\rho}, \hat{\zeta}_1, \hat{\zeta}_2, \dots, \hat{\zeta}_p\}$ являются OLS-оценками коэффициентов. Пусть также t является t -оценкой для гипотезы $\rho = 1$. Тогда

$$ADF \text{ } \rho\text{-статистика: } \frac{T \cdot (\hat{\rho} - 1)}{1 - \hat{\zeta}_1 - \hat{\zeta}_2 - \dots - \hat{\zeta}_p} \xrightarrow{d} DF_T, \quad (9.4.25)$$

$$ADF \text{ } t\text{-статистика: } t \xrightarrow{d} DF_T. \quad (9.4.26)$$

где DF_ρ и DF_t являются двумя случайными величинами, определенными в (9.3.6) и (9.3.9).

Тестирование гипотез о ζ

Доказательство, приводящее к этому удивительно простому результату является очень длинным, но оно имеет любопытный результат. Согласно обобщению (9.4.21) на авторегрессию p -го порядка является следующее:

$$\sqrt{T} \begin{bmatrix} \hat{\zeta}_1 - \zeta_1 \\ \hat{\zeta}_2 - \zeta_2 \\ \vdots \\ \hat{\zeta}_p - \zeta_p \end{bmatrix} \xrightarrow{d} N \left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{bmatrix} \right), \quad (9.4.27)$$

где γ_j является автоковариацией $\{\Delta y_t\}$ j -го порядка. По утверждению 6.7 это то же самое асимптотическое распределение, которое вы бы получили, если бы расширенная авторегрессия (9.4.3) оценивалась посредством OLS с ограничением $\rho = 1$, то есть если (9.4.1) оценивается посредством OLS. Легко показать (контрольный вопрос 6), что гипотезы, затрагивающие только коэффициенты при I(0)-регрессорах с нулевым средним, $(\zeta_1, \zeta_2, \dots, \zeta_p)$, можно тестировать обычным образом с обычной асимптотически управляемыми t и F -статистиками.

Для получения предельных распределений регрессионных коэффициентов мы использовали тот факт, что один из регрессоров является I(0)-процессом с нулевым средним, тогда как другой является I(1)-процессом без дрейфа. Систематическое исследование более общих случаев можно найти в [Sims, Stock and Watson, 1990] и [Watson, 1994, Section 2].

Что делать, когда p неизвестно

Утверждение 9.5 предполагает, что порядок авторегрессии p для Δy_t известен. Что мы можем сделать, когда порядок p неизвестен? Чтобы привести различие между истинным порядком p и количеством лагированных наблюдений, включенных в расширенную регрессию, мы в этом параграфе обозначаем последнее как $\hat{p}$. Мы рассмотрели аналогичную проблему в параграфе 6.4. Разлика состоит в том, что здесь DGP является стационарным процессом в первых разностях, а не в уровнях, как в параграфе 6.4, и что оцениваемое уравнение является расширенной авторегрессией со свободно оцениваемым ρ . Мы демонстрируем ниже без доказательства три результата для больших выборок относительно выбора $\hat{p}$, при которых наиболее строгим утверждение 9.6. Эти результаты применимы к более общему, чем предполагается в утверждении 9.6, классу процессов: $\{\Delta y_t\}$ является стационарным и обратимым

ARMA(p, q) с нулевым средним неизвестного порядка (хотя и с дополнительным предположением конечности четвертого момента ε_t)¹.

Таким образом, при записи процесса Δy_t в форме авторегрессии, порядок авторегрессии может быть бесконечным, как тогда, когда $q > 0$, или конечным, как тогда, когда $q = 0$.

Первый результат состоит в том, что заключенные утверждения 9.6 продолжает выполняться, когда $\hat{p}$, количество лагированных первых разностей в расширенной авторегрессии, увеличивается с объемом выборки с соответствующей скоростью.

(1) ([Said and Dickey, 1984], обобщение утверждения 9.6.) Предположим, что $\hat{p}$ удовлетворяет

$$\hat{p} \rightarrow \infty, \text{ но } \frac{\hat{p}}{T^{1/3}} \rightarrow 0 \text{ при } T \rightarrow \infty. \quad (9.4.28)$$

(То есть $\hat{p}$ стремится к бесконечности, но с более низкой скоростью, чем $T^{1/3}$.) Тогда две статистики, ADF ρ и ADF t , основанные на расширенной авторегрессии с $\hat{p}$ лагированными изменениями, имеют те же самые предельные распределения, что и указанные в утверждении 9.6.

Этот результат, однако, не дает нам практического ориентира для выбора глубины запаздываний, поскольку существует бесконечно много правил для выбора $\hat{p}$, удовлетворяющих (9.4.28). Естественный вопрос: может ли какое-либо из правил для выбора глубины запаздываний, рассмотренных в параграфе 6.4, последовательно (t -правило от общего к частному или правила, основанные на информационных критериях, используемых в данном контексте? Для освежения вашей памяти: правило, основанное на информационном критерии, полагает $\hat{p}$ равным значению j , которое минимизирует

$$\log\left(\frac{SSR_j}{j}\right) = (j+1) \times \frac{C(T)}{j}, \quad (9.4.29)$$

где SSR_j является суммой квадратов остатков в модели авторегрессии с j лагами:

$$y_t = \rho y_{t-1} + \varphi_1 \Delta y_{t-1} + \varphi_2 \Delta y_{t-2} + \dots + \varphi_j \Delta y_{t-j} + \varepsilon_t. \quad (9.4.30)$$

Компонента $C(T)/T$ убывает на $j+1$, поскольку уравнение имеет $j+1$ коэффициентов, включая коэффициент при лагированном y_t .

¹ См. параграф 8.2 относительно определения обратных ARMA(p, q)-процессов. Если ARMA(p, q)-процесс является обратным, то с ним можно работать в виде процесса авторегрессии бесконечного порядка.

Для информационного критерия Аканке (AIC) $C(T) = 2$, в то время как для байссовского информационного критерия (BIC), также называемого информационным критерием Шварца (SIC), $C(T) = \log(T)$. Для любого из этих правил $\hat{p}$ выбирается из множества $p = 0, 1, 2, \dots, p_{\max}$. В параграфе 6.4 эта верхняя граница $p_{\max}$ была подожжена равной некоторому известному целому числу, большому или равному истинному порядку процесса, в котором данные являются конечного порядка. Указанное $\hat{p}$, выбранное одним из этих правил, является функцией от данных (не только функцией от T) и поэтому случайным величина. Конечно, когда $\{\Delta_t y\}$ является стационарным и обратимым ARMA(p, q)-процессом, порядок модели авторегрессии бесконечен при $q > 0$, и верхняя граница очевидно не может быть установлена равной истинному порядку. Однако ее можно сделать возрастающей вместе с объемом выборки T . Чтобы подчеркнуть зависимость $p_{\max}$ от T , запишем это как $p_{\max}(T)$. Если указанные правила модифицировать таким образом, то остается ли верным обобщение Said — Dickey утверждения 9.6, если $\hat{p}$ выбирается одним из этих правил? Ответ, полученный в [Ng and Perron, 1995], утвердительный. Более конкретно:

- (2) Предположим, что $\hat{p}$ выбирается k -разом от общего k -критерию $k = p_{\max}(T)$, удовлетворяющим условию (9.4.28) (которое в обобщении Said — Dickey, приведенном выше, было условием для $\hat{p}$, $\hat{p} \leq n$ для $p_{\max}(T)$, и $p_{\max}(T) > c \cdot T^g$ для некоторых $c > 0$ и $0 < g < 1/3$). Тогда предельные распределения ADF ρ - и ADF t -статистик те же, что указаны в утверждении 9.6.¹
- (3) Предположим, что $\hat{p}$ выбирается по критерию AIC или BIC с $p_{\max}(T)$, удовлетворяющим условию (9.4.28). Тогда выполняется то же самое заключение².

Таким образом, не имеет значения, какие правила, использованные для выбора $\hat{p}$ — распределения ADF-статистик на больших выборках те же самые, как и в утверждении 9.6.

Предложение для выбора $p_{\max}(T)$

Распределение на конечных выборках зависит, однако, от того, какое правило мы используем, а также от выбора верхней границы $p_{\max}(T)$.

¹Это Theorem 3.3 в [Ng and Perron, 1995]. Мы используем AIC вместо (9.4.28). Из условия AIC² следует из второго условия указанного выше.

²Это Theorem 4.3 в [Ng and Perron, 1995]. Теорема не упоминает верхнюю границу $p_{\max}(T)$, но (как указал Peter Perron в частной беседе с автором) из работы [Perron and Delbort, 1988, Section 7.4.6] следует, что верхняя граница может быть (9.4.28).

Например, $\rho_{max}(T) = \lceil T^{1/4} \rceil$ (целая часть $T^{1/4}$) удовлетворяет условиям в результате (2) для последовательности t -тестов. То же относится и к $\rho_{max}(T) = \lceil 100T^{3/16} \rceil$. Важно поэтому, чтобы исследователи использовали одинаковое $\rho_{max}(T)$ при решении вопроса о порядке расширенной авторегрессии. В этом контексте актуальной является литература по методу Монте-Карло, изучающая смещение различных тестов на единичный корень в конечных выборках. Симуляции, проведенные в [Schwert, 1989], показывают, что для малых и умеренно больших выборок (от $T = 25$ до $T = 1000$) включение достаточного количества лагов в авторегрессию является важным для минимизации искажений размера теста¹. Выбор $\hat{p}$ (количества лагов, включенных в авторегрессию), который был более или менее успешным в контролировании фактического размера в исследовании Шверта, равен $\lceil 12 \cdot \left(\frac{T}{100}\right)^{1/4} \rceil$. Верхняя граница $\rho_{max}(T)$ должна, следовательно, давать правило выбора для возможности выбора значения $\hat{p}$ столь же большим. Поэтому в примерах и приложениях в этой главе в любом из правил для выбора глубины запаздываний мы используем функцию:

$$\rho_{max}(T) = \lceil 12 \cdot \left(\frac{T}{100}\right)^{1/4} \rceil \quad (\text{целая часть } 12 \cdot \left(\frac{T}{100}\right)^{1/4}). \quad (9.4.31)$$

Эта функция удовлетворяет условиям результата (2) и (3), приведенных выше.

Временной интервал выборки при выборе $\hat{p}$ включает моменты $t = \rho_{max}(T) + 2, \rho_{max}(T) + 3, \dots, T$. Первое t равно $\rho_{max}(T) + 2$, поскольку $\rho_{max}(T) + 1$ наблюдений необходимо для вычисления $\rho_{max}(T)$ лагированных приращений в расширенной регрессии. Поскольку используется только $T - \rho_{max}(T) - 1$ наблюдений для оценивания авторегрессии (9.4.30) для $j = 1, 2, \dots, \rho_{max}(T)$, объем выборки в целевой функции в (9.4.29) должен быть равным $T - \rho_{max}(T) - 1$, а не T .

$$\log\left(\frac{SSR_j}{T - \rho_{max}(T) - 1}\right) + (j+1) \times \frac{C(T - \rho_{max}(T) - 1)}{T - \rho_{max}(T) - 1}. \quad (9.4.32)$$

Включение в регрессию постоянной составляющей

Как и в DF-тестах, мы можем модифицировать ADF-тесты так, чтобы они были инвариантными к трибунлению константы к ряду. Мы допускаем, чтобы $\{y_t\}$ отличался на неспецифицированную константу от процесса

¹ Напомним, что искажения размера означают разность между фактическим, для толпы, размером (вероятностью отвержения верной нулевой гипотезы в конечных выборках) и минимальным размером (вероятностью отвергнуть верную нулевую гипотезу, когда объем выборки бесконечен).

который сводит расширенный авторегрессию без константы (9.4.3) Это сводится к замене y_t в (9.4.3) на $y_t - \alpha$, что приводит к

$$y_t = \alpha^* + \rho y_{t-1} + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \dots + \zeta_p \Delta y_{t-p} + \varepsilon_t, \quad (9.4.33)$$

где $\alpha^* = (1 - \rho)\alpha$. Как и в AR(1)-модели без константы, нулевая гипотеза H_0 является совместной гипотезой к том, что $\rho = 1$ и $\alpha^* = 0$. Однако тесты на единичный корень обычно фокусируются на единственном ограничении $\rho = 1$.

В качестве аналитического упражнения оставшем доказательство следующей

Утверждение 9.7 (расширенный тест Дики — Фуллера на единичный корень с постоянной составляющей): Предположим, что $\{y_t\}$ является ARIMA($p, 1, 0$)-процессом, так что $\{\Delta y_t\}$ — стационарный AR(p)-процесс с нулевым средним, следующий (9.4.1). Рассмотрим оцененные расширенной авторегрессии с постоянной составляющей, (9.4.33), и пусть $\{\hat{\rho}, \hat{\rho}^*, \hat{\zeta}_1, \hat{\zeta}_2, \dots, \hat{\zeta}_p\}$ являются OLS-оценками коэффициентов. Также пусть t^* будет t -значением для гипотезы $\rho = 1$. Тогда

$$ADF(\rho^*)\text{-статистика: } \frac{T \cdot (\hat{\rho}^* - 1)}{1 - \hat{\zeta}_1 - \hat{\zeta}_2 - \dots - \hat{\zeta}_p} \xrightarrow{d} DF_{\rho^*}, \quad (9.4.34)$$

$$ADF(t^*)\text{-статистика: } t^* \xrightarrow{d} DF_{t^*}, \quad (9.4.35)$$

где случайные величины DF_{ρ^*} и DF_{t^*} определены в (9.3.13) и (9.3.14).

Основная идея доказательства заключается в использовании теоремы Фриля — Ва, чтобы записать $\hat{\rho}^*$ и t -значение посредством формул, содержащих центрированные ряды, созданные из $\{y_t\}$.

Прежде чем перейти к примеру, дадим два комментария по поводу утверждения 9.7

- (Численная инвариантность.) Поскольку регрессия включает постоянную составляющую, тестовые статистики инвариантны к прибавлению константы, как следует из утверждения 9.4.
- (Обобщение Said — Dickey.) Обобщение Said — Dickey — Ng — Perron также применимо здесь: если $\{\Delta y_t\}$ является стационарным и обратным ARMA(p, q)-процессом (так что Δy_t можно записать в форме авторегрессии возможно бесконечного порядка), то ADF(ρ^*)- и t^* -статистики имеют те же самые предельные распределения, которые указаны в утверждении 9.7, при условии, что $\hat{\rho}$ (целое

лагированных приращений, включенных в авторегрессию) устанавливается либо последовательным t -правилом, либо AIC, либо BIC¹.

Пример 9.2 (ADF на ставку по казначейским обязательствам): В эмпирическом упражнении в главе 2 мы вскользь отметили, что номинальная процентная ставка R_t может иметь тренд для периода выборки, рассмотренного в [Fama, 1975]. Здесь мы проверяем, имеет ли этот процесс стохастический тренд (то есть является ли он процессом $I(1)$ без сноса). Набор данных, который мы использовали в эмпирическом приложении в главе 2, состоял из месячных наблюдений одномесячной ставки по казначейским обязательствам США. Мы берем период выборки с января 1953 г. по июль 1971 г. ($T = 223$ наблюдений), период выборки в исследовании [Fama, 1975]. В процентной ставке временной тренд не признается. Поэтому мы включаем в авторегрессию постоянную составляющую, но не включаем время.

Максимальная глубина запаздываний $p_{\max}(T)$ полагается равной 14 согласно функции (9.4.31). В процессе выбора фактической глубины запаздываний $\hat{p}$ мы фиксируем период выборки с $t = p_{\max} + 2 = 16$ (апрель 1954 г.) по $t = 223$ (июль 1971 г.) с объемом выборки, равным 208 ($= T - p_{\max}(T) - 1$). Последовательное t -правило выбирает $\hat{p}$ равным 14. Цельной функцией правила, основанного на информационных критериях, когда объем выборки фиксированный, является (9.4.32). AIC выбирает $\hat{p} = 14$, в то время как BIC выбирает $\hat{p} = 1$. При заданном $\hat{p} = 1$ мы оцениваем расширенную авторегрессию по максимальному периоду выборки, равному $t = \hat{p} + 2, \dots, T$ (с марта 1953 г. по июль 1971 г.), с 221 наблюдением. Оцененная авторегрессия имеет вид:

$$R_t = 0.088 + 0.97705 R_{t-1} \quad 0.20 \quad \Delta R_{t-1} \\ (0.037) \quad (0.0162) \quad (0.067)$$

$$R^2 = 0.944, \text{ объем выборки} = 221.$$

На основе этой оцененной регрессии мы можем вычислить

$$\text{ADF } \hat{\rho}' = 221 \times \frac{0.97705 - 1}{1 + 0.20} = -4.22, \quad \rho' = \frac{0.97705 - 1}{0.0162} = -1.42.$$

Из табл. 9.1, панель (b), 5-процентное критическое значение для ADF ρ' -статистики равно -14.1 , 5-процентное критическое значение для ADF ρ' -статистики равно -2.86 из табл. 9.2, панель (b). Для каждой из статистик нулевая гипотеза (1) не отвергается без сомнений.

¹Ng and Perron, 1996, доказали этот результат о выборе глубины запаздывания только для случая без достаточной составляющей. То, что это может быть применимо к случайной составляющей, а также на случай тренда, было подтверждено в частной беседе с одним из авторов статьи.

Включение линейного тренда

Как и в DF-тестах, здесь также можно допустить наличие линейного тренда. Заменяя ρ в расширенной авторегрессии (9.4.3) на $\rho_0 + \alpha + \delta \cdot t$, мы получаем следующую расширенную авторегрессию с линейным трендом:

$$y_t = \alpha^* + \delta^* \cdot t + \rho_0 y_{t-1} + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \dots + \zeta_p \Delta y_{t-p} + \varepsilon_t, \quad (9.4.36)$$

где

$$\alpha^* = (1 - \rho_0)\alpha + (\rho_0 - \zeta_1 - \zeta_2 - \dots - \zeta_p)\delta, \quad \delta^* = (1 - \rho_0)\delta. \quad (9.4.37)$$

Поскольку $\delta^* = 0$, когда $\rho = 1$, нулевая гипотеза $I(1)$ (о том, что процесс является $I(1)$ -процессом с или без тренда) имеет место, когда $\rho = 1$ и $\delta^* = 0$ в (9.4.36), а мы, как обычно, фокусируемся на единственном ограничении $\rho = 1$.

Объединяя методы, использованные для доказательства утверждений 9.6 и 9.7, легко доказать следующее:

Утверждение 9.8 (расширенный тест Дики — Фуллера на единичный корень с линейным временным трендом): Предположим, что $\{y_t\}$ является суммой линейного временного тренда и процесса $ARIMA(p, 1, 0)$, так что $\{\Delta y_t\}$ — стационарный $AR(p)$ -процесс, среднее которого может быть, а может и не быть равным нулю. Рассмотрим оценивание расширенной авторегрессии с трендом, (9.4.36), и пусть $(\hat{\alpha}, \hat{\delta}, \hat{\rho}^*, \hat{\zeta}_1, \hat{\zeta}_2, \dots, \hat{\zeta}_p)$ являются OLS-оценками коэффициентов. Также пусть t^* будет t значением для гипотезы $\rho = 1$. Тогда

$$ADF_{\rho^*}\text{-статистика: } \frac{T^* \cdot (\hat{\rho}^* - 1)}{1 - \hat{\zeta}_1 - \hat{\zeta}_2 - \dots - \hat{\zeta}_p} \rightarrow Df_{\rho^*}, \quad (9.4.38)$$

$$ADF_{t^*}\text{-статистика: } t^* \rightarrow Df_{t^*}, \quad (9.4.39)$$

где случайные величины Df_{ρ^*} и Df_{t^*} определены в (9.3.19) и (9.3.20).

Прежде чем перейти к примеру, сделаем три замечания:

- (Численная инвариантность к параметру тренда.) Как и в утверждении 9.5, поскольку регрессия включает постоянную составляющую α (трекя, α и δ не влияют на OLS-оценки $(\hat{\rho}, \hat{\zeta}_1, \dots, \hat{\zeta}_p)$ и на стандартные ошибки, так что (α, δ) не будут влиять на распределение ADF-статистик как в конечных, так и на больших выборках.

- (Обобщение Said — Dickey.) Обобщение Said — Dickey — Ng — Perron применимо и здесь: если $\{\Delta y_t\}$ является стационарным и обратимым процессом $ARMA(p, q)$ с возможно ненулевым средним, то $ADF \rho^*$ - и t^* -статистики имеют предельные распределения, указанные в утверждении 9.8, при условии, что $\hat{\rho}$ выбирается последовательным t -правилом или правилами AIC или BIC.
- (Нужно ли включать время в регрессию?) Тот же самый комментарий, который мы приводили по поводу выбора между DF-тестами с или без тренда, применим и к ADF-тестам. С одной стороны, если вы уверены, что процесс при нулевой гипотезе не имеет тренда, то должны использовать ADF ρ^* - и t^* -тесты из утверждения 9.7 с расширенной авторегрессией без использования времени, поскольку на конечных выборках мощность ADF-тестов в общем случае больше в расширенной авторегрессии, не включающей время. С другой стороны, если вы предполагаете, что процесс может иметь тренд, то должны использовать ADF ρ^* - и t^* -тесты из утверждения 9.8 с включением времени в расширенную авторегрессию.

Пример 9.3 (имеет ли ВВП единичный корень?): Как было показано на рис. 9.1, логарифм ВВП США имеет выраженный линейный тренд, поэтому мы включаем время в расширенную авторегрессию. Используя логарифм квартальных данных реального ВВП США с 1947:Q1 по 1998:Q1 (205 наб. значений), мы оцениваем расширенную авторегрессию (9.43б). Как и в примере 9.2, мы полагаем $p_{\max}(T) = \lfloor 2 \cdot (205/100)^{1/4} \rfloor$ (что равно 14). Так же, как и в примере 9.2, период выборки синхронизирован: $t = p_{\max}(T) - 2, \dots, T$ в процессе выбора количества лагированных приращений. Последовательное t -правило выбирает $\hat{\rho} = 12$, в то время как AIC и BIC выбирают $\hat{\rho} = 1$. При заданном $\hat{\rho} = 1$ расширенная авторегрессия, оцененная по максимальной выборке $t = \hat{\rho} + 2, \dots, T$ (с 1947:Q3 по 1998:Q1), имеет вид

$$\hat{y}_t = 0.22 + 0.00022 t + 0.0707 y_{t-1} + 0.318 \Delta y_{t-1}, \\ (0.10) \quad (0.0001) \quad (0.0137) \quad (0.006)$$

$$R^2 = 0.099, \text{ объем выборки} = 203.$$

На основе этой оцененной регрессии мы можем вывести

$$ADF \hat{\rho}^* = 203 \times \frac{0.0707388}{1 - 0.348} = -9.11, \quad \rho^* = \frac{0.0707388}{0.0137104} = 2.23.$$

Из табл. 9.1, панель (с), 5-процентное критическое значение для ADF $\hat{\rho}^*$ -статистики равно -21.7, из табл. 9.2, панель (с), 5-процентное критическое значение для ADF t^* -статистики равно -3.4). Для каждой из этих статистик нулевая гипотеза H_0 не отвергается без сомнения.

Данных, охватывающих чуть более 50 лет, может оказаться недостаточно для различения между $I(1)$ и $I(0)$ -альтернативами. (Как замечено в [Растол, 1991], среди прочего, именно протяженность данных, а не частота выборки, например квартальная против годовой, более важна для оценки мощности тестов на единичный корень, поскольку $1 - \hat{\rho}$ намного меньше для квартальных данных, чем для месячных.) Сейчас мы используем для ADF-тестов данные годового ВВП с 1869 по 1997 год¹. Поэтому $T = 129$. Как и прежде, мы полагаем $p_{\max}(T) = \lfloor 12 \cdot (129/100)^{1/4} \rfloor$ (что равно 12) и фиксируем размер выборки в процессе выбора глубины ла, заданный $\hat{p}$. Периодическое t -правило выбирает 9 для $\hat{p}$, в то время как значения выбранных AIC и BIC, равно 1. Когда $\hat{p} = 1$, регрессию, оцененную по максимальной выборке $t = \hat{p} + 2, \dots, T$ (с 1871 по 1997, 127 наблюдений), имеет вид:

$$y_t = 0.67 + 0.0014 t + 0.85886 y_{t-1} + 0.32 \Delta y_{t-1} \\ (0.18) \quad (0.0048) \quad (0.03989) \quad (0.086)$$

$$R^2 = 0.990, \text{ объем выборки} = 127.$$

ADF ρ^T и t^T -статистики равны -26.2 и -3.53 соответственно, что меньше (больше по абсолютному значению), чем соответствующие 5-процентные критические значения, равные -21.7 и -3.41 . На годовых данных, покрывающих самого более длинный диапазон времени, мы можем отвергнуть нулевую гипотезу $I(1)$ о том, что ВВП имеет стохастический тренд.

Примеры 9.1–9.3 создают впечатление легкости, с которой нулевая гипотеза $I(1)$ не отвергается, и это приводит к подозрению, что DF- и ADF-тесты могут иметь никакую мощность на конечных выборках. Проблема мощности будет изучаться в следующем параграфе.

Сводный обзор DF- и ADF-тестов и других тестов на единичный корень

Рассмотренные нами в этом и предыдущих параграфах различные $I(1)$ -процессы удовлетворяют модели (множеству DGP):

$$y_t = d_t + z_t, \quad z_t = \rho z_{t-1} + u_t, \quad \{u_t\} \sim \text{процесс с нулевым средним.} \quad (9.4.40)$$

Ограничения на (9.4.40), помимо ограничения $\rho \rightarrow 1$, которое характеризует нулевую гипотезу $I(1)$ для каждого случая, таковы:

- (1) Утверждение 9.3: $d_t = 0$, $\{u_t\}$ — независимый белый шум.
- (2) Утверждение 9.4: $d_t = \alpha$, $\{u_t\}$ — независимый белый шум.

¹ Данные годового ВВП с 1869 года доступны на сайте бюро экономического анализа (*Bureau of Economic Analysis*). Данные о ВВП из [Baik and Gordon, 1989] используются для экстраполяции ВВП до 1869 года.

- (3) Утверждение 9.5: $d_t = \alpha + \delta \cdot t$, $\{u_t\}$ — независимый белый шум.
- (4) Обобщение Said — Dickey утверждения 9.6: $d_t = 0$, $\{u_t\}$ — ARMA(p, q) с нулевым средним.
- (5) Обобщение Said — Dickey утверждения 9.7: $d_t = \alpha$, $\{u_t\}$ — ARMA(p, q) с нулевым средним.
- (6) Обобщение Said — Dickey утверждения 9.8: $d_t = \alpha + \delta \cdot t$, $\{u_t\}$ — ARMA(p, q) с нулевым средним.

Контрольные вопросы

1. (Нужна истинность различения на 10%-проценте с нулевым средним и I(1)-процесс без тренда.) Предположим, что в (9.4.4) $p = 2$. Покажите, что авторегрессию третьего порядка можно эквивалентно записать как

$$y_t = a_1 \cdot y_{t-1} + a_2 \cdot (y_{t-1} - y_{t-2}) + a_3 \cdot (y_{t-1} - y_{t-2}) + \varepsilon_t.$$

Как (a_1, a_2, a_3) связаны с (ϕ_1, ϕ_2, ϕ_3) ? Какой регрессор является I(0)-процессом с нулевым средним, а какой — I(1)-процессом без тренда? Проверьте, что OLS-оценка коэффициента при y_{t-2} идентична при тестировании этого уравнения, эквивалентно равна OLS-оценке для p из (9.4.3) с $p = 2$.

2. Пусть $\phi(z) = 1 - \phi_1 z - \dots - \phi_{p+1} z^{p+1}$ является полиномом, ассоциированным с AR($p+1$)-процессом (9.4.4). Покажите, что уравнение $\phi(z) = 0$ имеет $p+1$ корней внутри единичного круга, если $\rho > 1$. Укажите: $\phi(1) = 1 - (\phi_1 + \dots + \phi_{p+1}) = 1 - \rho < 0$, если $\rho > 1$; $\phi(0) = 1 > 0$.
3. Докажите, что $(9.4.16) \rightarrow \frac{1}{2} \sum_{t=1}^T W(1)^2 + \frac{1}{2} \tau^2$. Указание: Асимптотическое представление $\left(\frac{1}{\sqrt{T}}\right)' \sum_{t=1}^T \Delta y_{t-1} y_{t-1}$ то же самое, что и для $\frac{1}{2} \sum_{t=1}^T \Delta y_t y_t$. Кроме того $\Delta y_t y_t = \Delta y_t y_{t-1} + (\Delta y_t)^2$. Используйте утверждение 9.2(b).
4. Пусть $\{\Delta y_t\}$ является I(0)-процессом с нулевым средним, удовлетворяющим (9.2.1)–(9.2.3). Стационарный AR(p) процесс с нулевым средним для $\{\Delta y_t\}$, рассмотренный в утверждении 9.6, является его частным случаем.

(а) Покажите, что $\{\Delta y_{t-1}, \varepsilon_t\}$ является стационарной и квадратичной марковской разностью. Указание: Поскольку $\{\varepsilon_t\}$ и $\{\Delta y_t\}$ совместно эргодически стационарны, $\{\Delta y_{t-1}, \varepsilon_t\}$ является эргодически стационарным. Чтобы показать, что $\{\Delta y_{t-1}, \varepsilon_t\}$ является марковской разностью, используйте закон повторных математических ожиданий. В $\{\varepsilon_1, \varepsilon_2, \dots\}$ информация больше, чем в $\{\Delta y_{t-2}, \dots, \Delta y_t, \varepsilon_{t-2}, \dots\}$.

(б) Докажите (9.4.19). Указание: Вы должны показать, что $E[(\Delta y_{t-1} \varepsilon_t)^2] = \tau_{00} \cdot \sigma^2$.

5. Проверьте (9.4.21). Указание: Из (9.4.20) и (9.2.4) вытекает, что

$$\tau \cdot (\hat{\beta} - \beta) \rightarrow \frac{1}{\sigma^2} \sum_{t=1}^T \varepsilon_t W(1)^2 - \tau$$

Кроме того, по (9.4.15) и (9.2.4)

$$c = \sqrt{\frac{1}{(1,1) \text{ элемент } (A_T)^{-1}}} \rightarrow 1/\sqrt{c(1)^2} \int_0^1 W^2 d\tau.$$

Поскольку из условия стационарности $c(1) > 0$, то $\sqrt{c(1)^2} = c(1)$.

6. (τ -тест на ζ .) В тексте утверждается (сразу вслед за (9.4.27)), что обычное t -значение для каждого ζ имеет асимптотическое стандартное нормальное распределение. Проверьте это для случая $p = 1$. **Указание:** Для $p = 1$ (9.4.27) имеет вид:

$$\sqrt{T} \cdot (\hat{\zeta}_1 - \zeta_1) \xrightarrow{d} N(0, \sigma^2/\gamma_0).$$

Вам нужно показать, что умноженная на $\sqrt{T}$ стандартная ошибка $\hat{\zeta}_1$ сходится по вероятности к квадратному корню из σ^2/γ_0 . Умноженная на $\sqrt{T}$ стандартная ошибка $\hat{\zeta}_1$ равна $c \cdot \sqrt{(2,2) \text{ элемент } (A_T)^{-1}}$.

9.5. Какой тест на единичный корень использовать

Кроме DF и ADF-тестов, доступен и ряд других тестов на единичный корень (перечень их см. в: [Maddala and Kim, 1998, Chapter 3]). Самый известный из них — тест Филлипса [Phillips, 1987] для случая (4) (упомянутого в конце предыдущего параграфа) и тест Филлипса-Перрона [Phillips and Perron, 1988], обобщение теста Филлипса, покрывающее случаи (5) и (6). (Тесты Филлипса выводятся в аналитическом упражнении 6.) Эти тесты основаны на OLS-оценке коэффициента при $m-1$ в AR(1)-уравнении, а не в расширенной авторегрессии, с долговременной дисперсией $\{\Delta u_t\}$, оцененной по остаткам в AR(1)-уравнении. Мы не предоставляем их в удобном виде, поскольку свойства этих тестов в конечных выборках довольно плохие (см. ниже). Существует новое поколение тестов на единичный корень с разумно низкими искажениями размера и хорошей мощностью. Они включают ADF-GLS-тест [Elliott, Rothenberg and Stock, 1996] и M-тесты [Perron and Ng, 1996].

Асимптотики локальности к единице

Эти и большинство других тестов на единичный корень состоятельны против всех фиксированных $I(0)$ альтернатив¹. Какой состоятельный тест следует выбрать в сравнении с другим? Напомним, что в параграфе 2.4 мы определяли асимптотическую мощность против последовательности локальных альтернатив, сходящихся к нулевой альтернативе со скоростью $\sqrt{T}$. Оказывается, что в контексте единичных корней соответствующая скорость равна T , а не $\sqrt{T}$ (см., например: [Stock, 1994,

¹ То, что DF-тесты состоятельны против общих $I(0)$ -альтернатив, проверялось в контрольном задании 5 к параграфу 9.3. По поводу состоятельности ADF-тестов см., например: [Stock, 1994, с. 2770].

pp. 2772–2773]). То есть при T , стремящемся к бесконечности, вероятность отвергнуть нулевую гипотезу $\rho = 1$, когда истинный DGP имеет вид:

$$\rho = 1 + \frac{c}{T}, \quad (9.5.1)$$

имеет предел, заключенный между 0 и 1. Этот предел и есть асимптотическая мощность. Последовательность локальных альтернатив, описанных в (9.5.1), называется **локальными к единице альтернативами** (*local-to-unity alternatives*). По определению, асимптотическая мощность равна номинальному размеру при $c = 0$. Можно показать, что DF и ADF ρ -тесты являются более мощными, чем DF и ADF t -тесты против локальных к единице альтернатив (см.: [Stock, 1994, pp. 2774–2776]). То есть для любого номинального размера и любого c асимптотическая мощность DF и ADF ρ -статистик больше, чем мощность DF и ADF t -статистик. На этом основании мы должны предпочесть основанный на ρ -тест тесту, основанному на t , однако этот вердикт отменяется, когда мы изучаем их свойства в конечных выборках.

