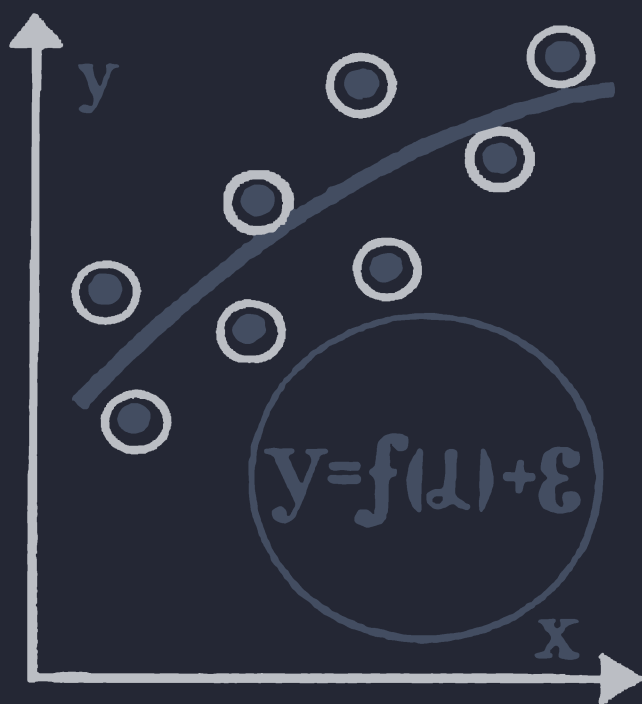


Е. З. Демиденко

ЛИНЕЙНАЯ И НЕЛИНЕЙНАЯ РЕГРЕССИИ



Е. З. Демиденко

ЛИНЕЙНАЯ
И НЕЛИНЕЙНАЯ
РЕГРЕССИИ

Москва «Финансы и статистика» 1981

Предисловие

Методы регрессионного анализа в последнее время стали широко популярными. Они с успехом применяются при анализе экспериментальных данных в различных областях науки: психологии, экономике, социологии, физике, химии, геологии, автоматике и др. В экономике, например, эти методы используются при построении многофакторных моделей производительности труда и функций спроса, производственных функций и экономико-статистических моделей. Во многом этому способствовало быстрое развитие вычислительной техники, которое позволило переложить на ЭВМ большую часть трудоемкой вычислительной работы. В то же время на практике часто применяются традиционные, иногда устаревшие, далекие от реальности подходы и методы. Цель книги — познакомить широкий круг читателей с современными методами оценивания в области регрессионного анализа: робастным (устойчивым) оцениванием в условиях разнородности наблюдений и возможного присутствия выбросов (загрязненные наблюдения); оцениванием в условиях мультиколлинеарности, т. е. сильной «коррелируемости», сопряженности независимых переменных регрессий; оцениванием параметров регрессии при наличии ошибок измерения, присутствующих практически в любой ситуации; нелинейным оцениванием, т. е. оцениванием параметров в нелинейных регрессиях, широко раздвигающим рамки применения регрессионного анализа.

В книге сделана попытка исследовать две схемы регрессии: регрессии как безусловного математического ожидания и регрессии как условного математического ожидания. В первом случае независимые переменные считаются неслучайными (детерминированными). Эта схема хорошо изучена. В схеме регрессии как условного математического ожидания независимые переменные, также как и зависимая переменная, являются случайными (стохастическими). Исследование подобных регрессий только начинается.

Книга состоит из трех частей. В первой части подробно исследуется классическая линейная регрессия, в которой основным моментом является предположение о детерминированности матрицы независимых переменных. Во второй

части рассматриваются альтернативные схемы регрессии и соответствующие им методы оценивания. Здесь изучаются регрессии со случайными независимыми переменными, оценивание в условиях засоренности, сильной коррелируемости факторов регрессии и присутствия ошибок измерения. Существенным является то, что схема линейной модели с ошибками в независимых переменных не является частным случаем регрессии как условного математического ожидания (параграф 4.1). Третья часть книги посвящена нелинейной регрессии. В этой части излагаются методы численного нахождения оценки метода наименьших квадратов и ее статистические свойства.

Автор выражает глубокую благодарность А. В. Коллемаеву, А. М. Дуброву и Г. Г. Пирогову за ряд ценных замечаний, высказанных в процессе подготовки рукописи к изданию.

Часть первая

ЛИНЕЙНАЯ РЕГРЕССИЯ КАК БЕЗУСЛОВНОЕ МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ

Глава 1

КЛАССИЧЕСКАЯ РЕГРЕССИЯ. СВОЙСТВА ОЦЕНКИ МНК

1.1. Основные предположения. Оценка МНК

Рассмотрим случайную величину, характеризующую некоторое явление. Обозначим эту величину y , а последовательность отдельных ее значений y_1, y_2, \dots, y_n . Допустим y зависит от целого ряда явлений, характеризующихся признаками x_1, x_2, \dots, x_m . Каждый из этих признаков описывается своим рядом значений. Естественно, что для анализа зависимости y от x_1, x_2, \dots, x_m регистрация значений признаков должна производиться одновременно. Итак, y — зависимая, а x_1, x_2, \dots, x_m — независимые переменные. Далее предположим, что между переменными имеется линейная связь.

Однако из-за влияния различных неучтенных факторов, а также воздействия случайности и помех наблюдения y будут в большей или меньшей мере отклоняться от линейной зависимости. В силу этого зависимость y от x_1, x_2, \dots, x_m будет не функциональная, а стохастическая. Последнюю можно записать в виде:

$$y_t = \alpha_1 x_{t1} + \alpha_2 x_{t2} + \dots + \alpha_m x_{tm} + \varepsilon_t, \quad t = 1, \dots, n. \quad (1.1)$$

В уравнении (1.1), которое в дальнейшем будем называть *регрессией*, t означает номер наблюдения; $\alpha_1, \alpha_2, \dots, \alpha_m$ — параметры, которые необходимо оценить; ε_t — случайное отклонение. Наличие ε_t в уравнении (1.1) приводит к тому, что эта зависимость будет стохастической. Анализ уравнения (1.1) и методика определения параметров станут более наглядными, а расчетные процедуры существ-

венно упрощаются, если воспользоваться матричной записью уравнения (1.1):

$$y = X\alpha + \varepsilon. \quad (1.2)$$

Здесь y — вектор зависимой переменной размерности $n \times 1$, представляющий собой n наблюдений значений y ; X — матрица независимых переменных, элементы которой суть $n \cdot m$ наблюдений значений m независимых переменных x_1, x_2, \dots, x_m ; размерность матрицы X равна $n \times m$; α — подлежащий оцениванию вектор неизвестных параметров размерности $m \times 1$; ε — вектор случайных отклонений (возмущений) размерности $n \times 1$. Таким образом,

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}, \quad \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix},$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Классический регрессионный анализ базируется на следующих предположениях, определяющих требования к параметрам α , случайным отклонениям ε и независимым переменным x_{ii} .

Предположение А. На вектор неизвестных параметров регрессии (1.1) не наложено ограничений. Это значит, что $\Theta = R^m$, где Θ — множество априорных значений параметров α .

Предположение Б. Вектор $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$ — случайный. Отсюда следует, что $y = (y_1, y_2, \dots, y_n)'$ — также случайный вектор.

Предположение В. Математическое ожидание (м. о.) ε_t равно нулю, т. е. $E(\varepsilon_t) = 0, t = 1, 2, \dots, n$; $E(\varepsilon) = 0$.

Предположение Г. Для любых $t_1 \neq t_2$ $E(\varepsilon_{t_1} \times \varepsilon_{t_2}) = 0, E(\varepsilon_t^2) = \sigma^2$ для всех $t = 1, 2, \dots, n$. Другими словами, $\text{cov}(\varepsilon) = \sigma^2 I_n$. Здесь σ^2 — дисперсия отклонений;

$\text{cov}(\varepsilon)$ — матрица ковариаций отклонений размерности $n \times n$; I_n — единичная матрица размерности $n \times n$, т. е.

$$\text{cov}(\varepsilon) = \begin{bmatrix} \sigma^2 & & & \\ & \sigma^2 & 0 & \\ & & \ddots & \\ & 0 & & \sigma^2 \end{bmatrix}.$$

Предположение Д. Матрица X детерминирована, т. е. x_{ti} не являются случайными переменными.

Предположение Е. $\text{rank}(X) = m$.

Эти предположения (они подробно обсуждаются в следующем параграфе) дают возможность исследовать свойства и статистическое содержание получаемых оценок вектора параметров α .

В дальнейшем выдвинутые предположения мы будем последовательно ослаблять. Так, в параграфе 2.3 рассмотрен случай, когда на вектор неизвестных параметров α наложены линейные ограничения. В параграфе 2.1 рассмотрена ситуация коррелируемых отклонений, имеющих разные дисперсии. Главы 3 и 4 книги посвящены изучению регрессии при случайной матрице X . Случай $\text{rank}(X) < m$ изучается в параграфе 6.2.

По предположению Д независимые переменные детерминированы, поэтому уравнение регрессии (1.2) может быть переписано следующим образом: $Ey = X\alpha$. Таким образом, регрессия, рассматриваемая в первой части книги, имеет вид *безусловного математического ожидания*.

Уравнение (1.1) содержит значения неизвестных параметров $\alpha_1, \alpha_2, \dots, \alpha_m$. Эти величины оцениваются на основе выборочных наблюдений, поэтому полученные расчетные показатели не являются истинными, а представляют собой лишь их статистические оценки. Модель линейной регрессии, в которой вместо истинных значений параметров подставлены их оценки (а именно такие регрессии и применяются на практике), имеет вид

$$y = Xa + e = \hat{y} + e,$$

где a — вектор оценок параметров; e — вектор «оцененных» отклонений регрессии, $e = y - Xa$; \hat{y} — оценка значений y , равная Xa .

Для оценивания неизвестного вектора параметров α воспользуемся методом наименьших квадратов (МНК). Со-

гласно этому методу минимизируется сумма квадратов отклонений

$$Q = Q(\alpha) = \sum_{t=1}^n (y_t - \alpha_1 x_{t1} - \dots - \alpha_m x_{tm})^2 =$$

$$= (y - X\alpha)'(y - X\alpha) = y'y - 2y'X\alpha + \alpha'X'X\alpha \Rightarrow \min. \quad (1.3)$$

Оценкой метода наименьших квадратов в линейной множественной регрессии называют вектор, минимизирующий сумму квадратов отклонений. Для нахождения минимума этой суммы продифференцируем (1.3) по α и приравняем полученное выражение нулю; получим (см. приложение П.2)

$$\partial Q / \partial \alpha = -2X'y + 2X'X\alpha = 0.$$

Если матрица наблюдений независимых переменных X имеет полный ранг, т. е. $\text{rang } X = m$, то, решая последнее уравнение относительно α , найдем

$$a = (X'X)^{-1}X'y. \quad (1.4)$$

Оценка (1.4) является оценкой метода наименьших квадратов.

Т е о р е м а 1.1. Если предположение E выполнено, то оценка МНК (1.4) единственна.

Д о к а з а т е л ь с т в о. Дважды дифференцируя сумму квадратов отклонений Q по α , получим (см. (П.6)) $\partial^2 Q / \partial \alpha^2 = 2X'X$. Матрица X в силу предположения E имеет полный ранг, поэтому матрица $X'X$ положительно определена. Значит, функция Q выпукла вниз и поэтому имеет единственный глобальный минимум, в котором $\partial Q(a) / \partial \alpha = 0$ (см. приложение П.3).

Геометрически $Q(\alpha)$ представляет собой параболоид в евклидовом пространстве R^m . Рассмотрим, например, случай $m = 2$ (рис. 1.1). Минимальное значение $Q = Q(a)$ характеризуется вектором $a = (a_1, a_2)'$. Предположим, что $\text{rang } X < m$, а именно $\text{rang } X = 1$. Тогда график $Q(\alpha)$ будет напоминать «бесконечное корыто», а минимум Q наблюдаться на всей прямой, образующей «дно корыта» (рис. 1, б). При этом в качестве минимизирующего вектора a будет выступать целое семейство точек, расположенных на прямой в плоскости (α_1, α_2) .

Иначе говоря, оценка МНК всегда существует, однако если $\text{rang } X < m$, то оценка МНК не единственна и образу-

ет целое семейство оценок. Формула (1.4) для нахождения оценок в этом случае неприменима (обобщение формулы (1.4) на случай $\text{rank } X < m$ рассмотрено в параграфе 6.2).

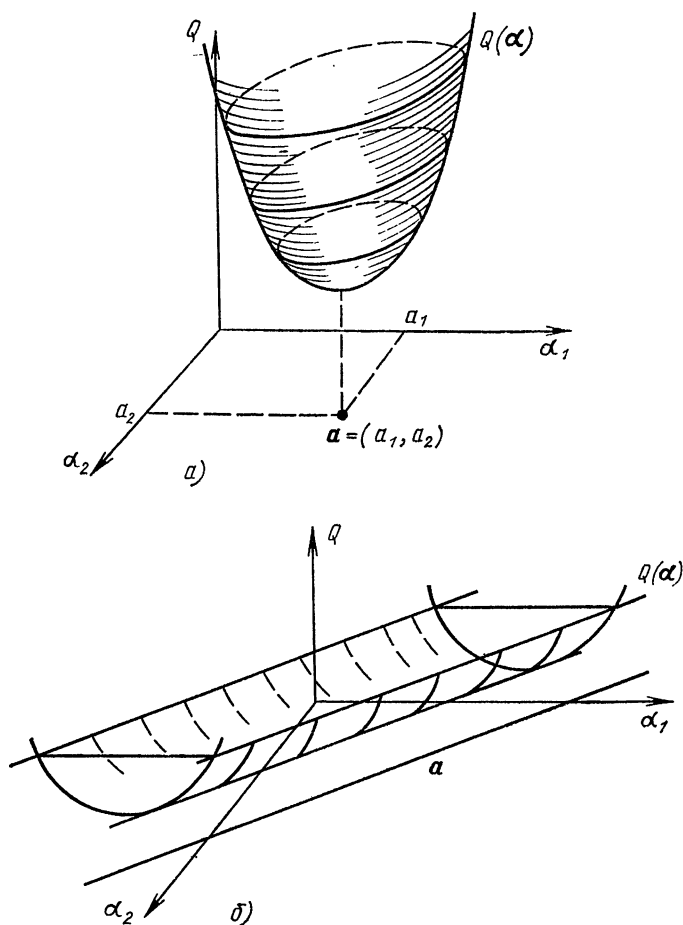


Рис. 1.1. Геометрическая интерпретация минимизации суммы квадратов отклонений, $m=2$: а) $\text{rank } X=2$; б) $\text{rank } X=1$

Для иллюстрации модели регрессии обратимся к примеру. При подборе данных для примера учитывалось основное требование классического регрессионного анализа: детерминированность независимых переменных (предположение Д).

Допустим, имеются три независимые переменные (факторы), воздействующие на зависимую переменную. Для большей конкретности рассмотрим химический эксперимент. Предположим, нас интересует результат реакции некоторого вещества B_1 с веществом B_2 . В ходе реакции получается вещество B_3 . Реакция B_1 и B_2 происходит в присутствии катализатора K . Результат эксперимента, назовем его выходом реакции (вещество B_3), зависит от пропорции веществ B_1 и B_2 . Пусть количество вещества B_2 во всех экспериментах одно и то же, количество же вещества B_1 будем менять. Проведем 15 экспериментов, в каждом из которых количество вещества B_1 варьирует. Каждый эксперимент проводим при некоторой температуре и некотором количестве катализатора K .

Очевидно, что схема линейной регрессии применима к разным ситуациям в различных областях практики. Например, прирост массы скота (Π) зависит от количества и соотношений кормов (K_1, K_2) и условий содержания скота.

Данные регрессионного анализа могут представлять собой временные ряды; они могут быть также элементами пространственно-структурной выборки. Так, y может быть числом дорожно-транспортных происшествий за определенный год по районам, x_1 — численностью населения данного района, x_2 — числом транспортных средств в районе, x_3 — общей протяженностью районных шоссе дорог, x_4 — численностью персонала ГАИ и т. п. Однако все возможные примеры должна объединять одна общая особенность: независимые переменные в таких регрессиях мы считаем детерминируемыми, т. е. не случайными. Случайные несут только зависимые переменные регрессии.

Вернемся к примеру с химическим экспериментом. Введем обозначения:

y_t — выход реакции (количество вещества B_3 , кг) в t -м эксперименте; замер выхода реакции делаем по истечении некоторого времени τ ;

x_{t1} — количество вещества B_1 (кг) в t -м эксперименте;

x_{t2} — температура реакции ($^{\circ}\text{C}$) в t -м эксперименте;

x_{t3} — количество катализатора (г) в t -м эксперименте.

Остальные условия проведения экспериментов остаются неизменными: количество вещества B_2 , величина давления, среда проведения эксперимента и т. п. Результаты и условия 15 экспериментов представлены в табл. 1.1.

Предположим, выход реакции, количество вещества B_1 , температура и количество катализатора связаны линейной

зависимостью типа (1.1). Таким образом, имеется уравнение регрессии

$$y_t = \alpha_1 x_{t1} + \alpha_2 x_{t2} + \alpha_3 x_{t3} + \alpha_4 x_{t4} + \varepsilon_t, \\ t = 1, \dots, 15, \quad (1.5)$$

где $x_{t4} \equiv 1$ для всех t (причины, по которым фактор x_{t4} введен в уравнение (1.5), объяснены ниже).

Оценим параметры регрессии (1.5) с помощью МНК, используя информацию, приведенную в табл. 1.1. По результатам 15 экспериментов получим следующие вектор y и матрицу X :

$$y = \begin{bmatrix} 140,28 \\ 142,02 \\ 149,90 \\ 147,12 \\ 163,62 \\ 173,40 \\ 178,86 \\ 186,26 \\ 183,53 \\ 198,76 \\ 205,30 \\ 206,77 \\ 198,42 \\ 216,48 \\ 221,45 \end{bmatrix}; \quad X = \begin{bmatrix} 252,36 & 96,67 & 8,37 & 1 \\ 262,54 & 100,07 & 9,07 & 1 \\ 285,70 & 96,78 & 9,35 & 1 \\ 277,52 & 101,30 & 9,67 & 1 \\ 307,95 & 100,35 & 9,45 & 1 \\ 322,44 & 104,8 & 10,12 & 1 \\ 334,88 & 106,17 & 10,35 & 1 \\ 350,11 & 109,2 & 11,03 & 1 \\ 346,10 & 104,48 & 10,38 & 1 \\ 374,91 & 106,88 & 12,15 & 1 \\ 378,49 & 113,14 & 12,98 & 1 \\ 397,48 & 112,38 & 11,34 & 1 \\ 378,39 & 109,07 & 10,95 & 1 \\ 393,44 & 114,45 & 12,89 & 1 \\ 403,84 & 115,23 & 13,71 & 1 \end{bmatrix}$$

Находим необходимые для оценивания α значения $X'X$, $(X'X)^{-1}$ и $X'y$ (матрицу $X'X$ часто называют матрицей плана):

$$X'X = \begin{bmatrix} \sum x_{i1}^2 & \sum x_{i1} x_{i2} & \sum x_{i1} x_{i3} & \sum x_{i1} \\ \sum x_{i2}^2 & \sum x_{i2} x_{i3} & \sum x_{i2} \\ *** & \sum x_{i3}^2 & \sum x_{i3} \\ & & n \end{bmatrix} = \\ = \begin{bmatrix} 1747425,6 & 541505,3 & 55592,1 & 5066,2 \\ & 169276,4 & 17254,3 & 1590,9 \\ & & *** & 1774,7 \\ & & & 15 \end{bmatrix};$$

$$(X' X)^{-1} = \begin{bmatrix} 0,0002735 & -0,001650 & 0,001926 & 0,1033 \\ & 0,02810 & -0,05288 & -1,1852 \\ & *** & 0,2696 & 3,3584 \\ & & & 125,4 \end{bmatrix};$$

$$X' y = \begin{bmatrix} \sum y_t x_{t1} \\ \sum y_t x_{t2} \\ \sum y_t x_{t3} \\ \sum y_t \end{bmatrix} = \begin{bmatrix} 935358,3 \\ 289935,3 \\ 29782,1 \\ 2712,2 \end{bmatrix};$$

$$a = (X' X)^{-1} X' y = \begin{bmatrix} 0,3947 \\ 0,2290 \\ 3,7460 \\ -17,083 \end{bmatrix}.$$

Оцененная регрессия выглядит следующим образом:
 $y = 0,3947x_1 + 0,2290x_2 + 3,7460x_3 - 17,083 + e$. (1.6)
 На рис. 1.2 показаны графики \tilde{y}_t и \hat{y}_t , $t = 1, \dots, 15$.

Таблица 1.1

Номер эк-сперимен-та	Выход реакции	Количество вещества B_1	Температу-ра T	Количество катализато-ра K	Оценки выхода реакции	Отклонение
	y	x_1	x_2	x_3	\hat{y}	e
1	140,28	252,36	96,67	8,37	136,01	4,27
2	142,02	262,54	100,07	9,07	143,42	-1,41
3	149,90	285,70	96,78	9,35	153,90	-4,00
4	147,12	277,52	101,30	9,67	147,09	0,03
5	163,62	307,95	100,35	9,45	162,83	0,79
6	173,40	322,44	104,8	10,12	172,08	1,32
7	178,86	334,88	106,17	10,35	180,41	-1,55
8	186,26	350,11	109,2	11,03	187,42	-1,16
9	183,53	346,10	104,48	10,38	182,32	1,21
10	198,76	374,91	106,88	12,15	200,87	-2,11
11	205,30	378,49	113,14	12,98	206,83	-1,53
12	206,77	397,48	112,38	11,34	208,00	-1,23
13	198,42	378,39	109,07	10,95	198,25	0,18
14	216,48	393,44	114,45	12,89	212,69	3,79
15	221,45	403,84	115,23	13,71	220,04	1,41

После того как получена оценка МНК, можно перейти к интерпретации регрессии. В регрессии (1.5) при фиксированных x_2 и x_3 увеличение x_1 на единицу его измерения ведет к изменению математического ожидания y на α_1 единиц. В таком случае говорят, что изменение самого значения y происходит «в среднем» и α_1 может служить оценкой величины изменения y при изменении x_1 .

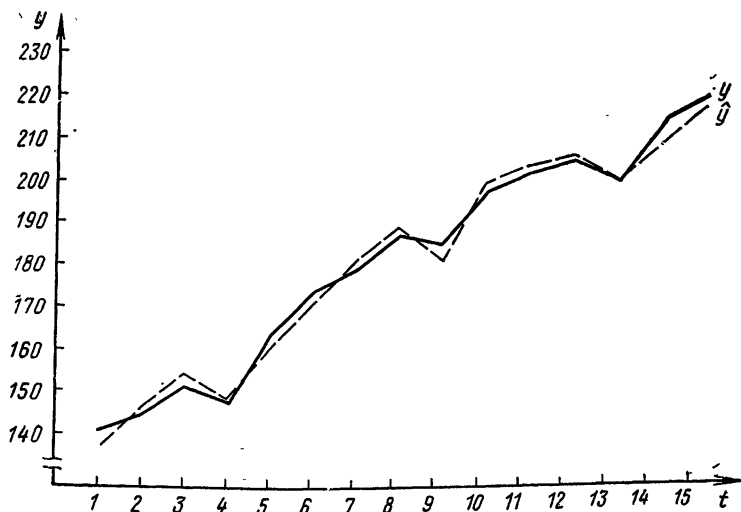


Рис. 1.2. Графики исходных значений y_t и значений \hat{y}_t , вычисленных на основе регрессии (1.6)

Таким образом, если в приведенном примере температуру проведения реакции и количество катализатора зафиксировать и изменять только количество вещества B_1 , то введение в реакцию дополнительно $(1 \text{ кг } B_1)$ приведет к увеличению выхода реакции приблизительно на $0,4 \text{ кг}$ («приблизительно», поскольку на выход реакции могут повлиять помехи ε). При фиксированном количестве B_1 и фиксированной температуре увеличение 1 г катализатора ведет к увеличению B_3 на $3,7 \text{ кг}$.

Однако такое толкование коэффициентов регрессии допустимо в весьма ограниченных пределах. В самом деле, пусть $x_1 = 0$, $x_3 = 0$, а $x_2 = 100$. Получим выход реакции $y = 22,9 - 17,1 = 5,8 \text{ кг}$. В действительности же в этих условиях ни о какой реакции речь не может идти (нет не-

обходимых составляющих). Уравнение регрессии, как правило, имеет смысл только в том диапазоне значений x_1, \dots, x_m , который имел место в эксперименте. В нашем примере — в диапазоне тех значений количества вещества B_1 , участвующего в реакции, температуры реакции и катализатора, на основе которых была оценена регрессия (1.5). Так как регрессия (1.5) учитывала только те значения независимых переменных, которые были охвачены наблюдениями, то

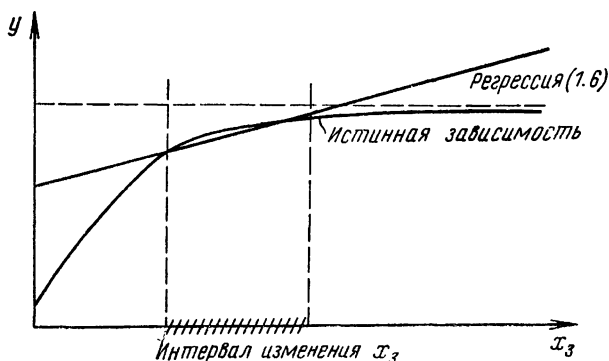


Рис. 1.3. Истинная зависимость y от x_3 и регрессия (1.6)

при выходе за эти рамки модель становится неадекватной действительности. Вполне вероятно, модель (1.5) нелинейна вне охваченного диапазона значений x_1, x_2, x_3 и ее использование в виде (1.6) в этом случае ошибочно. Так, на рис. 1.3 истинная зависимость между выходом реакции y и количеством катализатора x_3 существенно нелинейна вне заштрихованного интервала изменения x_3 .

У п р а ж н е н и я 1.1

1. Пусть X — матрица порядка $n \times m$, $\text{rank } X = r$. Покажите, что $\text{rank}(X'X) = \text{rank } X = r$. В частности, проверьте, что предположение E влечет $|X'X| \neq 0$.

2. Найдите семейство оценок МНК в регрессии $n = 6, m = 2$, где $y_t = 1, t = 1, 2, 3; y_t = 0, t = 4, 5, 6; x_{t1} = 1, x_{t2} = 2, t = 1, \dots, 6$.

3. Проверьте, что оценка МНК в общем случае есть множество $a = \{a \in R^m: X'Xa = X'y\}$. Пусть $\text{rank } X = r < m$; покажите, что тогда a — линейное многообразие размерности $m - r$.

4. Найдите оценку МНК, если $m = n$.

5. Чему равна оценка МНК для случая $m = 1, 2$?

6. Что означает предположение E для случая $m = 1$?

7. Постройте график суммы квадратов отклонений для парной регрессии без свободного члена $y = (1, 3, 4, 2, 5)'$, $x = (1, 2, 3, 4, 5)'$, $y_i = \alpha x_i + \varepsilon_i$; $i = 1, 2, 3, 4, 5$. Найдите оценку МНК.

8. По предположению Гаусса ε_i и ε_j ($i \neq j$) не коррелируют между собой. Верно ли это для их «оценок» e_i и e_j ? Будут ли «оценки» отклонений $\{e_i\}$ гомоскедастичны?¹

1.2. Геометрия МНК

На минимизацию суммы квадратов отклонений можно взглянуть по-иному. Имеется $m + 1$ векторов в евклидовом пространстве R^n : y, x_1, x_2, \dots, x_m . Обозначим S_X — линейное подпространство, натянутое на векторы x_1, x_2, \dots, x_m (семейство линейных комбинаций x_1, x_2, \dots, x_m). Из фор-

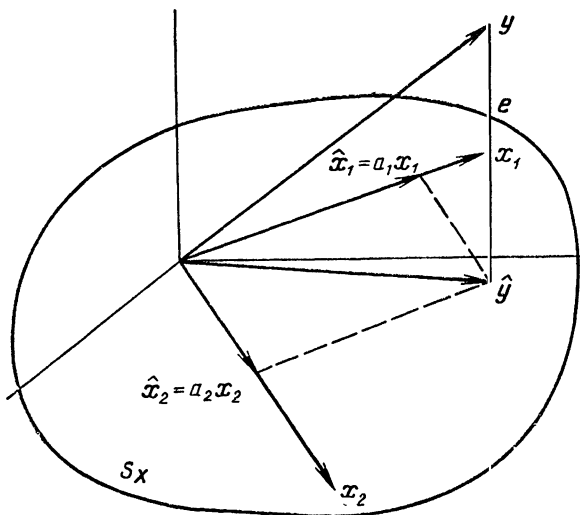


Рис. 1.4. Геометрия МНК, $n=3, m=2$

мулы (1.3) видим, что задача оценивания предполагает нахождение такого вектора $\hat{y} \in S_X$, для которого расстояние между y и \hat{y} минимально, т. е.

$$\|y - \hat{y}\|^2 = (y - X\alpha)'(y - X\alpha) \Rightarrow \min_{\hat{y} \in S_X}$$

Таким образом, для нахождения оценки МНК необходимо сначала найти \hat{y} , а затем разложить его по векторам

¹Отклонения называются гомоскедастичными, если они имеют одинаковую дисперсию.

x_1, \dots, x_m . Коэффициенты разложения и будут координатами оценки МНК a_i . Как известно, минимальное расстояние между вектором и гиперплоскостью есть длина перпендикуляра, опущенного из конца вектора на гиперплоскость (линейное многообразие). Итак, \hat{y} — проекция y на S_X (рис. 1.4), $\|y - \hat{y}\| = \hat{Q}$. Отсюда, в частности, следует, что вектор y единствен при любых ситуациях (т. е. и тогда, когда $\text{rang } X < m$). Для того чтобы найти оценку МНК, т. е. a , необходимо разложить затем \hat{y} по x_1, x_2, \dots, x_m . На рис. 1.4 $\hat{y} = \hat{x}_1 + \hat{x}_2$, где $\hat{x}_i = a_i x_i$. Если $\text{rang } X < m$, то такое разложение не будет единственным.

Мы рассмотрели задачу оценивания вектора неизвестных параметров с точки зрения наилучшего приближения — чисто алгебраической задачи. Однако в силу стохастичности y задача является вероятностной. В дальнейшем нас не столько будет интересовать значение a для фиксированного y , сколько средние характеристики оценки при варьировании y .

У п р а ж н е н и я 1.2

1. В задаче 2 упражнений 1.1 найдите \hat{y} и S_X .
2. Что геометрически означает $a = 0$?
3. Покажите, что в парной регрессии $y_t = \alpha x_t + \varepsilon_t$, $\alpha > 0$ тогда и только тогда, когда угол между векторами x и y острый.
4. Допустим, угол между y и x_i острый для всех $i = 1, \dots, m$. Верно ли тогда, что в множественной регрессии (1.1) $a_i > 0$ для всех $i = 1, \dots, m$? Будет ли это верно, если угол между любой парой x_i и x_j ($i = 1, \dots, m; j = 1, \dots, m$) острый?

1.3. Обсуждение предпосылок классической регрессии

Предположение А означает, что априори нам ничего неизвестно о параметре α . Иногда определенная информация все же существует. Например, в регрессии (1.5) мы вправе предположить, что $\alpha_1, \alpha_2, \alpha_3 > 0$. Априорные ограничения иногда выступают в виде линейных уравнений относительно α с известными коэффициентами. В этом случае оценка МНК будет отличаться от обычной. Исследование линейной регрессии (1.1) при линейных ограничениях на параметры отложим до параграфа 2.3.

Предположение Б констатирует стохастическую природу зависимой переменной. Так, в примере с химическим экспериментом y — случайная переменная. Слу-

чайность выхода реакции есть результат большого числа «неучтенных» факторов: чистоты вещества B_1 , B_2 и катализатора, присутствия посторонних веществ и т. п. В то же время математическое ожидание y есть линейная функция количества вещества B_1 , температуры и катализатора.

Предположение В означает, что $E(\varepsilon_t) = 0$ для всех $t = 1, \dots, n$, т. е. среднее каждого отклонения равно нулю.

При описании химического эксперимента мы требовали неизменность условий проведения эксперимента: количества вещества B_2 , давления, времени проведения реакции. Вообще говоря, эти условия также могут меняться. Например, можно представить ситуацию, когда трудно определить конец реакции или точно измерить давление, при котором происходит реакция. Для того чтобы предположение В выполнялось, достаточно потребовать, чтобы условия эксперимента изменялись случайным образом и независимо друг от друга. Чем сильнее вариация условий экспериментов, тем выше значение σ^2 . Это в свою очередь ведет к ухудшению свойств оценок (точность оценивания падает).

Если отклонения трактовать как суммарный эффект неучтенных факторов, то по предположению В требуется, чтобы этот эффект в среднем был равен нулю. Вообще говоря, такое предположение достаточно обременительно. Достаточно трудно подобрать факторы x_1, \dots, x_m так, чтобы оставшийся эффект «свести на нет». Можно ослабить рассматриваемое предположение и вместо него выдвинуть предположение В'.

Предположение В'. Математическое ожидание ε_t равно константе; $E(\varepsilon_t) = \alpha_{m+1}$ — неизвестный параметр; $-\infty < \alpha_{m+1} < \infty$. Таким образом, мы требуем просто, чтобы остаточный эффект в среднем был постоянен.

Покажем, как в условиях сделанного предположения свести задачу к предыдущей. Положим $\varepsilon'_t = \varepsilon_t - \alpha_{m+1}$, тогда $E(\varepsilon'_t) = 0$ и условие В выполнено. Рассмотрим регрессию

$$y_t = \alpha_1 x_{t1} + \dots + \alpha_m x_{tm} + \alpha_{m+1} x_{t, m+1} + \varepsilon'_t, \quad (1.7)$$

где $x_{t, m+1} \equiv 1$, $t = 1, 2, \dots, n$, т. е.

$$y_t = \alpha_1 x_{t1} + \dots + \alpha_m x_{tm} + \alpha_{m+1} + \varepsilon'_t. \quad (1.8)$$

Если дополненная система векторов $(x_1, \dots, x_m, x_{m+1}) = (x_1, \dots, x_m, 1)$ осталась линейно-независимой, то уравнение (1.7) полностью удовлетворяет всем предположениям

А—Е. При этом, оценивая вектор параметров $\alpha \in R^{m+1}$, мы найдем и оценку для α_{m+1} . В связи с вышеизложенным предпочтительнее пользоваться регрессией со свободным членом; при этом предположение В заменяется более слабым предположением В'. Коэффициент α_{m+1} трактуем тогда как суммарный эффект неучтенных факторов.

Теперь ясны причины, по которым в регрессии (1.5) присутствует постоянный член α_4 . Этим самым мы сможем оценить остаточный эффект воздействия на величину выхода реакции. Мы предполагаем, что этот эффект одинаков для всех 15 экспериментов. Как следует из (1.6), суммарный эффект неучтенных факторов равен -17 , т. е. неучтенные факторы оказывают отрицательное воздействие. Дадим возможное объяснение отрицательного значения «суммарного эффекта» неучтенных факторов для рассматриваемого примера.

Возможно, отрицательный знак a_4 есть результат того, что реакция между B_1 и B_2 не может начаться ниже определенной температуры T . Таким образом, если количество вещества B_1 и количество катализатора K равны нулю, а температура проведения реакции T , выход реакции будет равен нулю. Другими словами, уравнение (1.6) может быть переписано следующим образом:

$$y \approx 0,4x_1 + 0,23(x_2 - 73,9) + 3,75x_3,$$

т. е. пороговое значение температуры $T = 73,9^\circ$.

Вообще говоря, интерпретация свободного члена регрессии как суммарного эффекта неучтенных факторов возможна далеко не всегда. Такая интерпретация правильна, если регрессия продолжает оставаться адекватной в окрестности малых значений независимых переменных $x_1, x_2, \dots, \dots, x_m$.

Очень часто регрессию рассматривают со свободным членом, т. е. в виде (1.8). Если формальное нахождение всех параметров (1.8) по формуле (1.4) связано с обращением матрицы $(m+1) \times (m+1)$, то, используя специфический вид регрессии со свободным членом, можно свести нахождение a к обращению матрицы меньшего порядка $m \times m$. Процедура такова:

1) находят средние y и всех x :

$$\bar{y} = \frac{1}{n} \sum_t y_{ti}; \quad \bar{x}_i = \frac{1}{n} \sum_t x_{ti}, \quad i = 1, 2, \dots, m;$$

2) вычисляют новые векторы y и всех x :

$$\dot{y}_i = y_i - \bar{y}; \quad \dot{x}_{ti} = x_{ti} - \bar{x}_i;$$

3) находят оценку МНК для $\alpha = (\alpha_1, \dots, \alpha_m)'$: $a = (\dot{X}'\dot{X})^{-1}\dot{X}'\dot{y}$;

4) находят оценку для α_{m+1} : $a_{m+1} = \bar{y} - \sum_{i=1}^m a_i \bar{x}_i$.

Можно показать, что полученная оценка для $(\alpha_1, \alpha_2, \dots, \alpha_{m+1})'$ совпадает с оценкой МНК уравнения (1.7), которая непосредственно получается из формулы (1.4).

Предположение Γ означает, что отклонения регрессии (а значит, и сама зависимая переменная) не коррелируют. Иногда требуют большего — независимости отклонений. Условие некоррелируемости довольно ограничительно, например в случае временного ряда y_t . Тогда предположение Γ означает отсутствие автокорреляции ряда ε_t . Другим требованием к отклонениям в классической линейной регрессии является условие гомоскедастичности, т. е. $\sigma^2(\varepsilon_t) = \sigma^2(y_t) = \sigma^2 = \text{const}$ — однородности отклонений, в противном случае говорим о гетероскедастичности. Это условие также довольно часто не выполняется. Даже если ε_t трактовать как ошибки измерения, то вполне вероятно, что большим значениям y_t будет соответствовать и большее значение ε_t . Принятие гипотезы гомоскедастичности означает, что «величина» случайных отклонений в (1.1) должна быть постоянной¹. Параметр $\sigma^2 > 0$ неизвестен. Случай $\sigma^2 = 0$ тривиален. Действительно, тогда с вероятностью 1 $y = X\alpha$ и y — детерминированный вектор. Поскольку $\text{rank } X = m$, существует единственный вектор α , удовлетворяющий этому равенству, т. е. Θ вырождается в точку и оценка тривиальна. В дальнейшем будем предполагать $\sigma^2 > 0$.

Принятие предположения Γ в регрессии (1.5) означает, что, во-первых, в отклонениях нет автокорреляции, т. е. неучтенные факторы ε_t действуют случайно от эксперимента к эксперименту. Автокорреляция отклонений могла наблюдаться, если после каждого эксперимента реактор промывался недостаточно хорошо, и результат $(t + 1)$ -го эксперимента в определенной степени зависел от t -го эксперимента. Вторая часть предположения Γ означает, что разброс неучтенных факторов в (1.5) постоянен.

¹Под «величиной» мы подразумеваем разброс отклонений, т. е. эффект случайности y_t .

Предположение Д фиксирует матрицу независимых переменных. Матрицу X в условиях данного предположения можно рассматривать как систему заданных коэффициентов.

В регрессии (1.5) считаем, в частности, что данные пятнадцати экспериментов x_1, x_2, x_3 не содержат ошибок измерения.

Переменные x_1, x_2, \dots, x_m по предположению не являются стохастическими и часто контролируемы.

Предположение Е влечет единственность оценки МНК и применимость формулы (1.4). Случай, когда это условие не выполняется, рассмотрен в [4], [60], а также в параграфе 5.2. Предположение Е влечет $m \leq n$, т. е. число неизвестных параметров должно быть не больше числа наблюдений.

Можно проверить, что предположение Е выполняется для регрессии (1.5) (см. табл. 1.1).

У п р а ж н е н и я 1.3

1. Найдите оценку МНК в задаче 2 упражнений 1.1; если

$$\Theta = \{\alpha = (\alpha_1, \alpha_2) \in R^2 : \alpha_1 + \alpha_2 = 1\}.$$

2. Выпишите формулы для вычисления оценки МНК в парной регрессии со свободным членом.

3. Допустим, все предположения классической регрессии выполнены, за исключением Г. Известно, что отклонения не коррелируют, но гетероскедастичны: $\sigma^2(\varepsilon_t) = \sigma_t^2 \neq \text{const}$, $t = 1, \dots, n$; причем σ_t^2 известны с точностью до постоянного множителя. Как свести подобную регрессию к гомоскедастичной? Будут ли для нее выполнены все предположения?

4. Рассмотрим простейшую регрессию $y_t = \alpha x_t + \varepsilon_t$, для которой выполнены все предположения Б—Е (ε_t независимы). Априорное множество $\Theta = \{\alpha \geq 0\}$. Докажите, что с положительной вероятностью $a \notin \Theta$, где a — обычная оценка МНК. Верно ли аналогичное утверждение для множественной регрессии с $\Theta = \{\alpha \in R^m : \alpha_1 \geq 0, \dots, \alpha_m \geq 0\}$?

1.4. Методология статистического оценивания

Кратко остановимся на основных моментах современного подхода к теории оценивания. Это поможет читателю глубже понять статистические свойства оценок, изучаемых в книге. Более подробно соответствующие вопросы изложены в [33].

Допустим, имеется n -мерная случайная величина $y = (y_1, \dots, y_n)$, т. е. выборка, распределение которой неиз-

вестно и зависит от некоторого неизвестного параметра α , про который известно лишь, что он принадлежит некоторому априорному множеству Θ . Функцию распределения, соответствующую α , обозначим F_α . Статистикой, или оценкой, называют функцию на R^n , зависящую от y , но не зависящую от α . Рассмотрим одну из оценок, которую, например, обозначим $a = a(y) = a(y_1, \dots, y_n)$. Для простоты будем считать, что число неизвестных параметров равно единице, т. е. $\alpha \in R^1$. Как ввести критерий качества оценивания a неизвестного параметра α ? Предположим,

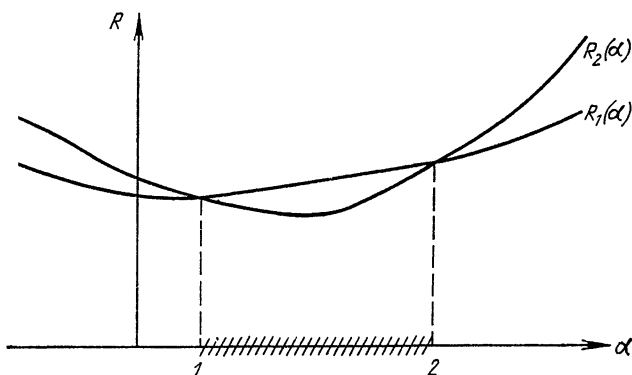


Рис. 1.5. Несравнимые функции риска

что α известно, тогда для заданного y можно выбрать квадратичный критерий (квадратичную функцию потерь), т. е. близость $a(y)$ и α будем измерять величиной $(a(y) - \alpha)^2$. Найдем для данного α усредненную точность оценивания. Она равна:

$$R_a(\alpha) = E_\alpha (a(y) - \alpha)^2 = \int_{R^n} (a(y) - \alpha)^2 dF_\alpha(y). \quad (1.9)$$

Функция $R_a(\alpha)$ называется функцией риска оценки a . Легко видеть, что для несмещенной оценки функция риска есть не что иное, как ее дисперсия. Для каждой оценки функция риска — неотрицательная функция параметра α . Теперь, казалось бы, имеется критерий, по которому можно измерять эффективность оценок и возможно определить наилучшую оценку. Однако многие оценки оказываются несравнимыми.

Действительно, рассмотрим функции риска $R_1(\alpha)$ и $R_2(\alpha)$, соответствующие оценкам $a_1(y)$ и $a_2(y)$ (рис. 1.5).

Какая из них лучше? Однозначно ответить на этот вопрос нельзя. Если истинное значение параметра лежит между 1 и 2, то предпочтительнее пользоваться оценкой a_2 , в противном случае — a_1 . Но мы как раз и не знаем, лежит ли неизвестный параметр в заданных границах или нет! Благоприятная ситуация для сравнения изображена на рис. 1.6. Здесь видно, что первая оценка лучше. Невозможность сравнения любых функций риска делает задачу нахождения оптимальной оценки в классе всех оценок неразрешимой. Действительно, допустим имеется некоторая

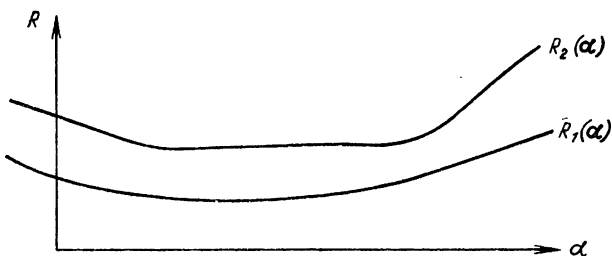


Рис. 1.6. Сравнимые функции риска

«наилучшая» оценка $a(y)$, функция риска которой $R_a(\alpha)$ минимальна, т. е. для любой другой оценки $b(y)$ имеем $R_a(\alpha) \leq R_b(\alpha)$ для всех α . Очевидно $R_a(\alpha) \neq 0$, так как в противном случае $a(y) = \alpha$ и Θ вырождается в точку.

Итак, пусть $R(\alpha_0) > 0$. В качестве новой оценки α положим $b = b(y) \equiv \alpha_0$ для всех $y \in R^n$. Найдем функцию риска такой оценки. По формуле (1.9) $R_b(\alpha) = E_\alpha (\alpha - \alpha_0)^2 = (\alpha - \alpha_0)^2$ (рис. 1.7). Очевидно, в заштрихованной окрестности α_0 оценка b будет лучше «наилучшей» оценки a .

Существует несколько способов обхождения трудностей, связанных с несравнимостью функций риска.

Байесовский подход. Как правило, не все точки априорного множества параметров для нас одинаковы. Так, если $\Theta = (0, 1)$, то весьма возможно, что гораздо важнее хорошо оценить неизвестный параметр, лежащий в окрестности 0,5, чем по краям отрезка $(0, 1)$. Таким образом, можно ввести функцию предпочтения, или весовую функцию. Ее можно рассматривать и как априорное распределение параметра $\alpha \in \Theta$. Так, в нашем случае в качестве функции предпочтения можно взять $p(\alpha) = \alpha(1 - \alpha)$. В общем случае $p(\alpha)$ есть неотрицательная функция на Θ . С по-

мощью функции предпочтения можно найти усредненную функцию риска $B_a = \int_0^1 R_a(\alpha) p(\alpha) d\alpha$. Теперь каждой оценке соответствует фиксированное число — задача сравнения оценок решена. Оптимальная байесовская оценка минимизирует усредненную функцию риска B_a .

Минимаксный подход. При этом подходе ориентируются на худшую возможность, т. е. на максимальное значение функции риска $M_a = \sup_{\alpha} R_a(\alpha)$. Наилучшей минимакс-

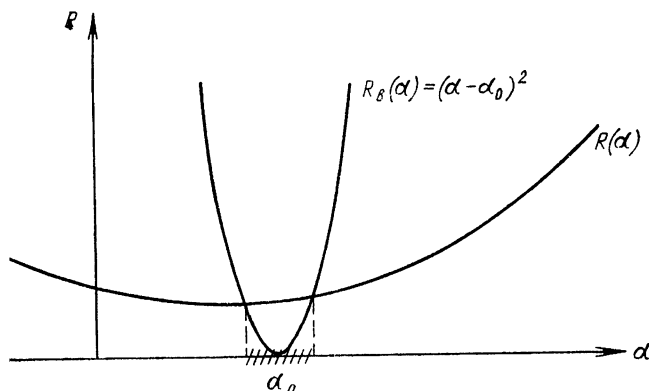


Рис. 1.7. Функция риска тривиальной оценки $b \equiv \alpha_0$

ной оценкой в этом смысле является та, которая минимизирует максимальный риск M_a .

Существует другой способ регуляризации задачи оценивания — сужение класса рассматриваемых оценок. Часто ограничивая класс оценок, мы можем найти оптимальную оценку из этого класса. Наиболее часто рассматривают класс *несмещенных оценок*. В этом случае даже можно вычислить нижнюю границу функции риска $R^*(\alpha)$; она называется границей Крамера—Рао. Для любой несмещенной оценки $a(y)$ имеем $R_a(\alpha) \geq R^*(\alpha)$ (рис. 1.8). Граница Крамера—Рао для дисперсии несмещенной оценки находится довольно просто: это обратная величина информации по Фишеру $I(\alpha)$, т. е. для любой несмещенной оценки $a(y)$ имеем

$$R_a(\alpha) = E_{\alpha} (a(y) - \alpha)^2 \geq 1/I(\alpha) = R^*(\alpha), \quad (1.10)$$

где

$$I(\alpha) = E_{\alpha} \left[\frac{\partial \ln f(\cdot; \alpha)}{\partial \alpha} \right]^2, \quad (1.11)$$

а. $f(\cdot; \alpha)$ — функция плотности выборки y_1, y_2, \dots, y_n ¹. Поэтому если мы докажем, что дисперсия некоторой оценки равна $1/I(\alpha)$ для любого $\alpha \in \Theta$, то эту оценку можно считать эффективной в классе несмещенных оценок.

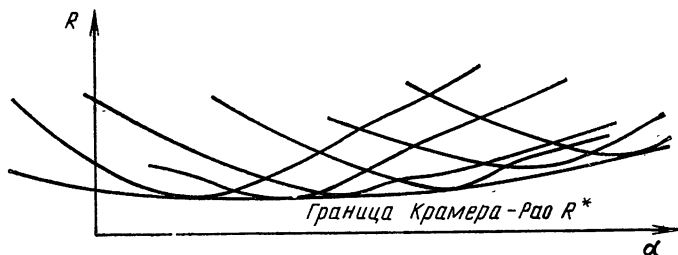


Рис. 1.8. Нижняя граница Крамера—Рао R^*

Можно рассмотреть еще более узкий класс — класс линейных несмещенных оценок. Каждую оценку из этого класса можно записать в виде

$$a = a(y) = \sum_{i=1}^n c_i y_i,$$

где c_i — некоторые константы, удовлетворяющие в силу несмещенности a условию $\sum c_i \mu_i(\alpha) = \alpha$, где $\mu_i(\alpha) = E y_i$. В этом классе также можно найти эффективную оценку. Она находится, как правило, весьма просто. Докажем, что средняя арифметическая является эффективной в этом классе оценкой математического ожидания при любой функции распределения с неизвестным математическим ожиданием μ и единичной дисперсией (y_1, y_2, \dots, y_n — случайная выборка, т. е. y_i независимы и одинаково распределены). Итак, пусть $m_1 = m_1(y) = \sum c_i y_i$, причем в силу несмещенности $\sum c_i = 1$. Тогда

$$R_{m_1}(\mu) = \sigma^2(m_1) = \sum_i c_i^2 \Rightarrow \min.$$

Легко показывается, что решением этой оптимизационной задачи является $c_1 = \dots = c_n = 1/n$, т. е. $m_1 = m = \sum y_i/n$.

¹При этом должны быть выполнены определенные условия регулярности (см., например, [63, с. 360]).

Если y_1, y_2, \dots, y_n нормально распределены, то можно показать, что средняя является эффективной оценкой в классе всех несмещенных оценок. Действительно, в этом случае

$$f(y_i; \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_i - \mu)^2}$$

— плотность распределения i -й координаты вектора y ; плотность распределения вектора y равна в силу независимости произведению таких плотностей:

$$f(y; \mu) = (2\pi)^{-n/2} \exp \left[-\frac{1}{2} \sum_i (y_i - \mu)^2 \right],$$

$$\ln f(y; \mu) = -\frac{n}{2} \ln 2\pi - \frac{1}{2} \sum_i (y_i - \mu)^2,$$

$$\frac{\partial \ln f(y; \mu)}{\partial \mu} = \sum_i (y_i - \mu).$$

Информационное количество в выборке y_1, y_2, \dots, y_n по Фишеру равно:

$$\begin{aligned} I(\mu) &= E_{\mu} \left(\frac{\partial \ln f(y; \mu)}{\partial \mu} \right)^2 = E \left[\sum_i (y_i - \mu)^2 \right] = \\ &= \sum_i E (y_i - \mu)^2 = n. \end{aligned}$$

Итак, нижняя граница дисперсии несмещенной оценки Крамера—Рао равна $1/n$, но дисперсия средней тоже равна $1/n$, откуда следует, что средняя — эффективная оценка в классе несмещенных оценок.

Мы доказали, что средняя является эффективной в классе несмещенных оценок, если выборка подчинена нормальному закону. Предположение нормальности существенно. Если взять другой закон распределения, то эта оценка может оказаться уже не эффективной.

Может быть введен еще один класс оценок: *оценки с ограниченной функцией риска*. Оценка $a = a(y)$ принадлежит классу оценок с ограниченной функцией риска, если найдется такое число M , что $R_a(\alpha) \leq M$ для всех $\alpha \in \Theta$. Класс оценок с ограниченной функцией риска вводится для того, чтобы не рассматривать оценки плохие при некоторых значениях неизвестных параметров.

Допустим, $y = (y_1, y_2, \dots, y_n)$ — случайная выборка из распределения F_α , $\alpha \in \Theta \subset R^1$. Предположим также, что $\sigma^2(y_i) = \sigma^2$, которая известна. Нетрудно показать, что тогда класс линейных несмещенных оценок параметра α уже класса линейных оценок с ограниченной функцией риска. Действительно, пусть $a = a(y)$ — линейная несмещенная оценка, т. е.

$$a(y) = \sum_{i=1}^n c_i y_i, \quad E a(y) = \alpha.$$

Тогда ее функция риска ограничена (не зависит от α):

$$R_a(\alpha) = \sigma^2(a) = \sigma^2\left(\sum_i c_i y_i\right) = \sigma^2 \sum_i c_i^2.$$

Если в качестве неизвестного параметра α выступает математическое ожидание y_i , т. е. $E y_i = \alpha$, то можно показать, что класс линейных несмещенных оценок и класс линейных оценок с ограниченной функцией риска совпадают.

Пример. Пусть $y = (y_1, \dots, y_n)$ — случайная выборка из совокупности с нормальным распределением с неизвестным математическим ожиданием μ и единичной дисперсией, т. е. $y_i \sim N(\mu, 1)$. Предположим, что априорное множество μ есть вся числовая прямая $(-\infty, +\infty)$. В качестве оценок μ рассмотрим три оценки: первая — хорошо известная средняя $m = m(y) = \sum y_i/n$; вторая — полусумма первого и последнего наблюдений $k = k(y) = (y_1 + y_n)/2$ и третья — «глупая» оценка $l = l(y) = y_1/n$. Найдем функцию риска каждой оценки. Прежде всего заметим, что первая и вторая оценки — несмещенные. Далее,

$$R_m(\mu) = E_\mu (m(y) - \mu)^2 = \sigma^2(m) = \frac{1}{n};$$

$$R_k(\mu) = E_\mu (k(y) - \mu)^2 = \sigma^2(k) = \frac{1}{2};$$

$$\begin{aligned} R_l(\mu) &= E_\mu (l(y) - \mu)^2 = E_\mu l^2 - 2\mu E_\mu l + \mu^2 = \\ &= \mu^2 \frac{(n-1)^2}{n^2} + \frac{1}{n^2}. \end{aligned}$$

Очевидно, для всех $n > 2$ первая оценка будет лучше второй (рис. 1.9). Может ли «глупая» оценка быть лучше средней? Легко проверить, что это возможно, если истинное значение математического ожидания заключено в пределах $(-1/\sqrt{n-1}, 1/\sqrt{n-1})$. На интервале $(-1/2, 1/2)$ «глупая» оценка будет лучше при $n = 5$.

Перейдем к случаю многомерного оценивания, т. е. $m > 1$. Пусть A — некоторая положительно определенная детерминированная (весовая) матрица, $a = a(y)$ — оценка, случайный вектор $m \times 1$ параметра $\alpha \in \Theta \subset R^m$. Обобщенной квадратичной функцией потерь назовем величину

$$R_a(\alpha) = E [(a(y) - \alpha)' A (a(y) - \alpha)]. \quad (1.12)$$

Вместо (1.12) можно рассмотреть матрицу средних квадратов отклонений (ошибок) (СКО):

$$E [(a(y) - \alpha) (a(y) - \alpha)']. \quad (1.13)$$

Оценка $a_1(y)$ не хуже оценки $a_2(y)$ в случае (1.13), если для всех истинных значений параметров α разность СКО между второй и первой оценками есть неотрицательно

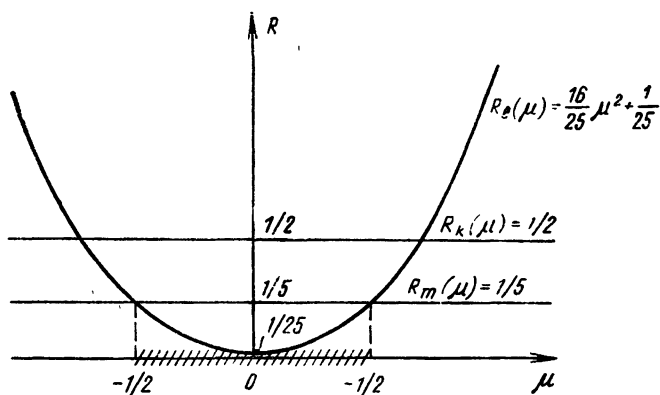


Рис. 1.9. Функции риска различных оценок средней при $n=5$

определенная матрица. Для несмещенных оценок матрица СКО превращается в матрицу ковариаций

$$\text{cov}(a) = E (a(y) - \alpha) (a(y) - \alpha)' = E (a(y) - Ea) \times \\ \times (a(y) - Ea)'$$

Таким образом, несмещенная многомерная оценка a_1 не хуже несмещенной оценки a_2 , если $\text{cov}(a_2) - \text{cov}(a_1)$ неотрицательно определена, что будем в дальнейшем записывать как $\text{cov}(a_1) \leq \text{cov}(a_2)$. Если разность $\text{cov}(a_2) - \text{cov}(a_1)$ — неотрицательно определенная и ненулевая матрица, то a_1 лучше a_2 . В этом случае будем иногда говорить, что a_1 имеет меньшую матрицу ковариаций, чем a_2 . Каждый из критериев эффективности многомерных оценок

(1.12) и (1.13) имеет свои преимущества и недостатки. Преимущества первого критерия: сравнимость эффективностей любых двух оценок. Недостаток (1.12): необходимо выбрать весовую матрицу \mathbf{A} , которая априори часто неизвестна. Недостаток (1.13) — несравнимость эффективностей некоторых оценок. Можно показать, что оба критерия в определенном смысле приводят к одним и тем же оптимальным оценкам: если оценка $\mathbf{a}_1(\mathbf{y})$ не хуже оценки $\mathbf{a}_2(\mathbf{y})$ в смысле (1.13), то $\mathbf{a}_1(\mathbf{y})$ не хуже $\mathbf{a}_2(\mathbf{y})$ в смысле (1.12) для любых весовых матриц \mathbf{A} , и наоборот (см. [191]).

Часто в критерии (1.12) в качестве весовой матрицы выбирают единичную. Тогда приходим к минимизации средней суммы квадратов ошибок (ССКО):

$$L_a(\alpha) = E[(\mathbf{a}(\mathbf{y}) - \alpha)'(\mathbf{a}(\mathbf{y}) - \alpha)] = \text{tr} E[(\mathbf{a}(\mathbf{y}) - \alpha)(\mathbf{a}(\mathbf{y}) - \alpha)']; \quad (1.14)$$

Неравенство Крамера—Рао (1.10) обобщается на многомерный случай. Пусть $\alpha \in R^m$, \mathbf{a} — несмещенная оценка, тогда

$$\text{cov}(\mathbf{a}) \geq \mathbf{I}^{-1}(\alpha), \quad (1.15)$$

где

$$\mathbf{I}(\alpha) = E_{\alpha} \left[\left(\frac{\partial \ln f(\cdot; \alpha)}{\partial \alpha} \right) \left(\frac{\partial \ln f(\cdot; \alpha)}{\partial \alpha} \right)' \right]; \quad (1.16)$$

здесь $\mathbf{I}(\alpha)$ — положительно определенная матрица $m \times m$. Неравенство (1.15) следует понимать так, что разность между левой и правой частями — неотрицательно определенная матрица.

У п р а ж н е н и я 1.4

1. Рассмотрим схему последовательных независимых испытаний. Пусть число испытаний равно n ; при каждом испытании событие A наступает с вероятностью θ и не наступает с вероятностью $1 - \theta$, $\theta \in \Theta = (0, 1)$. В качестве оценки θ выбирается $\hat{\theta} = m/n$, где m — число появлений события A . Покажите, что $\hat{\theta}$ — несмещенная оценка. Постройте ее функцию риска. Будет ли $\hat{\theta}$ линейной оценкой?

2. Пусть y_1, y_2, \dots, y_n — независимы и одинаково распределены, $y_i \sim R(0, \theta)$ — равномерное распределение на $(0, \theta)$, $\theta \in \Theta = (0, \infty)$. Требуется оценить θ . Рассмотрим оценку $\hat{\theta}_m = \max y_i$. Будет ли она несмещенной? Найдите ее функцию риска.

Найдите несмещенную оценку с минимальной дисперсией; чему рав-

на ее функция риска. Постройте линейную несмещенную оценку. Сравните ее функцию риска с функцией риска $\hat{\theta}_m$. Какая оценка лучше?

3. При условии задачи 1 нас интересует оценка θ^2 . Будет ли оценка $(m/n)^2$ несмещенной?

1.5. Теорема Гаусса—Маркова

Оценка МНК является статистикой, т. е. случайной величиной. Разные наблюдения y приводят к разным значениям оценки, причем зависимость a от y линейная (a —линейная оценка). Вычислим математическое ожидание и матрицу ковариаций оценки a . Используя (1.4), получим

$$Ea = E [(X'X)^{-1} X'y].$$

Так как по предположению $D X$ есть матрица детерминированная, то ее можно выносить за знак математического ожидания, т. е.

$$\begin{aligned} Ea &= E [(X'X)^{-1} X'y] = (X'X)^{-1} X'E(y) = \\ &= (X'X)^{-1} X'X\alpha = \alpha. \end{aligned} \quad (1.17)$$

Таким образом, математическое ожидание оценки равно истинному значению параметра. Итак, оценка МНК несмещена.

Найдем матрицу ковариаций оценки МНК. По определению

$$\text{cov}(a) = E (a - Ea) (a - Ea)'$$

Из несмещенности оценки МНК следует

$$\begin{aligned} a - Ea &= a - \alpha = (X'X)^{-1} X'(X\alpha + \varepsilon) - \alpha = \alpha + \\ &+ (X'X)^{-1} X'\varepsilon - \alpha = (X'X)^{-1} X'\varepsilon, \end{aligned}$$

откуда

$$\begin{aligned} \text{cov}(a) &= E [(X'X)^{-1} X'\varepsilon\varepsilon'X(X'X)^{-1}] = \\ &= (X'X)^{-1} X'E(\varepsilon\varepsilon')X(X'X)^{-1} = \\ &= (X'X)^{-1} X'\sigma^2 I_n X(X'X)^{-1} = \sigma^2 (X'X)^{-1}; \end{aligned}$$

окончательно

$$\text{cov}(a) = \sigma^2 (X'X)^{-1}. \quad (1.18)$$

Нельзя считать (1.18) статистикой, поскольку $\text{cov}(a)$ зависит от неизвестного параметра σ^2 . Позднее мы оценим σ^2 и найдем оценку для $\text{cov}(a)$.

Приведем известную теорему Гаусса—Маркова, в которой говорится о важнейших статистических свойствах оценки МНК.

Теорема Гаусса—Маркова 1.2. Пусть предположения А—Е выполняются, оценка МНК является

а) несмещенной,

б) эффективной в классе несмещенных оценок, линейных по y .

Доказательство. Несмещенность уже была нами доказана. Докажем линейную эффективность МНК. Пусть \mathbf{b} — другая несмещенная оценка, линейная по y . Тогда $\mathbf{b} = \mathbf{H}y$, где \mathbf{H} — некоторая детерминированная матрица $m \times n$. Из условия несмещенности \mathbf{b} имеем

$$\mathbf{E}\mathbf{b} = \mathbf{E}[\mathbf{H}y] = \mathbf{H}\mathbf{E}y = \mathbf{H}\mathbf{X}\alpha = \alpha \text{ для любого } \alpha \in R^m,$$

откуда

$$\mathbf{H}\mathbf{X} = \mathbf{I}_m. \quad (1.19)$$

Обозначим $\mathbf{C} = \mathbf{H} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Тогда из условия (1.19)

$$\begin{aligned} \mathbf{C}\mathbf{X} &= [\mathbf{H} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{X} = \mathbf{H}\mathbf{X} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \\ &= \mathbf{H}\mathbf{X} - \mathbf{I} = \mathbf{0}, \end{aligned}$$

$$\mathbf{b} - \alpha = \mathbf{H}y - \alpha = \mathbf{H}\mathbf{X}\alpha + \mathbf{H}\varepsilon - \alpha = \mathbf{H}\varepsilon.$$

Вычислим матрицу ковариаций для оценки \mathbf{b} :

$$\begin{aligned} \text{cov}(\mathbf{b}) &= \mathbf{E}(\mathbf{b} - \alpha)(\mathbf{b} - \alpha)' = \mathbf{E}(\mathbf{H}\varepsilon\varepsilon'\mathbf{H}') = \sigma^2\mathbf{H}\mathbf{H}' = \\ &= \sigma^2[\mathbf{C} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'][\mathbf{C}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] = \\ &= \sigma^2[\mathbf{C}\mathbf{C}' + \mathbf{C}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}' + \\ &+ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] = \sigma^2\mathbf{C}\mathbf{C}' + \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \\ &= \sigma^2\mathbf{C}\mathbf{C}' + \text{cov}(\mathbf{a}). \end{aligned}$$

Поскольку матрица $\mathbf{C}\mathbf{C}'$ неотрицательно определена, получаем $\text{cov}(\mathbf{b}) \geq \text{cov}(\mathbf{a})$, т. е. разница матриц ковариации любой линейной несмещенной оценки и оценки МНК неотрицательно определена (см. параграф 1.4).

З а м е ч а н и я: 1. Как видно из доказательства, $\mathbf{C} \neq \mathbf{0}$, поэтому хотя бы один диагональный элемент матрицы $\sigma^2\mathbf{C}\mathbf{C}'$ больше нуля. Это, в частности, означает, что если \mathbf{b} — какая-либо другая несмещенная линейная оценка, то дисперсия i -й координаты $\sigma^2(b_i) \geq \sigma^2(a_i)$ — дисперсии i -й координаты оценки МНК, причем хотя бы для одной координаты $\sigma^2(b_j) > \sigma^2(a_j)$.

2. Еще раз подчеркнем оптимальность свойств оценок МНК в своем классе. Как следует из предыдущего параграфа, можно построить смещенные оценки, которые для некоторых значений α будут лучше оценок МНК. Существенным условием является требование линейности оценок. Весьма вероятно, что можно построить несмещенные нелинейные оценки, которые будут для всех α более оптимальными, чем оценка МНК. Однако в дальнейшем нами показано, что если распределение ϵ нормально, то оценка МНК будет наилучшей в классе всех (линейных и нелинейных) несмещенных оценок.

Дж. Ходжес и Е. Леман доказали минимаксность оценки МНК [129].

Иногда нам интересны не сами оценки вектора параметров $\alpha_1, \alpha_2, \dots, \alpha_m$, а некоторые их линейные комбинации. Допустим, нас интересуют оценки вектора $\beta = \mathbf{V}\alpha$, где \mathbf{V} — известная матрица $k \times m$, β — неизвестный вектор, подлежащий оцениванию, размерности $k \times 1$. Можно доказать, что оценка $\mathbf{b} = \mathbf{V}\mathbf{a}$ является: а) несмещенной, б) эффективной в классе несмещенных оценок, линейных по \mathbf{y} . Это позволяет прояснить еще одно свойство оценки МНК. Предположим, в регрессии нас интересует только один параметр, скажем α_1 . Мы стараемся получить наилучшую линейную несмещенную оценку только этого параметра. Что это будет за оценка? Как следует из последнего факта, это будет оценка МНК. Действительно, положим $\mathbf{V} = (1, 0, \dots, 0)$ — матрица $1 \times m$, тогда $\beta = \alpha_1$ и наилучшей оценкой будет $b = \mathbf{V}\mathbf{a} = a_1$, т. е. первая координата оценки МНК. Итак, можно сделать вывод: МНК является одновременно эффективным и с точки зрения оценивания индивидуального параметра регрессии, и с точки зрения оценивания всех параметров совместно.

Можно также рассмотреть линейные несмещенные оценки, минимизирующие сумму дисперсий параметров. Эти оценки совпадают с оценками МНК.

До сих пор мы интересовались оцениванием параметров регрессии $\alpha_1, \alpha_2, \dots, \alpha_m$. Однако существует еще один неизвестный параметр — дисперсия отклонений регрессии σ^2 . Как найти удовлетворительную оценку для этого параметра? Обозначим $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\mathbf{a}$. Положим

$$s^2 = \frac{\mathbf{e}'\mathbf{e}}{n-m} = \frac{\sum \epsilon_i^2}{n-m} \quad (1.20)$$

Теорема 1.3. *Статистика s^2 несмещенно оценивает σ^2 .*
Доказательство дано в параграфе 1.11.

Как мы уже отмечали, выражение (1.18) нельзя считать статистикой, поскольку σ^2 неизвестно. Теперь, имея оценку s^2 , можно построить несмещенную оценку для матрицы ковариации оценки МНК:

$$\widehat{\text{cov}}(\mathbf{a}) = s^2 (\mathbf{X}' \mathbf{X})^{-1}. \quad (1.21)$$

На практике матрицу ковариаций довольно затруднительно интерпретировать, так как она зависит от единиц измерения оценок, что в свою очередь зависит от единиц измерения переменных y, x_1, \dots, x_m . С этой точки зрения удобнее пользоваться матрицей корреляций параметров. Она рассчитывается на основе (1.21), (i, j) -й элемент которой равен:

$$\text{cor}_{ij}(\mathbf{a}) = (\mathbf{X}' \mathbf{X})_{ij}^{-1} / \sqrt{(\mathbf{X}' \mathbf{X})_{ii}^{-1} (\mathbf{X}' \mathbf{X})_{jj}^{-1}},$$

где $(\mathbf{X}' \mathbf{X})_{ij}^{-1}$ — (i, j) -й элемент матрицы $(\mathbf{X}' \mathbf{X})^{-1}$. По значению cor_{ij} мы можем оценить, как i -я координата оценки МНК коррелирует с j -й координатой. Грубо говоря, мы как бы оцениваем линейную взаимозаменяемость параметров регрессии. Поскольку $\text{cor}(\mathbf{a})$ не зависит от выборки \mathbf{y} , ее можно считать «абсолютно точной оценкой». Это освобождает нас от необходимости проверять эти коэффициенты на значимость, так как мы сразу получаем истинные значения коэффициентов корреляции между a_i и a_j .

Найдем статистику s^2 , матрицу ковариаций и корреляций для регрессии (1.6). В предпоследней и последней колонках табл. 1.1 даны соответственно $\hat{\mathbf{y}}$ и $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$. Сумма квадратов отклонений $\hat{Q} = 68,46$, поэтому $s^2 = \hat{Q}/(n-m) = 68,46/11 = 6,22$ — несмещенная оценка дисперсии σ^2 . Используя матрицу $(\mathbf{X}' \mathbf{X})^{-1}$, вычисленную в параграфе 1.1, найдем оценку матрицы ковариаций оценки МНК:

$$\widehat{\text{cov}}(\mathbf{a}) = s^2 (\mathbf{X}' \mathbf{X})^{-1} =$$

$$= \begin{bmatrix} 0,001702 & -0,0103 & -0,01198 & 0,6425 \\ & 0,1747 & -0,3289 & -7,3712 \\ & & *** & 1,677 & 20,889 \\ & & & & 780,64 \end{bmatrix}.$$

Исходя из этой матрицы можно найти дисперсию и среднеквадратические отклонения для оценки α (табл. 1.2). Матрица корреляций параметров равна:

$$\text{cov}(\alpha) = \begin{bmatrix} 1,00 & -0,595 & -0,224 & 0,558 \\ & 1,000 & -0,608 & -0,987 \\ & & *** & 1,000 & 0,578 \\ & & & & 1,000 \end{bmatrix}.$$

Таблица 1.2

Дисперсия оценки, s_i^2	α_1	α_2	α_3	α_4
		0,001702	0,1747	1,677
Среднее квадратическое отклонение (стандартная ошибка), s_i	0,04126	0,4180	1,295	27,94

Таким образом, можно утверждать, что коэффициенты корреляции между a_1 и a_2 , a_1 и a_4 , a_2 и a_3 , a_3 и a_4 приблизительно одинаковы по абсолютной величине. Максимальный коэффициент корреляции равен $-0,987$ и соответствует корреляции между a_2 и a_4 . Как интерпретировать $r(a_i, a_j)$? Рассмотрим, например, a_1 и a_2 . Мы получили $r(a_1, a_2) = -0,595$. Это означает, что если повторять наблюдения за y (при фиксированной матрице X) и каждый раз вычислять оценку МНК, то, расположив пару (a_1, a_2) на плоскости, получим облако рассеяния с коэффициентом корреляции, равным $-0,595$. Знак минус означает, что при увеличении a_1 мы должны скорее ожидать уменьшение a_2 , и наоборот.

У п р а ж н е н и я 1.5

1. Покажите, что оценка МНК линейна по y .
2. Пусть A и B — матрицы $m \times m$. Докажите, что если $A \geq B$, то $A_{ii} \geq B_{ii}$ для $i = 1, \dots, m$. Верно ли обратное?
3. Пусть $b = Ny$ — несмещенная линейная оценка α . Как следует из доказательства теоремы Гаусса—Маркова, $\text{cov}(b) = \sigma^2 NN'$. Найдите b из условия $\text{tr} \text{cov}(b) = \sigma^2 \text{tr} NN' \Rightarrow \min$.
4. Под обобщенной дисперсией несмещенной оценки понимают определитель ее матрицы ковариаций. Покажите, что оценка МНК минимизирует обобщенную дисперсию.
5. Является ли оценка s несмещенной оценкой σ ?

6. Дана регрессия $y_t = \alpha x_t + \varepsilon_t$. В качестве оценки для α выбирается $b = \Sigma y_t / \Sigma x_t$ ($\Sigma x_t \neq 0$) — коэффициент наклона прямой, соединяющей начало координат и «среднюю» точку (\bar{x}, \bar{y}) ;
 а) докажите, что оценка b — линейная несмещенная оценка;
 б) найдите ее дисперсию. Во сколько раз оценка МНК эффективнее b ? Когда эффективности оценок совпадают?

7. Дана регрессия $y_t = \alpha_1 x_t + \alpha_2 + \varepsilon_t$. Разобьем пары наблюдений (x_t, y_t) на две группы. В первую войдут k первых наблюдений, во вторую — оставшиеся. Найдем $\bar{x}_1 = \Sigma_{t=1}^k x_t/k$, $\bar{y}_1 = \Sigma_{t=1}^k y_t/k$,

$\bar{x}_2 = \Sigma_{t=k+1}^n x_t/(n-k)$, $\bar{y}_2 = \Sigma_{t=k+1}^n y_t/(n-k)$. Проведем через две пары точек (\bar{x}_1, \bar{y}_1) и (\bar{x}_2, \bar{y}_2) прямую. В качестве оценки (α_1, α_2) выберем

$b_1 = (\bar{y}_2 - \bar{y}_1) / (\bar{x}_2 - \bar{x}_1)$, $b_2 = \bar{y} - b_1 \bar{x}$;

а) докажите, что эта оценка является линейной и несмещенной;
 б) найдите $\sigma^2(b_1)$, $\sigma^2(b_2)$, $\text{cov}(b_1, b_2)$. Во сколько раз оценка МНК a_1 эффективнее оценки b_1 ? Когда эффективности совпадают?

8. Используя метод множителей Лагранжа, найдите несмещенную оптимальную оценку $a_1 = \Sigma c_t y_t$ параметра α_1 в парной линейной регрессии $y_t = \alpha_1 x_t + \alpha_2 + \varepsilon_t$.

9. В задаче 7 $x_t = t$, $y_t = 1/t$, $n = 10$. Найдите s^2 , несмещенную оценку матрицы ковариации оценок МНК, стандартные отклонения оценок параметров α_1 , α_2 и матрицу коэффициентов корреляций параметров.

10. Пусть $Z^{k \times m}$ — любая подматрица матрицы $X^{n \times m}$. Как ее выбрать, чтобы оценка МНК, построенная на основе матрицы Z , была наилучшей?

11. Покажите, что точность оценивания по методу наименьших квадратов увеличивается при увеличении объема выборки (покажите, что $X'_{n+1} X_{n+1} \geq X'_n X_n$, где X_n — матрица независимых переменных, построенная на выборке объема n).

1.6. Коэффициент детерминации и его интерпретация

Этот коэффициент вводится в курсах математической статистики (см. [42], [63]).

Рассмотрим $m + 1$ случайную величину y , x_1, \dots, x_m . Регрессией y на x_1, \dots, x_m в математической статистике называют условное математическое ожидание y при фиксированных значениях x_1, x_2, \dots, x_m , которое обозначают как $E(y/x_1, x_2, \dots, x_m)$. Условной дисперсией y при заданных x_1, \dots, x_m называется дисперсия y относительно регрессии $E(y/x_1, \dots, x_m)$. Таким образом,

$$\sigma^2(y/x_1, \dots, x_m) = E[(y - E(y/x_1, \dots, x_m))^2/x_1, \dots, x_m]. \quad (1.22)$$

Предположим, что условная дисперсия (1.22) не зависит от значений, которые принимают независимые переменные,

и равна σ^2 . Коэффициент детерминации определяется следующим образом:

$$r^2 = 1 - \frac{\sigma^2(y/x_1, \dots, x_m)}{\sigma^2(y)} = 1 - \frac{\sigma^2}{\sigma^2(y)}, \quad (1.23)$$

где $\sigma^2(y)$ — дисперсия случайной величины y . Можно показать, что $\sigma^2 \leq \sigma^2(y)$, поэтому $0 \leq r^2 \leq 1$, причем $r^2 = 1$ тогда и только тогда, когда y и x_1, \dots, x_m связаны функциональной зависимостью, точнее, $y = E(y/x_1, \dots, x_m)$ с вероятностью 1. Чем ближе коэффициент детерминации (1.23) к 1, тем «функциональнее» статистическая зависимость между y и x_1, \dots, x_m . Таким образом, r^2 является показателем адекватности, или качества соответствия, регрессии $E(y/x_1, x_2, \dots, x_m)$ исходной системы случайных величин y, x_1, \dots, x_m . Величина $1 - r^2$ есть доля дисперсии y , которая не смогла быть «объяснена» с помощью регрессии $E(y/x_1, \dots, x_m)$, величина r^2 — доля «объясненной» дисперсии y .

В курсах по регрессионному анализу коэффициент детерминации вводится следующим образом.

Пусть дана линейная регрессия со свободным членом

$$y_t = \alpha_1 x_{t1} + \dots + \alpha_{m-1} x_{t, m-1} + \alpha_m + \varepsilon_t, \\ t = 1, \dots, n. \quad (1.24)$$

Оценка МНК равна: $\mathbf{a} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$. (Напомним, что последний столбец матрицы \mathbf{X} состоит из единиц.) Далее можно найти расчетный вектор зависимой переменной $\hat{\mathbf{y}} = \mathbf{X}\mathbf{a}$ и оценку вектора отклонений $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\mathbf{a}$.

По определению оценка МНК удовлетворяет уравнению $-\mathbf{X}'\mathbf{y} + \mathbf{X}'\mathbf{X}\mathbf{a} = \mathbf{0}$, откуда

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}\mathbf{a}) = \mathbf{X}'\mathbf{e} = \mathbf{0}. \quad (1.25)$$

Обозначим $\mathbf{1} = (1, 1, \dots, 1)'$ — вектор размерности n , все координаты которого равны единице. Тогда, переписывая (1.25) для последнего столбца матрицы \mathbf{X} , получим $\mathbf{1}'\mathbf{e} = \sum e_i = 0$, т. е. среднее $\bar{e} = 0$. Отсюда следует, что среднее \bar{y} равно среднему $\bar{\hat{y}}$, так как $\bar{y} = \bar{\hat{y}} + \bar{e} = \bar{\hat{y}} + 0 = \bar{\hat{y}}$.

Далее, $\mathbf{e}'\hat{\mathbf{y}} = \mathbf{e}'\mathbf{X}\mathbf{a} = 0$, поэтому $\mathbf{e}'\hat{\mathbf{y}} - \hat{\bar{y}}\mathbf{e}'\mathbf{1} = 0$. Выведенные уравнения используем для разложения суммы квадратов отклонений \mathbf{y} от средней:

$$\begin{aligned} \sum_t (y_t - \bar{y})^2 &= (\mathbf{y} - \bar{y}\mathbf{1})' (\mathbf{y} - \bar{y}\mathbf{1}) = [(\hat{\mathbf{y}} - \bar{y}\mathbf{1}) + \mathbf{e}]' \times \\ &\times [(\mathbf{y} - \bar{y}\mathbf{1}) + \mathbf{e}] = (\hat{\mathbf{y}} - \bar{y}\mathbf{1})' (\hat{\mathbf{y}} - \bar{y}\mathbf{1}) + 2(\hat{\mathbf{y}} - \bar{y}\mathbf{1})' \mathbf{e} + \mathbf{e}' \mathbf{e} = \\ &= (\hat{\mathbf{y}} - \bar{y}\mathbf{1})' (\hat{\mathbf{y}} - \bar{y}\mathbf{1}) + \mathbf{e}' \mathbf{e}, \end{aligned}$$

которое перепишем как

$$\sum_t (y_t - \bar{y})^2 = \sum_t (\hat{y}_t - \bar{y})^2 + \sum_t e_t^2. \quad (1.26)$$

Таким образом, разброс объясняемой переменной около средней равен сумме разброса, «объясняемого регрессией», и разброса, который не удалось объяснить.

Коэффициент детерминации регрессии (1.24) определяется следующим образом:

$$\begin{aligned} R^2 &= \frac{\text{объясняемая сумма квадратов}}{\text{вся сумма квадратов}} = \\ &= \frac{\sum (\hat{y}_t - \bar{y})^2}{\sum (y_t - \bar{y})^2} = 1 - \frac{\sum e_t^2}{\sum (y_t - \bar{y})^2}. \end{aligned} \quad (1.27)$$

Коэффициент детерминации легко интерпретируется геометрически. Перенесем начало координат выборочного пространства R^n в точку $(\bar{y}, \bar{y}, \dots, \bar{y})' = \bar{\mathbf{y}} \in R^n$. Тогда

$$R^2 = \frac{\sum_t (\hat{y}_t - \bar{y})^2}{\sum_t (y_t - \bar{y})^2}$$

есть отношение квадрата длины катета к квадрату гипотенузы. Таким образом, коэффициент детерминации равен квадрату косинуса угла между \mathbf{y} и $\hat{\mathbf{y}}$ или между \mathbf{y} и $S_{\mathbf{X}}$.

Часто величину (1.27) интерпретируют так же, как и коэффициент детерминации r^2 (1.23), т. е. как показатель адекватности модели, как долю дисперсии \mathbf{y} , объясняемой регрессией (1.24). Однако если независимые переменные детерминируемы, то такое толкование коэффициента (1.27) недопустимо. Покажем, почему это происходит. Коэффициент (1.27) может быть переписан следующим образом:

$$1 - \frac{\sum e_t^2/n}{\sum (y_t - \bar{y})^2/n}. \quad (1.28)$$

Числитель дроби в (1.28) есть оценка условной дисперсии σ^2 , знаменатель — оценка дисперсии \mathbf{y} , причем \bar{y} — оцен-

ка математического ожидания случайной величины y . Таким образом, для того чтобы статистика $\sum (y_t - \bar{y})^2/n$ была приемлемой оценкой дисперсии y , необходимо прежде всего, чтобы $E y_t = \text{const}$. Но в силу детерминированности x_{ti} это влечет

$$E y_t = \alpha_1 x_{t1} + \dots + \alpha_{m-1} x_{t,m-1} + \alpha_m = \text{const},$$

$$t = 1, \dots, n. \quad (1.29)$$

Уравнение (1.29) означает линейную зависимость вектор-столбцов матрицы X , что противоречит предположению Е.

Забвение того факта, что y_1, y_2, \dots, y_n в регрессии (1.24) имеют разное математическое ожидание, приводит к тому, что оценка $\sum (y_t - \bar{y})^2/n$ оказывается завышенной,

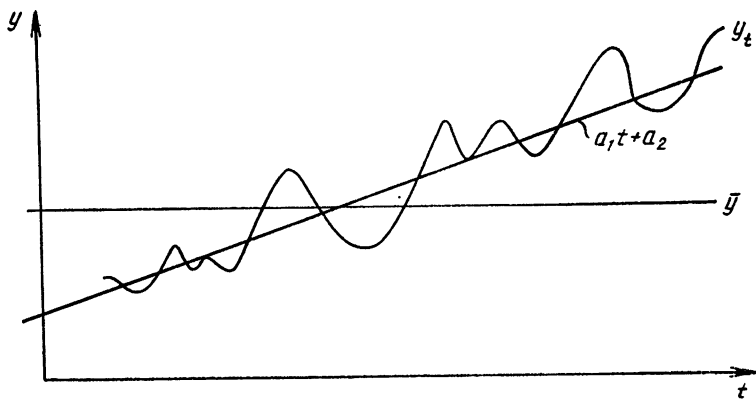


Рис. 1.10. Две модели для растущего ряда

а R^2 близким к 1. Особенно высокие значения R^2 возникают в тех случаях, когда y_t — временной монотонный (в среднем) ряд (рис. 1.10). Величина $\sum (y_t - \bar{y})^2$ будет тем больше, чем выше скорость возрастания (или убывания) ряда y_t . Таким образом, R^2 в регрессиях с монотонными зависимыми переменными скорее есть показатель не адекватности модели, а роста (падения) y_t . Автору приходилось сталкиваться с регрессиями, у которых коэффициент детерминации оказывался 0,998 и выше.

Толкование коэффициента (1.27) как показателя адекватности модели часто приводит к парадоксам, один из ко-

торых рассмотрим. Допустим, нас интересует зависимость объема выпуска некоторого предприятия от других его показателей, например от наличия основных фондов и фонда заработной платы. Имеется выборка соответствующих показателей за n лет. Предполагается следующая мультипликативная зависимость

$$y_t = e^{\alpha_1 t} K_t^{\alpha_2} L_t^{1-\alpha_2} \alpha_3' e^{\varepsilon_t}, \quad t = 1, \dots, n, \quad (1.30)$$

или

$$\ln \frac{y_t}{L_t} = \alpha_1 t + \alpha_2 \ln \frac{K_t}{L_t} + \alpha_3 + \varepsilon_t,$$

где $\alpha_3 = \ln \alpha_3'$. Предположения А—Е считаем выполненными. Коэффициент детерминации, отвечающий этой модели, обозначаем R_1^2 . Теперь рассмотрим конкурирующую модель тренда

$$y_t = e^{\beta_1 t} \beta_2' e^{\xi_t}, \quad (1.31)$$

или

$$\ln y_t = \beta_1 t + \beta_2 + \xi_t,$$

где $\beta_2 = \ln \beta_2'$.

Коэффициент детерминации, вычисленный для этой модели, обозначим через R_2^2 . Ясно, что модель (1.30) более правдоподобна, чем модель (1.31), так как выпуск предприятия должен зависеть от основных фондов и фондов заработной платы. В то же время наверняка $R_2^2 > R_1^2$. Объясним, почему это произойдет. Ряд y_t будет иметь резко выраженную тенденцию, таковым будет и ряд $\ln y_t$, поэтому $\sum_t (\ln y_t - \sum_t \ln y_t / n)^2 / n$ будет большой величиной, а R_2^2 будет близок к 1. Наоборот, ряды y_t и L_t скорее всего имеют одинаковую тенденцию, и поэтому ряд $\ln y_t / L_t$ не будет ни возрастающим, ни убывающим, т. е. R_1^2 не должен принимать очень больших значений.

Однако, что же означает R^2 , рассчитанный по формуле (1.27). Помимо исходной регрессии (1.24) рассмотрим другую регрессию — модель среднего

$$\tilde{y}_t = \beta_m + \xi_t, \quad t = 1, \dots, n. \quad (1.32)$$

Оценкой МНК уравнения (1.32) является $b_m = \bar{y} = \sum y_t / n$ с минимальной суммой квадратов, равной $\sum (y_t - \bar{y})^2$. Тогда отношение $\sum e_t^2 / \sum (y_t - \bar{y})^2$ есть показатель того, насколько модель среднего (1.32) лучше модели регрессии (1.24), т. е. показатель эффекта присутствия переменных

x_1, x_2, \dots, x_{m-1} : чем выше это отношение, тем меньше эффект введения переменных x_1, \dots, x_{m-1} в уравнение (1.32). Разность $1 - \Sigma e_t^2 / \Sigma (y_t - \bar{y})^2$ отражает, насколько модель (1.24) лучше модели среднего (1.32). Итак, в регрессиях с детерминированными независимыми переменными коэффициент детерминации необходимо трактовать как показатель, отражающий, насколько модель регрессии лучше модели среднего. Если принять такую трактовку R^2 , то становятся ясными большие значения этого коэффициента для растущих временных рядов. По определению модель среднего для таких рядов не удовлетворительна, а значит R^2 должен быть близким к 1.

Вместо коэффициента (1.27) для монотонных рядов y можно предложить другие более приемлемые показатели. Прежде всего отметим, что по тем же соображениям, по которым в уравнение регрессии мы вводили свободный член, в уравнение регрессии с монотонным рядом предлагается вводить член $\beta_{m-1}t + \beta_m$.

Поскольку y — возрастающий (убывающий) ряд, то, вероятно, таковыми будут и неучтенные факторы ε_t со средней тенденцией $\beta_{m-1}t + \beta_m$. Тогда уравнение регрессии будет выглядеть следующим образом:

$$y_t = \alpha_1 x_{t1} + \dots + \alpha_{m-2} x_{t,m-2} + \alpha_{m-1} t + \alpha_m + \varepsilon_t. \quad (1.33)$$

Вместо модели (1.32) уместно теперь рассмотреть другую, конкурирующую с (1.33), модель тренда:

$$y_t = \beta_{m-1} t + \beta_m + \xi_t. \quad (1.34)$$

Пусть b_{m-1} и b_m — оценки МНК модели-тренда (1.34). Тогда коэффициент детерминации предлагаем считать по формуле

$$R_T^2 = 1 - \Sigma e_t^2 / \Sigma (y_t - b_{m-1} t - b_m)^2,$$

где Σe_t^2 — минимальная сумма квадратов отклонений, соответствующая регрессии (1.33). R_T^2 отражает эффект присутствия переменных x_1, \dots, x_{m-2} в регрессии (1.33). Можно показать, что $0 \leq R_T^2 \leq 1$, причем если $R_T^2 = 0$, то $\alpha_1 = \dots = \alpha_{m-2} = 0$; чем R_T^2 ближе к 1, тем лучше модель (1.33) модели-тренда (1.34).

Трактуя R^2 как показатель адекватности модели исходным данным, иногда его предлагают выбирать в качестве критерия присутствия некоторой независимой переменной в регрессии. При этом не надо забывать одно обстоятельство: с добавлением новых переменных коэффициент детерми-

нации не уменьшается. Докажем это. Пусть регрессия y на x_1, \dots, x_m привела к коэффициенту детерминации R_m^2 , при этом $S_m = \langle x_1, \dots, x_m \rangle$, проекция y на S_m есть \hat{y}_m , $e_m = y - \hat{y}_m \perp \hat{y}_m$. Дополняя множество независимых переменных, приходим к множеству $S_{m+1} = \langle x_1, \dots, x_m, x_{m+1} \rangle$; проекцию y на S_{m+1} обозначим \hat{y}_{m+1} ,

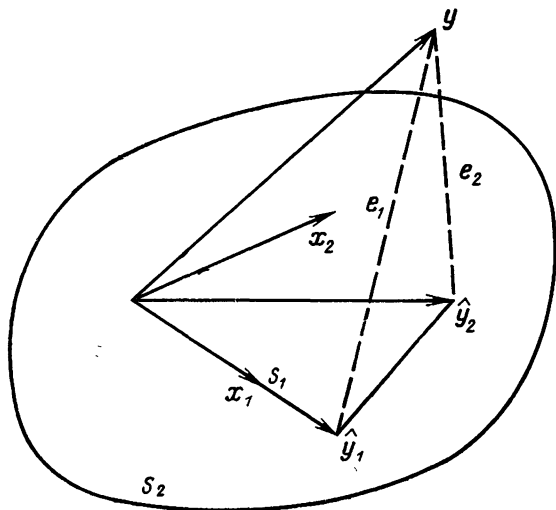


Рис. 1.11. К доказательству того, что коэффициент детерминации не уменьшается с добавлением переменных

$e_{m+1} = y - \hat{y}_{m+1}$. По определению $e_{m+1} \perp S_{m+1}$, но $\hat{y}_m - \hat{y}_{m+1} \in S_{m+1}$, поэтому $e_{m+1} \perp \hat{y}_m - \hat{y}_{m+1}$. По теореме Пифагора

$$\begin{aligned} \|e_m\|^2 &= \|y - \hat{y}_m\|^2 = \|y - \hat{y}_{m+1}\|^2 + \|\hat{y}_m - \hat{y}_{m+1}\|^2 \geq \\ &\geq \|y - \hat{y}_{m+1}\|^2 = \|e_{m+1}\|^2, \end{aligned}$$

т. е.

$$\|e_m\|^2 \geq \|e_{m+1}\|^2.$$

Из последнего неравенства следует требуемое:

$$\begin{aligned} R_m^2 &= 1 - \|e_m\|^2 / \sum (y_t - \bar{y})^2 \leq 1 - \|e_{m+1}\|^2 / \sum (y_t - \bar{y})^2 = \\ &= R_{m+1}^2. \end{aligned}$$

На рис. 1.11 показаны соответствующие величины для $m = 1, n = 3$. При этом $S_1 = \{\alpha x_1\}$, S_2 — плоскость, натянутая на x_1 и x_2 .

Часто вместо R^2 (формула (1.27)) рассматривают правленный коэффициент детерминации R_n^2 . Суть поправки сводится к тому, что вместо смещенных оценок для σ^2 и $\sigma^2(y)$ рассматривают несмещенные $s^2 = \frac{\sum e_i^2}{n-m}$ и $\hat{s}_y^2 = \frac{\sum (y_t - \bar{y})^2}{n-1}$. Таким образом,

$$R_n^2 = 1 - \frac{\sum e_i^2}{\sum (y_t - \bar{y})^2} \times (n-1)/(n-m) = (R^2(n-1) - m + 1)/(n-m). \quad (1.35)$$

Правленный коэффициент детерминации является лучшей оценкой истинного значения коэффициента детерминации, чем обычный коэффициент (1.27).

Для регрессии без свободного члена неявно предполагается, что $\bar{y} = 0$, поэтому формула (1.27) переписывается следующим образом:

$$R^2 = 1 - \frac{\sum e_i^2}{\sum y_i^2}. \quad (1.36)$$

Коэффициент детерминации (1.36) отражает, насколько модель без свободного члена $\hat{y}_t = a_1 x_{t1} + \dots + a_m x_{tm}$ лучше модели $\hat{y}_t = 0$.

У п р а ж н е н и я 1.6

1. Покажите, что коэффициент детерминации (1.27) в парной регрессии $y_t = \alpha_1 x_t + \alpha_2 + \varepsilon_t$ равен квадрату коэффициента корреляции между y и x .

2. Что геометрически означает $R^2 = 0$ и $R^2 = 1$?

3. Покажите, что оценка МНК соответствует максимальному значению R^2 (1.27).

4. Покажите, что $0 \leq R_T^2 \leq 1$. Что означает $R_T^2 = 0$ и $R_T^2 = 1$?

5. Верно ли, что $R_T^2 \leq R^2$?

6. Покажите, что $R_n^2 \leq R^2$.

7. Введите понятие правленного коэффициента R_T^2 .

8. В каком случае введение нового фактора в регрессию не изменит коэффициента детерминации? Дайте геометрическую интерпретацию.

9. Докажите, что коэффициент детерминации не зависит от выбора масштаба измерения переменных.

1.7. Состоятельность и асимптотическая нормальность оценки МНК

Рассмотрим поведение оценки МНК при увеличении числа наблюдений, т. е. остановимся на ее асимптотических свойствах.

По-видимому, самым слабым, и поэтому самым желательным, необходимым свойством любой оценки является состоятельность. Под состоятельностью оценки понимается возрастающая до бесконечности точность оценивания при увеличении числа наблюдений. Таким образом, статистика a_n состоятельно оценивает α (здесь индекс n указывает на то, что оценка a построена на основе n первых наблюдений y_1, y_2, \dots, y_n), если при $n \rightarrow \infty$ разброс a_n около истинного значения α стремится к нулю.

Исходя из различных толкований понятия «разброс» получают различные виды состоятельности. Приведем некоторые, наиболее часто встречающиеся виды состоятельности (сходимости).

1. *Просто состоятельность, или слабая состоятельность*, опирается на понятие сходимости случайных величин по вероятности. Последовательность (одномерных) оценок a_1, a_2, \dots состоятельно оценивает (одномерный) параметр α , если для любого числа $\varepsilon > 0$ вероятность того, что $|a_n - \alpha| > \varepsilon$, стремится к нулю при $n \rightarrow \infty$:

$$P\{|a_n - \alpha| > \varepsilon\} \rightarrow 0, \quad n \rightarrow \infty.$$

Это записываем как $\text{plim}_{n \rightarrow \infty} a_n = \alpha$.

2. *Сильная состоятельность* опирается на понятие сходимости с вероятностью 1. Так, последовательность a_1, a_2, \dots сильно состоятельно оценивает α , если вероятность того, что $\lim_{n \rightarrow \infty} a_n = \alpha$, равна 1, т. е.

$$P\{\lim_{n \rightarrow \infty} a_n = \alpha\} = 1.$$

3. *Состоятельность в среднем квадратичном*. a_n сходится к (оценивает) α в среднем квадратичном, если м. о. квадрата отклонения стремится к нулю при $n \rightarrow \infty$, т. е.

$$E(a_n - \alpha)^2 \rightarrow 0, \quad n \rightarrow \infty.$$

Это записываем как $\text{l.i.m.}_{n \rightarrow \infty} a_n = \alpha$.

И хотя эти три определения состоятельности различны с математической точки зрения, все они отражают одну качественную картину — «сведение на нет» разброса оценки около истинного значения при $n \rightarrow \infty$.

Первое определение самое слабое: состоятельность в среднем квадратичном и сильная состоятельность влекут

состоятельность в слабом смысле. Первое утверждение следует из неравенства Чебышева

$$P \{ |a_n - \alpha| > \varepsilon \} \leq \frac{E (a_n - \alpha)^2}{\varepsilon^2}. \quad (1.37)$$

При любом $\varepsilon > 0$, если $n \rightarrow \infty$, левая часть (1.37) стремится к нулю. Второе утверждение здесь доказывать не будем¹. Если нет дополнительной информации, нельзя доказать, что сильная состоятельность влечет состоятельность в среднем квадратичном, или наоборот.

В статистике чаще используют слабую состоятельность. Однако часто легче доказать состоятельность в среднем квадратичном. Состоятельность тогда следует из неравенства Чебышева (1.37). Многомерная статистика состоятельно оценивает многомерный параметр α в каком-либо смысле 1—3, если соответствующая состоятельность имеет место для каждой координаты α_i . В частности, легко показать, что $a_n \rightarrow \alpha \in R^m$ в среднем квадратичном тогда и только тогда, когда $E (a_n - \alpha) (a_n - \alpha)' \rightarrow 0$ — нулевая матрица $m \times m$. Этим фактом мы часто будем пользоваться.

Очевидно, для доказательства состоятельности оценки МНК необходимо сделать какие-либо ограничения на рост независимых переменных, т. е. на матрицу X_n . (В этом параграфе оценку МНК a и матрицу X будем сопровождать индексом n .) Как правило, используют условие сильной регулярности матриц X_n .

Сильная регулярность независимых переменных:

$$\lim_{n \rightarrow \infty} \frac{1}{n} X_n' X_n = A, \quad (1.38)$$

где A — невырожденная конечная матрица $m \times m$.

Теорема 1.4. *Если предположения А—Е выполнены, а матрицы X_n сильно регулярны, то оценка МНК состоятельна в среднем квадратичном.*

Доказательство. Теорема будет доказана, если мы докажем, что

$$E (a_n - \alpha) (a_n - \alpha)' = \text{cov} (a_n) = \sigma^2 (X_n' X_n)^{-1} \rightarrow 0, \\ n \rightarrow \infty.$$

¹Доказательство можно найти, например, в [63].

Из условия регулярности X_n и невырожденности A следует:

$$\lim_{n \rightarrow \infty} n (X_n' X_n)^{-1} = A^{-1},$$

поэтому $\lim_{n \rightarrow \infty} (X_n' X_n)^{-1} = 0$, что доказывает теорему.

Условие сильной регулярности является слишком ограниченительным. Оно, в частности, не выполняется в регрессиях на время, т. е. при выделении трендов. Можно доказать, что если r_1, r_2, \dots — последовательность неотрицательных чисел, причем

$$\lim_{n \rightarrow \infty} \bar{r}_n = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n r_t < \infty,$$

то последовательность r_1, r_2, \dots не стремится к ∞ , т. е. можно выбрать ограниченную подпоследовательность. Если последовательность r_1, r_2, \dots возрастающая, то ограниченность средних \bar{r}_n влечет ограниченность самой последовательности. Данные рассуждения могут быть перенесены на элементы матрицы X_n , если положить $r_t = x_{it}^2$.

Пример. Рассмотрим регрессию на время

$$y_t = \alpha t + \varepsilon_t, \quad t = 1, 2, \dots, n.$$

Оценка МНК $a = \Sigma y_t t / \Sigma t^2$ состоятельна, так как

$$E(a - \alpha)^2 = \sigma^2 (a) = \sigma^2 / \Sigma t^2 = 6\sigma^2 / [n(n+1) \times \\ \times (2n+1)] \rightarrow 0, \quad n \rightarrow \infty.$$

С другой стороны, условие сильной регулярности не выполняется:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \Sigma t^2 = \lim_{n \rightarrow \infty} \frac{(n+1)(2n+1)}{6} = \infty.$$

Аналогично можно показать, что в любой регрессии, в которой присутствует монотонно возрастающий фактор времени, предположение (1.38) неверно.

Существует необходимое и достаточное условие состоятельности оценки МНК, которое выражается через характеристические числа матрицы $X_n' X_n$.

Условие Эйкера [97]: минимальное характеристическое число матрицы $X_n' X_n$ стремится к $+\infty$ при $n \rightarrow \infty$:

$$\lambda_{\min}(X_n' X_n) \rightarrow \infty, \quad n \rightarrow \infty. \quad (1.39)$$

Теорема 1.5. Пусть предположения А—Е выполняются для всех n , начиная с некоторого $n_0 \geq t$. Тогда условие Эйкера (1.39) эквивалентно состоятельности оценки МНК в среднем квадратичном.

Доказательство этой теоремы дано в параграфе 1.11.

Насколько жестким является условие состоятельности оценки МНК? Насколько «вероятно» его выполнение на практике? Для того чтобы ответить на эти вопросы, заметим, что $\lambda_{\min}(\mathbf{X}'_n \mathbf{X}_n)$ есть квадрат минимальной длины вектора, являющегося линейной комбинацией вектор-столбцов x_1, x_2, \dots, x_m матрицы \mathbf{X}_n . Более строго, пусть $(\omega_1, \omega_2, \dots, \omega_m)' = \mathbf{w}$ — вектор коэффициентов линейной комбинации, $\|\mathbf{w}\| = 1$, тогда

$$\begin{aligned} \min_{\|\mathbf{w}\|=1} \|\mathbf{X}_n \mathbf{w}\|^2 &= \min_{\|\mathbf{w}\|=1} (\mathbf{X}_n \mathbf{w})' (\mathbf{X}_n \mathbf{w}) = \\ &= \min_{\|\mathbf{w}\|=1} \mathbf{w}' \mathbf{X}'_n \mathbf{X}_n \mathbf{w} = \lambda_{\min}(\mathbf{X}'_n \mathbf{X}_n). \end{aligned}$$

Таким образом, $\lambda_{\min}(\mathbf{X}'_n \mathbf{X}_n)$ можно трактовать как показатель вырожденности $\mathbf{X}'_n \mathbf{X}_n$ или как меру линейной зависимости переменных x_1, x_2, \dots, x_m (более подробно об измерении степени линейной зависимости см. параграф 5.1). Поэтому можно сказать, что оценка МНК состоятельна тогда и только тогда, когда степень линейной независимости x_1, x_2, \dots, x_m растет до бесконечности.

Использовать на практике условие Эйкера весьма сложно. Часто легче непосредственно установить факт $(\mathbf{X}'_n \mathbf{X}_n)^{-1} \rightarrow \mathbf{0}$, чем доказать (1.39). Можно предложить более простой критерий состоятельности оценки МНК. Для этого построим для матрицы \mathbf{X}_n матрицу сопряженности \mathbf{R}_n . По определению

$$(\mathbf{R}_n)_{ij} = \frac{\sum_t x_{ti} x_{tj}}{\sqrt{\sum_t x_{ti}^2 \cdot \sum_t x_{tj}^2}}, \quad i, j = 1, \dots, m. \quad (1.40)$$

Другими словами, $(\mathbf{R}_n)_{ij}$ есть косинус угла между векторами x_i и x_j в евклидовом пространстве R^n . Матрица сопряженности похожа на матрицу корреляций x_1, \dots, x_m , отличие лишь в том, что в матрице корреляций рассматриваются отклонения x_{ti} от соответствующей средней \bar{x}_i . Выбирая термин «матрица сопряженности», мы тем самым подчеркиваем, что x_1, \dots, x_m не случайные векторы, как это необходимо считать при вычислении матрицы корреляций.

Более точно, R_n есть матрица парных коэффициентов сопряженности.

Теорема 1.6. *Если:*

а) $\sum x_{it}^2 \rightarrow \infty$ для любого $i = 1, \dots, m$ при $n \rightarrow \infty$,

б) $R_n \rightarrow R$, $|R| = 0$,

то оценка МНК состоятельна (в среднем квадратичном).

Доказательство этой теоремы дано в параграфе 1.11.

Пример. Докажем с помощью теоремы 1.6 состоятельность оценки МНК в регрессии на время $y_t = \alpha_1 t + \alpha_2 + \varepsilon_t$, $t = 1, \dots, n$. Применим формулы $\sum t = n(n+1)/2$, $\sum t^2 = n(n+1)(2n+1)/6$. Условие а) очевидно выполняется, проверим выполнимость условия б).

Для данной регрессии

$$X_n' X_n = \begin{bmatrix} \frac{n(n+1)(2n+1)}{6} & \frac{n(n+1)}{2} \\ \frac{n(n+1)}{2} & n \end{bmatrix};$$

$$R_n = \begin{bmatrix} 1 & \frac{\sqrt{6}}{2} \sqrt{\frac{n+1}{2n+1}} \\ \frac{\sqrt{6}}{2} \sqrt{\frac{n+1}{2n+1}} & 1 \end{bmatrix} \rightarrow$$

$$\rightarrow \begin{bmatrix} 1 & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & 1 \end{bmatrix} = R, \quad |R| \neq 0,$$

т. е. условие б) теоремы 1.6 выполнено, значит, оценка МНК состоятельна.

Говоря об оценке МНК, мы имеем в виду оценку параметров регрессии α . Однако имеется еще один неизвестный параметр σ^2 . В качестве несмещенной оценки этого параметра мы предлагаем статистику s^2 (1.20). Замечательно, что состоятельность s^2 верна без какого-либо предположения об изменении матрицы X_n .

Теорема 1.7 [146]. Пусть все предположения А—Е выполнены для всех $n \geq n_0$. Тогда $\lim_{n \rightarrow \infty} s^2 = \sigma^2$.

Доказательство см. в параграфе 1.11.

Прежде чем приступить к исследованию асимптотической нормальности оценки МНК, дадим определение асимптотически нормальной последовательности оценок. Пусть

d_1, d_2, \dots — последовательность некоторых одномерных оценок. Эта последовательность асимптотически нормальна, если найдутся такие константы α_n, β_n ($\beta_n > 0$), что оценка $b_n = (d_n - \alpha_n)/\beta_n$ сходится по распределению к случайной нормальной величине $N(0, 1)$. Если математическое ожидание и дисперсия d_n конечны, то их можно использовать в качестве нормирующих констант, т. е. положить $\alpha_n = E d_n, \beta_n^2 = \sigma^2(d_n)$, тогда $E b_n = 0, \sigma^2(b_n) = 1$.

Аналогичное определение вводится и в многомерном случае; здесь нормирующими константами будут невырожденная матрица и вектор. Как правило, мы выбираем квадратный корень из матрицы¹ ковариаций d_n и вектор математического ожидания, тогда $b_n = S_n^{-1/2}(d_n - \mu_n)$, где $S_n = \text{cov}(d_n), \mu_n = E d_n$ и предельное распределение есть $N(0, I)$.

Легко проверить, что стандартизованной по такому правилу оценкой МНК является

$$b_n = \frac{1}{\sigma} (X_n' X_n)^{-1/2} X_n' \varepsilon = A_n \varepsilon, \quad (1.41)$$

$$\text{где} \quad A_n = \frac{1}{\sigma} (X_n' X_n)^{-1/2} X_n' \quad (1.42)$$

— матрица $m \times n$.

Теорема 1.8. *Допустим, $\{\varepsilon_i\}$ независимы и одинаково распределены. Оценка МНК асимптотически нормальна тогда и только тогда, когда для каждого $i = 1, \dots, m$*

$$\max \{A_{ni1}^2, A_{ni2}^2, \dots, A_{nin}^2\} \rightarrow 0, \quad n \rightarrow \infty, \quad (1.43)$$

где A_{nit} — (it) -й элемент матрицы A_n .

Доказательство см. в параграфе 1.11.

З а м е ч а н и я: 1. Так же как и в состоятельности, здесь накладывается определенное ограничение на изменение независимых переменных при $n \rightarrow \infty$. Это ограничение определяется условием (1.43), которое отражает равномерное убывание элементов матрицы A_n .

2. Для доказательства асимптотической нормальности оценки МНК мы требуем выполнения более жесткого условия, чем предположение Г: независимости и одинаковой распределенности отклонений.

¹Под квадратным корнем положительно определенной матрицы A понимается квадратная положительно определенная матрица $B = A^{1/2}$, такая, что $B^2 = A$. Можно доказать, что такая матрица существует и единственна [58].

Широко распространено ошибочное мнение, что оценка МНК асимптотически нормально распределена в условиях сильной регулярности матриц X_n (1.38). Э. Маленво даже «доказывает», что если отклонения независимы и одинаково распределены, а матрицы X_n сильно регулярны, то оценка МНК асимптотически нормальна. Его доказательство содержит ошибку [48, с. 232]. Дело в том, что из условия (1.38) не следует, что $\max_i x_{ii}^2/n$ стремится к нулю при $n \rightarrow \infty$

для любого i , а именно последнее условие необходимо для применения центральной предельной теоремы при доказательстве асимптотической нормальности оценки МНК. В параграфе 1.11 приведен соответствующий пример.

Т е о р е м а 1.9. Если $\{\varepsilon_i\}$ независимы и одинаково распределены, матрицы X_n сильно регулярны (уравнение (1.38)), причем для любого $i = 1, 2, \dots, m$ $\frac{1}{n} \max_{t=1, \dots, n} x_{it}^2 \rightarrow 0$, $n \rightarrow \infty$, то оценка МНК асимптотически нормальна, более того,

$$\sqrt{n} (a_n - \alpha) \xrightarrow{d} N(0, \sigma^2 A^{-1}).$$

Д о к а з а т е л ь с т в о. Докажем, что условия теоремы приводят к (1.43). Действительно,

$$A_n = \frac{1}{\sigma} (X_n' X_n)^{-1/2} X_n' = \frac{1}{\sigma} \left(\frac{X_n' X_n}{n} \right)^{-1/2} \frac{X_n'}{\sqrt{n}}.$$

Но, как следует из условия теоремы,

$$\frac{1}{\sigma} \left(\frac{X_n' X_n}{n} \right)^{-1/2} \rightarrow \frac{1}{\sigma} A^{-1/2}, \quad n \rightarrow \infty.$$

Из второго условия теоремы следует $\frac{1}{n} \max_{t=1, \dots, n} x_{it}^2 \rightarrow 0$, $n \rightarrow \infty$, что окончательно ведет к выполнению (1.43). Доказательство последнего утверждения теоремы предоставляем читателю.

П р и м е р. Рассмотрим регрессию $y_t = \alpha x_t + \varepsilon_t$, где $x_t = (\sqrt{d})^t$, $t = 1, 2, \dots, d$ — некоторое число, большее 1. Тогда

$$\sum_t x_t^2 = \sum_t d^t = d \frac{d^n - 1}{d - 1} \rightarrow \infty, \quad n \rightarrow \infty,$$

т. е. (1.39) выполнено, но

$$\max_t \frac{x_i^2}{\sum_t x_i^2} = \frac{1}{\sum_t x_i^2} \max_t x_i^2 = \frac{d^n}{d^n - 1} \cdot \frac{d-1}{d} \rightarrow \frac{d-1}{d} > 0,$$

т. е. (1.43) не выполнено. Значит, оценка МНК состоятельна, но не асимптотически нормальна. В данном примере это объясняется быстрым (экспоненциальным) ростом ряда x_t .

Применение теоремы 1.8 довольно затруднительно. Приведем более простое условие — достаточный критерий Андерсона [6, с. 35—37].

Т е о р е м а 1.10 (достаточный критерий асимптотической нормальности). Если $\{\varepsilon_t\}$ независимы, одинаково распределены и

а) $R_n \rightarrow R$, $|R| \neq 0$ (см. уравнение (1.40)),

б) $\max_t x_{it}^2 / \sum_t x_{it}^2 \rightarrow 0$ для любого $i = 1, \dots, m$ при $n \rightarrow \infty$, то оценка МНК асимптотически нормальна. Доказательство см. в параграфе 1.11.

П р и м е р. Докажем, применяя теорему 1.10, асимптотическую нормальность оценки МНК в регрессии на время: $y_t = \alpha_1 t + \alpha_2 + \varepsilon_t$, $t = 1, \dots, n$. Условие а) было проверено ранее при доказательстве состоятельности оценки. Рассмотрим условие б):

$$i = 1; \frac{\max_t x_{it}^2}{\sum_t x_{it}^2} = \frac{6n^2}{n(n+1)(2n+1)} \rightarrow 0, \quad n \rightarrow \infty,$$

$$i = 2; \frac{\max_t x_{it}^2}{\sum_t x_{it}^2} = \frac{1}{n} \rightarrow 0, \quad n \rightarrow \infty.$$

Условия теоремы выполняются, значит, оценка МНК асимптотически нормальна.

Что дают нам теоремы о состоятельности и асимптотической нормальности оценки МНК? С теоретической точки зрения первое свойство заключается в том, что с ростом числа наблюдений при определенных ограничениях на независимые переменные точность оценивания бесконечно возрастает. Второе свойство делает возможным при больших n и опять же при определенных условиях на независимые переменные считать распределение оценки МНК приблизительно нормальным $N(\alpha, \sigma^2 (X_n' X_n)^{-1})$. Это свойство

является весьма важным, так как построение удовлетворительных доверительных интервалов и проверка гипотез относительно параметров регрессии возможны только при известном распределении отклонений регрессии. Вместе с тем использовать условия состоятельности и асимптотической нормальности оценки МНК можно практически только в регрессиях, где правая часть есть функция времени (t), т. е. в регрессиях-трендах. В регрессиях планируемого эксперимента эксперименты необходимо ставить таким образом, чтобы с ростом n условия (1.39) и (1.43) выполнялись. Тогда оценка МНК будет сходиться к истинному значению параметров α , и распределение оценки будет близко к нормальному.

У п р а ж н е н и я 1.7

1. Докажите, что если $\text{plim } a_n = \alpha$, $\text{plim } b_n = \beta$, то $\text{plim } (a_n + b_n) = \alpha + \beta$, $\text{plim } a_n b_n = \alpha\beta$. Перенесите эти формулы на многомерный случай. Верны ли эти формулы для других видов сходимости?

2. Пусть a_n — m -мерная случайная величина. Докажите, что $\text{plim } a_n = \alpha$ тогда и только тогда, когда $\text{plim } \|a_n - \alpha\| = 0$.

3. Постройте последовательность случайных величин, которая сходится в смысле определений 1 и 2, но не сходится в смысле определения 3.

4. Покажите, что матрица A в условии (1.38) положительно определена.

5. Докажите непосредственно состоятельность оценки МНК в регрессии $y_t = \alpha_1 t + \alpha_2 + \varepsilon_t$. Выполняется ли условие регулярности независимых переменных? Используя теорему 1.6, докажите состоятельность оценки МНК в параболической регрессии $y_t = \alpha_1 t^2 + \alpha_2 t + \alpha_3 + \varepsilon_t$.

6. Является ли ограниченность матриц X_n достаточным условием для ее регулярности?

7. Покажите, что из условия сильной регулярности следует условие Эйкера.

8. Будут ли оценки МНК состоятельны для следующих регрессий:

$$\begin{aligned} \text{а) } y_t &= \alpha + \varepsilon_t; & \text{б) } y_t &= \alpha_1 / \sqrt{t} + \varepsilon_t; & \text{в) } y_t &= \alpha_1 / t + \alpha_1 + \varepsilon_t; \\ \text{г) } y_t &= \alpha_1 t + \alpha_2 / t + \varepsilon_t; & \text{д) } y_t &= \alpha_1 e^{\beta t} + \varepsilon_t; & \text{е) } y_t &= \alpha_1 \sin \omega t + \varepsilon_t, \end{aligned}$$

β ; ω — некоторые действительные числа.

9. Проверьте выполнение теоремы 1.6 для регрессии $y_t = (1 + 1/t)\alpha_1 + \alpha_2 + \varepsilon_t$. Состоятельна ли оценка МНК?

10. Докажите асимптотическую нормальность оценки МНК для регрессий на время: а) $y_t = \alpha t + \varepsilon_t$; б) $y_t = \alpha_1 t + \alpha_2 + \varepsilon_t$; в) $y_t = \alpha_1 t^2 + \alpha_2 t + \alpha_3 + \varepsilon_t$.

11. Для каких регрессий из задачи 8 оценка МНК будет асимптотически нормальной?

12. Будут ли и при каких условиях состоятельны и асимптотически нормальны оценки из задач 6 и 7 упражнения 1.5?

1.8. Свойства оценки МНК при нормальных отклонениях

До сих пор мы не делали какого-либо предположения относительно распределения отклонений ε_t . Сейчас мы предположим, что эти отклонения распределены по нормальному закону с нулевым математическим ожиданием и одинаковой неизвестной дисперсией σ^2 . Имея конкретное распределение для ε_t , а значит и для y_t , мы можем применить метод максимального правдоподобия (ММП).

Теорема 1.11. Если $\varepsilon_t \sim N(0, \sigma^2)$, то оценки МНК и ММП совпадают.