Свойства в конечных выборках

Обращаясь к вопросу о тестировании нулевой гипотезы $I(1)$ против фиксированной, а не дрейфующей, $I(0)$ -альтернативы, двумя желательными свойствами теста в конечных выборках являются достаточно малые искажения размера и высокая мощность против $I(0)$ -альтернатив. Существует большой объем исследований методом Монте-Карло свойств на конечных выборках DF, ADF и других тестов на единичный корень (в качестве ссылок см. статьи, цитированные на с. 2777 в [Stock, 1994], а также [Ng and Perron, 1998], по поводу сравнения ADF-GLS- и M -тестов). Основные выводы заключаются в следующем.

- Практически все тесты страдают от искажений размера, в частности, когда нулевая гипотеза в некотором смысле близка к $I(0)$. M -тест в общем случае имеет наименьшее искажение размера, следующий за ним — ADF t . Искажения размера у ADF ρ -теста намного больше, чем у ADF t -теста. Например, нулевая гипотеза $I(1)$, рассмотренная в [Schwert, 1989], имеет вид:

$$y_t - y_{t-1} = \varepsilon_t + \theta \varepsilon_{t-1}, \{\varepsilon_t\} \text{ — гауссовские i.i.d.} \quad (9.5.2)$$

Если θ близко к -1 , то этот $I(1)$ -процесс очень близок к процессу белого шума, поскольку AR-корень, равный единице, почти равен MA-корню $-\theta$. Симуляции в [Schwert, 1989] показали, что тест ADF ρ - и тесты Z_ρ и Z_t Филлипса — Перрона (см. аналитическое упражнение 6) слишком часто отвергают нулевую гипотезу $I(1)$ в конечных выборках, когда θ близко к -1 . В частности, для

тестов Филиппса – Перрона в случае, когда $\theta = -0.8$ и номинальный размер равен 5%, фактический размер превышает 90% даже для выборок столь большого размера, как 1000.

- Как в случае ADF-тестов выбранная глубина запаздываний $\hat{p}$ влияет на размер и мощность в конечных выборках? Для заданного правила выбора $\hat{p}$ мощность в общем случае выше для ADF ρ -тестов, чем для ADF t -тестов.
- Мощность в общем случае намного больше для ADF-GLS и M -тестов, чем для ADF t - и ρ -тестов.

В отличие от M -статистик статистику ADF-GLS можно легко вычислить в стандартных регрессионных пакетах. В эмпирическом упражнении в этой главе нам будет предложено использовать ADF-GLS для тестирования нулевой гипотезы $I(1)$.

9.6. Приложение: Паритет покупательной способности

Теория паритета покупательной способности, ППС (*purchasing power parity, PPP*) утверждает, что для любых двух стран обменный курс равен отношению уровней цен в этих странах. Иными словами, национальные уровни цен, конвертированные в общую валюту, должны быть равными. Пусть P_t — индекс цен для Соединенных Штатов, P_t^* — индекс цен для рассматриваемой зарубежной страны (скажем, Великобритании) и S_t — обменный курс в долларах за единицу иностранной валюты. ППС утверждает, что

$$P_t = S_t \cdot P_t^*, \quad (9.6.1)$$

или, переходя к логарифмам и используя для них строчные символы¹,

$$p_t = s_t + p_t^*. \quad (9.6.2)$$

С ППС связан, но отличается от него закон единой цены (*law of one price*), который утверждает, что (9.6.1) выполняется для любого товара, и при этом P_t представляет собой цену интересующего товара в долларах, P_t^* — цену этого товара в иностранной валюте. Если международный ценовой арбитраж работает без сбоев, закон единой цены должен выполняться для каждой даты t . Если закон единой цены имеет место для всех товаров и если корзина товаров для индекса цен одинакова для двух стран, то должен выполняться ППС.

¹Уравнения (9.6.1) и (9.6.2) являются формулировкой «общепринятой» ППС. То, что называется «реальной» ППС, требует только того, чтобы выполнялось первое уравнение от (9.6.2), то есть цены роста обменного курса должны быть пропорциональны разнице между темпами роста цен в местных и зарубежных индексах цен. Относительный ППС не следует путать с более слабой версией ППС, которая будетведена позже.

В связи с таким рядом факторов, ограничивающих международный финансовый арбитраж, ППС не выполняется в точности для каждой даты¹. Более слабая версия ЛПС заключается в том, что отклонение от ППС

$$z_t \equiv x_t - y_t - \mu_t^L, \quad (9.6.3)$$

которое является реальным обменным курсом, стабильно. Даже если это не обеспечивает выполнения закона единой цены в краткосрочной перспективе, международный арбитраж должен существовать в долговременной перспективе.

Смущающая жизнеспособность модели случайного блуждания

В течение многих лет эконометрики находили слишком откровенно ложную о том, что при резких скачках обменного курса реальные обменные курсы следуют случайному блужданию. Как отметил [Rogoff, 1996], это приводило к смеси рода смущения, поскольку каждая разумная статистическая модель предполагает выполнение слабой версии ППС. Тем не менее исследования с середины 1980-х годов, в которых применялись данные за более продолжительные периоды времени, смогли отвергнуть нулевую гипотезу случайного блуждания. Хорошим примером является работа [Luthiaz and Taylor, 1996]. Чтобы обнаружить, что нулевую гипотезу случайного блуждания можно отвергнуть, эта работа использует охватывающие два столетия, с 1791 по 1990 год, ежеквартальные данные по реальным обменным курсам доллар/стерлинг и фунт/стерлинг. В дальнейшем мы применим ADF *t*-тест к тем же данным по обменному курсу доллар/стерлинг (доллар за фунт)

Реальный обменный курс доллар/стерлинг, представленный как (9.6.3) и изображенный на рис. 9.4, не обнаруживает признаков тренда. Поэтому мы не включаем время в расширенную авторегрессию. Если бы мы начали дело со слабой версии закона единой цены, с I_t^L и I_t^F , обозначающими внутреннюю и зарубежную цену определенного товара, то тогда могли бы исключить из расширенной авторегрессии константу, поскольку изменение в логичном измерении товара не влияет на y_t . Но I_t^L и I_t^F здесь являются индексами цен и номинаны в базовом году, поэтому к прибавлению константы k и k^* . Чтобы сделать тест инвариантным к таким измерениям, в расширенную авторегрессию постоянная составляющая включается. Применяя ADF-, а не DF-тест, мы можем допустить серьезную корреляцию в первых разностях z_t при нулевой гипотезе, но глубина запаздывания, выбранная посредством BIC (с $p_{max}(T) = [12 \cdot (200/100)^{1/4}] = 14$), оказывается равной 0, так что ADF

¹См., например, Section 4 в [Rogoff, 1996], описательное перечисление факторов, которые препятствуют выполнению ППС.

t -тест сводится к DF t -тесту, а расширенная авторегрессия упрощается до AD(1)-уравнению. Оцененное AD(1)-уравнение имеет вид:

$$\xi_t = 0.179 + 0.8869 \xi_{t-1}, \quad SEH = 0.071, \quad R^2 = 0.999, \quad t = 1792, \dots, 1990, \\ (0.052) \quad (0.0326)$$

DF t -статистика (t^*) равна -3.5 ($=(0.8869 - 1)/0.0326$), что значимо на 1%-ном уровне. Таким образом, мы можем, несомненно, отвергнуть гипотезу о том, что реальный обменный курс стерлинг/доллар является случайным блужданием.



Рис. 9.4. Реальный обменный курс доллар/стерлинг, 1791–1990, значение 1914 года принимается равным 0

Очевидным недостатком этого результата является то, что период выборки 1791–1990 годы охватывает периоды и фиксированного, и плавающего курса. Здесь я утверждаю, что арбитраж мировой цены работает более эффективно при фиксированных курсовых режимах. Если это так, то на периодах фиксированного курса коэффициент при ξ_{t-1} должен быть более близким к 0. [Muth and Taylor, 1996] сообщают, что если использовать простой тест Чоу, то нельзя отвергнуть гипотезу о том, что коэффициент для ξ_{t-1} один и тот же до и после введения плавающего курса в 1973 году. В эмпирическом упражнении в этой главе вам будет предложено проверить их результат, а также использовать AD(1)-GLS для проверки той же самой гипотезы.

Набор задач для главы 9

Аналитические упражнения

1. Докажите утверждение 9.2(h). Указание: По определению $\Delta \xi_t$ мы имеем $\xi_t = \xi_{t-1} + \Delta \xi_t$. Поэтому

$$(\xi_t)^2 = (\xi_{t-1} + \Delta \xi_t)^2.$$

Из этого выведите соотношение.

$$\Delta \xi_t \cdot \xi_{t-1} = \left(\frac{1}{2}\right) [(\xi_t)^2 - (\xi_{t-1})^2 - (\Delta \xi_t)^2].$$

Просуммируйте это по $t = 1, 2, \dots, T$, чтобы получить:

$$\sum_{t=1}^T \Delta \xi_t \cdot \xi_{t-1} = \left(\frac{1}{2}\right) [(\xi_T)^2 - (\xi_0)^2] - \left(\frac{1}{2}\right) \sum_{t=1}^T (\Delta \xi_t)^2.$$

разделите обе стороны на T , чтобы получить:

$$\frac{1}{T} \sum_{t=1}^T \Delta \xi_t \cdot \xi_{t-1} = \frac{1}{2} \left(\frac{\xi_T}{\sqrt{T}} \right)^2 - \frac{1}{2} \left(\frac{\xi_0}{\sqrt{T}} \right)^2 - \frac{1}{2T} \sum_{t=1}^T (\Delta \xi_t)^2. \quad (*)$$

Примените утверждение 6.9 и Эргодическую теорему к слагаемым в правой части (*).

2. (Доказательство утверждения 9.4) Пусть $\hat{\rho}^n$ — OLS-оценка для ρ в AR(1)-уравнении с постоянной составляющей, (9.3.11), и определите канонизованный ряд $\{y_{t-1}^n\}$ ($t = 1, 2, \dots, T$) как

$$y_{t-1}^n \equiv y_{t-1} = \bar{y} \equiv \frac{y_0 + y_1 + \dots + y_{T-1}}{T} \quad (t = 1, 2, \dots, T). \quad (1)$$

Эти y_{t-1}^n равны OLS-оценкам регрессии y_{t-1} на константу для $t = 1, 2, \dots, T$.

(а) Выпишите формулу

$$\hat{\rho}^n = 1 + \frac{\sum_{t=1}^T \Delta y_t y_{t-1}^n}{\sum_{t=1}^T (y_{t-1}^n)^2}. \quad (2)$$

Указание: По теореме Фрида — Бау $\hat{\rho}^n$ — канонизованный OLS-оценочный коэффициент регрессии y_t на y_{t-1}^n (без постоянной составляющей). Таким образом,

$$\hat{\rho}^n = \frac{\sum_{t=1}^T y_t y_{t-1}^n}{\sum_{t=1}^T (y_{t-1}^n)^2} \text{ так как } \hat{\rho}^n = 1 + \frac{\sum_{t=1}^T (y_t - y_{t-1}^n) y_{t-1}^n}{\sum_{t=1}^T (y_{t-1}^n)^2}. \quad (3)$$

Используйте тот факт, что $\sum_{t=1}^T y_{t-1}^n = 0$, чтобы показать, что $\sum_{t=1}^T (y_t - y_{t-1}^n) y_{t-1}^n = \sum_{t=1}^T (y_t - \bar{y}) y_{t-1}^n$.

- (б) (Лектор.) Покажите, что при нулевой гипотезе о том, что $\{y_t\}$ является $I(1)$ -процессом без сноса (необязательно случайным блужданием без сноса).

$$T \cdot (\hat{\rho}^* - 1) \rightarrow \frac{\frac{\lambda^2}{2} ([W^H(1)]^2 - [W^H(0)]^2) - \gamma_0}{\lambda^2 \cdot \int_0^1 [W^H(r)]^2 dr} \quad (4)$$

где γ_0 является дисперсией Δy_t , λ^2 — долгосрочная дисперсия $\Delta y_t \sim W^H(\cdot)$ — центрированное стандартное броуновское движение, введенное в параграфе 9.2. **Указание:** Используйте утверждение 9.2(c) и (d) с $\xi_t = y_t$.

- (с) (Тригубинин.) Проверьте, что $T \cdot (\hat{\rho}^* - 1)$ сходится к DF_ρ^y (определенному в (9.3.13)), если $\{y_t\}$ является случайным блужданием без сноса. **Указание:** Если y_t — случайное блуждание, то $\lambda^2 = \gamma_0$.
- (d) (Дополнительно.) Пусть ε является стандартной ошибкой регрессии для AR(1) уравнения с постоянной составляющей, (9.3.11). Покажите, что s^2 является состоятельной оценкой для $\gamma_0 (\equiv \text{Var}(\Delta y_t))$ при нулевой гипотезе о том, что $\{y_t\}$ является $I(1)$ -процессом без сноса (так что Δy_t является $I(0)$ -процессом с нулевым средним, удовлетворяющим (9.2.1)–(9.2.3)).

Указание: Пусть $(\hat{\alpha}^*, \hat{\rho}^*)$ — OLS-оценка для (α^*, ρ) . Покажите, что $\hat{\rho}^*$ сходится к нулю по вероятности. Запишите s^2 как

$$\begin{aligned} s^2 &= \frac{1}{T-1} \sum_{t=1}^T [(\Delta y_t - \hat{\alpha}^*) - (\hat{\rho}^* - 1)y_{t-1}]^2 = \\ &= \frac{1}{T-1} \sum_{t=1}^T (\Delta y_t - \hat{\alpha}^*)^2 - \\ &\quad - \frac{2}{T-1} \cdot (T \cdot (\hat{\rho}^* - 1)) \cdot \frac{1}{T} \sum_{t=1}^T (\Delta y_t - \hat{\alpha}^*) \cdot y_{t-1} + \\ &\quad + \frac{1}{T-1} \cdot [T \cdot (\hat{\rho}^* - 1)]^2 \cdot \frac{1}{T^2} \sum_{t=1}^T (y_{t-1})^2. \quad (5) \end{aligned}$$

Поскольку $\rho^* = 0$, легко показать, что первое слагаемое в правой части (5) сходится по вероятности к $\gamma_0 = E[(\Delta y_t)^2]$. Для доказательства того, что второе слагаемое сходится к нулю по вероятности, скажите доказавшим, что

$$\frac{1}{\sqrt{T}} \cdot \frac{1}{T} \sum_{t=1}^T y_{t-1} \rightarrow_d \lambda \int_0^1 W(r) dr, \quad (6)$$

- (с) Покажите, что $\hat{\mu}^n \rightarrow_d 1/\sigma_\varepsilon^2$, если $\{y_t\}$ является случайным блужданием без сноса. Указание: Поскольку $\{y_t\}$ является случайным блужданием без сноса, $X' = 0 = o^p(\equiv \text{дисперсия } \varepsilon_1)$. Поэтому $\hat{\mu}$ является состоятельной оценкой для σ

$$\hat{\mu}^n = \frac{\hat{\sigma}^n - 1}{n \times \sqrt{(1, 2) \text{ элемент } (X'X)^{-1}}}. \quad (7)$$

Покажите, что квадратный корень из (2, 2) элемента в $(\sum_{t=1}^T \Delta y_t \Delta y_t')^{-1}$ равен

$$\frac{1}{\sqrt{\sum_{t=1}^T (\hat{y}_t^* - \bar{y}^*)^2}}$$

где $\hat{y}_t^* \equiv [1, y_{t-1}]'$.

3. (Доказательство утверждения 9.6) Пусть $\hat{\mu}^T$ является OLS-оценкой для β в AR(1) уравнении с лагированным трендом, (9.3.17), и определите лагированный ряд $\{\hat{y}_{t-1}^T : t = 1, 2, \dots, T\}$ как

$$\hat{y}_{t-1}^T \equiv y_{t-1} - \hat{\alpha} - \hat{\delta} \cdot t, \quad (8)$$

где $(\hat{\alpha}, \hat{\delta})$ являются OLS-оценками коэффициентов для регрессии y_{t-1} на $(1, t)$ для $t = 1, 2, \dots, T$.

- (а) Выведите

$$\hat{\sigma}^{T-1} = \frac{\sum_{t=1}^T \Delta y_t \hat{y}_{t-1}^T}{\sum_{t=1}^T (\hat{y}_{t-1}^T)^2}, \quad (9)$$

Указание: По теореме Фриша-Воя, $\hat{\sigma}^T$ является OLS-оценкой коэффициента при $\hat{y}_{t-1}^T$ в регрессии y_t на $\hat{y}_{t-1}^T$ (без лагированной составляющей для времени). Таким образом,

$$\hat{\sigma}^T = \frac{\sum_{t=1}^T y_t \hat{y}_{t-1}^T}{\sum_{t=1}^T (\hat{y}_{t-1}^T)^2} \quad \text{или} \quad \hat{\sigma}^{T-1} = \frac{\sum_{t=1}^T (y_t - \hat{\alpha} - \hat{\delta} \cdot t) \cdot \hat{y}_{t-1}^T}{\sum_{t=1}^T (\hat{y}_{t-1}^T)^2}. \quad (10)$$

Покажите, что $\sum_{t=1}^T (y_t - \hat{y}_{t-1}^T) \cdot \hat{y}_{t-1}^T = \sum_{t=1}^T (y_t - y_{t-1}) \cdot \hat{y}_{t-1}^T$. (По логичности, $\sum_{t=1}^T \hat{y}_{t-1}^T = 0$ и $\sum_{t=1}^T t \cdot \hat{y}_{t-1}^T = 0$ (см. нормальные уравнения).)

- (б) (Легко.) Покажите, что при предположении о том, что $y_t = \alpha + \delta \cdot t + \varepsilon_t$ с $\{\varepsilon_t\}$ является $I(1)$ -процессом без сноса

$$\hat{\sigma}^T \rightarrow_d (\hat{\sigma}^2 - 1) \cdot \frac{\lambda^2}{2} \frac{([W^T(1)]^2 - [W^T(0)]^2) - \frac{\gamma_0}{2}}{\lambda^2 \int_0^1 [W^T(1+\tau)]^2 d\tau}, \quad (11)$$

где η_t является дисперсией Δy_t , $W^2(\cdot)$ — долгосрочная дисперсия Δy_t , и $W^2(\cdot)$ — детрендированное стандартное броуновское движение, введенное в параграфе 9.2. **Указание:** Пусть ξ_{t-1}^T — остатки от гипотетической регрессии ξ_{t-1} на $(1, t)$ для $t = 1, 2, \dots, T$, где $\xi_t = y_t - \alpha - \delta \cdot t$ (эта регрессия гипотетическая, поскольку мы не знаем α и δ). Тогда $\xi_{t-1}^T = y_{t-1}^T$ для $t = 1, 2, \dots, T$, поскольку ξ_t отличается от y_t только на линейный тренд $\alpha + \delta \cdot t$. Используйте утверждения 9.2, пункты (e) и (f). Поскольку $\Delta y_t = \delta + \Delta \xi_t$ при нулевой гипотетической дисперсии $\Delta \xi_t$ равна дисперсии Δy_t , а долгосрочная дисперсия $\Delta \xi_t$ равна долгосрочной дисперсии Δy_t .

- (c) (Трижды так!) Примеры, см. 7. (b) сходится к $N(0, \sigma^2)$ (в предельной σ (9.3.19)), если $\{y_t\}$ является случайным блужданием с или без дрейфа.

4. (Доказательство (9.4.16)) В тексте осталось недоказанным, что

$$\frac{1}{T} \sum_{t=1}^T y_{t-1}^2 \rightarrow \frac{1}{2} \cdot \sigma^2 (1 + W^2(\cdot)^2 - 1) \quad (9.4.16)$$

Докажите это при условии, что $\{y_t\}$ является И(1)-процессом без дрейфа, так что $\{\Delta y_t\}$ является И(0)-процессом с нулевым средним, удовлетворяющим (9.2.1)–(9.2.3). **Указание:** y_t в (9.4.16) является процессом инноваций в МА-представлении для $\{\Delta y_t\}$. Используя разложение Бьервада — Нелсона, имеем:

$$y_t = \alpha(1)\eta_t + \alpha_0 + (y_0 - \alpha_0), \quad \eta_t = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t.$$

Поэтому

$$\frac{1}{T} \sum_{t=1}^T y_{t-1}^2 = \alpha_0^2 + \frac{1}{T} \sum_{t=1}^T y_{t-1}^2 + \frac{1}{T} \sum_{t=1}^T y_{t-1} \eta_t + (y_0 - \alpha_0) \frac{1}{T} \sum_{t=1}^T \varepsilon_t. \quad (*)$$

Покажите, что второе и третье слагаемые в правой части уравнения сходятся по вероятности к нулю. После того как вы выполните аналитические упражнения 1, легко показать, что

$$\frac{1}{T} \sum_{t=1}^T y_{t-1}^2 \rightarrow \frac{\sigma^2}{2} (W^2(\cdot)^2 - 1).$$

5. (Доказательно, доказательство утверждения 9.7) Чтобы упростить вычисления, положим $\rho = 1$, так что расширенная авторегрессия имеет вид:

$$y_t = \alpha^* + \rho y_{t-1} + \varepsilon_t \Delta y_{t-1} + \varepsilon_t.$$

Пусть $\{\hat{\alpha}^*, \hat{\rho}, \hat{\zeta}_1\}$ — OLS-оценки коэффициентов. Покажите, что

$$\frac{T \cdot (\hat{\rho} - 1)}{1 - \hat{\zeta}_1} \rightarrow D F_{\rho}^b.$$

Указание: По теореме Фриша — Во, $(\hat{\rho}^*, \hat{\zeta}_1^*)$ численно те же самые, что и OLS-оценки коэффициентов в регрессии y_t на $(y_{t-1}^{(p)}, (\Delta y_{t-1})^{(p)})$, где $y_{t-1}^{(p)}$ — остатки от регрессии y_{t-1} на константу, а $(\Delta y_{t-1})^{(p)}$ — остатки от регрессии Δy_{t-1} на константу для $t = 1, 2, \dots, T$. Следовательно, если $\mathbf{b} = (\hat{\rho}^*, \hat{\zeta}_1^*)'$, $\beta = (\rho, \zeta_1)'$ и $\mathbf{x}_t = (y_{t-1}^{(p)}, (\Delta y_{t-1})^{(p)})'$, то они удовлетворяют (9.4.8). Поэтому

$$(1,1)\text{-й элемент в } A_T = \frac{1}{T^2} \sum_{t=1}^T (y_{t-1}^{(p)})^2,$$

$$1\text{-й элемент в } c_T = \frac{1}{T} \sum_{t=1}^T y_{t-1}^{(p)}.$$

Примените утверждение 9.2, пункты (с) и (d) с $\zeta_1 = \rho$ к этим выражениям.

6. (Тесты Филлипса.) Предположим, что $\{y_t\}$ является $I(1)$ -процессом без тренда, так что Δy_t может быть стационарно коррелированным. Тем не менее оценим АР(1)-уравнение для $\{y_t\}$ без постоянной составляющей, а не расширенную авторегрессию без константы. Пусть $\hat{\rho}$ является OLS-оценкой коэффициента при y_{t-1} . Определим

$$Z_p = T \cdot (\hat{\rho} - 1) \cdot \frac{1}{2} \cdot \left(\frac{T^2 \cdot \hat{\sigma}_{\hat{\rho}}^2}{\lambda^2} \right) \cdot (\hat{\lambda}^2 - \hat{\gamma}_0),$$

$$Z_t = \frac{s}{\hat{\lambda}} \cdot t - \frac{1}{2} \cdot \left(\frac{T \cdot \hat{\sigma}_{\hat{\rho}}^2}{s \cdot \hat{\lambda}} \right) \cdot (\hat{\lambda}^2 - \hat{\gamma}_0),$$

где $\hat{\sigma}_{\hat{\rho}}^2$ — стандартная ошибка для $\hat{\rho}$, s — стандартная ошибка OLS регрессии, $\hat{\lambda}^2$ — состоятельная оценка для λ^2 , $\hat{\gamma}_0$ — состоятельная оценка для γ_0 и t — t -значение для гипотезы $\rho = 1$.

(а) Покажите, что Z_p и Z_t можно записать как

$$Z_p = \frac{\frac{1}{T} \sum_{t=1}^T \Delta y_t y_{t-1}}{\frac{1}{T^2} \sum_{t=1}^T (y_{t-1})^2} - \frac{1}{2} \frac{(\hat{\lambda}^2 - \hat{\gamma}_0)}{\lambda^2},$$

$$Z_t = \frac{s}{\hat{\lambda}} \cdot t - \frac{\frac{1}{2} (\hat{\lambda}^2 - \hat{\gamma}_0)}{\hat{\lambda} \times \sqrt{\frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2}}.$$

Указание: Из алгебраических преобразований метода множественных квадратов,

$$\frac{\hat{\sigma}_\varepsilon^2}{s^2} = \frac{1}{\sqrt{\sum_{i=1}^n (g_i - 1)^2}}.$$

(ii) Покажите, что

$$Z_{\alpha/2} \geq DF_{\alpha/2} \text{ и } Z_{\alpha/2} \geq F_{F, \alpha/2}.$$

Указание: Используйте утверждение 9.2, а также (a) и (b) с $\xi_1 = g_1$.

7. (Дополнительно.) Покажите, что $\sigma(L)$ в 19.2.5) абсолютно суммируемо.

Указание: Доказательство в одну строку имеет вид:

$$\sum_{i=1}^{\infty} \sigma_i = \sum_{i=1}^{\infty} \left| - \sum_{j=i+1}^{\infty} \lambda_j \right| \leq \sum_{j=1}^{\infty} \sum_{i=j-1}^{\infty} |\lambda_j| = \sum_{j=1}^{\infty} j |\lambda_j| < \infty.$$

Проверьте каждое из этих равенств и неравенств.

Упражнения на метод Монте-Карло

1. (При выполнении упражнения вы получите как бытую, так и острую.) В этой секции мы хотим проверить, что в конечных выборках мощность против стационарных альтернатив больше у $DF_{\alpha/2}$ -теста, чем у $DF_{F, \alpha/2}$ -теста, то есть включение времени в AR(1)-уравнение уменьшает мощность. DGP, который мы берем в качестве стационарной альтернативы, является стационарным AR(1)-процессом:

$$g_t = \rho g_{t-1} + \varepsilon_t, \quad \rho = 0.95, \quad \varepsilon_t \sim \text{i.i.d. } N(0, 1), \quad g_0 = 0.$$

Возьмите объем выборки T равным 100 и номинальный размер равным 5%. Ваша компьютерная программа для вычисления мощности на конечных выборках должна иметь следующие данные:

- (1) Углубите счетчики данными, чтобы Должны быть два счетчика: один — для $DF_{\alpha/2}$, а другой — для $DF_{F, \alpha/2}$. Они будут указывать количество случаев, когда (поверия) нулевая гипотеза отвергается.
- (2) Начните цикл с большим количеством повторений (скажем, 1 миллион). В каждом повторении проделайте следующее:
 - (i) Сгенерируйте вектор данных $(g_1, g_2, \dots, g_T)$.

Совет для пакета GAUSS: Как бы то ни было отмечено в упражнении 1 в метод Монте-Карло для главы 1, чтобы сгенерировать $(g_1, g_2, \dots, g_T)$, исключите создание цикла внутри

текущего цикла: используйте матричные операторы. Матричная формула имеет вид:

$$y_{(T \times 1)} = r_{(T \times 1)} \cdot y_0 + A_{(T \times T)(T \times 1)} \cdot \varepsilon$$

где

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}, \quad r = \begin{bmatrix} \rho \\ \rho^2 \\ \vdots \\ \rho^T \end{bmatrix},$$

$$A = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ \rho & 1 & 0 & \dots & 0 \\ \rho^2 & \rho & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \dots & \rho & 1 \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix}.$$

Чтобы определить r , используйте Gauss-команду `vecr`. Чтобы определить A , используйте `toeplitz` и `lowmat`. r и A должны определяться в шаге (1) вне текущего цикла.

- (ii) Вычислите t^{μ} и t^{τ} по $(y_0, y_1, \dots, y_T)$. (Поэтому объем выборки фактически равен $T+1$.)
 - (iii) Увеличьте счетчик для t^{μ} на 1, если $t^{\mu} < -2.86$ (5%-ное критическое значение для DF_t^{μ}). Увеличьте счетчик для t^{τ} на 1, если $t^{\tau} < -3.41$ (5%-ное критическое значение для DF_t^{τ}).
- (3) После окончания цикла разделите счетчик для каждой статистики на число повторений для вычисления частоты отсечения нулевой гипотезы. Эта частота сходится к мощности в конечных выборках при возрастании числа Монте-Карло повторений к бесконечности. Мощность должна оказаться около 0.124 для DF_t^{μ} -теста (так что только приблизительно в 12% случаев мы отвергаем неверную нулевую гипотезу) и 0.092 для DF_t^{τ} -теста. Поэтому действительно мощность оказывается меньше, когда в регрессию включается время, по крайней мере против этого частного DGP.

2. (Искажения размера.) В этих симуляциях Монте-Карли мы измеряем искажения размера для ADF ρ^{μ} - и t^{μ} -тестов. План DGP при нулевой гипотезе — один из рассмотренных в [Schwert, 1989]:

$$y_t = \rho y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}, \quad \rho = 1, \quad \theta = -0.8,$$

$$\varepsilon_t \sim \text{i.i.d. } N(0, 1), \quad y_0 = 0, \quad \varepsilon_0 = 0.$$

Для $T = 100$ и номинального размера 5% вычислите размер на конечных выборках, или точный размер, для ρ^h и ρ^u . Было бы полезно выбрать число лагированных приращений в расширенной регрессии одним из стандартных, зависящих от данных методов (последовательные F , AIC или BIC), но, чтобы упростить программу простейшей, а также чтобы сократить время вычислений, положите его равным 4. Таким образом, уравнение регрессии, в отличие от DGP, имеет вид:

$$y_t = \text{const} + \rho y_{t-1} + \zeta_1 \cdot (y_{t-1} - y_{t-2}) + \zeta_2 \cdot (y_{t-2} - y_{t-3}) + \dots + \zeta_4 \cdot (y_{t-4} - y_{t-5}) + \text{ошибка}.$$

Поскольку выборка включает y_0 , период выборки, который может использовать четыре лагированных приращения, исключает значения t с $t = 5$ (не $p + 2 = 6$) по $t = T$, и фактический объем выборки равен $T - 4$. Проверьте, что искажение размера меньше для ADF ρ^u -теста, чем для ADF ρ^h -теста. (Размер должен быть около 0.496 для ADF ρ^h и 0.290 для ADF ρ^u .)

Эмпирические упражнения

Прежде чем отвечать, прочитайте [Lothian and Taylor, 1996, pp. 488-509] (но пропустите их обсуждение тестов Филлипса — Перрона). Файл LT.ASC содержит следующие годовые данные:

Столбец 1: год

Столбец 2: обменный курс доллар/стерлинг (ценом его S_t)

Столбец 3: U.S. WPI (wholesale price index, индекс оптовых цен), нормализованный к 100 для 1914 года (P_t^U)

Столбец 4: U.K. WPI, нормализованный к 100 для 1914 года (P_t^U)

Период выборки с 1791 по 1990 год (200 наблюдений). Это те же данные, которые использовали [Lothian and Taylor, 1996], сделанные доступными для нас. Относительно источников данных и того, как авторы объединили их для построения последовательных временных рядов, см. приложение в [Lothian and Taylor, 1996]. (Согласно приложению, обменный курс (и, возможно, WPI) является годовыми средними. Это довольно неудачно, из-за проблемы агрегирования по времени нужно использовать данные на момент времени: если некоторая переменная следует случайному блужданию на основе дневных данных, то средние годовые значения этой переменной не следуют случайному блужданию. См. об этом более подробно часть (i) ниже.)

Вычислите реальный обменный курс доллар/стерлинг как

$$z_t = \ln(S_t) - \ln(P_t) + \ln(P_t^*),$$

где S_t выражается в долларах за фунт, его возрастание означает укрепление стерлинга. График реального обменного курса на рис. 9.4 не показывает явного временного тренда, поэтому мы применим ADF F -тест. Следовательно, оцениваемая расширенная авторегрессия имеет вид:

$$\Delta z_t = \text{const} + \rho z_{t-1} + \zeta_1(\Delta z_{t-1} + \Delta z_{t-2}) + \dots + \zeta_p(\Delta z_{t-p} + \Delta z_{t-p-1}) + \text{ошибка}_t,$$

где p — число лагированных приращений, включенных в расширенную авторегрессию (которое обозначалось как $\tilde{p}$ в тексте).

(а) (ADF с автоматическим выбором лагов.) Для всего периода выборки с 1791 по 1990 год примените для выбора p (число лагированных приращений, которые нужно включить) последовательное t -критерий, AIC и BIC. Возьмите $p_{\max}$ (максимальное число лагированных изменений в расширенной авторегрессии, обозначенное в тексте как $p_{\max}(T)$) равным (9.4.31) (так что $p_{\max} = 14$). Для последовательного t -теста используйте 10-процентный уровень значимости. Вы должны разбить $p = 14$ последовательным t , 14 по AIC и 0 по BIC.) Проверьте, что значение t -статистики для $p = 0$ (при оценивании по всей выборке) соответствует значению τ_d , приведенному в Table 1 [Banerjee and Taylor, 1996].

Советы для пакета RATS: RATS позволяет вам изменять количество регрессоров в модели. Поскольку RATS OLS процедура LINREG не позволяет получать стандартные ошибки, в этой программе нельзя использовать DF F . Поэтому используйте DF F -статистику как t -значение для коэффициента лага z_{t-1} в регрессии $\Delta z_t = \alpha + \dots$. В LINREG нельзя использовать также AIC и BIC. Поэтому используйте COMPUTE. Пакет в кодах RATS z является логарифмом реального обменного курса, а dz — его первым разложением. При применении критерия последовательного t , AIC и BIC зафиксируйте период выборки с 1800 по 1990 год (он соответствует $p_{\max} + 2, \dots, T$). Коды RATS для AIC и BIC таковы:

```
linreg dz 1806:1 1990:1;# constant z(1)
compute ssr=ssr
compute akaike = log(ssr/lnobs)+lnreg*2.0/lnobs
compute schwarz = log(ssr/lnobs)+lnreg*log(lnobs)
                  /lnobs
```

```
display akaike schwarz
< run ADF with 1 through 14 lagged changes
do p=1,14
  linreg dz 1906:1 1990:1;# constant z(1); dz(1 to p)
  compute akaike = log(%res/%nobs)+inreg*2.0/%nobs
  compute schwarz = log(%res/%nobs)+4*inreg
  (log(%nobs)/%nobs)
  display akaike schwarz
end do
```

Советы для пакета TSP: В отличие от RATS TSP не позволяет вам изменять период выборки и коэффициенты регрессии в цикле. Процедура OLSQ рассчитывает БК, но не АК, — потому АК этого пакета не существует, используя команду SET.