Доказательство. Из условия теоремы и независимости $\{\varepsilon_t\}$ следует $\xi \sim N(0, \sigma^2 I_n)$, значит, $y \sim N(X\alpha, \sigma^2 I_n)$ с плотностью, равной

$$f(y; \alpha, \sigma^2) = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \times \\ \times \exp \left[-\frac{1}{2\sigma^2} (y - X\alpha)' (y - X\alpha) \right]. \quad (1.44)$$

Оценка ММП соответствует максимуму этой функции. Возьмем логарифм этой функции, отбросим константу, не влияющую на максимизацию, и поменяем знак. Получим

$$\psi(y; \alpha, \sigma^2) = n \ln \sigma^2 + \frac{1}{\sigma^2} (y - X\alpha)' (y - X\alpha) \Rightarrow \min.$$

Необходимым условием минимума является обращение первых производных функции в точке минимума в нуль. Найдем производную по α :

$$\partial \psi / \partial \alpha = \frac{1}{\sigma^2} (-2X' y + 2X' X\alpha) = 0.$$

Решением этого уравнения и будет оценка ММП $\hat{\alpha} = (X'X)^{-1} X'y$, которая совпадает с оценкой МНК.

Итак, если отклонения нормально распределены, то оценка МНК совпадает с общестатистическим методом оценивания — методом максимального правдоподобия.

Иногда, ссылаясь на совпадение оценок МНК и ММП в условиях нормальных отклонений, утверждают, что все известные оптимальные свойства ММП приобретает в этом случае и МНК: состоятельность, асимптотическую нормальность и эффективность. Без дополнительных оговорок подобный перенос свойств оценок для регрессий не имеет места. Дело в том, что указанные асимптотические свойства

ва оценок ММП доказываются в условиях *одинакового распределения* наблюдений y_1, y_2, \dots, y_n . В регрессиях *наблюдения распределены неодинаково*, они имеют разные математические ожидания:

$$E y_t = \alpha_1 x_{t1} + \alpha_2 x_{t2} + \dots + \alpha_m x_{tm} \neq \text{const}, \quad t = 1, 2, \dots$$

Поэтому и состоятельность, и асимптотическую нормальность оценок метода наименьших квадратов необходимо перепроверять. Для того чтобы указанные асимптотические свойства выполнялись, необходимо наложить определенные ограничения на поведение независимых переменных на бесконечности, т. е. при увеличении объема выборки, что и было сделано в предыдущем параграфе.

Наиболее ценным с практической точки зрения является то, что в условиях нормальной гипотезы распределение оценки МНК принимает конкретный вид и является тоже нормальным.

Теорема 1.12. Если $\varepsilon \sim N(0, \sigma^2 I_n)$, то

а) оценка МНК имеет нормальное распределение $N(\alpha, \sigma^2 (X'X)^{-1})$, т. е. распределена по нормальному закону с м. о. α и матрицей ковариаций $\sigma^2 (X'X)^{-1}$;

б) статистика $(y - Xa)'(y - Xa)/\sigma^2 = (n - m)s^2/\sigma^2$ распределена по χ^2 с $n - m$ степенями свободы;

в) оценки a и s^2 независимы.

Доказательство. а) Оценка МНК $a = (X'X)^{-1}X'y$ является линейной по y , значит, имеет нормальное распределение. Ранее было показано, что $Ea = \alpha$, $\text{cov}(a) = \sigma^2 (X'X)^{-1}$, поэтому распределение a есть $N(\alpha, \sigma^2 (X'X)^{-1})$. б) Имеем

$$\begin{aligned} \varepsilon &= y - Xa = [I_n - X(X'X)^{-1}X']\varepsilon, \\ (y - Xa)'(y - Xa) &= \varepsilon'[I_n - X(X'X)^{-1}X'] [I_n - \\ &\quad - X(X'X)^{-1}X']\varepsilon = \varepsilon' A \varepsilon, \end{aligned}$$

где $A = I_n - X(X'X)^{-1}X'$ — идемпотентная матрица¹, причем

$$\begin{aligned} \text{tr } A &= \text{tr } I_n - \text{tr } (X(X'X)^{-1}X') = n - \text{tr } (X'X)^{-1}X'X = \\ &= n - \text{tr } I_m = n - m. \end{aligned}$$

Теперь воспользуемся утверждением (П.14) приложения, что и доказывает б).

¹Квадратная симметричная матрица A называется идемпотентной, если $A^2 = A$ [61].

в) Применим теорему (П.15). Для доказательства независимости \mathbf{a} и s^2 достаточно показать, что $\mathbf{V}\mathbf{A} = \mathbf{0}$, где $\mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. И действительно,

$$\mathbf{V}\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{0}.$$

Значение доказанной теоремы велико. Факт нормальности \mathbf{a} , независимости \mathbf{a} и s^2 дает нам возможность проверять статистические гипотезы, строить критерии, находить доверительные интервалы.

Большое значение имеет также следующая теорема.

Теорема 1.13. *Если отклонения нормальны, то оценка МНК является эффективной в классе всех несмещенных оценок с минимальной матрицей ковариаций, равной $\text{cov}(\mathbf{a}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.*

Доказательство см. в параграфе 1.11.

Теорема 1.13 утверждает, что если \mathbf{b} — другая несмещенная оценка $\boldsymbol{\alpha}$, то $\text{cov}(\mathbf{b}) \geq \text{cov}(\mathbf{a}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

Существует много критериев проверки гипотезы о нормальном распределении отклонений. Однако они требуют большого числа наблюдений n . С некоторыми из этих критериев можно познакомиться в [1].

У п р а ж н е н и я 1.8

1. Покажите, что оценкой ММП σ^2 при нормальных отклонениях будет $s^2 = (\mathbf{y} - \mathbf{X}\mathbf{a})'(\mathbf{y} - \mathbf{X}\mathbf{a})/n$.

2. Докажите, что необходимое условие минимума функции ψ является в данном случае и достаточным, т. е. что ψ имеет единственный минимум относительно $\boldsymbol{\alpha}$ и σ^2 .

3. Как найти оценку ММП для линейной регрессии, если отклонения подчиняются закону распределения Лапласа?

4. Методом максимального правдоподобия найдите оценку неизвестного параметра λ в распределении Пуассона $P\{x = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, 2, \dots$ Будет ли эта оценка несмещенной? Достигает ли она нижней границы Крамера—Рао?

1.9. Общие принципы проверки статистических гипотез и построения доверительных интервалов¹

Допустим, распределение n -мерной случайной величины \mathbf{y} зависит от некоторого (может быть многомерного) неизвестного параметра $\boldsymbol{\alpha}$, который принадлежит априори

¹Основные принципы проверки статистических гипотез более подробно изложены, например, в [45, 30 и 19].

заданному множеству Θ . Далее, задано разбиение множества Θ на два подмножества Θ_H и Θ_K ($\Theta_H \cup \Theta_K = \Theta, \Theta_H \cap \Theta_K = \emptyset$). Выдвигаем гипотезу H , которую мы впоследствии будем проверять: истинное значение параметра α принадлежит множеству Θ_H . Если Θ_H состоит из одной точки, гипотезу H называют простой. Каждый раз, когда значение случайной величины y известно (т. е. имеется выборка), мы должны определенно ответить, верна наша гипотеза или нет. Таким образом, выборочное пространство R^n разбивается в свою очередь на два подмножества E_H и E_K ($E_H \cup E_K = R^n, E_H \cap E_K = \emptyset$). Если y попадает в E_H , то гипотезу H принимаем, если $y \in E_K$ — отвергаем. При этом мы можем совершить ошибку двух родов: 1) гипотеза H верна, а мы ее отвергаем, 2) гипотеза H неверна, а мы ее принимаем.

Ошибки могут иметь разные последствия. Например, если гипотеза H состоит в наличии у пациента некоторой тяжелой болезни, то ошибка 2) не так существенна, как 1). Одновременно свести обе ошибки к минимуму невозможно. Целесообразно поступить следующим образом. Зададим верхнюю границу λ максимальной ошибки первого рода и при этом условии будем минимизировать ошибку второго рода; λ называют *уровнем значимости*. Как же подсчитать ошибку, совершаемую нами для каждого y ? Разумеется, в такой постановке ответить на такой вопрос невозможно. Однако мы можем подсчитать вероятность совершения ошибок первого и второго родов. Вероятность ошибки первого рода:

$$P_\alpha(\text{гипотеза } H \text{ неверна}) = P_\alpha(y \in E_K) = P_\alpha(E_K) \leq \lambda, \quad (1.45)$$

где α — истинное значение параметра, принадлежащее Θ_H (гипотеза на самом деле верна). Вероятность ошибки второго рода:

$$P_\alpha(\text{гипотеза } H \text{ верна}) = P_\alpha(y \in E_H) = 1 - P_\alpha(E_K) \Rightarrow \min,$$

где $\alpha \in \Theta_K$ (гипотеза на самом деле неверна). Минимизация предыдущего выражения эквивалентна максимизации вероятности отвергнуть гипотезу, когда она на самом деле неверна, т. е.

$$P_\alpha(E_K) \Rightarrow \max, \alpha \in \Theta_K. \quad (1.46)$$

Множество E_K называют *критической областью критерия*, или *критическим множеством*. Функцию

$$\beta(\alpha) = P_\alpha(E_K), \alpha \in \Theta_K \quad (1.47)$$

называют *функцией мощности*. Таким образом, критерий проверки гипотезы H полностью определяется подмножеством E_K выборочного пространства R^n . Множество E_K , разумеется, не должно зависеть от неизвестного параметра α .

Пример. Рассмотрим пример из параграфа 1.4. Допустим, мы хотим проверить простую гипотезу $H: \mu = 0$. В наших обозначениях $\Theta = \{-\infty < \mu < \infty\} = R^1$, $\Theta_H = \{\mu = 0\}$, $\Theta_K = \{\mu \neq 0\}$: Допустим, мы задались некоторым уровнем значимости λ (как правило, выбирают $\lambda = 0,05$ или $\lambda = 0,01$). В качестве критического множества рассмотрим множество $E_K = \{y \in R^n : |\Sigma y_i/n| > \varphi\}$, где φ — некоторое фиксированное число, зависящее от λ . Найдем функцию мощности предложенного критерия

$$\beta(\mu) = P_\mu(E_K) = (2\pi)^{-\frac{n}{2}} \int_{E_K} \exp\left[-\frac{1}{2} \Sigma (y_i - \mu)^2\right] dy_1 dy_2 \dots dy_n.$$

Заметим, что $\Sigma y_i/n \sim N(\mu, 1/n)$, поэтому

$$\begin{aligned} \beta(\mu) &= \frac{1}{\sqrt{2\pi/n}} \int_{|z| > \varphi} \exp\left[-\frac{n}{2} (z - \mu)^2\right] dz = \\ &= \frac{1}{\sqrt{2\pi/n}} \int_{|z - \mu| > \varphi} \exp\left(-\frac{nz^2}{2}\right) dz. \end{aligned}$$

В точке $\mu = 0$ функция мощности имеет минимум. Это очевидно, так как если истинное значение μ близко к нулю, то мы легко можем совершить ошибку второго рода (т. е. принять гипотезу $\mu = 0$, тогда как $\mu \neq 0$). Вероятность ошибки первого рода равна:

$$P_0(E_K) = P_{\mu=0}(E_K) = \frac{1}{\sqrt{2\pi/n}} \int_{|z| > \varphi} \exp\left(-\frac{nz^2}{2}\right) dz \leq \lambda.$$

Обозначим функцию распределения статистики $\Sigma y_i/n$ при $\mu = 0$ через $F(t)$, тогда $P_0(E_K) = 2F(-\varphi)$. Для заданного λ φ является решением уравнения $\lambda = P_0(E_K) = 2F(-\varphi)$ (рис. 1.12). Значение φ находится из таблиц нормального распределения.

Естественный путь сравнения различных критериев — сравнение их функций мощности. Чем выше функция мощности, тем лучше критерий. Однако здесь возникает та же проблема, что и при сравнении функций риска (см. параграф 1.4). Функции мощности могут оказаться несравнимыми, но оптимальные критерии все же могут существовать. Их функция мощности в каждой точке $\alpha \in \Theta_K$ имеет большее значение, чем у других критериев с фиксирован-

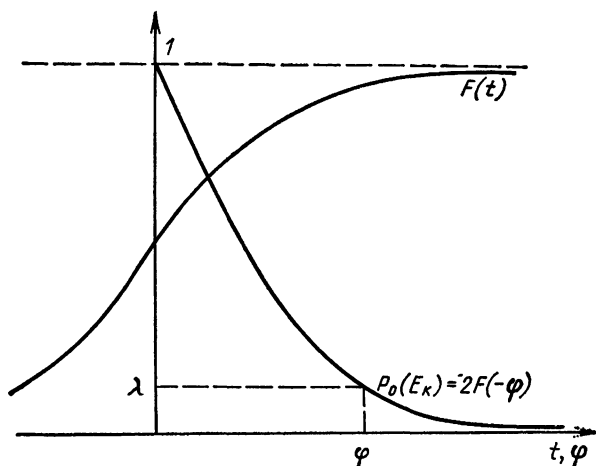


Рис. 1.12. Функция распределения средней $\Sigma y_i/n$ и зависимость $P_0(E_K)$ от φ

ным уровнем значимости. Такие критерии в математической статистике называют *равномерно наиболее мощными (РНМ)*.

Желаемым и естественным свойством критерия является следующее: вероятность принятия гипотезы, когда она неверна, меньше вероятности принятия гипотезы, когда она верна. В наших обозначениях $P_\alpha(E_H) \leq P_\beta(E_H)$ для любых $\alpha \in \Theta_K$, $\beta \in \Theta_H$. Такие критерии называются *несмещенными*. РНМ несмещенный критерий соответствует эффективной несмещенной оценке в теории оценивания.

Существует общий способ построения критериев проверки статистических гипотез. Он аналогичен методу максимального правдоподобия в статистическом оценивании и называется *критерием отношения правдоподобия*. Суть его заключается в следующем. Пусть плотность распределения y равна $f(y; \alpha)$, т. е. зависит от неизвестного вектора

параметров $\alpha \in \Theta$. Для каждого y найдем $\max_{\alpha \in \Theta_H} f(y; \alpha)$ и $\max_{\alpha \in \Theta} f(y; \alpha)$ (считаем, что максимум достигается).

В качестве критического множества при проверке гипотезы $H: \alpha \in \Theta_H$ выбираем

$$E_K = \left\{ y \in R^n: \max_{\alpha \in \Theta_H} f(y; \alpha) / \max_{\alpha \in \Theta} f(y; \alpha) < \varphi \right\}, \quad (1.48)$$

где φ — фиксированное число, зависящее от λ , которое в свою очередь задает верхнюю границу вероятности совершения ошибки первого рода $P_\alpha(E_K) \leq \lambda$ для всех $\alpha \in \Theta_H$. Статистика $\max_{\alpha \in \Theta_H} f(y; \alpha) / \max_{\alpha \in \Theta} f(y; \alpha)$ называется *статистикой критерия отношения правдоподобия*.

В случае простой случайной выборки, т. е. когда y_i независимы и одинаково распределены, известны асимптотические оптимальные статистические свойства критерия отношения правдоподобия [63].

Продолжение примера. Применим критерий отношения правдоподобия для проверки гипотезы $H: \mu = 0$. Функция плотности y равна:

$$f(y; \mu) = (2\pi)^{-\frac{n}{2}} \exp \left[-\frac{1}{2} \sum (y_i - \mu)^2 \right],$$

далее

$$\begin{aligned} \max_{\mu \in \Theta_H} f(y; \mu) &= \max_{\mu=0} (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2} \sum (y_i - \mu)^2} = \\ &= (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2} \sum y_i^2}; \end{aligned}$$

$$\max_{\mu \in \Theta} f(y; \mu) = \max_{\mu} (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2} \sum (y_i - \mu)^2} = 2\pi^{-\frac{n}{2}} e^{-\frac{1}{2} \sum (y_i - \bar{y})^2}$$

Последнее выражение следует из того, что минимальное значение $\sum (y_i - \mu)^2$ достигается при $\mu = \bar{y} = \sum y_i / n$. Критическим множеством будет:

$$\begin{aligned} E_K &= \left\{ y \in R^n: \exp \left[-\frac{1}{2} (\sum y_i^2 - \sum (y_i - \bar{y})^2) \right] < \varphi' \right\} = \\ &= \left\{ y \in R^n: \sum y_i^2 - \sum (y_i - \bar{y})^2 > -2 \ln \varphi' \right\} = \\ &= \left\{ y \in R^n: 2\bar{y} \sum y_i - n\bar{y}^2 > -2 \ln \varphi' \right\} = \\ &= \left\{ y \in R^n: |\bar{y}| > \sqrt{\frac{-2 \ln \varphi'}{n}} \right\} = \left\{ y \in R^n: |\bar{y}| > \varphi \right\}. \quad (1.49) \end{aligned}$$

Как видим, это множество совпадает с найденным ранее. Значение φ находится, как и раньше, из решения уравнения $\lambda = 2F(-\varphi)$.

Перейдем к построению доверительных интервалов. Пусть распределение случайного вектора $y \in R^n$ известно с точностью до $\alpha \in \Theta \subseteq R^m$. Оценим α с помощью доверительного интервала или в общем случае с помощью *доверительного множества*. Доверительное множество каждый раз зависит от наблюдения y , т. е. является его функцией. Обозначим его через $D(y)$. Важно, с какой вероятностью доверительное множество $D(y)$ покрывает истинный параметр. Аналогично проверке гипотез можно ввести коэффициент доверия доверительного множества. $D(y)$ является доверительным множеством с коэффициентом доверия не менее $1 - \lambda$, если для каждого $\alpha \in \Theta$:

$$P_{\alpha} \{y \in R^n : \alpha \in D(y)\} \geq 1 - \lambda. \quad (1.50)$$

Слева записана вероятность того, что $D(y)$ покроет α . Эта вероятность вычисляется при условии, что истинное значение параметра также равно α .

Вероятность ошибки, т. е. вероятность того, что доверительное множество не покроет истинное значение параметра, равна:

$$P_{\alpha} \{y \in R^n : \alpha \notin D(y)\} = 1 - P_{\alpha} \{y \in R^n : \alpha \in D(y)\} \leq 1 - (1 - \lambda) = \lambda.$$

Таким образом, λ можно трактовать как максимальную вероятность ошибки накрытия.

Можно ввести понятие *несмещенности доверительного множества*, аналогичное несмещенности статистического критерия. Допустим, α — истинное значение параметра, $\beta \neq \alpha$. Мы же ошибочно предполагаем, что именно β является истинным, и поэтому наше доверительное множество $D(y)$ направлено на оценивание параметра β . Вероятность такого (ошибочного) оценивания равна $P_{\alpha} \{\beta \in D(y)\}$. Естественно считать, что эта вероятность будет меньше, чем если бы мы не делали ошибки и $\alpha = \beta$; тогда вероятность накрытия равна $P_{\alpha} \{\alpha \in D(y)\} \geq 1 - \lambda$ (по условию (1.50)). Говорим, что D является несмещенным доверительным множеством, если для любых $\alpha, \beta \in \Theta$ имеем

$$P_{\alpha} \{y \in R^n : \beta \in D(y)\} \leq P_{\alpha} \{y \in R^n : \alpha \in D(y)\}. \quad (1.51)$$

Теперь дадим определение *наиболее точного доверительного множества*. Представим себе ситуацию, когда истинное значение параметра α нам неизвестно и мы ошибочно оцениваем β вместо α . Пусть имеются два метода доверительного оценивания, т. е. два доверительных множества $D(\mathbf{y})$ и $S(\mathbf{y})$; D будет точнее S , если D покрывает ошибочное значение β реже, чем S :

$$P_{\alpha}\{\mathbf{y} \in R^n : \beta \in D(\mathbf{y})\} \leq P_{\alpha}\{\mathbf{y} \in R^n : \beta \in S(\mathbf{y})\}. \quad (1.52)$$

В частном случае, когда $\Theta \subset R^1$, естественно вместо произвольных доверительных множеств рассматривать доверительные интервалы. Такой интервал будет характеризоваться парой статистик $a_1(\mathbf{y})$ и $a_2(\mathbf{y})$, причем $a_1(\mathbf{y}) \leq a_2(\mathbf{y})$ для любых \mathbf{y} . Вероятность накрытия равна:

$$P_{\alpha}\{\mathbf{y} \in R^n : a_1(\mathbf{y}) \leq \alpha \leq a_2(\mathbf{y})\}, \quad \alpha \in \Theta.$$

Задача доверительного оценивания теснейшим образом связана с проверкой простой гипотезы. Будем проверять простую гипотезу $H: \alpha = \beta$, где β — фиксированное значение из Θ . Допустим, имеется некоторый критерий проверки этой гипотезы. Область принятия гипотезы обозначим E_H или E_{β} . Предположим, что критерий имеет уровень значимости λ , причем $P_{\alpha}(E_H) = P_{\alpha}(E_{\beta}) = 1 - \lambda$ для всех $\alpha \in \Theta$. На основе E_{β} построим доверительное множество для оценивания α , которое обозначим $D(\mathbf{y})$. Положим $\mathbf{y} \in E_{\beta}$ тогда и только тогда, когда $\beta \in D(\mathbf{y})$; другими словами,

$$D(\mathbf{y}) = \{\beta \in \Theta : \mathbf{y} \in E_{\beta}\}. \quad (1.53)$$

Как объяснить выбор (1.53)? Множество E_{β} есть множество тех возможных наблюдений \mathbf{y} , которые «скорее всего» получаются, если в качестве неизвестного параметра выступает параметр β . Поэтому множество всех параметров, которые «порождают» это множество E_{β} , будет «близко расположенным» к β и будет хорошей доверительной оценкой этого параметра. Легко убедиться в том, что если критерий имеет уровень значимости λ , то построенное доверительное множество $D(\mathbf{y})$ (1.53) имеет коэффициент доверия $1 - \lambda$:

$$\begin{aligned} P_{\alpha}\{\mathbf{y} \in R^n : \alpha \in D(\mathbf{y})\} &= P_{\alpha}\{\mathbf{y} \in R^n : \mathbf{y} \in E_{\beta}\} = \\ &= P_{\alpha}(E_{\beta}) = 1 - \lambda. \end{aligned}$$

Основная связь между проверкой гипотез и доверительным оцениванием выражается в виде следующей теоремы, доказательство которой весьма просто.

Теорема 1.14. Пусть имеется РНМ несмещенный критерий проверки простой статистической гипотезы $H: \mu = \beta$ с уровнем значимости λ . Тогда доверительное множество $D(\mathbf{y})$ (1.53) является несмещенным, имеет коэффициент доверия $1 - \lambda$ и является равномерно наиболее точным (РНТ), т. е. эффективным. Обратно, РНТ несмещенные доверительные множества приводят к РНМ несмещенным критериям.

Сформулированная теорема предлагает нам большие возможности в построении доверительных множеств. Задачи построения гипотез и доверительного оценивания можно считать эквивалентными.

Пример. Несколько обобщим предыдущий пример. Предположим, мы хотим проверить простую гипотезу H : истинное значение μ равно μ_0 . Легко проверить, что область принятия гипотезы будет

$$E_\mu = \{ \mathbf{y} \in R^n : |\bar{y} - \mu| \leq \varphi \}. \quad (1.54)$$

Тогда

$$\begin{aligned} D(\mathbf{y}) &= \{ \mu : \mathbf{y} \in E_\mu \} = \{ \mu : |\bar{y} - \mu| \leq \varphi \} = \\ &= \{ \mu : \bar{y} - \varphi \leq \mu \leq \bar{y} + \varphi \}. \end{aligned}$$

Значение φ , как и раньше, является решением уравнения $\lambda = 2F(-\varphi)$. Множество $D(\mathbf{y})$ имеет коэффициент доверия $1 - \lambda$. Можно показать, что критерий (1.54) является несмещенным РНМ. Отсюда вытекает, что доверительный интервал (a_1, a_2) для оценивания μ , где $a_1 = \bar{y} - \varphi$, $a_2 = \bar{y} + \varphi$, является несмещенным РНТ.

У п р а ж н е н и я 1.9

1. В задаче проверки гипотезы $\mu = 0$ из примера параграфа 1.4 с неизвестной дисперсией σ^2 $\alpha = (\mu, \sigma^2) \in \Theta = \{ \text{верхняя полуплоскость в системе координат } (\mu, \sigma^2) \}$ покажите, что критерий отношения правдоподобия приводит к критическому множеству $\{ |\bar{y}|/s_1 > \varphi \}$, где $s_1^2 = \Sigma (y_i - \bar{y})^2/n$. Как найти φ для данного значения λ ?

2. Найдите в том же примере критерий проверки гипотезы $H: \mu = \mu_0$, где μ_0 — фиксированное значение.

1.10. Проверка гипотез и доверительное оценивание в линейной регрессии

Сначала построим критерий проверки общей линейной гипотезы относительно параметров регрессии, применяя критерий отношения правдоподобия. На основе критерия проверки гипотез мы сможем найти доверительные множества и интервалы.

Рассмотрим линейную гипотезу в самом общем виде H :

$$R\alpha = r, \quad (1.55)$$

где R — известная матрица $k \times m$, $\text{rang } R = k < m$; r — заданный вектор $k \times 1$. Таким образом, в гипотезе H на α накладывается k линейных независимых ограничений. Будем минимизировать сумму квадратов отклонений при условии (1.55). Другими словами, решаем следующую оптимизационную задачу с ограничениями $(y - X\alpha)' \times (y - X\alpha) \Rightarrow \min$ при условии $R\alpha = r$.

Можно показать, что оценка, приводящая сумму квадратов отклонений к минимуму при условии, что гипотеза H верна, равна:

$$\begin{aligned} a_R &= a + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - Ra) = \\ &= a + (X'X)^{-1}R'S(r - Ra), \end{aligned} \quad (1.56)$$

где $S = [R(X'X)^{-1}R']^{-1}$; a — оценка МНК; вывод оценки дается в параграфе 2.3. Минимальное значение суммы квадратов отклонений равно:

$$\begin{aligned} Q_R &= (y - Xa_R)'(y - Xa_R) = \\ &= [y - Xa - X(X'X)^{-1}R'S(r - Ra)]' \times \\ &\quad \times [y - Xa - X(X'X)^{-1}R'S(r - Ra)] = \\ &= (y - Xa)'(y - Xa) - (y - Xa)'X'(X'X)^{-1}R'S(r - Ra) + \\ &\quad + (r - Ra)'SR(X'X)^{-1}X'X(X'X)^{-1}R'S(r - Ra). \end{aligned}$$

Второе слагаемое равно нулю. Последнее слагаемое после сокращений равно $(r - Ra)'S(r - Ra)$, поэтому

$$Q_R = \hat{Q} + (r - Ra)'S(r - Ra), \quad (1.57)$$

где \hat{Q} соответствует сумме квадратов отклонений оценки МНК.

Для проверки гипотез и доверительного оценивания необходимо знать вид распределения y . Предположим, что y имеет нормальное распределение $N(X\alpha, \sigma^2 I_n)$. В этом слу-

чае плотность распределения y зависит от $m + 1$ неизвестных параметров и равна:

$$f(y; \alpha, \sigma^2) = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \times \\ \times \exp \left[-\frac{1}{2\sigma^2} (y - X\alpha)' (y - X\alpha) \right].$$

Для проверки гипотезы $H: R\alpha = r$ применим критерий отношения правдоподобия. Для этого необходимо найти $\max f(y; \alpha, \sigma^2)$ при условии (1.55). Максимуму f соответствует минимум суммы квадратов отклонений Q , что приводит к оценке a_R (1.56). Легко видеть, что значением σ^2 , обращающим f в максимум, является

$$s_R^2 = \frac{1}{n} (y - Xa_R)' (y - Xa_R) = \frac{1}{n} Q_R, \quad (1.58)$$

где Q_R задается (1.57); $\max f(y, \alpha, \sigma^2)$ без ограничений приводит к обычной оценке МНК a и s^2 , равной

$$s_M^2 = \frac{1}{n} (y - Xa)' (y - Xa) = \frac{1}{n} \hat{Q}. \quad (1.59)$$

Индекс M указывает здесь на то, что эта оценка является оценкой метода максимального правдоподобия.

По определению критическое множество критерия отношения правдоподобия равно (1.48), т. е.

$$E_K = \left\{ \max_{R\alpha=r} f(y; \alpha, \sigma^2) / \max_{\alpha, \sigma^2} f(y; \alpha, \sigma^2) < \varphi \right\} = \\ = \left\{ \frac{(2\pi)^{-n/2} (s_R^2)^{-n/2} e^{-n/2}}{(2\pi)^{-n/2} (s_M^2)^{-n/2} e^{-n/2}} < \varphi \right\} = \\ = \left\{ y : \left(\frac{s_R^2}{s_M^2} \right)^{-\frac{n}{2}} < \varphi \right\} = \left\{ y : \frac{s_R^2}{s_M^2} > \varphi' \right\}. \quad (1.60)$$

С учетом (1.57), (1.58) и (1.59), введя новую константу и разделив числитель на k , а знаменатель на $n - m^1$, получим:

$$E_K = \left\{ y : \frac{(r - Ra)' S (r - Ra) / k}{(y - Xa)' (y - Xa) / (n - m)} > F \right\} = \\ = \left\{ y : \frac{(Q_R - \hat{Q}) / k}{s^2} > F \right\}, \quad (1.61)$$

где s^2 — несмещенная оценка σ^2 (1.20).

¹Суть этой операции будет ясна из дальнейшего изложения.

Наша ближайшая задача — найти распределение статистики критерия (1.61). Прежде всего найдем распределение числителя. Имеем $\mathbf{r} - \mathbf{R}\mathbf{a} = \mathbf{r} - \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} =$

$$= \mathbf{r} - \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\alpha} - \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} = \\ = -\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon},$$

поэтому

$$(\mathbf{r} - \mathbf{R}\mathbf{a})' \mathbf{S}(\mathbf{r} - \mathbf{R}\mathbf{a}) = \boldsymbol{\varepsilon}' \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{S}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} = \\ = \boldsymbol{\varepsilon}' \mathbf{D}\boldsymbol{\varepsilon}, \quad (1.62)$$

где \mathbf{D} — идемпотентная матрица, причем

$$\text{tr } \mathbf{D} = \text{tr } \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{S} = \\ = \text{tr } \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{S} = \text{tr } \mathbf{I}_k = k.$$

Применяя формулу (П.14) приложения, утверждаем, что $\frac{1}{\sigma^2} \boldsymbol{\varepsilon}' \mathbf{D}\boldsymbol{\varepsilon} \sim \chi^2(k)$. Ранее было доказано, что $(\mathbf{y} - \mathbf{X}\mathbf{a})' \times \times (\mathbf{y} - \mathbf{X}\mathbf{a})/\sigma^2 \sim \chi^2(n - m)$. Теперь покажем, что две квадратичные формы независимы. Как следует из (П.16), для этого достаточно показать, что

$$\mathbf{D}(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \mathbf{O}. \quad (1.63)$$

Это равенство проверить нетрудно. Учитывая вышесказанное, можно утверждать, что статистика критерия отношения правдоподобия имеет распределение F с k и $n - m$ степенями свободы.

Рассмотрим процедуру проверки гипотезы $H: \mathbf{R}\mathbf{a} = \mathbf{r}$. Прежде всего необходимо задаться уровнем значимости λ . Для данного уровня λ находим соответствующее значение $F_\lambda(k, n - m)$ следующим образом: обозначим $f(t; k, n - m)$ — плотность распределения Фишера с k и $n - m$ степенями свободы. Найдем такое F_λ , чтобы $\int_{F_\lambda}^{\infty} f(t; k, n - m) dt = \lambda$. Значение F_λ находят из таблиц.

Если

$$(Q_R - \hat{Q})/ks^2 = (\mathbf{r} - \mathbf{R}\mathbf{a})' \mathbf{S}(\mathbf{r} - \mathbf{R}\mathbf{a})/ks^2 > F_\lambda(k, n - m), \quad (1.64)$$

то гипотезу $H: \mathbf{R}\mathbf{a} = \mathbf{r}$ отвергаем, в противном случае принимаем.

Рассмотрим прежде всего три специальных случая применения критерия (1.64). Сначала построим критерий проверки гипотезы $H: \alpha_i = \beta$, где β — некоторое фиксированное значение.

рованное число. Для этой гипотезы $k = 1$, $\mathbf{R} = (0, \dots, 1, \dots, 0)$, где на i -м месте стоит 1. Далее

$$\mathbf{S} = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1} = [(\mathbf{X}'\mathbf{X})_{ii}^{-1}]^{-1} = 1/(\mathbf{X}'\mathbf{X})_{ii}^{-1},$$

поэтому

$$(\mathbf{r} - \mathbf{R}\mathbf{a})' \mathbf{S} (\mathbf{r} - \mathbf{R}\mathbf{a}) = (\beta - a_i)^2 / (\mathbf{X}'\mathbf{X})_{ii}^{-1}.$$

Таким образом,

$$(\beta - a_i)^2 / (\mathbf{X}'\mathbf{X})_{ii}^{-1} s^2 \propto F(1, n - m),$$

а

$$(a_i - \beta) / s \sqrt{(\mathbf{X}'\mathbf{X})_{ii}^{-1}} \propto t(n - m)$$

— распределение Стьюдента с $n - m$ степенями свободы. С учетом общей формулы (1.61), если

$$|a_i - \beta| / s_i > t_\lambda, \quad (1.65)$$

где $s_i = s \sqrt{(\mathbf{X}'\mathbf{X})_{ii}^{-1}}$ — оценка стандартного отклонения a_i , гипотезу о равенстве $\alpha_i = \beta$ отвергаем, в противном случае принимаем. Если обозначить через $t(z, n - m)$ функцию плотности t -распределения с $n - m$ степенями свободы, то t_λ для данного уровня значимости λ находится из решения уравнения $\int_{|z| > t_\lambda} t(z; n - m) dz = \lambda$. Значение t_λ оп-

ределяют также из таблиц, например см. [37]. Статистику $|a_i - \beta| / s_i$ называют t -статистикой.

Для примера вернемся к регрессии (1.6). Рассмотрим поочередно гипотезы $H_i: \alpha_i = 0$, $i = 1, 2, 3, 4$. В данном случае $\beta = 0$; значения s_i вычислены в параграфе 1.5. t -статистики для каждого параметра равны: $t_1 = 9,632$, $t_2 = 0,548$, $t_3 = 2,893$, $t_4 = 0,611$. Зададимся 5%-ным уровнем значимости; тогда с учетом того, что степень свободы для нашей регрессии равна: $n - m = 15 - 4 = 11$, табличное значение t -статистики равно: $t_{5\%} = 2,201$. Таким образом, с 95%-ной уверенностью можем утверждать, что гипотезы $H_2: \alpha_2 = 0$ и $H_4: \alpha_4 = 0$ верны, $H_1: \alpha_1 = 0$ и $H_3: \alpha_3 = 0$ не верны.

Проверка гипотез $H_i: \alpha_i = 0$ имеет большое значение в регрессиях. От ответа на вопрос: «Считать ли параметр нулем?» — зависит, оставлять или выбросить переменную-фактор из уравнения регрессии. Если гипотеза $\alpha_i = 0$ подтвердилась с большой вероятностью, то фактор, как правило, удаляют из регрессии, если нет — оставляют. Подобную процедуру выбора существенных факторов называют иногда процедурой автоматического отсева переменных.

Рассмотрим гипотезы $H_i: \alpha_i = 0$ в свете общих идей проверки статистических гипотез. Не теряя общности, остановимся на случае $i = 1$, при этом гипотеза $H_1: \alpha_1 = 0$. В данном случае $\Theta = \{-\infty < \alpha_1 < \infty\}$, $\Theta_H = \{\alpha_1 = 0\}$, $\Theta_K = \{\alpha_1 \neq 0\}$, $E_K = \{y \in R^n : |a_1| > s_1 t_\lambda\}$ — критическая область; $E_H = \{y \in R^n : |a_1| \leq s_1 t_\lambda\}$ — область принятия гипотезы $\alpha_1 = 0$.

Ошибка первого рода: $\alpha_1 = 0$, а мы делаем вывод, что $\alpha_1 \neq 0$, и поэтому x_1 оставляем в уравнении регрессии. Вероятностью этой ошибки мы задаемся заранее; она составляет λ . Ошибка такого рода ведет к перебору факторов в регрессии.

Ошибка второго рода: $\alpha_1 \neq 0$, а мы делаем вывод, что $\alpha_1 = 0$, а поэтому x_1 исключаем из регрессии. Вероятность совершения этой ошибки не постоянна и зависит от конкретного значения α_1 . Ошибка второго рода ведет к недобору. *Недобор связан со смещением в оценках и является более серьезной ошибкой спецификации регрессии, чем перебор* (см. параграф 2.4).

Если при диагнозе тяжелой болезни (см. параграф 1.9) мы стараемся минимизировать ошибку первого рода, то в регрессионном анализе, наоборот, целесообразнее минимизировать ошибку второго рода.

Приведенные соображения показывают, что *не следует увлекаться низкими значениями λ , так как при этом повышается вероятность недобора, что ведет к смещению в оценках.*

Теперь рассмотрим второй специальный случай гипотезы (1.55). Пусть задана линейная регрессия со свободным членом

$$y_t = \alpha_1 x_{t1} + \dots + \alpha_{m-1} x_{t, m-1} + \alpha_m + \varepsilon_t. \quad (1.66)$$

Проверим линейную гипотезу $H: \alpha_1 = \alpha_2 = \dots = \alpha_{m-1} = 0$. Другими словами, мы проверяем на взаимную конкуренцию две модели: (1.66) и модель среднего

$$y_t = \alpha_m + \varepsilon_t. \quad (1.67)$$

Неформально мы проверяем, есть ли эффект от введения в уравнение (1.67) факторов x_1, \dots, x_{m-1} . Очевидно, $Q_R = \Sigma (y_t - \bar{y})^2$. Выразим \hat{Q} через R^2 (уравнение (1.27)):

$$\hat{Q} = \Sigma (y_t - \bar{y})^2 (1 - R^2), \text{ поэтому}$$

$$\frac{(Q_R - \hat{Q})/(m-1)}{\hat{Q}/(n-m)} = \frac{n-m}{m-1} \cdot \frac{R^2}{1-R^2}. \quad (1.68)$$

Таким образом, если для данного наблюдения y выражение (1.68) больше $F_\lambda(m-1, n-m)$ — значения, получаемого из F — распределения с $m-1$ и $n-m$ степенями свободы, соответствующего данному уровню значимости λ , то гипотеза $H: \alpha_1 = \dots = \alpha_{m-1} = 0$ отвергается, в противном случае принимается.

Рассмотрим проверку линейной гипотезы на нашем примере—регрессии. Для этой регрессии $R^2 = 0,9935$, и поэтому $(n-m)/(m-1) \cdot R^2/(1-R^2) = 560,4$. Как и прежде, выберем $\lambda = 5\%$, тогда $F_\lambda(m_1, m_2) = 8,76$ с $m_1 = 3$ и $m_2 = 11$ степенями свободы. Поскольку $560,4 > 8,76$, делаем вывод, что гипотезу $H: \alpha_1 = \alpha_2 = \alpha_3 = 0$ необходимо отвергнуть.

Рассмотрим третий случай применения линейной гипотезы (1.55). Допустим, кроме данного набора наблюдений имеется дополнительный. Являются ли регрессии, построенные по двум выборкам, одинаковыми? Формально задачу можно записать следующим образом. Имеются две регрессии:

$$y_1 = X_1 \alpha_1 + \varepsilon_1; \quad y_2 = X_2 \alpha_2 + \varepsilon_2, \quad (1.69), (1.70)$$

где y_1 и ε_1 — векторы размерности n_1 ; y_2 и ε_2 — векторы размерности n_2 ; X_1 — матрица $n_1 \times m$; X_2 — матрица $n_2 \times m$; α_1 и α_2 — векторы размерности m . Предполагаем, что ε_1 и ε_2 независимы, причем $\varepsilon_1 \sim N(0, \sigma^2 I_{n_1})$, $\varepsilon_2 \sim N(0, \sigma^2 I_{n_2})$. Регрессии (1.69) и (1.70) могут быть объединены в одну:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}. \quad (1.71)$$

Для этой регрессии проверяется линейная гипотеза

$$H: \alpha_1 = \alpha_2, \quad (1.72)$$

что в терминах R и r (1.55) означает

$$R = \begin{bmatrix} 1 & 0 & -1 & 0 \\ & 1 & & -1 \\ & \vdots & & \ddots \\ 0 & 1 & 0 & -1 \end{bmatrix} \text{— матрица } m \times 2m, r = 0.$$

Для проверки гипотезы (1.72) необходимо найти Q_R и \hat{Q} . С учетом (1.72) регрессии (1.69) и (1.70) могут быть переписаны следующим образом:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \alpha + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}, \quad (1.73)$$

Таким образом, Q_R равно сумме квадратов отклонений оценки МНК составной регрессии. Найдем теперь сумму квадратов отклонений оценки МНК регрессии (1.71).

Имеем:

$$\begin{aligned} Q &= \begin{bmatrix} y_1 - X_1 \alpha_1 \\ y_2 - X_2 \alpha_2 \end{bmatrix}' \begin{bmatrix} y_1 - X_1 \alpha_1 \\ y_2 - X_2 \alpha_2 \end{bmatrix} = \\ &= (y_1 - X_1 \alpha_1)' (y_1 - X_1 \alpha_1) + (y_2 - X_2 \alpha_2)' (y_2 - X_2 \alpha_2). \end{aligned}$$

Обозначим \hat{Q}_1 — сумму квадратов отклонений оценки МНК регрессии (1.69), \hat{Q}_2 — сумму квадратов отклонений регрессии (1.70). Тогда, как следует из предыдущего выражения, минимальное значение Q , т. е. \hat{Q} , равно $\hat{Q}_1 + \hat{Q}_2$. Итак, $Q_R - \hat{Q} = Q_R - \hat{Q}_1 - \hat{Q}_2$. Далее учтем, что

$$s^2 = \frac{\hat{Q}}{n_1 + n_2 - 2m} = \frac{\hat{Q}_1 + \hat{Q}_2}{n_1 + n_2 - 2m}.$$

Поэтому если

$$\frac{(Q_R - \hat{Q}_1 - \hat{Q}_2)/m}{(\hat{Q}_1 + \hat{Q}_2)/(n_1 + n_2 - 2m)} > F_\lambda(m, n_1 + n_2 - 2m),$$

гипотезу (1.72) отвергаем, в противном случае принимаем.

Остановимся на оптимальности критерия отношения правдоподобия, применяемого для регрессии с нормально распределенными отклонениями. Было показано, что не существует РНМ критерия для проверки общей линейной гипотезы (1.55). Более того, доказано, что не существует и РНМ несмещенного критерия для случая $k \geq 2$. В то же время для $k=1$ критерий отношения правдоподобия является эффективным, т. е. РНМ несмещенным критерием. И тем не менее для $k \geq 2$ критерий отношения правдоподобия является оптимальным в более узком смысле¹. Вышесказанное позволяет утверждать, что критерий (1.61) является достаточно эффективным.

¹Подробнее см. [37; с. 334, 342—344; 45].

Для построения доверительных интервалов и областей воспользуемся общим методом построения, изложенным в предыдущем параграфе. Суть его основывается на связи с критерием проверки гипотез. Начнем с индивидуальных доверительных интервалов для параметров регрессии. Как следует из (1.65), множеством принятия решения при проверке гипотезы $H_i : \alpha_i = \beta$ является $E_i = \{y \in R^n : |a_i - \beta| < t_{\lambda} s_i\}$. Таким образом, оптимальной доверительной областью с коэффициентом доверия $1 - \lambda$ является интервал

$$D_i = \{\beta : |a_i - \beta| < t_{\lambda} s_i\}. \quad (1.74)$$

Оптимальность этого интервала заключается в том, что он является несмещенным и равномерно наиболее точным (см. параграф 1.9). Значение t_{λ} находится так же, как и в случае проверки гипотез.

Доверительные 95%-ные интервалы для рассматриваемой регрессии (1.6) приведены в табл. 1.3.

Т а б л и ц а 1.3

Параметры	a_1	a_2	a_3	a_4
Нижняя граница	0,304	-0,691	0,896	--77,6
Верхняя граница	0,4854	1,149	6,596	-44,5

Как видим, доверительные интервалы для всех параметров, исключая первый, достаточно широки. Это позволяет выдвинуть гипотезу о преобладающем значении присутствия вещества B_1 в реакции с веществом B_2 . Остальные факторы имеют второстепенное значение (что является вполне естественным).

Найдем совместную доверительную область D для всех параметров $(\alpha_1, \alpha_2, \dots, \alpha_m)' = \alpha$. Для этого необходимо проверить гипотезу $H : \alpha_1 = \beta_1, \alpha_2 = \beta_2, \dots, \alpha_m = \beta_m$, т. е. $\alpha = (\beta_1, \dots, \beta_m)' = \beta$. Очевидно, $Q_R = (y - X\beta)' \times (y - X\beta)$, поэтому критическим множеством, как следует из (1.61), является

$$E_{\beta} = \left\{ y \in R^n : \frac{(Q_R - \hat{Q})/m}{\hat{Q}/(n-m)} \leq F_{\lambda}(m, n-m) \right\},$$

но

$$Q_R = (y - X\beta)'(y - X\beta) = [(y - Xa) + X(a - \beta)]' \times \\ \times [y - Xa) + X(a - \beta)] = \hat{Q} + 2(y - Xa)' X(a - \beta) + \\ + (\beta - a)' X' X(\beta - a).$$

Второе слагаемое равно нулю, поэтому $Q_R = (\beta - a)' X' X \times \\ \times (\beta - a) + \hat{Q}$ и

$$D = \left\{ \beta \in R^m : (\beta - a)' X' X(\beta - a) \leq \frac{m}{n-m} F_{\lambda} \cdot \hat{Q} \right\} = \\ = \{ \beta \in R^m : (\beta - a)' X' X(\beta - a) \leq ms^2 F_{\lambda} \} \quad (1.75)$$

— доверительная область с коэффициентом доверия $1 - \lambda$.

Очевидно, D есть эллипсоид в R^m . Центр его находится в a . Для $m > 3$ графическое построение эллипсоида практически невозможно, да и при $m = 3$ эта задача весьма затруднительна. Достаточное представление о расположении доверительного эллипсоида дает вычисление характеристических чисел и векторов матрицы плана $X'X$. В регрессии (1.6) ими будут (расположены в порядке возрастания характеристических чисел):

$$\lambda_1 = 0,00963; \quad s_1 = (0,000766; -0,0147; 0,00283; 0,999);$$

$$\lambda_2 = 5,61; \quad s_2 = (-0,0261; -0,0179; 0,999; -0,0285);$$

$$\lambda_3 = 1344,8; \quad s_3 = (-0,296; 0,955; 0,00979; 0,0140);$$

$$\lambda_4 = 1917144,1; \quad s_4 = (0,955; 0,296; 0,0303; 0,00277).$$

Величина, характеризующая обусловленность матрицы $X'X$, $\lambda_{\max}(X'X)/\lambda_{\min}(X'X) = 1917144,1/0,00963 \approx 2 \times 10^8$ очень велика; следовательно, матрица $X'X$ плохо обусловлена (подробнее см. параграф 6.1). Отношение максимального характеристического числа к минимальному равно отношению длины максимальной полуоси эллипсоида к минимальной. Таким образом, эллипсоид, соответствующий матрице $X'X$, сильно вытянут в одном направлении и сжат в другом. Направление, в котором вытянут эллипсоид, соответствует вектору s_4 . В свою очередь s_4 почти совпадает с направлениями оси β_1 . Следующая ось D направлена близко к β_2 и т. д. Таким образом, эллипсоид может быть достаточно хорошо аппроксимирован только одним фактором x_1 ; еще лучшую аппроксимацию даст введение факторов x_1, x_2 , и т. д. (подробнее см. параграф 6.1). Дей-

ствительно, если оставить только первый фактор x_1 , то регрессия будет не намного хуже первоначальной:

$$y_t = 0,535x_{t1} + e_t, \quad (1.76)$$

(0,0027)

$s^2 = 12,4$, что соответствует сумме квадратов отклонений 173,6 (полной регрессии (1.6) отвечает сумма квадратов отклонений $\hat{Q} = 68,46$). В регрессии y на x_2 сумма квадратов отклонений будет уже 4256.

Таким образом, если в рассматриваемом химическом эксперименте температура проведения реакции лежит в окрестности 100° , а количество катализатора, участвующего в реакции, находится в пределах 10 г, то вместо модели (1.6) возможно использование более грубой модели (1.76). Привлечение дополнительного килограмма вещества B_1 приведет к увеличению выхода реакции в среднем на 535 г.

Доверительная область в виде эллипсоида несет на себе большую информацию. Однако, как уже отмечалось, его построение и интерпретация затруднительны уже для $m = 3$. Встает вопрос о возможности построения доверительной области, которую легче интерпретировать. В качестве такой области (множества) выберем обобщенный прямоугольник (прямоугольный параллелепипед). Это множество будем строить на основе индивидуальных доверительных интервалов. Рассмотрим один из способов построения совместных доверительных интервалов [63, с. 302], предложенный Тьюки.

Т е о р е м а 1.15. Пусть D_i ($i = 1, 2, \dots, m$) — индивидуальные доверительные интервалы для параметров α_i (1.74) с коэффициентом доверия $1 - \lambda/m$. Эти интервалы будут совместными с коэффициентом доверия не менее $1 - \lambda$.

Д о к а з а т е л ь с т в о. Обозначим через A_i множество тех y , для которых D_i накрывает истинное значение параметров α_i , т. е. $A_i = \{y : \alpha_i \in D_i\}$. По условию теоремы $P(A_i) = 1 - \lambda/m$. Множество одновременного (совместного) накрытия равно: $A = A_1 \cap \dots \cap A_m$. Оценим вероятность этого множества:

$$\begin{aligned} P(A) &= P(A_1 \cap \dots \cap A_m) = 1 - P(A_1^c \cup \dots \cup A_m^c) \geq \\ &\geq 1 - P(A_1^c) - \dots - P(A_m^c) = 1 - m\lambda/m = 1 - \lambda, \end{aligned}$$

где A_i^c — дополнение к множеству A_i и $P(A_i^c) = 1 - P(A_i) = \lambda/m$.

Построим совместную доверительную область в виде обобщенного прямоугольника для регрессии-примера (1.6) с коэффициентом доверия $1 - \lambda = 95\%$, т. е. $\lambda = 5\%$. Для этого надо построить индивидуальные доверительные интервалы с коэффициентом доверия $1 - 0,05/4 = 0,9875$. Приближенное значение $t_{0,0125}$ при 11 степенях свободы равно 3, поэтому одновременно 95%-ным доверительным прямоугольником для $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ будет $(0,27366; 0,5212)$; $(-1,017; 1,475)$; $(-0,139; 7,631)$; $(-100,9; 66,74)$. Как видим, границы доверительных интервалов стали весьма широкими. Это результат того, что число степеней свободы в регрессии-примере (1.6) невелико и равно 11. Чем больше значение m , тем сильнее индивидуальные доверительные интервалы будут отличаться от совместных, построенных по методу Тьюки. В [63] обсуждается еще один метод построения совместных доверительных интервалов.

Важным моментом теории проверок гипотез и построения доверительных интервалов является предположение о нормальном распределении отклонений. Если отклонения регрессии не являются нормально распределенными, но выполняются условия асимптотической нормальности оценки МНК, можно показать, что построенные таким образом доверительные интервалы и критерии проверок гипотез являются асимптотически оптимальными. В некоторых работах исследуется, насколько эффективность проверок гипотез и доверительного оценивания теряется при отклонениях от нормальности. Если отклонения имеют распределения с легкими хвостами (см. гл. 5), то построенные в этом параграфе критерии и интервалы не теряют практически своих свойств. Для распределений с тяжелыми хвостами ситуация меняется. Вопросы устойчивости (робастности) см. в [37].

Можно предложить грубый метод построения доверительных интервалов, не зависящий от распределения ε , основанный на неравенстве Чебышева. Пусть a_i — i -я координата вектора оценки МНК, s_i — ее стандартная ошибка (вернее, оценка этой величины), тогда по неравенству Чебышева для любого $\tau > 0$

$$P \{ |a_i - \alpha_i| < \tau s_i \} \geq 1 - 1/\tau^2. \quad (1.77)$$

Доверительным интервалом для параметра α_i является интервал $(a_i - \tau s_i, a_i + \tau s_i)$ с коэффициентом доверия не менее $1 - 1/\tau^2$. Так, если положить $\tau = 3$, то коэффициент доверия будет не менее $8/9 \approx 0,889$. Точность выполнения

неравенства (1.77) зависит от того, насколько s^2 эффективно оценивает истинное значение стандартной ошибки отклонений σ^2 . Применяя процедуру Тьюки, на основе (1.77) можно построить совместные доверительные интервалы.

У п р а ж н е н и е 1.10

1. Используя метод Лагранжа, найдите a_R (1.56).

1.11. Доказательства

1. Доказательство теоремы 1.3. Имеем

$$\begin{aligned} \mathbf{e} &= \mathbf{y} - \hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} - \mathbf{X}\mathbf{a} = \mathbf{X}(\boldsymbol{\alpha} - \mathbf{a}) + \boldsymbol{\varepsilon} = \\ &= -\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon} = [-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{I}_n]\boldsymbol{\varepsilon} = \mathbf{A}\boldsymbol{\varepsilon}, \end{aligned} \quad (1.78)$$

где $\mathbf{A} = -\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{I}_n$; \mathbf{I}_n — единичная матрица $n \times n$. Матрица \mathbf{A} идемпотентна:

$$\begin{aligned} \mathbf{A}^2 &= [\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] [\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = \\ &= \mathbf{I}_n + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \\ &\quad - 2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{A}. \end{aligned} \quad (1.79)$$

Применяя формулу (П.13), получим

$$\mathbf{E}\mathbf{e}'\mathbf{e} = \mathbf{E}(\boldsymbol{\varepsilon}'\mathbf{A}\boldsymbol{\varepsilon}) = \sigma^2 \operatorname{tr} \mathbf{A} = \sigma^2 (\operatorname{tr} \mathbf{I}_n - \operatorname{tr} \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'),$$

но $\operatorname{tr} \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \operatorname{tr} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = m$, поэтому $\mathbf{E}\mathbf{e}'\mathbf{A}\boldsymbol{\varepsilon} = \sigma^2 (n - m)$, откуда $\mathbf{E}s^2 = \sigma^2$, что и требовалось доказать.

2. Доказательство теоремы 1.5. 1) Пусть $\lambda_{\min}(\mathbf{X}'_n\mathbf{X}_n) \rightarrow \infty$, $n \rightarrow \infty$. Тогда

$$\lambda_{\max}(\mathbf{X}'_n\mathbf{X}_n)^{-1} = 1/\lambda_{\min}(\mathbf{X}'_n\mathbf{X}_n) \rightarrow 0, \quad n \rightarrow \infty.$$

Применим левое неравенство (П.17) для матрицы $(\mathbf{X}'_n\mathbf{X}_n)^{-1}$; имеем

$$\max_{i,j} \{ |(\mathbf{X}'_n\mathbf{X}_n)^{-1}_{ij}| \} \leq \lambda_{\max}(\mathbf{X}'_n\mathbf{X}_n)^{-1} \rightarrow 0,$$

откуда $(\mathbf{X}'_n\mathbf{X}_n)^{-1} \rightarrow 0$, т. е. $\operatorname{cov}(\mathbf{a}_n) \rightarrow 0$, $n \rightarrow \infty$.

2) Пусть \mathbf{a}_n сходится к $\boldsymbol{\alpha}$ в среднем квадратичном, т. е. $\operatorname{cov}(\mathbf{a}_n) = \sigma^2(\mathbf{X}'_n\mathbf{X}_n)^{-1} \rightarrow 0$, $n \rightarrow \infty$. Применим правое неравенство (П.17) для матрицы $(\mathbf{X}'_n\mathbf{X}_n)^{-1}$, получим

$$\lambda_{\max}(\mathbf{X}'_n\mathbf{X}_n)^{-1} \leq \sum_{i,j} |(\mathbf{X}'_n\mathbf{X}_n)^{-1}_{ij}| \rightarrow 0,$$

откуда

$$\lambda_{\max}(\mathbf{X}'_n \mathbf{X}_n)^{-1} \rightarrow 0 \text{ и } \lambda_{\min}(\mathbf{X}'_n \mathbf{X}_n) \rightarrow \infty.$$

3. Доказательство теоремы 1.6. Обозначим

$$\mathbf{D}_n = \begin{bmatrix} \sqrt{\sum_i x_{i1}^2} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\sum_i x_{im}^2} \end{bmatrix};$$

тогда $\mathbf{R}_n = \mathbf{D}_n^{-1} \mathbf{X}'_n \mathbf{X}_n \mathbf{D}_n^{-1}$ и $\mathbf{X}'_n \mathbf{X}_n = \mathbf{D}_n \mathbf{R}_n \mathbf{D}_n$. Поскольку $|\mathbf{R}| \neq 0$, то $\lambda_{\min}(\mathbf{R}) > \delta > 0$, и начиная с некоторого n_0 для всех n $\lambda_{\min}(\mathbf{R}_n) \geq \delta^1$. Поэтому, применяя неравенство (П.12), получим $\lambda_{\min}(\mathbf{X}'_n \mathbf{X}_n) = \lambda_{\min}(\mathbf{D}_n \mathbf{R}_n \mathbf{D}_n) \geq \geq \lambda_{\min}(\mathbf{D}_n^2) \cdot \delta = \delta \min_i \sum_i x_{ii}^2 \rightarrow \infty$ по условию теоремы.

Итак, условие Эйкера выполнено, поэтому оценка МНК состоятельна.

4. Доказательство теоремы 1.7. Как следует из (1.78) и (1.79),

$$\begin{aligned} s^2 &= \frac{1}{n-m} \boldsymbol{\varepsilon}' (\mathbf{I}_n - \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n) \boldsymbol{\varepsilon} = \\ &= \frac{\sum \varepsilon_i^2}{n-m} - \frac{\boldsymbol{\varepsilon}' \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \boldsymbol{\varepsilon}}{n-m}. \end{aligned} \quad (1.80)$$

В силу некоррелируемости $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ первое слагаемое по закону больших чисел при $n \rightarrow \infty$ стремится по вероятности к σ^2 . Докажем, что второе слагаемое стремится по вероятности к нулю. В силу положительной определенности $\mathbf{X}'_n \mathbf{X}_n$ найдется невырожденная матрица \mathbf{P}_n , такая, что $\mathbf{P}_n \mathbf{P}'_n = (\mathbf{X}'_n \mathbf{X}_n)^{-1}$, откуда $\mathbf{P}'_n \mathbf{X}'_n \mathbf{X}_n \mathbf{P}_n = \mathbf{I}_m$.

Обозначим $\xi_n = \frac{1}{\sqrt{n-m}} \mathbf{P}'_n \mathbf{X}'_n \boldsymbol{\varepsilon}$, тогда второй член в (1.80) перепишется:

$$\frac{1}{n-m} \boldsymbol{\varepsilon}' \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \boldsymbol{\varepsilon} = \xi'_n \xi_n,$$

¹Мы пользуемся фактом непрерывности характеристических чисел матрицы от ее элементов [53; с. 206].

но $E\xi_n = 0$, поэтому

$$\text{cov}(\xi_n) = E\xi_n \xi_n' = \frac{\sigma^2}{n-m} P_n' X_n' X_n P_n = \frac{\sigma^2}{n-m} I_m \rightarrow 0,$$

т. е. $\text{plim} \xi_n = 0$, поэтому $\text{plim} s_n^2 = \sigma^2$.

5. Доказательство теоремы 1.8 основано на теореме сходимости суммы элементарной системы случайных величин (см. [18, с. 288]). Фиксируем $i \in [1, m]$. Тогда i -я координата стандартизированной оценки МНК равна: $b_{ni} = \sum_{t=1}^n A_{nit} \varepsilon_t = \sum_{t=1}^n \xi_{nt}$, где $\xi_{nt} = A_{nit} \varepsilon_t$, A_{nit} — (i, t) -й элемент матрицы A_n , $1 \leq t \leq n$. Проверим выполнимость аксиом, позволяющих считать систему случайных величин $\{\xi_{nt}\}$ элементарной [18, с. 283]. Прежде всего заметим, что $\sum_t A_{nit}^2 = 1/\sigma^2$, что следует из $A_n A_n' = I/\sigma^2$.

Далее,

- 1) $\sigma^2(\xi_{nt}) = A_{nit}^2 \cdot \sigma^2 \leq 1$;
- 2) $\sigma^2(b_{ni}) = 1$;
- 3) $\max \{A_{ni1}^2, A_{ni2}^2, \dots, A_{nin}^2\} \rightarrow 0$

— условие теоремы. Применим теперь теорему 2 [18, с.288]. Условие ее переписывается следующим образом:

$$\lim_{n \rightarrow \infty} \sum_{t=1}^n A_{nit}^2 \int_{|x| > \tau / |A_{nit}|} x^2 dF = 0, \quad (1.81)$$

где F — функция распределения ε_i . Таким образом, для доказательства теоремы достаточно показать, что условия (1.81) и (1.43) эквивалентны.

Достаточность. Докажем сначала, что для любого $\varepsilon > 0$ найдется такое N , что для всех $n \geq N$ и $t \leq n$

$$\int_{x^2 > \tau^2 / A_{nit}^2} x^2 dF < \varepsilon. \quad (1.82)$$

Пусть $\varepsilon_1 > 0$ такое, что $\int_{x^2 > \tau^2 / \varepsilon_1^2} x^2 dF < \varepsilon$. Пусть N такое, что для любого $n \geq N$ $\max_t \{A_{nit}^2\} \leq \varepsilon_1^2$, т. е. $A_{nit}^2 \leq \varepsilon_1^2$ для $t \leq n$, $n \geq N$. Имеем

$$\int_{x^2 > \tau^2 / A_{nit}^2} x^2 dF \leq \int_{x^2 > \tau^2 / \varepsilon_1^2} x^2 dF < \varepsilon,$$

откуда и следует (1.82). Далее, начиная с N для любого $n \geq N$

$$\sum_{i=1}^n A_{nit}^2 \int_{|x| > \tau / |A_{nit}|} x^2 dF \leq \varepsilon \sum_{i=1}^n A_{nit}^2 = \varepsilon / \sigma^2,$$

т. е. условие (1.81) выполнено.

Необходимость. Допустим, (1.43) не имеет места. Тогда найдется такая последовательность $1 \leq t_n \leq n$, что $A_{nit_n}^2 \geq \varepsilon_1^2$ для всех $n \geq N$. В таком случае

$$\int_{x^2 > \tau^2 / A_{nit_n}^2} x^2 dF \geq \int_{x^2 > \tau^2 / \varepsilon_1^2} x^2 dF \geq \delta > 0.$$

Поэтому

$$\begin{aligned} \sum_{i=1}^n A_{nit}^2 \int_{|x| > \tau / |A_{nit}|} x^2 dF &\geq A_{nit_n}^2 \int_{|x| > \tau / |A_{nit_n}|} x^2 dF \geq \\ &\geq \varepsilon_1^2 \delta > 0, \end{aligned}$$

т. е. условие (1.81) не выполняется. Теорема доказана.
З а м е ч а н и е. Доказательство теоремы нельзя считать полным в достаточной мере, так как мы доказали асимптотическую нормальность только i -й координаты оценки МНК. Для полноты доказательства теперь необходимо рассмотреть линейную комбинацию координат оценок МНК и доказать, что она асимптотически нормальна. Доказательство будет аналогично приведенному. По известной теореме тогда и весь вектор оценки МНК будет асимптотически нормален.

6. П р и м е р р е г р е с с и и, в которой матрица X_n сильно регулярна, а оценка МНК не асимптотически нормальна. Для простоты рассмотрим регрессию с одним оцениваемым параметром: $y_t = \alpha x_t + \varepsilon_t$. Отклонения ε_t считаем независимыми и одинаково распределенными. Пусть последовательность x_t такова:

$$x_t = \begin{cases} \text{если } t = 2^k, \text{ то } \sqrt{t}, \\ \text{в противном случае } 1, \end{cases}$$

где $k = 0, 1, 2, \dots$ Тогда

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n x_t^2 &= \frac{1}{n} \left(\sum_{t=2^k}^n x_t^2 + \sum_{t \neq 2^k}^n x_t^2 \right) = \\ &= \frac{1}{n} \left(\sum_{k < \log_2 n} k + \sum_{t \neq 2^k}^n 1 \right) = \frac{1}{n} [\log_2(n)] ([\log_2(n)] + 1)/2 + \\ &+ \frac{1}{n} (n - \log_2 n) = 1 + \frac{1}{n} [\log_2 n] ([\log_2 n] + 1)/2 - \frac{\log_2 n}{n}, \end{aligned}$$

где $[\cdot]$ означает целую часть числа.

Нетрудно проверить, что последняя сумма имеет предел 1 при $n \rightarrow \infty$. С другой стороны, если взять $n = 2^k$, то для построенной последовательности $\frac{1}{n} \max_{t \leq n} x_t^2 = 1$. Таким образом,

условие теоремы 1.8 не выполняется, и оценка МНК не будет асимптотически нормальной.

7. Доказательство теоремы 1.10. Прежде всего более подробно рассмотрим, как условие (1.39) связано с (1.43). Для этого обозначим \mathbf{X}_{nt} — t -ю вектор-строку $1 \times m$ матрицы \mathbf{X}_n ; \mathbf{A}_{nt} — t -й вектор-столбец $m \times 1$ матрицы \mathbf{A}_n , $t = 1, 2, \dots, n$. По определению (1.42)

$$\begin{aligned} \mathbf{A}_{nt} &= \frac{1}{\sigma} (\mathbf{X}'_n \mathbf{X}_n)^{-1/2} \mathbf{X}'_{nt}; \\ \|\mathbf{A}_{nt}\|^2 &= \frac{1}{\sigma^2} \mathbf{X}_{nt} (\mathbf{X}'_n \mathbf{X}_n)^{-1/2} (\mathbf{X}'_n \mathbf{X}_n)^{-1/2} \mathbf{X}'_{nt} = \\ &= \frac{1}{\sigma^2} \mathbf{X}_{nt} (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_{nt}. \end{aligned} \quad (1.83)$$

Теперь покажем, что условие (1.43) эквивалентно

$$\max_t \|\mathbf{A}_{nt}\|^2 \rightarrow 0, \quad n \rightarrow \infty. \quad (1.84)$$

Этот факт следует из очевидных неравенств

$$A_{nit}^2 \leq \|\mathbf{A}_{nt}\|^2 \leq m \cdot \max_i A_{nit}^2.$$

Условие (1.84) теперь позволяет понять, почему (1.43) сильнее (1.39). Из (1.83) следует

$$\begin{aligned} \|\mathbf{A}_{nt}\|^2 &\leq \frac{1}{\sigma^2} \|\mathbf{X}_{nt}\|^2 \cdot \lambda_{\max} (\mathbf{X}'_n \mathbf{X}_n)^{-1} = \\ &= \frac{1}{\sigma^2} \|\mathbf{X}_{nt}\|^2 / \lambda_{\min} (\mathbf{X}'_n \mathbf{X}_n). \end{aligned} \quad (1.85)$$

Для того чтобы выполнялось условие (1.84), необходимо, чтобы не только $\lambda_{\min}(\mathbf{X}'_n \mathbf{X}_n) \rightarrow \infty$, но чтобы отношение $\|\mathbf{X}_{nt}\|^2 / \lambda_{\min}(\mathbf{X}'_n \mathbf{X}_n)$ равномерно стремилось к нулю.

По определению \mathbf{R}_n имеем $(\mathbf{X}'_n \mathbf{X}_n)^{-1} = \mathbf{D}_n^{-1} \mathbf{R}_n^{-1} \mathbf{D}_n^{-1}$, откуда по неравенству (П.11)

$$\begin{aligned} \lambda_{\max}(\mathbf{X}'_n \mathbf{X}_n)^{-1} &\leq \lambda_{\max}(\mathbf{D}_n^{-2}) \cdot \lambda_{\max}(\mathbf{R}_n^{-1}) = \\ &= 1/\lambda_{\min}(\mathbf{R}_n) \cdot \min_i \sum_t x_{ti}^2 \leq 1/\delta \min_i \sum_t x_{ti}^2. \end{aligned}$$

Поэтому из (1.85) следует

$$\max_t \|\mathbf{A}_{nt}\|^2 \leq \frac{1}{\delta \cdot \sigma^2} \cdot \frac{\max_t \|\mathbf{X}_{nt}\|^2}{\min_i \sum_t x_{ti}^2} \leq \frac{m}{\delta \cdot \sigma^2} \cdot \frac{\max_{t,t} x_{ti}^2}{\min_i \sum_t x_{ti}^2}.$$

Сходимость (1.84) следует из условия б) теоремы.

8. Доказательство теоремы 1.13. Логарифм функции плотности (1.44) равен: $\ln f(\mathbf{y}; \alpha, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\alpha)' (\mathbf{y} - \mathbf{X}\alpha)$, ее производные по α и σ^2 равны:

$$\frac{\partial \ln f}{\partial \alpha} = \frac{1}{\sigma^2} (\mathbf{X}' \mathbf{y} - \mathbf{X}' \mathbf{X} \alpha) = \frac{1}{\sigma^2} \mathbf{X}' (\mathbf{y} - \mathbf{X} \alpha) = \frac{1}{\sigma^2} \mathbf{X}' \boldsymbol{\varepsilon};$$

$$\frac{\partial \ln f}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{(\mathbf{y} - \mathbf{X}\alpha)' (\mathbf{y} - \mathbf{X}\alpha)}{2\sigma^4} = -\frac{n}{2\sigma^2} + \frac{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{2\sigma^4}.$$

Число искомых параметров равно $m + 1$; обозначим общий вектор параметров $\theta \in R^{m+1}$:

$$\theta = \begin{bmatrix} \alpha \\ \sigma^2 \end{bmatrix};$$

$$\text{тогда } \frac{\partial \ln f}{\partial \theta} = \begin{bmatrix} \frac{1}{\sigma^2} \mathbf{X}' \boldsymbol{\varepsilon} \\ -\frac{n}{2\sigma^2} + \frac{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{2\sigma^4} \end{bmatrix}.$$

Найдем

$$\mathbf{E}_\theta (\partial \ln f / \partial \theta) (\partial \ln f / \partial \theta)' = \mathbf{E} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \mathbf{I}(\theta).$$

Имеем

$$EA_{11} = E \frac{1}{\sigma^4} X' \varepsilon \varepsilon' X = \frac{1}{\sigma^4} X' E \varepsilon \varepsilon' X = \frac{1}{\sigma^2} (X' X);$$

$$A_{12} = \frac{1}{2\sigma^4} X' \varepsilon (-n + \varepsilon' \varepsilon / \sigma^2) \in R^m;$$

$$\begin{aligned} E(A_{12})_t &= E \frac{1}{2\sigma^4} \sum_t x_{ti} \varepsilon_t \left(-n + \sum_t \varepsilon_t^2 / \sigma^2 \right) = \\ &= -\frac{n}{2\sigma^2} E \sum_t x_{ti} \varepsilon_t + \frac{1}{2\sigma^6} E \sum_t x_{ti} \varepsilon_t \cdot \sum_t \varepsilon_t^2. \end{aligned}$$

В силу того что $E\varepsilon_t = 0$, первое слагаемое обращается в нуль. В силу независимости ε_t и ε_s $E x_{ti} \varepsilon_t \varepsilon_s^2 = x_{ti} E \varepsilon_t \times E \varepsilon_s^2 = 0$, поэтому второе слагаемое равно:

$$\frac{1}{2\sigma^6} E \sum_t x_{ti} \varepsilon_t^3 = \frac{1}{2\sigma^6} \sum_t x_{ti} E \varepsilon_t^3 = 0,$$

так как для нормального распределения 3-й момент равен нулю. Итак, $EA_{12} = 0$, поэтому

$$I^{-1}(\theta) = \begin{bmatrix} (EA_{11})^{-1} & 0 \\ 0 & (EA_{22})^{-1} \end{bmatrix} = \begin{bmatrix} \sigma^2 (X' X)^{-1} & 0 \\ 0 & (EA_{22})^{-1} \end{bmatrix}.$$

Но матрица ковариации оценки МНК как раз равна $\sigma^2 (X' X)^{-1}$, т. е. нижняя граница в неравенстве Крамера—Рао (1.15), (1.16) достигается, и оценка МНК будет эффективна в классе всех несмещенных оценок — теорема доказана.

Глава 2

ДРУГИЕ ВОПРОСЫ ЛИНЕЙНОЙ РЕГРЕССИИ

2.1. Взвешенный МНК. Оценка Эйткена

В классической регрессии предполагается, что отклонения в регрессии не коррелируют друг с другом (предположение Γ параграфа 1.1). Это довольно жесткое условие, которое весьма редко выполняется для временных рядов; вероятно тогда, что отклонения t -го момента времени тесно связаны с отклонениями $(t+1)$ -го момента. Ослабим

предположение Γ и допустим, что ковариационная матрица отклонений не обязательно имеет вид $\sigma^2 \mathbf{I}$, хотя и известна с точностью до постоянного множителя.

Предположение Γ' . $\text{cov}(\varepsilon) = \sigma^2 \Omega$, где Ω — известная положительно определенная матрица, $\sigma^2 > 0$ — неизвестный параметр. На протяжении этого параграфа будем считать, что предположение Γ' выполнено.

Уравнение регрессии

$$y = X\alpha + \varepsilon, \text{cov}(\varepsilon) = \sigma^2 \Omega \quad (2.1)$$

некоторым преобразованием может быть сведено к классическому

$$v = Z\alpha + \xi, \text{cov}(\xi) = \sigma^2 I_n. \quad (2.2)$$

Известно, что любую положительно определенную матрицу Ω можно представить в виде $\Omega = \mathbf{T}\mathbf{T}'$, где \mathbf{T} — невырожденная матрица, алгоритм вычисления которой дается в [27, с. 287]. Положим

$$v = \mathbf{T}^{-1}y; Z = \mathbf{T}^{-1}X; \xi = \mathbf{T}^{-1}\varepsilon. \quad (2.3)$$

Легко проверить, что для нового уравнения регрессии выполняются все предположения А — Е параграфа 1.1, при этом параметры, которые необходимо оценивать, остаются без изменения. Естественно для их оценивания воспользоваться МНК, который применим к уравнению (2.2):

$$Q(\alpha) = (v - Z\alpha)'(v - Z\alpha) = (y - X\alpha)' \Omega^{-1} (y - X\alpha) \Rightarrow \min. \quad (2.4)$$

Как видно из этого выражения, для оценивания α в исходном уравнении (2.1) необходимо применять *взвешенный* МНК. Суть его нетрудно понять. Матрица ковариаций Ω представляет собой совокупность ковариаций и дисперсий y_t . Чем больше разброс y_t , тем менее это наблюдение должно учитываться при оценивании параметров регрессии. Оценку взвешенного (обобщенного) МНК можно получить, непосредственно минимизируя (2.4) или же как обычную оценку МНК из «нормализованной» регрессии (2.2):

$$\begin{aligned} \hat{\alpha} &= (Z'Z)^{-1}Z'v = [X'(\mathbf{T}^{-1})'\mathbf{T}^{-1}X]^{-1}X'(\mathbf{T}^{-1})'\mathbf{T}^{-1}y = \\ &= [X'(\mathbf{T}\mathbf{T}')^{-1}X]^{-1}X'(\mathbf{T}\mathbf{T}')^{-1}y = \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y. \end{aligned} \quad (2.5)$$

В экономико-статистической литературе оценку (2.5) называют оценкой Эйткена [71], в теоретической литературе по математической статистике — оценкой Гаусса—Мар-

кова. Мы ее будем называть *оценкой Эйткена*. Читатель может убедиться в том, что

$$\text{cov}(\hat{\alpha}) = \sigma^2 (\mathbf{X}' \mathbf{\Omega}^{-1} \mathbf{X})^{-1}. \quad (2.6)$$

Аналогично теореме Гаусса—Маркова в случае $\text{cov}(\varepsilon) = \sigma^2 \mathbf{\Omega}$ может быть доказана следующая теорема.

Теорема 2.1 (обобщенная теорема Гаусса—Маркова). *Оценка Эйткена является:*

- а) *несмещенной;*
- б) *эффективной в классе несмещенных оценок, линейных по у.* Несмещенной оценкой σ^2 в нашем случае является

$$\hat{\sigma}^2 = (\mathbf{y} - \mathbf{X}\hat{\alpha})' \mathbf{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\hat{\alpha}) / (n - m). \quad (2.7)$$

Теорема 2.2. *Если отклонения нормальны, то:*

- а) *оценка Эйткена совпадает с оценкой ММП;*
- б) *оценка Эйткена имеет распределение $N(\alpha, \sigma^2 \times (\mathbf{X}' \mathbf{\Omega}^{-1} \mathbf{X})^{-1})$;*
- в) $\hat{\sigma}^2 (n - m) / \sigma^2 \sim \chi^2 (n - m)$;
- г) $\hat{\sigma}^2$ и $\hat{\alpha}$ *независимо распределены;*
- д) *оценка Эйткена является эффективной в классе всех несмещенных оценок.*

Аналогично (1.39) в условиях предположения Γ' возможно разложение

$$\begin{aligned} (\mathbf{y} - \hat{\mathbf{y}}\mathbf{1})' \mathbf{\Omega}^{-1} (\mathbf{y} - \bar{\mathbf{y}}\mathbf{1}) &= (\hat{\mathbf{y}} - \bar{\mathbf{y}}\mathbf{1})' \mathbf{\Omega}^{-1} (\hat{\mathbf{y}} - \bar{\mathbf{y}}\mathbf{1}) + \\ &+ (\mathbf{y} - \mathbf{X}\hat{\alpha})' \mathbf{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\hat{\alpha}). \end{aligned} \quad (2.8)$$

Можно определить эффект перехода от модели $y_t = \alpha_m + \varepsilon_t$ к модели (2.1) со свободным членом

$$R^2 = 1 - \frac{(\mathbf{y} - \mathbf{X}\hat{\alpha})' \mathbf{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\hat{\alpha})}{(\mathbf{y} - \bar{\mathbf{y}}\mathbf{1})' \mathbf{\Omega}^{-1} (\mathbf{y} - \bar{\mathbf{y}}\mathbf{1})}. \quad (2.9)$$

Процедура проверок гипотез и доверительного оценивания также практически мало изменится при переходе к модели (2.1). Всюду вместо обычной суммы квадратов отклонений необходимо брать взвешенную. Например, совместный доверительный интервал для α с коэффициентом доверия $1 - \lambda$ равен:

$$\mathbf{D} = \{ \beta \in R^m: (\beta - \hat{\alpha})' \mathbf{X}' \mathbf{\Omega}^{-1} \mathbf{X} (\beta - \hat{\alpha}) \leq m \hat{\sigma}^2 F_{\lambda}(m, n - m) \}, \quad (2.10)$$

где $\hat{\sigma}^2$ рассчитывается по формуле (2.7).

Из асимптотических свойств оценки Эйткена остановимся только на состоятельности.

Теорема 2.3. При условии Эйкера (1.39) и условии $\lambda_{\max}(\Omega_n) \leq d$ оценка Эйткена состоятельна.

Доказательство опирается на неравенство (П.12). Используя его, получим

$$\begin{aligned} \lambda_{\min}(\mathbf{X}'_n \Omega_n^{-1} \mathbf{X}_n) &\geq \lambda_{\min}(\Omega_n^{-1}) \cdot \lambda_{\min}(\mathbf{X}'_n \mathbf{X}_n) \geq \\ &\geq \frac{1}{d} \lambda_{\min}(\mathbf{X}'_n \mathbf{X}_n) \rightarrow \infty, \end{aligned}$$

откуда следует, что

$$\text{cov}(\hat{\mathbf{a}}_n) = \sigma^2 (\mathbf{X}'_n \Omega_n^{-1} \mathbf{X}_n)^{-1} \rightarrow \mathbf{0}, \quad n \rightarrow \infty.$$

Использование эффективной оценки Эйткена предполагает знание матрицы Ω . На практике она, как правило, неизвестна. Оценить Ω на основе наблюдения y также невозможно хотя бы уже потому, что для этого необходимо оценить $n(n+1)/2$ величин по имеющейся информации из n наблюдений.

Рассмотрим, какие свойства будет иметь обычная оценка МНК, если все-таки $\Omega \neq \mathbf{I}$.

Теорема 2.4. Если все предположения А — Е (за исключением Г) верны, $\text{cov}(\mathbf{e}) = \sigma^2 \Omega$, то оценка МНК является:

а) несмещенной;

б) состоятельной, если $\lambda_{\max}(\Omega_n) \leq d < \infty$.

Доказательство. а) $\mathbf{a} = \boldsymbol{\alpha} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}$, $\mathbf{E}\mathbf{a} = \boldsymbol{\alpha}$.

б) Прежде всего найдем матрицу ковариаций оценки МНК в условиях предположения Г'. Имеем

$$\begin{aligned} \text{cov}(\mathbf{a}) &= \mathbf{E}(\mathbf{a} - \boldsymbol{\alpha})(\mathbf{a} - \boldsymbol{\alpha})' = \mathbf{E}(\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{e} \mathbf{e}' \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} = \\ &= \sigma^2 (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \Omega_n^{-1} \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1}. \end{aligned}$$

Применяя неравенство (П.11), получим

$$\begin{aligned} \lambda_{\max} \text{cov}(\mathbf{a}) &\leq \sigma^2 \lambda_{\max}((\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1}) \lambda_{\max}(\Omega_n) = \\ &= \sigma^2 \lambda_{\max}(\mathbf{X}'_n \mathbf{X}_n)^{-1} \times \lambda_{\max}(\Omega_n) \leq \\ &\leq \sigma^2 \lambda_{\max}(\mathbf{X}'_n \mathbf{X}_n)^{-1} \cdot d \rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

Условие состоятельности довольно естественно: необходимо, чтобы дополнительная информация, даваемая наблюдениями при $n \rightarrow \infty$, не убывала до нуля.

В литературе предлагаются достаточные условия состоятельности оценки МНК в случае, когда отклонения

регрессии являются стационарными, т. е. когда матрица $\sigma^2 \Omega$ имеет вид [3]:

$$\sigma^2 \Omega = \sigma^2 \begin{bmatrix} \omega_0 & \omega_1 & \omega_2 & \dots & \omega_n \\ & \omega_0 & \omega_1 & \dots & \omega_{n-1} \\ & & \dots & \dots & \\ & & & \dots & \\ & & & & \omega_0 \end{bmatrix}. \quad (2.11)$$

Эти условия похожи на условия теоремы 1.6.

Рассмотрим частный случай использования взвешенного МНК, на котором отчетливо видна суть взвешивания. Допустим, в предположении Γ' матрица Ω является диагональной и $\Omega_{tt} = \sigma_t^2$. Другими словами, мы по-прежнему, как и в классическом предположении Γ , считаем отклонения $\{\varepsilon_t\}$ (дисперсии которых известны статистику с точностью до постоянного множителя) некоррелируемыми, однако имеющими разную дисперсию. Такие отклонения называют гетероскедастичными. Взвешенный МНК приводит к минимизации (2.4) или

$$\sum_t (y_t - \alpha_1 x_{t1} - \dots - \alpha_m x_{tm})^2 / \sigma_t^2 = \sum_t (y_t / \sigma_t - \alpha_1 x_{t1} / \sigma_t - \dots - \alpha_m x_{tm} / \sigma_t)^2. \quad (2.12)$$

Для нахождения оценки параметров можно было бы воспользоваться формулой (2.5). Однако, как видно из (2.12), оценка Эйткена (2.5) совпадает с оценкой МНК взвешенной регрессии

$$\frac{y_t}{\sigma_t} = \alpha_1 \frac{x_{t1}}{\sigma_t} + \dots + \alpha_m \frac{x_{tm}}{\sigma_t} + \frac{\varepsilon_t}{\sigma_t},$$

или

$$v_t = \alpha_1 z_{t1} + \dots + \alpha_m z_{tm} + \xi_t, \quad t = 1, \dots, n \quad (2.13)$$

Легко видеть, что все предположения А — Е для взвешенной регрессии (2.13) выполнены, причем $\text{cov}(\xi) = \sigma^2 I_n$. Можно показать, что $\text{cov}(\hat{\alpha})$, R^2 , доверительные интервалы для уравнения (2.1) равны соответствующим величинам из взвешенной регрессии (2.13).

Вернемся к регрессии-примеру (1.5). Относительно отклонения ε_t примем следующую гипотезу: отклонения не коррелируют друг с другом, но имеют разные дисперсии. Будем считать, что стандартное отклонение y_t пропорционально (приближенно) самому значению y_t . Это имеет основание, например, если принимается, что y_t содержит

только ошибки измерения. Тогда мы считаем, что относительная ошибка измерения y_t постоянна. Поскольку y_t растет одновременно с номером эксперимента, можно положить $\sigma_t = \sigma_0 t$, где $\sigma_0 = \text{const}$, т. е. $\sigma_t^2 = \sigma_0^2 t^2$. Примем эту гипотезу. Для того чтобы найти оценки МНК, рассмотрим регрессию

$$\frac{y_t}{t} = \alpha_1 \frac{x_{t1}}{t} + \alpha_2 \frac{x_{t2}}{t} + \alpha_3 \frac{x_{t3}}{t} + \alpha_4 \frac{1}{t} + \frac{\varepsilon_t}{t}, \quad t=1, \dots, 15.$$

Оценим ее методом наименьших квадратов:

$$y_t = 0,453 \quad x_{t1} + 0,917 \quad x_{t2} + 4,17 \quad x_{t3} + 78,8, \\ (0,0721) \quad (0,909) \quad (4,41) \quad (58,4)$$

$\hat{\sigma}_0^2 = 0,63$. Все параметры, за исключением a_1 , оказываются незначимыми.

Иногда гетероскедастичность отклонений связывают с одним или несколькими независимыми переменными. Например, Дж. Джонстон [26] в парной регрессии $y_t = \alpha_1 x_t + \alpha_2 + \varepsilon_t$ предлагает выбирать $\sigma_t^2 = \sigma_0^2 x_t^2$. Р. Парк [170] обобщает зависимость σ_t от x_t и предлагает считать $\sigma_t^2 = \sigma_0^2 x_t^\nu$, где $x_t > 0$ и ν — неизвестный параметр, который находится следующим образом. Сначала находим обычную оценку МНК a_1 и a_2 ; строим отклонения $e_t = y_t - a_1 \hat{x}_t - a_2$. Далее рассуждаем следующим образом: e_t^2 может служить оценкой σ_t^2 , поэтому можно приближенно записать $e_t^2 \approx \sigma_0^2 x_t^\nu$ или $\ln e_t^2 = \ln \sigma_0^2 + \nu \ln x_t + \eta_t$, $t = 1, 2, \dots, n$. Рассматривая это равенство как регрессию $\ln e_t^2$ на $\ln x_t$, можно найти оценку МНК для σ_0^2 и ν . Затем применяем взвешенный МНК к исходной регрессии с весом x_t^ν . В [111] предлагается схема, в которой σ_t^2 есть функция некоторого полинома от x_t . Очевидно, эти приемы можно перенести и на случай множественной регрессии.

И все же описанные спецификации гетероскедастичности нам кажутся искусственными и мало правдоподобными. Почему дисперсия отклонений зависит от x_t ? По нашему предположению x_t детерминирован. Если же x_t содержит ошибки измерения, то оценка МНК вообще неприемлема (это показано в гл. 4). Почему дисперсия отклонений пропорциональна степени независимой переменной? Чем это аргументируется? Нам кажется, что при спецификации гетероскедастичности ε_t необходимо отталкиваться от зависимой переменной y_t , а не от независимых x_1, x_2, \dots, x_m , по-

скольку в наших предположениях только y_t может содержать ошибки измерения.

Предположим, σ_t пропорционально $|Ey_t|$, т. е.

$$\sigma_t^2 = \sigma_0^2 (\alpha_1 x_{t1} + \dots + \alpha_m x_{tm})^2, \quad t = 1, \dots, n. \quad (2.14)$$

Гипотеза (2.14) весьма естественна, например, когда отклонения ε_t трактуются как ошибки измерения зависимой переменной. Тогда (2.14) утверждает, что относительная ошибка постоянна, т. е. отношение $\sigma_t/|Ey_t| = \text{const}$. Гипотеза (2.14), очевидно, противоречит независимости Ω от α . Для нахождения оценок α можно воспользоваться различными методами. Большинство из них приводят к нелинейному оцениванию (см. часть III книги). Однако можно предложить следующую двухшаговую процедуру, которая использует только МНК:

1) оцениваем обычным МНК исходную регрессию;

2) находим $\hat{y}_t = a_1 x_{t1} + \dots + a_m x_{tm}$; далее полагаем $\hat{\sigma}_t = |\hat{y}_t|$ и применяем взвешенный МНК.

В [180] предложена более общая схема, в которой стандартная ошибка отклонений регрессии есть линейная комбинация независимых переменных. Другими словами,

$$\sigma_t = |\beta_0 + \beta_1 x_{t1} + \dots + \beta_m x_{tm}|, \quad (2.15)$$

где $\beta_0, \beta_1, \dots, \beta_m$ — неизвестные коэффициенты, подлежащие оцениванию. Схема (2.15) является более общей, чем схема (2.12): если в последней мы считаем, что σ_t зависит от Ey_t , т. е. является линейной комбинацией x_1, \dots, x_m с коэффициентами регрессии, то в (2.15) эти коэффициенты могут быть любыми. Схема (2.15) применима, если предполагается, что гетероскедастичность зависит от переменных регрессионной модели x_1, \dots, x_m , но вид зависимости неизвестен. Для оценивания регрессии в предположении (2.15) можно применить ММП. Предположим, отклонения ε_t нормальны, т. е. $\varepsilon_t \sim N(0, \sigma_t^2) = N(0, (\beta_0 + \beta_1 x_{t1} + \dots + \beta_m x_{tm})^2)$. Тогда можно найти функцию плотности распределения выборки y_1, \dots, y_n , которая зависит от $2m + 1$ неизвестных параметров $\alpha_1, \dots, \alpha_m, \beta_0, \beta_1, \dots, \beta_m$. Максимизируя эту плотность, найдем оценки ММП для α и β .

В литературе предлагаются некоторые методы выявления гетероскедастичных наблюдений и борьбы с ними [75, 76]. Иногда для выявления гетероскедастичности строят график квадратов отклонений ε_t .

В некоторых случаях оценка МНК совпадает с оценкой Эйткена, даже если $\Omega \neq I$.

Теорема 2.5. Оценка Эйткена (2.5) и оценка МНК совпадают тогда и только тогда, когда существуют такие невырожденная матрица $C^{m \times m}$ и матрица $V^{n \times m}$, причем вектор-столбцы последней линейно независимы и являются характеристическими векторами матрицы Ω , что

$$X = VC. \quad (2.16)$$

Доказательство этой теоремы дано в [6].

Выражение (2.16) можно трактовать так: независимые переменные суть линейная комбинация некоторых характеристических векторов матрицы Ω . Рассмотрим для примера простейшую регрессию $y_t = \alpha x_t + \varepsilon_t$, где $t = 1, 2$ и $\text{cov}(\varepsilon) = \sigma^2 \Omega$ ($m = 1, n = 2$). Допустим, характеристический вектор, соответствующий максимальному характеристическому числу Ω , равен e_1 , минимальному — e_2 . Оценки МНК и Эйткена совпадают, если $x = (x_1, x_2)$ лежит либо на e_1 , либо на e_2 . В противном случае оценка Эйткена эффективнее.

С помощью теоремы 2.5 можно находить условия на матрицу независимых переменных X , для которой оценка МНК и оценка Эйткена совпадают.

Пример. Рассмотрим модель регрессии (2.1). Отклонения регрессии считаем гетероскедастичными и независимыми. Другими словами, $\text{cov}(\varepsilon) = \sigma^2 \Omega$ — диагональная матрица:

$$\Omega = \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \sigma_n^2 \end{bmatrix}. \quad (2.17)$$

Значения σ_i^2 отличны друг от друга: $\sigma_i^2 \neq \sigma_j^2$, если $i \neq j$. Спрашивается, когда оценка МНК и оценка Эйткена будут совпадать? С помощью теоремы 2.5 ответ найти нетрудно: $n - m$ строк матрицы X должны быть нулевыми. Докажем это утверждение. Прежде всего отметим, что характеристическими векторами матрицы (2.17) являются e_1, \dots, e_n , где $e_i = (0, \dots, 1, \dots, 0)$, а единица расположена на i -м месте. Вектору e_i соответствует характеристическое число σ_i^2 . Пусть вектор-столбцы матрицы $V^{n \times m}$ составлены из e_1, \dots, e_n — матрица полного ранга. Оценка МНК и

оценка Эйткена совпадают тогда и только тогда, когда матрица X может быть представлена в виде произведения:

$$VC = \begin{bmatrix} 0 & 0 & \dots & 0 \\ \vdots & 1 & & \vdots \\ 1 & 0 & & \\ \vdots & \vdots & & 1 \\ 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} C_{11} & \dots & C_{1m} \\ \vdots & & \vdots \\ C_{m1} & \dots & C_{mm} \end{bmatrix} = X.$$

Число нулевых строк матрицы V равно $n - m$. Столь же нулевых строк будет иметь матрица X .

В частности, при $m = 1$ оценка МНК и оценка Эйткена совпадают, только если ряд $\{x_t\}$ содержит лишь одно ненулевое значение. В задачах 7 и 8 упражнения 2.1 исследуются другие возможности совпадения двух оценок.

Нетрудно догадаться, что чем больше матрица ковариаций-отклонений регрессии имеет кратных корней, тем вероятнее оценки МНК и Эйткена будут совпадать. Однако совпадение двух оценок — факт весьма редкий на практике.

У п р а ж н е н и я 2.1

1. Получите оценку Эйткена, непосредственно минимизируя (2.4).

2. Докажите, что в условиях предположения Е (параграф 1.1) оценка Эйткена единственна.

3. Покажите, что доверительные интервалы параметров регрессий (2.1) и (2.2) с преобразованиями (2.3) совпадают.

4. В модели среднего $y_t = \alpha_m + \varepsilon_t$, где Ω — диагональная матрица, $\sigma_t^2 = \sigma_0^2 t^2$. Покажите, что ни оценка Эйткена, ни оценка МНК не будут состоятельными.

5. Можно показать, что если $\text{cov}(\varepsilon) = \sigma^2 I$, то оценка МНК в регрессии $y_t = \alpha/t + \varepsilon_t$, несостоятельна. Предположим теперь, что $\text{cov}(\varepsilon)$ — диагональная матрица и $\sigma^2(\varepsilon_t) = \sigma_0^2/t^2$. Покажите, что оценка МНК по-прежнему будет несостоятельной, а оценка Эйткена — состоятельной.

6. Чему равна оценка МНК в регрессии $y_t = \alpha x_t + \varepsilon_t$, где $\sigma^2(\varepsilon_t) = kx_t^2$, $\{\varepsilon_t\}$ не коррелируют?

7. Допустим, отклонения гетероскедастичны и $\sigma_1^2 = \dots = \sigma_k^2 \neq \sigma_{k+1}^2 = \dots = \sigma_n^2$. Какой должна быть матрица X , чтобы оценка МНК совпадала с оценкой Эйткена? Начните со случая $m = 1$.

8. При каких $\{x_t\}$ оценки МНК и Эйткена в регрессии задачи 6 совпадают?

2.2. Прогноз по регрессии

На основе регрессионной модели

$$y_t = \alpha_1 x_{t1} + \dots + \alpha_m x_{tm} + \varepsilon_t, \quad t = 1, \dots, n \quad (2.18)$$

можно находить прогноз зависимой (эндогенной) переменной y , зная соответствующие значения независимых (экзогенных) переменных x_1, \dots, x_m . Допустим, для линейной регрессии (2.18) выполняются все предположения А — Е. По статистике за прошлое мы можем оценить вектор α . Перед нами стоит задача оценки y_τ , где $\tau > n$. Предположим, прогноз вектора независимых переменных $x_\tau = (x_{\tau 1}, \dots, x_{\tau m})'$ известен точно. В качестве оценки y_τ естественно рассмотреть

$$\hat{y}_\tau = x_\tau' a = \sum_{i=1}^m a_i x_{\tau i}. \quad (2.19)$$

Поскольку y_τ есть величина случайная, то употребление термина «оценивание» будет несколько некорректно. Говорим, что некоторая оценка несмещенно оценивает y_τ , если математическое ожидание этой оценки равно $E y_\tau = \alpha_1 x_{\tau 1} + \dots + \alpha_m x_{\tau m}$.

Т е о р е м а 2.6. Прогноз (2.19) является несмещенным, эффективным в классе линейных несмещенных прогнозов с дисперсией¹

$$\sigma_\tau^2 = \sigma^2 (1 + x_\tau' (X' X)^{-1} x_\tau). \quad (2.20)$$

Д о к а з а т е л ь с т в о. Несмещенность прогноза доказывается просто:

$$E \hat{y}_\tau = x_\tau' E a = x_\tau' \alpha = E y_\tau.$$

Докажем его эффективность. По условию прогноз y_τ будем искать в классе линейных несмещенных прогнозов. Необходимо найти наилучший прогноз \hat{y}_τ . По договоренности \hat{y}_τ может быть записан в виде линейной комбинации предыдущих наблюдений y_1, y_2, \dots, y_n , т. е.

$$\tilde{y}_\tau = \sum_i c_i y_i = c' y, \quad (2.21)$$

где $c = (c_1, \dots, c_n)'$ — вектор коэффициентов при y . В силу несмещенности (2.21) $E \tilde{y}_\tau = c' E y = c' X \alpha = x_\tau' \alpha$ для

¹В данном параграфе под дисперсией несмещенного прогноза \hat{y}_τ понимается величина $E (\hat{y}_\tau - y_\tau)^2$.

любых $\alpha \in R^m$. Поэтому условие несмещенности может быть записано:

$$X'c = x_\tau. \quad (2.22)$$

Найдем дисперсию прогноза (2.21). Имеем

$$E(\tilde{y}_\tau - y_\tau)^2 = E(\tilde{y}_\tau - x'_\tau \alpha - \varepsilon_\tau)^2 = E(\tilde{y}_\tau - x'_\tau \alpha)^2 + E\varepsilon_\tau^2$$

в силу некоррелируемости ε_τ и y_1, \dots, y_n . Далее

$$E(\tilde{y}_\tau - y_\tau)^2 = \sigma^2(\tilde{y}_\tau) + \sigma^2 = c' \text{cov}(y) c + \sigma^2 = \sigma^2(c'c + 1).$$

Итак, задача оптимального линейного несмещенного прогноза сводится к нахождению такого вектора $c \in R^n$, что $c'c \Rightarrow \min$ при условии несмещенности (2.22). Построим функцию Лагранжа

$$F(c, \lambda) = c'c - \lambda(X'c - x_\tau), \lambda \in R^m. \quad (2.23)$$

Тогда $\partial F/\partial c = 2c - X\lambda = 0$, откуда $\lambda = 2(X'X)^{-1}X'c$. Подставим это значение λ в выражение (2.23) и снова продифференцируем по c , получим

$$\partial F/\partial c = 2c - 4X(X'X)^{-1}X'c + 2X(X'X)^{-1}x_\tau = 0,$$

но с учетом (2.22)

$$2c - 4X(X'X)^{-1}x_\tau + 2X(X'X)^{-1}x_\tau = 0,$$

откуда окончательно $c = X(X'X)^{-1}x_\tau$, а это как раз соответствует c при использовании оценки МНК.

Таким образом, *оптимальный линейный несмещенный прогноз также приводит к оценке МНК.*

З а м е ч а н и я: 1. При построении прогноза естественно допускается, что y_τ удовлетворяет уравнению регрессии (2.18). Таким образом, считается, что структура модели в будущем не изменится.

2. Вместо «оценивания» реального значения y_τ можно оценивать его м. о., т. е. $x'_\tau \alpha$. Тогда за счет отсутствия случайной ошибки ε_τ дисперсия прогноза (2.19) будет на σ^2 меньше, т. е. будет равна:

$$\sigma_\tau^2 = \sigma^2 x'_\tau (X'X)^{-1} x_\tau. \quad (2.24)$$

3. Формулы (2.20) не годятся в качестве статистик, так как содержат неизвестный параметр σ^2 . Несмещенной оценкой дисперсии прогноза в форме (2.20) будет

$$s_\tau^2 = s^2(1 + x'_\tau (X'X)^{-1} x_\tau),$$

где s^2 — несмещенная оценка σ^2 ;

$$s^2 = \frac{1}{n-m} (y - \hat{y})' (y - \hat{y}) = \frac{1}{n-m} e' e.$$

Наиболее часто регрессионную модель используют в двух целях:

1) исследователя интересуют сами коэффициенты $\alpha_1, \alpha_2, \dots, \alpha_m$ — качественный анализ регрессии. При этом оптимальным является метод наименьших квадратов;

2) коэффициенты необходимы постольку, поскольку они необходимы для построения прогноза для y_τ . Как следует из теоремы 2.6, оптимальной оценкой здесь также будет оценка МНК.

Разумнее прогноз делать не точечный, а интервальный. Для построения доверительного интервала необходимо задаться распределением ошибок. Допустим, как и ранее, $\varepsilon \sim N(0, \sigma^2 \mathbf{I})$, отклонение $\varepsilon_\tau \sim N(0, \sigma^2)$ независимо от ε . Можно показать, что доверительный интервал

$$D_\tau = \{\hat{y}_\tau - t_\lambda s_\tau < y_\tau < \hat{y}_\tau + t_\lambda s_\tau\} \quad (2.25)$$

является несмещенным, равномерно наиболее точным с коэффициентом доверия $1 - \lambda$, т. е. $P(D_\tau) = 1 - \lambda$.

У п р а ж н е н и я 2.2

1. Какой вид принимает формула (2.20) при $m = 1, 2$?

2. Предположим, прогноз x_τ найден не точно, т. е. $\sigma^2(\hat{x}_{\tau i}) = \sigma_i^2 > 0$, $E\hat{x}_{\tau i} = x_{\tau i}$, причем $\hat{x}_{\tau i}$ не коррелируют между собой и не коррелируют с «прошлым» y_1, \dots, y_n и ε_τ . Найдите дисперсию прогноза $y_\tau = a'x_\tau$.

3. Допустим, имеется k экспертных несмещенных прогнозов $\hat{y}_{\tau 1}, \dots, \hat{y}_{\tau k}$ с одинаковой дисперсией v^2 . Считаем, что $\hat{y}_{\tau i}$ не коррелируют между собой и с y_1, \dots, y_n . Как экспертные прогнозы учесть при построении прогноза по регрессии (2.18)?

4. Обобщите формулу (2.20) на случай, когда $\text{cov}(\varepsilon) = \sigma^2 \Omega$, причем ε_τ не коррелирует с y_1, \dots, y_n .

5. Предыдущую задачу решите для случая, когда $E(\varepsilon_t \varepsilon_t) = r_t$ известно, $t = 1, \dots, n$.

2.3. Регрессия с ограничениями на параметры

В этом параграфе ослабим предположение А. Будем считать, что априорное множество Θ не совпадает со всем пространством R^m , а является лишь его частью. Рассмотрим случай, когда Θ представляет собой гиперплоскость в R^m некоторой размерности $k < m$. Это означает, что на истинный вектор параметров наложено k линейно-независимых априорных ограничений в виде уравнений с известными

коэффициентами. С такого рода ограничениями мы сталкивались в параграфе 1.10 при проверке линейной гипотезы (1.55). Итак, предположим, неизвестный вектор параметров α удовлетворяет соотношению

$$R\alpha = r, \quad (2.26)$$

где $R^{k \times m}$, $\text{rank } R = k$, r заданы. Остальные предположения Б — Е считаем выполненными.

Т е о р е м а 2.7. *Оценка, минимизирующая сумму квадратов отклонений при ограничениях (2.26), равная*

$$a_R = a + (X'X)^{-1}R'S(r - Ra), \quad (2.27)$$

где $S = [R(X'X)^{-1}R']^{-1}$, a — оценка МНК, является несмещенной и эффективной в классе несмещенных оценок, линейных по y^1 , удовлетворяющих (2.26); матрица ковариации (2.27) равна:

$$\text{cov}(a_R) = V[I - R'SR(X'X)^{-1}] \quad (2.28)$$

где $V = \text{cov}(a) = \sigma^2(X'X)^{-1}$.

Д о к а з а т е л ь с т в о. Для нахождения a_R воспользуемся МНК при ограничениях (2.26). Таким образом, минимизируем $Q_R(\alpha) = (y - X\alpha)'(y - X\alpha)$ при условии $R\alpha = r$. Введем множители Лагранжа и построим функцию

$$Q_R(a; \lambda) = (y - X\alpha)'(y - X\alpha) - \lambda'(R\alpha - r); \quad (2.29)$$

$$\partial Q_R / \partial \alpha = -2X'y + 2X'X\alpha - R'\lambda = 0.$$

Выразим отсюда α и подставим его в (2.26), откуда найдем $\lambda = 2[R(X'X)^{-1}R']^{-1}(r - Ra)$, что при подстановке в (2.29) ведет к оценке (2.27).

Несмещенность оценки (2.27) доказывается просто: в силу (2.26)

$$Ea_R = Ea + (X'X)^{-1}R'SE(r - Ra) = \\ = a + (X'X)^{-1}R'S(r - R\alpha) = a.$$

Доказательство линейной эффективности оценки предоставим читателю (см. также [58, с. 202—204]). Найдем $\text{cov}(a_R)$:

$$\text{cov}(a_R) = \text{cov}(a - (X'X)^{-1}R'SRa) = \\ = \text{cov}\{(I - (X'X)^{-1}R'SR)a\} = \text{cov}(Pa) = \\ = P \text{cov}(a) = P' = \sigma^2 P(X'X)^{-1}P'.$$

¹В данном случае оценку b называем линейной по y , если $b = Hy + c$, где H — матрица $m \times n$, c — вектор $m \times 1$.

Далее

$$\begin{aligned} P(X'X)^{-1}P' &= [I - (X'X)^{-1}R'SR](X'X)^{-1}[I - \\ &- R'SR(X'X)^{-1}] = (X'X)^{-1} - 2(X'X)^{-1}R'SR(X'X)^{-1} + \\ &+ (X'X)^{-1}R'SR(X'X)^{-1}R'SR(X'X)^{-1} = \\ &= (X'X)^{-1} - (X'X)^{-1}R'SR(X'X)^{-1} = \\ &= (X'X)^{-1}[I - R'SR(X'X)^{-1}], \end{aligned}$$

откуда следует (2.28).

Несмещенной оценкой для σ^2 является

$$s^2 = (y - Xa_R)'(y - Xa_R) / (n - m + k). \quad (2.30)$$

В условиях нормально распределенных отклонений можно построить доверительные интервалы для параметра α и проверять статистические гипотезы [58].

Вкратце рассмотрим регрессии с ограничениями на параметры в виде неравенств. Вместо (2.26) будем предполагать, что

$$R\alpha \leq r, \quad (2.31)$$

где условия на R и r те же, что и в (2.26)¹. Оценка МНК минимизирует сумму квадратов отклонений $Q = (y - X\alpha)'(y - X\alpha)$ при ограничениях (2.31). Нетрудно показать, что во-первых, эта оценка b единственна; во-вторых, либо b совпадает с обычной оценкой МНК a , что верно, если $Ra \leq r$, либо $(Ra)_i = r_i$ хотя бы для одного $i = 1, \dots, m$. Итак, в тривиальном случае $b = a$, в противном случае некоторые неравенства в (2.31) оценка b обращает в равенства. Последнее обстоятельство является нежелательным. Действительно, допустим, исследуется взаимосвязь производительности труда y и материального стимулирования x : $y_t = \alpha_1 x_t + \alpha_2 + \varepsilon_t$. Разумно предположить, что $\alpha_1 \geq 0$. Минимизируя Q при ограничении $\alpha_1 \geq 0$, мы столкнемся с двумя ситуациями: а) b совпадает с оценкой МНК; в этом случае ограничение $\alpha_1 \geq 0$ становится «лишним», б) $a_1 < 0$, тогда $b_1 = 0$, что малопримемлемо.

Спецификация (2.31) неудовлетворительна тем, что в случае неравенств с ненулевой вероятностью оценка b принимает крайние значения. Неслучайно поэтому в некоторых случаях оценка МНК a оказывается лучше b [179]. Исследование оценки b , ее свойств и вычисление можно

¹ $a \leq b$ тогда и только тогда, когда $a_i \leq b_i$ для всех i .

найти в [153]. Там же есть ссылки на другие работы по этому вопросу.

Можно показать, что оценка МНК при ограничениях (2.31) является смещенной.

У п р а ж н е н и я 2.3

1. Докажите, что ближайшей оценкой (в смысле расстояния), удовлетворяющей (2.26), к оценке МНК является $\mathbf{b} = \mathbf{a} + \mathbf{R}'(\mathbf{R}'\mathbf{R})^{-1}(\mathbf{r} - \mathbf{R}\mathbf{a})$. Докажите, что эта оценка несмещенная. Найдите ее матрицу ковариаций. Покажите, что ее матрица ковариаций больше, чем $\text{cov}(\mathbf{a}_R)$.

2. Докажите, что оценка \mathbf{a}_R , минимизирующая $Q(\alpha)$ и удовлетворяющая (2.26), единственна.

3. Пусть предположение E не выполняется. Докажите, что единственность \mathbf{a}_R обеспечивается условием $\text{rang} \begin{bmatrix} \mathbf{X} \\ \mathbf{R} \end{bmatrix} = m$.

4. Так же как и в задаче 1, будем искать ближайшую оценку. В качестве меры близости рассмотрим $(\mathbf{b} - \mathbf{a})'\mathbf{X}'\mathbf{X}(\mathbf{b} - \mathbf{a})$, где $\mathbf{R}\mathbf{b} = \mathbf{r}$ и \mathbf{a} — оценка МНК. Докажите, что тогда $\mathbf{b} = \mathbf{a}_R$. Используя это, дайте геометрическую интерпретацию \mathbf{a}_R . Напомним, что $\text{cov}(\mathbf{a}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

5. Докажите, что \mathbf{a}_R — линейно эффективная несмещенная оценка.

6. Докажите, что s^2 — несмещенная оценка σ^2 .

7. Покажите, что оценка МНК, минимизирующая Q при ограничениях (2.31), является смещенной (начните со случая $m = 1$).

2.4. Перебор и недобор факторов в регрессии

Как правило, исследователю неизвестна истинная модель регрессии, т. е. неизвестно, какие факторы входят в регрессию. Поэтому ошибка перебора факторов или их недобора является весьма вероятной. Исследуем, к чему приводят эти ошибки спецификации, а также выясним, какие из них ведут к более серьезным последствиям.

Перебор. Истинная модель есть

$$y_t = \alpha_1 x_{t1} + \dots + \alpha_m x_{tm} + \varepsilon_t, \quad t = 1, \dots, n, \quad (2.32)$$

или

$$\mathbf{y} = \mathbf{X}\alpha + \varepsilon, \quad \text{cov}(\varepsilon) = \sigma^2 \mathbf{I}_n,$$

относительно которой выполнены все предположения А — Е. Мы предполагаем, что

$$y_t = \delta_1 x_{t1} + \dots + \delta_m x_{tm} + \varphi_1 p_{t1} + \dots + \varphi_k p_{tk} + \varepsilon_t, \quad (2.33)$$

или

$$y = Z\beta + \varepsilon, Z = [XP], \beta' = (\delta', \varphi'),$$

где $P^{n \times k}$, $\varphi^{k \times 1}$, $\delta^{m \times 1}$, $Z^{n \times (m+k)}$, $\beta^{(m+k) \times 1}$. Оценкой МНК в (2.33) является

$$b = (Z'Z)^{-1} Z'y. \quad (2.34)$$

Докажем, что b несмещенно оценивает α . Другими словами, если обозначим $b' = (d'f')$, то $Ed = \alpha$, $Ef = 0$; d и f — вектор-столбцы размерности m и k соответственно. Обозначим $F = P'P - P'X(X'X)^{-1}X'P$; тогда по формуле Фробениуса (П.2)

$$\begin{aligned} Eb &= \begin{bmatrix} X'X & X'P \\ P'X & P'P \end{bmatrix}^{-1} \begin{bmatrix} X'X \\ P'X \end{bmatrix} \alpha = \\ &= \begin{bmatrix} (X'X)^{-1} + (X'X)^{-1}X'PF^{-1}P'X(X'X)^{-1} & -(X'X)^{-1}X'PF^{-1} \\ -F^{-1}P'X(X'X)^{-1} & F^{-1} \end{bmatrix} \times \\ &\quad \times \begin{bmatrix} X'X \\ P'X \end{bmatrix} \alpha = \\ &= \begin{bmatrix} I + (X'X)^{-1}X'PF^{-1}P'X - (X'X)^{-1}X'PF^{-1}P'X \\ -F^{-1}P'X + F^{-1}P'X \end{bmatrix} \alpha = \\ &= \begin{bmatrix} I_m \\ 0 \end{bmatrix} \alpha = \begin{bmatrix} \alpha \\ 0 \end{bmatrix}, \quad (2.35) \end{aligned}$$

что доказывает несмещенность (2.34). Математическое ожидание оценок «лишних» параметров $\varphi_1, \varphi_2, \dots, \varphi_k$ равно нулю, т. е. они также оцениваются несмещенно.

Можно доказать, что оценка $(y - Zb)'(y - Zb)/(n - m - k)$ является несмещенной оценкой σ^2 .

Докажем, что оценка МНК в случае перебора является также состоятельной. Для этого предположим, что матрица Z сильно регулярна. Таким образом, накладываем ограничение на матрицу дополнительных переменных P : в пределе лишние переменные p_1, \dots, p_k линейно независимы с истинными переменными x_1, \dots, x_m . Итак, пусть

$$\frac{Z'Z}{n} = \begin{bmatrix} \frac{X'X}{n} & \frac{X'P}{n} \\ \frac{P'X}{n} & \frac{P'P}{n} \end{bmatrix} \rightarrow \begin{bmatrix} A_{11} & A_{12} \\ A'_{12} & A_{22} \end{bmatrix} = A, |A| \neq 0$$

при $n \rightarrow \infty$. Найдем сначала предельную ковариационную матрицу «лишних» параметров Φ . Учитывая формулу (2.35), получим

$$\frac{F}{n} = \frac{P'P}{n} - \frac{P'X}{n} \left(\frac{X'X}{n} \right)^{-1} \frac{X'P}{n} \rightarrow A_{22} - A'_{12} A^{-1}_{11} A_{12}, \quad n \rightarrow \infty.$$

Корректность этого выражения следует из того, что $|A_{11}| \neq 0$. Далее, предельная матрица для F/n невырождена в силу разложения определителя $|A|$ по формуле (П.3). Таким образом, $\text{cov}(\mathbf{f}) = \sigma^2 F^{-1} \rightarrow 0$.

Теперь распишем ковариационную матрицу для параметров \mathbf{d} :

$$\begin{aligned} & \left(\frac{X'X}{n} \right)^{-1} + \left(\frac{X'X}{n} \right)^{-1} \frac{X'P}{n} \left(\frac{F}{n} \right)^{-1} \frac{P'X}{n} \left(\frac{X'X}{n} \right)^{-1} \rightarrow \\ & \rightarrow A_{11} + A_{11} A_{12} (A_{22} - A'_{12} A^{-1}_{11} A_{12}) A'_{12} A_{11}, \end{aligned}$$

это также невырожденная матрица, т. е. $\text{cov} \mathbf{d} \rightarrow 0, n \rightarrow \infty$.

Итак, вектор \mathbf{b} в среднем квадратичном сходится к истинному значению $(\delta, 0)'$.

Оценка, получаемая в регрессиях с «лишними» независимыми переменными, как было показано, обладает по-прежнему рядом оптимальных свойств. Однако *точность при переборе теряется*. Покажем, почему это происходит. Матрица ковариаций оценки МНК для истинной модели равна $\sigma^2 (X'X)^{-1}$, а матрица ковариаций оценки \mathbf{d} для модели (2.33) с использованием (2.35) равна:

$$\begin{aligned} \text{cov}(\mathbf{d}) &= \sigma^2 [(X'X)^{-1} + (X'X)^{-1} X' P F^{-1} P' X (X'X)^{-1}] = \\ &= \text{cov}(\mathbf{a}) + \sigma^2 (X'X)^{-1} X' P F^{-1} P' X (X'X)^{-1}. \quad (2.36) \end{aligned}$$

Матрица F неотрицательно определена, так как $F = \mathbb{I} = P' (I_n - X (X'X)^{-1} X') P$, где $I_n - X (X'X)^{-1} X'$ — симметричная идемпотентная, а значит, и неотрицательно определенная матрица. Неотрицательная определенность F влечет и неотрицательную неопределенность второго слагаемого в (2.36), т. е. $\text{cov}(\mathbf{d}) \geq \text{cov}(\mathbf{a})$.

Нетрудно заметить, что если «лишнее» множество независимых переменных ортогонально истинному набору переменных, т. е. $X'P = 0$, то $\text{cov}(\mathbf{d}) = \text{cov}(\mathbf{a})$. Отсюда можно сделать вывод: если круг основных независимых переменных очерчен, то дополнительные переменные (которые могут оказаться лишними) надо стараться вводить в уравнение (2.32) так, чтобы они не «коррелировали» сильно с основным множеством переменных. Тогда потери точности

при переборе будут незначительны. Наоборот, если лишние переменные сильно сопряжены с исходным множеством переменных x_1, x_2, \dots, x_m , то потери точности будут велики (матрица F^{-1} , а значит и $\text{cov}(\hat{d})$, будет принимать большие значения).

Недобор. В этом случае истинным уравнением является (2.32), а мы оцениваем регрессию

$$y_t = \gamma_1 \omega_{t1} + \dots + \gamma_k \omega_{tk} + \xi_t, \quad (2.37)$$

или

$$y = W\gamma + \xi, \quad X = [WV];$$

другими словами, $\omega_1 = x_1, \dots, \omega_k = x_k$ входят в регрессию (2.37), а остальные $m - k$ переменных $v_1 = x_{k+1}, \dots, v_{m-k} = x_m$ в регрессии отсутствуют. Докажем, что тогда используемая оценка

$$g = (W'W)^{-1} W'y \quad (2.38)$$

в общем случае является *смещенной*. Действительно,

$$\begin{aligned} E g &= (W'W)^{-1} W'E y = (W'W)^{-1} W' X \alpha = \\ &= (W'W)^{-1} W' [WV] \begin{bmatrix} \alpha^1 \\ \alpha^2 \end{bmatrix}, \end{aligned}$$

где α^1 и α^2 — подвекторы вектора α размерности k и $m - k$ соответственно. Перемножая члены в последнем равенстве, получим

$$E g = [I_k \quad (W'W)^{-1} W'V] \begin{bmatrix} \alpha^1 \\ \alpha^2 \end{bmatrix} = \alpha^1 + (W'W)^{-1} W'V \alpha^2. \quad (2.39)$$

Поскольку второе слагаемое в уравнении (2.39) отлично от нуля, делаем вывод, что оценка (2.38), т. е. оценка МНК для регрессии с недобором (2.37), является смещенной оценкой подвектора α^1 . Оценка g будет несмещенной, когда матрицы переменных W и V ортогональны, т. е. $W'V = 0$.

Случай ортогональности является идеальным для двух рассмотренных возможностей: перебора и недобора. Однако на практике независимые переменные сильно сопряжены. Если же в планируемом эксперименте мы полностью контролируем значения независимых переменных, то их желательно брать близко к ортогональным.

Объясним суть смещения (2.39). Прежде всего заметим, что j -й столбец матрицы $(W'W)^{-1}W'V$ формально является оценкой МНК в регрессии $v_i = W\tau + \delta$. Рассмотрим для примера случай $j = 1$. Тогда смещение g_1 равно $\sum t_i \alpha^2_i$, где t — первая вектор-строка матрицы $(W'W)^{-1}W'V$, а t_i — оценка МНК неизвестной τ_1 в предыдущей регрессии. Ничего нет удивительного в том, что оценка МНК в регрессии с недобором является смещенной: неучтенная часть уравнения регрессии равномерно распределяется в оценке g с помощью «довесков». Суть этих «довесков» — регрессия неучтенных факторов на учтенные (2.37).