- (b) (DF-тест на период плавающего курса.) Примените DF t^* -тест к периоду плавающего обменного курса 1974–1990 годы. Из-за лагированной зависимой переменной, z_t — первое уравнение соответствует $t = 1975$.

$$\Delta r_{USD} = \alpha_0 + \alpha_1 r_{USD} + \text{ошибка}.$$

Потому объем выборки равен 16. Можете ли вы отвергнуть нулевую гипотезу случайного блуждания на 5%-ном уровне?

- (c) (DF-тест за период золотого стандарта.) Примените DF t^* -тест к периоду золотого стандарта 1870–1913 годы. Положите $k = 1$ для 1871 года, так что объем выборки равен 43. Можете ли вы отвергнуть нулевую гипотезу случайного блуждания на 5%-ном уровне? Назначается ли оценка для ρ (коэффициент при z_{t-1} в регрессии z_t на константу и z_{t-1}) ближе к нулю, чем в (b)?

- (d) (Тест Чоу.) Отвергнув нулевую гипотезу $I(1)$, мы действуем в предположении, что логарифм реального объема производства следует стационарному AR(1)-процессу. Для тестирования нулевой гипотезы о том, что в AR(1)-уравнении нет структурного изменения, разделим выборку на два подпериода, 1791–1973 и 1974–1990 годы, и применим тест Чоу. Вы можете убедиться, что тест Чоу изначально разработан для строго экзогенных регрессоров. Если K — количество регрессоров, включая константу (2 в данном случае), — это F -тест с $(K, T - 2K)$ степенями свободы. Этот результат неприменим здесь непосредственно, поскольку в AR(1)-уравнении регрессоры являются лагированной зависимой переменной, которая не является строго экзогенной. Однако, как вы проверяли в аналитическом

упражнения 11 в главе 2 если переключить регрессоры и зависимая переменная эргодически стационарна (что выполняется, значит, поскольку $\{z_t\}$ является эргодически стационарным, когда $|\alpha| < 1$). Но на больших выборках с возрастающим объемом каждой из двух подвыборок $K \cdot F$ асимптотически имеет распределение $\chi^2(K)$. Этот результат не требует, чтобы регрессоры были строго экзогенными. Вычислите $K \cdot F$ и проверьте утверждение (см. р. 498–502 в [Lothian and Taylor, 1996]) о том, что за течение периода плавающего курса нет признаков структурного сдвига в параметрах. В другом случае, с наличием латентной зависимой переменной, возникает вопрос о том, куда включать уравнение для 1974 года,

$$\Delta \ln x_t = \alpha_0 + \alpha_1 \ln x_{t-1} + \text{ошибка},$$

в правую часть выбросив из второй. На больших выборках это не имеет значения, поместите это уравнение в первую подвыборку (а в этой выборке первая подвыборка равна 155). (Замечание. $K \cdot F$ равна бы бы около 1,26.)

- (c) (ADF-GLS.) Как было упомянуто в параграфе 9.5, ADF-GLS тест достигнет существенно более высокой мощности какой только несколько большего доказательства размера. ADF-GLS μ - и t^* -статистики вычисляются следующим образом. Как и в ADF μ - и t^* -тестах, процент нулевой гипотезе имеет вид:

$$y_t = \alpha_0 + \alpha_1 z_t, \quad z_t = \rho z_{t-1} + \alpha_2, \quad \rho \neq 1, \\ \{\alpha_t\} \text{ является } I(0)\text{-процессом с нулевым средним.} \quad (3)$$

Заметим, что $\alpha_1 = \alpha_2/\theta$. В случае, когда α_2 является константой, $\alpha_2 = 1$ и $\theta = \rho - \alpha_1$, в то время как в случае, когда α_2 является линейным трендом, $\alpha_2 = (1, t)'$ и $\theta = (\alpha, \beta)'$. Мы предполагаем, что выборка имеет объем T и это $(y_1, y_2, \dots, y_T)$.

- (i) (GLS-оценивание θ .) При оценивании регрессионного уравнения

$$y_t = \alpha_1' \theta + \text{ошибка}, \quad (t = 1, 2, \dots, T), \quad (4)$$

таким-то GLS, предполагая, что процесс ошибок является $AR(1)$ -процессом с авторегрессионных коэффициентами $\bar{\rho} = 1 + \sigma_1^2/\sigma_2^2$ (Значения σ будут специфицированы через моменты.) То есть оцените θ посредством GLS, регрессируя

$$y_1, y_2, \dots, y_T - \bar{\rho} y_1, \bar{\rho} y_2, \dots, \bar{\rho} y_{T-1} - \bar{\rho} y_T = 1$$

на

$$\alpha_1, \alpha_2, \dots, \bar{\rho} \alpha_1, \bar{\rho} \alpha_2, \dots, \bar{\rho} \alpha_{T-1}, \bar{\rho} \alpha_T = 1$$

(В отличие от истинного GLS, первые наблюдения, y_1 и x_1 , здесь не вычисляются посредством $\sqrt{1 - \bar{r}^2}$, но это не имеет значения на больших выборках.) Положите $\bar{r}$ равным

$$\bar{r} = -5$$

в центрированном случае (с $x_i = 1$),

$$\bar{r} = -13.5$$

в детрендированном случае. (Грубо говоря, если $\rho = 1 + c/T$, то такой выбор $\bar{r}$ гарантирует, что асимптотическая мощность при $c = c$ составляет примерно 50% асимптотической мощности, достигаемой идеальным тестом, построенным специально для этой альтернативы, $c = \bar{r}$.) Назовем полученную оценку $\tilde{\beta}_{GLS}$ и определим $\tilde{y}_t = y_t - x_t' \tilde{\beta}_{GLS}$. (Эти не следует путать с GLS-остатком, который равен $(y_t - \bar{\rho}y_{t-1}) - (x_t - \bar{\rho}x_{t-1})' \tilde{\beta}_{GLS}$.)

- (ii) (ADF без постоянной составляющей, основанный на остатках.) Оцените расширенную авторегрессию без постоянной составляющей, использующую $\tilde{y}_t$ вместо y_t :

$$\tilde{y}_t = \rho \tilde{y}_{t-1} + \zeta_1 (\tilde{y}_{t-1} - \tilde{y}_{t-2}) + \dots + \zeta_p (\tilde{y}_{t-p} - \tilde{y}_{t-p-1}) + \text{ошибка}.$$

t -значения для $\rho = 1$ равно ADF-GLS t -статистике; назовем ее ADF-GLS t^u , если $x_t = 1$ и ADF-GLS t^x , если $x_t = (1, t)'$. Предельное распределение ADF-GLS t^u такое же, как и у DF_t (оно табулировано в табл. 9.2, панель (a)), в то время как предельные распределения ADF-GLS t^x табулированы в Table I.C. в [Elliot et al., 1996]. Воспроизведем их левосторонние критические значения:

$$\begin{aligned} & -3.48 \text{ для } 1\%, -3.15 \text{ для } 2.5\%, \\ & 2.89 \text{ для } 5\% \text{ и } 2.57 \text{ для } 10\%. \end{aligned}$$

[Elliot et al., 1996] показали для ADF-GLS-теста, что его асимптотическая мощность выше, чем у ADF ρ - и t -тестов, и это, в сравнении с ADF тестами, его мощность на конечных выборках существенно выше, лишь с незначительным ухудшением искажений размера.

Примените ADF-GLS t^u -тест к выборке с малым стандартом 1870–1913 годов. Выберите глубину заглаживания по BIC с p_{max} (максимальное число лагированных приращений в расширенной авторегрессии, обозначенное как $p_{max}(T)$ в тексте) согласно (9.4.3) (так что $p_{max} = 9$). Можете ли вы отвергнуть нулевую гипотезу

о том, что реальный обменный курс является $I(1)$ процессом без сноса? (Напомните: BIC выберет 0 для ρ . Статистика должна быть равна -2.24 . Вы должны быть в состоянии отвергнуть нулевую гипотезу $I(1)$ на 3%-ем уровне.)

Советы для лапелла TSP: Пример кода TSP, предполагающих в этой части, что $\rho = 0$.

```
?
? (c) ADE-GG8
?
? Step 1: GLS demeaning
?
sol_rhobar=1-7/(1913-1870+1);
smpl 1870 1913;
const=1;
ot=1;
zt=z; % z is the log real exchange rate
smpl 1871 1913;
z1=z-rhobar*z(-1);
ot=const-rhobar*const(-1);
smpl 1870 1913;
clag zt ot;
z=z-@coef(1);
? %coef(1) is the GLS estimate of the mean
?
? Step 2: DF with demeaned variable
?
smpl 1871 1913;
olag z z(-1);
set t=(@coef(1)-1)/@aes(1); print t;
```

- (f) Линейная регрессионная аппроксимация h_t выражается по времени. Предположим, что y_t является случайным блужданием без сноса. Создайте на его основе выражение по времени ряда h_t формулы.

$$\bar{y}_1 = \frac{y_1 + y_2 + \dots + y_n}{n}, \quad \bar{y}_2 = \frac{y_2 + y_3 + \dots + y_n}{n}, \dots$$

(Думайте об y_i как об обменном курсе в день i , а о $\bar{y}_j$ — как о среднем дневных курсах за год j .) Покажите, что $\{\bar{y}_j\}$ не является случайным блуждением. Более конкретно, покажите, что

$$\text{Cov}(\bar{y}_{j+1} - \bar{y}_j, \bar{y}_{j+2} - \bar{y}_{j+1}) > 0.25 \text{ при } n \rightarrow \infty.$$

Ответы на избранные вопросы

Аналитические упражнения

1. Поскольку $E(\xi_1^2) < \infty$, $\xi_0/\sqrt{T}$ сходится в среднем квадратическом (и, следовательно, по вероятности) к 0 при $T \rightarrow \infty$. Таким образом, второе слагаемое в правой части (*) в указании можно игнорировать. Что касается первого слагаемого, то мы имеем:

$$\frac{\xi_T}{\sqrt{T}} = \frac{\xi_0}{\sqrt{T}} + \frac{\Delta\xi_1 + \Delta\xi_2 + \dots + \Delta\xi_T}{\sqrt{T}}.$$

Первое слагаемое в правой части этого уравнения можно игнорировать. По утверждению 6.9, второе слагаемое сходится к $N(0, \lambda^2)$, поскольку $\Delta\xi_t$ является линейным процессом с абсолютно суммируемыми МА-коэффициентами. Поэтому первое слагаемое в правой части (*) сходится по распределению к $(\lambda^2/2)X$, где $X \sim \chi^2(1)$. Поскольку $\{\Delta\xi_t\}$ является эргодически стационарным, последнее слагаемое в правой части (*) сходится по вероятности к $\frac{1}{2} E[(\Delta\xi_1)^2]$, то эргодической теореме.

5. (1.1)-й элемент A_T сходится к случайной величине, указанной в утверждении 9.2(с). Что касается первого элемента, a_T , то, по разложению Бездриджа — Нельсона:

$$y_t = c(1)y_t - y_t + (y_0 - y_1), \quad a_t \equiv x_1 + x_2 + \dots + x_t. \quad (1)$$

Следовательно,

$$y_{t-1}^p = c(1)y_{t-1}^p - y_{t-1}^p, \quad (2)$$

где

$$y_{t-1}^p = y_{t-1} - \bar{y}, \quad x_t = \frac{y_t - \dots + y_{t-1}}{T} \text{ для } t = 1, \dots, T$$

и

$$\frac{1}{T} \sum_{t=1}^T y_{t-1}^p \cdot \varepsilon_t = c(1) \frac{1}{T} \sum_{t=1}^T y_{t-1}^p \cdot \varepsilon_t - \frac{1}{T} \sum_{t=1}^T y_{t-1}^p \cdot \varepsilon_t \quad (3)$$

Второе слагаемое в правой части этого уравнения можно игнорировать асимптотически. По утверждению 2.9(d) для $\xi_t = w_t$ (случайное блуждание без сноса), первое слагаемое в правой части сходится к

$$\psi(1) \left(\frac{\sigma_w^2}{2} \right) \{ [W^w(1)]^2 - [W^w(0)]^2 - 1 \}.$$

Отсюда мы имеем:

$$\frac{1}{T} \sum_{t=1}^T y_{t-1}^p \cdot \varepsilon_t \rightarrow \frac{1}{2} \cdot \sigma^2 \cdot \psi(1) \{ [W^w(1)]^2 - [W^w(0)]^2 - 1 \}. \quad (4)$$

Матрица A_T асимптотически диагональна также и для центрированного случая. Внедиагональные элементы A_T равны $1/\sqrt{T}$, умноженному на

$$\frac{1}{T} \sum_{i=1}^T (\Delta y_{i-1})^{(p)} y_{i-1}^p. \quad (5)$$

Поскольку $\sum_{t=1}^T y_t^p = 0$, эти равны:

$$\frac{1}{T} \sum_{i=1}^T \Delta y_{i-1} y_{i-1}^p. \quad (6)$$

Путем небольших алгебраических манипуляций, аналогичных применявшимся в контрольном вопросе 3 в параграфе 9.4 и утверждении 9.2(б), легко доказать, что (6) имеет предельное распределение. Поэтому (6), умноженное на $1/\sqrt{T}$, что является внедиагональным элементом A_T , сходится к нулю по вероятности. Тогда, используя те же аргументы, которые мы использовали в тексте для случая без константы, легко доказать, что $\hat{\zeta}$ состоятельна для ζ_1 . Поэтому $\psi(1)$ состоительно оценивается посредством $1/(1 - \hat{\zeta}_1)$. Собирая все вместе, получаем:

$$\frac{T \cdot (\hat{\rho}^p - 1)}{1 - \hat{\zeta}} \rightarrow DF_p^p.$$

Эмпирические упражнения

(а) При $p = 0$ оцените AR(1)-уравнение

$$\hat{\rho} = \text{const} + \frac{0.8809}{(0.0326)} \varepsilon_{t-1}, \quad SSR = 0.995135, \quad R^2 = 0.790, \quad T = 199.$$

Значение $\hat{\rho}$ равно -3.47 , что соответствует значению в Table 1 [Lobato and Taylor, 1996]. 5-процентное критическое значение равно -2.86 по табл. 9.2, панель (b).

(b) Оцененное AR(1)-уравнение для периода 1974–1990 годов имеет вид:

$$z_t = \text{const} + 0.7703 z_{t-1}, \quad SSR = 0.130904, \quad R^2 = 0.539, \quad T = 16, \\ (0.1904)$$

$t^* = 1.41 > -2.50$. Мы не можем отвергнуть нулевую гипотезу на 5%-ном уровне.

(c) Оцененное AR(1)-уравнение для периода 1870–1913 годов имеет вид:

$$z_t = \text{const} + 0.7222 z_{t-1}, \quad SSR = 0.060429, \quad R^2 = 0.558, \quad T = 43, \\ (0.1045)$$

$t^* = -2.60 > -2.80$. Мы не можем отвергнуть нулевую гипотезу на 5%-ном уровне.

(d) Оцененное AR(1)-уравнение для периода 1791–1974 годов имеет вид:

$$z_t = \text{const} + 0.8903 z_{t-1}, \quad SSR = 0.837772, \quad R^2 = 0.805, \quad T = 183, \\ (0.0329)$$

Из этого результата и результатов в (a) и (b)

$$K \cdot F = \frac{0.203135 - (0.837772 + 0.130904)}{\frac{0.837772 + 0.130904}{(183-4)}} = 1.2\%.$$

5-процентное критическое значение для $\chi^2(2)$ равно 5.99. Поэтому мы принимаем нулевую гипотезу о стабильности параметров.

(e) ADF-GLS $t^* = -2.24$. 5%-ное критическое значение в табл. 9.2, строка (a) равно -1.95. Поэтому мы можем отвергнуть нулевую гипотезу (1)

Литература

- Bajbou, R., 1996, "Long Memory Processes and Fractional Integration in Econometrics," *Journal of Economics*, 73, 5–39.
- Balks, N. and R. Gordon, 1989, "The Estimation of Prewar Gross National Product: Methodology and New Evidence," *Journal of Political Economy*, 97, 389–42.
- Chan, N. and C. Wei, 1988, "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *Annals of Statistics*, 16, 367–401.
- Dickey, D., 1976, "Estimator and Hypothesis Testing in Nonstationary Time Series," Ph.D. Dissertation, Iowa State University.
- Dickey, D. and W. Fuller, 1979, "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

- Ellert, G., T. Rotherberg, and J. Stock, 1996, "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813-836.
- Fama, E., 1975, "Short-Term Interest Rates as Predictors of Inflation," *American Economic Review*, 65, 269-281.
- Fuller, W., 1976, *Introduction to Statistical Time Series*, New York: Wiley.
- , 1996, *Introduction to Statistical Time Series* (2d ed.), New York: Wiley.
- Hal, P., and C. Heyde, 1980, *Martingale Limit Theory and Its Application*, New York: Academic.
- Hamilton, J., 1994, *Time Series Analysis*, Princeton: Princeton University Press.
- Hannan, E. J., and M. Deistler, 1968, *The Statistical Theory of Linear Systems*, New York: Wiley.
- Kwiatkowski, D., P. Phillips, P. Schmidt, and Y. Shin, 1992, "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root," *Journal of Economics*, 34, 139-175.
- Laybourne, S., and B. McCabe, 1991, "A Consistent Test for a Unit Root," *Journal of Business and Economic Statistics*, 12, 157-166.
- Lawson, J., and M. Taylor, 1996, "Real Exchange Rate Behavior: The Recent Find from the Perspective of the Past Two Centuries," *Journal of Political Economy*, 104, 488-509.
- Maddala, G. S., and In-Mun Kim, 1998, *Unit Roots, Cointegration, and Structural Change*, New York: Cambridge University Press.
- Ng, S., and P. Perron, 1995, "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association*, 90, 268-281.
- , 1998, "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," working paper, mimeo.
- Perron, P., 1991, "Test Consistency with Varying Sampling Frequency," *Econometric Theory*, 7, 341-368.
- Perron, P., and S. Ng, 1996, "Useful Modifications to Unit Root Tests with Dependent Errors and Their Local Asymptotic Properties," *Review of Economic Studies*, 63, 435-463.
- Phillips, P., 1987, "Time Series Regression with a Unit Root," *Econometrica*, 55, 277-301.
- Phillips, P., and S. Durlauf, 1990, "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473-490.
- Phillips, P., and P. Perron, 1988, "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335-346.
- Rogoff, K., 1996, "The Purchasing Power Parity Puzzle," *Journal of Economic Literature*, 34, 647-668.
- Said, S., and D. Dickey, 1984, "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71, 599-607.
- Sargan, G., and A. Bhargava, 1984, "Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk," *Econometrica*, 51, 153-171.
- Schwert, G., 1989, "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 7, 147-158.

- Sims, C., J. Stock, and M. Watson. 1990. "Inference in Linear Time Series Models with Some Unit Roots." *Econometrica*, 58, 113–144.
- Stock, J., 1994. "Unit Roots, Structural Breaks and Trends," Chapter 46 in R. Engle and D. McLeod (eds.), *Handbook of Econometrics*, Volume IV, New York: North Holland.
- Tanaka, K., 1990, "Testing for a Moving Average Unit Root," *Econometric Theory*, 6, 453–474.
- . 1996. *Time Series Analysis*, New York: Wiley.
- Watson, M., 1994. "Vector Autoregressions and Cointegration," Chapter 47 in R. Engle and D. McLeod (eds.), *Handbook of Econometrics*, Volume IV, New York: North Holland.

Глава 10. Коинтеграция

Как иллюстрируют примеры из предыдущего параграфа, многие экономические временные ряды могут характеризоваться как $I(1)$. Но если часто их линейные комбинации оказываются стационарными. Такие перекрестные вызовываются **коинтегрированными** (*cointegrated*), а веса в линейной комбинации называются **коинтегрирующим вектором** (*cointegrating vector*). В этой главе изучаются коинтегрированные системы, а именно векторные $I(1)$ -процессы, элементы которых коинтегрированы.

В предыдущей главе мы определили одностепенные $I(0)$ - и $I(1)$ -процессы. Первый параграф данной главы представляет обобщение этих понятий на векторные процессы и определяет понятие ранга для векторных $I(1)$ -процессов. В параграфе 10.2 приводятся альтернативные представления коинтегрированной системы. В параграфе 10.3 более подробно изучается тест приверженности, является ли векторный $I(1)$ -процесс коинтегрированным. Техника для оценивания коинтегрирующих векторов и получения статистических выводов о них разрабатывается в параграфе 10.4.

В качестве приложения в параграфе 10.5 рассматривается функция спроса на деньги. Авторафин увеличит предложение денег, номинальная процентная ставка и логарифм реального дохода называются $I(1)$, но существование стабильной функции спроса на деньги подразумевает, что эти переменные коинтегрированы, с коэффициентами в функции спроса на деньги, формирующими коинтегрирующий вектор. Методы параграфа 10.4 используются для оценивания этого коинтегрирующего вектора.

В отличие от других глав мы не представляем результаты в виде ряда утверждений, поскольку уровень предмета таков, что формулировать предположения трудно без введения дополнительных технических терминов. Однако для читателей, которые ищут решения технических деталей, мы указываем в тексте ссылки на источники, в которых эти предположения сформулированы формально.

Замечание об обозначениях: В отличие от других глав символ $\sim$ используется в качестве размерности векторных процессов.

и, как и в предыдущих частях, символ « T » используется для обозначения «выборки», а « t » — индекс наблюдения. Также, если y_t является t -м наблюдением n -мерного векторного процесса, его j -й элемент будет обозначаться как y_{jt} вместо y_{jt} . Мы прокиваем эти комментарии, чтобы соответствовать обозначениям, используемым в большинстве оригинальных работ по коинтеграции.

10.1. Коинтегрированные системы

На рис. 10.1(а) изображены логарифмы квартальных реального совокупного личного располагаемого дохода на душу населения и личных расходов на потребление для США с 1947:Q1 по 1998:Q1. Оба ряда имеют

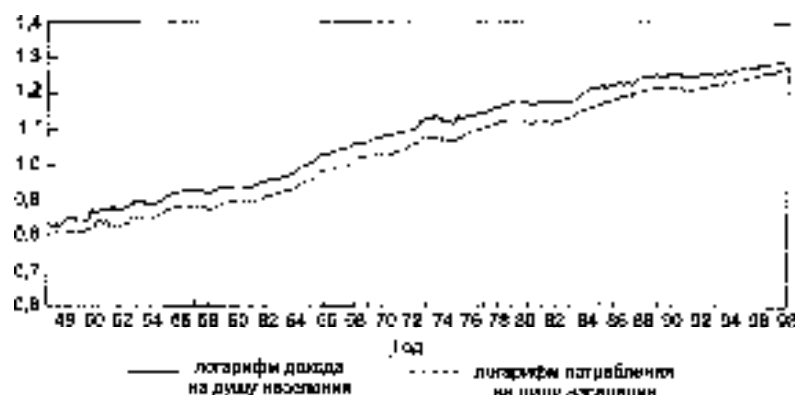


Рис. 10.1(а). Логарифмы дохода и потребления

очевидные тренды и — как будет показано в примере ниже — стохастические тренды. Однако эти два $I(1)$ ряда проявляют тенденцию двигаться совместно, и это наводит на мысль о том, что их разность стационарна. Это пример **коинтеграции** (*cointegration*). Коинтеграционных соотношений в экономике предостаточно. В самом деле, многие из переменных, которые мы изучали в этой книге, являются коинтегрированными: цены на стир и тот же параметр — зарплата двух связанных странах (разность может быть стационарной при более слабой версии параметра — покупательной способности), долгосрочная и краткосрочная процентные ставки (даже хотя они имеют тренды, дифференциал доходности может быть стационарным), форвардный и спотовый обменные курсы и так далее. Коинтеграция может определяться для более чем двух переменных. Например, существование стабильной функции спроса на деньги подразумевает, что линейная комбинация логарифма реальной

долежной массы, номинальной процентной ставки и лотарейная пропорция дохода может быть стационарной, несмотря на то, что каждая из трех переменных является $I(1)$.

Мы начинаем этот параграф обобщая определение одномерных $I(0)$ - и $I(1)$ -процессов из предыдущей главы на векторные процессы. После этого будет формально введено понятие коконтрагации для многомерных $I(1)$ -процессов.

Линейные векторные $I(0)$ - и $I(1)$ -процессы

Для определения векторных $I(0)$ и $I(1)$ -процессов следует [Патил, оп., 1994] и [Джонсон, 1995]. Для введения многомерного обобщения линейного $I(0)$ -процесса с нулевым средним рассмотрим марковский VMA (векторное обобщение скользящего среднего)-процесс с нулевым средним:

$$x_t = \Psi(L) \varepsilon_t, \quad \Psi(L) \equiv I_n - \Psi_1 L + \Psi_2 L^2 - \dots, \quad (10.1.1)$$

где ε_t является $i.i.d.$ с

$$E(\varepsilon_t) = 0, \quad E(\varepsilon_t \varepsilon_s') = \Omega, \quad \Omega \text{ положительно определена.} \quad (10.1.2)$$

(Пример ошибок ε_t может быть векторной мультикатой разностью, удовлетворяющей определенным свойствам, но мы не будем допускать такую степень общности.) Как и в одномерном случае, мы накладываем для ограничения Первое является I -суммируемостью¹.

$$\{\Psi_j\} \text{ является } I\text{-суммируемой.} \quad (10.1.3)$$

Поскольку $\{\Psi_j\}$ абсолютно суммируема, когда выполняется I -суммируемость, многомерная версия утверждения 6.11d¹ означает, что x_t является (строго) стационарным и эргодическим. Второе ограничение на VMA-процесс с нулевым средним имеет вид

$$\Psi(1) \neq 0 \quad (n \times n \text{ матрица нулей}). \quad (10.1.4)$$

То есть по крайней мере один элемент $n \times n$ матрицы $\Psi(1)$ ненулевой. Это естественное многомерное обобщение условия (9.2.3ii) и выражает непостоянность $I(0)$ -процессов на с. 599.

(Линейный) n -мерный векторный $I(0)$ -процесс с нулевым средним (*linear zero-mean n dimensional vector $I(0)$ process*), или **(линейная)**

¹ Последовательность матриц $\{\Psi_j\}$ называется I -суммируемой, если $\{\Psi_j\}$ I -суммируема (то есть $\sum_{j=0}^{\infty} \| \Psi_j \| < \infty$) для всех $L, l = 1, 2, \dots, n$, где $\Psi_{j,l} = (L, l)$ -й элемент $n \times n$ матрицы Ψ_j .

I(0)-система с нулевым средним (*linear I(0) system*) — это VMA-процесс, удовлетворяющий (10.1.1)–(10.1.4). Для заданной I(0)-системы u_t (линейный) **I(0)-процесс** (*linear I(0) process*), или (линейную) **I(0)-систему** (*linear I(0) system*), можно записать как

$$\delta_t = u_t,$$

где δ_t — n -мерный вектор, представляющий единичное стационарное процесс. Из основанных формул (6.5.6) и (6.5.15), долгосрочная ковариационная матрица I(0)-системы имеет вид:

$$\Psi(1)\Omega\Psi(1)'. \quad (10.1.5)$$

Поскольку Ω положительно определена и $\Psi(1)$ удовлетворяет (10.1.4), по крайней мере один из диагональных элементов этой долгосрочной ковариационной матрицы положительный, из чего вытекает, что по крайней мере один из элементов I(0)-системы сам по себе является I(0) (то есть I(0) как одномерный процесс). Мы не требуем, чтобы Ψ была обязательно невырожденной. По существу, допущение этой вырожденности составляет суть только контеграции. Следовательно, долгосрочная ковариационная матрица также может быть вырожденной.

Пример 10.1 (двумерная I(0)-система): Рассмотрим следующий двумерный VMA процесс первого порядка:

$$\begin{cases} u_{1t} = \varepsilon_{1t} - \varepsilon_{2,t-1} - \gamma\varepsilon_{2,t-1}, & \text{или } \varepsilon_t = \varepsilon_t + \Psi_1\varepsilon_{t-1}, \\ u_{2t} = \varepsilon_{2t}, \end{cases} \quad (10.1.6)$$

где

$$\varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}, \Psi_1 = \begin{bmatrix} -1 & \gamma \\ 0 & 0 \end{bmatrix}. \quad (10.1.7)$$

Подобно любому VMA конечного порядка, условие I-стационарности выполняется тривиальным образом. Требование (10.1.4) выполняется, поскольку

$$\Psi(1) = I_2 + \Psi_1 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \quad (10.1.8)$$

Поскольку этот размер соответствует рангу определителя I(0)-системы. Если $\gamma = 0$, то первый элемент u_{1t} фактически является одномерным I(1)-процессом.

n -мерную I(1)-систему (*n-dimensional I(1) system*) y_t , связанную с I(0) системой u_t с нулевым средним, можно определить через соотношение:

$$\Delta y_t = \delta_t + u_t, \quad (10.1.9)$$

Поэтому средние Δy_t равны δ . Поскольку не каждая компонента соответствующего $I(0)$ процесса с нулевым средним y_t должна сама по себе быть $I(0)$, некоторые компоненты $I(1)$ -системы y_t сами по себе могут не быть $I(1)$ ¹. Подставляя (10.1.1) в этот выражение, мы получим:

$$\Delta y_t = \delta + \Psi(L)\varepsilon_t. \quad (10.1.10)$$

Это называется **представлением $I(1)$ -системы в форме векторного скользящего среднего (VMA)** (*vector moving average representation*). В уровнях y_t можно записать как

$$y_t = y_0 + \delta \cdot t + u_1 + u_2 + \dots + u_t, \quad (10.1.11)$$

или, записывая подробно,

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{bmatrix} = \begin{bmatrix} y_{1,0} \\ y_{2,0} \\ \vdots \\ y_{n,0} \end{bmatrix} + \begin{bmatrix} \delta_1 \cdot t \\ \delta_2 \cdot t \\ \vdots \\ \delta_n \cdot t \end{bmatrix} + \begin{bmatrix} u_{1,1} \\ u_{2,1} \\ \vdots \\ u_{n,1} \end{bmatrix} + \begin{bmatrix} u_{1,2} \\ u_{2,2} \\ \vdots \\ u_{n,2} \end{bmatrix} + \dots + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ \vdots \\ u_{n,t} \end{bmatrix}.$$

Что касается начального значения y_0 , мы предполагаем, что это вектор фиксированных констант или что он случайный, но не зависящий от ε_t для всех t .

Пример 10.2 (двумерная $I(1)$ -система): $I(1)$ -система, для которой соответствующий $I(0)$ -процесс с нулевым средним является двумерным процессом из примера 10.1, имеет вид:

$$\begin{cases} \Delta y_{1t} = \delta_1 + \varepsilon_{1t} - \varepsilon_{1,t-1} + \gamma(\varepsilon_{2,t-1} + \dots + \varepsilon_{n,t-1}), \\ \Delta y_{2t} = \delta_2 + \varepsilon_{2t}, \end{cases} \quad \text{или} \quad \Delta y_t = \delta + \varepsilon_t + \Psi \cdot \varepsilon_{t-1}, \quad (10.1.12)$$

где Ψ задается в (10.1.7). В уровнях:

$$\begin{cases} y_{1t} = y_{1,0} + \delta_1 \cdot t + (\varepsilon_{1t} - \varepsilon_{1,0}) + \gamma \cdot (\varepsilon_{2,0} + \varepsilon_{2,1} + \dots + \varepsilon_{n,t-1}), \\ y_{2t} = y_{2,0} + \delta_2 \cdot t + (\varepsilon_{2,1} + \varepsilon_{2,2} + \dots + \varepsilon_{2,t}). \end{cases} \quad (10.1.13)$$

Второй элемент y_t есть $I(1)$ как одномерный процесс. Если $\gamma = 0$, то первый элемент y_{1t} является стационарным относительно тренда, поскольку отклонение от функции тренда $(y_{1,0} - \varepsilon_{1,0}) + \delta_1 \cdot t$ является стационарным (фактически $i.i.d.$). В противном случае y_{1t} является $I(1)$.

¹Во многих учебниках и также в [Engle and Granger, 1987] все компоненты $I(1)$ -системы сами по себе предполагаются $I(1)$, но такое предположение не является необходимым для разработки теории коинтеграции. Этот вопрос придется особая роль, изложенная в [McLeod, 1993, pp. 352–353] и [Johansen, 1995, Chapter 5].

Разложение Боверидажа — Нальсона

Для $I(1)$ -системы, связанной с $I(0)$ -процессом с нулевым средним является m , удовлетворяющий (10.1.1)–(10.1.4), где α — получить разложение Боверидажа — Нальсона (БН). Напомним из параграфа 9.2, что одностороннее БН-разложение основано на парадоксальной MA -сходимости (9.2.5) на $c = 600$. Очевидная минимизирующая $\alpha(L)$ имеет вид:

$$\begin{aligned}\Psi(L) &= \Psi(1) + (1 - L)\alpha(L), \quad \alpha(L) = \sum_{j=0}^{\infty} \alpha_j L^j, \\ \alpha_j &= (\Psi_{j+1} + \Psi_{j+2} + \dots) \quad (10.1.14)\end{aligned}$$

Поэтому многомерный аналог (9.2.6) на $c = 600$ имеет вид:

$$y_t = \Psi(1)\varepsilon_t + \eta_t, \quad \eta_t = \sum_{j=0}^{\infty} \eta_j = \alpha(L)\varepsilon_t. \quad (10.1.15)$$

Поскольку $\Psi(L)$ является I -суммируемым, то $\alpha(L)$ абсолютно суммируем, и η_t является ковариационно-стационарным процессом. Подставляя это в (10.1.11), мы получаем многомерную версию БН разложения:

$$y_t = \delta(L) + \Psi(1)(\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t) + \eta_t = (y_t^* + \eta_t). \quad (10.1.16)$$

Как и в одномерном случае, $I(1)$ -система y_t раскладывается на линейный тренд $\delta(L)$, стохастический тренд $\Psi(1)(\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t)$, стационарную компоненту η_t и начальное условие $y_0 = \eta_0$. По построению η_0 является случайной величиной. Поэтому, если y_0 не полностью коррелировано с η_0 , начальное условие будет случайным.