Теперь сравним, какая из оценок — g или b — является более приемлемой с точки зрения точности оценивания истинного параметра α . Поскольку d несмещенно оценивает α , то $E(d - \alpha)(d - \alpha)'$ дается выражением (2.36). В качестве оценки α в регрессии (2.37) фактически выбирается $\bar{g} = \begin{bmatrix} g \\ 0 \end{bmatrix}$, поэтому

$$E(\bar{g} - \alpha)(\bar{g} - \alpha)' = \begin{bmatrix} E(g - \alpha^1)(g - \alpha^1)' & -(Eg - \alpha^1)\alpha^{2'} \\ -\alpha^2(Eg - \alpha^1)' & \alpha^2\alpha^{2'} \end{bmatrix}.$$

Но

$$g = (W'W)^{-1}W'X\alpha + (W'W)^{-1}W'\varepsilon = Eg + (W'W)^{-1}W'\varepsilon,$$

поэтому

$$E(g - \alpha^1)(g - \alpha^1)' = E[(g - Eg) + (Eg - \alpha^1)] \times \\ \times [(g - Eg)' + (Eg - \alpha^1)'] = \sigma^2(W'W)^{-1} + \\ + (W'W)^{-1}W'V\alpha^2\alpha^{2'}V'W(W'W)^{-1}.$$

Окончательно

$$E(\bar{g} - \alpha)(\bar{g} - \alpha)' = \begin{bmatrix} \sigma^2(W'W)^{-1} + (W'W)^{-1}W'V\alpha^2\alpha^{2'}V'W(W'W)^{-1} & \\ & -(W'W)^{-1}W'V\alpha^2\alpha^{2'} \\ -\alpha^2\alpha^{2'}V'W(W'W)^{-1} & \alpha^2\alpha^{2'} \end{bmatrix}. \quad (2.40)$$

Сравнивая эту матрицу с матрицей (2.36), делаем вывод: в общем случае нельзя утверждать, что (2.36) меньше или больше (2.40). Например, если $\alpha^2 = 0$, то легко видеть, что разница между (2.36) и (2.40) будет положительно опреде-

лена. Значит, если α^2 достаточно близко к 0, то оценка \mathbf{g} предпочтительнее. Наоборот, при $\alpha^2 \rightarrow \infty$ матрица (2.40) неограниченно возрастает и оценка \mathbf{b} лучше. Таким образом, единственное, что можно утверждать, это то, что в некоторой окрестности $\alpha_2 = 0$ (2.38) лучше (2.34), вне этой окрестности наоборот (2.34) лучше (2.38). Здесь наблюдается такая же ситуация, как в параграфе 1.4. Для любой несмещенной оценки можно найти тривиальную смещенную оценку, которая в некоторой окрестности неизвестного параметра будет лучше несмещенной.

Докажем, что оценка МНК в случае недобора является несостоятельной. Доказательство для простоты будем проводить в условиях сильной регулярности матрицы \mathbf{X} . Итак, предположим .

$$\frac{\mathbf{X}'\mathbf{X}}{n} = \begin{bmatrix} \frac{\mathbf{W}'\mathbf{W}}{n} & \frac{\mathbf{W}'\mathbf{V}}{n} \\ \frac{\mathbf{V}'\mathbf{W}}{n} & \frac{\mathbf{V}'\mathbf{V}}{n} \end{bmatrix} \rightarrow \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}'_{12} & \mathbf{B}_{22} \end{bmatrix} = \mathbf{B}, \quad |\mathbf{B}| \neq 0.$$

Тогда

$$\begin{aligned} (\mathbf{W}'\mathbf{W})^{-1} &= \frac{1}{n} \left(\frac{\mathbf{W}'\mathbf{W}}{n} \right)^{-1} \rightarrow 0, \quad n \rightarrow \infty, \\ (\mathbf{W}'\mathbf{W})^{-1} \mathbf{W}'\mathbf{V} \alpha^2 \alpha^{2'} \mathbf{V}'\mathbf{W} (\mathbf{W}'\mathbf{W})^{-1} &= \\ = \left(\frac{\mathbf{W}'\mathbf{W}}{n} \right)^{-1} \frac{\mathbf{W}'\mathbf{V}}{n} \alpha^2 \alpha^{2'} \frac{\mathbf{V}'\mathbf{W}}{n} \left(\frac{\mathbf{W}'\mathbf{W}}{n} \right)^{-1} &\rightarrow \\ \rightarrow \mathbf{B}_{11}^{-1} \mathbf{B}_{12} \alpha^2 \alpha^{2'} \mathbf{B}'_{12} \mathbf{B}_{11}^{-1} &\neq 0. \end{aligned}$$

Таким образом,

$$\begin{aligned} & \mathbf{E}(\bar{\mathbf{g}} - \alpha)(\bar{\mathbf{g}} - \alpha)' \rightarrow \\ & \rightarrow \begin{bmatrix} \mathbf{B}_{11}^{-1} \mathbf{B}_{12} \alpha^2 \alpha^{2'} \mathbf{B}'_{12} \mathbf{B}_{11}^{-1} & -\mathbf{B}_{11}^{-1} \mathbf{B}_{12} \alpha^2 \alpha^{2'} \\ -\alpha^2 \alpha^{2'} \mathbf{B}'_{12} & \mathbf{I}_1 \end{bmatrix} \neq 0 \end{aligned}$$

при $n \rightarrow \infty$ и оценка $\bar{\mathbf{g}}$ не состоятельна.

Подведем итоги: в случае перебора оценка МНК теряет в эффективности, зато остается несмещенной и состоятельной; в случае же недобора оценка МНК является смещенной и несостоятельной. По изложенным выше причинам недобор считаем более существенной ошибкой спецификации, чем перебор, так как он ведет к более тяжелым последствиям.

Проблема оптимального выбора множества независимых из данного набора переменных исследуется в [114, 93, 157, 182, 28].

В качестве показателя правильности выбора множества переменных в [61] предлагается брать статистику s^2 . Результат основывается на следующем: пусть, как и ранее, (2.32) будет истинной моделью, тогда как мы предполагаем, что модель имеет вид:

$$y = Q\tau + \xi, \quad (2.41)$$

где $Q^{n \times l}$, $\tau^{l \times 1}$, $E\xi = 0$.

Строим оценки $s_{\tau}^2 = \frac{1}{n-l} (y - Qt)' (y - Qt)$ и $s_{\alpha}^2 = \frac{1}{n-m} (y - Xa)' (y - Xa)$, где t и a — оценки МНК для регрессий (2.41) и (2.32) соответственно. Показано, что $Es_{\tau}^2 \geq Es_{\alpha}^2$. Поэтому для неправильно специфицированных моделей в среднем оценка s_{τ}^2 будет больше, чем для правильно специфицированных (см. также [52]). Изложенный результат может быть применен к частным случаям неправильно специфицированных моделей: перебору и недобору. И в том, и в другом случае оценка s_{τ}^2 будет (в среднем) больше оценки s_{α}^2 .

2.5. Псевдонезависимые регрессии

До сих пор мы имели дело с оцениванием одной регрессии. Часто требуется оценить одновременно несколько регрессий, некоторое семейство их. Совокупность таких регрессий будем называть *псевдонезависимыми регрессиями* (seemingly unrelated regressions). Этот термин был впервые введен А. Зеллнером в [197], откуда и начинается систематическое изучение псевдонезависимых регрессий. Псевдонезависимыми считаем регрессии потому, что, во-первых, независимые переменные могут одновременно входить сразу в несколько регрессий; во-вторых, отклонения разных регрессий могут коррелировать. В то же время псевдонезависимые регрессии отличаются от эконометрических моделей (синхронных регрессий), поскольку зависимые переменные в первом случае присутствуют только в левых частях уравнений.

Псевдонезависимые регрессии допускают применение МНК к каждому уравнению регрессии в отдельности. Однако в данном случае такой метод не будет эффективным в классе несмещенных линейных оценок. Дело в том, что отклонения в псевдонезависимых регрессиях в общем случае

коррелируют для разных уравнений, поэтому оценка Эйткена является более эффективной, чем МНК. Ранее (см. параграф 2.1) метод Эйткена был практически неприменим из-за отсутствия весовой матрицы или ее оценки. В случае псевдонезависимых регрессий удастся построить состоятельную оценку этой матрицы, использование которой приводит к стохастическому эквиваленту оценки Эйткена — оценке Зеллнера. Естественным обобщением оценки Зеллнера является итеративная оценка Зеллнера s -го порядка, которая по своей природе близка к оценке ММП в предположении нормальности отклонений.

Не всегда исследование свойств оценок псевдонезависимых регрессий проводилось на достаточно строгом уровне. Многие предположения опускались, использовалось большое число недоказанных результатов (см., например, [197]). По возможности мы старались доказательство делать наиболее полными, а утверждения формулировать строго.

Итак, пусть имеется k линейных регрессий

$$y_i = X_i \alpha_i + \varepsilon_i, \quad i = 1, \dots, k, \quad (2.42)$$

где $y_i^{n \times 1}$, $X_i^{n \times m_i}$, $\alpha_i^{m_i \times 1}$, $\varepsilon_i^{n \times 1}$ — соответственно вектор зависимой переменной, матрица независимых переменных, вектор неизвестных параметров, подлежащий оцениванию, вектор случайных отклонений. Систему (2.42) будем называть системой псевдонезависимых регрессий. Относительно X_i , α_i , ε_i всюду в дальнейшем будем предполагать, что: а) X_i — детерминированные матрицы, $\text{rang } X_i = m_i$; б) $\alpha_i \in \Theta_i$ — априорное множество параметров совпадает с R^{m_i} ; в) $E\varepsilon_i = 0$, $E\varepsilon_{ti}\varepsilon_{tj} = \sigma_0^2 \omega_{ij}$, $E\varepsilon_{ti}\varepsilon_{\tau j} = 0$, $t \neq \tau$. Предположение в) влечет стохастическую зависимость регрессий системы (2.42). Если через ε^t обозначить вектор-столбец отклонений, соответствующий номеру наблюдения или моменту времени t и составленный из отклонений системы (2.42), то $\text{cov}(\varepsilon^t) = \sigma_0^2 \Omega^1$. Естественно, мы считаем $|\Omega| \neq 0$. Таким образом, необходимо оценить дополнительно $k(k+1)/2$ неизвестных параметров.

Система (2.42) может быть сведена к одной регрессии. Для этого построим новые векторы и матрицы размерности

¹В целях идентифицируемости будем считать, что по заданной матрице $T^{k \times k}$ можно единственным образом найти σ_0^2 и Ω так, чтобы $\sigma_0^2 \Omega = T$.

$\times 1, nk \times m, kn \times 1$ и $m \times 1$ соответственно, где $m = \sum_{i=1}^k m_i$:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}, X = \begin{bmatrix} X_1 & & & 0 \\ & X_2 & & \\ & & \ddots & \\ 0 & & & X_k \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix}, \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix}.$$

Система (2.42) переписывается в виде

$$y = X\alpha + \varepsilon. \quad (2.43)$$

Легко видеть, что $\text{rank } X = m$, $\alpha \in \Theta = R^m$ и $E\varepsilon = 0$.

Найдем матрицу ковариаций вектора ε . По определению имеем

$$\begin{aligned} \text{cov}(\varepsilon) &= E\varepsilon\varepsilon' = E \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix} [\varepsilon_1' \ \varepsilon_2' \ \dots \ \varepsilon_k'] = \\ &= E \begin{bmatrix} \varepsilon_1 \varepsilon_1' & \varepsilon_1 \varepsilon_2' & \dots & \varepsilon_1 \varepsilon_k' \\ \vdots & \vdots & & \vdots \\ \varepsilon_k \varepsilon_1' & \varepsilon_k \varepsilon_2' & \dots & \varepsilon_k \varepsilon_k' \end{bmatrix} = \\ &= \sigma_0^2 \begin{bmatrix} \omega_{11} I_n & \omega_{12} I_n & \dots & \omega_{1k} I_n \\ \vdots & \vdots & & \vdots \\ \omega_{k1} I_n & \omega_{k2} I_n & \dots & \omega_{kk} I_n \end{bmatrix} = \sigma_0^2 (\Omega \otimes I_n), \end{aligned}$$

где \otimes — знак кронекерова произведения матриц (см., например, [9]); I_n — единичная матрица $n \times n$.

Если матрица Ω известна, то, применяя обобщенный метод наименьших квадратов, приходим к оценке Эйткена, которая является несмещенной и линейно эффективной (параграф 2.1)¹:

$$\begin{aligned} \hat{b} &= [X'(\Omega \otimes I)^{-1} X]^{-1} X'(\Omega \otimes I)^{-1} y = \\ &= [X'(\Omega^{-1} \otimes I) X]^{-1} X'(\Omega^{-1} \otimes I) y. \end{aligned} \quad (2.44)$$

¹Воспользуемся следующим свойством кронекерова произведения: $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ [9].

Оценка МНК для составной регрессии (2.43) также является несмещенной, но уже не будет эффективной:

$$\mathbf{a} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (2.45)$$

Легко проверяется, что

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{X}'_1\mathbf{X}_1 & & & \\ & \ddots & & \\ & & \mathbf{0} & \\ & & & \ddots & \\ & \mathbf{0} & & & & \\ & & & & & \mathbf{X}'_k\mathbf{X}_k \end{bmatrix},$$

и поэтому

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} (\mathbf{X}'_1\mathbf{X}_1)^{-1} & & & \\ & \ddots & & \\ & & \mathbf{0} & \\ & & & \ddots & \\ & \mathbf{0} & & & & \\ & & & & & (\mathbf{X}'_k\mathbf{X}_k)^{-1} \end{bmatrix}.$$

Тогда (2.45) переписывается следующим образом:

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_1\mathbf{X}_1 & & & \\ & \ddots & & \\ & & \mathbf{0} & \\ & & & \ddots & \\ & \mathbf{0} & & & & \\ & & & & & \mathbf{X}'_k\mathbf{X}_k \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'_1\mathbf{y}_1 \\ \mathbf{X}'_2\mathbf{y}_2 \\ \vdots \\ \mathbf{X}'_k\mathbf{y}_k \end{bmatrix} = \\ = \begin{bmatrix} (\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{y}_1 \\ (\mathbf{X}'_2\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{y}_2 \\ \vdots \\ (\mathbf{X}'_k\mathbf{X}_k)^{-1}\mathbf{X}'_k\mathbf{y}_k \end{bmatrix}.$$

Другими словами, оценка (2.45) совпадает с оценкой МНК, примененного к каждому уравнению системы (2.42) в отдельности.

При некоторых условиях оценка Эйткена (2.44) совпадает с оценкой МНК (2.45), как показано в [96]. Ясно, что эти оценки будут совпадать, если $\mathbf{\Omega} = \mathbf{I}_k$, т. е. когда корреляции между разными уравнениями регрессий отсутствуют. В этом случае, применяя МНК к каждому уравнению (2.42), получаем несмещенные линейно эффективные оценки. Однако существует другая нетривиальная ситуация, когда $\mathbf{a} = \mathbf{b}$. Покажем, что если все матрицы \mathbf{X}_i совпа-

дают, то $\mathbf{a} = \mathbf{b}$. Итак, пусть $\mathbf{X}_i = \mathbf{Z}^{n \times m_i}$, $i = 1, \dots, k$. Преобразуем сначала оценку МНК¹:

$$\begin{aligned} \mathbf{a} &= [(\mathbf{I} \otimes \mathbf{Z})' (\mathbf{I} \otimes \mathbf{Z})]^{-1} (\mathbf{I} \otimes \mathbf{Z})' \mathbf{y} = [\mathbf{I} \otimes (\mathbf{Z}' \mathbf{Z})]^{-1} (\mathbf{I} \otimes \mathbf{Z}') \mathbf{y} = \\ &= [\mathbf{I} \otimes (\mathbf{Z}' \mathbf{Z})^{-1}] (\mathbf{I} \otimes \mathbf{Z}') \mathbf{y} = [\mathbf{I} \otimes (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}'] \mathbf{y}. \end{aligned}$$

Распишем оценку Эйткена:

$$\begin{aligned} \mathbf{b} &= [\mathbf{X}' (\mathbf{\Omega}^{-1} \otimes \mathbf{I}) \mathbf{X}]^{-1} \mathbf{X}' (\mathbf{\Omega}^{-1} \otimes \mathbf{I}) \mathbf{y} = \\ &= [(\mathbf{I} \otimes \mathbf{Z})' (\mathbf{\Omega}^{-1} \otimes \mathbf{I}) (\mathbf{I} \otimes \mathbf{Z})]^{-1} (\mathbf{I} \otimes \mathbf{Z})' (\mathbf{\Omega}^{-1} \otimes \mathbf{I}) \mathbf{y} = \\ &= [(\mathbf{I} \otimes \mathbf{Z}') (\mathbf{\Omega}^{-1} \otimes \mathbf{I}) (\mathbf{I} \otimes \mathbf{Z})]^{-1} (\mathbf{I} \otimes \mathbf{Z}') (\mathbf{\Omega}^{-1} \otimes \mathbf{I}) \mathbf{y} = \\ &= [\mathbf{\Omega}^{-1} \otimes \mathbf{Z}'] (\mathbf{I} \otimes \mathbf{Z})^{-1} (\mathbf{\Omega}^{-1} \otimes \mathbf{Z}' \mathbf{y}) = \\ &= [\mathbf{\Omega}^{-1} \otimes (\mathbf{Z}' \mathbf{Z})^{-1}] (\mathbf{\Omega}^{-1} \otimes \mathbf{Z}' \mathbf{y}) = [\mathbf{\Omega} \otimes (\mathbf{Z}' \mathbf{Z})^{-1}] (\mathbf{\Omega}^{-1} \otimes \mathbf{Z}' \mathbf{y}) = \\ &= [\mathbf{I} \otimes (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}'] \mathbf{y}, \end{aligned}$$

что совпадает с оценкой \mathbf{a} . Таким образом, если матрицы \mathbf{X}_i одинаковы во всех регрессиях, то оценка МНК эффективна.

Как правило, матрица $\mathbf{\Omega}$ неизвестна. Однако в случае псевдонезависимых регрессий можно построить весьма удовлетворительную оценку для $\mathbf{\Omega}$. Пусть $\mathbf{a}_i = (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{y}_i$ — оценка МНК параметра α_i . Обозначим

$$\mathbf{e}_i = \mathbf{y}_i - \mathbf{X}_i \mathbf{a}_i = (\mathbf{I}_n - \mathbf{X}_i (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i') \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, k.$$

Положим

$$\begin{aligned} \hat{\sigma}_0^2 \omega_{ij} &= \frac{1}{n} \mathbf{e}_i' \mathbf{e}_j = \frac{1}{n} \boldsymbol{\varepsilon}_i' (\mathbf{I}_n - \mathbf{X}_i (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i')' \times \\ &\times (\mathbf{I}_n - \mathbf{X}_j (\mathbf{X}_j' \mathbf{X}_j)^{-1} \mathbf{X}_j') \boldsymbol{\varepsilon}_j. \end{aligned} \quad (2.46)$$

Можно доказать, что при весьма слабых предположениях оценка матрицы $\hat{\sigma}_0^2 \mathbf{\Omega}$ на основе (2.46) является состоятельной. При этом мы не будем накладывать на матрицу \mathbf{X} каких-либо ограничений при $n \rightarrow \infty$.

Теорема 2.8. Допустим, отклонения ε_{ti} для разных t независимы и имеют конечный четвертый момент ν_4 . Тогда оценка (2.46) является состоятельной оценкой $\hat{\sigma}_0^2 \mathbf{\Omega}$ в смысле сходимости в среднем квадратичном.

Доказательство теоремы дано в [23].

¹Воспользуемся следующими свойствами кронекерова произведения: $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$ и $(\mathbf{A} \otimes \mathbf{C})(\mathbf{B} \otimes \mathbf{D}) = \mathbf{AB} \otimes \mathbf{CD}$ (см. [9]).

Состоятельную оценку Ω можно подставить в формулу (2.44). Соответствующую оценку будем называть оценкой Zellнера:

$$z = [X' (\hat{\Omega} \otimes I)^{-1} X]^{-1} X' (\hat{\Omega} \otimes I)^{-1} y. \quad (2.47)$$

Получив оценку Zellнера, можно по аналогии построить целый класс оценок неизвестного вектора параметра α , которые назовем итеративными оценками Zellнера s -го порядка; эти оценки строятся рекурсивно: пусть z^{s-1} — итеративная оценка Zellнера $(s-1)$ -го порядка; построим на ее основе оценку $\sigma_0^2 \Omega$:

$$\hat{\sigma}_0^2 \omega_{ij} = \frac{1}{n} (y_i - X_i z_i^{s-1})' (y_j - X_j z_j^{s-1}), \quad (2.48)$$

тогда z^s определим как

$$z^s = [X' (\hat{\Omega}^{s-1} \otimes I)^{-1} X]^{-1} X' (\hat{\Omega}^{s-1} \otimes I)^{-1} y, \quad (2.49)$$

где $\hat{\Omega}^{s-1}$ рассчитывается по формуле (2.48); z^0 примем за оценку МНК; z^1 — оценка Zellнера (2.47). Теоретически процесс можно продолжать до бесконечности. Если предел z^s , $s \rightarrow \infty$ существует, назовем его итеративной оценкой Zellнера

$$z^\infty = \lim_{s \rightarrow \infty} z^s. \quad (2.50)$$

В работе [23] доказано, что при выполнении условия Эйкера оценка Zellнера состоятельна и что оценки (2.49) будут состоятельны, если матрицы X_n сильно регулярны. При этом условии может быть доказана асимптотическая нормальность оценки Эйткена, итеративной оценки Zellнера z^s и оценки z^∞ [64, 197].

Если закон распределения отклонений с точностью до матрицы ковариаций известен, то возможно применение метода максимального правдоподобия. Предположим, что ε — нормально распределенные случайные векторы, т. е. $\varepsilon^t \sim N(0, \sigma_0^2 \Omega)$. Обозначим $S = (\sigma_0^2 \Omega)^{-1}$, тогда, как нетрудно проверить, функция плотности распределения вектора y запишется:

$$f(y; \alpha, S) = (2\pi)^{-\frac{nk}{2}} |S \otimes I_n|^{1/2} \exp \left[-\frac{1}{2} (y - X\alpha)' \times \right. \\ \left. \times (S \otimes I) (y - X\alpha) \right].$$

Оценка ММП соответствует максимуму $f(y; \alpha, S)$, т. е. минимуму функции

$$p(y; \alpha, S) = (y - X\alpha)' (S \otimes I) (y - X\alpha) - n \ln |S| = \\ = \varepsilon' (S \otimes I) \varepsilon - n \ln |S|. \quad (2.51)$$

Объясним, что мы понимаем под аргументом функции p . Матрица S представляется набором $k(k+1)/2$ чисел, т. е. вектором из $R^{k(k+1)/2}$; обозначим через $E \subset R^{k(k+1)/2}$ множество тех векторов, которые соответствуют положительно определенным матрицам. Общий вектор-аргумент β функции p принадлежит $F = R^m \times E \subset R^{m+k(k+1)/2}$. Можно показать, что F — открытое множество.

Минимизация функции (2.51) почти наверное корректна. А именно функция (2.51) ограничена снизу почти для всех ε (см. задачу 6 упражнения 2.5). Однако можно показать, что эта функция не является выпуклой вниз, поэтому минимизировать ее необходимо с определенной осторожностью (см. приложение П.3).

Необходимым условием обращения (2.51) в минимум в некоторой точке является равенство нулю в этой точке производных p по α и S . Легко проверить, что $\frac{dp}{d\alpha} = 2X'(S \otimes I)X\alpha - 2X'(S \otimes I)y$ (приложение П.2). Далее очевидно, что

$$\frac{\partial(S \otimes I)}{\partial S_{ij}} = \begin{bmatrix} 0 & & 0 \\ & I_n & \\ & \vdots & \\ 0 & & j & 0 \end{bmatrix} i, i, j = 1, \dots, k, \quad (2.52)$$

где I_n — единичная матрица, расположенная в (i, j) -блоке. Имеем

$$\frac{\partial p}{\partial S_{ij}} = (y - X\alpha)' \frac{\partial(S \otimes I)}{\partial S_{ij}} (y - X\alpha) - n \frac{\partial \ln |S|}{\partial S_{ij}}.$$

Но по формуле (П.9) $\frac{d \ln |S|}{d S_{ij}} = (S^{-1})_{ij}$, поэтому с учетом (2.52)

$$\begin{aligned} \partial p / \partial S_{ij} &= (y_i - X_i \alpha_i)' (y_j - X_j \alpha_j) - n S_{ij}^{-1} = \\ &= (y_i - X_i \alpha_i)' (y_j - X_j \alpha_j) - n \sigma_0^2 \Omega_{ij}. \end{aligned} \quad (2.53)$$

Допустим, значение $S = \Omega^{-1} / \sigma_0^2$ известно, тогда, приравнявая $dp/d\alpha$ к нулю, получаем оценку

$$\hat{\alpha} = [X'(S \otimes I)X]^{-1} X'(S \otimes I)y. \quad (2.54)$$

Обратно, если α известно, то, приравнявая (2.53) к нулю, найдем

$$\hat{\sigma}_0^2 \Omega_{ij} = \frac{1}{n} (y_i - X_i \alpha_i)' (y_j - X_j \alpha_j). \quad (2.55)$$

На нулевом шаге итерационной процедуры в качестве приближения можно взять оценку МНК, по ней построить по формуле (2.55) оценку $\hat{\sigma}_0^2 \Omega$, затем снова найти оценку (2.55) и т. д. Процедуры (2.54) и (2.55) полностью совпадают с итеративной оценкой Zellнера s -го порядка, а предел, если он существует, равен z^∞ .

Естественно встает вопрос о сходимости итеративной процедуры (2.54) и (2.55) к оценке ММП. Если бы функция (2.51) была выпуклой вниз, то сходимость установить было бы нетрудно. Однако это не так. Поэтому вместо сходимости z^s к глобальному минимуму функции (2.51) можно говорить лишь о сходимости точек, градиент в которых равен нулю. Доказательство того, что если z^∞ существует, то это значение удовлетворяет уравнению $dp/d\beta = 0$, дано в [168].

Доказано [172], что найдется такая окрестность истинного вектора параметров α , что почти для всех y существует $n = n(y)$, начиная с которого последовательность z^s сходится к оценке ММП, если начальное приближение лежит в выбранной окрестности α .

До сих пор мы рассматривали асимптотические свойства оценок, наибольший интерес из которых представляет оценка Zellнера z . Остановимся на свойствах оценки z при конечных объемах выборки. Единственное, что доказано в общем случае, это несмещенность оценки Zellнера при условии, что ε имеют симметричное распределение (см. [142]). Этот факт доказывается весьма просто: оценку (2.47) можно переписать как

$$\begin{aligned} z &= [X' (\hat{\Omega} \otimes I)^{-1} X]^{-1} X' (\hat{\Omega} \otimes I)^{-1} X \alpha + \\ &+ [X' (\hat{\Omega} \otimes I)^{-1} X]^{-1} X' (\hat{\Omega} \otimes I)^{-1} \varepsilon = \\ &= \alpha + [X' (\hat{\Omega} \otimes I)^{-1} X]^{-1} X' (\hat{\Omega} \otimes I)^{-1} \varepsilon. \end{aligned} \quad (2.56)$$

В силу симметричности распределения ε второе слагаемое в (2.56) также симметрично распределено, а значит, $Ez = \alpha$. Осталось показать, что

$$E \{ [X' (\hat{\Omega} \otimes I)^{-1} X]^{-1} X' (\hat{\Omega} \otimes I)^{-1} \} < \infty.$$

Обозначим случайную матрицу, стоящую под знаком математического ожидания, через A . Тогда $AX = I$ и $(EA)X = I$. Поэтому все элементы матрицы EA конечны, так как в противном случае не было бы выполнено тождество $(EA)X = I$.

Других свойств z , доказанных в общем случае, найдено не было.

В работах [175, 164, 198] исследуется система псевдонезависимых регрессий для случая $k = 2$. В [175] разбирается частный случай, когда X_1 является подматрицей X_2 . Более того, берется так называемая «оценка $\sigma_0^2 \Omega$ без ограничений», т. е. y_i регрессирует на множество всех x системы (2.42). Оценка (2.46) является оценкой $\sigma_0^2 \Omega$ с учетом ограничений. Первую оценку Ω обозначим $\tilde{\Omega}$, а соответствующую оценку Зеллнера — через \tilde{z} . Далее предполагается нормальность отклонений. При сделанных предположениях доказано, что оценка \tilde{z}_2 (т. е. оценка Зеллнера для второго уравнения) совпадает с оценкой МНК a_2 , в то же время $\tilde{z}_1 \neq a_1$, т. е. \tilde{z}_1 является линейно эффективной оценкой. Основным результатом, полученный в [175], состоит в том, что если $\rho^2 > > 1/(n - 1)$, то оценка \tilde{z}_1 более эффективна, чем a_1 . Наоборот, при малых значениях ρ оценка МНК может быть эффективнее оценки Зеллнера. Полученный результат можно было предугадать. При уменьшающихся значениях ρ оценка МНК приближается к линейно эффективной, а в условиях «нормальной» гипотезы — к эффективной несмещенной оценке.

Аналогичное исследование проведено в [164]; там приняты те же предположения, что и в [175], но не считается, что X_1 является подматрицей X_2 , а рассматривается общий случай. Результаты, полученные этими авторами, также похожи на результаты [175]: при малых $|\rho|$ более эффективен МНК, при высоких $|\rho|$ — метод Зеллнера. Кроме значения ρ , на сравнительную эффективность МНК и метода Зеллнера оказывает зависимость между независимыми переменными в первом и втором уравнениях. В качестве коэффициента зависимости между X_1 и X_2 в [164] предложено брать характеристические числа матрицы $X_1' X_2 (X_2' X_2)^{-1} \times \times X_2' X_1 (X_1' X_1)^{-1}$. Можно показать, что $0 \leq \lambda_i \leq 1$. В [199] показано, что λ_i — квадраты коэффициентов канонической корреляции между X_1 и X_2 . В [164] приведена таблица, с помощью которой для значений $n = 3, 5, 9, 13$ и 23 для различных значений ρ и λ можно вычислить относитель-

ную эффективность оценки МНК по отношению к оценке \tilde{z} . Например, оказывается, что если $13 + K \leq n \leq 23 + K$, где K — общее число независимых переменных, и $|\rho| \geq \geq 0,5$ и $\lambda_i \leq 0,7$, то оценка \tilde{z} заметно эффективнее оценки МНК; с другой стороны, если $|\rho| < 0,5$ или $\lambda_i > 0,7$, то выигрыш будет либо равен нулю, либо незначителен. Для малых выборок, т. е. если $3 + K < n < 13 + K$, область «неэффективности» \tilde{z} весьма широка: $|\rho| \leq 0,8$. Окончательный вывод, к которому приходят авторы, следующий: если $|\rho| > 0,3$ и $n - K \geq 23$, то предпочтительнее выбирать \tilde{z} , а не оценку МНК; если $|\rho| < 0,3$, эффективность \tilde{z} будет не высока по сравнению с a ; при малых выборках \tilde{z} будет по-прежнему более эффективна, если $|\rho|$ достаточно велико, а λ_i близки к нулю.

В [198] рассматривается другой частный случай, когда экзогенные переменные в разных уравнениях ортогональны друг другу, т. е. $X_1'X_2 = 0$. В этом случае $\lambda_i = 0$, поэтому рассмотренная там ситуация является наиболее благоприятной для оценки \tilde{z} . Там же табулируется функция сравнительной эффективности z для разных значений $\rho = \omega_{12}/\sqrt{\omega_{11}\omega_{22}}$ и n .

В работе [147] проведено тщательное исследование методом Монте-Карло конкурирующих оценок псевдонезависимых регрессий: оценки МНК, оценки Zellnera, итеративной оценки ММП. Авторы рассмотрели 4 модели: первая модель состояла из двух уравнений, вторая — небольшая модификация первой, третья — из четырех уравнений и пятая — из двух уравнений, где в качестве некоторых независимых переменных рассматривались независимые переменные с лагом. Кроме этого было выбрано 10 различных спецификаций для случайных отклонений. Например, в одной из спецификаций $\sigma_0^2 \omega_{11} = \sigma_0^2 \omega_{22} = 1$, для первой, второй и четвертой моделей ρ ($\varepsilon^1, \varepsilon^2$) полагался равным 0,925 и 0,6; для третьей модели — 0,941; 0,640 и 0. В других спецификациях $\omega_{11} \neq \omega_{22}$.

Основной вывод, к которому пришли авторы, следующий: оценка Zellnera, итеративная оценка Zellnera и оценка ММП оказались практически одинаковыми. По этой причине *предпочтительнее оценка Zellnera*, как наиболее простая из трех оценок. Оценка МНК оказалась менее эффективной, чем оценка Zellnera для большинства экспериментов.

У п р а ж н е н и я 2.5

1. Докажите, что если условие а) (см. с. 99) выполняется, то ранг составной матрицы X равен $m = \sum m_i$.

2. Используя теорему 2.5, докажите, что если матрица независимых переменных одна и та же для всех регрессий системы (2.42), то оценка Эйткена и оценка МНК совпадают.

3. Докажите состоятельность (2.46) при условии сильной регулярности матриц X_i .

4. Докажите, что $\lambda_{\min}(X_i' X_i) \rightarrow \infty$ для всех $i = 1, \dots, k$ влечет $\lambda_{\min}(X X') \rightarrow \infty$. Верно ли обратное?

5. Докажите, что матрицы X_i , $i = 1, \dots, k$, сильно регулярны тогда и только тогда, когда матрица X сильно регулярна.

6. Найдите условия ограниченности снизу функции (2.51). Начать со случая $k = 1$.

7. Докажите, что функция (2.51) не является выпуклой вниз функцией. (Сведите функцию (2.51) к функциям одного аргумента, для которой затем найдите вторую производную — см. приложение П.3.)

2.6. Вычислительные трудности МНК

Оценка МНК является решением системы нормальных уравнений

$$X' X a = X' y. \quad (2.57)$$

Таким образом, вычисление оценки МНК соответствует решению системы линейных уравнений. В настоящее время число методов решения линейных систем очень велико, исследованы свойства этих методов, разработаны многочисленные программы. В регрессионном анализе вычисляется не только оценка МНК, но и ее матрица ковариаций $\sigma^2 (X' X)^{-1}$. Поэтому задачу нахождения a в системе (2.57) целесообразно решать обращением матрицы плана $X' X$, что ведет к отысканию эффективного алгоритма обращения симметричной матрицы. Основные трудности МНК возникают, когда матрица плана плохо обусловлена. В параграфе 6.1 введены меры плохой обусловленности матрицы $X' X$. Под плохой обусловленностью матрицы можно понимать, например, близость ее определителя к нулю. Показателем обусловленности служит также отношение максимального характеристического числа (х.ч.) матрицы к минимальному (см. [165]), т. е.

$$\kappa(X' X) = \lambda_{\max}(X' X) / \lambda_{\min}(X' X). \quad (2.58)$$

Чем больше отношение (2.58), тем хуже обусловленность матрицы $X' X$. Вычисление обратной матрицы, производится ли оно вручную или на ЭВМ, несет на себе ошибки округле-

ния промежуточных результатов. В [165] показано, что чем хуже обусловленность матрицы $(X'X)$, тем сильнее ошибки округления влияют на конечный результат $(X'X)^{-1}$. Это заставляет плохо обусловленные матрицы обращаться с большим количеством знаков в промежуточных вычислениях. Часто точности в 8 знаков не хватает для нахождения хорошего приближения оценки МНК в регрессиях. Так, уже при расчетах оценки МНК в параболическом тренде

$$y_t = \alpha_1 t^2 + \alpha_2 t + \alpha_3 + \varepsilon_t \quad (2.59)$$

на ЭВМ с обычной точностью (8 знаков) результаты оказываются очень грубыми. Только счет с двойной точностью (16 знаков) позволяет найти приемлемое приближение к истинной оценке МНК на ЭВМ.

Проверка различных программ, реализующих МНК, дана в [154]. Рассмотрена линейная регрессия от семи факторов: y — общее число занятых в экономике США (тыс. чел.); x_1 — дефлятор (индекс) цен (%); x_2 — валовой национальный продукт (млрд. дол.); x_3 — общее число безработных (тыс. чел.); x_4 — число военнослужащих (тыс. чел.); x_5 — неработающее население от 14 лет (тыс. чел.); x_6 — год; $x_7 \equiv 1$. В табл. 2.1 приведены статистические данные регрессии Дж. Лонгли [154]. Расчеты проводились на разных ЭВМ по разным алгоритмам. Часто результаты, т. е.

Таблица 2.1

y	x_1	x_2	x_3	x_4	x_5	x_6
60323	83,0	234,289	2356	1590	107608	1947
61122	88,5	259,426	2325	1456	108632	1948
60171	88,2	258,054	3682	1616	109773	1949
61187	89,5	284,599	3351	1650	110929	1950
63221	96,2	328,975	2099	3099	112075	1951
63639	98,1	346,999	1932	3594	113270	1952
64989	99,0	365,385	1870	3547	115094	1953
63761	100,0	363,112	3578	3350	116219	1954
66019	101,2	397,469	2904	3048	117388	1955
67857	104,6	419,180	2822	2857	118734	1956
68169	108,4	442,769	2936	2798	120445	1957
66513	110,8	444,546	4681	2637	121950	1958
68655	112,6	482,704	3813	2552	123366	1959
69564	114,2	502,601	3931	2514	125368	1960
69331	115,7	518,173	4806	2572	127852	1961
70551	116,9	554,894	4007	2827	130081	1962

оценки МНК, отличались не только первыми и вторыми значащими цифрами, но и знаками. Даже вычисления с двойной точностью не приводили к удовлетворительному результату. Истинное значение оценки МНК для данной регрессии удалось получить с помощью специального настольного калькулятора, работавшего с 40 значащими цифрами. Оцененная регрессия на этом калькуляторе имела следующий вид

$$y = 15,0619 x_1 - 0,0358 x_2 - 2,0202 x_3 - 1,0332 x_4 - \\ - 0,0511 x_5 + 1829,15 x_6 - 3482258,635. \quad (2.60)$$

В большинстве программ, рассмотренных Дж. Лонгли, матрица плана вычислялась по формуле ¹

$$(X'X)_{ij} = \sum_t (x_{ti} - \bar{x}_i)(x_{tj} - \bar{x}_j). \quad (2.61)$$

При переходе к эквивалентной записи

$$(X'X)_{ij} = \sum_t x_{ti} x_{tj} - n \bar{x}_i \bar{x}_j$$

удалось число верных значащих цифр в оценке МНК увеличить в некоторых случаях на две.

Как мы уже говорили, вычислительные трудности МНК связаны с обращением матрицы плана. Обращению матриц посвящена обширная литература (см., в частности, [68]). Существует много методов обращения матриц. Лонгли обнаружил, что в регрессионных задачах лучше работает алгоритм ортогонализации Грамма — Шмидта, в частности алгоритм обращения матриц, основанный на исключении по Гауссу. Основной причиной расхождения результатов для разных алгоритмов и ЭВМ Лонгли считает то обстоятельство, что алгоритмы обращения матриц, реализованных на ЭВМ, предназначены для ручных вычислений или с применением калькуляторов. При переходе на ЭВМ алгоритмы теряют свою эффективность.

Часто исследователь не подозревает, что «машинное решение» неверно. Хорошим индикатором правильного решения в случае регрессий со свободным членом является сумма отклонений регрессии $\sum e_t$, где $e_t = y_t - a_1 x_{t1} - \dots - a_m x_{tm}$. Как следует из параграфа 1.6, эта сумма должна быть равна нулю. Если значение $\sum e_t$ отличается от нуля достаточно заметно, то регрессия оценена неверно. Однако

¹Имеются в виду регрессии со свободным членом.

равенство нулю $\sum e_i$ еще не означает, что регрессия оценена правильно. Часто для проверки правильности решения бывает полезно поменять местами независимые переменные и сравнить результаты. Решения будут содержать меньшие ошибки, если матрицу плана перед обращением привести к корреляционной, т. е. вычисления производить с матрицей $(X'X)_{ij}/\sqrt{(X'X)_{ii}(X'X)_{jj}}$. Этот метод будет наиболее эффективен, если независимые переменные имеют разные масштабы измерения.

Результаты, полученные Дж. Лонгли, а также практика автора показывают, что к полученным с помощью ЭВМ оценкам МНК, особенно для многофакторных регрессий, следует относиться осторожно, не делая скоропалительных выводов. В любом случае вычисления следует проводить с двойной точностью.

В работе [84] вновь обращено внимание на регрессию Лонгли (2.60). Был поставлен следующий вопрос: является ли решение Лонгли, проведенное с 40 значащими цифрами, в действительности удовлетворительным? Авторы [84] справедливо заметили, что поскольку данные представляют собой результат округления и если $x_{11} = 83$, то истинное значение x_{11} вполне может находиться в интервале (82,5; 83,4). Аналогичное замечание верно для всех независимых переменных $x_1 \div x_5$. Авторы просчитали 1000 регрессий. В каждом варианте значения независимых переменных $x_1 \div x_5$ отличались от опубликованных (табл. 2.1) на случайную величину с равномерным распределением в разряде, следующем за истинным значением. Так, для x_{11} значение независимой переменной выбиралось на интервале (82,5; 83,499). Результаты 1000 регрессий превзошли все ожидания. В табл. 2.2 приводится выдержка табл. 4 из [84]. Как видим, средние значения 1000 регрессий далеки от «истинных» значений, найденных Лонгли. Значения оценок МНК лежат в очень широких границах. Результаты испытаний говорят сами за себя. Регрессию (2.60) нельзя считать удовлетворительной. Вероятно, регрессию Лонгли вообще не имеет смысла оценивать методом наименьших квадратов, так как в данных присутствует мультиколлинеарность.

Идея проверки регрессии на устойчивость относительно ошибок округления, примененная в [84] к регрессии (2.60), заслуживает внимания. Аналогичные расчеты могут быть проделаны для любой другой регрессии. Для каждого числа (конечной десятичной дроби) N может быть указан интервал чисел, каждое из которых после округления дает

Таблица 2.2

Статистика	α_1	α_2	α_3	α_4	α_5	α_6	α_7
Регрессия Лонгли	15,0619	-0,0358	-2,0202	-1,0332	-0,0511	1829,15	-3482258
Средняя оценка	-26,4404	0,0344	-0,9637	-0,2804	-0,2804	637,098	-1152648
Стандартное отклонение	59,7	0,024	0,351	0,15	0,161	326	637918
Нижняя граница	-232,3	-0,089	-2,423	-1,326	-0,939	-1707	-3483281
Верхняя граница	237,0	0,196	1,767	0,363	0,481	1800	3452563
Процент соответствия ¹	2	3,9	0,7	6,4	94,6	0,9	1
plim	-32,455	0,0449	-0,8104	-0,6794	-0,3148	460,01	-806545

¹Т. е. процент регрессий, у которых оценка совпадает с оценкой Лонгли в первом знаке.

число N . Так, если $N = 25,3$, то таким интервалом будет $[25,25; 25,35)$. Аналогичный интервал может быть построен для каждого x_{ti} , $t = 1, \dots, n$, $i = 1, \dots, m$. Таким образом, матрице независимых переменных X соответствует целое множество матриц, каждая из которых после округления дает X . Обозначим это множество Γ — множество матриц $n \times m$. На место оценки $a_X = (X'X)^{-1} X'y$ приходит семейство оценок a_X , $X \in \Gamma$. Вообще говоря, матрица X равноправна с любой другой матрицей из Γ , поэтому имеет смысл найти минимальные и максимальные координаты оценки МНК при $X \in \Gamma$. Итак, обозначим через a_i i -ю координату оценки МНК, тогда можно определить

$$\min_{X \in \Gamma} a_i(X), \max_{X \in \Gamma} a_i(X), i = 1, \dots, m. \quad (2.62)$$

Вообще говоря, можно найти более общее множество

$$a(\Gamma) = \{a \in R^m : a = (X'X)^{-1} X'y, X \in \Gamma\}. \quad (2.63)$$

Прежде всего необходимо выяснить, существует ли матрица $X \in \Gamma$, для которой $\text{rang } X < m$. Ясно, что если такая матрица существует, то оценка МНК будет некорректной. Далее, если для всех $X \in \Gamma$ $\text{rang } X = m$, могут быть найдены значения (2.62). Однако отыскание этих значений весьма сложно. Для их приближенного определения можно применить процедуру статистических испытаний: случайным образом отбирать $X \in \Gamma$ и вычислять значения (2.62). Минимальное значение i -й координаты оценки МНК по всем испытаниям соответствует $\min_{X \in \Gamma} a_i(X)$, максимальное значение — $\max_{X \in \Gamma} a_i(X)$. Эта процедура может служить хорошей проверкой устойчивости оценки МНК по матрице независимых переменных.

Перейдем к определению эффекта от округления значений независимых переменных. В [84] введен «индекс возмущения» (perturbation index) следующим образом. Пусть, как и прежде, матрица X — матрица независимых переменных регрессии. На основе имеющейся матрицы X будем образовывать новые матрицы $Z = X + \Lambda$, так что после округления матрица Z превращается в X , Λ — матрицу отклонений, причем $E\Lambda = 0$,

$$\frac{1}{n} E\Lambda' \Lambda = D = \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \sigma_m^2 \end{bmatrix}.$$

Пусть имеется k матриц Z_i . Обозначим

$$Z_* = \begin{bmatrix} X + \Lambda_1 \\ \vdots \\ X + \Lambda_k \end{bmatrix}, y_* = \begin{bmatrix} y \\ \vdots \\ y \end{bmatrix}$$

— составные матрицы независимых и зависимой переменных размерности $nk \times m$ и $nk \times 1$ соответственно. Найдем

$$a_* = (Z_*' Z_*)^{-1} Z_*' y_* = \left[kX'X + X' \sum_i \Lambda_i + \sum_i \Lambda_i' X + \sum_i \Lambda_i' \Lambda_i \right]^{-1} \left[kX' y + \sum_i \Lambda_i' y \right].$$

Найдем вероятностный предел a_* :

$$\begin{aligned} \text{plim}_{k \rightarrow \infty} a_* &= \text{plim} \left[X'X + X' \frac{\sum \Lambda_i}{k} + \frac{\sum \Lambda_i'}{k} + \frac{\sum \Lambda_i' \Lambda_i}{k} \right]^{-1} \times \\ &\times \left[X' y + \frac{\sum \Lambda_i'}{k} y \right] = \left[X'X + X' \text{plim} \frac{\sum \Lambda_i}{k} + \right. \\ &+ \text{plim} \frac{\sum \Lambda_i'}{k} X + \left. \text{plim} \frac{\sum \Lambda_i' \Lambda_i}{k} \right]^{-1} \left[X' y + \text{plim} \frac{\sum \Lambda_i'}{k} y \right] = \\ &= (X'X + D)^{-1} X' y \end{aligned}$$

в силу закона больших чисел. Тогда

$$a - \text{plim} a_* = [I - (X'X)^{-1} D]^{-1} a. \quad (2.64)$$

Ясно, что чем ближе матрица $(X'X)^{-1} D$ к нулю, тем меньше будет разность (2.60), т. е. тем менее будет отличаться a от a_* . Индексом возмущения авторы [84] назвали след матрицы $(X'X)^{-1} D$, т. е.

$$PI = \text{tr} (X'X)^{-1} D. \quad (2.65)$$

Если регрессия устойчива к округлениям независимых переменных, то PI должен быть близким к нулю. Для регрессии (2.60) индекс (2.61) был равен 2,98.

С окончательным выводом авторов [84] о том, что вычислительная программа часто является далеко не самым важным фактором при вычислении регрессий, в некоторых случаях целесообразнее вообще отказаться от вычислений, нельзя не согласиться. Часто ошибки в данных на порядок выше ошибок, привносимых программой. Для устойчивых регрессий даже самые плохие программы давали хорошие результаты. Одним из примеров, когда необходимо вообще отказаться от вычислений по методу наименьших квадратов, как показано в [84], является регрессия Лонгли.

Часть вторая

АЛЬТЕРНАТИВНЫЕ СХЕМЫ И МЕТОДЫ ОЦЕНИВАНИЯ

Глава 3

РЕГРЕССИЯ КАК УСЛОВНОЕ МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ

3.1. Основные предположения

В первых двух главах независимые переменные считались детерминированными. В качестве независимых переменных выступали либо контролируемые, управляемые величины, которые задавались экспериментатором — регрессии управляемого эксперимента, либо функции номеров наблюдений — регрессии-тренды. В этой главе будем считать зависимые, и независимые переменные случайными (стохастическими). В математической статистике под регрессией случайной величины y на случайную величину x понимают условное математическое ожидание $E(y/x)$. В смысле этого определения модель (1.1) нельзя считать регрессией. Правильнее было бы назвать ее линейной моделью математического ожидания (не условного).

Однако в силу традиций и установившейся терминологии модель (1.1) была названа регрессией. В данной главе мы будем изучать модель условного математического ожидания, т. е. термин «регрессия» будем употреблять в корректном смысле.

Итак, примем следующие предположения:

Предположение D' . X — случайная матрица.

Предположение E' . $\text{rang} X = m$ с вероятностью 1.

Предполагаем также, что условное математическое ожидание y при заданном X есть линейная функция, неизвестная с точностью до коэффициентов $\alpha = (\alpha_1, \dots, \alpha_m)' \in R^m$: $E(y/X) = X\alpha$. Схема «математической регрессии» обходится без введения отклонений ϵ , поэтому предположение В здесь излишне. Предположение Г необходимо переписать в терминах условного математического ожидания.

Предположение Γ'' . Ковариационная матрица у при фиксированной матрице X имеет вид $\sigma^2 I_n$, т. е.

$$\text{cov}(y/X) = E\{(y - X\alpha)(y - X\alpha)' / X\} = \sigma^2 I_n.$$

Итак, в данной главе считаем предположения А, Б (параграф 1.1), Γ'' , D' и E' выполненными.

Распределение матрицы X может быть известно или неизвестно с точностью до конечного числа неизвестных параметров.

3.2. Свойства оценки МНК

Прежде всего в схеме регрессии

$$E(y/X) = X\alpha \quad (3.1)$$

изучим свойства оценки метода наименьших квадратов

$$\alpha = (X'X)^{-1}X'y. \quad (3.2)$$

Предположение E' позволяет утверждать, что оценка (3.2) существует почти наверное, т. е. с вероятностью 1. Для дальнейшего исследования свойств оценки МНК нам понадобится следующая формула. Пусть x и y — две случайные величины и $z = z(x, y)$ — третья случайная величина, которая является их функцией. Тогда

$$E z(x, y) = E_x \{E(z(x, y)/x)\}. \quad (3.3)$$

В фигурных скобках стоит условное м. о. z при фиксированном x . Для получения м. о. (безусловного) случайной величины z необходимо взять м. о. (безусловное) $E(z(x, y)/x)$ функции x . Доказательство (3.3.) для абсолютно непрерывных функций плотностей x и y приводится в параграфе 3.4. Очевидным образом формула (3.3.) переносится на многомерный случай.

Докажем несмещенность оценки МНК. Применяя многомерный аналог формулы (3.3), получим

$$\begin{aligned} E\alpha &= E_X(E(a/X)) = E_X[E((X'X)^{-1}X'y)/X] = \\ &= E_X((X'X)^{-1}X'E(y/X)) = E_X((X'X)^{-1}X'X\alpha) = E_X\alpha = \alpha. \end{aligned}$$

Легко находится матрица ковариаций оценки МНК:

$$\begin{aligned} \text{cov}(a) &= E(a - \alpha)(a - \alpha)' = E_X\{E((a - \alpha)(a - \alpha)' / X)\} = \\ &= E_X\sigma^2(X'X)^{-1} = \sigma^2 E(X'X)^{-1}. \end{aligned} \quad (3.4)$$

Формула (3.4) отличается от соответствующей формулы в случае постоянной матрицы X наличием знака математического ожидания. Таким образом, для вычисления матрицы ковариаций оценки МНК в регрессии (3.1) необходимо знание истинного распределения случайных величин x_1, x_2, \dots, x_n . Попытки оценивать матрицу $E(X'X)^{-1}$ на основе оценки матрицы $E(X'X)$ приводят к занижению результата. Это следует из матричного аналога неравенства Коши:

$$I = E[(X'X)(X'X)^{-1}] \leq E(X'X) \cdot E(X'X)^{-1},$$

откуда

$$[E(X'X)]^{-1} \leq E(X'X)^{-1},$$

где знак неравенства понимается в том смысле, что разность между правой и левой частями есть неотрицательно определенная матрица.

В качестве иллюстрации рассмотрим следующий пример. Пусть x_1, x_2, \dots, x_n — независимые одинаково распределенные случайные величины, $x_i \sim N(0, \theta^2)$, т. е. математическое ожидание x_i равно нулю, а дисперсия $\theta^2 > 0$. Зависимая переменная y связана с x_1, x_2, \dots, x_n таким образом, что $E(y_i/x_1, x_2, \dots, x_n) = \alpha x_i, i = 1, \dots, n$. Найдем дисперсию оценки МНК. По формуле (3.4) для этого необходимо найти $E \frac{1}{\sum_{i=1}^n x_i^2}$. Воспользуемся следующей формулой:

лой:

$$\int_{R^n} \psi \left(\sqrt{\sum_{i=1}^n x_i^2} \right) dx_1 \dots dx_n = 2 \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)} \int_0^\infty r^{n-1} \psi(r) dr,$$

где ψ — некоторая непрерывная функция (см., например, [7, с. 307]), Γ — гамма-функция. По этой формуле

$$\begin{aligned} E \frac{1}{\sum x_i^2} &= (2\pi\theta^2)^{-\frac{n}{2}} \int_{R^n} \frac{1}{\sum x_i^2} e^{-\frac{1}{\theta^2} \sum x_i^2} dx_1 \dots dx_n = \\ &= (2\pi\theta^2)^{-\frac{n}{2}} \frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)} \int_0^\infty r^{n-3} e^{-\frac{r^2}{2\theta^2}} dr. \end{aligned}$$

По формуле 130 [7, с. 297] находим окончательно

$$E \frac{1}{\Sigma x_t^2} = 2 \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)} (\theta \sqrt{2})^{n-2} \cdot \frac{1}{2} \Gamma\left(\frac{n-2}{2}\right) (2\pi\theta^2)^{-n/2} = \\ = \frac{1}{\theta^2 (n-2)}.$$

Итак, дисперсия оценки МНК равна $\sigma^2/(n-2)\theta^2$. Легко видеть, что для $n \leq 2$ дисперсия оценки МНК является бесконечной. Подсчет по формуле $1/E \Sigma x_t^2$ приводит к значению $1/\theta^2 n < 1/\theta^2 (n-2)$, т. е. дает заниженное значение дисперсии. При больших n разница, однако, будет ничтожна.

Оценка s^2 (1.20) также остается несмещенной:

$$E s^2 = E_X (E (s^2/X)) = E_X \sigma^2 = \sigma^2.$$

Перейдем к оптимальным свойствам оценки МНК в смысле ее эффективности. Естественен вопрос: верна ли теорема Гаусса — Маркова для схемы условного математического ожидания (3.1). Полностью ответить на этот вопрос мы не можем. Покажем, что если известны некоторые характеристики случайной матрицы X , то существует линейная несмещенная оценка, которая имеет меньшую матрицу ковариаций, чем оценка МНК. Построение такой оценки начнем со случая $m = 1$. Все математические ожидания будем предполагать конечными. Пусть $\{y_t, x_t\}$ — система случайных величин, причем

$$E (y_t/x_1, x_2, \dots, x_n) = \alpha x_t, \quad t = 1, \dots, n. \quad (3.5)$$

Оценкой МНК для регрессии (3.5) является $a = \Sigma x_t y_t / \Sigma x_t^2$ с дисперсией $\sigma^2(a) = \sigma^2 E \frac{1}{\Sigma x_t^2}$. Пусть $a_1 = \Sigma c_t y_t$, $c_t = c_t(x)$ — другая линейная по y несмещенная оценка. Несмещенность a_1 влечет для любых α равенство

$$E a_1 = E \Sigma c_t y_t = E_x (E (\Sigma c_t y_t/x)) = E_x (\Sigma c_t x_t) \alpha = \alpha.$$

Таким образом, a_1 будет несмещенной оценкой, только если $E (\Sigma c_t x_t) = 1$.

Обозначим $b_t = b_t(x) = c_t - x_t / \Sigma x_t^2$, тогда условие несмещенности переписывается:

$$E \Sigma b_t x_t = 0. \quad (3.6)$$

Далее нетрудно показать, что $\sigma^2(a_1) = \sigma^2 \cdot E \Sigma c_t^2 +$

+ $\alpha^2 \sigma^2 (\sum c_t x_t)$. Выражая c_t через b_t , получим

$$\begin{aligned} \sigma^2(a_1) &= \sigma^2 \cdot E \sum_t \left(b_t + \frac{x_t}{\sum_t x_t^2} \right)^2 + \alpha^2 \sigma^2 (\sum x_t b_t) = \\ &= \sigma^2 \cdot E \frac{1}{\sum x_t^2} + \sigma^2 \cdot E \left[\sum b_t^2 + 2 \frac{\sum b_t x_t}{\sum x_t^2} \right] + \alpha^2 \sigma^2 (\sum x_t b_t) \end{aligned}$$

или

$$\sigma^2(a_1) - \sigma^2(a) = \sigma^2 \cdot \left[E \sum b_t^2 + 2E \frac{\sum b_t x_t}{\sum x_t^2} \right] + \alpha^2 \sigma^2 (\sum x_t b_t), \quad (3.7)$$

где $\sigma^2 (\sum x_t b_t)$ — дисперсия случайной величины $\sum x_t b_t$. Для того чтобы оценка a_1 имела меньшую дисперсию, чем оценка МНК, необходимо и достаточно показать, что правая часть (3.7) отрицательна. Найдем такие $b_t = b_t(x)$, что (3.7) будет отрицательно. Обозначим $g(x) = E \frac{1}{\sum x_t^2} - \frac{1}{\sum x_t^2}$. Легко показать, что $Eg(x) = 0$ и

$$E \frac{g(x)}{\sum x_t^2} = E^2 \frac{1}{\sum x_t^2} - E \frac{1}{(\sum x_t^2)^2} = -Eg^2(x) \leq 0, \quad (3.8)$$

где $Eg^2(x)$ — дисперсия случайной величины $1/\sum x_t^2$. Предположим теперь, что $\sum x_t^2 \neq \text{const}$ с вероятностью 1. Тогда неравенство (3.8) будет строгим, так как в противном случае с вероятностью 1

$$\frac{1}{\sum x_t^2} = \text{const} = \frac{1}{k} > 0; \quad \sum x_t^2 = k = \text{const}.$$

Допустим, $E 1/\sum x_t^2$ известно и $p > 0$ — некоторое число, положим

$$b_t = pg(x)/x_t. \quad (3.9)$$

Тогда $\sum b_t x_t = npg(x)$ и условие несмещенности (3.6) будет выполнено: $E \sum b_t x_t = 0$. Далее из (3.9) находим

$$\sum b_t^2 = p^2 g^2(x) \sum \frac{1}{x_t^2},$$

откуда

$$E \frac{\sum b_t x_t}{\sum x_t^2} = n p E \frac{g(x)}{\sum x_t^2} = -n p E g^2(x) < 0.$$

Правая часть выражения (3.7) переписывается следующим образом:

$$\begin{aligned} & \sigma^2 p^2 E \left[g^2(x) \sum_t \frac{1}{x_t^2} \right] - 2np E g^2(x) + \alpha^2 p^2 E g^2(x) = \\ & = \left\{ \sigma^2 E \left[g^2(x) \sum_t \frac{1}{x_t^2} \right] + \alpha^2 E g^2(x) \right\} p^2 - 2np E g^2(x). \end{aligned}$$

Если p выбрать из интервала

$$0 < p < 2n E g^2(x) \left/ \left\{ E \left[g^2(x) \sum_t \frac{1}{x_t^2} \right] + \alpha^2 E g^2(x) \right\} \right.,$$

то выражение (3.7) будет отрицательным и оценка a_1 будет иметь меньшую дисперсию, чем оценка МНК. Таким образом, нами доказана следующая теорема.