Пример 10.2 (продолжение): Для двумерной $I(1)$ -системы в примере 10.2 матричная версия $\alpha(L)$ равна $-\Psi_1$, так что стохастическая компонента η_t в БН-разложении имеет вид:

$$\eta_t = \begin{bmatrix} \varepsilon_{1t} - \gamma_1 \varepsilon_{2t} \\ 0 \end{bmatrix}. \quad (10.1.17)$$

(Вам нетрудно будет проверить, что исходя из матричной версии (9.2.5) 1) Двумерный стохастический тренд действительно как

$$\Psi(1)(\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t) = \begin{bmatrix} 0 & \gamma_1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sum_{j=1}^t \varepsilon_{1j} \\ \sum_{j=1}^t \varepsilon_{2j} \end{bmatrix} = \begin{bmatrix} \gamma_1 \sum_{j=1}^t \varepsilon_{2j} \\ \sum_{j=1}^t \varepsilon_{2j} \end{bmatrix}. \quad (10.1.18)$$

Таким образом, двумерный стохастический тренд приводится в движение одним общим стохастическим трендом, $\sum_{j=1}^t \varepsilon_{2j}$. Это пример представления общих трендов в [Stock and Watson, 1993] (см. подраздел 2 ниже).

Определение кointеграции

$I(1)$ -система y_t не является стационарной, поскольку содержит стохастический тренд $\Psi(1) \cdot (\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t)$, и, если $\delta \neq 0$, содержит детерминированный тренд $\delta \cdot t$. Детерминированные и стохастические тренды, однако, могут исчезать, если мы берем подходящую линейную комбинацию элементов y_t . Для осуществления этой идеи умножим обе стороны ВН-разложения (10.1.16) на векторный подходящий вектор α , чтобы получить:

$$\alpha' y_t = \alpha' \delta \cdot t + \alpha' \Psi(1)(\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t) + \alpha' \eta_t = \alpha'(y_0 - \eta_0). \quad (10.1.19)$$

Если α удовлетворяет условию

$$\alpha' \Psi(1) = 0', \quad (10.1.20)$$

то стохастический тренд удаляется и $\alpha' y_t$ принимает вид:

$$\alpha' y_t = \alpha' \delta \cdot t + \alpha' \eta_t = \alpha'(y_t - \eta_0). \quad (10.1.21)$$

Строго говоря, этот процесс не является стационарным относительно тренда, поскольку начальное условие $\alpha'(y_0 - \eta_0)$ может коррелировать с последующими значениями $\alpha' \eta_t$ (см. контрольный вопрос 3 ниже в качестве иллюстрации). Процесс будет стационарным относительно тренда, если, например, начальное значение y_0 будет таково, что $\alpha'(y_0 - \eta_0) = 0$.

Теперь мы готовы определить кointеграцию³.

Определение 10.1 (кointеграция): Пусть y_t — n -мерная $I(1)$ -система, связанная с $I(0)$ -системой нулевым средним η_t , удовлетворяющей (10.1.1)–(10.1.4). Мы говорим, что система y_t является «кointегрированной», с n -мерным «кointегрирующим вектором» α , если $\alpha \neq 0$ и $\alpha' y_t$ можно сделать стационарным относительно тренда путем подходящего выбора его начального значения y_0 .

Определение кointеграции, таким образом, хотя и продиктовано логической последовательностью, необязательно означает, что теория кointеграции требует, чтобы начальное значение y_0 было выбрано так, как указано в определении. Процесс в (10.1.21) не является стационарным, поскольку η_t и η_0 коррелированы. Однако, так как $\eta_t = \alpha(L)\varepsilon_t$ и $\alpha(L)$ абсолютно суммируем, η_t и η_0 будут становиться асимптотически независимыми при росте t . В этом смысле процесс (10.1.21) является

¹Идея кointеграции была предложена в: [Kwiatkowski, 1981]. См. также [Alogoskoufis, 1988] и [Banerjee and Taylor, 1977].

²Наше определение то же самое, что и Definition 3.4 в [Johansen, 1988].

«асимптотически стационарным», и асимптотическая стационарность — это все, что нужно для оценивания и получения статистических выводов для кointегрированных $I(1)$ -систем¹.

С определенной таким образом кointеграцией мы можем определить следующие связанные понятия:

- (Ранг кointеграции.) **Ранг кointеграции** (cointegrating rank) — это число линейно независимых кointегрирующих векторов, а **кointеграционное пространство** (cointegrating space) — это пространство, натянутое на кointегрирующие векторы. Как показывает предыдущее обсуждение, n -мерный вектор α является кointегрирующим вектором тогда и только тогда, когда выполняется (10.1.20). Поскольку если ранг кointеграции равен k , существует k таких линейно независимых векторов, то из этого вытекает, что

$$\text{rank} [\Psi(1)] = n - k. \quad (10.1.22)$$

Иными словами, ранг кointеграции k равен $n - \text{rank} [\Psi(1)]$.

- (Кointеграция внутри подмножества y_t .) По крайней мере один из элементов кointегрирующего вектора ненулевой. Предположим, без потери общности, что первый элемент кointегрирующего вектора отличен от нуля. Тогда мы пишем, что y_{1t} (первый элемент y_t) **кointегрирован с** (cointegrated with) y_{2t} (оставшиеся $n - 1$ элементов y_t), или что y_{1t} является **частью кointеграционного соотношения** (part of a cointegrating relationship). Мы можем также определить кointеграции для подмножества y_t . Например, $n - 1$ переменных в y_{2t} некointегрированы, если не существует $(n - 1)$ -мерного вектора $\delta \neq 0$, такого что (10.1.20) выполняется с $\alpha' = (0, \delta')$. Следовательно, y_{2t} некointегрированы тогда и только тогда, когда последние $n - 1$ строк в $\Psi(1)$ являются линейно независимыми.
- (Стохастическая кointеграция.) Заметим, что детерминированный тренд $\alpha'\delta \cdot t$ не удаляется из $\alpha'y_t$, если только кointегрирующий вектор не удовлетворяет соотношению:

$$\alpha'\delta = 0. \quad (10.1.23)$$

Во многих приложениях кointегрирующий вектор удаляет и стохастический, и детерминированный тренды в (10.1.19). То есть

¹Различие между стационарностью и строгой точечковой стационарностью обсуждается в [Lütkepohl, 1993, pp. 348–349].

вектор π , который удовлетворяет (10.1.20), обязательно удовлетворяет (10.1.23), так что $\alpha'y_t$, который можно теперь записать как

$$\alpha'y_t = \alpha'\pi + \alpha'(y_t - \pi), \quad (10.1.24)$$

является стационарным (не только стационарным относительно тренда) для подходящего выбора π . Из этого следует, что δ является линейной комбинацией столбцов $\Psi(1)$, так что

$$\text{rank} \begin{bmatrix} \delta & \Psi(1) \end{bmatrix} = n - h. \quad (10.1.25)$$

(h < n+1)

Если не указать обратное, мы будем предполагать, что (10.1.25), так же как и (10.1.22), выполняются, когда ранг коинтеграции равен h . Если мы хотим описать случай, когда коинтегрирующий вектор удаляет стохастический тренд, но необязательно удалит детерминированный тренд, то будем использовать термин «стохастическая коинтеграция» (*stochastic cointegration*). Конечно, если π не содержит детерминированный тренд (то есть если $\delta = 0$), то отсюда между коинтеграцией и стохастической коинтеграцией нет

Пример 10.2 (продолжение): В двумерной $I(1)$ -системе (10.1.12) в примере 10.2 матрица $\Psi(1)$ задается в (10.1.8), поэтому

$$\begin{bmatrix} \delta & \Psi(1) \end{bmatrix} = \begin{bmatrix} \delta_1 & 0 & \gamma \\ \delta_2 & 0 & 1 \end{bmatrix}.$$

Ранг $\Psi(1)$ равен 1, так что ранг коинтеграции равен 1 ($2 - 1$). Все коинтегрирующие векторы можно записать как $(\alpha_1, -\alpha_2)'$, $\alpha_1 \neq 0$. Предположим, что коинтегрирующий вектор также удалит детерминированный тренд, можно записать как $\alpha_1 = -\alpha_2\delta_2/\delta_1$, так $\delta_1 = -\alpha_2\delta_2/\alpha_1$, из чего следует, что ранг матрицы $\begin{bmatrix} \delta & \Psi(1) \end{bmatrix}$ указанной выше, равен единице.

Из определения коинтеграции вытекают следующие важные следствия в отношении $I(1)$ -систем. Например:

- ($h < n$). Для n -мерной $I(1)$ -системы ранг коинтеграции не может быть равным n . Если бы он был равным n , то тогда $\delta = \text{rank}[\Psi(1)] = 0$ и $\Psi(1)$ была бы матрицей из нулей, что исключено требованиями (10.1.4).
- (Следствие коинтеграцией определяется долгосрочной дисперсией Δy_t .) Долгосрочная ковариационная матрица Δy задается как $\Psi(1)' \Omega \Psi(1)$ (см. (10.1.5)), которая положительно определена тогда и только тогда, когда $\Psi(1)$ невырождена. Следовательно, y_t

некointегрирован тогда и только тогда, когда долговременная ковариационная матрица Δy_t положительно определена¹. Поскольку долговременная дисперсия каждого элемента Δy_t положительна, если долговременная ковариационная матрица Δy_t положительно определена, то из этого следует, что каждая компонента y_t сама по себе является $I(1)$ (то есть является одномерным $I(1)$ -процессом), если y_t некointегрирован. То же самое верно для подмножества y_t . Например, рассмотрим y_{2t} , последние $n-1$ компонент в y_t . Долговременная ковариационная матрица Δy_{2t} задается как $\Psi_2(1)\Omega\Psi_2(1)'$, где $\Psi_2(1)$ — последние $n-1$ строк в $\Psi(1)$. Она положительно определена тогда и только тогда, когда строки $\Psi_2(1)$ линейно независимы или y_{2t} некointегрирован. В частности, каждая компонента y_{2t} сама по себе является $I(1)$, если y_{2t} некointегрирован.

- Предположим, что y_{1t} кointегрирован с y_{2t} . Тогда y_{2t} некointегрирован, если $h=1$, и кointегрирован, если $h>1$. Причина заключается в следующем. Из кointеграции y_{1t} с y_{2t} следует то, что существует кointегрирующий вектор, (назовем его α_1), первый элемент которого ненулевой. Если $h=1$, то не должно существовать $(n-1)$ -мерного вектора $b \neq 0$, такого что $\alpha_2' = (0, b')$ является бы кointегрирующим вектором, поскольку α_1 и α_2 линейно независимы. Такой вектор α_2 можно найти, если $h>1$.
- (VAR в первых разностях?) Если y_t является стационарным в разностях (без сноса), возникает искушение моделировать его как стационарную VAR конечного порядка $\Phi(L)\Delta y_t = \varepsilon_t$, где $\Phi(L)$ — матричный полином степени p , такой что все корни $|\Phi(z)| = 0$ лежат вне единичного круга. Но тогда y_t не может быть кointегрированным. Причина заключается в следующем. Если полином $\Phi(L)$ удовлетворяет условию стационарности, то последовательность матриц коэффициентов $\{\Psi_t\}$ обратного полинома $\Psi(L) = \Phi(L)^{-1}$ ограничена по метрически убывающей последовательностью (см. параграф 6.3). Поэтому $\Psi(L)$ является 1-суммируемым, и $\Delta y_t = \Psi(L)\varepsilon_t$ является VMA-процессом, удовлетворяющим (10.1.2) и (10.1.3). Кроме того,

$$|\Psi(1)| = |\Phi(1)|^{-1} = |\Phi(1)|^{-1} \neq 0.$$

Поэтому $\Psi(\cdot)$ невырождена и долговременная ковариационная матрица положительно определена.

¹ Эта эквивалентность специфична для линейных процессов. В общем случае полнотеловичная определенность долговременной ковариационной матрицы Δy_t достаточна, но не всегда необходима, для того чтобы y_t не был кointегрирован. См. [Primer, 1986, p. 33].

Контрольные вопросы

1. Пусть (y_1, y_2) будет как в примере 10.2 с $\gamma = 0$. Покажите, что $(y_1 + y_2)$ является $I(1)$. **Указание:** Вы должны показать, что долговременная дисперсия $\{\Delta y_1 + \Delta y_2\}$ положительна.
2. (Представление общих трендов в [Stock and Watson, 1998]) Пусть $\pi_t \equiv \pi_1 + \pi_2 + \dots + \pi_T$ так что BN-разложение имеет вид:

$$y_t = \delta \cdot t + \Psi(1)w_t + \pi_t + (y_0 - \pi_0).$$

Результат из линейной алгебры утверждает, что

если C является $n \times n$ матрицей ранга $n - h$, то существует $n \times n$ невырожденная матрица G и $n \times (n - h)$ матрица F такого ранга, такая что

$$\begin{pmatrix} C & G \end{pmatrix}_{(n \times n)(n \times n)} = \begin{pmatrix} F & \vdots & 0 \end{pmatrix}_{(n \times (n-h)) \times n}.$$

Покажите, что перпендикулярную компоненту $\Psi(1)w_t$ можно записать как $F\pi_t$, где $F = n \times (n - h)$ матрица полного ранга по строкам, и $\pi_t = (n - h)$ -мерное случайное блуждание с положительно определенной $\text{Var}(\Delta \pi_t)$. **Указание:** $\Psi(1)w_t = \Psi(1)GG^{-1}w_t$. Пусть π_t является первыми $n - h$ компонентами $G^{-1}w_t$. Это представление делает понятным, что $I(1)$ -система с рангом конкатрации h имеет $n - h$ общих стохастических трендов.

3. (y_t^* не является стационарным.) Чтобы доказать, что прожект в (10.12) не вполне стационарный, рассмотрим простой случай, где $\pi_t = \varepsilon_t - \varepsilon_{t-1}$, $\delta = 0$ и $y_0 = 0$. Покажите, что

$$\text{Cov}(\pi_t, \pi_0) = \begin{cases} 2\Omega & \text{для } t = 0, \\ -\Omega & \text{для } t = 1, \\ 0 & \text{для } t > 1, \end{cases} \quad \text{Var}(\alpha'y_t) = \begin{cases} 0 & \text{для } t = 0, \\ 6\alpha'\Omega\alpha & \text{для } t = 1, \\ 4\alpha'\Omega\alpha & \text{для } t > 1. \end{cases}$$

где $\Omega = \text{Var}(\varepsilon_t)$. Проверьте, что $\{\alpha'y_t\}$ является стационарным для $t = 2, 3, \dots$.

4. (Что, если некоторые элементы стационарны?) Пусть y_t является $I(1)$ -системой, и предположим, что первый элемент y_1 является стационарным. Покажите, что в этом случае ранг конкатрации как минимум равен 1.
5. (Матрица конкатрирующих векторов.) Предположим, что ранг конкатрации n -мерной $I(1)$ -системы π_t равен h , и пусть A является $n \times h$ матрицей, содержащей h линейно независимых конкатрирующих векторов. Пусть B является любой $h \times h$ невырожденной матрицей. Покажите, что столбцы AB есть h линейно независимых конкатрирующих векторов. **Указание:** Вы можете не невырожденную матрицу не изменить ранга.

10.2. Альтернативные представления коинтегрированных систем

В дополнение к представлению в виде триады (см. контрольный вопрос 2 в предыдущей главе) существуют три других полезных представления коинтегрированных векторных I(1) процессов: треугольное представление Филлипса [Phillips, 1991], VAR-представление и VECM (векторная модель коррекции ошибок, *vector error correction model*) в [Davidson et al., 1993]. Эти представления охватываются в данном параграфе.

Треугольная система Филлипса

Это представление удобно для статистических коинтегрированных тестов. Представление объяснено для случая единичной коинтеграции, но мы изначально предположим, что ранг коинтеграции k равен 1. Пусть α является коинтегрирующим вектором, и предположим, без ущерба общности, что первый элемент α ненулевой (если он равен нулю, то можно поменять порядок переменных в системе, так что первый элемент после перепорядочивания будет ненулевым). Разобьем y_t следующим образом:

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ (n-1) \times 1 \end{bmatrix}. \quad (10.2.1)$$

Таким образом, y_{1t} коинтегрирующая y_{2t} . Поскольку коинтегрирующий вектор, умноженный на скаляр, также является коинтегрирующим вектором, мы можем нормализовать коинтегрирующий вектор α так, чтобы первый элемент α равнялся единице:

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} 1 \\ -\gamma \\ (n-1) \times 1 \end{bmatrix}. \quad (10.2.2)$$

Мы начнем с предыдущего параграфа, то есть коинтегрирующий вектор α является не только стохастическим трендом (то есть $\alpha' \Psi(1) < 0$), но также и детерминированным трендом (то есть $\alpha' \delta = 0$), то тогда $\alpha' y_t$ можно записать как (10.1.24). Поскольку α в (10.1.24) является коинтегрирующей вектору (10.2.2), мы получим:

$$y_{1t} = \gamma' y_{2t} + \varepsilon_t^* \quad (10.2.3)$$

где

$$\varepsilon_t^* = (1 - \gamma' \eta_t), \quad 0 \leq (1 - \gamma')' y_{2t} = 0. \quad (10.2.4)$$

Поскольку y_t совместно стационарный, x_t^* является стационарным. Это уравнение, где значения x_t^* рассматриваются как ошибки, а μ как константа, называют **кointегрирующей регрессией** (*cointegrating regression*). Ее регрессионные коэффициенты γ , связывающие перманентную компоненту в y_{1t} с перманентными компонентами, входящими в y_{2t} , можно интерпретировать как описывающие долгосрочное соотношение между y_{1t} и y_{2t} . Кointегрирующий вектор будет иметь компоненту тренда $\alpha'\delta + t$ в качестве дополнительного регрессора, если кointегрирующий вектор не удаляет детерминированный тренд в $\alpha'y_t$. **Треугольное представление** (*triangular representation*) — это система из n уравнений, состоящая из кointегрирующей регрессии в последних $n-1$ строках в (10.1.9).

$$\Delta y_{2t} + \delta_2 + \alpha_2 = \begin{bmatrix} \delta_2 \\ \alpha_2 \end{bmatrix} + \Psi_2(L) \varepsilon_t, \quad (10.2.5)$$

где δ_2 и α_2 — последние $n-1$ элементов n -мерного вектора δ и α соответственно, а $\Psi_2(L)$ — последние $n-1$ строк в $\Psi(L)$ в VMA-представлении (10.1.10). Как было отмечено в предыдущем параграфе, следствием предположения о том, что $h = 1$, является то, что y_{2t} неограничен (он был бы кointегрированным, если бы $h > 1$). В частности, каждый элемент y_{2t} является $I(1)$.

Рассмотрим общий случай, когда ранг кointеграции h необязательно равен 1. Изменяя порядок элементов в y_t , если это необходимо, можно (см., например [Hamilton, 1994, pp. 576–577]) выбрать h линейно независимых кointегрирующих векторов, $\alpha_1, \alpha_2, \dots, \alpha_h$, таких что

$$A \equiv \begin{bmatrix} \alpha_1 & \dots & \alpha_h \end{bmatrix} = \begin{bmatrix} I_h \\ \alpha^{(h)} \\ 0 \\ \alpha^{(n-h)} \end{bmatrix}. \quad (10.2.6)$$

Разобьем y_t соответственно, как

$$y_t \equiv \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix}, \quad (10.2.7)$$

Умножив обе стороны ВМ разложения (10.1.16) слева на матрицу A' и замечая, что $A'\Psi(1) = 0$ (поскольку столбцы в A являются кointегрирующими векторами), $A'\delta = 0$ (если эти кointегрирующие векторы также удаляют детерминированный тренд) и $A'y_t = y_{2t} + 1'y_{3t}$, мы получим следующие h кointегрирующих регрессий.

$$y_{1t} = \begin{bmatrix} I' \\ 0 \end{bmatrix} y_{2t} + \mu + x_t^*, \quad (10.2.8)$$

где $\mu = A'(y_2 - \pi_2)$ и $z_1^* = A'y$. Поскольку π_2 известно с высокой точностью, это выполняется и для z_1^* . Треугольное представление — это эти k контегрирующих регрессий, дополненные оставшейся частью VMA представления:

$$\Delta y_{2t} = \delta_2 + \alpha_{2t} = \delta_2 + \Psi_2(L)' \varepsilon_t, \quad (10.2.9)$$

$$(\alpha - \delta) \times 1 = (\alpha - \delta) \times 1 + (x - A)x_t, \quad (\alpha - \delta) \times 1 = (x - A)x_t + \alpha \times 1.$$

Здесь $\Psi_2(L)$ — это полином в L степеней $\Psi(L)$ в VMA-представлении (10.1.10). Легко показать (внутренний вопрос 1), что y_{2t} неинтегрирован.

Чтобы проиллюстрировать треугольное представление и то, как его можно получить из VMA-представления, рассмотрим:

Пример 10.3 (из VMA в треугольное): В двумерной системе (10.1.12) примера 10.2 ранг контегрировки равен 1. Контегрирующий вектор, второй элемент которого равен единице, имеет вид: $(1, -\gamma)'$. Так что z_1^* и μ в (10.2.3) имеют вид:

$$z_1^* = (1, -\gamma)\eta_t = \eta_t - \gamma\eta_{2t},$$

$$\mu = (\dots, \alpha_t)(y_t - \eta_t) = (x_t A - \gamma y_{2t} A) = (x_{1t} A - \gamma x_{2t} A), \quad (10.2.10)$$

и треугольное представление имеет вид:

$$\begin{cases} y_{1t} = \alpha + \gamma y_{2t} + (x_{1t} - \gamma x_{2t})', \\ \Delta y_{2t} = \delta_2 + \alpha_{2t}. \end{cases} \quad (10.2.11)$$

VAR и коинтеграция

В стационарном случае мы нашли удобным и удобным обозначить векторный процесс как VAR конечного порядка. Хотя, как видно выше, коинтегрированную $I(1)$ -систему нельзя представить в виде VAR в разностях конечного порядка, некоторые коинтегрированные системы могут допускать представление VAR конечного порядка в *уровнях*. Поэтому предположим, что коинтегрированную $I(1)$ -систему y_t можно записать как

$$y_t = \alpha + \delta \cdot t + \xi_t, \quad \Phi(L)' \xi_t = \varepsilon_t,$$

$$\Phi(L)' = I_n - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p. \quad (10.2.12)$$

Для дальнейшего использования мы уберем отсюда ξ_t , получая

$$y_t = \alpha + \delta \cdot t + \Phi y_t - \Phi_1 y_{t-1} - \dots - \Phi_p y_{t-p} + \varepsilon_t. \quad (10.2.13)$$

где

$$\alpha_{(n \times 1)}^* \equiv (\Phi_1 + 2\Phi_2 + \dots + p\Phi_p)\delta - \Phi(1)\alpha, \quad \delta_{(n \times 1)}^* \equiv \Phi(1)\delta. \quad (10.2.14)$$

Как мы узнаем, что эта VAR в уровнях конечного порядка является кointегрированной I(1)-системой? Очевидный способ узнать это — получить VMA-представление. $\Delta \xi_t = \Psi(L)\varepsilon_t$ из VAR-представления, $\Phi(L)\xi_t = \varepsilon_t$, и посмотреть, удовлетворяет ли $\Psi(L)$ определению кointегрированной системы. Получить это несколько сложнее, поскольку VMA-представление дается в первых разностях, но это можно сделать довольно легко следующим образом. Беря первую разность от обеих частей $\Phi(L)\xi_t = \varepsilon_t$ и замечая, что $\Delta = 1 - L$ и $(1 - L)\Phi(L) = \Phi(L)(1 - L)$, мы получаем:

$$\Phi(L)\Delta \xi_t = (1 - L)\varepsilon_t. \quad (10.2.15)$$

Подставляя $\Delta \xi_t = \Psi(L)\varepsilon_t$ в это уравнение, получаем:

$$\Phi(L)\Psi(L)\varepsilon_t = (1 - L)\varepsilon_t. \quad (10.2.16)$$

Это уравнение должно выполняться для любой реализации ε_t , так что

$$\Phi(L)\Psi(L) = (1 - L)I_n. \quad (10.2.17)$$

Последнее уравнение можно решить относительно $\Psi(L)$, умножая обе части слева на $\Phi(L)^{-1}$, обращение $\Phi(L)^{-1}$. Это дает $\Psi(L) = \Phi(L)^{-1}(1 - L)$. Вопрос заключается в том, при каких условиях на $\Phi(L)$ эта последовательность $\Psi(L)$ 1-суммируема и так ли $|\Phi(1)|^{-1} = n - h$?

Мы можем легко вывести необходимое условие. Полагая $L = 1$ в равенстве (10.2.17), получаем

$$\Phi(1)^{-1}\Phi(1) = \frac{0}{(n-h)(n-h)}. \quad (10.2.18)$$

Поскольку ранг $\Psi(1)$ равен $n - h$, когда ранг кointеграции ξ_t равен h , ранг $\Phi(1)$ не больше, чем h . Как показано ниже, этот ранг фактически равен h . Для формулирования необходимого и достаточного условия пусть $U(L)$ и $V(L) = n \times n$ матричные латентные полиномы, все корни которых лежат вне единичного круга, и пусть $M(L)$ — матричный полином, имеющий вид:

$$M(L) = \begin{bmatrix} (1-L)I_{n-h} & 0 \\ 0 & (1-L)^{n-h}I_h \end{bmatrix} \quad (10.2.19)$$

Обращение существует, поскольку $\Phi_n = I_n$ невырожден; см. параграф 6.3. Поскольку мы не предвзято включаем условие единственности $\Phi(L)$, обратный объект может не быть абсолютно суммируемым.

То есть первые $n - h$ диагональных элементов диагональной матрицы $M(L)$ равны $(1 - L)$, а остальные h диагональных элементов равны единице. Необходимое и достаточное условие для того, чтобы VAR-процесс конечного порядка $\{\xi_t\}$, следующий $\Phi(L)\xi_t = \varepsilon_t$, был контегрированной $I(1)$ -системой ранга h , состоит в том, что $\Phi(L)$ можно факторизовать как $\Phi(L) = U(L)M(L)V(L)^T$. Следовательно, все корни $|\Phi(z)| = 0$ лежат на или за пределами единичной окружности, и все те, которые находятся на единичной окружности, являются единичными корнями ($z = 1$). Тот, что $\Phi(L)$ имеет $n - h$ единичных корней, а остальные корни лежат за пределами единичной окружности, недостаточно: см. Пример в контрольном вопросе 4, где $\Phi(z)$ имеет два единичных корня и один корень вне единичной окружности, однако эта система не $I(1)$. $n - h$ единичных корней должны быть разложены в системе каким-либо образом, указываемым разложением $\Phi(z) = U(z)M(z)V(z)$.

Пологая $z = 1$ в этой факторизации, мы получаем $\Phi(1) = U(1)M(1) \times V(1)$. Поскольку все корни $U(z)$ и $V(z)$ лежат за пределами единичной окружности, $U(1)$ и $V(1)$ невырождены, и ранг $\Phi(1)$ равен рангу $M(1)$, который равен h . То есть

$$\text{rank}[\Phi(1)] = h. \quad (10.2.20)$$

Тогда из основ линейной алгебры следует, что существуют две $n \times h$ матрицы полного столбцового ранга, B и A , такие что

$$\Phi(1) = \begin{pmatrix} B & A' \\ 0 & 0 \end{pmatrix} \quad (10.2.21)$$

Это иногда называют **условием пониженного ранга** (*reduced rank condition*). Выбор B и A не единственный, если $K = h \times h$ невырожденная матрица, то $B(K^{-1})'$ и AK вместо B и A также удовлетворяют (10.2.21). Подставляя (10.2.21) в (10.2.18), мы получаем $BA'\Psi(1) = 0$. Поскольку B имеет полный столбцовый ранг, из этого уравнения следует, что $A'\Psi(1) = 0$. Поэтому столбцы $n \times h$ матрицы A являются контегрирующими векторами.

Векторная модель коррекции ошибок (VECM)

Для одномерного случая мы получили расширенную авторегрессию (9.4.3) на с. 621 из AR-уравнения. Ту же самую идею можно применить здесь к VAR. Это вполне простой алгебры — показать, что VAR-представление $\Phi(L)\xi_t = \varepsilon_t$ в (10.2.12) можно эквивалентно записать как

$$\xi_t = \rho\xi_{t-1} + \zeta_1\Delta\xi_{t-1} + \zeta_2\Delta\xi_{t-2} + \dots + \zeta_{p-1}\Delta\xi_{t-p+1} + \varepsilon_t. \quad (10.2.22)$$

¹Этот результат является следствием леммы (Лемма 80), сформулированной в [Engle and Yoo, 1991]. Поступило изложение есть в [Watson, 1994, pp. 2870–2875].

где

$$\rho = \Phi_1 + \Phi_2 + \dots + \Phi_p, \quad (10.2.23)$$

$$\zeta_\alpha = -(\Phi_{\alpha+1} + \Phi_{\alpha+2} + \dots + \Phi_p) \quad \text{для } \alpha = 1, 2, \dots, p-1. \quad (10.2.24)$$

Вычитая ξ_{t-1} из обеих частей (10.2.22) и замечая, что $\rho = I_n + -(\Phi_1 + \Phi_2 + \dots + \Phi_p) = -\Phi(1)$, мы можем переписать (10.2.22) как

$$\begin{aligned} \Delta \xi_t = & -\Phi(1)\xi_{t-1} + \zeta_1 \Delta \xi_{t-1} + \zeta_2 \Delta \xi_{t-2} + \dots + \zeta_{p-1} \Delta \xi_{t-p+1} + \varepsilon_t = \\ & -BA'\xi_{t-1} + \zeta_1 \Delta \xi_{t-1} + \zeta_2 \Delta \xi_{t-2} + \dots + \zeta_{p-1} \Delta \xi_{t-p+1} + \varepsilon_t. \end{aligned} \quad (10.2.25)$$

Используя соотношение $y_t = \alpha + \delta \cdot t + \xi_t$, его можно переписать в уравнении для y_t :

$$\begin{aligned} \Delta y_t = & \alpha^* + \delta^* \cdot t - \Phi(1)y_{t-1} - \\ & + \zeta_1 \Delta y_{t-1} + \dots + \zeta_{p-1} \Delta y_{t-p+1} + \varepsilon_t, \end{aligned} \quad (10.2.26)$$

$$\begin{aligned} = & \alpha^* + \delta^* \cdot t - \underbrace{B}_{(k \times n)(n \times 1)} z_{t-1} + \\ & + \zeta_1 \Delta y_{t-1} + \dots + \zeta_{p-1} \Delta y_{t-p+1} + \varepsilon_t, \end{aligned} \quad (10.2.27)$$

где α^* и δ^* выражаются как в (10.2.14), а z_t задается как

$$z_t = \underbrace{A'}_{(k \times n)} \underbrace{y_t}_{(n \times 1)}. \quad (10.2.28)$$

Поскольку, как было отмечено выше, $n \times k$ матрица A содержит k континтегрирующих векторов, z_t является стационарным относительно тренда (с подходящим выбором начального значения y_0)¹. Это представление называется **векторной моделью коррекции ошибок (VECM) (vector error-correction model)**, или **представлением в форме коррекции ошибок (error-correction representation)**. Если бы в VECM не было слагаемого Bz_{t-1} , то процесс, выраженный как VAR в первых разностях,

¹ На тот случай, если вас одолеет вопрос, почему y_t падает, Действительно, мы можем решить (10.2.27) относительно z_t как

$$z_t = -(B'B)^{-1}B'[\alpha^* + \delta^* \cdot t - \Delta y_t + \zeta_1 \Delta y_{t-1} + \dots + \zeta_{p-1} \Delta y_{t-p+1} + \varepsilon_t]$$

Потому вы можете подумать, что z_t является стационарным относительно тренда вне зависимости от выбора y_0 . Строго говоря, это не так, поскольку, в отличие от VMA-представления, Δy_t в VAR/VECM-представлении не определено для $t = -1, -2, \dots$. Однако, Δy_t ($t = 1, 2, \dots$) стационарны и можно так же проанализировать поведение y_t . Этот вопрос обсуждается в [Борнштейн, 1993, гл. 10, пункт 4.2].

не был бы коинтегрированной VECM, эмбрион в себя k коинтегрирующих соотношений путем объединения k линейных комбинаций уровней переменных. Если в коинтегрирующих соотношениях нет временного тренда, то есть если $A'\delta = 0$, то $\delta' = \Phi(1)\delta = BA'\delta = 0$. Поэтому VAR- и VECM-представления не исключают временной тренд, за исключением всемогущего существования временных трендов в элементах y_t .

То, что одна и та же I(1)-процесс имеет VMA-, VAR- и VECM-представление, известно как *Теорема Грейнджера о представлении* (Granger Representation Theorem).

Пример 10.4 (из VMA в VAR/VECM): В предыдущем примере мы вывели треугольные представления из VMA-представления (10.1.12). В этом примере мы выводим VAR и VECM-представления из того же самого VMA-представления. Для $\Phi(L)$ в (10.1.12) легко проверить, что (10.2.17) выполняется с

$$\Phi(L) = I_2 - \Phi_1 L, \quad \Phi_1 = \begin{bmatrix} 0 & \gamma \\ 0 & 1 \end{bmatrix}. \quad (10.2.29)$$

Поэтому VMA можно представить как VAR конечного порядка. Для этой VAR мы имеем:

$$\Phi(1) = I_2 - \Phi_1 = \begin{bmatrix} 1 & -\gamma \\ 0 & 0 \end{bmatrix}, \quad (10.2.30)$$

что можно записать как (10.2.21) с

$$B = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{и} \quad A = \begin{bmatrix} 1 \\ -\gamma \end{bmatrix} \quad (\text{например}). \quad (10.2.31)$$

Из (10.2.27) и (10.2.28) вытекает, что VECM для такого выбора B и A имеет вид:

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} \gamma\delta_2 + \alpha_1 & -\alpha_2 \\ \delta_2 & 0 \end{bmatrix} + \begin{bmatrix} \delta_1 & \gamma\delta_2 \\ 0 & 0 \end{bmatrix} y_{t-1} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} y_{t-1} + \varepsilon_t, \\ y_t = A'y_t = y_{1t} - \gamma y_{2t}. \quad (10.2.32)$$

Детерминированный тренд исчезает, если $\delta_1 = \gamma\delta_2$, то есть если коинтегрирующий вектор также y_t имеет детерминированный тренд.

ML-процедура Йохансена

В начале этого параграфа, введя VECM-представление, мы рассмотрели основные приложения предложенной в [Johansen, 1988] методики идентификации коинтегрированных систем методом максимального правдоподобия (ML). Впрочем, и VAR-представления (10.2.13). Для простоты будем считать, что в коинтегрирующих соотношениях тренд отсутствует, так

что $\delta' = \mathbf{0}'$. В параграфе 8.7 мы рассматривали условное ML-оценивание VAR (условное на начальные значения $\mathbf{y}$). Если вектор ошибок ε_t является совместно нормальным $N(\mathbf{0}, \Omega)$ и если у нас есть выборка из $T + p$ наблюдений $(\mathbf{y}_{-p+1}, \mathbf{y}_{-p+2}, \dots, \mathbf{y}_T)$, тогда (среднее) логарифмическое правдоподобие для $(\mathbf{y}_1, \dots, \mathbf{y}_T)$, условное на начальные значения $(\mathbf{y}_{-p+1}, \mathbf{y}_{-p+2}, \dots, \mathbf{y}_0)$, задается как

$$Q_T(\theta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(|\Omega^{-1}|) - \frac{1}{2T} \sum_{t=1}^T \{(\mathbf{y}_t - \Pi' \mathbf{x}_t)' \Omega^{-1} (\mathbf{y}_t - \Pi' \mathbf{x}_t)\}, \quad (10.2.33)$$

где n является размерностью системы (не объемом выборки), $\theta = (\Pi, \Omega)$ и

$$\Pi' = \begin{bmatrix} \alpha' & \Phi_1 & \Phi_2 & \cdots & \Phi_p \\ (n \times 1) & (n \times n) & (n \times n) & & (n \times n) \end{bmatrix}, \quad \mathbf{x}_t = \begin{bmatrix} 1 \\ \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \vdots \\ \mathbf{y}_{t-p} \end{bmatrix} \quad (10.2.34)$$

Пусть

$$\xi_0 \equiv -\Psi(1) = -[I_n - (\Phi_1 + \cdots + \Phi_p)], \quad (10.2.35)$$

Тогда условие пониженного ранга состоит в том, что $\xi_0 = -BA'$, и VECM (10.2.26) можно записать как

$$\Delta \mathbf{y}_t = \alpha' + \xi_0 \mathbf{y}_{t-1} + \xi_1 \Delta \mathbf{y}_{t-1} + \cdots + \xi_{p-1} \Delta \mathbf{y}_{t-p+1} + \varepsilon_t. \quad (10.2.36)$$

Уравнения (10.2.36) и (10.2.24) дают взаимно однозначное отображение между $(\Phi_1, \dots, \Phi_p)$ и $(\xi_0, \dots, \xi_{p-1})$. Таким образом, то же самое среднее логарифмическое правдоподобие $Q_T(\theta)$ можно записать в терминах параметров VECM как

$$-\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(|\Omega^{-1}|) - \frac{1}{2T} \sum_{t=1}^T \{(\Delta \mathbf{y}_t - \tilde{\Pi}' \tilde{\mathbf{x}}_t)' \Omega^{-1} (\Delta \mathbf{y}_t - \tilde{\Pi}' \tilde{\mathbf{x}}_t)\}, \quad (10.2.37)$$

где

$$\tilde{\Pi}' \equiv \begin{bmatrix} \alpha' & \xi_0 & \xi_1 & \cdots & \xi_{p-1} \\ (n \times 1) & (n \times n) & (n \times n) & & (n \times n) \end{bmatrix}, \quad \tilde{\mathbf{x}}_t \equiv \begin{bmatrix} 1 \\ \mathbf{y}_{t-1} \\ \Delta \mathbf{y}_{t-1} \\ \vdots \\ \Delta \mathbf{y}_{t-p+1} \end{bmatrix} \quad (10.2.38)$$

¹ В качестве подробного рассмотрения временного тренда в VECM см.: Johansen, 1995, Sections 5.7 and 6.2.

ML-оценки параметров VECM предоставляет набор $(\alpha^*, \zeta_0, \dots, \zeta_{n-1}, \hat{\Omega})$, который максимизирует эту целевую функцию при ограничениях, связанных с пониженным рангом и состоящих в том, что $\zeta_n = -BA'$ для некоторых $n \times h$ матриц полного ранга A и B . Это ограничение вынуждает h контегрирующих соотношений в $\{I\}$ -системе. Очевидно, что максимизированное логарифмическое правдоподобие тем больше, чем больше предполагаемый ранг контегриации. Таким образом, мы можем использовать статистику отношения правдоподобий для тестирования нулевой гипотезы $h = h_0$ против альтернативной гипотезы о более высоком ранге контегриации. В отличие от случая стационарной VAR (предельное распределение этой статистики отношения правдоподобий нестандартно). При определенном таким образом ранге контегриации h можно получить ML-оценку для h контегрирующих векторов как оценку для A . Относительно подробного изложения этой процедуры см.: [Hamilton, 1994, Chapter 20], а также [Johansen, 1995, Chapter 6].

Контрольные вопросы

1. (Вектор ошибок в треугольном представлении.) Распишите вектор ошибок $(\varepsilon_1', \varepsilon_2')$ в треугольном представлении (10.2.6) и (10.2.9). Его можно записать как

$$\begin{bmatrix} \varepsilon_1' \\ \varepsilon_2' \\ \varepsilon_3' \\ \vdots \\ \varepsilon_n' \end{bmatrix} = \Psi^* (L) \varepsilon, \quad \Psi^* (L) = \begin{bmatrix} \Psi_1'(L) \\ \Psi_2'(L) \\ \vdots \\ \Psi_n'(L) \end{bmatrix}, \quad (10.2.39)$$

где

$$\Psi_i'(L) = (I_n - \Gamma') \alpha_i(L), \quad (10.2.40)$$

где $\alpha_i(L)$ из ВУ-разложения, а $\Psi_2'(L)$ — это последние $n - h$ строк $\Psi_1'(L)$ в VMA-представлении.

- (a) Покажите, что $\Psi_1'(L)$ — это, полный строковый ранг (то есть $n - h$ строк $\Psi_2'(L)$ линейно независимы). Укажите: Предположим, что верно утверждение, что существует $(n - h)$ -мерный вектор $b \neq 0$, такой что $b'\Psi_2'(L) = 0'$. Покажите, что тогда $(0', b')$ был бы $(n - h)$ -мерный контегрирующий вектор, а ранг контегриации был бы не менее $n - 1$.
 - (b) Проверьте, что $\Psi^*(L)$ блочно суммируем.
 - (c) Запишите $\Psi^*(L) = \Psi_2' + \Psi_1'(L) = \Psi_2' + L^2 + \dots$. Для двучленного члена в примере 10.3 проверьте, что Ψ_2' не является диагональной матрицей. (Словом сказать, даже если элементы e_i некоррелированы, ε_1' и ε_2' могут быть коррелированными.)
2. (От треугольного представления к VMA.) В примере 10.3 начните с треугольного представления (10.2.11) и восстановите VMA-представление. Укажите: Выпишите видную разность идентифицирующей регрессии

3. (Альтернативное разложение $\Phi(1)$.) Для двумерной системы в примере 10.4 проверьте, что

$$B = (1, 0)', \quad A = (1/\alpha, -1)'$$

является альтернативным разложением $\Phi(1)$. Запишите VECM, соответствующую этому выбору B и A .

4. (Трёхмерная VAR, являющаяся I(2).) Рассмотрим трёхмерную VAR, заданную как

$$\begin{cases} y_{1t} = 2y_{1,t-1} - y_{1,t-2} + \varepsilon_{1t} & (\text{мы считаем } \Delta^2 y_{1t} = \varepsilon_{1t}), \\ y_{2t} = \rho y_{2,t-1} + (1-\rho)\varepsilon_{2t} & \text{с } |\rho| < 1, \text{ и} \\ y_{3t} = \varepsilon_{3t}. \end{cases}$$

Проверьте, что это I(2)-система, записав $\Delta^2 y_t$ как векторный процесс стационарного ранга. Запишите $\Phi(L)$ для этой системы и покажите, что $\Phi(1)$ имеет три корня, два из которых единичные, а один лежит за пределами единичной окружности.

10.3. Тестирование нулевой гипотезы об отсутствии коинтеграции

После того как мы узнали понятие коинтеграции, нам нужно разобраться с двумя проблемами. Первая — как определить ранг коинтеграции, а вторая — как оценить коинтегрирующие векторы и делать статистические выводы на их основе. Первая проблема будет обсуждаться в текущем параграфе, а вторая будет темой следующего. Существует несколько процедур определения ранга коинтеграции. Среди них тест отношения правдоподобия Йоханссона [Johansson, 1988]), основанный на процедуре минимизации методах максимального правдоподобия VECM (кратко рассмотренной в конце предыдущего параграфа), а процедура обобщения Филлипса и Стока и Уатсона [Stock and Watson, 1988]). Эти процедуры позволяют нам тестировать нулевую гипотезу $h = h_0$, где h_0 — некоторое произвольное число между 0 и $n - 1$. Мы не рассматриваем эти процедуры¹. Здесь мы рассматриваем только простой тест, предложенный в [Engle and Granger, 1987] и обобщенный в [Phillips and Ouliaris, 1990]. В этом тесте нулевая гипотеза имеет вид: $h = 0$ (отсутствие коинтеграции), а альтернативная — $h \geq 1$.

Линейные регрессии

Тест, предложенный Энглом и Грейнджером [Engle and Granger, 1987], основывается на OLS-оценке регрессии

$$y_{1t} = \mu + \gamma' y_{2t} + \varepsilon_t \quad (10.3.1)$$

¹ См. [Maddala and Kim, 1998, Section 7] и каталог доступных процедур

где y_{1t} — первый элемент y_t , y_{2t} — вектор остальных $n - 1$ элементов, а ε_t — ошибка.

Эта регрессия была бы коинтегрирующей регрессией, если бы $h = 1$, и y_{1t} был частью коинтегрирующего соотношения. Однако при нулевой гипотезе $h = 0$ (отсутствие коинтеграции) эта регрессия не представляет коинтегрирующее соотношение. Пусть $(\hat{\beta}, \hat{\gamma})$ — OLS-оценки коэффициентов для (β, γ) . Оказывается, что $\hat{\gamma}$ не дает состоятельных оценок для любых потенциальных параметров в системе¹. Например, даже если y_{1t} не связан с y_{2t} (в том смысле, что Δy_{1t} и Δy_{2t} независимы для всех t), t - и F -статистики, ассоциированные с OLS-оценками, становятся произвольно большими при росте объема выборки, создавая ложное впечатление, что между y_{1t} и y_{2t} существует тесная связь. Этот феномен, называемый **ложной регрессией** (*spurious regression*), был впервые обнаружен в исследованиях методом Монте-Карль, представленных в [Granger and Newbold, 1974]. [Phillips, 1986] теоретически получил распределение на больших выборках статистик для ложных регрессий. Например, распределение t -значения, деленного на $\sqrt{T}$, сходится к некоторому невырожденному распределению.

Тест на коинтеграцию, основанный на остатках

Тем не менее регрессия (10.3.1) предоставляет полезный метод тестирования нулевой гипотезы об отсутствии коинтеграции, поскольку OLS-остатки, $y_{1t} - \hat{\beta} - \hat{\gamma}'y_{2t}$, должны иметь стохастический тренд, если y_t неинтегрирован, и быть стационарными в противном случае. Энгл и Грейнджер [Engle and Granger, 1987] предложили применить для тестирования нулевой гипотезы об отсутствии коинтеграции ADF t -тест для остатков. На основе использования остатков этот тест называется **тестом на коинтеграцию, основанным на остатках** (*residual-based test for cointegration*). Асимптотические распределения тестовых статистик для некоторых ведущих тестов на единичный корень были получены теоретически, и (асимптотические) критические значения для них были табулированы в [Phillips and Ouliaris, 1990] и [Johansen, 1992a].

В отличие от одномерных тестов на единичный корень эти асимптотические распределения (и, следовательно, асимптотические критические значения) зависят от размерности n системы. Это происходит потому, что остаток зависит от $(\hat{\beta}, \hat{\gamma})$, которые, будучи оценками на основе данных, являются случайными величинами. Здесь мы указываем соответствующие асимптотические критические значения для случая, когда тестом на единичный корень, применяемым к остаткам, является ADF t -тест из утверждения 9.6 для авторегрессий без константы или тренда. То есть ADF t -статистика является t -значением коэффициента при x_{t-1} в следующей расширенной авторегрессии, оцениваемой по остаткам

$$\Delta x_t = (\rho - 1)x_{t-1} + \zeta_1 \Delta x_{t-1} + \dots + \zeta_p \Delta x_{t-p} + \text{ошибка}, \quad (10.3.2)$$

где x_t здесь является остатком от регрессии (10.3.1). Включать константу в эту расширенную авторегрессию нет необходимости, поскольку при уже включенной в регрессию константе выборочное среднее остатков гарантированно равно нулю. Также нет необходимости включать время, поскольку переменные в регрессии (10.3.1) неявно или явно включают временные тренды (более подробно об этом см. ниже). Число лагированных изменений, p , должно возрастать к бесконечности с объемом выборки T , но со скоростью более низкой, чем $T^{1/3}$. Более точно¹,

$$p \rightarrow \infty, \text{ но } \frac{p}{T^{1/3}} \rightarrow 0 \text{ при } T \rightarrow \infty. \quad (10.3.3)$$

Критические значения те же самые, если вместо ADF t используется тест Филлипса Z_t (см. аналитическое упражнение 6 в предыдущей главе). Надо рассмотреть три случая.

1. $E(\Delta y_{2t}) = 0$ и $E(\Delta y_{1t}) = 0$, так что ни один элемент $I(1)$ -системы не имеет сноса. Соответствующие критические значения приведены в табл. 10.1(a), которая воспроизводит Table IIb в [Phillips and Ouliaris, 1990]. Относительно формулирования условий на VMA-представление, при которых получено асимптотическое распределение, см.: [Hamilton, 1994, Proposition 19.4].
2. $E(\Delta y_{2t}) \neq 0$, но $E(\Delta y_{1t})$ может быть, а может и не быть нулевым. Этот случай обсуждался в [Hansen, 1992a]. Пусть $g(\equiv n - 1)$ является количеством регрессоров помимо константы в регрессии (10.3.1). Некоторые из g регрессоров имеют снос. Поскольку линейные тренды из различных регрессоров можно скомбинировать в один², регрессию (10.3.1) можно записать как регрессию y_{1t} на константу, $g - 1$ $I(1)$ -регрессоров без сноса и один $I(1)$ -регрессор со сномом. Теперь, поскольку линейные тренды доминируют над стохастическими трендами (в смысле, точно указанном при обсуждении BN-разложения в предыдущей главе), $I(1)$ -регрессор с трендом ведет себя на больших выборках практически как время. Поэтому остатки получаются «асимптотически теми же самыми».

¹См.: [Phillips and Ouliaris, 1990, Theorem 4.2]. Глубина запаздываний должна быть случайной величиной, поскольку она может зависеть от данных. Условие то же самое, что и в обобщении Said – Dickey теста ADF в предыдущей главе (см. параграф 9.4), за исключением того, что p здесь может быть случайной величиной.

²Например, пусть $n = 3$, так что есть два регрессора, y_{2t} и y_{3t} , с коэффициентами γ_1 и γ_2 соответственно. Если δ_2 и δ_3 являются сносками в y_{2t} и y_{3t} соответственно, то линейные тренды в регрессии (10.3.1) равны $\gamma_1 \delta_2 t$ и $\gamma_1 \delta_3 t$ и их можно объединить в единственный временной тренд $(\gamma_1 \delta_2 + \gamma_2 \delta_3)t$.

что и остатки от регрессии с константой, $g - 1$ $I(1)$ -переменными без slopes и времени как регрессорах, в том смысле, что предельное распределение статистики, основанной на остатках от первой регрессии, то же самое, что и для основанной на остатках от второй регрессии.

Таблица 10.1

Критические значения для ADF t -статистики, примененной к остаткам

Оцененная регрессия $y_{1t} = \mu + \gamma' y_{2t}$				
Количество регрессоров, исключающих константную	1%	2,5%	5%	10%
(a) Регрессоры не имеют slopes				
1	-3,96	-3,64	-3,37	-3,07
2	-4,31	-4,02	-3,77	-3,45
3	-4,73	-4,37	-4,11	-3,83
4	-5,07	-4,71	-4,45	-4,16
5	-5,28	-4,98	-4,71	-4,43
(b) Некоторые регрессоры имеют slopes				
1	-3,96	-3,67	-3,41	-3,13
2	-4,30	-4,07	-3,80	-3,52
3	-4,65	-4,39	-4,16	-3,84
4	-5,04	-4,77	-4,49	-4,20
5	-5,30	-5,02	-4,74	-4,46

Источник: Для панели (a) — [Phillips and Ouliaris, 1990, Table IIc]. Для панели (b) — первая строка из [Follett, 1996, Table 10 A 2], остальные строки из [Phillips and Ouliaris, 1990, Table IIc].

Критические значения для ADF t -теста, основанные на последней регрессии, приведены в Table IIc в [Phillips and Ouliaris, 1990]. Следовательно, чтобы найти соответствующее критическое значение, когда регрессия (10.3.1) включает константу и g регрессоров, но не включает время, обратитесь к этой таблице для $g - 1$ регрессоров. Если регрессия имеет только один регрессор помимо константы (то есть если $g = 1$), то эта регрессия асимптотически эквивалентна регрессии y_{1t} на константу и время. Для этого случая предельное распределение ADF t -статистики, вычисленный по остаткам, оказывается распределением Дики — Фуллера (DF^*) в утверждении 9.8. В табл. 10.1(b) объединяются эти распределения: для случая единственного $I(1)$ -регрессора ($g = 1$) она показывает ADF t^* -распределение, а для случая $g > 1$ -регрессоров она показывает критические значения из Table IIc в [Phillips and Ouliaris, 1990] для $g - 1$ регрессоров. Например, если $g = 2$, то 5%-ное критическое значение равно -3,80, что равно 5%-ному критическому значению в Table IIc в [Phillips and Ouliaris, 1990] для одного регрессора.

3. Остается случай, когда $E(\Delta y_{2t}) = 0$ и $E(\Delta y_{1t}) \neq 0$. Поскольку y_{1t} имеет снос, а y_{2t} не имеет, мы должны включить в регрессию (10.3.1) время, чтобы удалить линейный тренд из остатков. Обсуждение для предыдущего случая делает понятным, что ADF t -статистика, вычисленная при использовании остатков из регрессии y_{2t} на константу, d $I(1)$ -регрессоров y_{1t} и время, имеет асимптотическое распределение, приведенное в табл. 10.1(b) для $d + 1$ регрессоров (или Table 1(c) в [Phillips and Ouliaris, 1990] для d регрессоров). Например, если $d = 2$, регрессия (10.3.1) включает константу, два $I(1)$ -регрессора и время, то критические значения можно найти из табл. 10.1(b) для трех регрессоров.

Несколько комментариев к тестам на коинтеграцию, основанным на остатках.

- (Состоятельность теста.) Альтернативная гипотеза состоит в том, что y_t коинтегрирован (то есть $h \geq 1$). Тест является состоятельным против этой альтернативы при условии, что y_{1t} коинтегрирован с y_{2t} ¹. Причина (мы обсудим ее более подробно ниже для случая $h = 1$) заключается в том, что OLS-остатки в регрессии (10.3.1) будут сходиться к стационарному процессу. Однако если y_{1t} является $I(1)$ и не является частью коинтеграционного соотношения, то тест может не иметь мощности против этой альтернативы о коинтеграции, поскольку OLS-остатки, $y_{1t} - \hat{\gamma}' y_{2t}$, с ненулевым (единичным) коэффициентом при $I(1)$ -переменной y_{1t} не будут сходиться к стационарному процессу. Таким образом, выбор нормализации (того, какую переменную следует использовать как зависимую переменную) имеет значение для состоятельности теста.
- (Нужно ли включать в регрессию время?) Если время включено в регрессию (10.3.1), то снос в y_{1t} , $E(\Delta y_{1t})$, влияет только на коэффициент при времени, делая численные значения остатков (и, следовательно, ADF t -значения) инвариантными к $E(\Delta y_{1t})$. Это означает, что случай 3 процедуры, в котором время включается в число регрессоров, можно использовать для случая 1, где $E(\Delta y_{1t})$ равно нулю. То есть если вы включаете время в регрессию (10.3.1), то соответствующие критические значения для случая 1 даны в табл. 10.1(b) для $d + 1$ регрессоров. Эта процедура обоснована также для случая 2, поскольку если время включено в дополнение к константе и d $I(1)$ -регрессоров со сном, то регрессию можно

¹Для случая $h = 1$ это является следствием Теоремы A.1 в [Phillips and Ouliaris, 1990]. Их применение (d) к этой теореме показывает, что тест является состоятельным, когда $h > 1$, и следовательно, y_t коинтегрирован.

переписать как регрессию с константой, g $I(1)$ -регрессорами и временем, которое объединяют сносы в g $I(1)$ -регрессорах. Эта регрессия попадает под случай 3, и применяются критические значения, заданные в табл. 10.1(b) для $g+1$ регрессоров. Следовательно, когда время включено в регрессию (10.3.1), можно использовать те же самые критические значения вне зависимости от местоположения снос. Возможным недостатком является пониженная мощность на конечных выборках. Мощность на конечных выборках очевидно ниже, по крайней мере для DGP, изученных в [Hansen, 1992a] в его симуляциях.

- (Выбор глубины запаздываний.) Требование (10.3.3) не дает практического правила на конечных выборках для выбора глубины запаздываний p в расширенной авторегрессии, и ее нужно оценивать по остаткам. Представляется, что в контексте тестов на основе остатков нет рабл, сравнимых с рабл [Ng and Perron, 1995] для одномерных $I(1)$ -процессов¹. Обычная практика здесь действовать так же, как и в одномерном контексте, который использует информационный критерий Акаике (AIC) или байесовский информационный критерий (BIC), называемый также информационным критерием Шварца (SIC), для определения числа лагированных изменений.
- (Рассмотрение для конечных выборок.) Помимо основанного на остатках ADF t -теста, имеется еще ряд тестов для проверки нулевой гипотезы об отсутствии коинтеграции. Таковыми являются тесты, предложенные в [Phillips and Ouliaris, 1990], [Stock and Watson, 1988] и [Johansen, 1988]. [Haug, 1996] приводит недавнее исследование методов Монте-Карло, изучающее свойства в конечных выборках этих и других тестов. Это исследование выявляет компромисс между мощностью и искажением размера (то есть тесты с наименьшим искажением размера, как правило, имеют малую мощность). Основанный на остатках ADF t -тест с глубиной запаздывания, выбранным по AIC, хотя и менее мощный, чем некоторые другие тесты, имеет наименьшее искажение размера для рассмотренных DGP.

В следующем примере основанный на остатках тест с ADF t -статистикой применяется к отношению потребление-доход:

Пример 10.6 (коинтегрированы ли потребление и доход?): Как уже было отмечено в связи с рис. 10.1(a), логарифм дохода (y_t) и логарифм потребления (c_t), по всей видимости, коинтегрированы с коинтегрирующим вектором $(1, -1)$. Этот рисунок, однако, несколько обманчив.

¹ Теперь такая работа есть. См. [Kim, 1997].

График разности $y_t - c_t$ (которая равна $-\log(1 - x_t)$, где x_t — норма сбережения) на рис. 10.1(б) показывает тенденцию к повышению сразу после войны и понижательную тенденцию с середины 1980-х годов. (Последнее является получившим широкую область фактик, что норма личиных сбережений в США снижается.) Как видно ниже, результаты теста зависят от того, включаются эти периоды или нет. Сначала мы сосредоточимся на периоде с 1950:Q1 по 1986:Q4 (148 наблюдений).

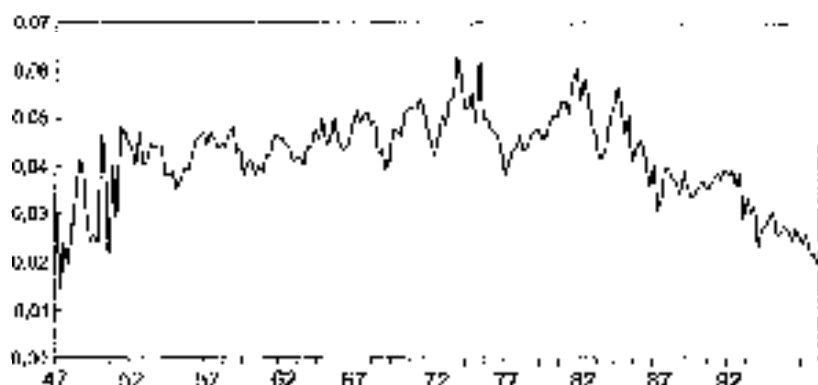


Рис. 10.1(б). Логарифм дохода минус логарифм потребления

Мы сначала проведем, являющийся указанный для ряда индивидуального $I(1)$, производя ADF t -тест с константой и линейным трендом и расширенной авторегрессии. Помимо BIC для выбора количества лагов лагированных изменений в расширенной авторегрессии, мы следуем той же самой практике, что и в предыдущей главе: максимальная глубина $p_{\max}$ равна $\lfloor 12 \cdot (T/100)^{1/4} \rfloor$ (целая часть $12 \cdot (T/100)^{1/4}$), в которой выбор глубины лагирования p после выбора фиксирован и включает моменты времени $t = p_{\max} - 2, p_{\max} - 3, \dots, T$, а после задания p для оценивания расширенной авторегрессии с p лагированными экзогенными используется максимальная выборка $t = p - 2, p - 3, \dots, T$. Для текущего объема выборки $p_{\max}$ равно 13.

Для располагаемого дохода BIC выбирает глубину запаздываний, равную 0, и ADF t -статистика для α равна -1.80 . 5%-ное критическое значение из табл. 9.2(c) равно -3.41 , поэтому мы не отвергаем нулевую гипотезу о том, что d является $I(1)$. Для лагирования глубины запаздываний p BIC равен 1 и ADF t -статистика равна -1.09 . Мы не отвергаем нулевую гипотезу H_0 на 5%-ном уровне. Таким образом, оба ряда можно считать как $I(1)$ со сном.

Обратившись к тесту на кointеграцию, основанному на остатках, находим, что OLS-оценивание статистической регрессии дает следующую ре-

зультат:

$$c_t = -0.046 - 0.975 y_t, \quad R^2 = 0.998, \quad t = 1950:Q1 \text{ to } 1986:Q4. \\ (0.009) \quad (0.0037) \quad (10.3.4)$$

Чтобы провести основанный на остатках ADF t -тест, оценивается расширенная авторегрессия без константы или времени по остаткам с глубиной запаздываний, равной 0, согласно BIC. ADF t -статистика равна -5.49 . Поскольку ряды имеют временные тренды, то, для того чтобы найти критические значения, мы обращаемся к табл. 10.1(b), а не к табл. 10.1(a). Для $\alpha = 1\%$ 5%-ное критическое значение равно -3.41 , поэтому мы можем отвергнуть нулевую гипотезу об отсутствии коинтеграции.

Если основанный на остатках тест проводится по всей выборке с 47:Q1 по 98:Q1, то ADF t -статистика с глубиной запаздываний 1, определенной согласно BIC, равна -2.94 и, таким образом, мы не можем отвергнуть гипотезу о том, что регрессия потребления на доход является ложной!

Тестирование нулевой гипотезы о коинтеграции

В тестах, рассмотренных выше, коинтеграция берется в качестве альтернативной, а не нулевой, гипотезы. Но очень часто в экономике гипотезой, представляющей экономический интерес, является коинтегрированность рассматриваемых переменных, так что было бы желательно разработать тесты, где нулевой является гипотеза $h = 1$, а не $h = 0$. Совсем недавно было предложено несколько тестов для проверки нулевой гипотезы о наличии коинтеграции. В качестве каталога таких тестов см.: [Maddala and Kim, 1998, Section 4.5]. Как было верно и в отношении тестирования нулевой гипотезы о том, что одномерный процесс является $I(0)$ -процессом, все еще не существует единственного теста на наличие коинтеграции, используемого большинством исследователей¹.

Контрольные вопросы

1. Проверьте, что если рассматривается тест на коинтеграцию, основанный на регрессии (10.3.1), не включающей время в качестве регрессора, то критические значения для случаев 1 и 2 отличаются незначительно.

¹Использовать для тестирования нулевой гипотезы о наличии коинтеграции против альтернативы об отсутствии коинтеграции процедуры [Johansen, 1988] и [Stock and Watson, 1988] нельзя, поскольку в их тестах альтернативная гипотеза специфицирует более высокий ранг коинтеграции, чем нулевая.

10.4. Статистические выводы о коинтегрирующих векторах

В предыдущем параграфе мы проверили, является ли $I(1)$ -система коинтегрированной. В данном параграфе мы предполагаем, что система коинтегрирована и что ранг коинтеграции и соответствующее треугольное представление известны¹. Наша цель — оценить коинтегрирующие векторы в треугольном представлении и получить статистические выводы о них. По большей части мы акцентируем на частном случае, где ранг коинтеграции равен 1; общий случай кратко обсуждается в конце параграфа.

Оценка OLS

Для случая $h = 1$ треугольное представление имеет вид:

$$\begin{cases} y_{1t} = \mu + \gamma' y_{2t} + \varepsilon_t^1, \\ \Delta y_{2t} = \delta_2 + \omega_{2t}, \end{cases} \quad \begin{bmatrix} z_t^* \\ \omega_{2t} \end{bmatrix}_{\substack{(1 \times 1) \\ ((n-1) \times 1)}} = \Psi^*(L) \varepsilon_t, \quad \begin{matrix} (n \times n) & (n \times 1) \end{matrix} \quad (10.4.1)$$

где y_{2t} неинтегрирован. (Это просто воспроизведение (10.2.3) и (10.2.5) с (10.2.39) для $h = 1$.) Поэтому теперь регрессия (10.3.1) является коинтегрирующей. Следовательно, существует единственный $(n-1)$ -мерный вектор γ , такой что $y_{2t} - \tilde{\gamma}' y_{2t}$ является стационарным процессом (ε_t^1) плюс инвариантная к времени случайная величина (μ), когда $\tilde{\gamma} = \gamma$, и имеет стохастический тренд, когда $\tilde{\gamma} \neq \gamma$. Это говорит о том, что OLS-оценка $(\hat{\mu}, \hat{\gamma})$ вектора коэффициентов, которая минимизирует сумму квадратов остатков, является состоятельной. Фактически, как показано в [Phillips and Durlauf, 1986] и [Stock, 1987] для случая, в котором δ_2 в (10.4.1) (= $E(\Delta y_{2t})$) равен нулю, и в [Hansen, 1992a, Theorem 1(b)] для случая $\delta_2 \neq 0$, OLS-оценка $\hat{\gamma}$ является суперсостоятельной для коинтегрирующего вектора γ , так что ошибка этой оценки сходится к 0 со скоростью, более высокой, чем обычная скорость $\sqrt{T^2}$. При этом

¹ На практике мы часто обобщаем таким образом теорему о коинтегрирующей системе с h коинтегрирующими соотношениями: обычно предполагаем, что предельная процессная теорема (такая как упомянутая в предыдущем параграфе), предназначенная для проверки ранга коинтеграции, это означает **проблему предельного тестирования (limit problem)**, поскольку распределение оцененных коинтегрирующих векторов не зависит во внимание неопределенности относительно ранга коинтеграции. Эта проблема не разрешается однозначно.

² Скорость сходимости ранга $\mu = 0$, если $\delta_2 = 0$, и $T^{1/2}$, если $\delta_2 \neq 0$. Все приведенные здесь исследования предполагают, что μ является фиксированной константой. То же самое заключение будет выполняться, даже если μ рассматривается как случайная величина. Как было отмечено в параграфе 10.1 (см. также прямо под (10.1.21)), строго говоря, процесс $\mu + \varepsilon_t^1$ является нестационарным даже если ε_t^1 является стационарным.

R^2 сходятся к единице¹. OLS-оценку для γ из коинтегрирующей регрессии мы будем называть **статической OLS (SOLS)** («static» OLS) оценкой коинтегрирующего вектора. Поскольку SOLS-оценка состоятельна, остатки сходятся к стационарному процессу с нулевым средним. Таким образом, если одномерный тест на единичный корень, такой как ADF-тест, применяется к остаткам, тест будет отвергать нулевую гипотезу $I(1)$ на больших выборках. Вот почему основанный на остатках тест из предыдущего параграфа является состоятельным против альтернативной гипотезы о коинтеграции.

Этот факт — то, что OLS-оценки коэффициентов являются состоятельными, когда регрессоры являются $I(1)$ -процессами и некоинтегрированы является замечательным результатом, резко контрастирующим со случаем стационарных регрессоров. Чтобы оценить по достоинству это отличие, вспомните из главы 3, что нужно, чтобы получить состоятельную оценку γ , когда регрессоры y_{it} являются стационарными: для того чтобы OLS-оценка была состоятельной, ошибки ε_i^* не должны коррелировать с y_{it} ; в противном случае нам нужны инструментальные переменные для y_{it} . В отличие от этого, если y_{it} коинтегрирован с y_{it} и если $I(1)$ -регрессоры y_{it} некоинтегрированы, как здесь, мы не должны беспокоиться о смещении из-за одновременности, по крайней мере на больших выборках, даже если ошибки ε_i^* и $I(1)$ -регрессоры коррелированы².

Однако в конечных выборках смещение SOLS-оценки (разность между математическим ожиданием оценки и истинным значением коэффициента) может быть большим, что было отмечено в [Banerjee et al., 1986] и [Stock, 1987]. Другой недостаток SOLS состоит в том, что асимптотическое распределение соответствующей t -статистики зависит от мешающих параметров (которыми являются коэффициенты в Ψ^*I , в (10.4.1)), что затрудняет статистические выводы. Позднее в этом параграфе мы введем другую оценку (называемую DOLS-оценкой), которая является эффективной, и при этом связанные с ней тестовые статистики (такие как t -статистика и статистика Вальда) имеют обычные асимптотические распределения.

Однако, поскольку y_{it} и ε_i^* являются строго стационарными, при $t \rightarrow \infty$ процесс y_{it} $I(1)$ асимптотически стационарный, и это все, что требуется здесь для асимптотических результатов.

¹Относительно формулировки слабей, эти условия не являются разумными, примененными, см. [Hamilton, 1994, Proposition 19.2]. Эта условка является ограничением на $\Psi^*(I)z$, в (10.4.1).