Теорема 3.1. *Если математическое ожидание $1/\sum_t x_t^2$ известно, то оценка МНК в классе линейных несмещенных оценок в условной регрессии ($m = 1$) не является эффективной, а линейная несмещенная оценка*

$$\begin{aligned} a_1 &= \frac{\sum y_t x_t}{\sum x_t^2} + p \left\{ E \frac{1}{\sum x_t^2} - \frac{1}{\sum x_t^2} \right\} \sum_t \frac{y_t}{x_t} = \\ &= a + pg(x) \sum_t \frac{y_t}{x_t}, \end{aligned} \quad (3.10)$$

где $g(x) = E \frac{1}{\sum x_t^2} - \frac{1}{\sum x_t^2}$, $p > 0$ достаточно мало, причем $\sum x_t^2 \neq \text{const}$ с вероятностью 1 и $P\{x_t = 0\} = 0$ для всех $t = 1, \dots, n$, для данного α имеет меньшую дисперсию, чем оценка МНК.

З а м е ч а н и я: 1. Можно найти непрерывные случайные величины x_1, x_2, \dots, x_n , для которых $\sum x_t^2 = \text{const}$, и поэтому $E \frac{1}{\sum x_t^2} - \frac{1}{\sum x_t^2} = 0$, что в свою очередь означает $a_1 = a$.

Действительно, пусть x_1 — случайная величина, равномерно распределенная на $(0,1)$. Положим $x_2 = \sqrt{1-x_1^2}$. Как легко заметить, $x_1^2 + x_2^2 = 1$ с вероятностью 1.

2. Если найдется такое t_0 , что x_{t_0} будет неограниченной случайной величиной, т. е. $P\{|x_{t_0}| > A\} > 0$ для любого A , то условие $\sum x_t^2 \neq \text{const}$ автоматически выполняется. Предположим противное: $\sum x_t^2 = k = \text{const}$, причем, например, x_2 — неограниченная случайная величина. Тогда

$x_1^2 = k - x_2^2 - \sum_{i=3}^n x_i^2 \geq k - x_2^2$. Если взять $A = \sqrt{k+1}$, то приходим к противоречивому соотношению

$$0 < P \{x_2^2 > k + 1\} \leq P \{x_2^2 < -1\}.$$

Теорема 3.1 без труда переносится на многомерный случай регрессии (3.1). Пусть X — матрица порядка $n \times m$, причем $\text{rang } X = m < n$. Найдется такая матрица X^- порядка $n \times m$, что $(X^-)' X = I_m$ — единичная матрица. Построение X^- осуществляем следующим образом. Обозначим x_1, x_2, \dots, x_m — вектор-столбцы матрицы X . По предположению они линейно независимы. Пусть e_1 — перпендикуляр, опущенный из конца вектора x_1 на линейное подпространство, порождаемое остальными векторами x_2, x_3, \dots, x_m . Очевидно, $e_1 \neq 0$, так как в противном случае x_1 линейно выражался бы через x_2, \dots, x_m . Нормируем e_1 таким образом, чтобы $e_1' x_1 = 1$. Аналогично построим векторы e_2, \dots, e_m . Матрица X^- тогда состоит из вектор-столбцов e_1, e_2, \dots, e_m . Легко проверяется, что $(X^-)' X = I_m$.

Введем обозначение

$$G = G(X) = E(X'X)^{-1} - (X'X)^{-1}. \quad (3.11)$$

Предыдущая теорема может быть перенесена на многомерный случай.

Теорема 3.2. Если $E(X'X)^{-1}$ известно, то оценка МНК в классе линейных несмещенных оценок в условной регрессии не является эффективной, а несмещенная линейная оценка

$$a_1 = (X'X)^{-1} X' y + p [E(X'X)^{-1} - (X'X)^{-1}] X^- y = a + pGX^- y,$$

где $p > 0$ достаточно мало и $G \neq 0$, с вероятностью 1 для любого заданного a имеет меньшую матрицу ковариаций, чем оценка МНК (в смысле $\text{cov}(a_1) \leq \text{cov}(a)$).

Доказательство теоремы приведено в [25].

Практическое использование теорем 3.1 и 3.2 весьма проблематично, но ценность их в том, что они показывают возможное ухудшение свойств оценки МНК при переходе к модели (3.1).

Остановимся на результатах, приведенных в [85]. Вместо класса линейных несмещенных оценок был рассмотрен класс линейных оценок с ограниченной функцией риска

(см. параграф 1.4), т. е. с ограниченной матрицей средних квадратов отклонений

$$E(\mathbf{b} - \alpha)(\mathbf{b} - \alpha)', \quad (3.12)$$

где \mathbf{b} — линейная статистика по y . Матрица (3.12) есть функция α . Статистика \mathbf{b} будет принадлежать к классу статистик с ограниченной функцией риска, если матрица (3.12) ограничена постоянными матрицами, не зависящими от $\alpha \in \Theta$ для любого распределения матрицы \mathbf{X} из заданного класса распределений. В параграфе 1.4 показано, что если в качестве неизвестного параметра выступает математическое ожидание, то класс линейных несмещенных оценок и класс линейных оценок с ограниченной функцией риска совпадают. Покажем, что это не происходит в схеме регрессии (3.1). Можно доказать, что класс линейных оценок с ограниченной функцией риска уже класса линейных несмещенных оценок. Для этого рассмотрим лемму.

Л е м м а 3.1. *Если $\mathbf{b} = \mathbf{c}(\mathbf{X})$ — линейная по y оценка неизвестного вектора параметров α с ограниченной функцией риска, то*

$$E(\mathbf{b}/\mathbf{X}) = \alpha, \quad (3.13)$$

или, что то же самое, $\mathbf{C}\mathbf{X} = \mathbf{I}_m$ почти для всех \mathbf{X} .

Легко проверяется и обратное: если \mathbf{b} — линейная статистика и $\mathbf{C}\mathbf{X} = \mathbf{I}_m$, то \mathbf{b} имеет ограниченную функцию риска. Доказательство леммы см. в [85].

Покажем, почему класс линейных оценок с ограниченной функцией риска уже класса линейных несмещенных оценок. Пусть \mathbf{b} — линейная оценка по y , т. е. $\mathbf{b} = \mathbf{C}y$, где $\mathbf{C} = \mathbf{C}(\mathbf{X})$. Условие несмещенности оценки \mathbf{b} записывается в виде

$$E\mathbf{b} = E_{\mathbf{X}}\{E(\mathbf{b}/\mathbf{X})\} = E_{\mathbf{X}}\{E(\mathbf{C}y/\mathbf{X})\} = E_{\mathbf{X}}(\mathbf{C}\mathbf{X}\alpha) = E(\mathbf{C}\mathbf{X})\alpha = \alpha$$

для всех α , т. е.

$$E(\mathbf{C}\mathbf{X}) = \mathbf{I}_m, \quad (3.14)$$

тогда как условие ограниченности функции риска по лемме 3.1 — в виде (3.13). Ясно, что условие (3.13) более сильное. Если оно выполняется, то выполняется и (3.14).

На основе (3.13) нетрудно показать, что в более узком классе — классе линейных оценок с ограниченными функциями риска — оценка МНК уже будет оптимальной, что и составляет содержание следующей теоремы.

Теорема 3.3. *Оценка МНК является эффективной оценкой в классе линейных оценок с ограниченной функцией риска.*

Доказательство. Пусть $\mathbf{b} = \mathbf{C}\mathbf{y}$ — линейная оценка с ограниченной функцией риска. По лемме 3.1 эта оценка также является несмещенной, причем равенство (3.13) выполняется почти для всех матриц \mathbf{X} . Найдем матрицу ковариаций оценки \mathbf{b} :

$$\text{cov}(\mathbf{b}) = E_{\mathbf{X}} \{ E(\mathbf{b} - \boldsymbol{\alpha})(\mathbf{b} - \boldsymbol{\alpha})' / \mathbf{X} \} = \sigma^2 E(\mathbf{C}\mathbf{C}'). \quad (3.15)$$

Оптимальной является оценка, которая приводит к минимальной матрице (3.15) при ограничении (3.13)¹. Для фиксированной матрицы \mathbf{X} оценке МНК соответствует матрица $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, которая приводит к минимальному значению $\mathbf{C}\mathbf{C}'$. Теперь воспользуемся тем, что из условия

$$\mathbf{C}\mathbf{C}' \leq \mathbf{C}_1\mathbf{C}'_1 \quad (3.16)$$

следует

$$E(\mathbf{C}\mathbf{C}') \leq E(\mathbf{C}_1\mathbf{C}'_1). \quad (3.17)$$

Теорема доказана.

Нетрудно убедиться, что оценка a_1 (теоремы 3.1 и 3.2) не принадлежит к классу оценок с ограниченной функцией риска.

С помощью некоторых дополнительных предположений в [85] доказывается минимаксность оценки МНК в классе линейных оценок с ограниченной функцией риска.

Подведем итоги: оценка МНК в классе линейных несмещенных оценок перестает быть оптимальной в регрессии (3.1). В классе линейных оценок с ограниченной функцией риска оценка МНК продолжает быть оптимальной. Таким образом, если не рассматривать оценок, которые являются весьма плохими для некоторых значений параметров регрессии, то оценка МНК будет эффективной.

Коротко остановимся на асимптотических свойствах оценки МНК. Аналогично детерминированному случаю можно показать, что условие Эйкера (1.39) эквивалентно квадратичной сходимости.

Теорема 3.4. *Предположим, $E(\mathbf{X}'_n\mathbf{X}_n) < \infty$, тогда оценка МНК сходится к истинному значению в среднем квадратичном тогда и только тогда, когда с вероятностью 1*

$$\lambda_{\min}(\mathbf{X}'_n\mathbf{X}_n) \rightarrow \infty, \quad n \rightarrow \infty. \quad (3.18)$$

Доказательство теоремы приведено в параграфе 3.4.

¹Метод сравнения матриц, принятый в книге, изложен в параграфе 1.4.

Можно было бы найти условия асимптотической нормальности оценки МНК в регрессии (3.1), однако это представляет лишь теоретический интерес. Асимптотическая нормальность далее исследована в частном и важном случае случайной повторной выборки (параграф 3.2).

Рассмотрим, как влияет схема условного математического ожидания (3.1) на другие характеристики регрессии. Начнем с коэффициента детерминации. Раньше этот коэффициент терял статистический смысл в силу того, что $E y_t \neq \text{const}$, что в свою очередь являлось результатом того, что независимые переменные x_1, x_2, \dots, x_m представляли собой неслучайные (детерминированные) числа. В схеме регрессии (3.1) условие постоянства математического ожидания y_t может быть выполнено, например, в случае, когда $E x_{ti} = c_i$, т. е. не зависит от t . При этом условие невырожденности (ситуа́ция E') с этим фактом не связано. Подобная ситуация наблюдалась в приведенном выше примере, где $E x_t = E y_t = 0, t = 1, \dots, n$. Далее, если $\sigma^2(y_t) = E(y_t - E y_t)^2 = \text{const}$, т. е. не зависит от t , то коэффициент детерминации несет на себе первоначальный статистический смысл, а именно является показателем функциональности, адекватности регрессии. В частности, R^2 показывает, какая часть дисперсии была объяснена зависимостью (3.1). Если хотя бы одно из перечисленных условий не выполняется, R^2 теряет статистический смысл и остается показателем, отражающим, насколько модель регрессии лучше модели среднего. Поэтому для временных рядов, имеющих тренд, коэффициент детерминации регрессий как условных м.о. имеет тот же смысл, что и для классической регрессии.

Далее очевидно, распределение оценки МНК в схеме (3.1) перестает быть нормальным, даже если отклонения нормально распределены. Распределение a зависит от распределения независимых переменных. В схеме регрессии как условного математического ожидания оценка МНК не будет эффективной в классе несмещенных оценок с нормально распределенными отклонениями.

Для регрессии (3.1) изменится и критерий отношения правдоподобия. Он будет также зависеть от распределения независимых переменных. Критерии проверок гипотез, разработанные в параграфе 1.10, также не будут равномерно наиболее мощными. Однако очень важно, что доверительные интервалы, построенные для случая, когда матрица X стохастическая, имеют по-прежнему коэффициент до-

верия $1 - \lambda$. Это следует из того, что если множество $D = D(y, X)$ имеет постоянную условную вероятность $P(D(y, X)/X) = 1 - \lambda$, то $P(D(y, X)) = 1 - \lambda$. Это утверждение следует непосредственно из формулы (3.3). Таким образом, доверительные интервалы, построенные в параграфе 1.10, переставая быть равномерно наиболее точными, по-прежнему покрывают истинное значение параметров с вероятностью $1 - \lambda$. Дополнительная информация о распределении X может сузить доверительный интервал (или увеличить мощность критерия), однако если дисперсия x_{ii} очень мала, то построенные критерии проверки гипотез будут близки к оптимальным.

У п р а ж н е н и я 3.2

1. Эквивалентна ли запись (3.5) записи $E(y_t/x_t) = \alpha x_t$? Рассмотрите пример авторегрессии первого порядка $E(y_t/y_{t-1}) = \alpha y_{t-1}$, где $x_t = y_{t-1}$, $t = 1, 2, \dots, n$.

2. При каких условиях $E(g^2 \Sigma 1/x_t^2) < \infty$?

3. Постройте аналог оценки a_1 для случая, когда найдется $t_0 \in [1, n]$, что $P\{x_{t_0} = 0\} \neq 0$ и для всех $t \neq t_0$ $P\{x_t = 0\} = 0$.

4. Докажите, что матрица X^{-1} имеет ранг, равный m .

5. Докажите, что $E(b - \alpha)(b - \alpha)' \leq D$ не зависит от α тогда и только тогда, когда для любой матрицы $\Omega E(b - \alpha)' \Omega(b - \alpha) \leq d(\Omega)$ не зависит от α .

6. Докажите, что $E(X'X)^{-1} < \infty$ влечет $E(X'X)^{-1}X' < \infty$.

3.3. Схема случайной выборки

Важнейшим частным случаем схемы регрессии (3.1) является случай независимых, одинаково распределенных наблюдений, т. е. случайной выборки. Итак, имеется $m + 1$ случайных величин y, x_1, x_2, \dots, x_m . Эти величины имеют свою функцию распределения, математические ожидания, дисперсии и т. д. (считаем, что все они конечны). Обозначим $\mu_0 = Ey$, $\mu_i = E x_i$, $i = 1, \dots, m$, $\sigma^2(y) = \sigma_0^2 = E(y - \mu_0)^2 > 0$, $C = \text{cov}(x) = E(x - \mu)(x - \mu)'$ положительно определена и $x = (x_1, \dots, x_m)'$, $\mu = (\mu_1, \dots, \mu_m)'$. Относительно случайных величин y, x известно, что

$$E(y/x) = E(y/x_1, \dots, x_m) = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m = \alpha'x, \quad (3.19)$$

где $(\alpha_1, \alpha_2, \dots, \alpha_m)'$ — вектор неизвестных параметров. Таким образом, регрессия y на x линейна и неизвестна

с точностью до своих коэффициентов. Предположим, разброс около регрессии постоянен, т. е.

$$\sigma^2 (y/x) = E_x \{E(y - E(y/x))^2/x\} = \text{const} = \sigma^2 \quad (3.20)$$

также неизвестно. Из случайной величины $(y, x) \in R^{m+1}$ производится случайная выборка (y_t, x_t) , $t = 1, \dots, n$, т. е. (y_t, x_t) независимы и одинаково распределены. Наблюдения образуют вектор $y \in R^n$ и матрицу $X^{n \times m}$. В силу независимости $\{y_t, x_t\}$ равенство (3.19) эквивалентно уравнению $E(y/X) = X\alpha$. Далее можно показать, что в силу положительной определенности $C \doteq \text{cov}(x) \text{ rank } X = m$ с вероятностью 1. Из независимости $\{y_t\}$ следует выполнимость предположения Γ^n (параграф 1.1). Итак, все предположения для схемы случайной выборки выполняются.

Существует много задач, в которых наблюдения можно считать результатом случайной выборки. Так, если исследуется зависимость урожая некоторой однолетней сельскохозяйственной культуры от средней температуры лета и количества выпавших осадков, то наблюдения за n лет дают нам: x_{t1} — среднюю температуру лета в году t ; x_{t2} — количество осадков, выпавших в течение вегетации культуры в году t . Мы предполагаем, что урожай y_t так зависит от x_{t1} и x_{t2} , что $E(y_t/x_{t1}, x_{t2}) = \alpha_1 x_{t1} + \alpha_2 x_{t2} + \alpha_3$. Ясно, что величины (y_t, x_{t1}, x_{t2}) можно считать независимыми по t .

В экономике большое число примеров рассмотренной схемы регрессии со случайной выборкой имеется в исследованиях пространственно-структурных зависимостей. Рассмотрим один из таких примеров. Допустим, изучается производительность труда в некоторой отрасли. Отрасль представляется совокупностью предприятий. Обозначим y — производительность труда; y имеет свое распределение (генеральная совокупность — показатели производительности труда всех предприятий). Имеется средняя производительность, разброс производительности труда и т. д. Далее, рассматриваются три случайные величины: x_1 — основные фонды (генеральная совокупность — множество значений основных фондов всех предприятий отрасли); x_2 — фондовооруженность (генеральная совокупность — множество значений фондовооруженности всех предприятий); x_3 — энерговооруженность (генеральная совокупность — множество значений энерговооруженности всех предприятий). Предположим, что при фиксированных x_1, x_2, x_3 ,

математическое ожидание y_t есть линейная функция x_1 , x_2 и x_3 , т. е.

$$E(y/x_1, x_2, x_3) = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4. \quad (3.21)$$

Свободный член введен в это уравнение регрессии по причинам, объясняемым в параграфе 1.3. Считаем, что условная дисперсия y (3.20) постоянна. Допустим, из множества всех предприятий отрасли (генеральная совокупность) отобрано определенное число предприятий, для каждого из них известны статистические данные по производительности труда (y), основным фондам (x_1), фондовооруженности (x_2) и энерговооруженности (x_3). Выборку считаем случайной. Наблюдения образуют вектор y и матрицу X , связанные между собой уравнением регрессии (3.1). Можно было бы рассмотреть исследуемую отрасль в целом, а в качестве статистики — показатели y , x_1 , x_2 , x_3 за ряд лет. Даже если наблюдения будут независимы, они уже не будут одинаково распределены! Производительность (средняя) наверняка будет повышаться (изменяется μ_0). Скорее всего будут повышаться средние и остальных показателей: x_1 , x_2 , x_3 — «временная» регрессия не подпадает под схему случайной выборки.

Свойства оценки МНК, доказанные в общем случае, будут верны и для схемы случайной выборки: несмещенность, неэффективность в классе несмещенных линейных оценок и эффективность в классе линейных оценок с ограниченной функцией риска.

В параграфе 1.6 отмечено, что в схеме классической регрессии (матрица X детерминирована) коэффициент детерминации не имеет статистического смысла. В схеме случайной выборки R^2 восстанавливает свое истинное статистическое содержание.

Обычный коэффициент детерминации

$$R^2 = 1 - \frac{\sum e_t^2}{\sum (y_t - \bar{y})^2} = 1 - \frac{\sum e_t^2/n}{\sum (y_t - \bar{y})^2/n} \quad (3.22)$$

теперь является естественной оценкой истинного значения (1.23): $\sum e_t^2/n$ является оценкой $\sigma^2(y/x)$, $\sum (y_t - \bar{y})^2/n$ — оценкой σ_y^2 . Коэффициент детерминации (3.22) можно трактовать как показатель «функциональности» зависимости y и x или показатель адекватности регрессии, R^2 есть доля объясняемой дисперсии y .

В схему случайной выборки естественно вписываются асимптотические свойства. В силу независимости и одинаковой распределенности переменных почти наверное (п.н.)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}' \mathbf{X} = \mathbf{C} + \mu\mu'. \quad (3.23)$$

Предел (3.23) легко доказывается с помощью усиленного закона больших чисел. Состоятельность оценки МНК можно доказать непосредственно, используя (3.23) или применив теорему 3.4.

Т е о р е м а 3.5. *Оценка МНК в схеме случайной выборки асимптотически нормальна.*

Д о к а з а т е л ь с т в о. Обозначим $\varepsilon = y - \mathbf{X}\alpha$, тогда

$$\sqrt{n}(\mathbf{a} - \alpha) = \sqrt{n}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \varepsilon = \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right)^{-1} \frac{\mathbf{X}' \varepsilon}{\sqrt{n}},$$

но из условия (3.23) следует п.н. $\left(\frac{\mathbf{X}'\mathbf{X}}{n} \right)^{-1} \rightarrow (\mathbf{C} + \mu\mu')^{-1}$.

Исследуем вектор $\mathbf{X}'\varepsilon/\sqrt{n}$. Его i -я компонента равна:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n x_{ii} \varepsilon_i = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i, \quad (3.24)$$

где $\xi_i = x_{ii}\varepsilon_i$. Но

$$E\xi_i = E_{x_{ii}} \{E(\xi_i/x_{ii})\} = E_{x_{ii}} \{x_{ii} E(\varepsilon_i/x_{ii})\} = 0;$$

$\{\xi_i\}$ — независимы, одинаково распределены. Найдем их дисперсию:

$$\begin{aligned} \sigma^2(\xi_i) &= E\xi_i^2 = E_{x_i} \{E(\xi_i^2/x_i)\} = E_{x_i} \{x_i^2 E(\varepsilon_i^2/x_i)\} = \\ &= E_{x_i} \{x_i^2 E(y - E(y/x))^2/x_i\} = \sigma^2 E x_i^2 = \sigma^2 C_{ii}. \end{aligned}$$

По центральной предельной теореме последовательность (3.24) асимптотически нормальна. Окончательно можно записать:

$$\sqrt{n}(\mathbf{a} - \alpha) \xrightarrow{\text{п.н.}} N(0, \sigma^2(\mathbf{C} + \mu\mu')^{-1}). \quad (3.25)$$

З а м е ч а н и е. Для доказательства асимптотической нормальности оценки МНК в схеме случайной выборки мы не налагаем условий на поведение матриц \mathbf{X}_n при $n \rightarrow \infty$. Асимптотические свойства оценки являются результатом предположений о независимости и одинаковой распределенности наблюдений $(y_t, \mathbf{x}_t) = (y_t, x_{t1}, \dots, x_{tm})$, $t = 1, \dots, n$.

Иногда в схеме регрессии предполагают нормальность случайных величин y и x . Нормальное распределение обладает одним хорошим свойством: регрессия y на x оказывается линейной. Напомним, что линейность регрессии в (3.19) в общем случае является достаточно строгим условием.

Как было отмечено, оценка МНК не является, вообще говоря, в схеме «математической регрессии» оценкой максимального правдоподобия. Однако, если наблюдения независимы и одинаково распределены по нормальному закону, оценки МНК и ММП совпадают.

Т е о р е м а 3.6. *В схеме случайной выборки из нормального распределения оценки МНК и ММП совпадают.*

Доказательство теоремы дано в параграфе 3.3.

В заключение отметим, что схема регрессии как условного математического ожидания является более общей, а поэтому и более сложной. Окончательные свойства тех или иных оценок или процедур зависят от распределения независимых переменных. Классическая регрессия является первым шагом схемы (3.1) — исследование линейной зависимости для фиксированных значений матрицы X . Второй шаг — исследование полученных свойств при флуктуациях X , которая изменяется в соответствии с законами распределения, задаваемыми априорно исследователем. Ясно, что дисперсия оценок, полученных при фиксированных X , будет меньше, чем при ее флуктуировании. Зато мы освобождаемся от конкретного вида X , и наши оценки будут более общими.

Часто трудно решить, является ли данная совокупность X случайной или детерминированной. Так, в примере с анализом производительности труда в отрасли матрицу данных X можно считать либо случайной, либо детерминированной в зависимости от поставленных целей.

В чем конкретно заключается разница между безусловной и условной регрессиями? В безусловной регрессии наши выводы касаются только данного, имеющегося в наличии набора независимых переменных, тогда как в условной регрессии полученные выводы и оценки имеют более общий характер; эти выводы могут быть распространены на всю генеральную совокупность независимых переменных.

У п р а ж н е н и я 3.3

1. Используя свойства условных математических ожиданий; покажите, что (3.19) влечет (3.1).

2. Докажите, что $|C| \neq 0$ влечет $\text{rang } X = m$ почти наверное.

3.4. Доказательства

1. Доказательство $Ez = E_x(E(z/x))$. Обозначим $f(x, y)$ — плотность распределения случайной величины (x, y) , тогда

$$Ez = Ez(x, y) = \iint z(x, y) f(x, y) dx dy. \quad (3.26)$$

Известно, что плотность распределения условной случайной величины $(x, y)/x$ есть $f(x, y)/f(x)$, поэтому

$$E(z(x, y)/x) = \int z(x, y) \frac{f(x, y)}{f(x)} dy.$$

Окончательно

$$\begin{aligned} E_x(E(z/x)) &= \int f(x) \left[\int z(x, y) \frac{f(x, y)}{f(x)} dy \right] dx = \\ &= \iint z(x, y) f(x, y) dx dy, \end{aligned}$$

что совпадает с (3.26). Формула $Ez = E_x(E(z/x))$ доказана.

2. Доказательство теоремы 3.4. Для доказательства теоремы достаточно показать, что (3.18) эквивалентно

$$E(\mathbf{X}'_n \mathbf{X}_n)^{-1} \rightarrow 0, \quad n \rightarrow \infty. \quad (3.27)$$

Достаточность. Пусть (3.18) имеет место. В силу неотрицательной определенности матрицы $(\mathbf{X}'_n \mathbf{X}_n)^{-1}$ для любого $i = 1, \dots, m$

$$(\mathbf{X}'_n \mathbf{X}_n)_{ii}^{-1} \leq \lambda_{\max}(\mathbf{X}'_n \mathbf{X}_n)^{-1} = 1/\lambda_{\min}(\mathbf{X}'_n \mathbf{X}_n) \rightarrow 0.$$

Нетрудно показать, что $(\mathbf{X}'_{n+1} \mathbf{X}_{n+1})_{ii}^{-1} \leq (\mathbf{X}'_n \mathbf{X}_n)_{ii}^{-1}$, $i = 1, \dots, m$. Применяя теорему об интегрировании монотонной последовательности, получаем $E(\mathbf{X}'_n \mathbf{X}_n)_{ii}^{-1} \rightarrow 0$, а значит, и $E(\mathbf{X}'_n \mathbf{X}_n)^{-1} \rightarrow 0$.

Необходимость. Пусть условие состоятельности (3.27) имеет место. Воспользуемся следующим фактом: пусть $Z_n \geq 0$ — случайные интегрируемые величины, причем $Z_{n+1} \leq Z_n$ и $EZ_n \rightarrow 0$, тогда $Z_n \rightarrow 0$ п. н. (Доказательство: пусть $Z = \lim Z_n$, положим $A = \{Z > 0\}$ и Z_n^A — сужение Z на A , тогда $EZ_n \geq EZ_n^A \rightarrow 0$, но $EZ_n^A \rightarrow EZ$: мера A равна нулю.) Поскольку последовательность $(\mathbf{X}'_n \mathbf{X}_n)_{ii}^{-1}$ невозрастающая, то $(\mathbf{X}'_n \mathbf{X}_n)_{ii}^{-1} \rightarrow 0$ п. н., значит, $\lambda_{\min}(\mathbf{X}'_n \mathbf{X}_n) \rightarrow \infty$.

3. Доказательство теоремы 3.6. Обозначим $\mathbf{z} = (y, x_1, \dots, x_m)$. По условию $\mathbf{z} \sim N(\mathbf{v}, \Omega)$, где $\mathbf{v} = (\mu_0, \mu_1, \dots, \mu_m)'$,

$$\Omega = \begin{bmatrix} \sigma_0^2 & \mathbf{p}' \\ \mathbf{p} & \mathbf{C} \end{bmatrix},$$

$\rho = \text{cov}(y, x)$. Регрессия y на x будет равна [5]:

$$E(y|x) = \mu_0 + \rho' C^{-1}(x - \mu) = \rho' C^{-1}x + \mu_0 - \rho' C^{-1}\mu.$$

Поскольку в регрессии свободный член отсутствует, то

$$\rho' C^{-1}\mu = \mu_0. \quad (3.28)$$

Пусть z_1, z_2, \dots, z_n — независимы, одинаково распределены $\in N(\nu, \Omega)$. Известно (см., например, [5]), что оценками ММП для μ_0, μ, C, ρ являются соответственно $\bar{y}, \bar{x}, \frac{1}{n} X'X - \bar{x}\bar{x}', \frac{1}{n} X'y - \bar{x}\bar{y}$. Далее воспользуемся следующим элементарным фактом: если $\hat{\theta}$ — оценка ММП параметра θ , то $g(\hat{\theta})$ является оценкой ММП для $g(\theta)$. Поэтому для доказательства равенства оценок МНК и ММП достаточно показать, что

$$a = (X'X)^{-1} X'y = \hat{C}^{-1} \hat{\rho}. \quad (3.29)$$

Формально перепишем (3.29).

$$(X'X - n\bar{x}\bar{x}')^{-1} (X'y - n\bar{y}\bar{x}) = (X'X)^{-1} X'y. \quad (3.30)$$

Умножим (3.30) слева на матрицу $X'X - n\bar{x}\bar{x}'$:

$$X'y - n\bar{y}\bar{x} = X'y - n\bar{x}\bar{x}'(X'X)^{-1}X'y,$$

или $\bar{y}\bar{x} = \bar{x}\bar{x}'(X'X)^{-1}X'y$. Но для оценки МНК имеет место равенство $\bar{y} = \hat{\bar{y}}$ (см. параграф 1.6), т. е. $\bar{y} = \frac{1}{n} 1'X(X'X)^{-1}X'y = \bar{x}(X'X)^{-1}X'y$. Делая обратные преобразования, приходим к (3.29) — теорема доказана.

Глава 4

ОШИБКИ В НЕЗАВИСИМЫХ ПЕРЕМЕННЫХ

4.1. Постановка задачи. Оценка МНК,

В этой главе обобщается классическая регрессия на случай, когда независимые переменные измеряются с ошибкой. В принципе случайные отклонения в независимых переменных можно трактовать шире, однако в любом случае эти отклонения должны удовлетворять определенным требованиям.

Важное место в классической регрессии (независимые переменные детерминированы) занимают регрессии плани-

руемого эксперимента. Основным требованием к ним является, как уже указывалось в начале книги, отсутствие ошибок измерения управляемых, т. е. независимых переменных. На практике они чаще всего все же существуют.

Система предположений в схеме с ошибками в независимых переменных такова. Существуют истинные значения переменных (они детерминированы) $y_t, x_{ti}, t=1, \dots, n; i=1, \dots, m$. Между этими *ненаблюдаемыми* переменными существует функциональная линейная связь

$$y_t = \alpha_1 x_{t1} + \alpha_2 x_{t2} + \dots + \alpha_m x_{tm}, \quad t = 1, \dots, n, \quad (4.1)$$

где $\alpha_1, \alpha_2, \dots, \alpha_m$ нам неизвестны и подлежат оцениванию. Истинные значения y_t и x_{ti} нам также не известны, зато известны наблюдения v_t, w_{ti} , которые отличаются от первых на случайные отклонения, т. е. на ошибки измерения:

$$v_t = y_t + \xi_t, \quad w_{ti} = x_{ti} + \varepsilon_{ti}, \quad i = 1, \dots, m; \quad t = 1, \dots, n, \quad (4.2)$$

Поскольку ξ_t, ε_{ti} мы интерпретируем как ошибки измерения, то естественны следующие гипотезы:

$$E\xi_t = E\varepsilon_{ti} = 0; \quad \sigma^2(\xi_t) = \sigma_1^2; \quad \sigma^2(\varepsilon_{ti}) = \sigma_{i+1}^2, \quad (4.3)$$

$\{\xi_t, \varepsilon_{ti}\}$ — система независимых случайных величин.

Первое уравнение в (4.3) означает, что измерения y_t и x_{ti} не содержат смещений. Вторая система предположений (4.3), касающаяся дисперсий, означает, что каждый ряд измеряется своим инструментом, чувствительность которого не зависит от номинальной величины ряда. Например, если y есть сила тока в цепи и эта величина измеряется амперметром с делением h , то, предполагая равномерное распределение ошибки на $[-h/2, h/2]$, мы можем утверждать, что $\sigma_1^2 = h^2/12$.

Вообще говоря, некоторые из переменных могут измеряться без ошибок, например если эти переменные — функции номера эксперимента или наблюдения; тогда $\sigma_i^2 = 0$. В частности, если $\sigma_1^2 > 0$, а $\sigma_i^2 = 0, i=2, \dots, m+1$, то приходим к классической регрессии $v_t = \alpha_1 x_{t1} + \dots + \alpha_m x_{tm} + \xi_t$. В дальнейшем, как правило, будем считать $\sigma_i^2 = 0$ для всех i .

Для иллюстрации продолжим рассмотрение примера, начатого в параграфе 1.1. Будем считать, что если все другие условия эксперимента фиксированы, а y_t, x_{t1}, x_{t2} и x_{t3} измерены без ошибок, то между ними наблюдается точная функциональная связь:

$$y_t = \alpha_1 x_{t1} + \alpha_2 x_{t2} + \alpha_3 x_{t3} + \alpha_4, \quad t = 1, \dots, 15.$$

Другими словами, нам известна формула реакции и известен закон взаимодействия веществ с точностью до неизвестных параметров $\alpha_1, \alpha_2, \alpha_3, \alpha_4$. Однако в силу ошибок измерения на место функциональной связи приходит статистическая. Считаем, что математическое ожидание ошибок измерения равно нулю. Если выход реакции (y_t), количество вещества B_1 (x_{t1}) и количество катализатора (x_{t3}) измеряются на одних весах, то можно считать $\sigma_1^2 = \sigma_2^2 = \sigma_4^2$. Приближенно можно считать даже $\sigma_1^2 = h^2/12$, где h — цена деления весов.

«Функциональность» уравнения (4.1) на первый взгляд может привести к мысли, что изложенная схема пригодна только для функциональных же зависимостей, т. е. зависимостей, встречающихся только в естественнонаучных дисциплинах. На самом деле это не так. Перепишем уравнение (4.1) в более привычном виде:

$$v = \alpha_1 x_{t1} + \alpha_2 x_{t2} + \dots + \alpha_m x_{tm} + \xi_t.$$

Оно очень похоже на уравнения регрессий, с которыми мы работали до сих пор. При этом ξ_t не обязательно есть ошибка измерения. Случайные величины ε_{ti} также не обязательно трактовать как ошибки измерения. Так, Э. Маленво приводит пример [48, с. 391, 392], где y_t — расход на данный продукт в семье t , x_{t1} — доход семьи t и случайные отклонения ξ_t и ε_{ti} трактуются не как ошибки измерения. Единственные условия, которым должны подчиняться ошибки, это условия (4.3) и условие независимости.

Схема (4.1)—(4.3), в которой y_t и x_{ti} детерминированы, называется функциональной. В другой схеме y_t и x_{ti} — случайные величины. Последняя схема называется *структурной*. Можно показать, что структурная схема в простейшем случае сводится к функциональной. Обозначим в структурной схеме $E x_{ti} = p_{ti}$, $x_{ti} - p_{ti} = \psi_{ti}$, $E y_t = q_t$, $y_t - q_t = \tau_t$.

Переходя к математическому ожиданию в (4.1), получим

$$q_t = \alpha_1 p_{t1} + \dots + \alpha_m p_{tm},$$

где $v_t = q_t + \delta_t$, $w_{ti} = p_{ti} + \theta_{ti}$, $\delta_t = \xi_t + \tau_t$, $\theta_{ti} = \varepsilon_{ti} + \psi_{ti}$. Если ψ_{ti} и τ_t независимы и гомоскедастичны, то таковыми будут и ξ_t и ε_{ti} . В литературе изучался простейший случай структурной зависимости $y_t = \alpha_1 x_t + \alpha_2$, где $x_t \sim N(\mu, \sigma_2^2)$, $\{x_t\}$ — независимы. В матричных обозначениях (4.1) и (4.2) переписываются следующим образом:

$$y = X\alpha; v = y + \xi; W = X + \varepsilon, \quad (4.4), (4.5), (4.6)$$

где y , X , α , v , ξ , W и ε имеют порядок соответственно $n \times 1$, $n \times m$, $m \times 1$, $n \times 1$, $n \times 1$, $n \times m$ и $n \times m$. Как и ранее, будем предполагать, что $\text{rang } X = m$. Подставляя (4.5) и (4.6) в уравнение (4.4) получим

$$v = W\alpha + (\xi - \varepsilon\alpha) = W\alpha + \eta, \quad (4.7)$$

где $\eta = \xi - \varepsilon\alpha$.

Схему (4.4) не надо путать со схемой условного математического ожидания, рассмотренной в предыдущей главе. Различие вытекает из того факта, что вектор отклонений η зависит от ε . Более подробно:

$$E(v/W) = W\alpha + E(\eta/W) = W\alpha + E(\xi - \varepsilon\alpha/W) = W\alpha + \\ + E(\xi/W) - E(\varepsilon/W)\alpha.$$

В силу независимости ξ и ε средний член в последнем равенстве обращается в нуль. Из равенства (4.6) окончательно следует

$$E(v/W) = W\alpha - \varepsilon\alpha = X\alpha \neq W\alpha,$$

что противоречит схеме регрессии как условного математического ожидания (3.1).

По наблюдениям W и v можно построить оценку МНК¹:

$$a_{\text{НК}} = (W'W)^{-1}W'v. \quad (4.8)$$

В классической регрессии оценка МНК была несмещенной, линейно эффективной в классе несмещенных оценок и при некоторых условиях состоятельной. Покажем, что в схеме с ошибками в независимых наблюдениях эти свойства оценки пропадают. Начнем с несмещенности. Строго доказать смещенность оценки (4.8) затруднительно. Относительно смещенности оценки МНК приведем следующие соображения. Перепишем (4.8) следующим образом:

$$a_{\text{НК}} = (W'W)^{-1}W'(W\alpha + \eta) = \alpha + (W'W)^{-1}W'\eta,$$

но

$$EW'\eta = E(X' + \varepsilon')(\xi - \varepsilon\alpha) = E(X'\xi - X'\varepsilon\alpha + \varepsilon'\xi - \\ - \varepsilon'\varepsilon\alpha) = EX'\xi - EX'\varepsilon\alpha + E\varepsilon'\xi - E\varepsilon'\varepsilon\alpha = -nD\alpha,$$

где D — диагональная матрица $m \times m$ с i -м диагональным элементом $\sigma_{\varepsilon_i}^2$. Вероятно, что $E(W'W)^{-1}W'\eta$ также отлично от нуля.

¹В этой главе мы рассмотрим несколько оценок, каждую из которых снабжаем своим индексом.

Оценка МНК не будет состоятельной даже при весьма сильных предположениях. Покажем, что если матрица X сильно регулярна, оценка МНК не будет состоятельна. Распишем (4.8) следующим образом:

$$a_{\text{НК}} = \alpha + (W' W)^{-1} W' \eta = \alpha + \left(\frac{W' W}{n} \right)^{-1} \left(\frac{W' \eta}{n} \right).$$

Рассмотрим первый предел по вероятности:

$$\begin{aligned} \text{plim } \frac{1}{n} W' W &= \text{plim } \frac{1}{n} (X + \varepsilon)' (X + \varepsilon) = \text{plim } \frac{1}{n} X' X + \\ &+ \text{plim } \frac{1}{n} X' \varepsilon + \text{plim } \frac{1}{n} \varepsilon' X + \text{plim } \frac{1}{n} \varepsilon' \varepsilon. \end{aligned}$$

По предположению

$$\text{plim } \frac{1}{n} X' X = A, \quad |A| \neq 0. \quad (4.9)$$

Легко показать, что в условиях сильной регулярности из закона больших чисел следует

$$\text{plim } \frac{1}{n} X' \varepsilon = \text{plim } \frac{1}{n} \varepsilon' X = 0; \quad (4.10)$$

$$\text{plim } \frac{1}{n} \varepsilon' \varepsilon = D, \quad (4.11)$$

где D — диагональная матрица, описанная ранее. Таким образом,

$$\text{plim } \frac{1}{n} W' W = A + D.$$

Второй предел по вероятности

$$\begin{aligned} \text{plim } \frac{1}{n} W' \eta &= \text{plim } \frac{1}{n} (X + \varepsilon)' (\xi - \varepsilon \alpha) = \text{plim } \frac{1}{n} X' \xi - \\ &- \text{plim } \frac{1}{n} X' \varepsilon \alpha + \text{plim } \frac{1}{n} \varepsilon' \xi - \text{plim } \frac{1}{n} \varepsilon' \varepsilon \alpha. \end{aligned}$$

Из закона больших чисел следует, что первые три слагаемых в последнем выражении равны нулю, таким образом

$$\text{plim } \frac{1}{n} W' \eta = -D\alpha,$$

окончательно

$$\text{plim } a_{\text{НК}} = \alpha - (A + D)^{-1} D\alpha. \quad (4.12)$$

Итак, систематическое смещение оценки МНК не равно нулю, если $\alpha \neq 0$. Как следует из формулы (4.12), если $\mathbf{D} = \mathbf{O}$, т. е. ошибки в \mathbf{X} отсутствуют, оценка МНК будет состоятельной.

В [177] подробно исследована оценка МНК для зависимости $y_t = \alpha_1 x_t + \alpha_2$ при конечном n . Обозначим $\tau = \frac{1}{n\sigma_1^2} \sum_t (x_t - \bar{x}_t)^2$.

Авторы показывают, что относительное смещение $(E a_{\text{НК}} - \alpha)/\alpha$ есть убывающая функция τ , более того, $(E a_{\text{НК}} - \alpha)/\alpha \rightarrow 0$ при $\tau \rightarrow \infty$, $(E a_{\text{НК}} - \alpha)/\alpha \rightarrow 1$ при $\tau \rightarrow 0$ для любого конечного n . Если y_t ограничены, то для больших n

$$(E a_{\text{НК}} - \alpha)/\alpha = -\frac{1}{1+\tau} \left[1 - \frac{2}{n} \cdot \frac{\tau^2}{(1+\tau)^2} - O(n^{-2}) \right],$$

$$\sigma^2(a_{\text{НК}}) = \frac{1}{n-2} \left[\frac{\sigma_1^2/\sigma_2^2}{1+\tau} + \frac{\beta^2 \tau (1+\tau^2)}{(1+\tau)^4} \right] + o(n^{-2}).$$

В частности, при $\tau \rightarrow \infty$ $a_n \rightarrow \alpha$. Там же приводится точное распределение $a_{\text{НК}}$ для нормальных ξ_t и ε_t .

У п р а ж н е н и я 4. 1

1. Допустим, все предположения классической регрессии выполнены. Изменится ли схема классической регрессии в случае ошибок измерения только у?

2. В каком случае оценка (4.8) п. н. существует? Начните со случая $m = 1$.

3. Попробуйте доказать смещенность оценки (4.8) для случая $m = 1$.

4. Будет ли оценка МНК состоятельной при других предположениях о поведении матриц \mathbf{X}_n при $n \rightarrow \infty$. Начните со случая $m = 1$.

4.2. Ортогональная регрессия

Естественно в схеме (4.4)—(4.6) искать другие методы оценивания параметра α , отличные от МНК. Сначала рассмотрим простейший случай — регрессию $y_t = \alpha_1 x_t + \alpha_2$ (формально можно считать переменной при α_2 $x_{t2} \equiv 1$, тогда $\sigma_2^2 = 0$). Оценка МНК параметров α_1 и α_2 получается в результате минимизации суммы квадратов отклонений $\sum \varepsilon_t^2$, $\varepsilon_t = y_t - \alpha_1 x_t - \alpha_2$, причем отклонения рассматриваются вдоль оси y , так как именно y может нести на себе ошибки измерения. В случае же, когда и x имеет ошибки измерения, выбор направления, параллельного оси y ,

становится спорным. Теперь и y , и x попадают в одинаковую ситуацию. Так, можно минимизировать отклонение от прямой вдоль оси x (рис. 4.1). На этом рисунке AB соответствует отклонению вдоль оси x (ошибки только в независимой переменной), AD — отклонению вдоль оси y (ошибки только в зависимой переменной). Компромиссом здесь может быть, например, отклонение, равное расстоянию от точки, отвечающей выборке y , до прямой. Регрессия, оцененная

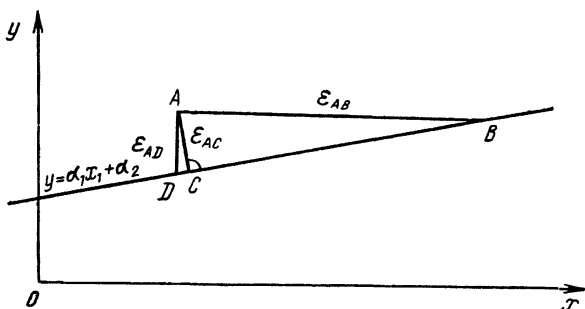


Рис. 4.1. Минимизация отклонений для разных направлений

минимизацией суммы квадратов расстояний от точек выборки до прямой, или в общем случае плоскости, называется ортогональной¹.

Перейдем к общему случаю. Поскольку y и x_t в схеме ошибок в независимых переменных становятся с теоретической точки зрения равноправными, следующие переобозначения являются целесообразными. Объединим все наблюдения в одну матрицу, т. е. обозначим

$$Z = [y, X], \quad V = [v, W], \quad \Theta = [\xi, \varepsilon], \quad (4.13)$$

$$\beta = (\beta_1, \dots, \beta_k)' = (-1, \alpha_1, \dots, \alpha_m)', \quad (4.14)$$

где $k = m + 1$. Соотношения (4.4), (4.5) и (4.6) переписываются следующим образом:

$$Z\beta = 0; \quad V = Z + \Theta, \quad (4.15) \quad (4.16)$$

причем $E\theta_{ii} = 0$, $\sigma^2(\theta_{ii}) = E\theta_{ii}^2 = \sigma_i^2$, $i = 1, 2, \dots, k = m + 1$. При нахождении оценок ортогональной регрес-

¹Часто регрессию, оцененную таким методом, называют взвешенной.

сии удобнее пользоваться другой нормировкой вектора β , а именно положим $\|\beta\| = 1$, т. е.

$$\sum_{i=1}^k \beta_i^2 = 1. \quad (4.17)$$

Можно было бы оставить условие $\beta_1 = -1$, но для простоты исследования (4.17) более целесообразно. Для отыскания оценок ортогональной регрессии нам понадобится следующая лемма.

Лемма 4.1. Пусть в R^k задана гиперплоскость (линейное подпространство размерности $k - 1$):

$$\Pi = \{r \in R^k : \beta_1 r_1 + \dots + \beta_k r_k = \beta' r = 0\}, \quad (4.18)$$

где вектор β фиксирован, $\|\beta\| = 1$. Тогда расстояние от произвольной точки $s \in R^k$ до плоскости Π есть

$$\rho(s, \Pi) = \min_{r \in \Pi} \|s - r\| = |s' \beta|.$$

Доказательство леммы несложно, его можно провести с помощью множителей Лагранжа.

Принцип ортогональной регрессии заключается в минимизации суммарных квадратов расстояний от точек выборки, т. е. $v^t = (v_{t1}, v_{t2}, \dots, v_{tk})' \in R^k$, до гиперплоскости Π ортогональной регрессии (4.18). В силу леммы 4.1

$$\begin{aligned} S(\beta) &= \sum_{i=1}^n \rho^2(v^i, \Pi) = \sum_{i=1}^n (\beta' v^i)^2 = \\ &= \sum_{i=1}^n (\beta_1 v_{i1} + \dots + \beta_k v_{ik})^2. \end{aligned} \quad (4.19)$$

Минимизируем $S(\beta)$ при условии $\|\beta\| = 1$. Для этого построим функцию Лагранжа:

$$\Phi(\beta, \lambda) = \sum_i (\beta' v^i)^2 - \lambda (\beta' \beta - 1).$$

Ее производные по β_i равны:

$$\partial \Phi / \partial \beta_i = 2 \sum_t (\beta_1 v_{t1} + \dots + \beta_k v_{tk}) v_{ti} - 2\lambda \beta_i = 0,$$

или в матричном виде:

$$V' V \beta - \lambda \beta = 0.$$

Для того чтобы последнее уравнение имело решение относительно β , необходимо и достаточно, чтобы

$$|V' V - \lambda I_k| = 0;$$

при этом значение β , минимизирующее $S(\beta)$, является характеристическим вектором (х. в.) матрицы $V'V$, а λ — соответствующим характеристическим числом. Однако число разных характеристических векторов и характеристических чисел может быть равно k . Какое же из них выбрать? Покажем, что для минимизации $S(\beta)$ необходимо выбрать минимальное х. ч. матрицы $V'V$ и соответствующий минимальный х.в. Действительно,

$$V'V = \sum_t v^t (v^t)',$$

но

$$\begin{aligned} S(\beta) &= \sum_t (\beta' v^t)^2 = \sum_t \beta' v^t (v^t)' \beta = \beta' \sum_t v^t (v^t)' \beta = \\ &= \beta' VV\beta. \end{aligned} \quad (4.20)$$

Минимальное значение квадратичной формы будет наблюдаться, если в качестве β взять характеристический вектор, соответствующий минимальному характеристическому числу матрицы $V'V$. Обозначим его b_{OP} , тогда

$$\min_{\|\beta\|=1} S(\beta) = b'_{OP} V' V b_{OP} = \lambda_{\min}(V'V).$$

Возвращаясь к старым обозначениям, можно найти оценку ортогональной регрессии параметра α :

$$(a_{OP})_i = -(b_{OP})_{i+1} / (b_{OP})_1, \quad i = 1, \dots, m.$$

Рассмотрим геометрический смысл ортогональной регрессии. Для простоты остановимся на случае $m = 1$, т. е.

$$z_{t1} \beta_1 + z_{t2} \beta_2 = 0, \quad t = 1, \dots, n.$$

Матрица $V'V$ задает характеристический эллипсоид

$$E_\gamma = \{r \in R^k : r' V' V r = \gamma\}, \quad \gamma > 0.$$

Величина γ определяет размеры характеристического эллипсоида. Каждый эллипсоид из семейства E_γ является пропорциональным растяжением другого. В случае $m = 1$, т. е. $k = 2$, эллипсоид превращается в эллипс. Ортогональная регрессия имеет направление характеристического вектора E_γ , отвечающего его максимальному характеристическому числу. Диаметром эллипса, сопряженным данному направлению p , как известно, называется геометрическое место точек середин отрезков параллельных p , отсекаемых

эллипсом. Очевидно, ортогональная регрессия есть диаметр эллипса, сопряженный направлению x . в., отвечающего минимальному x . ч. матрицы $V'V$ (рис. 4.2). Регрессия оценки МНК есть диаметр характеристического эллипса E_y , который сопряжен с направлением оси y . Регрессия, соответствующая оценке МНК « x на y », есть диаметр, сопряженный направлению оси x .

Длина отрезка $OP = \lambda_{\min}$, длина отрезка $OT = \lambda_{\max}$. Вектор \overline{OP} соответствует характеристическому вектору s

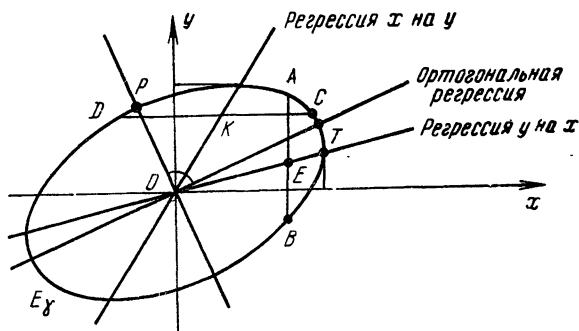


Рис. 4.2. Ортогональная регрессия и регрессии, сопряженные данному направлению для $m=1$

λ_{\min} , вектор $\overline{OT} = \lambda_{\max}$. Регрессия y на x делит отрезок AB в точке E пополам, регрессия x на y делит отрезок CD в точке K пополам, угол POC — прямой.

В качестве примера рассмотрим регрессию (1.5). Допустим, данные табл. 1.1 есть результаты измерений величин $y_t, x_{t1}, x_{t2}, x_{t3}$. Считаем, что в отсутствии ошибок измерения количество вещества B_1 , получившегося в результате реакции, есть линейная функция количества вещества B_2 , температуры и количества катализатора, т. е.

$$y_t = \alpha_1 x_{t1} + \alpha_2 x_{t2} + \alpha_3 x_{t3} + \alpha_4. \quad (4.21)$$

Ошибки измерения делают эту зависимость стохастической. Перепишем (4.21) в виде (4.15), т. е.

$$\beta_1 z_{t1} + \beta_2 z_{t2} + \beta_3 z_{t3} + \beta_4 z_{t4} = 0. \quad (4.22)$$

Для того чтобы избавиться от постоянного члена, перейдем к центрированным рядам, т. е. вместо v_{ti} будем рассматри-

вать $v_{ti} - \bar{v}_i$, где $v_i = \sum_t v_{ti}/n$. Матрица $V'V$ в регрессии (1.5) равна:

$$\begin{bmatrix} 10459,0 & 19341,0 & 2269,9 & 597,3 \\ 19341,0 & 36367,2 & 4165,8 & 1076,9 \\ 2269,9 & 4165,8 & 533,7 & 134,4 \\ 597,3 & 1076,9 & 134,4 & 37,78 \end{bmatrix}.$$

Ниже приведены характеристические векторы и числа этой матрицы в порядке их убывания:

$$(0,4684; 0,8773; 0,1012; 0,0263) \quad \lambda_4 = 47206$$

$$(0,7889; -0,4683; 0,3710; 0,1439) \quad \lambda_3 = 154,2$$

$$(-0,3808; 0,0964; 0,9190; 0,0319) \quad \lambda_2 = 34,7$$

$$(-0,1150; 0,04173; -0,0864; 0,9887) \quad \lambda_1 = 2,06$$

Выбираем последний вектор, отвечающий $\lambda_{\min}(V'V) = 2,06$:

$$b_{OP} = (-0,115; 0,04173; -0,0864; 0,989)'$$

Оценка α ортогональной регрессии равна:

$$(a_{OP})_1 = -(b_{OP})_2 / (b_{OP})_1 = 0,363$$

$$(a_{OP})_2 = -(b_{OP})_3 / (b_{OP})_1 = -0,751$$

$$(a_{OP})_3 = -(b_{OP})_4 / (b_{OP})_1 = 8,600$$

$$(a_{OP})_4 = \bar{v}_1 - (a_{OP})_1 \bar{v}_2 - (a_{OP})_2 \bar{v}_3 - (a_{OP})_3 \bar{v}_4 = 3,034$$

Итак, зависимость (4.21), оцененная методом ортогональной регрессии, равна;

$$y_t = 0,363 x_{t1} - 0,751 x_{t2} + 8,600 x_{t3} + 3,034, \quad t=1,$$

..., n .

Универсальным методом оценивания параметров является метод максимального правдоподобия. К каким оценкам приведет этот метод, если применить его к задаче (4.15), (4.16)? Применение ММП еще отчетливее позволит понять трудности оценивания в модели с ошибками в независимых переменных.

¹Целесообразность процедуры центрирования объяснена в следующем параграфе.

1. Найдите оценку ортогональной регрессии для случая $y_t = \alpha x_t$, $y_t = \alpha_1 x_t + \alpha_2$ и оценку регрессии x на y в этих случаях.

2. Найдите оценку регрессии, когда направление минимизирующих ошибок p произвольно. Сравните ее с оценкой МНК, с оценкой ортогональной регрессии, с оценкой регрессии x на y . Дайте геометрическую интерпретацию такой оценки.

4.3. Метод максимального правдоподобия

Предположим, случайные отклонения θ_{ti} имеют нормальное распределение, другими словами, $\theta_{ti} \sim N(0, \sigma_i^2)$. Тогда v_{ti} также нормально распределены, причем $v_{ti} \sim N(z_{ti}, \sigma_i^2)$. Параметрами, которые необходимо оценить в первую очередь, являются $\alpha_1, \alpha_2, \dots, \alpha_m$, так называемые «структурные параметры». Остальные параметры мешающие: $\sigma_1^2, \dots, \sigma_k^2, z_{ti}$. Таким образом, всего необходимо оценить $2k - 1 + nk$ неизвестных параметров, тогда как число наблюдений равно nk . Ясно, что все параметры удовлетворительно оценить невозможно. Для нас основной интерес представляют структурные параметры $\alpha_1, \dots, \alpha_m$. Однако и для них далеко не всегда существует оценка ММП. Можно показать, что она существует, только если σ_i^2 известны по крайней мере с точностью до постоянного множителя [37]. Поэтому предположим, что известен вектор весов $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_k)$, причем $\sigma_i^2 = \sigma^2 \Phi_i^2$, где σ^2 — неизвестный параметр. Нормализуем переменные v_{ti}, z_{ti}, β_i , т. е. вместо них рассмотрим

$$v_{ti}/\Phi_i, z_{ti}/\Phi_i, \beta_i \Phi_i, i = 1, \dots, k; t = 1, \dots, n. \quad (4.23)$$

Нормализованные переменные по-прежнему удовлетворяют исходным уравнениям (4.15) и (4.16), причем $\sigma^2 (v_{ti}) = \sigma^2 (\theta_{ti}) = \sigma^2$. Поскольку v_{ti} независимы, то их совместная плотность равна:

$$p(V; \beta, \sigma^2, Z) = (2\pi)^{-\frac{kn}{2}} (\sigma^2)^{-\frac{kn}{2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{t,i} (v_{ti} - z_{ti})^2 \right].$$

Возьмем логарифм этой функции, отбросим постоянные члены и поменяем знак, получим

$$p_1(V; \beta, \sigma^2, Z) = nk \ln \sigma^2 + \frac{1}{\sigma^2} \sum_{t=1}^n \sum_{i=1}^k (v_{ti} - z_{ti})^2 \rightarrow \min. \quad (4.24)$$

Необходимым условием экстремума функции является обращение ее частных производных в точке экстремума в нуль. Таким образом,

$$\frac{\partial p_1}{\partial \sigma^2} = \frac{f nk}{\sigma^2} - \frac{1}{\sigma^4} \sum_{t,i} (v_{ti} - z_{ti})^2 = 0,$$

откуда

$$\hat{\sigma}^2 = \frac{1}{nk} \sum_{t,i} (v_{ti} - z_{ti})^2.$$

Подставим это значение в минимизирующую функцию (4.24), получим

$$p_2(\mathbf{V}; \boldsymbol{\beta}, \mathbf{Z}) = \sum_{t,i} (v_{ti} - z_{ti})^2 = \sum_t \|\mathbf{v}^t - \mathbf{z}^t\|^2,$$

где $\mathbf{v}^t = (v_{t1}, \dots, v_{tk})'$, $\mathbf{z}^t = (z_{t1}, \dots, z_{tk})'$ — t -е строки матриц \mathbf{V} и \mathbf{Z} соответственно. По определению z_{ti} удовлетворяют уравнению

$$\boldsymbol{\beta}' \mathbf{z}^t = \beta_1 z_{t1} + \dots + \beta_k z_{tk} = 0,$$

т. е. \mathbf{z}^t принадлежат гиперплоскости Π (4.18). Величина $\|\mathbf{z}^t - \mathbf{v}^t\|^2$ есть квадрат расстояния от точки $\mathbf{v}^t \in R^k$ до точки $\mathbf{z}^t \in \Pi$. Это расстояние по условию необходимо минимизировать. Минимальная величина $\|\mathbf{v}^t - \mathbf{z}^t\|$ есть $\rho(\mathbf{v}^t, \Pi)$, по лемме 4.1 она равна $|\boldsymbol{\beta}' \mathbf{v}^t|$. Таким образом, получаем

$$p_3(\mathbf{V}; \boldsymbol{\beta}) = \sum_t (\boldsymbol{\beta}' \mathbf{v}^t)^2 \Rightarrow \min. \quad (4.25)$$

Как видим, p_3 совпадает с S (4.19). Поэтому оценка ММП для стандартизованной зависимости (4.23) совпадает с ортогональной регрессией. Оценка ММП является характеристическим вектором стандартизованной матрицы $\mathbf{V}'\mathbf{V}$, отвечающим минимальному х.ч. этой матрицы. Если $\varphi_i = 1$, то оценка ортогональной регрессии $\mathbf{b}_{ор}$ и оценка ММП $\mathbf{b}_{МП}$ совпадают. Для получения оценок исходной зависимости необходимо произвести операции, обратные (4.23), т. е. положить $(\mathbf{b}_{МП})_i / \varphi_i$, $i = 1, 2, \dots, k$. При желании новый вектор $\mathbf{b}_{МП}$ может быть пронормирован так, чтобы $\|\mathbf{b}_{МП}\| = 1$ или $(\mathbf{b}_{МП})_1 = 1$.