²Если y_{it} коинтегрированы, то $h > 2$ и регрессия (10.3.1) — с $m - 1$ объясняемой — не будет более канонической регрессией (коинтегрирующая регрессия должна иметь $n - h$ регрессоров, см. треугольное представление (10.2.8)). В этом случае можно применить общую методологию, предложенную в [Sims, Stock and Watson, 1990]. По поводу доступного изложения этой методологии см.: [Hamilton, 1994, Chapter 18] или [Watson, 1994, Section 2].

Двумерный пример

Чтобы прояснить потопляя и природу корреляций между ошибками в I(1)-регрессорах, а также для того чтобы проложить путь к введению DOLS-методики, рассмотрим двумерную версию (10.4.1). Предположим на время, что $\Psi^*(L) = \Psi_0^*$ (так что отсутствует сериальная корреляция в $(\varepsilon_1^*, \varepsilon_{2t}^*)$), $\beta_2 = 0$ и $\mu = 0$. Треугольное представление можно записать как

$$\begin{cases} y_{1t} = \alpha_{y1} + \varepsilon_t^*, \\ \Delta y_{2t} = \alpha_{y2}, \end{cases} \quad \begin{bmatrix} \varepsilon_t^* \\ \alpha_{y2} \end{bmatrix} = \Psi_0^* \varepsilon_t. \quad (10.4.2)$$

Ошибка ε_t^* и I(1)-регрессор y_{1t} в коинтегрирующей регрессии в (10.4.2) могут быть коррелированными, поскольку

$$\begin{aligned} \text{Cov}(y_{2t}, \varepsilon_t^*) &= \text{Cov}(y_{2,t-1} + \Delta y_{2,t-1} + \Delta y_{2,t-2} + \dots + \Delta y_{2t}, \varepsilon_t^*) = \\ &= \text{Cov}(\Delta y_{2,t-1} + \Delta y_{2,t-2} + \dots + \Delta y_{2t}, \varepsilon_t^*) \\ &\quad (\text{так как } \text{Cov}(y_{2,t-1}, \varepsilon_t^*) = 0) \\ &= \text{Cov}(\Delta y_{2,t-1} + \Delta y_{2,t-2} + \dots + \Delta y_{2t}, \varepsilon_t^*) \quad (\text{так как } \Delta y_{2t} = 0) \\ &= \text{Cov}(\varepsilon_{2t}, \varepsilon_t^*) \quad (\text{так как } (\varepsilon_{2t}, \varepsilon_t^*) \text{ независимы}). \end{aligned} \quad (10.4.3)$$

Поскольку требование диагональности Ψ_0^* и Σ не было, ε_t^* и ε_{2t} могут быть одновременно коррелированными.

Чтобы изолировать эту возможную корреляцию, рассмотрим проекции наклонных квадратов ε_t^* на константу и ε_{2t} ($= \Delta y_{2t}$). Вспомогательный параграф 2.9, что $\tilde{E}^*(y|1, \varepsilon_{2t}) = [E(y) - \beta_0 E(\varepsilon_{2t}) + \beta_2 \varepsilon_{2t}]$, где $\beta_0 = \text{Cov}(y, y)/\text{Var}(y)$ и β_2 — ошибка проекции наклонных квадратов, $y = \tilde{E}^*(y|1, \varepsilon_{2t})$, имеет нулевое среднее и не коррелирован с ε_{2t} . Здесь среднее равно нулю и для ε_t^* и для ε_{2t} . Поэтом $\tilde{E}^*(\varepsilon_t^*|1, \varepsilon_{2t}) = \beta_0 \varepsilon_{2t}$, если мы обозначим ошибку проекции метода наклонных квадратов как $v_t = \varepsilon_t^* - \tilde{E}^*(\varepsilon_t^*|1, \varepsilon_{2t})$, то имеем:

$$\begin{aligned} \varepsilon_t^* &= \beta_0 \varepsilon_{2t} + v_t = \beta_0 \Delta y_{2t} + v_t, \\ E(v_t) &= 0, \quad \text{Cov}(\varepsilon_{2t}, v_t) = \text{Cov}(\Delta y_{2t}, v_t) = 0. \end{aligned} \quad (10.4.4)$$

Подставляя это в коинтегрирующую регрессию, получаем:

$$y_{1t} = \alpha_{y1} + \beta_0 \Delta y_{2t} + v_t. \quad (10.4.5)$$

Эта регрессия будет называться **расширенной коинтегрирующей регрессией** (*augmented cointegrating regression*). Сейчас мы покажем, что I(1)-регрессор y_{2t} является *слабо экзогенным* в том смысле, что для всех t , $\text{Cov}(y_{2t}, v_t) = 0$. По построению, $\text{Cov}(\Delta y_{2t}, v_t) = 0$. Для $t \neq 1$ и для $\text{Cov}(\Delta y_{2t}, v_1)$ имеем:

$$\begin{aligned} \text{Cov}(\Delta y_{2t}, v_1) &= \text{Cov}(\varepsilon_{2t}, v_1) = \text{Cov}(\varepsilon_{2t}, \varepsilon_1^* - \beta_0 \varepsilon_{21}) = \\ &= \text{Cov}(\varepsilon_{2t}, \varepsilon_1^*) - \beta_0 \text{Cov}(\varepsilon_{2t}, \varepsilon_{21}) = 0. \end{aligned} \quad (10.4.6)$$

поскольку $(1, 1, \dots, 1, 0, \dots, 0)$ является л.д. Поэтому $\Delta y_{2,t}$ является строго экзогенным. Строгая экзогенность $\Delta y_{2,t}$ влечет то же самое и для $y_{2,t}$, поскольку

$$\begin{aligned} \text{Cov}(y_{2,t}, y_t) &= \text{Cov}(y_{2,t-1} + \Delta y_{2,t-1} + \Delta y_{2,t-2} + \dots + \Delta y_{2,t-1}, y_t) = \\ &= \text{Cov}(\Delta y_{2,t-1} + \Delta y_{2,t-2} + \dots + \Delta y_{2,t-1}, y_t) \\ &= 0 \quad \text{поскольку } \text{Cov}(y_{2,t-1}, y_t) = 0 \\ &= 0 \quad \text{для всех } t, \text{ а по (10.4.6).} \end{aligned} \quad (10.4.7)$$

Продолжение двумерного примера

Разминая, мы показали, что в расширенной коинтегрирующей регрессии (10.4.3) $I(1)$ -регрессор является строго экзогенным, если $\Psi^*(L) = \Psi^*$. Пусть $(\hat{\beta}_1, \hat{\beta}_2)$ является OLS-оценкой для (β_1, β_2) из расширенной коинтегрирующей регрессии. (От $\hat{\beta}_1$ следует отделить от 3OLS-оценки для β_1 в стандартной регрессии y (10.4.2).) Эта регрессия (10.4.3), опять же, основана на дисперсионном векторе регрессии (9.4.6) в том, что если из $I(1)$ -регрессора вычитать $I(0)$ с нулевым средним, то другой $I(1)$ будет иметь $I(0)$ в качестве дисперсионного вектора. Это матрица « $X'X$ », если она правильно стандартизирована на T и $\sqrt{T}$, тогда она асимптотически диагональна, так что матрица $I(0)$ -регрессора можно моделировать предельное нормальное распределение $\hat{\beta}_1$. OLS-оценки коэффициентов при $I(1)$ -регрессоре. То же самое верно и здесь, в такой же самой аргументации, основанной на асимптотической диагональности, после нормализованной соответствующим образом матрицы « $X'X$ », доказывает, что обычное t -значение для тестирования гипотезы $\beta_2 = 0$ асимптотически эквивалентно:

$$\bar{t} = \frac{\frac{1}{T} \sum_{t=1}^T y_{2,t} \Delta y_{2,t}}{\sqrt{\frac{\sigma_y^2}{T} \sum_{t=1}^T (\Delta y_{2,t})^2}}. \quad (10.4.8)$$

где σ_y^2 является дисперсией y_t . То есть разность между обычным t -значением к $\bar{t}$ в (10.4.8) сходится к нулю по вероятности, а поэтому предельные распределения $\bar{t}$ и $\bar{t}$ одинаковы.

В расширенной авторегрессии, используемой для ADF-теста, предельное распределение ADF t -статистики было стандартным (она является распределением Дик — Фуллера). В отличие от этого асимптотическое распределение $\bar{t}$ (и, следовательно, обычных t -значений) является стандартным нормальным. Чтобы увидеть, почему это так, вспомним, что $I(1)$ -регрессор $y_{2,t}$ является строго экзогенным в том смысле, что $\text{Cov}(y_{2,t}, y_t) = 0$ для всех t, y . Чтобы развить интуицию, временно предположим, что $(\varepsilon_1, \varepsilon_2)$ совместно нормальны. Тогда $y_{2,t}$ и y_t являются независимыми для всех t, y , а не просто некоррелированными. Следовательно,

ю, распределение u_t условное относительно $(y_{2,1}, y_{2,2}, \dots, y_{2,T})$, то же самое, что и безусловное распределение, которое есть $N(0, \sigma_u^2)$. Поэтому распределение числителя (10.4.8), условное относительно $(y_{2,1}, \dots, y_{2,T})$, имеет вид

$$N\left(0, \frac{\sigma_u^2}{T} \sum_{t=1}^T (y_{2t})^2\right).$$

Но стандартное стандартное нормальное распределение дано знаменателем $\hat{\sigma}_u$, и поэтому

$$\tilde{t} \mid (y_{2,1}, y_{2,2}, \dots, y_{2,T}) \sim N(0, 1).$$

Поскольку это условное распределение $\tilde{t}$ не зависит от $(y_{2,1}, y_{2,2}, \dots, y_{2,T})$, безусловным распределением $\tilde{t}$ также будет $N(0, 1)$. (Этот способ доказательства не является новым для нас: см. ящик, содержащий (1.4.3), в главе 1). Вспомнив из главы 2, что, даже если предположение нормальности отбросить, распределение обычного t -значения является стандартным нормальным, хотя и асимптотически, на больших выборках. Тот же самый вывод верен и здесь: предельное распределение $\tilde{t}$ есть $N(0, 1)$, даже если (z_t^*, y_{2t}) не являются совместно нормальными. Доказательство этого требует аргументов, отличных от тех, которые были использованы в главе 2, поскольку y_{2t} здесь имеет стохастический тренд. Мы не будем давать здесь формальное доказательство; только краткое резюме. Доказательство состоит из двух частей. Первая заключается в том, чтобы получить предельное распределение числителя (10.4.8) (что можно сделать, используя результат, сформулированный, например, Proposition 18.1(c) в [Hamilton, 1994] или Lemma 2.3(c) в [Watson, 1994]). Это распределение является нестандартным. Вторая состоит в том, чтобы показать, что нормализация (10.4.8) преобразует нестандартное распределение в нормальное распределение (Lemma 5.1 в [Park and Phillips, 1988]).

Этот пример двумерной $I(1)$ системы является частным в нескольких отношениях: (a) нет серийной корреляции в процессе ошибок (z_t^*, u_{2t}) , (b) $I(1)$ -регрессор y_{2t} является скалярным, (c) g_{2t} не имеет смысла, и (d) $\mu = 0$. Конечно, ослабление (a) требует некоторого размышления.

Допущение серийной корреляции

Если предположение $\Psi^*(L) = \Phi_0^*$, сделанное в (10.4.2), не выполняется, то (z_t^*, u_{2t}) является скалярно коррелированным. Следовательно, $I(1)$ -регрессор y_{2t} в расширенной коинтегрирующей регрессии (10.4.5) больше не будет строго экзогенным, поскольку $\text{Cov}(\Delta z_{2t-1}, y_{2t})$, оставаясь нулевой для $t = 1$, больше не является гарантированно нулевой для $t \neq 1$. Чтобы устранить эту корреляцию, рассмотрим проекцию метода наименьших квадратов z_t^* на текущие, прошлые и будущие значения y_{2t} , а не только

на текущее значение y_t . Замечая, что $\Delta y_{2t} = y_{2t}$, эту прокладку можно записать как

$$z_t' = \beta(L)\Delta y_{2t} - y_{2t}, \quad \beta(L) = \sum_{j=-\infty}^{\infty} \beta_j L^j \quad (10.4.9)$$

(Эта y_t отличается от y_t в (10.4.5)). По построению, член y_{2t} имеет $E(y_t) = 0$ и $\text{Cov}(\Delta y_{2t}, y_t) = 0$. Другая причина, по которой $\beta(L)$ может иметь бесконечный порядок, то предположим, что для любой момент, что это не так и $\beta_j = 0$ для $|j| > p$. Таким образом,

$$z_t' = \beta_0 \Delta y_{2t} + \beta_1 \Delta y_{2t-1} + \dots + \beta_{-p} \Delta y_{2t+p} + \beta_1 \Delta y_{2t-1} + \dots + \beta_p \Delta y_{2t+p} - y_{2t}. \quad (10.4.10)$$

Подставляя это в коинтегрирующую регрессию в (10.4.2), мы получаем (заметим!) расширенную коинтегрирующую регрессию:

$$y_{1t} = \gamma y_{1t} + \beta_0 \Delta y_{2t} + \beta_1 \Delta y_{2t-1} + \dots + \beta_{-p} \Delta y_{2t+p} + \beta_1 \Delta y_{2t-1} + \dots + \beta_p \Delta y_{2t+p} + \varepsilon_t \quad (10.4.11)$$

Поскольку $\text{Cov}(\Delta y_{2t}, y_t) = 0$ для всех i и k , Δy_{2t} является строго экзогенным, что означает (как было показано выше), что регрессор в уравнении y_{1t} также является строго экзогенным. Таким образом, включив в расширенную коинтегрирующую регрессию не только текущее значение, но также прошлые и будущие значения $I(1)$ -регрессора, мы в состоянии сохранить строгую экзогенность y_{2t} . Оценку коинтегрирующего вектора γ , основанную на расширенной коинтегрирующей регрессии, назовем **динамической OLS (DOLS)** (dynamic OLS), чтобы отличать ее от SOLS-оценки, основанной на коинтегрирующей регрессии без приращений в y_2 .

У расширенной регрессии имеется $2 - 2p$ регрессоров, первый из которых есть $I(1)$ без наклона, а остальные — $I(0)$ с нулевым средним. Как и выше, при надлежащем масштабировании на T и $\sqrt{T}$, $I(1)$ -регрессор асимптотически не коррелирован с $2p + 1$ $I(0)$ -регрессорами с нулевым средним, так что соответствующим образом масштабированная матрица $X'X$ является блочно-диагональной, в $I(0)$ -регрессоры с нулевым средним можно игнорировать при анализе предельного распределения DOLS-оценки для γ . Следовательно, выражение для случайной величины, асимптотически эквивалентной T -значению для $\gamma = \gamma_0$, снова задается как (10.4.6), выражение, полученное для случая, в котором (z_t', y_{1t}) строго некоррелированы.

Однако, когда (z_t', y_{1t}) строго коррелированы, вывод асимптотическое распределение (10.4.8) отличается от случая, когда (z_t', y_{1t}) строго некоррелированы, поскольку ошибки ε_t могут быть теперь строго коррелированными, поскольку (10.4.9), устраняя корреляции между

Λ_{η_2} и η_2 для всех k, l , не устраняет собственную сериальную корреляцию в η_2 . С целью изучения асимптотического распределения, пусть V ($1 \times T$) является автоковариационной матрицей T последовательных значений η_2 , а $\lambda_{\eta_2}^2$ является долговременной дисперсией η_2 . Также пусть X в оставшейся части этого подпараметра является $T \times 1$ матрицей $(y_{21}, y_{22}, \dots, y_{2T})'$. И опять, чтобы развить интуицию, временно предположим, что (z_1^t, η_{2t}) является совместно нормальным. Тогда, поскольку y_{21} строго экзогенный, распределение числителя в (10.4.8), условное относительно X , является нормальным с нулевым средним и условной дисперсией

$$\frac{1}{T^2} X' V X. \quad (10.4.12)$$

Чтобы сделать отношение стандартным нормальным на конечных выборках, пришлось бы заменить знаменатель в (10.4.8) на квадратный корень из этой величины. К счастью, оказывается (см.: Спинулату 2.7 в [Phillips, 1988]), что на больших выборках распределение числителя является тем же самым, что и распределение, которое вы бы получили, если бы η_2 был бы сериально некоррелированным, но с дисперсией $\lambda_{\eta_2}^2$, а не $\sigma_{\eta_2}^2$.

Следовательно, все, что требуется, чтобы модифицировать отношение (10.4.8) для учета сериальной корреляции в η_2 , — это заменить $\sigma_{\eta_2}^2$ ($\equiv \text{Var}(\eta_2)$) на $\lambda_{\eta_2}^2$ (долговременная дисперсия η_2), а именно если $\hat{t}$ задается как (10.4.8) с $\sigma_{\eta_2}^2$, все еще находящейся в знаменателе, то шкалированное значение

$$\left(\frac{\sigma_{\eta_2}}{\lambda_{\eta_2}} \right) \cdot \hat{t} \quad (10.4.13)$$

распределено асимптотически как $N(0, 1)$. Поскольку OLS t -значение для γ асимптотически эквивалентно $\hat{t}$, то из этого следует, что

$$\left(\frac{\sigma_{\eta_2}}{\lambda_{\eta_2}} \right) \cdot t \rightarrow_d N(0, 1), \quad (10.4.14)$$

где λ_{η_2} является некоторой состоятельной оценкой для λ_{η_2} , а σ_{η_2} — обычная стандартная ошибка OLS регрессии (легко показать, что σ_{η_2} является состоятельной для σ_{η_2}). Иными словами, если мы шкалируем обычную стандартную ошибку для $\hat{\gamma}$ (OLS-ошибка коэффициента при η_2 в (10.4.11)) как

шкалированная стандартная ошибка —

$$= \left(\frac{\lambda_{\eta_2}}{\sigma_{\eta_2}} \right) \times \text{обычная стандартная ошибка}, \quad (10.4.15)$$

то t -значение, основанное на шкалированной стандартной ошибке, является асимптотически $N(0, 1)$.

Приведенная аргументация применима только к коэффициенту при $I(1)$ -регрессоре. Поэтому t -значения для значений β в (10.4.11), даже когда их стандартные ошибки шкалированы так, как только что было описано, несомненно являются $N(0,1)$ на больших выборках.

Поскольку регрессоры строго экзогенны, обсуждение в параграфах 1.6 и 6.7 приводит нас к мысли, что можно применить GLS. Однако из отмеченного выше факта, что $X'VX$ ведет себя асимптотически как $\lambda_n X'X$, следует, что GLS-оценка для γ (называемая **DGLS**-оценкой) асимптотически эквивалентна OLS-оценке (или более точно T , умноженное на их разность, сходится к 0 по вероятности) (см. [Phillips, 1988] и [Phillips and Park, 1988]). Следовательно, выстраивая в коррелятивности от коррекции на серийную корреляцию в ошибках e_t посредством GLS нет

Состоятельную оценку для λ_n получить несложно. Вспомним, что в параграфе 6.5 мы использовали для оценивания долгосрочной ковариационной матрицы векторного процесса процедуру VARHAC. Эту же самую процедуру можно применять в данном случае скалярного процесса ошибок. Рассмотрим набор $AR(p)$ -процесса с остатком $\hat{e}_t$ от расширенной коинтегрирующей регрессии

$$\hat{e}_t = \rho_1 \hat{e}_{t-1} + \rho_2 \hat{e}_{t-2} + \dots + \rho_p \hat{e}_{t-p} + \varepsilon_t \quad (t = p+1, \dots, T). \quad (10.4.16)$$

И здесь же следует отметить, что в расширенной коинтегрирующей регрессии (10.4.16) выбор глубины p принципиально не используется. ВК посредством (10.4.16) может использоваться для любых лагов $0, 1, \dots, T^{1/2}$. Используя соотношение (6.2.21) на $\varepsilon_t = \hat{e}_t$ и допуская, что одновременно дисперсия ряда значений преобразованной функции автокорреляций при $\tau \rightarrow \infty$, долгосрочную дисперсию λ_n^2 для γ можно оценить как

$$\hat{\lambda}_n^2 = \frac{\hat{\sigma}_\varepsilon^2}{(1 - \hat{\alpha}_1 - \dots - \hat{\alpha}_p)^2} \approx \frac{1}{T - p} \sum_{i=p+1}^T \hat{e}_i^2, \quad (10.4.17)$$

где $\hat{\alpha}_j$ ($j = 1, 2, \dots, p$) — оценочные $AR(p)$ коэффициенты, а $\hat{e}_i$ — остатки от $AR(p)$ уравнения (10.4.16).

Общий случай

Обратимся теперь к общему случаю n -мерной коинтегрированной системы с возможно ненулевым сносом. Мы фокусируемся на случае $h = 1$, поскольку обобщение на случай, в котором $h > 1$, следует непосредственно. Так что треугольное представление имеет вид (10.4.1), а расширенная коинтегрирующая регрессия имеет вид

$$\begin{aligned} y_t = & \gamma + \gamma' y_{1t} + \beta_1' \Delta y_{1t} + \beta_2' \Delta y_{1t+1} + \dots + \beta_{p-1}' \Delta y_{1t+p-1} + \\ & + \beta_1' \Delta y_{2t-1} + \dots + \beta_p' \Delta y_{2t-p} + \varepsilon_t. \end{aligned} \quad (10.4.18)$$

Здесь β_j ($j = 0, 1, 2, \dots, p, -1, -2, \dots, -p$) являются коэффициентами проекции метода наименьших квадратов в проекции z_t^* (ошибка в кointегрирующей регрессии) на текущие, прошлые и будущие значения Δy_{2t} . DOLS-оценка кointегрирующего вектора γ — это просто OLS-оценка для γ из этой расширенной кointегрирующей регрессии.

Следующие результаты доказаны в [Saikkonen, 1991] и [Stock and Watson, 1993]:¹

1. Описанные выше результаты для двумерного случая остаются в силе. То есть DOLS-оценка для γ является суперсостоятельной, и надлежащим образом масштабированные t -статистика и статистика Вальда для тестирования гипотез относительно γ имеют обычные асимптотические распределения (стандартное нормальное и хи-квадрат). (Правильное масштабирование — умножить обычное t -значение на $(s/\hat{\lambda}_n)$ и статистику Вальда на квадрат $(s/\hat{\lambda}_n)$. Однако это простое масштабирование не работает для t -значений и статистики Вальда в случае тестирования гипотез, включающих ρ или β_j).
2. Если двухсторонний фильтр $\beta(L)$ в проекции z_t^* на опережающие и запаздывающие значения Δy_{2t} имеет бесконечный порядок, то тогда в состав ошибки u_t входит остаток от усечения:

$$\sum_{j=p-1}^{\infty} \beta'_{-j} \Delta y_{2t-1-j} + \sum_{j=p+1}^{\infty} \beta'_j \Delta y_{2t+j}.$$

Все результаты, только что упомянутые выше, выполняются при условии, что p в (10.4.18) полагается таким, чтобы его значение возрастало с увеличением объема выборки T со скоростью, меньшей, чем $T^{1/3}$. См. [Saikkonen, 1991, Section 4] относительно формулирования требуемых условий регулярности.

3. Рассматриваемая оценка является эффективной в некотором точном смысле². Она асимптотически эквивалентна другим эффективным оценкам, таким как процедура метода максимального правдоподобия Йоханссона, основанная на VECM, «полностью модифицированная» оценка (*Fully Modified estimator*) Филлипса и Хансена

¹ Они также предполагают, что p является фиксированной константой. Однако те же самые выводы сохраняются, даже если p является порядком $1/\log T$ (или даже случайной функцией).

² Это не означает, что u_t асимптотически $N(0, 1)$, но сама DOLS-оценка не является асимптотически нормальной. Поэтому обычный критерий сравнения двусторонней асимптотической эффективности в классе асимптотически нормальных оценок здесь неприменим. См. [Saikkonen, 1991, Section 5] относительно доказательств распределения «эффективности».

[Phillips and Hansen, 1990}], оценка нелинейного метода наименьших квадратов [Phillips and Loretan, 1991] (а также DGLS, как было отмечено выше). Для всех этих эффективных оценок t -статистика и статистика Вальда для тестирования гипотез относительно γ , правильно образом скаляризованные, если это необходимо, имеют стандартные асимптотические распределения (более подробно см.: [Watson, 1994, Section 3.4.3]).

Другие оценки и свойства на конечных выборках

[Stock and Watson, 1993] изучали свойства только что упомянутых оценок (за исключением оценки Филлипса - Лоретана) в конечных выборках. Для различных DGP и объемов выборки ($T = 100$ и 300), которые они изучали, вырисовывались следующие результаты. 1) Смещение является наименьшим для оценки Йоханссона и наибольшим для SOLS. 2) Однако дисперсия оценки Йоханссона намного больше, чем дисперсия других эффективных оценок, и DGLS имеет наименьшую среднеквадратичную ошибку (RMSE). 3) Для всех изученных оценок распределение t -значений (скаляризованных правильно, если необходимо) показывает тенденцию иметь более длинные хвосты, по сравнению с $N(0,1)$, указывая, что нулевая гипотеза будет отвергаться слишком часто. Эти выводы в целом соответствуют выводам на основании исследований методом Монте-Карло в [Maddala and Kim, 1998, Section 5.7].

Контрольные вопросы

1. Как мы подчеркивали, существует сходство между расширенной авторегрессией (9.4.6) на $n = 622$ и расширенной коинтегрирующей регрессией (10.4.5). Тогда почему t -значение при (1) -регрессии в расширенной коинтегрирующей регрессии (10.4.5) асимптотически распределено как $N(0,1)$, в то время как в (9.4.6) это не так?

10.5. Приложение: Спрос на деньги в США

Литература по спецификациям функции спроса на деньги очень обширна (см.: [Goldfeld and Nighel, 1990], в качестве обзора). Уравнение спроса на деньги, обычно оцениваемое в литературе, имеет вид:

$$m_t - p_t = \mu + \gamma_y y_t + \gamma_R R_t + \varepsilon_t^m, \quad (10.5.1)$$

где m_t — логарифм денежной массы в период t , p_t — логарифм уровня цен, y_t — логарифм дохода, R_t — номинальная процентная ставка и ε_t^m — некоторая ошибка. γ_y является эластичностью спроса на деньги по доходу, а γ_R — полуэластичностью спроса на деньги по процентной ставке

(это полужластичность, поскольку процентная ставка входит в уравнение спроса на деньги в уровнях, а не в логарифмах). Большинство эмпирических исследований, которые предпочтительнее литературе по единичным корням и коинтеграции, страдают от двух недостатков. Первый, как будет показано ниже, заключается в том, что единичные переменные, m_t — y_t , R_t и R_{t-1} , по-видимому, содержат стохастические тренды. Если переменные имеют тренды, обычные статистические выводы, опирающиеся на предположение о стационарности, не будут обоснованными. Второй — регрессоры могут быть эндогенными. В случае спроса на деньги это может быть серьезной проблемой, поскольку практически во всех макроэкономических моделях сдвиг в спросе на деньги, представленный ошибками ε_t^d , влияет на номинальную процентную ставку и, возможно, доход. Разрешение этих проблем обеспечивается эконометрической техникой оценивания коинтегрирующих регрессий, представленной в предыдущем параграфе.

Другая проблема, которая рассматривается в этом параграфе, — это проблема стабильности функции спроса на деньги. Консенсусом в литературе является то, что спрос на деньги нестабилен. Это инспирируется в [Lucas, 1988], который, используя годовые данные с 1900 года, утверждает, что полужластичность процентной ставки стабильна, если наложить ограничение, что эластичность по доходу равна единице. Мы проверим утверждение Лукаса, применяя современные эконометрические методы, развитые в этой главе.

Данные

Наш анализ основан на годовых данных, исследованных в [Lucas, 1988] и расширенных в [Stock and Watson, 1993], чтобы покрыть период 1900–1989 годов. Их мерами денежной массы, дохода и номинальной процентной ставки служат M1, чистый национальный продукт (ВНП) и ставка по шестимесячным векселям (в процентах к годовым исчислениям) соответственно. Использование *чистого* национального продукта, а не валового национального продукта, следует традиции, восходящей к [Friedman, 1959]. Переменная *массовая* деятельность (m_t) в уравнении спроса на деньги должна быть мерой богатства, а не мерой объема сделок. Поскольку официальная статистика по M1 до 1900 года отсутствует, приходится прибегнуть к эмпирическому оцениванию рядов из нескольких источников. Для денежной массы и процентной ставки месячные данные усредняются, чтобы получить годовые наблюдения. (Подробнее о построении данных см. Appendix B в [Stock and Watson, 1993].)

$(m_t - p_t, y_t, R_t)$ как коинтегрированная система

На рис. 10.2 и 10.3 изображены графики рядов $m_t - p_t$, y_t и R_t . Все три ряда имеют явные тренды. Логарифм реальной денежной массы ($m_t - p_t$),

с одной стороны, быстро возрастал в течение первой половины века, но не испытывал почти никакого роста в дальнейшем до 1981 года. Логарифм дохода (y), с другой стороны, неуклонно рос в течение всего периода выборки с серьезным нарушением этого роста с 1930-х до начала 1940-х, поэтому скорость обращения $M1$ ($y - (m - p)$), также изображенная на рис. 10.3, сократалась в течение этого периода, а затем неуклонно возрастала до 1981 года. За этим движением скорости обращения $M1$ довольно тесно следует номинальная процентная ставка, что приводит нас к мысли о том, что $m - p$, y и R являются интегрированными с коинтегрирующим вектором, компонента которого, соответствующая y , близка к 1.

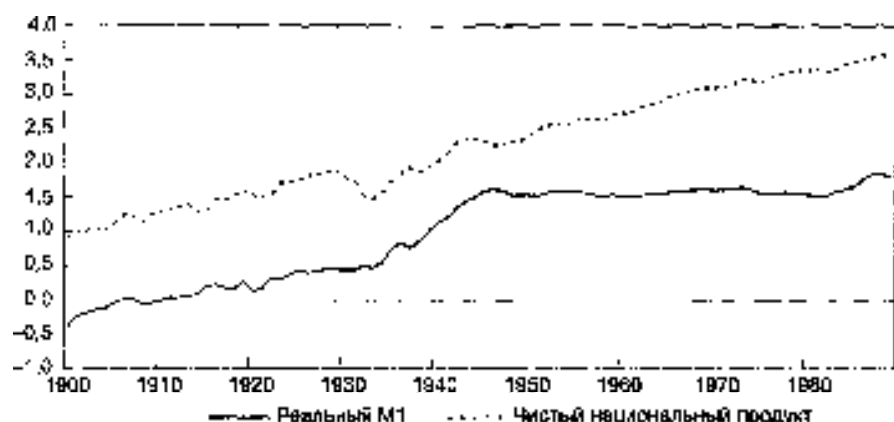


Рис. 10.2. Реальный ВНП США и реальная $M1$ в логарифмах

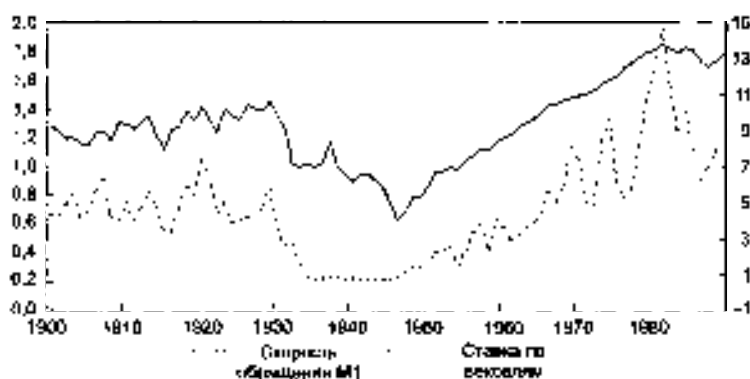


Рис. 10.3. Логарифм скорости обращения $M1$ (левая шкала) и краткосрочная ставка по векселям (правая шкала)

Анализ этих рисунков подсказывает, что эти три переменные, $m - p$, y и R , индивидуально можно хорошо описать $I(1)$ -процессом. Действительно, [Stock and Watson, 1993] на основе ADF-тестов сообщают, что $m - p$ и y каждый является $I(1)$ -процессом со сроком, а то время как

R является $I(1)$ -процессом без сноса. Они также сообщают, основываясь на тестировании общих трендов в [Stock and Watson, 1988], что эта трехмерная система является кointегрированной с рангом кointеграции, равным 1. В дальнейшем мы исходим из предположения, что $m - p$ кointегрирована с y и R . (В эмпирическом упражнении нас попросят проверить справедливость этой предположки.)

DOLS

Первая строка табл. 10.2 показывает SOLS-оценку кointегрирующей регрессии (10.5.1). (Период выборки — 1903–1987 годы, поскольку в расширенную кointегрирующую регрессию DOLS строкой ниже будут включены два опережающих и два запаздывающих изменения.) Стандартные ошибки не приводятся, поскольку асимптотическое распределение

Таблица 10.2

SOLS- и DOLS-оценки, 1903–1987 годы

	γ_m	γ_R	R^2	Стд. ошибки регрессии
SOLS-оценки	0,945	-0,082	0,959	0,135
DOLS-оценки	0,970	-0,101	0,952	0,096
(Опережающие стандартные ошибки) (0,046) (0,013)				

Источник: Точечные оценки и стандартные ошибки из Table III в [Stock and Watson, 1995]. R^2 и стандартная ошибка регрессии из авторских расчетов.

соответствующего t -отношения, зависящее от масштабирующих параметров, неизвестно. SOLS-оценка эластичности по доходу, равная 0,945, близка к единице, но мы не можем сказать, является ли она незначимо отличной от единицы. Чтобы быть в состоянии делать статистические выводы, мы обращаемся к DOLS-оценкам кointегрирующего вектора, который должно дать оценки параметров $(\mu, \gamma_y, \gamma_R)$ кointегрирующей регрессии (10.5.1) за счет добавления Δy_t , ΔR_t и их опережающих и запаздывающих значений. Следуя [Stock and Watson, 1993], число опережающих и запаздывающих значений (произвольно) выбрали равным 2, поэтому расширенная кointегрирующая регрессия, связанная с (10.5.1), имеет вид:

$$\begin{aligned}
 m_t - p_t = & \mu + \gamma_y y_t + \gamma_R R_t + \\
 & + \beta_{y0} \Delta y_t + \beta_{y-1} \Delta y_{t-1} + \beta_{y-2} \Delta y_{t-2} + \beta_{y1} \Delta y_{t+1} + \beta_{y2} \Delta y_{t+2} + \\
 & + \beta_{R0} \Delta R_t + \beta_{R-1} \Delta R_{t-1} + \beta_{R-2} \Delta R_{t-2} + \\
 & + \beta_{R1} \Delta R_{t+1} + \beta_{R2} \Delta R_{t+2} + v_t. \quad (10.5.2)
 \end{aligned}$$

Это обусловливает период выборки $t = 1903, \dots, 1987$. Точечные DOLS-оценки $(\hat{\beta}_1, \hat{\beta}_2)$, приведенные во второй строке табл. 10.2, получены в результате подбора к этому уравнению посредством OLS.