Если переменные имеют разные дисперсии, то оценки ММП и ортогональной регрессии будут отличаться друг от друга. Другими словами, выбирая направление минимального х.ч. матрицы $\mathbf{V}'\mathbf{V}$ в ортогональной регрессии, мы тем

самым неявно предполагаем, что переменные (y, x_1, x_2, \dots, x_m) имеют одинаковые ошибки (имеются в виду стандартные отклонения). Если ошибки будут разными, направление должно быть другим, учитывающим эту разницу.

Оценки ММП в схеме с ошибками в независимых переменных более гибкие, чем оценки ортогональной регрессии. Действительно, представим ситуацию, когда ошибки в x незначительны по сравнению с ошибками в y . Тогда направление минимизации ошибок должно быть близко к y , в то время как в ортогональной регрессии оно не зависит от величин ошибок в y и x . С учетом того, что $\sigma_1^2 \gg \sigma_2^2$, направление в ММП, как нетрудно проверить, будет близко к направлению y .

Иногда помимо переменных, в которых присутствуют ошибки, в зависимость (4.1) входят переменные, которые измеряются без ошибок. В первую очередь это относится к уравнениям со свободным членом. Нахождение оценок ММП в этом случае несложно. Рассмотрим подход, предложенный в [90].

Итак, пусть

$$y_t = \alpha_1 x_{t1} + \dots + \alpha_m x_{tm} + \gamma_1 s_{t1} + \dots + \gamma_p s_{tp}, \quad (4.26)$$

где относительно y_t, x_{ti} выполняются все предположения, сделанные ранее, а s_{tj} детерминированы, $\text{rank } \mathbf{S} = p$. Обозначим $\mathbf{M} = \mathbf{I}_n - \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$. Найдем х.в. матрицы $\mathbf{W}'\mathbf{M}\mathbf{W}$; характеристический вектор, соответствующий ее минимальному х. ч., обозначим δ . Тогда, как и ранее, $a_i = -\delta_{i+1}/\delta_i, i = 1, \dots, m$. Оценка для γ равна:

$$\mathbf{g} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'(\mathbf{y} - \mathbf{W}\mathbf{a}).$$

На основе этого результата легко показать, что если в уравнении (4.26) $p = 1$ и $s_{t1} \equiv 1$, т. е. γ_1 — свободный член, то для нахождения вектора α необходимо сначала центрировать матрицу \mathbf{W} , т. е. положить $\bar{w}_{ti} = w_{ti} - \bar{w}_i$, где $\bar{w}_i = \frac{1}{n} \sum_t w_{ti}$; оценка \mathbf{g}_1 равна $\bar{y} - a_1 \bar{x}_1 - \dots - a_m \bar{x}_m$.

Найдем оценки метода максимального правдоподобия для регрессии-примера. В табл. 4.1 приведены 9 вариантов стандартных отклонений ($\sigma_1, \sigma_2, \sigma_3, \sigma_4$) соответственно для переменных y_t, x_{t1}, x_{t2} и x_{t3} . Выбор значений объясняется в параграфе 4.7. Варианты $S_1 - S_5$ соответствуют равным дисперсиям, поэтому оценка ММП в этом случае совпадает с оценкой ортогональной регрессии. В табл. 4.2 при-

Таблица 4.1

	σ_1	σ_2	σ_3	σ_4
S_1	0,00289	0,00289	0,00289	0,00289
S_2	0,0289	0,0289	0,0289	0,0289
S_3	0,289	0,289	0,289	0,289
S_4	2,89	2,89	2,89	2,89
S_5	28,9	28,9	28,9	28,9
S_6	2,89	0,289	0,289	0,289
S_7	0,289	2,89	0,289	0,289
S_8	0,289	0,289	2,89	0,289
S_9	0,289	0,289	0,289	2,89

ведены оценки ММП для остальных вариантов $S_6 - S_9$. Для сравнения в нижней строке приведены оценки ортогональной регрессии.

Проанализируем полученные оценки. Как видим, оценка ММП во всех вариантах, за исключением S_8 , для первого параметра α_1 относительно стабильна. Максимальное отклонение оценок ММП от оценок ортогональной регрессии и вообще от всей массы оценок наблюдаются для варианта S_8 . Другими словами, оценки ММП наиболее чувствительны к ошибкам измерения в x_2 . Оценки второго и четвертого параметров резко отличаются друг от друга в зависимости от дисперсий, поэтому являются наиболее нестабильными и резко реагирующими на априорные значения стандартных отклонений $\sigma_1, \sigma_2, \sigma_3, \sigma_4$. Оценки третьего параметра относительно устойчивы (хотя и не в такой степени, как оценка α_1), опять же за исключением варианта S_8 .

Остановимся несколько подробнее на свойствах оценки ортогональной регрессии и оценки ММП. Во-первых, можно

Таблица 4.2

	α_1	α_2	α_3	α_4
S_6	0,3895	0,0925	4,449	-0,5618
S_7	0,393	-0,484	6,640	1,878
S_8	-0,259	11,92	-19,4	-52,48
S_9	0,3594	-0,7384	8,680	2,957
$S_1 - S_5$	0,363	-0,753	8,598	3,034

показать, что оценка b_{OP} будет смещена, даже если $\varphi_i = 1$, а θ_{ti} нормально распределены. Будет ли оценка b_{OP} обладать самым необходимым и слабым свойством — свойством состоятельности? Ответ положительный.

Теорема 4.1. Если $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$, а матрицы Z_n сильно регулярны, т. е. $\lim_{n \rightarrow \infty} \frac{1}{n} Z_n' Z_n = A$, где A — детерминированная матрица, $\text{rank } A = m$, то оценка b_{OP} сильно состоятельна, т. е.

$$\lim_{n \rightarrow \infty} b_{OP} = \beta \text{ с вероятностью } 1.$$

Доказательство теоремы дано в параграфе 4.8.

З а м е ч а н и я: 1. Поскольку b_{OP} — ограниченные случайные величины ($\|b_{OP}\| = 1$), то $E b_{OP} \rightarrow \beta$, $E \|b_{OP} - \beta\|^2 \rightarrow 0$, т. е. оценки b_{OP} являются асимптотически несмещенными и состоятельными в среднем квадратичном.

2. Если $\beta_1 \neq 0$, то сильно состоятельными являются также оценки a_{OP} .

3. Существенным условием теоремы является равенство дисперсий σ_i^2 . В противном случае оценки b_{OP} и a_{OP} не являются состоятельными.

4. В теореме не делается предположений о конкретном виде распределения θ_{ti} . Поэтому оценки ММП (при известных φ_i) также будут состоятельны.

В условиях регулярности матриц Z_n может быть исследовано асимптотическое распределение $\lambda_{\min}(V'V)$ (см., например, [48, с. 382]).

Можно показать, что оценкой ММП для σ^2 является $\lambda_{\min}(V'V)/nk$, где V — нормализованная матрица. Используя асимптотическую нормальность оценки ММП, Э. Маленко доказывает, что матрица асимптотических ковариаций C оценки $\sqrt{n}(b_{MP} - \beta)$ определяется из системы уравнений

$$\frac{1}{n^2} Z' Z C Z' Z = \frac{1}{n} Z' Z + \sigma^2 I_k - \sigma^4 \beta \beta', \quad C \beta = 0, \quad (4.27)$$

где $E \theta_{ti}^2 = \sigma_i^2$. Непосредственно использовать (4.27) невозможно, так как C участвует в этом выражении неявно. В параграфе 4.8 показано, как матрица C может быть выражена из (4.27) явно. Там же даются приближенные формулы для нахождения дисперсий оценок a_{OP} .

В [72] подробно изучена оценка ММП для регрессии $y_t = \alpha_1 x_t + \alpha_2$ для конечного n , в частности построена приближенная функция распределения этой оценки.

У п р а ж н е н и я 4.3

1. Найдите оценку ММП для α в регрессии $y_t = \alpha x_t$ и $y_t = \alpha_1 x_t + \alpha_2$; $\sigma_i^2 = \sigma^2 \varphi_i$, $i = 1, 2$.
2. Найдите оценку ММП величин z_{ti} .
3. Найдите оценку ММП для σ^2 .
4. Примените ММП для случая, когда σ_i^2 полностью известны.
5. Как направление минимизации отклонений зависит от φ_i ?
6. Докажите, что для нахождения оценки ММП в регрессии со свободным членом надо сначала центрировать матрицу наблюдений.
7. Докажите, что ранг матрицы в теореме 4.1 не более m .

4.4. Метод группировки

Этот метод был впервые предложен А. Вальдом для случая парной регрессии [193]. Суть его заключается в том, что наблюдения разбиваются на две группы, а оценкой регрессии является прямая линия, проходящая через центры групп. Итак, допустим

$$\begin{aligned} v_t &= y_t + \xi_t; & \omega_t &= x_t + \varepsilon_t; \\ y_t &= \alpha_1 x_t + \alpha_2, & t &= 1, \dots, n. \end{aligned} \quad (4.28)$$

Схема (4.28) соответствует общей схеме (4.1) для $m = 2$ со свободным членом α_2 . Разобьем n пар наблюдений (ω_t, v_t) на две группы G_1 и G_2 ; число элементов группы G_1 обозначим через n_1 , G_2 — через n_2 , $n_1 + n_2 = n$. Найдём центры этих групп:

$$\begin{aligned} c_1^1 &= \frac{1}{n_1} \sum_{t \in G_1} \omega_t; & c_2^1 &= \frac{1}{n_1} \sum_{t \in G_1} v_t; \\ c_2^1 &= \frac{1}{n_2} \sum_{t \in G_2} \omega_t; & c_2^2 &= \frac{1}{n_2} \sum_{t \in G_2} v_t. \end{aligned}$$

Соединим центры прямой; угловой коэффициент этой прямой a_1^1 берем в качестве оценки метода группировки (МГ) α_1 ; свободный коэффициент прямой a_2^1 — оценка МГ параметра α_2 ¹ (рис. 4.3). Очевидно, $a_1^1 = (c_1^2 - c_2^2)/(c_1^1 - c_2^1)$. В дальнейшем будем интересоваться только оценкой угло-

¹В задаче 7 упражнения 1.5 необходимо было показать, что в классической регрессии подобный метод приводит к несмещенным и состоятельным оценкам, но менее эффективным, чем оценки МНК.

вого наклона α_1^Γ . В [193] доказано, что если группировка не зависит от ошибок ε_t и

$$\liminf_{n \rightarrow \infty} |\bar{x}_2 - \bar{x}_1| > 0, \quad (4.29)$$

где

$$\bar{x}_2 = \frac{1}{n_2} \sum_{t \in G_2} x_t, \quad \bar{x}_1 = \frac{1}{n_1} \sum_{t \in G_1} x_t,$$

то оценка α_1^Γ состоятельна. Доказательство весьма простое. Ниже мы его проведем для более общего случая. А. Вальд построил доверительные интервалы для параметра α_1 [193].

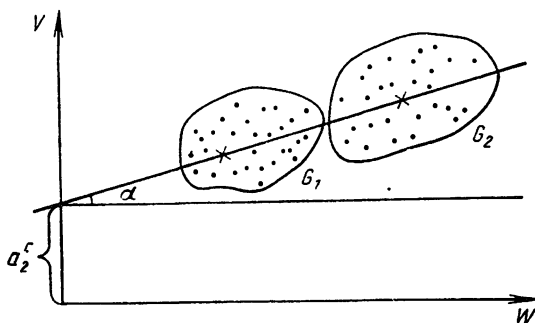


Рис. 4.3. Оценка Вальда, $\alpha_1^\Gamma = \operatorname{tg} \alpha$

М. Бартлетт, например, предложил вместо двух групп рассмотреть три [81]. Он разбивает n пар наблюдений (w_t, v_t) на три группы: G_1, G_2, G_3 с соответственным числом элементов n_1, n_2, n_3 ; $n_1 + n_2 + n_3 = n$. В качестве оценки α_1 Бартлетт рассмотрел

$$b = \frac{\sum_{t \in G_3} v_t/n_3 - \sum_{t \in G_1} v_t/n_1}{\sum_{t \in G_3} w_t/n_3 - \sum_{t \in G_1} w_t/n_1}. \quad (4.30)$$

Обозначим числитель (4.30) через b_2 , знаменатель — b_1 , аналогично

$$d_1 = \sum_{G_3} x_t/n_3 - \sum_{G_1} x_t/n_1; \quad d_2 = \sum_{G_3} y_t/n_3 - \sum_{G_1} y_t/n_1;$$

легко проверить, что $\alpha_1 = d_2/d_1$. Докажем, что при условии (4.29) оценка b состоятельна. Для этого покажем, что

$\text{plim } b_2 = d_2$, $\text{plim } b_1 = d_1$. Поскольку ε_t в разных группах независимы, то

$$E(b_2 - d_2)^2 = \sigma_1^2 (1/n_3 + 1/n_1) \rightarrow 0;$$

$$E(b_1 - d_1)^2 = \sigma^2 (1/n_3 + 1/n_1) \rightarrow 0,$$

если $\min(n_1, n_3) \rightarrow \infty$ при $n \rightarrow \infty$. Далее, $\text{plim } b = = \text{plim } b_2 / \text{plim } b_1 = d_2 / d_1 = \alpha_1$, если предел в знаменателе не равен нулю, т. е. если

$$\liminf \left| \sum_{G_3} x_t / n_3 - \sum_{G_1} x_t / n_1 \right| > 0, \quad (4.31)$$

что совпадает с условием (4.29) для двух групп. Условие (4.31) означает, что средние x для разных групп должны быть *асимптотически различимы*. Интуитивно понятно, что в противном случае информация основывалась бы только на случайных ошибках ε_t и ожидать состоятельности b трудно.

Условия (4.31) и независимость ε_t при разбиениях на группы—довольно жесткие условия. Например, разбиение на группы не может быть случайным: условие (4.31) при этом нарушится. Формально нельзя воспользоваться разбиением на группы после ранжирования w_t , так как при этом ε_t будут зависимыми¹.

Оптимальному выбору групп разбиения посвящен ряд работ [110, 190, 155]. Авторы их сходятся на том, что оптимальным разбиением является разбиение на три равные группы. Приведем табл. 4.3 из [155].

Способ группировки, основанный на ранжировании, предложен в [166], где рассмотрена структурная схема, т. е. x_t и y_t случайны, причем ε_t и x_t одинаково распределены (ε_t имеет функцию распределения F_1 , x_t — F_2). Итак, пусть w_t ранжированы; авторы предлагают две оценки для α_1 . Первый метод: допустим, известны такие два числа A и B , что $P\{w_t \leq A\} > 0$, $P\{w_t > B\} > 0$. Определим две группы следующим образом:

$$G_1 = \{(\omega_t, v_t) : \omega_t \leq A\}, G_3 = \{(\omega_t, v_t) : \omega_t > B\}.$$

Оценка b (4.30) при таком разбиении приводит нас к оценке f_1 . Второй метод: пусть заданы два положительных числа

¹Можно проверить, что $\varepsilon_{(1)} = \min_t \varepsilon_t$, $\varepsilon_{(n)} = \max_t \varepsilon_t$ будут зависимыми при независимых $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$.

Таблица 4.3

Плотность распределения f_t	Область изменения x	Оптимальные значения		Эффективность
		n_1/n	n_2/n	
Нормальное $\frac{1}{\sqrt{2\pi}} \times \exp(-x^2/2)$	$-\infty < x < \infty$	0,27	0,27	0,81
Прямоугольное $1/2$	$ x < 1$	0,33	0,33	0,89
Полуокружность $3/4 \times (1-x^2)$	$ x < 1$	0,31	0,31	0,86
U -образное $10/9 (1/4 + x^4)$	$ x < 1$	0,39	0,39	0,93
J -образное e^{-2-x}	$-2 < x < \infty$	0,45	0,15	0,79
Скошенное $96x^3 e^{-x/2}$	$0 < x < \infty$	0,36	0,19	0,80

$p_1, p_2; p_1 + p_2 \leq 1$. Группы G_1 и G_3 определяются следующим образом:

$$G_1 = \{(w_t, v_t): t \leq np_1\}, \quad G_3 = \{(w_t, v_t): t \geq (1 - p_2)n\}.$$

Соответствующую оценку обозначим f_2 . В [166] доказаны следующие теоремы о состоятельности оценок f_1 и f_2 .

Теорема 4.2. Пусть $\mu \leq v$ такие, что $P\{\mu \leq \varepsilon_t \leq v\} = 0$. Тогда f_1 состоятельна тогда и только тогда, когда

$$P\{A - v < x_t \leq A - \mu\} = P\{B - v < x_t \leq B - \mu\} = 0.$$

Теорема 4.3. Пусть φ^{p_1} и $\varphi^{p_2} - p_1$ - и p_2 -квантили распределения w_t . Тогда f_2 состоятельна тогда и только тогда, когда

$$\begin{aligned} & P\{\varphi^{p_1} - v < x_t \leq \varphi^{p_1} - \mu\} = \\ & = P\{\varphi^{1-p_2} - v < x_t \leq \varphi^{1-p_2} - \mu\} = 0. \end{aligned}$$

Как видно из теорем, основным условием состоятельности оценок является ограниченность ошибок ε_t и истинных значений x_t . Эти условия не выполняются, когда ошибки распределены по нормальному закону.

Найдем оценку метода группировки для регрессии-примера (1.5). Для этого 15 наблюдений разобьем на 4 группы. Пусть в первую группу войдут первые четыре наблюдения, во вторую — вторые четыре, в третью — третьи четыре, в четвертую — последние три наблюдения. Матрица \bar{C} и вектор \bar{c} при таком разбиении будут следующими:

$$\bar{C} = \begin{bmatrix} 122,3 & 14,21 & 3,652 \\ 63,03 & 7,787 & 2,129 \\ 17,63 & 3,697 & 0,8042 \end{bmatrix}, \quad \bar{c} = \begin{bmatrix} 67,29 \\ 36,58 \\ 13,53 \end{bmatrix}.$$

Оценка группового метода для (a_1, a_2, a_3) равна $(0,229; 1,49; 4,93)$, оценка $a_4 = -107,9$, т. е. $y_t = 0,229 x_{t1} + 1,49 x_{t2} + 4,93 x_{t3} - 107,9$.

Оценка группового метода отличается от оценки МНК. Каковы основные преимущества и недостатки группового

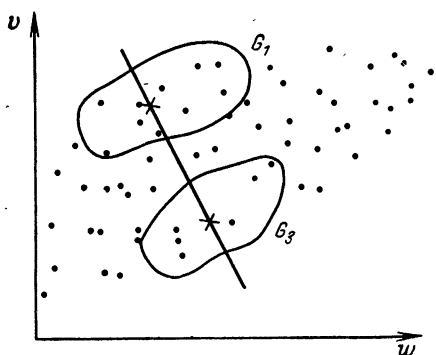


Рис. 4.4. Пример неудовлетворительного разбиения на группы

метода? Преимущества два: 1) простота, 2) при некоторых условиях разбиения этот метод дает состоятельные оценки. Однако групповой метод имеет один серьезный недостаток: *эффективность его резко зависит от разбиения наблюдений на группы*. В частности, не любое разбиение ведет к состоятельным оценкам.

Для того чтобы разбиение было эффективным, необходима дополнительная априорная информация о значениях x_t , $t = 1, 2, \dots, n$.

На рис. 4.4 показана ситуация плохого разбиения: полученная прямая при таком разбиении, очевидно, является неудовлетворительной.

Иногда, в особенности для временных рядов, реальным является наличие в ряде x_t тренда. В этом случае априорная информация заключается в том, что ряд x_t «в среднем»

можно считать монотонным. Удовлетворительным тогда является следующее разбиение:

$$G_1 = \{(x_t, y_t), t = 1, \dots, n_1\}, G_2 = \{(x_t, y_t), t = n_2, \dots, n_2 + 1, \dots, n\}.$$

Не пытайтесь быть строгими в рассуждениях, покажем, что если x_t возрастают с большой вероятностью, то условие состоятельности (4.31) при указанном разбиении на группы выполняется. В силу возрастания ряда $\{x_t\}$ для каждого t $x_{t+1} - x_t \geq d > 0$, поэтому приблизительно

$$\sum_{t \in G_2} x_t / (n - n_2) - \sum_{G_1} x_t / n_1 \geq d \left(n - \frac{n_1 + n_2}{2} \right) > 0,$$

и условие (4.31) заведомо выполняется.

У п р а ж н е н и я 4.4

1. Докажите, что если группировка не зависит от ε_t и (4.29) выполняется, то оценка Вальда состоятельна.

2. Допустим, оценка Вальда a_1^Γ состоятельна. Будет ли оценка $a_2^\Gamma = \bar{y} - a_1^\Gamma x$ состоятельной оценкой α_2 ?

3. Будет ли оценка Барлетта параметра α_2 состоятельна, если условие (4.31) выполняется?

4.5. Метод инструментальных переменных

Допустим, помимо переменных y_t и x_{ti} имеется другая переменная s_{ti} , которая также измеряется с ошибкой, а наблюдаемая величина равна:

$$p_{ti} = s_{ti} + \varphi_{ti}, \quad (4.32)$$

где φ_{ti} для простоты, опять интерпретируем как ошибки измерений, т. е. $\{\varphi_{ti}\}$ независимы между собой, независимы с $\{\xi_t, \varepsilon_{ti}\}$ и $E \varphi_{ti} = 0$, $\sigma^2(\varphi_{ti}) = v_i^2$. В матричной записи (4.32) переписывается следующим образом:

$$P = S + \psi, \quad (4.33)$$

где P , S и ψ — матрицы порядка $n \times m$, первая и последняя стохастические, а вторая детерминированная. Другим существенным предположением, накладываемым на S , является поведение этой матрицы при $n \rightarrow \infty$:

$$\lim \frac{1}{n} S' X = B, \quad |B| \neq 0, \quad \lim \frac{1}{n} S' S = H, \quad |H| \neq 0. \quad (4.34)$$

Переменные, удовлетворяющие (4.33) и (4.34), называем *инструментальными* (ИП). Объяснение термина дано ниже. Теперь вместо оценки МНК рассмотрим другую оценку:

$$a_{И} = (P'W)^{-1}P'V. \quad (4.35)$$

Допустим, X сильно регулярна, т. е. $\lim (X'X)/n = A$. Докажем, что оценка $a_{И}$ при условии (4.34) в отличие от оценки МНК будет состоятельна. Имеем

$$a_{И} = (P'W)^{-1}P'(W\alpha + \eta) = \alpha + (P'W)^{-1}P'\eta.$$

Найдем предел по вероятности:

$$\begin{aligned} \text{plim } \frac{1}{n} P'W &= \text{plim } \frac{1}{n} S'X + \text{plim } \frac{1}{n} S'\varepsilon + \text{plim } \frac{1}{n} \psi'X + \\ &+ \text{plim } \frac{1}{n} \psi'\varepsilon. \end{aligned}$$

Из существования пределов (4.34) следует, что последние три слагаемые равны нулю, поэтому

$$\text{plim } \frac{1}{n} P'W = B.$$

Далее,

$$\text{plim } \frac{1}{n} P'\eta = \text{plim } \frac{1}{n} S'\eta + \text{plim } \frac{1}{n} \psi'\eta.$$

В силу независимости ψ от η и условия (4.34) этот предел также равен нулю, поэтому

$$\text{plim } a_{И} = \alpha + \text{plim } \left(\frac{1}{n} P'W \right)^{-1} \text{plim } \frac{1}{n} P'\eta = \alpha,$$

т. е. оценка $a_{И}$ состоятельна.

Что означает условие (4.34)? Покажем, что оно влечет асимптотическую сопряженность X и S . Сначала точно определим, что понимается под сопряженностью двух матриц X и S . Матрица R порядка $m \times m$ называется матрицей сопряженности матриц X и S , если (i, j) -й элемент этой матрицы равен коэффициенту сопряженности между i -м столбцом матрицы X и j -м столбцом матрицы S . Обозначим

$$D_X = \begin{bmatrix} \sum_t x_{t1}^2 & & 0 \\ & \ddots & \\ 0 & & \sum_t x_{tm}^2 \end{bmatrix}; \quad D_S = \begin{bmatrix} \sum_t s_{t1}^2 & & 0 \\ & \ddots & \\ 0 & & \sum_t s_{tm}^2 \end{bmatrix}.$$

Тогда матрица сопряженности между \mathbf{X} и \mathbf{S} равна:

$$\mathbf{R} = \mathbf{R}(\mathbf{X}, \mathbf{S}) = \mathbf{D}_X^{-1/2} \mathbf{X}' \mathbf{S} \mathbf{D}_S^{-1/2}.$$

Покажем, что при $n \rightarrow \infty$ матрица \mathbf{R} имеет предел. Действительно,

$$\mathbf{R} = \left(\frac{\mathbf{D}_X}{n} \right)^{-1/2} \frac{\mathbf{X}' \mathbf{S}}{n} \left(\frac{\mathbf{D}_S}{n} \right)^{-1/2}.$$

Тогда

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{R} &= \left[\lim_{n \rightarrow \infty} \frac{\mathbf{D}_X}{n} \right]^{-1/2} \lim_{n \rightarrow \infty} \frac{\mathbf{X}' \mathbf{S}}{n} \left[\lim_{n \rightarrow \infty} \frac{\mathbf{D}_S}{n} \right]^{-1/2} = \\ &= \text{Diag}(\mathbf{H})^{-1/2} \mathbf{B}' \text{Diag}(\mathbf{H})^{-1/2} = \mathbf{R}^*, \end{aligned} \quad (4.36)$$

очевидно $|\mathbf{R}^*| \neq 0$.

Теперь понятно, почему матрицу \mathbf{S} называют матрицей инструментальных переменных. Переменные s_{ti} являются «заменителями» переменных x_{ti} , причем их ошибки измерения независимы от ошибок измерения матрицы \mathbf{X} . Это влечет независимость наблюдаемых переменных \mathbf{W} и \mathbf{P} . Если случайные отклонения трактуются как ошибки измерения, то подобная независимость естественна. В противном случае мы не вправе ожидать независимости \mathbf{P} и \mathbf{W} . То, что \mathbf{S} является удовлетворительным заменителем \mathbf{X} , обеспечивается условием сопряженности (4.34). Таким образом, \mathbf{S} выступают в качестве инструмента измерения \mathbf{X} , поэтому соответствующие переменные и называют инструментальными.

В качестве примера инструментальной переменной рассмотрим регрессию из примера с химическим экспериментом параграфа 1.1. Допустим, измерение температуры реакции термометром невозможно. В качестве инструментальной переменной для температуры реакции может выступать спектр некоторого вещества. Таким образом, вместо температуры в регрессии (4.21) подставляем некоторую характеристику спектра. При этом считаем, что ошибки в определении спектра независимы с остальными ошибками. Сопряженность температуры и спектра следует из их тесной взаимосвязи.

Таким образом, инструментальные переменные должны обладать двумя непререкаемыми условиями: а) для каждого t вектор $(p_{t1}, p_{t2}, \dots, p_{tm})$ не зависит от вектора $(\varepsilon_{t1}, \varepsilon_{t2}, \dots, \varepsilon_{tm})$; б) матрицы \mathbf{S} и \mathbf{X} асимптотически сопряжены. Проверить эти условия на практике невозможно.

Поэтому неудивительно, что метод инструментальных переменных неоднократно подвергался критике. Можно указать на три трудности применения этого метода: во-первых, выбор инструментальной переменной произволен, поэтому имеется возможность получения большого спектра оценок, соответствующих разным инструментальным переменным. Во-вторых, очень трудно проверить предположение о независимости инструментальной переменной от ошибок измерения. В-третьих, подход инструментальных переменных возводит свойство состоятельности в ранг особой важности, которое не является таковым в случае больших выборочных дисперсий. В общем случае применение метода ИП весьма проблематично.

В случае же временных динамических рядов применение метода ИП может быть иногда весьма эффективным. Для таких рядов матрица X , как правило, самосопряжена. Поэтому, если в качестве инструментальной переменной взять матрицу, сдвинутую на единицу времени, то мы вправе ожидать хороших свойств оценок метода ИП. Обозначим $X^{(n-1) \times m}$ — матрицу X без последней строки; пусть $x^0 = (x_{01}, \dots, x_{0m})$ — вектор-строка истинных значений x , соответствующая первому моменту времени. Обозначим

$$X_0 = \begin{bmatrix} x^0 \\ X_1 \end{bmatrix}$$

— матрица $n \times m$. Аналогично введем $n \times m$ матрицы W_0 и ε_0 . Предположим, что сопряженность между X_0 и X ненулевая, в частности существует предельная матрица сопряженности R :

$$\lim D_X^{-1/2} X'_0 X D_X^{-1/2} = R, \quad |R| \neq 0, \quad (4.37)$$

где $D_X = \text{Diag}(X'X)$. Далее легко проверить, что из условия регулярности матрицы X следует регулярность и матрицы X_0 , т. е. условие (4.34) выполнено:

$$\lim \frac{1}{n} X'_0 X_0 = \lim \frac{1}{n} X'X = A. \quad (4.38)$$

Больше того,

$$\lim \frac{1}{n} X'_0 X = \lim \left(\frac{D_X}{n} \right)^{1/2} R \left(\frac{D_X}{n} \right)^{1/2} = D_A^{1/2} R D_A^{1/2}, \quad (4.39)$$

где $D_A = \text{Diag}(A)$.

В качестве оценки ИП рассматриваем

$$a_0 = (W'_0 W)^{-1} W'_0 v. \quad (4.40)$$

Матрица ИП W_0 не обладает в точности теми свойствами, которые мы требовали от инструментальных переменных. Ошибки измерения для W_0 , $\{\varepsilon_{it}^0 = \varepsilon_{t-1,i}\}$, вообще говоря, зависят от ошибок измерения $\{\xi_t, \varepsilon_{it}\}$, однако для каждого t эти ошибки независимы. Оценка (4.40) состоятельна. Рассмотрим теорему.

Т е о р е м а 4.4. *Если X строго регулярна, а условие (4.37) выполняется, то оценка (4.40) состоятельна.*

Доказательство дано в параграфе 4.8

Может быть доказана асимптотическая нормальность оценки (4.40). Запишем

$$\sqrt{n}(a_0 - \alpha) = \left(\frac{W'_0 W}{n} \right)^{-1} \frac{W'_0 \eta}{\sqrt{n}}.$$

Из доказательства теоремы (4.4) следует, что вероятностный предел $(W'_0 W/n)^{-1}$ — невырожденная матрица. Вектор $W'_0 \eta / \sqrt{n}$ разлагается на сумму четырех векторов:

$$W'_0 \eta / \sqrt{n} = X'_0 \xi / \sqrt{n} - X'_0 \varepsilon \alpha / \sqrt{n} + \varepsilon'_0 \xi / \sqrt{n} - \varepsilon'_0 \varepsilon \alpha / \sqrt{n}.$$

Легко проверить, что первые три вектора имеют нормальное распределение с нулевым математическим ожиданием. Можно показать, что и последнее слагаемое асимптотически нормально¹. Таким образом, можно считать асимптотически нормальным и сам вектор $\sqrt{n}(a_0 - \alpha)$.

Для практических применений оценки a_0 необходимо, хотя бы приближенно, знать ее матрицу ковариаций. В параграфе 4.8 дана асимптотическая матрица ковариаций вектора $\sqrt{n}(a_0 - \alpha)$. На ее основе может быть получена приближенная матрица ковариаций вектора $a_0 - \alpha$:

$$E(a_0 - \alpha)(a_0 - \alpha)' \approx (W'_0 W)^{-1} [h(W'_0 W) + n(h + s^2)S] (W'_0 W_0)^{-1}, \quad (4.41)$$

где

$$h = s_1^2 + \sum_{i=1}^m a_i^2 s_{i+1}^2;$$

$$S = \text{Diag}(s_2^2, s_3^2, \dots, s_{m+1}^2),$$

а s_i^2 — оценки σ_i^2 или же их точные значения, если σ_i^2 известны. Можно показать, что при $\sigma_i^2 = 0$, $i = 2, \dots, m+1$,

¹Необходимо применить центральную предельную теорему с зависимыми слагаемыми. Доказательство проводится так же, как и в работе [143].

формула (4.41) превращается в точную. Приближение ее будет тем точнее, чем меньше σ_i^2 , $i = 2, \dots, m + 1$.

В [101] предлагаются два метода оценивания зависимостей с ошибками в переменных, каждый из которых есть комбинация метода ИП и МНК. Автор рассматривает простейший случай зависимости:

$$\begin{aligned} y_t &= \alpha x_t, \quad t = 1, \dots, n; \\ v_t &= y_t + \xi_t, \quad \sigma^2(\xi_t) = \sigma^2; \\ \omega_t &= x_t + \varepsilon_t, \quad \sigma^2(\varepsilon_t) = \sigma_1^2, \end{aligned}$$

причем x_t регулярны, т. е. $\lim_{n \rightarrow \infty} \sum x_t^2/n = A > 0$. Пусть z_t — инструментальная переменная, такая, что $z_t = \rho x_t + v_t$, где v_t — случайная ошибка, $E v_t = 0$, $\sigma^2(v_t) = A^1$, независима со всеми остальными ошибками ξ_t, ε_t . Для оценки МНК a и инструментальной переменной $a_0 = \sum_t v_t z_t / \sum z_t \omega_t$ находится асимптотический квадрат отклонения. На основе сравнения асимптотического квадрата отклонения предлагается выбирать либо a_0 , либо a . Вторая оценка есть линейная комбинация оценки МНК и ИП:

$$\lambda a + (1 - \lambda) a_0, \quad 0 \leq \lambda \leq 1.$$

Вычисляется асимптотический квадрат отклонения этой оценки от истинного значения и λ выбирается таким образом, что эта величина обращается в минимум.

Оценки сравниваются здесь методом Монте-Карло. Вторая оценка оказалась более предпочтительной.

У п р а ж н е н и я 4.5

1. Приведите примеры инструментальных переменных в экономике.
2. Будет ли оценка (4.40) несмещенной?

4.6. Оценка Картни—Вайссмана

В [143] Е. Картни и И. Вайссман предложили оценку параметров в регрессии с ошибками в независимых переменных $y_t = \alpha_1 x_t + \alpha_2$ в предположении (4.37). Обобщим их результат и получим соответствующую оценку в общем

¹ Это условие не ограничивает рассмотрения.

виде. Итак, допустим предел (4.37) имеет место, матрица X сильно регулярна. Рассмотрим вероятностный предел

$$\text{plim} \frac{1}{n} (W_0 - W)' (W_0 - W), \quad (4.42)$$

где $W_0 = X_0 + \varepsilon_0$, $W = X + \varepsilon$.

Вычисление этого предела проводится стандартным образом, он равен

$$2A + 2D - D_A^{1/2} (R + R') D_A^{1/2}, \quad (4.43)$$

где матрицы A , D_A и R имеют прежний смысл; D — диагональная матрица, (i, i) -й элемент которой равен σ_{i+1}^2 . Теперь найдем предел

$$\text{plim} \frac{1}{n} (W_0 - W)' v, \quad (4.44)$$

который вычисляется так же, как и (4.42). Он равен

$$D_A^{1/2} R D_A^{1/2} \alpha - A \alpha. \quad (4.45)$$

Введем следующие обозначения:

$$A_1 = \frac{1}{n} W'W; \quad A_2 = \frac{1}{n} W'v; \quad A_3 = \frac{1}{n} (W_0 - W)' (W_0 - W); \quad (4.46)$$

$$A_4 = \frac{1}{n} (W_0 - W)' (v_0 - v); \quad P = 2A - D_A^{1/2} (R + R') D_A^{1/2}.$$

Тогда формально можно записать:

$$A_1 = A + D; \quad A_2 = A \alpha; \quad A_3 = 2D + P; \quad A_4 = P \alpha. \quad (4.47)$$

Решая эту систему относительно α , найдем

$$\alpha = (2A_1 - A_3)^{-1} (2A_2 - A_4). \quad (4.48)$$

Оценкой Картни — Вайссмана назовем статистику

$$a_{\text{КВ}} = [2W'W - (W_0 - W)' (W_0 - W)]^{-1} [2W'v - (W_0 - W)' (v_0 - v)], \quad (4.49)$$

которая получается после подстановки (4.46) в (4.48). Состоятельность (4.49) можно проверить непосредственно, все необходимые пределы найдены ранее (см. параграф 4.5). В [143] доказывается, что оценка (4.49) для случая простой регрессии асимптотически нормальна; там же приводится асимптотическая дисперсия оценки. Картни и Вайссман доказали, что их оценка более эффективна, чем оценка a_0 в асимптотическом смысле.

У п р а ж н е н и е 4.6

1. Найдите оценку $a_{\text{КВ}}$ для регрессий $y_t = \alpha x_t + \varepsilon_t$ и $y_t = \alpha_1 x_t + \alpha_2 + \varepsilon_t$.

4.7. Сравнение оценок

Отличительной чертой всех рассмотренных оценок является требование наличия априорной информации. Априорная информация о системе (4.1) может выступать в разных видах. Так, в случае ММП мы должны знать дисперсии величин хотя бы с точностью до постоянного множителя. Если таковой информации не имеется, то оценок ММП не существует, а оценки ортогональной регрессии не являются состоятельными. В методе разбиения выборки на группы (А. Вальд, М. Бартлетт) мы должны иметь дополнительную информацию о неизвестных истинных значениях независимых переменных. В противном случае оценка группового метода будет неудовлетворительной. В методе инструментальных переменных информация выступает в виде знания инструментальных переменных, на которые накладываются жесткие условия. В частности, в оценках a_0 и $a_{\text{КВ}}$ мы требовали асимптотической автосопряженности независимых переменных. По-видимому, без дополнительной информации невозможно вообще состоятельно оценить параметры зависимости (4.1), даже в предположении сильной регулярности матриц X .

Основным критерием допустимости той или иной оценки является ее состоятельность. Однако состоятельность — свойство теоретического характера. Ясно, что несостоятельная оценка для данного, конкретного n может быть намного лучше состоятельной оценки. И это скорее будет наблюдаться для не очень больших n . Этот вывод, в частности, подтверждается следующими расчетами.

Каждая переменная регрессия (1.5), по предположению, измерялась с ошибкой. Стандартные отклонения этих ошибок приведены в табл. 4.1. Объясним их выбор. Допустим, точность весов, измеряющих выход реакции, равна 10 г. Тогда $v_1 = 140$ кг 280 г, а ошибка измерения лежит в пределах от $-0,005$ до $0,005$ г. Поэтому можно предположить, что, например, ξ_1 равномерно распределена на интервале $(-0,005; 0,005)$, v_1 равномерно распределена на интервале $(140, 275; 140, 285)$, а истинное значение $140, 275 < y_1 < 140, 285$. При этом стандартное отклонение ξ_1 равно: $\sigma_1 = 0,01/\sqrt{12} = 0,00289$. Аналогичное рассуждение про-

водим для остальных случайных величин v_i , w_{ii} . Таким образом, случай S_1 соответствует предположению о точности измерения до 0,01. Случай S_2 соответствует гипотезе о точности измерения до 0,1 и т. д. Случай S_5 соответствует экстремальной ситуации, когда точность весов равна 100 кг, а точность термометра 100° . Набор стандартных отклонений S_8 отвечает ситуации, когда y , x_1 , x_3 измеряются на одних весах с точностью деления 1 кг. Температура при этом замеряется неточно, ее точность равна 10° .

Для данной регрессии для различных вариантов $S_1 - S_9$ имитировались измерения y_i , x_{t1} , x_{t2} и x_{t3} . В качестве истинных значений выбирались значения из табл. 1.1. Для каждой имитации (всего было 500 испытаний) вычислялись оценки МНК, ММП, ИП, ортогональной регрессии (ОР), Картни — Вайссмана (КВ), группового метода (ГМ). В групповом методе в первую группу вошли наблюдения первых четырех экспериментов, во вторую — наблюдения вторых четырех экспериментов и т. д., в последнюю — последних трех экспериментов. В качестве «истинных» параметров были взяты параметры $\alpha_1 = 0,3974$; $\alpha_2 = 0,229$; $\alpha_3 = 3,746$; $\alpha_4 = -17,083$. После 500 испытаний были вычислены средние квадратов отклонений оценок от истинных значений, т. е.

$$\frac{1}{500} \sum_{i=1}^{500} (a_i^j - \alpha_i)^2,$$

где a_i^j — значение j -й оценки ($j = 1, 2, 3, 4, 5, 6$) в i -м испытании. Во всех случаях лучшей оказалась оценка МНК. Распределение мест показано на рис. 4.5. Чем можно объяснить на первый взгляд странный вывод при сравнении эффективности шести оценок? Ведь все оценки, за исключением оценки МНК, были при некоторых условиях состоятель-

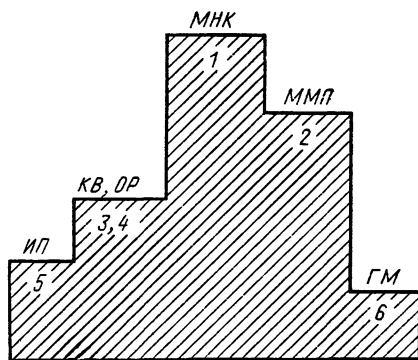


Рис. 4.5. Распределение мест 6 оценок параметров регрессии-примера

ны. Дело в том, что объем выборки в нашей регрессии очень невелик, равен 15. А при таких объемах несостоятельные оценки могут быть лучше состоятельных. Приведенные результаты статистических испытаний с регрессией (1.5) — хорошая иллюстрация этому. Если бы объем выборки в регрессии (1.5) был равен 100 или более, то результаты по сравнению оценок, возможно, были бы другие.

Исследование статистических свойств оценок, рассмотренных в этой главе, практически не проводилось. Исследованы лишь простейшие зависимости $y_t = \alpha x_t$ и $y_t = \alpha_1 x_t + \alpha_2$ (оценка МНК [177], оценка ММП [72]), которые на практике применяются весьма редко.

4.8. Доказательства

1. Доказательство теоремы 4.1. Прежде всего докажем следующий факт: пусть \mathbf{A}_n — стохастическая симметричная матрица порядка $m \times m$, причем $\text{plim } \mathbf{A}_n = \mathbf{A}$ — детерминированная матрица, тогда $\text{plim } \lambda_{\text{min}}(\mathbf{A}_n) = \lambda_{\text{min}}(\mathbf{A})$ следует из непрерывности характеристических чисел относительно элементов матрицы [53, с. 206]. Отсюда

$$\text{plim } \lambda_{\text{min}}(\mathbf{A}_n) = \lambda_{\text{min}}(\text{plim } \mathbf{A}_n) = \lambda_{\text{min}}(\mathbf{A}). \quad (4.50)$$

Перейдем к доказательству теоремы. Поскольку размерность матрицы \mathbf{A} равна $k \times k$, а ее ранг $m = k - 1$, то $|\mathbf{A}| = 0$. Для любого истинного значения вектора параметров β , удовлетворяющих (4.15), имеем

$$\mathbf{A}\beta = \lim \frac{1}{n} \mathbf{Z}'_n \mathbf{Z}_n \beta = 0, \quad \|\beta\| = 1.$$

Поскольку $\text{rank } \mathbf{A} = k - 1$, то существует единственный вектор $\|\beta\| = 1$, для которого $\mathbf{A}\beta = 0$. Пусть $(\mathbf{b}_{OP})_k$ — любая сходящаяся подпоследовательность последовательности $(\mathbf{b}_{OP})_n$, причем $P\{\lim (\mathbf{b}_{OP})_k = \gamma\} = 1$. Сильная сходимость $(\mathbf{b}_{OP})_k$ к γ влечет сходимость по вероятности, т. е. $\text{plim } (\mathbf{b}_{OP})_k = \gamma$. Докажем, что тогда $\gamma = \beta$. Для этого достаточно показать, что $\mathbf{A}\gamma = 0$.

Ранее было показано, что

$$\begin{aligned} \text{plim } \frac{1}{n} \mathbf{V}' \mathbf{V} &= \mathbf{A} + \sigma^2 \mathbf{I}_k, \text{ поэтому } \mathbf{A}\gamma = \left(\text{plim } \frac{1}{k} \mathbf{V}' \mathbf{V} - \right. \\ &\left. - \sigma^2 \mathbf{I}_k \right) \text{plim } (\mathbf{b}_{OP})_k = \text{plim } \frac{1}{k} \mathbf{V}' \mathbf{V} (\mathbf{b}_{OP})_k - \sigma^2 \gamma = \\ &= \text{plim } \lambda_{\text{min}} \left(\frac{1}{k} \mathbf{V}' \mathbf{V} \right) (\mathbf{b}_{OP})_k. \end{aligned}$$

С учетом (4.50)

$$\text{plim } \lambda_{\min} \left(\frac{1}{k} \mathbf{V}'\mathbf{V} \right) = \lambda_{\min} (\mathbf{A} + \sigma^2 \mathbf{I}_k) = \sigma^2,$$

поэтому

$$\mathbf{A}\boldsymbol{\gamma} = \lambda_{\min} \left(\frac{1}{k} \mathbf{V}'\mathbf{V} \right) \text{plim } (\mathbf{b}_{\text{OP}})_k - \sigma^2 \boldsymbol{\gamma} = \sigma^2 \boldsymbol{\gamma} - \sigma \boldsymbol{\gamma} = \mathbf{0}.$$

Теперь покажем, что $\lim_{n \rightarrow \infty} (\mathbf{b}_{\text{OP}})_n = \boldsymbol{\beta}$ с вероятностью 1.

Пусть \mathbf{w} — любая точка выборочного пространства. Поскольку, $\| \mathbf{b}_{\text{OP}}(\mathbf{w}) \| = 1$, то найдется хотя бы одна сходящаяся подпоследовательность $(\mathbf{b}_{\text{OP}})_k \rightarrow \boldsymbol{\gamma}$. Но было показано, что тогда $\boldsymbol{\gamma} = \boldsymbol{\beta}$. Значит, с вероятностью 1 $\lim_{n \rightarrow \infty} (\mathbf{b}_{\text{OP}})_n = \boldsymbol{\beta}$.

2. Нахождение матрицы \mathbf{C} из выражения (4.27). Перепишем (4.27), подставляя вместо \mathbf{Z} ее оценку $\hat{\mathbf{Z}}$, вместо $\boldsymbol{\beta}$ — его оценку $\mathbf{b}_{\text{OP}} = \mathbf{b}$, вместо σ^2 — оценку $\hat{\sigma}^2 = \lambda \min(\mathbf{V}'\mathbf{V})/nk$:

$$\frac{1}{n^2} \hat{\mathbf{Z}}' \hat{\mathbf{Z}} \hat{\mathbf{C}} \hat{\mathbf{Z}}' \hat{\mathbf{Z}} = \frac{1}{n} \hat{\mathbf{Z}}' \hat{\mathbf{Z}} + \hat{\sigma}^2 \mathbf{I}_k - \hat{\sigma}^4 \mathbf{b} \mathbf{b}', \quad (4.51)$$

причем \mathbf{C} находим из условия $\mathbf{C} \mathbf{b} = \mathbf{0}$. Обозначим $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ — характеристические числа матрицы $\hat{\mathbf{Z}}' \hat{\mathbf{Z}}$, $\lambda_k = \lambda_{\min}(\hat{\mathbf{Z}}' \hat{\mathbf{Z}}) = 0$. Характеристические векторы, соответствующие этим числам, обозначим $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k = \mathbf{b}$. Пусть \mathbf{P} — ортогональная матрица, столбцы которой — суть х.в. матрицы $\hat{\mathbf{Z}}' \hat{\mathbf{Z}}$. По определению $\hat{\mathbf{Z}}' \hat{\mathbf{Z}} = \mathbf{P} \boldsymbol{\Lambda} \mathbf{P}'$, где $\boldsymbol{\Lambda} = \text{Diag}(\lambda_1, \dots, \lambda_m, 0)$. Умножим (4.51) слева на \mathbf{P}' и справа на \mathbf{P} , получим

$$\frac{1}{n^2} \mathbf{P}' \hat{\mathbf{C}} \mathbf{P} \boldsymbol{\Lambda} = \frac{1}{n} \boldsymbol{\Lambda} + \hat{\sigma}^2 \mathbf{I} - \hat{\sigma}^4 \mathbf{P}' \mathbf{b} \mathbf{b}' \mathbf{P}. \quad (4.52)$$

Обозначим $\boldsymbol{\Lambda}^- = \text{Diag}(1/\lambda_1, \dots, 1/\lambda_m, 0)$, $\mathbf{E}^- = \text{Diag}(1, \dots, 1, 0)$. Умножим (4.52) слева и справа на $\boldsymbol{\Lambda}^-$, получим

$$\frac{1}{n^2} \mathbf{E}^- \mathbf{P}' \hat{\mathbf{C}} \mathbf{P}' \mathbf{E}^- = \frac{1}{n} \boldsymbol{\Lambda}^- + \hat{\sigma}^2 (\boldsymbol{\Lambda}^-)^2 - \hat{\sigma}^2 \boldsymbol{\Lambda}^- \mathbf{P}' \mathbf{b} \mathbf{b}' \mathbf{P} \boldsymbol{\Lambda}^-. \quad (4.53)$$

Обозначим $\mathbf{P} \mathbf{E}^- = \mathbf{P}_0$ — матрица, совпадающая с \mathbf{P} , за исключением последнего столбца, который равен нулю. Очевидно, $\mathbf{P}'_0 \mathbf{P}_0 = \mathbf{P}_0 \mathbf{P}'_0 = \mathbf{E}^-$. Умножим (4.53) слева на \mathbf{P}_0 и справа на \mathbf{P}'_0 , получим

$$\begin{aligned} \frac{1}{n^2} \mathbf{C}_0 &= \frac{1}{n^2} \mathbf{E}^- \mathbf{C} \mathbf{E}^- = \frac{1}{n} \mathbf{P}_0 \boldsymbol{\Lambda}^- \mathbf{P}'_0 + \hat{\sigma}^2 \mathbf{P}_0 (\boldsymbol{\Lambda}^-)^2 \mathbf{P}'_0 - \\ &\quad - \hat{\sigma}^4 \mathbf{P}_0 \boldsymbol{\Lambda}^- \mathbf{b} \mathbf{b}' \mathbf{P} \boldsymbol{\Lambda}^- \mathbf{P}'_0, \end{aligned} \quad (4.54)$$

где матрица C_0 совпадает с матрицей C , за исключением последней строки и последнего столбца, которые в матрице C_0 равны нулю. По построению матрицы P , в которой последний столбец равен b , имеем $Pb = (0, 0, \dots, 1)'$ откуда $\Lambda^{-1}Pb = 0$, что в свою очередь ведет к тому, что последнее слагаемое в выражении (4.54) обращается в нуль. Окончательно матрица ковариаций для (b_1, b_2, \dots, b_m) может быть *приближенно* определена как

$$\text{cov}(b_1, \dots, b_m) \approx P_0 \Lambda^{-1} P_0' + \hat{\sigma}^2 P_0 (\Lambda^{-1})^2 P_0' n. \quad (4.55)$$

Для того чтобы найти дисперсию $b_k - \beta_k$ (т. е. c_{kk}), воспользуемся условием $Cb = 0$. Разбивая матрицу V и вектор b на подвекторы, получим

$$C = \begin{bmatrix} C_0 & c_0 \\ c_0' & c_{kk} \end{bmatrix} \begin{bmatrix} b_0 \\ b_k \end{bmatrix} = \begin{bmatrix} C_0 b_0 + c_0 b_k \\ c_0' b_0 + c_{kk} b_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \quad (4.56)$$

здесь C_0 — матрица $m \times m$, c_0 — вектор-столбец $m \times 1$, b_0 — вектор-столбец $m \times 1$. Из выражения (4.56) получаем $c_0 = -\frac{1}{b_k} C_0 b_0$, откуда $c_{kk} = \frac{1}{b_k^2} b_0' C_0 b_0$.

Часто нас интересуют параметры, нормализованные условием $\beta_1 = 1$, что соответствует уравнению исходной модели (4.1). Поэтому нас также интересуют асимптотические дисперсии b_i/b_1 , $i = 2, \dots, k$. Их можно вычислить на основе матрицы (4.54). Рассмотрим отношение b_i/b_1 как функцию двух переменных. Разложим эту функцию в ряд Тейлора до линейных членов в окрестности (β_i, β_1) , тогда

$$\frac{b_i}{b_1} \approx \frac{\beta_i}{\beta_1} + (b_i - \beta_i) \frac{1}{\beta_1} - (b_1 - \beta_1) \frac{\beta_i}{\beta_1^2},$$

откуда

$$\begin{aligned} E \left(\frac{b_i}{b_1} - \frac{\beta_i}{\beta_1} \right)^2 &= E \left[(b_i - \beta_i) \frac{1}{\beta_1} - (b_1 - \beta_1) \frac{\beta_i}{\beta_1^2} \right]^2 = \\ &= \frac{E(b_i - \beta_i)^2}{\beta_1^2} - 2 \frac{\beta_i}{\beta_1^2} E(b_i - \beta_i)(b_1 - \beta_1) + \frac{\beta_i^2}{\beta_1^4} E(b_1 - \beta_1)^2 = \\ &= \frac{1}{\beta_1^2} \left(C_{ii} - \frac{2\beta_i}{\beta_1} C_{i1} + \frac{\beta_i^2}{\beta_1^2} C_{11} \right), \quad i = 2, \dots, m. \quad (4.57) \end{aligned}$$

С помощью формулы (4.57) можно находить приближенные дисперсии оценок b_i . Для этого необходимо вместо истинных значений β_i подставить в (4.57) их оценки b_i .

3. Доказательство теоремы 4.4. Имеем

$$a_0 = \alpha + (W'_0 W)^{-1} W'_0 \eta = \alpha + \left(\frac{W'_0 W}{n} \right)^{-1} \frac{W'_0 \eta}{n}.$$

Далее

$$\begin{aligned} \text{plim } \frac{W'_0 W}{n} &= \text{plim } \frac{1}{n} (X'_0 + \varepsilon'_0) (X + \varepsilon) = \lim \frac{1}{n} X'_0 X_0 + \\ &+ \text{plim } \frac{1}{n} X'_0 \varepsilon + \text{plim } \frac{1}{n} \varepsilon'_0 X + \text{plim } \frac{1}{n} \varepsilon'_0 \varepsilon. \end{aligned}$$

Первое слагаемое равно (4.39); второе и третье — нули, нетрудно видеть, что и последнее слагаемое — нуль. Итак,

$$\text{plim } \frac{W'_0 W}{n} = D_A^{1/2} R D_A^{1/2}$$

— невырожденная матрица $m \times m$. Далее,

$$\begin{aligned} \text{plim } \frac{1}{n} W'_0 \eta &= \text{plim } \frac{1}{n} X'_0 \xi - \text{plim } \frac{1}{n} X'_0 \varepsilon \alpha + \\ &+ \text{plim } \frac{1}{n} \varepsilon'_0 \xi - \text{plim } \frac{1}{n} \varepsilon'_0 \varepsilon \alpha. \end{aligned}$$

Первые три предела, очевидно, равны нулю, последний также равен нулю. Окончательно

$$\text{plim } a_0 = \alpha + \left(\text{plim } \frac{W'_0 W}{n} \right)^{-1} \text{plim } \frac{W'_0 \eta}{n} = \alpha.$$

4. Доказательство формулы (4.41). Сначала найдем

$$\lim \frac{1}{n} E(W'_0 \eta)(W'_0 \eta)' = \lim \frac{1}{n} W'_0 \eta \eta' W_0.$$

Имеем

$$\begin{aligned} \frac{1}{n} (W'_0 \eta)(\eta' W_0) &= \frac{1}{n} (X'_0 \xi - X'_0 \varepsilon \alpha + \varepsilon'_0 \xi - \varepsilon'_0 \varepsilon \alpha) (\xi' X_0 - \\ &- \alpha' \varepsilon' X_0 + \xi' \varepsilon_0 - \alpha' \varepsilon' \varepsilon_0) = \frac{1}{n} X'_0 \xi \xi' X_0 - \frac{1}{n} X'_0 \xi \alpha' \varepsilon' X_0 + \\ &+ \frac{1}{n} X'_0 \xi \xi' \varepsilon_0 - \frac{1}{n} X'_0 \xi \alpha' \varepsilon' \varepsilon_0 - \frac{1}{n} X'_0 \varepsilon \alpha \xi' X_0 + \\ &+ \frac{1}{n} X'_0 \varepsilon \alpha \alpha' \varepsilon' X_0 - \frac{1}{n} X'_0 \varepsilon \alpha \xi' \varepsilon_0 + \frac{1}{n} X'_0 \varepsilon \alpha \alpha' \varepsilon' \varepsilon_0 - \end{aligned}$$

$$\begin{aligned}
& -\frac{1}{n} \varepsilon'_0 \varepsilon \alpha \xi' X_0 + \frac{1}{n} \varepsilon'_0 \varepsilon \alpha \alpha' \varepsilon' X_0 - \frac{1}{n} \varepsilon'_0 \varepsilon \alpha \xi' \varepsilon_0 + \\
& + \frac{1}{n} \varepsilon'_0 \varepsilon \alpha \alpha' \varepsilon' \varepsilon_0 + \frac{1}{n} \varepsilon'_0 \xi \xi' X_0 - \frac{1}{n} \varepsilon'_0 \xi \alpha' \varepsilon' X_0 + \\
& + \frac{1}{n} \varepsilon'_0 \xi \xi' \varepsilon_0 + \frac{1}{n} \varepsilon'_0 \xi \alpha' \varepsilon' \varepsilon_0.
\end{aligned}$$

Рассмотрим первое слагаемое. В силу предположения $E\xi\xi' = \sigma^2 I_n$, поэтому

$$\frac{1}{n} E X_0' \xi \xi' X_0 = \sigma^2 \frac{1}{n} X_0' X_0 \rightarrow \sigma^2 A.$$

Математические ожидания второго—пятого слагаемых равны нулю в силу независимости ξ и ε . Далее имеем

$$E(\varepsilon \alpha \alpha' \varepsilon')_{ik} = E \sum_{j, s}^m \varepsilon_{ij} \varepsilon_{ks} \alpha_j \alpha_s, \quad i, k = 1, \dots, n.$$

В этой сумме слагаемые не равны нулю, только если $j = s$ и $i = k$. При этом $E(\varepsilon \alpha \alpha' \varepsilon)_{ii} = \sum_j E \varepsilon_{ij}^2 \alpha_j^2 = \sum_j \sigma_j^2 \alpha_j^2$, поэтому

$$\frac{1}{n} E(X_0' \varepsilon \alpha \alpha' \varepsilon' X_0) \rightarrow \sum_{j=1}^m \sigma_j^2 \alpha_j^2 \cdot A.$$

Математическое ожидание седьмого слагаемого в силу независимости ε и ξ также равно нулю. В восьмом слагаемом типичным является присутствие ε_{ii}^2 . Предположим, что $E\varepsilon_{ii}^2 = 0$, для этого достаточно предположить, что ε_{ii} имеют симметричное распределение относительно нуля. Тогда м.о. восьмого слагаемого также равно нулю. Девятое слагаемое равно нулю в силу независимости ε и ξ . М. о. десятого и одиннадцатого слагаемых равны нулю по тем же причинам, что и восьмого. Далее имеем для $i, p = 1, \dots, m$

$$E(\varepsilon'_0 \varepsilon \alpha \alpha' \varepsilon' \varepsilon_0)_{ip} = E \sum_{s, k, j, r} \varepsilon_{j-1, i} \varepsilon_{jk} \varepsilon_{s-1, p} \varepsilon_{sr} \alpha_k \alpha_r.$$

Слагаемые в последней сумме не равны нулю, только если $j = s, i = p, r = k$. Тогда

$$\begin{aligned}
E(\varepsilon'_0 \varepsilon \alpha \alpha' \varepsilon' \varepsilon_0)_{ii} &= E \sum_{j=1}^n \sum_{k=1}^m (\varepsilon_{j-1, i})^2 \varepsilon_{j, k}^2 \alpha_k^2 = \\
&= n \sigma_i^2 \sum_k \sigma_k^2 \alpha_k^2.
\end{aligned}$$

Поэтому

$$\frac{1}{n} E (\varepsilon'_0 \boldsymbol{\varepsilon} \boldsymbol{\alpha} \boldsymbol{\alpha}' \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}_0) \rightarrow \sum_{j=1}^m \sigma_j^2 \alpha_j^2 \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_m^2 \end{bmatrix}.$$

М. о. тринадцатого и четырнадцатого слагаемых в силу независимости ξ и ε равны нулю. Для пятнадцатого слагаемого

$$E (\varepsilon'_0 \xi' \xi \boldsymbol{\varepsilon}_0)_{ik} = E \sum_{j,s} \varepsilon_{j-1, i} \varepsilon_{s-1, k} \xi_j \xi_s = n \sigma^2 \sigma_i^2$$

при $i = k$ и равно нулю при $i \neq k$, т. е.

$$\frac{1}{n} E (\varepsilon'_0 \xi' \xi \boldsymbol{\varepsilon}_0) \rightarrow \sigma^2 \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_m^2 \end{bmatrix}.$$

М. о. последнего слагаемого в силу независимости ξ и ε равно нулю. Обозначим

$$h = \sigma^2 + \sum_{i=1}^m \alpha_i^2 \sigma_i^2, \quad \mathbf{S} = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_m^2 \end{bmatrix},$$

тогда окончательно

$$\begin{aligned} \lim \frac{1}{n} E \mathbf{W}'_0 \boldsymbol{\eta} \boldsymbol{\eta}' \mathbf{W}_0 &= h \mathbf{A} + h \mathbf{S} + \sigma^2 \mathbf{S} = h \mathbf{A} + (h + \sigma^2) \mathbf{S}, \\ E n (\mathbf{a}_0 - \boldsymbol{\alpha}) (\mathbf{a}_0 - \boldsymbol{\alpha})' &\rightarrow (\mathbf{D}_A^{1/2} \mathbf{R} \mathbf{D}_A^{1/2})^{-1} \times \\ &\times [h \mathbf{A} + (h + \sigma^2) \mathbf{S}] (\mathbf{D}_A^{1/2} \mathbf{R}' \mathbf{D}_A^{1/2})^{-1}. \end{aligned}$$

Глава 5

РОБАСТНЫЕ ОЦЕНКИ

5.1. Робастные оценки параметра положения

В условиях нормальной гипотезы метод наименьших квадратов является оптимальным. Отметим характерную особенность нормального распределения — основная масса распределения сосредоточена на конечном интервале

(— 3σ , 3σ), Вне этого интервала находится лишь 0,27 % распределения¹. Другими словами, нормальное распределение имеет «легкие хвосты».