Чтобы вывести соответствующие стандартные ошибки, необходимо оценить долговременную дисперсию ошибки σ^2 . С этой целью мы подбираем к DOLS-остаткам процесс авторегрессии. Порядок авторегрессии (снова произвольно) подгадывается равным 2. С DOLS-остатками, вычисленными для $t = 1903, \dots, 1987$, периодом выборки для оценивания АК(2) ставится $t = 1905, \dots, 1987$ (объем выборки = 81). Оцененная авторегрессия имеет вид:

$$\hat{\varepsilon}_t = 0.03806 \hat{\varepsilon}_{t-1} + 0.10941 \hat{\varepsilon}_{t-2}, \quad SSR = 0.19843. \quad (10.5.3)$$

Затем мы используем формулу (10.4.17), т.е. чтобы учесть возможность воспроизводимости DOLS-оценок (Stock and Watson, 1995, Table F1), для вычисления $\hat{\sigma}_r^2$ мы делим SSR не на $T - p$, а на $T - p - K$, где K означает количество регрессоров в расширенной коинтегрированной регрессии (которое равно здесь 15). Таким образом, $\hat{\sigma}_r^2 = 0.19843/(83 - 2 - 15) = 0.0029181$, а

$$\hat{\sigma}_r = \frac{0.0029181}{(1 - 0.03806 + 0.13311)^2} = 0.07047. \quad (10.5.4)$$

Оценка $\hat{\sigma}_r$ является квадратным корнем из этой величины, которая равна 0,277. Поскольку стандартная ошибка DOLS регрессии равна 0,096, множитель в формуле (10.4.15) для скаляризации обычных стандартных ошибок OLS из расширенной коинтегрированной авторегрессии равен $0.277/0.096 = 2.89$. Числа в скобках в табл. 10.2 являются скорректированными стандартными ошибками, вычисленными таким образом. Теперь мы можем видеть, что оцененная точечность по доходу не является значимо отличной от единицы.

Нестабильный спрос на деньги?

В табл. 10.3 приводятся оценки параметров по SOLS и DOLS для двух подвыборок, 1903–1945 и 1946–1987 годы. (См. также [Stock and Watson, 1993], дата события 1946 года была выбрана из-за причины естественного события в конце Второй мировой войны, т.е. по причине того, что она разделяет выборку почти пополам.) В резком контрасте с оценками, основанными на первой половине выборки, послевоенные оценки сильно отличаются от тех, которые основаны на полной выборке.

[Lucas, 1988] утверждает, что оценка, подобная этой, соответствует стабильному спросу на деньги с единичной эластичностью по доходу. В поддержку этого суждения он указывает на рис. 10.4. На этом рисунке изображен график зависимости $m_t - \mu - y_t$ (логарифм обратной

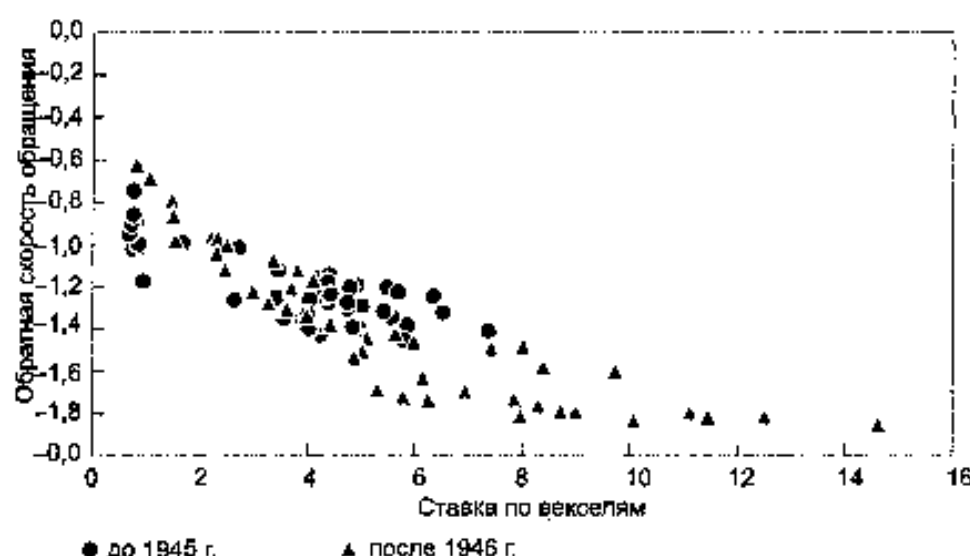
скорости обращения, которая была бы зависимой переменной в уравнении спроса на деньги (10.5.1), если бы эластичность по доходу γ_y полагалась равной единице) от процентной ставки. При этом послевоенные наблюдения указаны другим набором символов, нежели довоенные наблюдения. Отличительной чертой является то, что, хотя послевоенные процентные ставки значительно выше, послевоенные наблюдения лежат на линии, определяемой предвоенными наблюдениями¹.

Таблица 10.3

Спрос на деньги в двух подвыборках

	1903–1945		1946–1967	
	γ_y	γ_R	γ_y	γ_R
SOLS-оценки	0,919	–0,085	0,192	–0,016
DOLS-оценки	0,887	–0,104	0,269	–0,027
(Пересчет стандартных ошибок)	(0,197)	(0,038)	(0,213)	(0,025)

Источник: Table III в [Stock and Watson, 1993].

Рис. 10.4. График зависимости $m - p - y$ от R

Этот факт позволяет предположить, что послевоенные оценки страдают от проблемы сильной мультиколлинеарности: из-за схожих трендов в y и R в послевоенном периоде (показанном на рис. 10.2 и 10.3), γ_y и γ_R , оцененные по послевоенному периоду, сильно коррелированы. Хотя для каждого элемента послевоенная оценка (γ_y , γ_R) отличается от предвоенной, они могут не быть совместно значимо отличными. Эту возможность можно было бы легко проверить тестом Чоу на структурный сдвиг, по крайней мере если бы регрессоры являлись стационарными.

¹ В оригинальном графике в [Lucas, 1988] период выборки – 1900–1985 годы, а дата сдвига – 1957 год, но оба графика говорят, по существу, об одном и том же.

В тесте Чоу вы могли бы оценить

$$m_t - p_t = \mu + \gamma_y y_t + \gamma_R R_t + \delta_D D_t + \delta_{yD} D_t + \delta_{RD} D_t + z_t^*, \quad (10.5.5)$$

где D_t является дамми-переменной, значение которой равно 1, если $t \geq 1946$ и 0 в противном случае. Таким образом, константа и коэффициенты при y_t и R_t равны $(\mu, \gamma_y, \gamma_R)$ до 1945 года и равны $(\mu + \delta_D, \gamma_y + \delta_{yD}, \gamma_R + \delta_{RD})$ после 1945 года. Если регрессоры являются стационарными, то статистика Вальда для тестирования нулевой гипотезы о том, что $\delta_D = 0$, $\delta_{yD} = 0$, $\delta_{RD} = 0$ асимптотически распределена по $\chi^2(3)$, когда обе подвыборки растут с сохранением неизменным относительного объема (см. аналитическое упражнение 12 в главе 2). Теперь, когда регрессоры являются $I(1)$ -процессами и неинтегрированы, как в данном случае, в [Nansen, 1992b, Theorem 3(a)] показано, что статистика Вальда (правильным образом локализованная, если необходимо), имеет то же самое асимптотическое распределение (χ^2), если параметры коинтегрирующей регрессии оценены эффективным методом, таким как DOLS (который добавляет Δy_t , ΔR_t и их опережающие и запаздывающие значения в коинтегрирующую регрессию (10.5.5)), полностью модифицированный метод наименьших квадратов (FMOLS — *Fully Modified estimation*) или ML Йохансена¹. DOLS-оценки регрессии (10.5.5), дополненной двумя опережающими и двумя запаздывающими значениями Δy и ΔR , приведены в табл. 10.4. Незначимая статистика Вальда поддерживает точку зрения Лукаса о том, что спрос на деньги в Соединенных Штатах в XX веке был стабильным.

Таблица 10.4

Тест на структурный сдвиг

	γ_y	γ_R	δ_D	δ_{yD}	δ_{RD}	Статистика Вальда
DOLS-оценка	0.921	-0.060	1.36	0.52	0.048	5.95
III стандартные отклонения	(0.142)	(0.026)	(0.72)	(0.51)	(0.034)	(p-значение = 0.11)
близко						

Источник: Вычисления автора, которые нужно проверить в эмпирическом упражнении.

¹ Мы отметили в предыдущем параграфе, что статистика Вальда не распределена асимптотически как хи-квадрат даже после стандартизации, если гипотеза включает в себя ρ . Однако результаты в [Nansen, 1992b] показывают, что статистика Вальда для тестирования гипотезы о равенстве ρ асимптотически распределена как хи-квадрат [Nansen, 1992b] также распределена тест на структурный сдвиг к неизвестному сдвигу ρ сдвига.

Набор задач для главы 10

Эмпирические упражнения

(Прежде чем выполнять эти задания, просмотрите Stock and Watson, 1993, Section 7 и Appendix B.) В файле MPYR.ASC находятся следующие годовые данные:

- Столбец 1: натуральный логарифм $M1$ (упомянутый как $m1$).
- Столбец 2: натуральный логарифм дефлятора BIL1 (y).
- Столбец 3: натуральный логарифм ВВП (g).
- Столбец 4: ставка по кредитам в процентах к годовому исчислению (R).

Период выборки — с 1900 по 1989 год. Это те же самые данные, которые использовались в (Stock and Watson, 1993) в их исследовании парков на деньги США. См. их Appendix B в качестве источника данных.

Вопросы (а) и (б) заключаются в проверке предположения того, что ряд $m - y$ кointегрирован с y и R . Остальные вопросы связаны с оцениванием кointегрирующего вектора.

- (а) (Однократные тесты на единичный корень.) В (Stock and Watson, 1993, Appendix A) сообщается, что ADF $I(0)$ - и $I(1)$ -статистики, вычисленные с использованием 2 и 4 лагированных первых разностей, не в состоянии отвергнуть гипотезу единичного корня для каждого из рядов $m - y$ и R на 10%-ном уровне и что гипотеза единичного корня не отвергается для y с 4 лагами, но отвергается на 10%-ном (но на 5%-ном) уровне с 2 лагами. Проверьте эти выводы. Используйте F -тест для $H_0: \rho = 1$ для $m - y$ и y и используйте асимптотические критические значения (критические значения для $T = \infty$).
- (б) (Тесты, основанные на остатках.) Проведите тест Энгла — Грейнджера для проверки нулевой гипотезы о том, что ряд $m - y$ не кointегрирован с y и R . Поместите $y = 1$ (выбрано по BIC). Если так не включаете время в регрессию, то ADF t -статистики, полученные по остаткам, должны быть равной примерно -1.7 . Следует применить случай 2, рассмотренный в параграфе 10.3. Постою 5%-ное критическое значение должно быть равным -3.61 .)
- (в) (DOLS.) Воспроизведите SOLS и DOLS оценки кointегрирующего вектора, приведенные в табл. 10.2.

- (d) (Тест Чэу) Воспроизведите DOLS оценки, приведенные в табл. 10.4. Они основаны на следующей расширяющей коинтегрирующей регрессии:

$$\begin{aligned} m_t - p_t = & \mu + \gamma_0 u_t + \gamma_1 R_t + \delta_0 D_t + \alpha_0 p_t D_t + \delta_1 H_t D_t + \\ & - \beta_{20} \Delta p_t - \beta_{21} \Delta p_{t-1} + \beta_{22} \Delta p_{t-2} + \beta_{31} \Delta p_{t-1} + \beta_{32} \Delta p_{t-2} + \\ & - \beta_{40} \Delta R_t + \beta_{41} \Delta R_{t-1} + \beta_{42} \Delta R_{t-2} + \beta_{51} \Delta R_{t-1} + \beta_{52} \Delta R_{t-2} + v_t. \end{aligned}$$

Нулевая гипотеза состоит в том, что $\delta_0 = \delta_1 = \delta_2 = 0$. Положите ρ в (10.4.16) равным двум. Вычислите статистику Вальда для тестирования нулевой гипотезы об отсутствии структурного сдвига.

Литература

- Aoki, Masamitsu, 1968, "Control of Large Scale Dynamic Systems by Aggregation," *IEEE Transactions on Automatic Control/IEEE Transactions on Automatic Control*, AC-13, 240-253.
- Banerjee, A., J. Dolado, D. Hendry, and G. Smith, 1988, "Exploiting Equilibrium Relationships in Econometrics through Static Models: Some Monte Carlo Evidence," *Oxford Bulletin of Economics and Statistics*, 48, 253-277.
- Box, G., and G. Tiao, 1977, "A Canonical Analysis of Multiple Time Series," *Biometrika*, 64, 335-365.
- Davidson, J., D. Hendry, F. Smith, and S. Yeo, 1978, "Econometric Modeling of the Aggregate Time-Series Relationships between Consumers' Expenditure and Income in the United Kingdom," *Economic Journal*, 88, 661-692.
- Engle, R., and C. Granger, 1987, "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251-276.
- Engle, R., and S. Yeo, 1991, "Cointegrated Economic Time Series: An Overview with New Results," in R. Engle and C. Granger (eds.), *Long-Run Economic Relations: Readings in Cointegration*, New York: Oxford University Press.
- Friedman, M., 1958, "The Demand for Money: Some Theoretical and Empirical Results," *Journal of Political Economy*, 67, 327-351.
- Fuller, W., 1996, *Introduction to Statistical Time Series* (2d ed.), New York: Wiley.
- Fildes, R. S., and D. Stekler, 1990, "The Demand for Money," Chapter 5 in B. Friedman and F. Hahn (eds.), *Handbook of Monetary Economics*, Volume 1, New York: North-Holland.
- Granger, C., 1981, "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Economics*, 10, 121-130.
- Granger, C., and P. Newbold, 1974, "Spurious Regressions in Econometrics," *Journal of Economics*, 2, 111-120.
- Harrison, J., 1994, *Time Series Analysis*, Princeton: Princeton University Press.

- Hansen, B., 1992a, "Efficient Estimation and Testing of Cointegrating Vectors in the Presence of Deterministic Trends," *Journal of Econometrics*, 53, 81-121.
- , 1992b, "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics*, 10, 321-333.
- Haug, A., 1994b, "Tests for Cointegration: A Monte Carlo Comparison," *Journal of Econometrics*, 71, 89-115.
- Johansen, S., 1988, "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231-254.
- , 1995a, *Likelihood Based Inference in Cointegrated Vector Auto Regressive Models*, New York: Oxford University Press.
- Lucas, R., 1988, *Money Demand in the United States: A Quantitative Review*, *Carnegie-Rochester Conference Series on Public Policy*, Volume 29, 137-168.
- Lütkepohl, H., 1993, *Introduction to Multiple Time Series Analysis* (2d ed.), New York: Springer-Verlag.
- Maddala, G. S., and Im-Kee Kim, 1998, *Unit Roots, Cointegration, and Structural Change*, New York: Cambridge University Press.
- Ng, S., and P. Perron, 1995, "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncated Lag," *Journal of the American Statistical Association*, 90, 268-281.
- Park, J., and P. Phillips, 1998, "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory*, 14, 468-497.
- Phillips, P., 1986, "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 35, 331-240.
- , 1988, "Weak Convergence to the Matrix Stochastic Integral $R(1)0$ and $R(1)0$," *Journal of Mathematical Analysis*, 24, 252-264.
- , 1991, "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283-306.
- Phillips, P., and S. Perron, 1986, "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473-495.
- Phillips, P., and B. Hansen, 1990, "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies*, 57, 99-126.
- Phillips, P., and M. Loretan, 1991, "Estimating Long-Run Economic Equilibrium," *Review of Economic Studies*, 58, 407-436.
- Phillips, P., and S. Chelvaris, 1990, "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 58, 165-188.
- Phillips, P., and J. Park, 1988, "Asymptic Inefficiency of OLS and GLS in Regressions with Integrated Regressors," *Journal of the American Statistical Association*, 83, 111-115.
- Safkekanca, P., 1991, "Asymptotically Efficient Estimation of Cointegration Regressions," *Econometric Theory*, 7, 1-21.
- Sims, C., J. Stock, and M. Watson, 1990, "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58, 113-144.

- Stock, J. 1987. "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors." *Econometrica*, 55, 1035-1056.
- Stock, J., and M. Watson, 1988, "Testing for Common Trends." *Journal of the American Statistical Association*, 83, 1097-1107.
- . 1993, "A Single Estimator of Cointegrating Vectors in Higher Order Integrated Systems." *Econometrica*, 61, 783-820.
- Watson, M., 1994, "Vector Autoregressions and Cointegration," Chapter 47 in H. Engle, and D. McFadden (eds.), *Handbook of Econometrics*, Volume IV. New York: North Holland.

Приложение А. Блочные (клеточные) матрицы и произведения Кронекера

Блочные (клеточные) матрицы

Иногда удобно разбивать элементы матрицы на $M \times N$ подматриц как

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1N} \\ \vdots & \ddots & \vdots \\ A_{M1} & \cdots & A_{MN} \end{bmatrix}.$$

Матрицу в таком виде называют **блочной (клеточной) матрицей** (*partitioned matrix*). Построчные индексы подматриц определяются тем же самым способом, что и элементы матрицы. Например, мы можем записать

$$A = \begin{bmatrix} 1 & 1 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

где

$$A_{11} = \begin{bmatrix} 1 & 1 \\ 2 & 5 \end{bmatrix}, \quad A_{12} = \begin{bmatrix} 7 \\ 8 \end{bmatrix}, \quad A_{21} = \begin{bmatrix} 3 & 6 \end{bmatrix}, \quad A_{22} = 9.$$

Наиболее известный частный случай — когда $M = N$ и диагональные блоки (A_{ii} для $i \neq j$) являются скалярными матрицами:

$$A = \begin{bmatrix} A_{11} & & \\ & \ddots & \\ & & A_{NN} \end{bmatrix}.$$

Это называется **блочно-диагональной матрицей** (*block diagonal matrix*).

Сложение и умножение блочных (клеточных) матриц

Сложение и умножение матриц обобщается на случай блочных (клеточных) матриц. Таким образом,

$$A + B = \begin{bmatrix} A_{11} + B_{11} & \cdots & A_{1N} + B_{1N} \\ \vdots & \ddots & \vdots \\ A_{M1} + B_{M1} & \cdots & A_{MN} + B_{MN} \end{bmatrix}. \quad (A.1)$$

$$\begin{aligned}
 AB &= \begin{bmatrix} A_{11} & \cdots & A_{1N} \\ \vdots & & \vdots \\ A_{M1} & \cdots & A_{MN} \end{bmatrix} \begin{bmatrix} B_{11} & \cdots & B_{1L} \\ \vdots & & \vdots \\ B_{N1} & \cdots & B_{NL} \end{bmatrix} = \\
 &= \begin{bmatrix} \sum_{n=1}^N A_{1n}B_{n1} & \cdots & \sum_{n=1}^N A_{1n}B_{nL} \\ \vdots & & \vdots \\ \sum_{n=1}^N A_{Mn}B_{n1} & \cdots & \sum_{n=1}^N A_{Mn}B_{nL} \end{bmatrix}. \quad (A.2)
 \end{aligned}$$

Во всех этих выражениях матрицы должны иметь согласованные размеры для выполнения операций. Так, при сложении размеры A_{mn} и B_{nl} должны быть одинаковыми для всех $m (= 1, 2, \dots, M)$ и $n (= 1, 2, \dots, N)$. При умножении количество столбцов в A_{mn} должно быть равно количеству строк в B_{nl} для всех $m (= 1, 2, \dots, M)$, $n (= 1, 2, \dots, N)$, $l (= 1, 2, \dots, L)$. Частный случай умножения, когда B является **штабелированным вектором** (*stacked vector*) c :

$$Ac = \begin{bmatrix} A_{11} & \cdots & A_{1N} \\ \vdots & & \vdots \\ A_{M1} & \cdots & A_{MN} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_N \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^N A_{1n}c_n \\ \vdots \\ \sum_{n=1}^N A_{Mn}c_n \end{bmatrix} \quad (A.3)$$

Несколько часто встречающихся случаев имеют вид:

$$\begin{bmatrix} A_{11} & & \\ & \ddots & \\ & & A_{MM} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_M \end{bmatrix} = \begin{bmatrix} A_{11}c_1 \\ \vdots \\ A_{MM}c_M \end{bmatrix}. \quad (A.4)$$

$$\begin{aligned}
 \begin{bmatrix} A_{11} & & \\ & \ddots & \\ & & A_{MM} \end{bmatrix} \begin{bmatrix} B_{11} & \cdots & \\ \vdots & & \vdots \\ B_{M1} & \cdots & B_{MM} \end{bmatrix} = \\
 = \begin{bmatrix} A_{11}B_{11} & \cdots & \\ \vdots & & \vdots \\ A_{MM}B_{M1} & \cdots & A_{MM}B_{MM} \end{bmatrix}. \quad (A.5)
 \end{aligned}$$

$$\begin{bmatrix} A_{11} & & \\ & \ddots & \\ & & A'_{MM} \end{bmatrix} \begin{bmatrix} B_{11} & \cdots & B_{1M} \\ \vdots & & \vdots \\ B_{M1} & \cdots & B_{MM} \end{bmatrix} \begin{bmatrix} A_{11} & \cdots & \\ \vdots & & \vdots \\ A_{MM} & \cdots & \end{bmatrix} =$$

$$= \begin{bmatrix} A'_{11} B_{11} A_{11} & \cdots & A'_{1M} B_{1M} A_{1M} \\ \vdots & & \vdots \\ A'_{M1} B_{M1} A_{11} & \cdots & A'_{MM} B_{MM} A_{MM} \end{bmatrix}, \quad (\text{A.6})$$

$$\begin{aligned} & \begin{bmatrix} A'_{11} & & \\ & \ddots & \\ & & A'_{MM} \end{bmatrix} \begin{bmatrix} B_{11} & \cdots & B_{1M} \\ \vdots & & \vdots \\ B_{M1} & \cdots & B_{MM} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_M \end{bmatrix} = \\ & = \begin{bmatrix} A'_{11} B_{11} c_1 + \cdots + A'_{1M} B_{MM} c_M \\ \vdots \\ A'_{M1} B_{M1} c_1 + \cdots + A'_{MM} B_{MM} c_M \end{bmatrix}. \quad (\text{A.7}) \end{aligned}$$

$$\begin{aligned} A_1 \cdots A_{MM} \begin{bmatrix} B_{11} & \cdots & B_{1M} \\ \vdots & & \vdots \\ B_{M1} & \cdots & B_{MM} \end{bmatrix} \begin{bmatrix} A_{11} \\ \vdots \\ A_{MM} \end{bmatrix} \\ = \sum_{m=1}^M \sum_{n=1}^M A'_{mm} B_{1n} A_{1n}. \quad (\text{A.8}) \end{aligned}$$

где c_m ($m = 1, 2, \dots, M$) являются векторными столбцами.

Обращение блочных (клеточных) матриц

Матрица, обратная к блочно-диагональной матрице, имеет вид:

$$\begin{bmatrix} A_1 & & \\ & \ddots & \\ & & A_{MM} \end{bmatrix}^{-1} = \begin{bmatrix} A_1^{-1} & & \\ & \ddots & \\ & & A_{MM}^{-1} \end{bmatrix}. \quad (\text{A.9})$$

при условии, что A_{mm} ($m = 1, 2, \dots, M$) обратимы. Это можно проверить непосредственным умножением:

Для блочной (клеточной) матрицы размерности 2×2 этого вида блочная обратная (partitioned inverse) матрица имеет вид:

$$\begin{bmatrix} A_1 & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_1^{-1} - A_1^{-1} A_{12} F_2 A_{21} A_1^{-1} & -A_{12} A_{22} F_2 \\ -F_2 A_{21} A_1^{-1} & F_2 \end{bmatrix}, \quad (\text{A.10})$$

где $F_2 = (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1}$. Это можно легко проверить, умножив A на обратную ей матрицу. Благодаря симметрии вычисления верхний левый блок можно записать также как:

$$F_2 = (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1}.$$

Произведения Кронекера

Для матриц $A (M \times N)$ и $B (K \times L)$ удобно задать **произведение Кронекера** (*Kronecker product*) следующим образом:

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1N}B \\ \vdots & & \vdots \\ a_{M1}B & \cdots & a_{ML}B \end{bmatrix}. \quad (\text{A.11})$$

Это матрица размера $ML \times NL$. В частном случае, когда A и B являются векторами:

$$a \otimes b = \begin{bmatrix} a_1 \\ \vdots \\ a_M \end{bmatrix} \otimes \begin{bmatrix} b_1 \\ \vdots \\ b_L \end{bmatrix} = \begin{bmatrix} a_1 b_1 \\ \vdots \\ a_M b_L \end{bmatrix}. \quad (\text{A.12})$$

Очень просто, но трюмально проверить, что

$$(A \otimes B)(C \otimes D) = AC \otimes BD, \quad (\text{A.13})$$

при условии, что A и C имеют согласованные размеры, и B и D имеют согласованные размеры, и что

$$(A \otimes B)^T = A^T \otimes B^T, \quad (\text{A.14})$$

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}, \quad (\text{A.15})$$

при условии обратимости A и B .

Предметный указатель

А

Абсолютно суммируемые коэффициенты:

- $AR(p)$ как $MA(\infty)$ с 409–410
- $MA(\infty)$ с 397–398

Абсолютно суммируемые обращения лапласа полиномов 404

Абсолютно суммируемый фильтр 400

Автоковариационная матрица 127

Автоковариация 127

Автоковариация порядка j 127

Автокорреляции коэффициент 127

Автокорреляции коэффициент порядка j 127

Авторегрессия модель первого порядка 421

Авторегрессия модель первого порядка:

- автоковариации 408
- значение 31
- оценивание 423
- - методом максимального правдоподобия 579–583
- распределение оценки наименьших квадратов для конечных выборок 612

Авторегрессия процесс первого порядка 405–408

Авторегрессия процесс порядка p 405–408

Авторегрессия: оценивание 429

Авторегрессия: оценивание:
векторных авторегрессий 427–428

- границы аппроксимаций для оценивания 424–425
- модели $AR(1)$ 421–423
- модели $ARMA(p, q)$ 428–429
- последовательная процедура от общего к частному, использующая F -тесты 424–425

Анализ примененных рядов:

- квадратные скобки 124
- дискретизацию стационарные процессы 126–127
- мартингалы 129–130
- необходимость предположений стационарности 124–125
- последовательность мартингал-разностей 131–132
- процессы белого шума 128
- процессы с авторегрессионной гетеросkedастичностью 126–127
- случайные блуждания 130
- строгая эргодичность 31–32

- фиксированные моменты сериальной зависимости 133–134

- эргодическая теорема 128–129

Анализ методом Монте-Карло 107

Аннулятор 40

Асимптотики:

- двухшагового метода наименьших квадратов 257
- для выборочных средних процессов с сериальной корреляцией 430–431
- локальные к единице 637
- мощность состоятельного теста 148–149
- оценки обобщенности метода моментов 244
- оценка максимального правдоподобия с полной информацией 568
- с фиксированными регрессорами 194

Асимптотическая дисперсия в теории для больших выборок 121

Асимптотическая мощность 148

Асимптотическая нормальность:

- M -оценок 502–503
- в линейной регрессионной модели 497–498
- выражение для ошибки оценки 515–518
- многомерных регрессий 556
- моделей QR 542
- обобщенного метода моментов 511–514
- обобщенного метода моментов для одного уравнения 232–233
- обобщенного метода моментов для нескольких уравнений 294
- оценки максимального правдоподобия для пробит-модели 498–499
- оценка максимального правдоподобия с полной информацией 567
- состоятельного оценивания асимптотический дисперсии 505
- сравнение с обобщенным методом моментов 514–515
- усеченных регрессионных моделей 547–550
- условного максимального правдоподобия 506–511
- экстремальных оценок 501–519

Асимптотическая теория 114

см. также асимптотическая диспер-

сия в теории для больших выборок
Асимптотически более эффективная 165

Асимптотически нормальные оценки 121

Асимптотическое распределение:

- оценки обобщенного метода моментов 236–239
- оценок наименьших квадратов 141, 143, 190–191
- Фишера-модели 558

Б

Байесовский информационный критерий BIC 425–427

Бейерштедт - Польсен разложение BN 650–661, 665–666

Белого шума процесс 128, 394, 416

Белого шума процесс векторный 416

Ближарной отклонка модель 539

Блочная матрица

- обращение 710–711
- описание 708
- скалярное и умножение 708–710

Бокс - Льюис Q-статистика 171, 173, 202–203

Бройде - Годфрей тест 178, 202, 206

Бруновского движения процессы 600

В

Вальда критерий 65

Великая статистика 522–524, 527

Вектор вклада для наблюдения t 502

Вектор данных 28

Векторной авторегрессии коэффициент оценивания 427–428

Векторной авторегрессии процесс VAR:

- оценивание методом условного максимального правдоподобия 583–584
- порядка p 414

Векторные процессы 416

Векторный процесс ARMA(p, q) 419–420

Векторный процесс скользящего среднего бесконечного порядка 417

Включенная регрессия 82

Взвешенного метода наименьших квадратов оценки:

- с известным α 163–164
- с оцененным α 164–165
- сравнение с обычными оценками наименьших квадратов 165

Взвешенный метод наименьших квадратов 81

Взвешенный метод наименьших квадратов или взвешенная регрессия 82

Взвешивающая матрица 236

Винеровский процесс 600

Влиятельность вклада 45

Влиятельные наблюдения 43–46

«Внутри» оценки 360

Внутреннее условие:

- для функций правдоподобия линейной регрессии 493
- для функции правдоподобия пробит-модели 492
- после репараметризации 492

Безразличие к среднему 506

Временной тренд:

- инкорпорирование в расширенный тест Дики - Фуллера 633–635
- описание 190

Временные ряды:

- графики серийно некоррелированных 164
- описание 124

Выборка 25

Выборка Власкина - Пыомарка 280

Выборочная автокорреляция j -го порядка 171

Выборочное среднее 122

Выборочный коэффициент корреляции i -го порядка 127

Выражение для ошибки оценки 237, 368

Г

Гаусса - Маркова теорема 48, 49, 60

Гаусса - Неймана алгоритм 530

Гауссовская модель векторной авторегрессии 583

Гауссовский процесс авторегрессии первого порядка 579

Гауссовский процесс авторегрессии порядка p 583

Гессинг для наблюдений t :

- для M -оценок 502
- для моделей с качественными откликами 549–551
- для TDD-моделей 613

Гиперсострательные оценки 192

Гипотеза эффективности рынка 419, см также форвардная премия рыночного равновесия модели

Гибкая спецификация:

- для оценивания авторегрессий 424–425
- использование информационных критериев Акаике, Шварца (Байесовского) 425–427

Гомоскедастичность безусловная:

- описание 14
- сравнение с условной 146

Гомоскедастичность:

- варианты асимптотических тестов при условной гомоскедастичности 157–158
- обобщенный метод моментов для нелинейных уравнений

- и условная гомоскедастичность 303–304

- обобщенный метод моментов и условная гомоскедастичность 265-264
 - определение условной и безусловной гомоскедастичности 34-35
 - оценивание с параметризованной условной гомоскедастичностью 166
 - оценивание с параметризованной условной гомоскедастичностью 161
 - оценки фиксированных эффектов и ослабление требований условной гомоскедастичности 353-366
 - распределение на больших выборках и условная гомоскедастичность 156-157
 - редукция к формулам для конюгированных выборок 153-156
 - тестирование условной гомоскедастичности 159-161
 - Уайта тест на условную гомоскедастичность 160-161
 - условная и безусловная 156
- Горизонтальная
- для эргодических стационарных процессов 432-433
 - ограничение степени серийной корреляции 436-437

II

- Двухмерная и две 94
- Данные 25
- Дарбина-Уотсона статистика 68
- Дарбина-Уотсона статистика 68
- Двухмерная $U(0)$ -система 663
- Двухмерная $U(1)$ -система 664
- Двухмерная система 692
- Двухсторонний фактор 696
- Двухсторонний тест 69
- Двухшаговый метод наименьших квадратов
- асимптотические свойства 259
 - асимптотические свойства 257-258
 - из асимптотического GMM 256-257
 - как две регрессии 260-261
 - как оценка инструментальных переменных 260
 - коррекция смещения с формулой 271-272
 - критерии 221-223
 - свойства на конечных выборках 258-259
 - сравнение с оценкой максимального правдоподобия в ограниченной модели 575-576

«Дельта-метод» 126

Детерминированный тренд 592

Джир Фуллера t -статистика в расширенной модели, критические значения 683

Дикс — Фуллера тесты:

- интегрированное время-тренд 610-619
- модель авторегрессии первого порядка 608-619
- обзор 635-636
- расширенные 621-636
- случайных блужданий без сноса 611-615

Добавление переменных тест 253-255

Докристинный интервал 61

Докристинная дисперсия 431

Докристинная ковариационная матрица S

- когда автокорреляция коррелируется и оценивается 437-438
- представление через матрицу данных 444-445
- процедура VARMAX для оценивания 440-442
- условная гомоскедастичность и оценивание 443-447
- ядра для оценивания 438-440

Доллар/Йена

- график $\Delta 30$ против фьючерсной ставки 455
- корреляция 452
- ошибка прогноза 451
- спотовая ставка 1973-80, 454

Доллар/стерлинг: реальные обменные курсы 641

Доминирование слабое 190

Доступный обобщенный метод наименьших квадратов 83

Дэвидсона — Мак-Кинлэна коррекция 209

III

Длинный корень $U(1)$ 634-635, см. также длинный

IV

Зависимая переменная 55

Зависимая переменная: латентная $U(1)$

Зависимых переменных: слабый 122

Зависимых переменных: усиленный 122

Зависимых переменных: 639-640

Зависимых переменных: математическая модель 30, 196, 197

Зависимых переменных:

- Чоубовца 122
- второй Колмогорова 122
- для ковариационно-стационарных процессов 430-431
- описание 122
- усиленные и слабые 122

Запаздываний глубина:
для оценивания авторегрессий
424–425

- использование информационных критериев Акаике, Паркса (байесовского) 425–427

Запаздывания оператор 399

Заработной платы уравнение:
рифическое 267

Заработной платы уравнение:

- IQ как мера стабильности 268
- модель компонент ошибки 355
- обзор 561
- обнаружение влияния длительности пребывания в должности 33
- описание 26–27 217
- полулогарифическое 267
- система уравнений 288–289
- смещение из-за пропуска переменной 267 268
- тестирование предположений о стабильности образования 250–251
- упрощенная версия 224 230

И

Идентифицируемость 232

Иена/доллар:

- график 230 против фиксированной ставки 455
- корреляция 452
- оценка прогноза 451
- спотовая ставка 454

Инвариантности принцип 603

Инвариантность:

- максимального правдоподобия 483–484
- элемент максимального правдоподобия с полной информацией 568–569

Индивидуальный эффект: неоднородность индивидов или фиксированный эффект 374–380

Инструмент релевантный 221

Инструмент: свой эффект на сам ин-
струмент:

- описание 221
- требования предположений о стабильности 231

Инструментальные переменные оценки 221 235

Инфляция

- гипотеза эффекта на а-рыча 131–132, 185–187
- реальная процентная ставка и инфляция 183
- равнение теста для больших выборок и теории для конечных выборок 187 189

Информационная матрица 74

- для наблюдения 506

Информационное множество 120

Информационное неравенство Кулиба-
ка — Лейблера 494

Искажение размера 146

Исключения страхи 564

Й

Йенген неравенство 534

Йохансен процедура максимального
правдоподобия 677 679

К

Каждые неслучайные регрессии,
итерационная процедура 570

Каждые неслучайные регрессии:

- асимптотические свойства 310
- как обобщенный метод моментов для одного уравнения 18
- как обобщенный метод моментов для нескольких уравнений 309–307

как специализация оценки
максимального правдоподобия
с полной информацией 569 570

сравнение с обобщенным мето-
дом моментов для нескольких
уравнений и методом наимень-
ших квадратов 310–312

Каналы США — веб-сайт 206

Квадратично суммирующая 459 460

Квази- (или псевдо-) максимальное
правдоподобие 76

Классическая линейная модель регрес-
сии:

- Крамера — Рао граница 73–75
- для случайных выборок 34
- матричная запись 27–28
- обобщенная 78
- описание 25, 660
- предположение линейности 25 27
- предположение строгой экзогенности 28 32
- статистика отношения правдопо-
добий 253

функция правдоподобия 71

Колье — Дуглас технология:

- идентификация 88
- параметризованная функция ла-
гранжа, выведенная на ее осно-
ве 85 88
- транзитивная функция ла-
гранжа как обобщение 94–95

Ковариационно-стационарные процессы
зависимости 450–451

- процесс белого шума 125, 384–385

фильтрация 388–389

Контрастная рента:

- описание 667
- определение 680 687

Контрастирующая сила 667

Контрастирующее пространство 667

Контраграция в векторной авторегрессии конъюгированного порядка 673–675

Контрагированные системы:

- альтернативных представлений 571–679
- Безаиджа — Нельсона разложения и контрагированные системы 600–601
- векторные авторегрессионные контрарегрессионные системы 673–675
- Иоханссона — Гиндеса канонического правдоподобия 677–679
- и u и v как контрагированная система 595–700
- линейные векторные $I(0)$ и $I(1)$ -процессы 602–603
- общий случай и жорновой канонизированной системы 695–697
- определены 500–670
- основание динамическим методом наименьших квадратов 689, 690, 693, 695, 696, 700, 701, 703
- система SOLS 688–689, 700
- представление в форме канонической модели координатных ошибок 675–677
- приложение к спросу на деньги в США 697–700
- проверка нулевой гипотезы об отсутствии контрагирования 687
- свойства оценки в конечных выборках 697
- статистические выводы о векторах 688
- тест, основанный на остатках 681–687
- треугольное представление Финдеса 671–673

Контрагированная регрессия 672

Контрагированный вектор 667

Модифицированный закон больших чисел 122

Концентрированная логарифмическая функция правдоподобия 558

Концентрированная логарифмическая функция правдоподобия:

- для линейной модели регрессии 55
- для метода максимального правдоподобия с полной информацией 559
- для многомерной модели регрессии 557

Коррекция степеней свободы 154

Коррелограмма 127

Корреляция между уравнениями, или одновременная корреляция 288

Косинус последовательности 404

Косинус — Шварца неравенство 198

Коэффициент автокорреляции k -го порядка 127

Коэффициент автокорреляции 42–43

Коэффициент детерминации детерминированный 42–43

Коэффициент проксимии метода наименьших квадратов 168

Коэффициенты проксимии метода наименьших квадратов 168

Критерия — Рэн:

- граница 73–75
- аксиомы границы 70

Критерий выбора теста на линейный характер:

- аксиомы локальные к единице 637
- свойства на малых выборках 638–639

Критерий выбора 196

Критерий произведения 711

Л

Линейная модель:

- абсолютно суммируемые образцы 404
- образцы 402–404

Лагранжевы полиномы степени μ 359

Лагранжевы и другие статистики 425–426

Линдберга — Левы центральная предельная теорема 134

Линдберга — Феллера центральная предельная теорема 194

Линейная по переменным 532

Линейная регрессионная модель:

- асимптотическая нормальность 509–510
- идентификация в методе максимального правдоподобия 497–498
- с нормальными ошибками 482

Линейный гипотезы проверка 145–146

Линейные векторные $I(0)$ и $I(1)$ процессы 602–604

Линейные оценки обобщенного метода моментов 485–486

Линейные процессы:

- линейные $I(0)$ -процессы 599–600
- описание 394–395
- скользящего среднего бесконечного порядка 395–398
- скользящего среднего порядка p 396

Линейный процесс скользящего среднего бесконечного порядка 395–398, 409–410, 431

Линейный процесс скользящего среднего порядка p 395

Логарифм дохода минус логарифм потребления 656

Логарифмическая функция правдоподобия 71, 72

Логарифмическое условное правдоподобие 481

Логит-модель 540
 Логлинейное уравнение 87
 Ложная регрессия 680–681
 Ложное детрендривание 605
 Локальные к единице асимптотики 638
 Лонгитюдные данные 352
 Лунга — Бокса Q -статистика 171–173

М

М-оценки:

- асимптотическая нормальность 502–506
- как экстремальная оценка 477–478
- описание 477
- состоятельность 490–492

Максимального правдоподобия принцип:

- для оценивания параметров модели 69
- основная идея 70
- связь с обобщенным методом моментов 18

Мартингал по отношению к $\{z_t\}$ 129

Мартингал-разность 131

Мартингалы 129–130

Математические обозначения 21

Матрица данных 28

Матричные обозначения 27–28

Мера риска для стандартного нормального распределения 310

Метод

- минимального расстояния 236
- моментов 235
- наименьших квадратов с ограничениями 66

Модели с качественным откликом:

- асимптотическая нормальность 542
- вклад и тесса для наблюдения i 540–541
- описание 539–540
- состоятельность 541

Модели:

- определение 25
- правильно специфицированные 478

Модель компонент ошибок:

- индивидуальный эффект 354
- описание 352, 353
- отклонения от групповых средних и модель компонент ошибок 356
- оценки фиксированных эффектов и модель компонент ошибок 358–370
- репараметризация модели 357–358

Модель коррекции ошибок векторная 671

Модель с ограниченной зависимой переменной 543

Модель с множественным выбором 539

Мощность критерия 147

Н

Наблюдаемый уровень значимости, p -значение 61

Наилучшая линейная несмещенная оценка 49

Наилучшая несмещенная оценка 75

Неоднородность индивидов 354

Неравенство Крамера — Рао 73–74

Неравенство треугольника 158

Номинальный размер теста 146

О

Обменные курсы 616,

см. также реальный обменный курс

Обнуление:

- пропущенные наблюдения 367
- против сжатия 368–369

Обобщенная нелинейная оценка инструментальных переменных 485

Обобщенного метода моментов для одного уравнения оценки:

- асимптотическая мощность 243–244
- асимптотическая теория 238–258
- асимптотические свойства 238–244
- асимптотическое распределение 238–239
- в случае преопределенных регрессоров 261
- для конечных выборок 244
- определение 236
- оценивание дисперсии ошибки 239–240
- оценивание долговременной ковариационной матрицы 241
- получение двухшаговой оценки наименьших квадратов из эффективного обобщенного метода моментов для одного уравнения 256–261
- последствия условной гомоскедастичности 255–264
- тестирование гипотез 240–245
- трехшаговый метод наименьших квадратов 18
- эффективная 242–243
- эффективная с ограничением 251

Обобщенного метода моментов для нескольких уравнений оценки:

- асимптотическая теория 296–299
- выпод 294–296
- инструментальная оценка с полной информацией 304–305
- кажущиеся несвязанными регрессии 308–310
- как оценки случайных эффектов 318–322

- общие коэффициенты 314-323
- объединенная OLS-оценка 319-320 322
- оценка трехшагового метода наименьших квадратов 305-308
- преимущества 287
- сравнение SUR и OLS 310-312
- сравнение со случаем одного уравнения 299 303
- условная гомоскедастичность 303-304
- условная гомоскедастичность эффективной оценки 317-320
- Обобщенного метода моментов для одного уравнения оценки:
 - сравнение с обобщенным методом моментов для нескольких уравнений 299 303
- Обобщенный метод моментов:
 - инкорпорирование сериальной корреляции 436-442
 - инкорпорирование сериальной корреляции и связь с обобщенным методом наименьших квадратов 445-447
 - общая формулировка 228-233
 - описание 234-237
 - рисунки 18
 - процедуры оценивания 18
 - связь между методом максимального правдоподобия и обобщенным методом моментов 18
 - смещение вследствие эндогенности 218-223
 - см. также обобщенного метода моментов для нескольких уравнений оценки
- Обобщенный метод наименьших квадратов:
 - инкорпорирование сериальной корреляции в обобщенный метод моментов и связь с обобщенным методом наименьших квадратов 445 447
 - ограничивающее свойство обобщенного метода наименьших квадратов 82-83
 - описание 78
 - последствия ослабления предположения 79
 - свойства на конечных выборках 48-53
 - эффективное оценивание с известной матрицей V 79-81
- Образование:
 - отдача от образования 265 273
 - смещение из-за способностей 271-272
 - тестирование предопределенности в уравнении заработной платы 250
 - уравнение заработной платы 217
 - эндогенность 272
- Обратимости условие 412
- Обратное отношение Миллса 510, 545
- Обращение блочных матриц 710 711
- Обращения:
 - лаговых полиномов 402-404
 - фильтра 401 402
 - фильтров для векторных процессов 418
- Общие коэффициенты:
 - как неограничивающее условие в обобщенном методе моментов для нескольких уравнений 322-323
 - обобщенный метод моментов для нескольких уравнений и общие коэффициенты 314-323
- Объединенная оценка наименьших квадратов 319-320
- Объясняющие переменные 25
- Обычный метод наименьших квадратов 18
- Односторонний t -тест 69
- Ортогональная 30
- Отклонения
 - от групповых средних 356
 - от среднего 169
- Отношения правдоподобия принцип:
 - F -отношение 66
 - тестирование гипотез 251-255
- Отношения правдоподобия статистика 526 527
- Отношения правдоподобия тест 76
- Оценивание методом максимального правдоподобия:
 - асимптотическая нормальность условного 506-511
 - идентификация в линейной модели регрессии 497-498
 - идентификация условной плотности 494-495
 - инвариантность 483 484
 - Йохансена 677 679
 - как M -оценка 477
 - максимальное правдоподобие с ограниченной информацией 542, 571
 - максимальное правдоподобие с полной информацией 559-570
 - многомерные регрессии 539-542
 - модели с качественным откликом 539-542
 - процессов $AR(p)$ и $VAR(p)$ 583-584
 - процессов $AR(1)$ 579-583
 - сериально коррелированные наблюдения 576-584
 - состоятельность 495-497
 - сравнение с асимптотической нормальностью обобщенного метода моментов 514-515

- тобит (цензурированная регрессия) модели 550-553
 - усеченные модели регрессии 543-550
 - условное 481-483
 - Оценивание спроса на деньги 697-703
 - Оценивания процедура 24
 - Оценка динамического метода наименьших квадратов 689, 690, 693, 695, 696, 700-701, 703
 - Оценка k -класса 575
 - Оценка ковариационная 360
 - Оценка максимального правдоподобия точная 579
 - Оценка максимального правдоподобия, условная относительно y_0 481
 - Оценка наименьших квадратов статистическая 689-700
 - Оценки максимального правдоподобия с ограниченной информацией:
 - вычисление 573-575
 - описание 571-573
 - сравнение с двухэтаговым методом наименьших квадратов 575
 - Оценки максимального правдоподобия с полной информацией:
 - исходная информация 559-563
 - оценивание методом максимального правдоподобия 569-570
 - полная система одновременных уравнений 562-563
 - свойства 567-569
 - связи между коэффициентами 563-564
 - тестирование сверхидентифицирующих ограничений 566-567
 - функция правдоподобия 564-565
 - Оценки минимального расстояния классические 478
 - Оценки с ограничениями 321
 - Ошибки в переменных 224-226
- П**
- Панель несбалансированная:
 - асимптотические свойства оценки фиксированных эффектов 369-370
 - «обнуление» пропущенных наблюдений 367-368
 - «обнуление» против сжатия 368-369
 - описание 367
 - смещение из-за отбора 367, 369
 - Панельное исследование динамики доходов 225
 - Панельные данные:
 - модель компонент ошибки 359-370
 - несбалансированные 367-370
 - описание 279, 289, 352
 - пространство параметров 476
 - Параметризация функции плотности 70
 - Паритет покупательной способности 639-640
 - Перекрестные ограничения 327
 - Переменная в левой части 25
 - Переменные в правой части 25
 - Перманентная компонента 597
 - Перманентного дохода гипотеза 224-226
 - Питменовская последовательность 149
 - Питменовский снос 149
 - Поддерживаемая гипотеза 55
 - Подобранное значение 40
 - Подход
 - в частотной области 414
 - во временной области 414
 - Покрытый процентный паритет, уравнение 449
 - Полный столбцовый ранг 32
 - Пониженного ранга условие 675
 - Последовательное t -правило от общего к частному 424-425
 - Последовательности случайных величин:
 - взгляд на оценки как на последовательности случайных величин 121
 - обзор предельных теорем 115-121
 - Последовательность локальных альтернатив 148
 - Постоянные реальные процентные ставки 181
 - Правдоподобие безусловное 70-71
 - Правильно специфицированная 478
 - Правильно специфицированная модель 302
 - Предел по вероятности 115
 - Предельная склонность к потреблению 225-226
 - Предельное распределение 117
 - Предопределенные регрессоры 137, 229, 261
 - Представление в виде скользящего среднего 405-408
 - Представление в форме векторного скользящего среднего 662-664
 - Представление в форме коррекции ошибок 675
 - Преходящая компонента 596
 - Приведенная форма 563
 - Приведенной формы коэффициенты 563
 - Пробит-модель:
 - асимптотическая нормальность оценки максимального правдоподобия 510-511
 - идентификация при оценивании методом максимального правдоподобия 498-499
 - условное максимальное правдоподобие 482-483
 - Прогноза ошибка, иена/доллар 451
 - Проекционная матрица P 40-41

Проекция методом наименьших квадратов:

- наилучший линейный прогноз, использующий проекцию методов наименьших квадратов 167-169
- описание 166-167

Производящая функция автокорреляций 413-415

Пронтегрированный 3 раз процесс авторегрессии — скользящего среднего 602

Пронтегрированный процесс авторегрессии — скользящего среднего 502

Полная регрессионная модель 26

Плоходействие параметров 476

Процентные ставки:

- инфляция и процентные ставки 183
- номинальные 187-188
- реальные и процентные ставки 188

Процесс авторегрессии порядка p 353-364

Процесс независимого белого шума 128

Процесс порожждения данных 137-139

Процесс с единичным корнем 304

Процессы авторегрессии - скользящего среднего:

- $ARIMA(p,q)$ 410-411, 426-429
- автокорреляции $AR(1)$ 405-412
- как инструмент эконометрических единичных корней 601-602
- обратимость 412
- описание 405
- производящая функция автокорреляций и спектр 413-415
- процесс авторегрессии первого порядка 405-408

Процессы с авторегрессионной гетероскедастичностью 132

Процессы с авторегрессионной гетероскедастичностью 133

Р

Равенство для информационной матрицы 73

Равномерная сходимость по вероятности 488

Равномерный закон больших чисел 490

Разложение в ряд значений 500

Расширенная авторегрессия 621-632

Расширенная коинтегрирующая регрессия:

- дополнение сепаратной корреляции 622-625
- пример с двумя переменными 630-632

Расширенные тесты Дики — Фуллера

- без постоянной составляющей 626-627

- включение в регрессию постоянной составляющей 630-632

- включение временного тренда 633-635

- вывод статистик тестов 625-627

- для квадратичных обязательств 632

- правильное распределение остатков наименьших квадратов 622-623

- предложение для выбора $AR(1)$ 629-630

- при неизвестном значении p 627-635

- расширенная авторегрессия 621-622

- с постоянной составляющей 630-632

Реализация или выборочная траектория 124

Реальный обменный курс 641.

см. также Обменные курсы

Регрессионные коэффициенты 25

Регрессии на время:

- оценивание методом наименьших квадратов 192
- предположения 190
- тестирование гипотез 192-193

Регрессионная функция 25

Регрессируемая переменная 23

Регрессия 25

Регрессии с ограничениями 66

Регрессоры:

- описание 25
- предопределенные 261, 229
- фиксированные 34, 194
- латентные 218-219, 229

Результативность:

- мощность теста 492
- максимизация правдоподобия 483-484
- тождество 550
- оценочные модели регрессии 547

Робастная стандартная ошибка 146

Робастное тестирование 146-147

Рост вложений в оборудование и инвестиции в Восточной Азии

С

Саргана статистика 367-368

Саргана — Бхаргавы статистика 360

Сбалансированные панели 367

Свертка 400

Сверхидентифицировано 39, 232

Сдвиг предположения

- скользящее из-за эндогенности 223-227

Уоркинга пример 224

Серьезная корреляция

- Плотная условные ограничивающие степени корреляции 430-437
- асимптотика для выборочных средних 430-435
- векторные процессы 305-447
- инкогнорирование и объединенный метод моментов 435-442
- инкогнорирование и объединенный метод моментов и связь с объединенным методом наименьших квадратов 442-447
- линейные процессы 394-404
- оценивание авторегрессий 421-429
- оценивание долговременной ковариационной матрицы 437-447
- процессы авторегрессии
- скользящего среднего 405-415
- рекуррентная коинтегрирующая регрессия и 692-693
- форвардная премия 448-455
- Серийный корреляционный тестирование:
 - Бокса — Пирса и Льюнга — Бокса 171-173
 - Брайда — Макфрей 178
 - выборочные автокорреляции, выполняемые на остатках 173-175
 - описание 170
 - основанное на вспомогательной регрессии 176-178
 - с предполагаемыми регрессионными 174-176
- Скользящего среднего процесс безкомнатного порядка 397
- Скользящего среднего процесс второго порядка 391-392
- Скользящего среднего процесс порядка k 395
- Скорость конвергенции 371
- Слабо (или коварационно) стационарный случайный процесс 126
- Случайного теорема 115
- Случайные блуждания:
 - аппроксимация $I(1)$ процесса $I(0)$ 600-601
 - жизнеспособность 610-611
 - как пример мартинала 560
 - обменные курсы как случайные блуждания 615-616
 - описание 130
 - со сносом 407
 - тест Дики — Фуллера на отсутствие сноса 608-610
- Случайные выборки 34
- Случайный эффект метод:
 - как объединять данные моментов для нескольких уравнений 315
 - как частный случай обобщенного метода моментов 15
 - сравнение случайных в фиксированных эффектах 363-365
- Смещение вelasticity одновременно 220
- Смещение из-за отбора 367-369
- Смещение из-за способности 271-273
- Смещение из-за эндогенности:
 - приближенные оценки предсказания 220-223
 - ошибки 218-220
 - ошибки в персональных данных пример 224-226
 - пример с количественными переменными 224
 - производственная функция как пример 226-227
- Смещение, обусловленное селективностью выборки 545
- Смещение при оценивании 465
- Советы управляющих федеральной резервной системы США веб-сайт 206
- Сопоставляющая форма 1411-1461
- Состоятельная при гетероскедастичности и автокоррелированности 437
- Состоятельная при гетероскедастичности и автокоррелированности оценка выборочной авторегрессии 440-442
- Состоятельная при гетероскедастичности стандартная ошибка 145-146
- Система равенств для экстремальных случаев:
 - для M -оценок 400-402
 - теорема 487-489
- Состоятельные оценки 121
- Состоятельные тесты 147-148
- Спектральная плотность (функция) 414
- Спецификационный тест 247
- Спрос и предложение модели:
 - наблюдаемые уравни предсказания 220-223
 - смещение вelasticity одновременно 218-219
- Среднее из выборок 124
- Среднеквадратичная ошибка 166
- Стабильности условия 403
- Ставки по ипотечным обязательствам, разнородный тест Дики-Фуллера 611-612
- Стандартная ошибка:
 - оценки наименьших квадратов 34
 - ошибки (SER) 41
 - уравнения (SIC) 41
- Стандартный Бакерский процесс:
 - экстраполированный 605
 - лентрированный 607
- Статистический эффект из M -системы на финансов 206
- Стационарные условия 406
- Стационарность 126
- Стационарные относительные цены:
 - процесс и 521-527

- Стационарный в разностях 126
- Стационарный в разностях процесс 594
- Стационарный относительно тренда 126, 190
- Степени свободы 151
- Стерлинг/доллар реальные обменные курсы 641
- Стохастическая коинтеграция 668
- Стохастические процессы:
 - ковариация 126-127
 - описание 124
 - различные классы 125-128
 - см. также последовательность случайных величин
- Стохастический тренд 592
- Строгая мультиколлинеарность 32
- Строгая экзогенность 28-32
- Структурная форма 562
- Структурной формы параметры 562
- Сумма квадратов
 - остатков 37
 - остатков (SSR) 36
 - остаточная (RSS) 36
 - ошибок (ESS) 36
- Суперсостоятельные оценки 192, 610
- Сходимость
 - в среднеквадратичном 116
 - по вероятности точечная 487
 - по распределению 116
 - почти наверное 115
- Сходится по вероятности:
 - асимптотически эквивалентные последовательности 119
 - описание 115
 - почти наверное 115
 - Случайного теорема 119
 - сохранение для непрерывных преобразований 118
- США:
 - единичный корень в ВВП 634
 - индексированные на инфляцию облигации: проданные с аукционом 1997 г. 188
 - реальные обменные курсы стерлинг/доллар 641
 - спрос на деньги 697-703
 - электроэнергетическая отрасль в США 84-91, 334
 - см. также доллар/цена

Т

- Тейлора разложение 515-518, 521-522
- Телескопическая сумма 433-434
- Теорема Гаусса — Маркова, 48, 49, 80
- Теорема Грейнджера о представлении 677
- Теорема о среднем значении 501, 503, 511
- Теория для больших выборок:
 - для обобщенного метода моментов для одного уравнения 238-258

- инфляция и теория для конечных выборок 187-189
- описание 114
- оценивание с параметризованной условной гетероскедастичностью 161-166
- последствия условной гомоскедастичности 154-158
- предельные теоремы для последовательностей случайных величин 115-124
- проекция методом наименьших квадратов 166-167
- распределение оценки наименьших квадратов 136-144
- регрессии на время 190-197
- состоятельное оценивание 151-153
- тестирование
 - - линейных гипотез 145-149
 - - на сериальную корреляцию 170-178
 - - нелинейных гипотез 150-151
 - - условной гомоскедастичности 159-161
- эконометрика рациональных ожиданий 179-189

Теория для конечных выборок 187

Теория для больших выборок:

- для обобщенного метода моментов для нескольких уравнений 296-299

Тест Чоу 105, 641

Тест на коинтегрированность, основанный на остатках 681-685

Тестирование гипотез:

- в предположении нормальности 55-57
- в экстремальных оценках 520-528
- для регрессий на время 192-193
- добавление переменной 253-255
- линейных 61-63
- о предопределенности образования в уравнении заработной платы 250-251
- о подмножествах условий ортогональности 248-251, 262-263
- о сверхидентифицирующих ограничениях 246-251
- об отдельных коэффициентах регрессии 57-59
- при оценивании обобщенным методом моментов 240-245
- с использованием принципа отношения правдоподобия 251-255
- условной гомоскедастичности 159-162, 264

Тестирование подмножества условий ортогональности 248-253, 264, 265

- Тестирование сверхидентифицирующих ограничений 246
- Тестовая статистика вклада 525
- Тобит (цензурированная регрессия) модели 550-553
- Точный размер теста 146
- Транслогарфическая функция издержек:
- доли факторов 326-327
 - многомерная регрессия с перекрестными ограничениями 331-333
 - описание 94, 325-326
 - природа ограничений 330-331
 - свойства 328-329
 - стохастические спецификации 329-330
 - уравнения для 333
 - эластичности замещения 327-328
 - электрические сети общего пользования в США 334
- Треугольное представление 672
- Трехшаговый метод наименьших квадратов:
- как обобщенный метод моментов для одного уравнения 18
 - как обобщенный метод моментов для системы уравнений 305-308
- Триада статистик резюме 527-528

У

- Уайта стандартная ошибка 145-146
- Уайта тест на условную гомоскедастичность 160
- Уравнение недоидентифицированное 232
- Уравнение сверхидентифицированное 89
- Уравнение идентифицируемое точно 232
- Уравнения правдоподобия 78
- Уровни IC :
- как мера способностей 268
 - ошибки в переменных 224-271
- Усечения эффект 544
- Усеченное ядро 439
- Усеченные модели регрессии:
- описание 543
 - проверка состоятельности и асимптотической нормальности 547-549
 - распределения 544-546
 - репараметризация функции правдоподобия 547
 - функция правдоподобия 546-547
- Условие Линдберга 194
- Условие идентифицируемости:
- порядковое 232
 - ранговое 230-232
- Условная гомоскедастичность:
- безусловная и условная 155

- варианты асимптотического теста при условной гомоскедастичности 157-158
- обобщенного метода наименьших квадратов для системы уравнений 303-304
- определение 34
- оценивание долговременной ковариационной матрицы при условной гомоскедастичности 443-447
- оценивание параметризованной условной гомоскедастичности 161-166
- оценка фиксированных эффектов и ослабление требований условной гомоскедастичности 356
- оценки обобщенного метода моментов и последствия условной гомоскедастичности 256-264
- собственная 133
- тестирование 159-161, 264
- эффективная оценка обобщенного метода моментов для нескольких уравнений и условная гомоскедастичность 317-320
- см. также гомоскедастичность
- устойчивое оценивание долговременной ковариационной матрицы при условной гомоскедастичности 438-439

Условное максимальное правдоподобие 481-483

Условное правдоподобие 70-71

Ф

- Фактический размер теста 146
- Федеральная энергетическая комиссия 85
- Федерального резерва бюллетень (Federal Reserve Bulletin) 206
- Фиксированные регрессоры:
- асимптотики с фиксированными регрессорами 194
 - описание 34
- Фиксированный эффект 354
- Фиксированных эффектов оценка:
- несбалансированные панели 367-370
 - описание 356
 - ослабление требований условной гомоскедастичности 363-366
 - свойства на больших выборках 360-362
 - сравнение со случайными эффектами 363-365
 - формула 359-360
- Филлипса тест 637
- Филлипса треугольное представление 671-673

Феллипи: срочна тест 637

Филтер коммутативный 400

Физический:

- коммутируемость стационарных процессов 126
- описание 398
- представляющая функция автокорреляционная 413

Филтеры:

- абсолютно суммируемые 400
- векторные процессы 416-418
- коммутативные 400
- многоканальные 416-417
- обращения 401-402
- описание 399
- произведение фильтров 400-401

Фирм гетерогенность 86

Формация преемник:

- гипотеза эффективного рынка 449-450
- как оптимальные предикторы 449
- описание 449
- регрессионные тесты 450-451
- средние темпы изменения за период 1973-89, 453
- тестирование равенства нулю безразличного среднего 450-453

Формализация отсутствия серьезной зависимости 133

Формализация теоремы 98-99

Функции правдоподобия:

- для многомерных регрессий 555-558
- для регрессионных моделей 71
- квази-псевдо-максимум 76
- максимизация через концентрацию 71-73
- метод максимального правдоподобия 478-480
- описание 70
- оценка максимального правдоподобия с полной информацией 564-565
- робастная модель 561-562
- условной модели регрессии 566-567
- условные и безусловные 70-71

Функциональная центральная предельная теорема 598, 604

Функция:

- потребления 26
- риска 345

Х

Хансена тест на сверхидентифицирующие ограничения 247-248

Характеристические корни 465

Характеристики статистики 263

Характеристики марковская гипотеза 130

Характеризованная модель регрессии 545

Характерные предельные теоремы:

- Гурана для эргодических случайных процессов с нулевым значением 434
- Лундберга-Лева 122, 134-135, 432
- Лундберга-Феллера 194 для $MA(x)$ 431-432
- для нормально распределенных ошибок 55-57
- для эргодических стационарных последовательностей мартигал-разностей 134-135
- функциональные MO , OM

Ч

Частота 414

Числовая слабая закон больших чисел 122, 431

Числами сходящиеся зайной 46

Чоу тест 106

Чоу тест на структурные изменения 100

Ш

Шварца (информационный критерий) правило 425-426

Ширинский 439

Э

Эйлера уравнение 485-486

Эконометрика единичных корней:

- выбор теста на единичный корень 617-619
- Дики-Фуллера тесты 608-619
- инструментальный 598-602
- использование авторегрессионного процесса 602-605
- использование линейных 100 -двоичных 599
- использование моделей авторегрессии с константой сдвига 601-602
- моделирование тренда 591-598
- приложение: паритет покупательной способности 639-641
- тренды для интегрированных процессов 593-595
- упражнения на метод Монте-Карло 647-649

Эконометрика рациональных ожиданий:

- гипотеза эффективности рынка 179-187
- начальная пропорция ставки как оптимальный предиктор 185, 187
- описание 179
- последующее развитие 188-189
- тестирование на серьезную корреляцию 183

Экстремальные оценки:

- М-оценки как экстремальные оценки 477-480
- асимптотическая нормальность 501-523
- измеримость 476-477
- классы 477-478
- обобщенный метод наименьших квадратов как экстремальная оценка 477
- описание 476
- оценка обобщенного метода моментов как экстремальная оценка 477-478
- проверка гипотез 520-523
- согласованность 487-500
- теоремы о состоятельности 487-489
- численная оптимизация 524-532

Электронная база данных (СДБА):

- данные 54-83
- метод наименьших квадратов с ограничением 59
- определение выпуска 56
- особенности 64
- применение технологии Кобба — Дугласа 56-88
- применение эконометрики 85-88
- результаты использования транслогарифмической функции 56-88
- тестирование универсальности функции издержек 80-91
- эластичность выпуска 93
- см. также США Эластичные переменные 560

Эндогенные регрессоры:

- вызываемые или смещения 218
- описание 218-219, 229

Энергетический спектр 414

Эргодическая стационарность:

- анализ временных рядов и эргодическая стационарность 124-125
- Горшны условие для 431-432
- описание 128

Эргодическая теорема 128-129

Эффективная оценка инструментальных переменных с полной информацией:

- обобщенного метода моментов для нескольких уравнений 294-296
- свойства на больших выборках 305

Эффективная оценка обобщенного метода моментов с ограничениями 251

Эффективности рынка гипотеза 179-181, 185-187

Я

Ядерные оценки 438-440

Ядро спектральное квадратичное 440

ADP-GLS-тест 637

CAN 121

DGLS-оценка 695

F-отношение:

- вывод из основе принципа отличия правдоподобий 66
- распределение 63-65
- формулы, содержащие остаточные суммы квадратов в регрессии без ограничений и с ограничениями 65-66

F-тест:

- в сравнении с t-тестом 66-67
- используемые принципы 63-66
- как тест отношения правдоподобий 76

I(0) процесс 503

I(1)-процессы:

- аппроксимация случайным блужданием 600-601
- свойства 595-597
- идентификация 597
- хointегрированные системы и I(1)-процессы 601-604
- со сномом 594
- типы 593-595

I(1)-процесс без сноса 594

I(2)-процессы 593

Journal of Business and Economic Statistics 248

KWW-тест 265

Linn observations 351

Monthly Labor Review 206

OLS-оценка-оценка 360

QR-рекомпозиция 46

t-значение 55

t-отношение:

- описание 58
- распределение 53-60
- разбросное 148-147

t-тест

Ю

Юнг — Уокера уравнения 408, 409

– двусторонний 69
– односторонний 69
– решающее правило 30–60

– сравнение с t -тестом 66–67

Young Men's Cohort of National
Longitudinal Survey (NLS-Y) 266

Учебное издание

Серия «Академический учебник»

Фумин Хайяни
Эконометрика

Главный редактор *Ю.А. Акимович*
Заведующая редакцией *Ю.В. Бандурова*
Выпускающий редактор *Е.В. Пирогов*
Редактор *О.В. Чирковская*
Художник *Е.Н. Сивоскил*
Оформил макет *О.З. Уголов*
Верстка *С.А. Куликов*
Корректор *О.Г. Солодова*

Подписано в печать 16.01.2011. Формат 70х108 1/16

Гарантия Авторского Усл. печ. л. 63,7

Тираж 1000 экз. Изд. № 106 Заказ № 1417

Издательский дом «Дело» РАНХиГС

*19571, Москва, г. г. Вернадского, 92

Компьютерный центр – тел. (495) 438-2619, (495) 433-2502

www.delo.ru

info@delo.ru

Отпечатано в АО «Первая Образцовая типография»

Фрунзеал «Левковский Точный Двор»

142300, Московская область, г. Чехов, ул. Полиграфистов, д.1

Сайт: www.shrd.ru E-mail: sales@shrd.ru, тел. 8(499)770-73-50