Таким образом, принимая гипотезу нормальности, мы автоматически предполагаем, что основная масса отклонений сосредоточена на некотором интервале. Вероятность большого отклонения при этом весьма мала. В реальной ситуации эта гипотеза является чересчур жесткой. Дело в том, что предполагаемая модель редко является абсолютно точно специфицированной; в частности, наблюдения могут быть засорены. Разумнее поэтому предположить, что отклонения с большей вероятностью могут принимать и большие значения. Это заставляет нас отказаться от распределения с легкими хвостами (в частности, от нормального распределения) и перейти к распределениям с тяжелыми хвостами. Оценки, ориентированные на распределения с легкими хвостами (в частности, оценка МНК), в новой ситуации оказываются далекими от эффективных. В распределениях с тяжелыми хвостами более эффективными будут менее чувствительные оценки, а именно такие, которые не меняют резко своих значений при возникновении больших отклонений (выбросов). Такие оценки будем называть *робастными* (от английского слова *robust* — устойчивый), или устойчивыми. Робастные оценки устойчивы относительно априорного распределения отклонений. Если отклонения не засорены, т. е. вероятность больших отклонений мала, робастные оценки будут менее эффективны, зато если отклонения содержат выбросы, то эти оценки будут малочувствительны к ним, а потому более удовлетворительными. Таким образом, переходя к распределениям с более тяжелыми хвостами, мы теряем в эффективности, но приобретаем в надежности. Соответствующие методы будут менее чувствительны к ошибкам спецификации отклонений регрессии.

Специально проблеме робастного (устойчивого) оценивания посвящена книга Б. А. Смоляка и Б. П. Титаренко [62], а также работы [56] и [32].

Чтобы не усложнять проблему робастного оценивания техническими деталями, рассмотрим сначала простейший случай: оценивание параметра положения.

¹Как следует из неравенства Чебышева; максимальный процент расположения вне этого интервала равен $1/11$, что соответствует распределению; сосредоточенному в двух точках.

Итак, пусть перед нами стоит следующая статистическая задача: наблюдения y_1, y_2, \dots, y_n — независимы и одинаково распределены с функцией распределения $F(x; \theta) = \Phi(x - \theta)$, где θ — параметр положения; $\theta \in (-\infty, \infty)$. Параметр θ , определяющий «центр» распределения случайной величины, подлежит оцениванию. Для простоты параметр масштаба σ считаем равным единице¹. Известно, что если Φ — пропорциональна $e^{-x^2/2}$, т. е. выборка извлечена из нормальной генеральной совокупности, то средняя $\bar{y} = \Sigma y_i/n$ является эффективной оценкой в классе несмещенных оценок θ (см. параграф 1.4). Однако если Φ имеет тяжелые хвосты, оценка \bar{y} уже не будет эффективной. Действительно, оценка \bar{y} направлена на минимизацию суммы квадратов отклонений:

$$\sum_i (y_i - \theta)^2. \quad (5.1)$$

Если же Φ имеет тяжелые хвосты, то весьма вероятно получение больших отклонений $\varepsilon_i = y_i - \theta$; возведение их в квадрат в сумме (5.1) приведет к резкому смещению \bar{y} в сторону больших отклонений.

Наиболее простой способ нивелировки \bar{y} состоит в следующем: отбросим минимальное и максимальное наблюдения в выборке y_1, y_2, \dots, y_n . На основе оставшихся наблюдений найдем новую среднюю

$$\bar{y}^{(2)} = \frac{1}{n-2} \sum_{i=2}^{n-1} y_{(i)},$$

где $y_{(i)}$ — ранжированный ряд, составленный из первоначальной выборки. Можно отбросить первые два и последние два члена ранжированного ряда и затем построить новую среднюю и т. д. Наконец, можно задаться долей $\beta > 0$ и отбрасывать члены вариационного ряда, для которых $i < \beta n$, $i > n(1 - \beta)$. Полученная средняя называется β -усеченная средняя и будет робастной (устойчивой) оценкой.

Обычная средняя также будет неэффективной в случае, когда y_1, \dots, y_n распределены неодинаково, и некоторые наблюдения имеют большую дисперсию, т. е. являются выбро-

¹Для наглядности под параметром масштаба можно понимать стандартное отклонение наблюдений.

сами. Сильно реагируя на такие выбросы, \bar{y} будет иметь большую дисперсию.

Другой робастной оценкой является медиана. Напомним, что медиана выборки y_1, y_2, \dots, y_n есть величина, по левую и по правую стороны от которой лежит одинаковое количество наблюдений. Интуитивно ясно, что медиана выборки будет устойчивее к виду распределения генеральной совокупности, чем обычная средняя. Вес наблюдения при построении медианы не зависит от его значения и равен 1. Поэтому даже большие отклонения не так резко изменяют значение медианы, как это произойдет в средней. Мы еще вернемся к медиане как робастной оценке параметра положения при рассмотрении оценок ММП.

П. Хьюбер [136] предложил целый класс робастных оценок (M -оценки). Вместо квадратичной функции в сумме (5.1) он рассмотрел минимизацию суммы вида

$$\sum_{i=1}^n \rho(y_i - \theta), \quad (5.2)$$

где ρ — некоторая выпуклая функция. Значение θ , которое обращает (5.2) в минимум для некоторой функции ρ , называется M -оценкой. Легко видеть, что M -оценку θ можно рассматривать как оценку метода максимального правдоподобия. Действительно, пусть y_i имеют функцию плотности $f(y_i - \theta)$. Тогда в силу независимости и одинаковой распределенности y_1, y_2, \dots, y_n функция плотности выборки равна:

$$F(y, \theta) = \prod_{i=1}^n f(y_i - \theta).$$

Если обозначить $\rho = -\ln f$, то после логарифмирования F приходим к выражению

$$-\ln F(y; \theta) = \sum_{i=1}^n \rho(y_i - \theta) \Rightarrow \min.$$

Для того чтобы оценка $\hat{\theta} = \hat{\theta}(y)$ была робастной, необходимо чтобы ρ была «менее возрастающей», чем x^2 . Например, Хьюбер предложил следующую функцию:

$$\rho(x) = \begin{cases} \frac{1}{2} x^2, & \text{если } |x| < c; \\ c|x| - \frac{1}{2} c^2, & \text{если } |x| \geq c. \end{cases} \quad (5.3)$$

Идея заключалась в том, чтобы вклад значений $y_i - \theta$, которые меньше по абсолютной величине некоторого порогового значения $c > 0$, в сумму измерять в квадратах от-

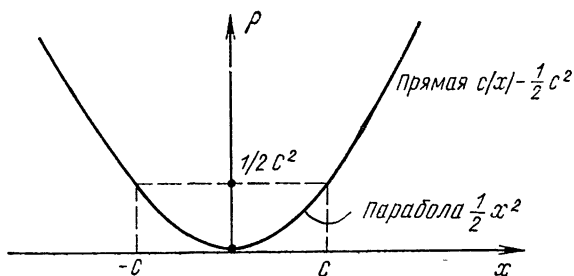


Рис. 5.1. Функция Хюбера $\rho(x)$ (5.3)

клонений (на рис. 5.1 этим значениям соответствует интервал $[-c, c]$); для наблюдений, для которых $|y_i - \theta|$ больше c , вклад измерять в более умеренных единицах — пропорционально $|y_i - \theta|$, на рисунке этим значениям соответствует интервал $(-\infty, c)$ и $(c, +\infty)$. Очевидно, что если $c = +\infty$, то приходим к оценке МНК. Можно рассмотреть целый класс оценок, соответствующих функциям $\rho_\nu(x) = |x|^\nu$, $0 < \nu < 2$, что приводит к минимизации сумм вида (рис. 5.2)

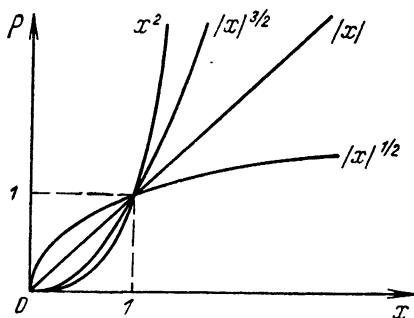


Рис. 5.2. Различные функции $\rho(x) = |x|^\nu$, $x > 0$

$$\sum_i |y_i - \theta|^\nu, \quad 0 < \nu < 2. \quad (5.4)$$

Оценки $\hat{\theta}$, получаемые в результате минимизации суммы (5.4), будем называть L_ν -оценками. Эти оценки малочувствительны к большим отклонениям $y_i - \theta$. Чем меньше значение ν , тем эта чувствительность меньше. В частности, можно показать, что медиана есть L_1 -оценка параметра положения. В качестве ρ можно предложить и другие функции. Большой набор функций ρ предлагается в [74]. Там

же исследуется эффективность робастного оценивания для некоторых из вводимых функций.

До сих пор мы считали параметр масштаба σ известным. В случае когда он неизвестен, минимизируемая сумма (5.1) трансформируется в

$$\sum_{i=1}^n \rho((y_i - \theta)/\sigma), \quad (5.5)$$

где σ также подлежит оцениванию. Для функции ρ типа $|x|^v$ введение параметра масштаба не меняет оценки (σ выносится за знак суммы). Если же ρ есть, например, функция Хьюбера, то σ может существенно повлиять на оценку. Для оценивания σ можно взять следующую статистику:

s = медиана $\{|y_i - M|, i = 1, 2, \dots, n\}$, где M — медиана выборки y_1, y_2, \dots, y_n .

Помимо M -оценок, введенных Хьюбером, существуют два других класса робастных оценок: L -оценки, основанные на упорядоченной выборке $y_{(1)}, y_{(2)}, \dots, y_{(n)}$ [138], и R -оценки, основанные на критериях рангов, впервые предложенные Ходжесом и Леманом [128] (см. также [86], [141], [70]).

У п р а ж н е н и я 5.1

1. Дайте точное определение распределения с тяжелыми хвостами.

2. Допустим, наблюдения y_1, y_2, \dots, y_n независимы и нормально распределены, $y_i \sim N(\mu, \sigma_i^2)$, где μ — неизвестный параметр; подлежащий оцениванию, а σ_i известны. Является ли оценка среднего $m = \sum y_i/n$ эффективной оценкой? Допустим, $\sigma_i^2 = \text{const}$, $i = 1; \dots; n - 1$; $\sigma_n^2 \gg \sigma_i^2$. Будет ли оценка m робастной оценкой? Какую оценку можно предложить еще?

5.2. Простейшие методы робастного оценивания регрессии

Допустим возможность присутствия в ряду наблюдений y_1, \dots, y_n выбросов, т. е. наблюдений, удовлетворяющих исходной регрессии и имеющих большие дисперсии либо вообще неудовлетворяющих исходной модели. И в том и другом случаях включение таких наблюдений в ряд равноправных членов выборки y_1, \dots, y_n приведет к заметному смещению оценок параметров и ухудшению их свойств. Как отмечает П. Хьюбер [137], даже одно далеко лежащее от общей массы наблюдение может испортить оценку МНК;

более того, выбросы в случае регрессии приводят к большому искажению, чем в задаче оценивания параметра положения. Выбросы могут быть результатом нарушения условия эксперимента, неправильного измерения, засорения данных и т. п. В случае оценивания параметров положения наиболее простой способ получения робастных, т. е. устойчивых, оценок заключался в отбрасывании ряда экстремальных значений выборки и оценивании параметра положения по усеченной выборке. Этот метод легко обобщается на случай регрессии. Доля отсекаемых экстремальных значений не должна быть очень высокой. Разумеется, если все отброшенные наблюдения в действительности оказались выбросами, то эффективность усеченной оценки только возрастет. Однако если среди отброшенных наблюдений есть и «хорошие», т. е. удовлетворяющие гипотезам классической регрессии, то эффективность новой оценки снизится. Простейшим компромиссом здесь может служить отбрасывание тех наблюдений, которые приводят к максимальному и минимальному отклонениям после оценивания МНК. Полезную информацию при этом дает график отклонений регрессии. Возможно мы и ошибемся, если примем некоторые наблюдения за выбросы, однако в любом случае расчет новой регрессии весьма полезен.

Найдем робастную оценку по указанному правилу для регрессии-примера (1.5). В табл. 1.1 приведены отклонения $e_t = y_t - \hat{y}_t$ после применения МНК. Максимальное отклонение, равное 4,27, соответствует первому, а минимальное — 4,00 — соответствует третьему наблюдению. Подозревая первое и третье наблюдения в выбросах, пересчитаем уравнение регрессии МНК по оставшейся выборке из 13 наблюдений; получим

$$\hat{y}_t = 0,397x_{t1} + 0,310x_{t2} + 3,56x_{t3} - 24,6$$

(0,034) (0,327) (0,997) (21,6)

На рис. 5.3 пунктирной линией показаны отклонения новой регрессии, сплошной — оцененной по МНК. Как видим, в результате отброса экстремальных наблюдений характер отклонений слабо изменился: разброс отклонений уменьшился.

Получение робастных оценок методом исключения выбросов имеет один недостаток. Как было отмечено, оценка МНК резко реагирует на наличие выбросов в исходной информации. Поэтому выявление выбросов с помощью регрессии,

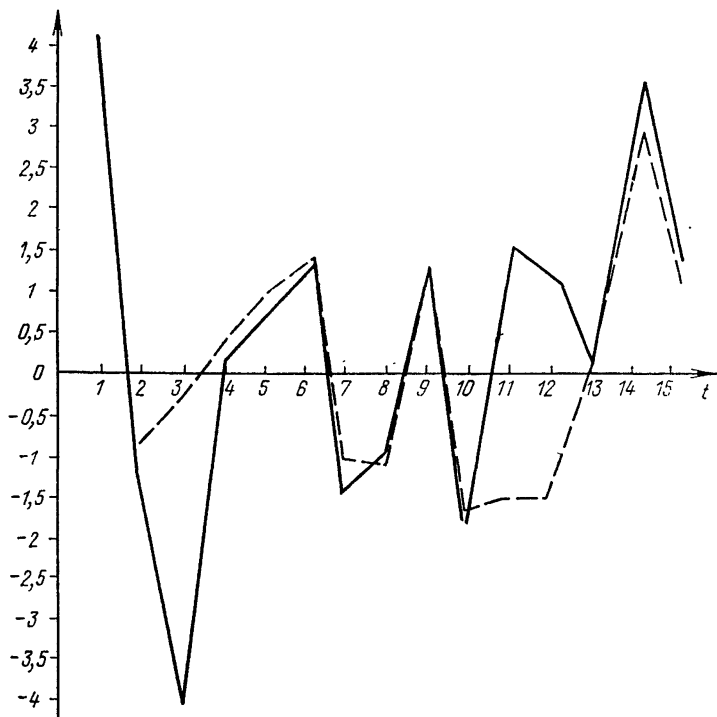


Рис. 5.3. Отклонения регрессии-примера

оцененной методом наименьших квадратов, может привести к тому, что подозреваемые наблюдения окажутся на самом деле «хорошими». Для удовлетворительной «оценки выбросов» необходимо пользоваться оценкой, малочувствительной к ним, т. е. робастной, а ее мы как раз и хотим найти.

5.3. L_ν -оценки

Перейдем к L_ν -оценкам регрессий, являющимся частным случаем оценок ММП Хьюбера. Прежде всего заметим, что функция плотности

$$f(x; \nu) = A e^{-\sigma |x|^\nu}, \quad 0 < \nu < 2 \quad (5.6)$$

имеет более тяжелые хвосты, чем функция плотности нормального распределения $f(x; 2) = A e^{-\sigma x^2}$. При $\nu = 1$ распределение (5.6) называется распределением Лапласа.

Предположим, $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ — независимы, одинаково распределены с функцией плотности (5.6). В силу независимости функция плотности выборки y_1, y_2, \dots, y_n равна:

$$f(y; \nu) = A^n \exp \left[-\sigma \sum_{t=1}^n |y_t - \alpha_1 x_{t1} - \dots - \alpha_m x_{tm}|^\nu \right].$$

Максимум функции f соответствует оценкам ММП. Легко видеть, что оценка ММП минимизирует сумму

$$\begin{aligned} Q_\nu(\alpha) &= \sum_{t=1}^n |y_t - \alpha_1 x_{t1} - \dots - \alpha_m x_{tm}|^\nu = \\ &= \sum_{t=1}^n |\varepsilon_t|^\nu. \end{aligned} \quad (5.7)$$

Оценки, минимизирующие (5.7), назовем L_ν -оценками.

Устойчивость суммы (5.7) относительно больших отклонений по сравнению с соответствующей суммой квадратов отклонений $\sum \varepsilon_t^2$ очевидна. Действительно, допустим, $\nu = 1$, $n = 100$ и основная масса отклонений сосредоточена на отрезке $(-1, 1)$, а одно отклонение равно 3, тогда этот член в сумме квадратов отклонений соответствует 9, а в сумме (5.7) — 3. Поэтому при нахождении оценки в сумме (5.7) это слагаемое не должно произвести значительного эффекта, тогда как при $\nu = 2$ эффект будет весьма существенным. В этом смысле ν можно интерпретировать как фильтр больших отклонений (выбросов).

Остановимся на существовании и единственности минимума функции (5.7). Можно легко убедиться, что функция действительного переменного $\psi(u) = |u|^\nu$ для $\nu \geq 1$ выпукла вниз, для $0 < \nu < 1$ выпукла вверх. Как сумма выпуклых вниз функций, таковой будет и функция $\psi(u_1, \dots, u_n) = \sum_i |u_i|^\nu$, а значит, и (5.7) как функция, совпадающая с ψ на подмножестве R^n — линейном многообразии размерности m . Таким образом, для функции (5.7) при $\nu \geq 1$ существует единственный локальный минимум, который совпадает с глобальным. Для $0 < \nu < 1$ это неверно, и минимизируемая функция может иметь несколько локальных минимумов (см., например, [51]).

При $\nu = 1$ необходимо минимизировать сумму абсолютных модулей отклонений (невязок)

$$\sum_{t=1}^n |\varepsilon_t|. \quad (5.8)$$

Минимизация (5.8) сводится к задаче линейного программирования (см., например, [192], [187]). Решение ее намного упростится, если перейти к двойственной задаче линейного программирования. Число ограничений задачи будет равно числу оцениваемых параметров, т. е. m , [192]. Таким образом, при нахождении параметров регрессии, которые минимизируют сумму абсолютных отклонений, можно воспользоваться стандартными программами линейного программирования. В [31] предложен простой итеративный метод минимизации (5.8), который основан на том, что регрессия, минимизирующая (5.8), проходит через $m+1$ точек выборки $(y_t, x_{t1}, \dots, x_{tm})$, $t = 1, \dots, n$.

Для нахождения минимума функции $Q_v(\alpha)$ для $v \geq 1$ можно, разумеется, применить общие методы оптимизации: градиентный, метод Ньютона, метод сопряженных градиентов и т. д. Однако существует более простой метод минимизации $Q_v(\alpha)$, опирающийся на обычный МНК. Этот метод впервые был предложен Р. Флетчером, Дж. Грантом и Х. Хебленом [102] и получил название «итеративного МНК». Идея его заключается в следующем. Найдем частные производные функции $Q_v(\alpha)$ по параметру α и приравняем их к нулю. Поскольку $v \geq 1$, то решение соответствующей системы уравнений приведет к точке глобального минимума. Имеем

$$\partial Q_v / \partial \alpha_i = -v \sum_t \text{sign}(\varepsilon_t) |\varepsilon_t|^{v-1} x_{ti} = 0, \quad i = 1, \dots, m,$$

откуда

$$\sum_t \varepsilon_t x_{ti} |\varepsilon_t|^{v-2} = \sum_t \varepsilon_t x_{ti} \omega_t = 0,$$

где $\omega_t = |\varepsilon_t|^{v-2}$. Данная система приводит нас к системе линейных нормальных уравнений относительно $\alpha_1, \dots, \alpha_m$:

$$\alpha_1 \sum_t x_{t1}^2 \omega_t + \dots + \alpha_m \sum_t x_{t1} x_{tm} \omega_t = \sum_t y_t x_{t1} \omega_t$$

(5.9)

$$\alpha_1 \sum_t x_{t1} x_{tm} \omega_t + \dots + \alpha_m \sum_t x_{tm}^2 \omega_t = \sum_t y_t x_{tm} \omega_t$$

Нетрудно заметить, что эта система линейных уравнений соответствует схеме взвешенного МНК с весами ω_t , при этом веса могут быть «оценены» на основе параметров, полученных из предыдущей итерации. Таким образом, на

нулевой итерации оцениваем регрессию каким-либо методом (например, МНК), получаем вектор оценок a^0 . Исходя из этого вектора «оцениваем» веса

$$\omega_t = |e_t|^{v-2} = |y_t - a_1^0 x_{t1} - \dots - a_m^0 x_{tm}|^{v-2}.$$

Решая систему линейных уравнений (5.9), т. е. применяя взвешенный МНК с весами ω_t , находим следующее значение вектора оценок a^1 и т. д. Этот же способ предлагают В. И. Мудров и В. Л. Кушко [51], называя его методом вариационно-взвешенных квадратических приближений. Аргументация применения итеративного МНК здесь намного проще. Сумму (5.7) перепишем следующим образом:

$$\sum_t |e_t|^v = \sum_t \varepsilon_t^2 |e_t|^{v-2} = \sum_t \varepsilon_t^2 \omega_t,$$

что опять приводит к взвешенному МНК с весами ω_t . В [51] доказана сходимость итеративного МНК для $1 \leq v < 2$.

Найдем оценки по этому методу для $v = 1$ и 1,5 нашей регрессии-примера (табл. 5.1). Метод сходится уже на первой итерации. Это говорит о высокой эффективности его. Наиболее мобильной оказалась оценка второго параметра. Легко заметить также, что для $v = 1,5$ оценки ближе к оценкам МНК, чем для $v = 1$, что вполне естественно. Даже при уменьшении чувствительности к большим отклонениям оценка параметра α_1 увеличивается, а α_2 и α_3 уменьшаются (по абсолютной величине).

Таблица 5.1

v	Номер итерации	α_1	α_2	α_3	α_4
1	0	0,395	2,229	3,747	-17,14
	1	0,399	0,171	3,749	-12,42
	2	0,399	0,171	3,750	-12,42
1,5	0	0,395	0,229	3,747	-17,14
	1	0,397	0,200	3,748	-14,80
	2	0,397	0,200	3,748	-14,80

Иногда полезно выяснить, что происходит с оценкой, если, наоборот, ориентироваться на большие отклонения, в частности найти оценку, которая минимизирует максимальное отклонение. Такую оценку вправе назвать антиробастной. Подобным оценкам отвечают большие значения v ,

т. е. $\nu \geq 2$. Остановимся на минимизации (5.7) для $\nu > 2$. При этих значениях Мудрову и Кушко не удалось доказать сходимость итеративного МНК. Однако Флетчером, Грантом и Хебленом было предложено в этом случае поправку к новому вектору оценок брать не полностью, а только ее часть [102]. Пусть \mathbf{a}^k — значение вектора оценок на k -й итерации. Применяя итеративный МНК, можно найти следующий вектор \mathbf{b}^{k+1} , при этом поправка $\mathbf{d}^{k+1} = \mathbf{b}^{k+1} - \mathbf{a}^k$. Следующее значение \mathbf{a} полагается равным:

$$\begin{aligned} \mathbf{a}^{k+1} &= \mathbf{a}^k + \frac{1}{\nu-1} \mathbf{d}^k = \left(1 - \frac{1}{\nu-1}\right) \mathbf{a}^k + \\ &+ \frac{1}{\nu-1} \mathbf{b}^{k+1} = (1-\gamma) \mathbf{a}^k + \gamma \mathbf{b}^{k+1}, \end{aligned} \quad (5.10)$$

где $\gamma = 1/(1-\nu)$. Оказалось, что модифицированный итеративный МНК уже является сходящимся с квадратичной скоростью сходимости для $\nu \geq 3$. Там же показано, что метод (5.10) совпадает с методом Ньютона.

В табл. 5.2 приведены L_ν -оценки регрессии-примера для $\nu = 3, 4, 10, 20$. Отметим равномерное снижение значений оценки параметра α_1 . L_ν -оценки для $\nu = 10$ и 20 параметров $\alpha_2, \alpha_3, \alpha_4$ сильно отличаются от предыдущих значений. Число итераций, необходимых для получения L_ν -оценок при $\nu = 10$ и 20, намного больше, чем для $\nu = 3$ и 4.

Таблица 5.2

ν	Общее число итераций	α_1	α_2	α_3	α_4
3	3	0,384	0,298	3,84	-21,8
4	3	0,380	0,269	4,08	-19,9
10	17	0,344	0,000605	6,19	-2,311
20	16	0,335	-0,0959	6,82	4,04

К идее минимизации суммы (5.7) можно подойти с другой точки зрения. Естественным методом оценивания параметров регрессии является следующий. Обозначим оценку параметров \mathbf{a} , тогда $\mathbf{X}\mathbf{a}$ является оценкой, приближением вектора данных (выборки) $\mathbf{y} \in R^n$. Поэтому минимизация расстояния между \mathbf{y} и $\mathbf{X}\mathbf{a}$ приведет нас к соответствующей оценке. Метод оценивания будет различным в зависимости от того, как мы будем измерять расстояние в R^n . В частности, если брать евклидово расстояние, то придем

к МНК. Большую группу составляют L_ν -метрики, где расстояние между $z, y \in R^n$ задается по формуле

$$\|z - y\| = \left(\sum_i |z_i - y_i|^\nu \right)^{1/\nu}, \quad \nu > 0.$$

Если $\nu = 1$, приходим к минимизации (5.8), в общем случае получаем L_ν -оценки.

В. Хоганом была доказана несмещенность L_ν -оценки при некоторых условиях на распределение вектора $(\epsilon_1, \dots, \epsilon_n)'$ [135]. Условие регулярности, которое накладывается на распределение отклонений, является следующим: математическое ожидание ϵ при условии, что $\epsilon \in L$ — линейному подпространству R^n , равно нулю для любого L , короче говоря, $E(\epsilon/\epsilon \in L) = 0$. Хоган доказал, что если ϵ удовлетворяют условию регулярности, то L_ν -оценка для $\nu \geq 1$ несмещена, т. е. $Ea = \alpha$, $a \in L_\nu$. Проанализируем условие регулярности Хогана. Ясно, что для $n = 1$ оно совпадает с условием $E\epsilon = 0$. Для $n > 1$ это условие сильнее $E\epsilon = 0$. Однако нетрудно проверить, что если $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ — независимы и одинаково распределены с симметричной функцией плотности, то ϵ удовлетворяет условию регулярности. Действительно, тогда функция плотности ϵ равна $f(\epsilon_1), \dots, f(\epsilon_n)$ и также симметрична, значит и $E(\epsilon/L) = 0$ для любого $L \subset R^n$. Таким образом, L_ν -оценки с распределениями Коши, Лапласа, Гаусса (нормальное), равномерным и т. д. будут несмещенными.

У п р а ж н е н и я 5.3

1. Почему выброс в случае регрессии может привести к большим искажениям оценки МНК, чем в случае оценивания параметра положения?

2. Пусть в регрессии $\sigma^2(\epsilon_t) = \sigma^2$, $t = 1; \dots; n - 1$; $\sigma^2(y_n) = \sigma_n^2 \gg \sigma^2$. В каком случае оценка МНК, построенная на выборке $t = 1, \dots; n - 1$, лучше оценки МНК, построенной по всей выборке $t = 1, 2, \dots, n$?

3. Покажите, что найдется такое u_0 , что $P_{\text{Лапласа}}(\xi \geq u) > P_{\text{норм}}(\eta \geq u)$ для всех $u \geq u_0$.

5.4. Оценки Хюбера, Андруса и Рамсея

Каждой M -оценке в случае оценивания параметра положения соответствуют две оценки: минимизирующая сумму (5.5) и минимизирующая сумму (5.2). Введение параметра масштаба σ в случае L_ν -оценок не играет роли: обе оценки совпадают. Однако в других случаях это не так.

Для того чтобы полученные M -оценки были робастными, необходимо, чтобы на интервале, содержащем 0, т. е. на интервале $(-c, c)$, функция ρ была близка к параболе (пропорциональна квадрату аргумента), а вне интервала скорость роста ρ заметно снижалась и зависимость ρ от x становилась меньше. Так, в функции Хюбера (5.3) на интервале $(-c, c)$ $\rho(x)$ совпадает с квадратичной функцией (скорость роста пропорциональна аргументу), а вне этого интервала — равна линейной функции (скорость роста постоянна). Значение c , т. е. то значение аргумента, при котором происходит уменьшение скорости роста ρ , исследователю неизвестно и его тоже приходится каким-либо образом оценивать. Если исследователь имеет представление о возможных отклонениях, то в качестве c можно выбрать критическое значение «нормального» отклонения, начиная с которого вклад в минимизируемую сумму должен измеряться не квадратичной, а менее чувствительной функцией. Если же априорное значение о возможных отклонениях отсутствует, то имеет смысл воспользоваться робастной оценкой с привлечением параметра масштаба σ . Тогда отклонения измеряются в стандартных единицах. Используя σ , можно, например, предложить следующее правило определения c . Пусть в функции ρ c — значение аргумента, при котором квадратичность ρ переходит в функцию меньшего роста. Полагаем $c = 3s$, где s — оценка σ . Предложенное правило напоминает известное правило «трех сигм».

Итак, пусть $\rho(x)$ — непрерывная, кусочно дифференцируемая, симметричная относительно нуля, возрастающая на $(0, \infty)$ функция, $\rho(0) = 0$. M -оценкой назовем вектор, обращающий сумму

$$\sum_i \rho(\varepsilon_i) = \sum_i \rho(y_i - \alpha_1 x_{i1} - \dots - \alpha_m x_{im} / \sigma) \quad (5.11)$$

в минимум, где σ — параметр масштаба, также подлежащий оцениванию. По сути дела, (5.11) определяет целый класс оценок; каждая оценка зависит от выбора функции ρ . Обозначим $\rho' = \varphi$ — производную функцию ρ . Вместо минимизации суммы (5.11) можно искать решение уравнений

$$\sum_i \varphi\left(\frac{y_i - \alpha_1 x_{i1} - \dots - \alpha_m x_{im}}{\sigma}\right) x_{ti} = 0, \quad i = 1, \dots, m. \quad (5.12)$$

Для оценивания σ выберем простейшую статистику: s равно медиане $\{e_i\}$, $e_i = y_i - \alpha_1 x_{i1} - \dots - \alpha_m x_{im}$.

При минимизации (5.11) в принципе можно воспользоваться общими методами минимизаций функций многих переменных (градиентный метод, метод Ньютона и др.). Однако трудно обобщить известный нам итеративный МНК и на более общий случай (5.12). Если же для функции $\rho(x) = |x|^v$ подходы Флетчера — Гранта — Хеблена и Мудрова — Кушко совпадают, то для произвольной функции они приведут, вообще говоря, к разным итерационным процессам.

Итеративный МНК Флетчера — Гранта — Хеблена. Преобразуем уравнения (5.12) следующим образом:

$$\sum_t \varphi(\varepsilon_t/\sigma) x_{ti} = \sum \varepsilon_t x_{ti} w_{t1},$$

где веса равны:

$$w_{t1} = \varphi(\varepsilon_t/\sigma)/\varepsilon_t, \quad t = 1, \dots, n. \quad (5.13)$$

Итерации производятся так же, как в итеративном МНК.

Метод вариационно-взвешенных квадратических приближений Мудрова — Кушко. Функция (5.11) сводится к сумме квадратов отклонений:

$$\begin{aligned} \sum_t \rho(\varepsilon_t/\sigma) &= \sum_t \varepsilon_t^2 [\rho(\varepsilon_t/\sigma)/\varepsilon_t^2] = \sum_t \varepsilon_t^2 w_{t2}; \\ w_{t2} &= \rho(\varepsilon_t/\sigma)/\varepsilon_t^2, \quad t = 1, \dots, n. \end{aligned} \quad (5.14)$$

Ясно, что в общем случае веса w_{t1} (5.13) и w_{t2} (5.14) будут отличны друг от друга, что приведет и к разным итерационным процессам. Однако и в том и в другом случаях мы должны получить одинаковые пределы для оценок параметров. Легко проверить, что для $\rho(x) = |x|^v$ $w_{t1} = w_{t2}$.

Рассмотрим робастные оценки ММП, обращающие сумму (5.11) в минимум для разных ρ .

Оценка Хьюбера (5.3). Правило «трех сигм» предлагает c взять равным 3. Использование итеративного МНК Флетчера — Гранта — Хеблена для регрессии-примера не привело к оценке, процесс оказался расходящимся. С другой стороны, метод Мудрова — Кушко (вернее, его обобщение) сошелся уже на второй итерации. Полученные оценки мало отличаются от оценок МНК. Вектор оценок равен (0,395; 0,228; 3,75; — 17,12). Оценка параметра масштаба равна 1,33.

Оценка Андруса [73]. Автор обобщает M -оценку, введенную им для оценивания параметра положения на случай

регрессии. В качестве ρ Андриус предлагает следующую функцию (рис. 5.4):

$$\rho(x) = \begin{cases} c(1 - \cos(x/c)), & |x| \leq \pi c; \\ 2c, & |x| > \pi c. \end{cases} \quad (5.15)$$

Идея заключается в том, что на интервале $(-\pi c/2, \pi c/2)$ функция ρ близка к квадратичной, на интервалах $(\pi c/2, \pi c)$ и $(-\pi c, -\pi c/2)$ скорость роста функции (5.15) уменьшается. Для $|x| > \pi c$ вклад отклонений не зависит от

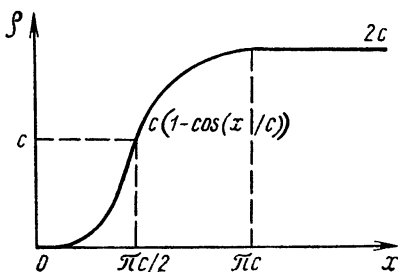


Рис. 5.4. Функция Андриуса для $x > 0$

их величины и равен $2c$. Используя правило «трех сигм», найдем значение c . Робастные свойства функции (5.15) проявляются для значений $|x| \geq \pi c/2$, причем в точках $\pm \pi c/2$ вторая производная (5.15) обращается в нуль. Таким образом, полагаем $\pi c/2 = 3$, откуда $c = 6/\pi \approx 1,91$. Использо-

вание метода Мудрова — Кушко после пяти итераций в регрессии-примере привело к оценке $(0,398; 0,212; 3,71; -16,1)$, которая также незначительно отличается от оценки МНК.

Оценка Рамсея [173] основана на функции

$$\rho(x) = \frac{1}{\gamma^2} [1 - (1 + \gamma|x|)e^{-\gamma|x|}], \quad \gamma > 0. \quad (5.16)$$

Легко показать, что при $\gamma \rightarrow 0$ функция Рамсея переходит в квадратичную. Для малых значений $|x|$ функция (5.16) близка к функции $|x|$. При $x \rightarrow \infty$ ρ имеет асимптоту, равную $1/\gamma^2$. Найдем значение γ , используя правило «трех сигм». Для этого найдем вторую производную функции Рамсея и приравняем ее к нулю: $\rho''(x) = e^{-\gamma|x|} (1 - \gamma|x|) = 0$, откуда $x = \pm 1/\gamma$ — точки перегиба функции (5.16). Полагаем $1/\gamma = 3$, откуда $\gamma = 1/3$. После восьми итераций по методу Мудрова — Кушко для регрессии-примера была получена следующая оценка параметров: $(0,404; 0,159; 3,73; -12,6)$; $s = 1,21$. Все три функции привели к параметрам, мало отличающимся от оценок МНК. Это говорит о том, что в регрессии-примере нет ярко выраженных выбросов, поэтому оценка МНК вполне приемлема.

5.5. Сравнение оценок методом статистических испытаний

Сравнение начнем с исследования свойств L_ν -оценок. В [176] исследована эффективность оценки, минимизирующей (5.7) для разных ν и разных распределений отклонений. Рассмотрено семь альтернативных распределений: равномерное, квадратный корень¹, треугольное², квадратное³, нормальное, Лапласа, Коши. Результаты статистических испытаний представлены графически на (рис. 5.5). Анализ графиков а) — ж) показывает, что если распределение близко к равномерному, то более оптимальной будет оценка, соответствующая норме L_∞ , т. е. для которой

$$\|x\| = \max_i |x_i|,$$

что приводит к критерию

$$\min_{\alpha} \max_t |\varepsilon_t|, \quad (5.17)$$

где $\varepsilon_t = y_t - \alpha_1 x_{t1} - \dots - \alpha_m x_{tm}$.

Задача (5.17) сводится к задаче линейного программирования [192]. Для распределений с конечной дисперсией и легкими хвостами (типа нормального) наиболее эффективной будет оценка МНК ($\nu = 2$). Для распределений с тяжелыми хвостами оптимальная оценка достигается для значений ν , близких к 1, причем чем тяжелее хвост, тем ближе ν к 1. Вывод, к которому приходят авторы, следующий: оптимальный выбор ν зависит от распределения ε . Нельзя заранее сказать, какое значение ν приведет к эффективной оценке. В случае выбросов хорошо зарекомендовала себя оценка, минимизирующая сумму абсолютных отклонений (5.8).

В работе [103] отклонения имеют смешанное нормальное распределение с плотностью

$$f(x; G, R, S) = G \frac{1}{\sqrt{2\pi}} e^{-x^2/2} + (1-G) \frac{1}{R\sqrt{2\pi}} e^{-(x-S)^2/2R^2},$$

¹ Функция плотности равна $2\sqrt{1-x^2}/\pi$ для $|x| \leq 1$ и 0 для $|x| > 1$.

² Функция плотности равна $1 - \sqrt{2}|x|$ для $|x| \leq 1/\sqrt{2}$ и 0 для $|x| > 1/\sqrt{2}$.

³ Функция плотности равна $3(x - \sqrt{3}/2)^2/2$ для $-\sqrt{3}/2 \leq x \leq -1/2\sqrt{3}$, $3/4 - 3x^2$ для $|x| < 1/2\sqrt{3}$ и $3(x + \sqrt{3}/2)^2/2$ для $1/2\sqrt{3} \leq x \leq \sqrt{3}/2$ и 0 для $|x| \geq \sqrt{3}/2$.

где $0 \leq G \leq 1$ — параметр смеси; R — стандартное отклонение второй смеси; S — математическое ожидание второй смеси. Если $S \neq 0$, распределение будет асимметричным. Если $R \gg 1$, то $1 - G$ есть доля выбросов в распределении с плотностью f . В [103] оценка МНК для парной регрес-

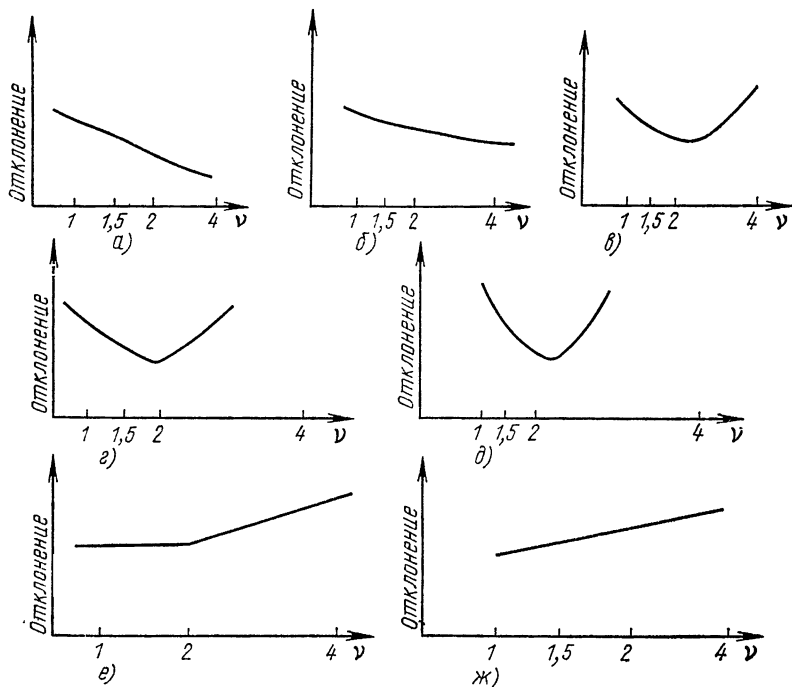


Рис. 5.5. Зависимость средних абсолютных отклонений оценок, минимизирующих $\sum |\varepsilon_i|^\nu$, от ν для разных распределений: а) равномерное; б) квадратный корень; в) треугольное; г) квадратное; д) нормальное; е) Лапласа; ж) Коши

сии со свободным членом $y_i = \alpha_1 x_i + \alpha_2 + \varepsilon_i$ сравнивается с L_ν -оценкой для $\nu = 1,25; 1,5$ и $1,75$. Величины относительных отклонений $\sum (\alpha_1 - \alpha_1)^\nu / \sum (\alpha_1^\nu - \alpha_1)^\nu \cdot 100\%$, рассчитанные по методу Монте-Карло, приведены в табл. 5.3.

Приведенная таблица позволяет сделать следующие выводы:

1. Даже если распределение отклонений нормально, но загрязнено другим нормальным распределением с большей дисперсией или с м.о., не равным нулю, эффективность оценки МНК значительно падает,

Таблица 5.3

ν	$S=0, R=4$ при $G=$					$S=4, R=1$ при $G=$				
	1,000	0,975	0,950	0,925	0,900	1,000	0,975	0,950	0,925	0,900
1,25	90	105	115	145	170	90	100	120	125	140
1,50	95	115	120	150	160	95	100	125	125	135
1,75	100	110	115	130	135	100	105	115	115	115
2,00	100	100	100	100	100	100	100	100	100	110

2. Оценка L_ν , соответствующая, например, $\nu = 1,5$ в случае незагрязненного нормального распределения (идеальный случай), теряет лишь 5% эффективности по сравнению с оценкой МНК, зато для загрязненных распределений выигрыш по сравнению с оценкой МНК доходит до 50%.

Таким образом, оценка $L_{1,5}$ является хорошим компромиссом между классической оценкой МНК и робастными оценками.

В работе [173] Дж. Рамсей методом Монте-Карло сравнивает эффективность различных методов оценивания парной регрессии $y_t = \alpha_1 x_t + \alpha_2 + \varepsilon_t$. Он рассмотрел три группы M -оценок Хьюбера: E_ν -оценки Рамсея ($\nu = 0,1; 0,3; 1,0$), оценку Андруса и L_ν -оценки ($\nu = 2,0; 1,8; 1,5; 1,0$). Отклонения регрессии ε_t были независимы с распределением, представляющим смесь нормальных распределений $N(0,1)$ и $N(0, \sigma^2)$ с параметром смеси q . Другими словами, распределение ε_t было равно $(1-q)F_1 + qF_2$, где F_1 — функция распределения $N(0,1)$, F_2 — то же $N(0, \sigma^2)$. Значение n выбиралось равным 5, 20 и 50, значение q — равным 0; 0,01; 0,05; 0,1; 0,25, значение σ — равным 3 и 10. Основные выводы, к которым пришел автор, следующие:

а) оценки E_1 и L_1 оказались неудовлетворительными из-за низкой эффективности для $q < 0,25$;

б) оценка МНК (L_2) также оказалась неудовлетворительной из-за низкой эффективности при высоких значениях q ;

в) оценки $E_{0,1}$ и $L_{1,8}$ близки к оценке МНК, поэтому их нельзя считать в полной мере робастными; при высоких значениях q их эффективность резко падала;

г) оценка $E_{0,5}$ оказалась слишком плохой в случае нормального распределения ($q = 0$);

д) оценки $E_{0,3}$ и $L_{1,5}$ показали себя одинаково хорошо; их эффективность не падала слишком низко при больших значениях q и была достаточно высокой для q , близких к нулю; однако почти во всех вариантах оценки $E_{0,3}$ были лучше $L_{1,5}$;

е) оценки $E_{0,3}$ оказались более устойчивыми по отношению к априорному распределению отклонений регрессии, т. е. более робастными по сравнению с оценкой Андруса, однако оценка Андруса более эффективна для нормального распределения ($q = 0$).

Таким образом, автор [173] заключает, что равномерно эффективными оценками оказались оценки $E_{0,3}$, оценка Андруса и отчасти оценка $E_{0,5}$. В исследовании, проведенном Рамсеем, L_q -оценки оказались в большинстве случаев хуже E_q -оценок и оценки Андруса.

Глава 6

МУЛЬТИКОЛЛИНЕАРНОСТЬ. СМЕЩЕННЫЕ ОЦЕНКИ

6.1. Мультиколлинеарность и ее измерение

Мультиколлинеарность — одно из основных препятствий эффективного применения аппарата регрессионного анализа. Что такое «мультиколлинеарность» и в чем ее смысл? Как и чем измерять мультиколлинеарность? Попытаемся ответить на поставленные вопросы. Под *мультиколлинеарностью* в дальнейшем будем понимать *сопряженность независимых переменных*¹. Мультиколлинеарность обычно называют сильной (harmfull), если оценки параметров или проверки гипотез зависят скорее от взаимозависимости независимых переменных модели регрессии, чем от зависимости y и x_1, x_2, \dots, x_m . Д. Фаррар и Р. Глаубер [100] считают мультиколлинеарность сильной, если коэффициент корреляции одной из пар независимых переменных больше коэффициента корреляции регрессии. Обозначим через x_i i -й вектор-столбец матрицы независимых переменных X ; тогда мультиколлинеарность означает «почти линейную за-

¹Мы употребляем термин «сопряженность независимых переменных» вместо более распространенного «коррелируемость», тем самым подчеркивая детерминированность независимых переменных регрессии.

зависимость» векторов x_1, \dots, x_m , т. е. существование чисел v_1, \dots, v_m , таких, что

$$v_1 x_1 + v_2 x_2 + \dots + v_m x_m \approx 0, \quad (6.1)$$

Чем ближе левая часть (6.1) к нулевому вектору из R^n , тем сильнее мультиколлинеарность. Предельный случай соответствует точному равенству в (6.1). Тогда говорим о *строгой мультиколлинеарности*. Этот случай разобран в следующем параграфе. Сейчас же мы по-прежнему считаем, что строгой линейной зависимости между вектор-столбцами матрицы X не существует, т. е. $\text{rang } X = m$, однако имеет место приближенная зависимость (6.1). Трудность установления факта мультиколлинеарности связана с тем, что на практике равенство (6.1) никогда не бывает точным. В первую очередь этому мешает наличие, может быть, с практической точки зрения незначительных ошибок измерения x_1, \dots, x_m . Точное равенство может отсутствовать также из-за ошибок округления. Поэтому мы согласны с Фарраром, Глаубером [100], которые говорят, что мультиколлинеарность есть не вопрос существования, а вопрос степени.

Чем опасна мультиколлинеарность? Для регрессионного анализа она опасна тем, что оценки МНК становятся малоэффективными, т. е. дисперсия оценок будет весьма большой.

Для наглядности рассмотрим случай $m = 2$. Сначала предположим, что x_1 и x_2 имеют одинаковый масштаб измерения. Этого можно добиться нормировкой векторов x_1 и x_2 , т. е. рассмотреть $x'_1 = x_1/\|x_1\|$, $x'_2 = x_2/\|x_2\|$. В таком случае матрица плана является матрицей сопряженности, т. е.

$$X'X = \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix},$$

где

$$r_{12} = \cos(x_1, x_2) = \sum x_{i1} x_{i2} / \sqrt{\sum x_{i1}^2 \cdot \sum x_{i2}^2}.$$

Тогда

$$(X'X)^{-1} = \begin{bmatrix} 1/(1-r_{12}^2) & -r_{12}/(1-r_{12}^2) \\ -r_{12}/(1-r_{12}^2) & 1/(1-r_{12}^2) \end{bmatrix}$$

и

$$\sigma_1^2 = \sigma^2(a_1) = \sigma^2/(1-r_{12}^2) = \sigma_2(a_2) = \sigma_2^2$$

есть дисперсия оценок МНК для α_1 и α_2 в регрессии $y = \alpha_1 x_1 + \alpha_2 x_2 + \varepsilon$. Если между переменными x_1 и x_2 существует тесная линейная зависимость, т. е. r_{12}^2 близок к 1, то, как следует из последнего выражения, дисперсии $\sigma_1^2 = \sigma_2^2$ оценок МНК будут иметь большие значения. В общем случае при наличии (6.1) матрица $X'X$ становится плохо обусловленной, в частности $|X'X| \approx 0$, т. е. $\lambda_{\min}(X'X) = \lambda_1 \approx 0$.

В качестве критерия качества оценки выберем сумму квадратов дисперсий

$$\sum_{i=1}^m \sigma_i^2 = \sigma^2 \operatorname{tr}(X'X)^{-1} = \sigma^2 \sum_{i=1}^m 1/\lambda_i. \quad (6.2)$$

Если $\lambda_1 \approx 0$, то значение суммы (6.2) будет велико. В частности, если $\lambda_1 \rightarrow 0$, то $\sum \sigma_i^2 \rightarrow \infty$.

Остановимся на вопросе измерения мультиколлинеарности. Вообще, мультиколлинеарность — понятие достаточно многогранное, и трудно предложить меру, которая была бы во всех отношениях хороша. Рассмотрим пять различных мер. Первые три из них — характеристики матрицы плана $X'X$.

а. Определитель матрицы плана $X'X$. Поскольку при наличии приближенной линейной зависимости (6.1) матрица $X'X$ становится плохо обусловленной, т. е. близкой к вырожденной (имеющей нулевой определитель), то определитель $|X'X| = \prod_{i=1}^m \lambda_i$ может выступать в качестве меры мультиколлинеарности. При рассмотрении проблемы мультиколлинеарности большую наглядность дает геометрический подход, который возможен для $m = 2, 3$. В дальнейшем мы часто будем изображать ту или иную оценку геометрически для случая $m = 2$. Особенно нагляден при геометрическом подходе *характеристический эллипс* (эллипсоид при $m = 3$). Как известно, геометрическим местом точек, составляющих доверительное множество при $m = 2$, в случае нормального распределения отклонений регрессии является внутренность эллипса

$$S_\gamma = \{ \alpha \in R^m : (\alpha - a)' X'X (\alpha - a) = \gamma \}, \quad (6.3)$$

где $a = (a_1, a_2)'$ — оценка МНК¹. Нетрудно показать, что S_γ является уровнем суммы квадратов отклонений $Q(\alpha)$.

¹Э. Маленко [48] определяет характеристический эллипс несколько по-иному.

Характеристические числа матрицы $X'X$ суть длины полуосей эллипса S_γ , а характеристические векторы матрицы плана — направления соответствующих осей эллипса. Центр эллипса находится в точке плоскости, отвечающей оценке МНК. Характеристический эллипс (6.3) является геометрическим представителем матрицы $X'X$. Выбор γ не имеет решающего значения, так как форма и положение эллипса от него не зависят.

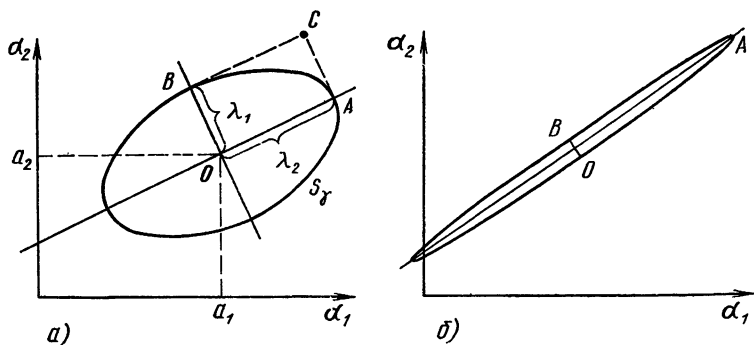


Рис. 6.1. Характеристический эллипс регрессии: а) мультиколлинеарность отсутствует; б) мультиколлинеарность

Можно показать, что $|X'X|$ пропорциональна площади внутренности характеристического эллипса S_γ . Поэтому малость $|X'X|$ означает малость площади, охватываемой S_γ . На рис. 6.1, а показан некоторый характеристический эллипс; O — центр эллипса имеет координаты (a_1, a_2) — оценка МНК. Длина отрезка OA — большая полуось, равна $\lambda_2 > \lambda_1$ — длины отрезка OB — малая полуось эллипса. Вектор OA совпадает с направлением х.в. матрицы $X'X$, отвечающим λ_2 , вектор OB — с направлением х.в., отвечающим λ_1 . Определитель $|X'X|$ численно равен площади прямоугольника $OBCA$. На рис. 6.1, б показан характеристический эллипс другой регрессионной задачи. Этот эллипс более вытянут в одном направлении и сжат в другом. При этом минимальное х.ч. уменьшилось в 4 раза, а максимальное х.ч. увеличилось в 4 раза. Ясно, что в случае б) мультиколлинеарность сильнее, но выбранный критерий $|X'X|$ приводит к одному и тому же значению. Форма характеристического эллипса позволяет сделать вывод, насколько «идентифицируемы» параметры α_1 и α_2 . Если эллипс сильно

вытянут в одном направлении и сжат в другом, то можно утверждать, что параметры слабо различимы. Грубо говоря, параметры приближенно линейно зависимы, т. е. плохо идентифицируемы: при оценивании может произойти «перелив» из одного параметра в другой. Так, для рис. 6.1, б мы можем грубо записать: $\alpha_2 - \alpha_1 - 1 \approx 0$.

б. *Минимальное характеристическое число матрицы плана.* Очевидно, чем меньше $\lambda_1 = \lambda_{\min}(\mathbf{X}'\mathbf{X})$, тем сильнее мультиколлинеарность (см. теорему 6.1). На λ_1 не оказывают влияния другие характеристические числа матрицы $\mathbf{X}'\mathbf{X}$. Имеется еще одно веское обстоятельство использования величины λ_1 как показателя мультиколлинеарности. Обозначим левую часть (6.1) через w . Если в качестве приближения w к 0 взять квадрат евклидова расстояния $\|w - 0\|^2 = \|w\|^2$, то, переходя к матричной форме, получим $\|w\|^2 = \|v_1 x_1 + \dots + v_m x_m\|^2 = \mathbf{v}'\mathbf{X}'\mathbf{X}\mathbf{v}$, откуда

$$\min_{\|v\|=1} \|w\|^2 = \lambda_{\min}(\mathbf{X}'\mathbf{X}) = \lambda_1.$$

Вектор $\mathbf{v} = (v_1, \dots, v_m)'$, соответствующий минимальному характеристическому числу λ_1 , дает тот набор коэффициентов, который приводит к максимально приближенной к нулю линейной комбинации векторов x_1, \dots, x_m .

О другом положительном качестве λ_1 мы уже упоминали. Оно связано с непосредственным влиянием λ_1 на точность оценивания регрессии МНК (6.2). Помимо того, что λ_1 отражает взаимную сопряженность независимых переменных x_1, \dots, x_m , она несет на себе эффект выбора масштаба измерения этих переменных. Действительно, пусть, например, x_1, \dots, x_m измерены в рублях, и мы решили перейти в другие единицы, например млн. рублей, тогда $\mathbf{X}^* = \delta \mathbf{X} = 10^{-6} \mathbf{X}$. Поэтому

$$\lambda_{\min}(\mathbf{X}^{*'}\mathbf{X}^*) = 10^{-12} \lambda_{\min}(\mathbf{X}'\mathbf{X}).$$

Если судить по величине минимального х. ч. матрицы плана, то преобразованная модель будет мультиколлинеарна. Полезно будет рассмотреть модель, в которой отсутствует сопряженность независимых переменных (ортогональная модель), а масштаб измерения x_1, \dots, x_m одинаков. Математически это означает: $\mathbf{X}'\mathbf{X} = \delta \mathbf{I}_m$, где $\delta > 0$. Тогда $\lambda_1 = \dots = \lambda_m = \delta$ и оценка МНК равна $\mathbf{a} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{y}/\delta$, т. е.

$$a_i = d_i/\delta, \quad d_i = (\mathbf{X}'\mathbf{y})_i, \quad i = 1, \dots, m.$$

Найдем t -статистику для каждого параметра

$$t_i = \frac{|a_{il}|}{s \sqrt{(X'X)^{-1}_{ii}}} = \frac{|d_{il}|}{\delta} \cdot \frac{d}{s} = \frac{|\alpha_{il}|}{s}, \quad i = 1, \dots, m. \quad (6.4)$$

Таким образом, даже если $\delta \approx 0$, т. е. сумма дисперсий (6.2) будет иметь большое значение, относительная точность оценивания (t -статистика (6.4)) может быть достаточно высока (например, больше 2). Поэтому мера λ_1 есть показатель качества оценивания в абсолютном смысле.

в. Мера обусловленности матрицы по Нейману — Голдстейну [165]. Дж. Нейман и Х. Голдстейн, исследуя методы обращения матриц, заметили, что удобной характеристикой вырожденности матрицы является отношение максимального х.ч. к минимальному. Эту меру можно предложить и для измерения мультиколлинеарности. Геометрически отношение $\lambda_{\max}(X'X)/\lambda_{\min}(X'X)$ означает, насколько сжат характеристический эллипсоид в одном направлении и вытянут в другом. А чем отличается новая мера от предыдущей? Во-первых, так же, как и λ_1 , отношение λ_m/λ_1 может нести на себе эффект сильной сопряженности независимых переменных. Однако этого может и не быть. Тогда λ_1 и λ_m/λ_1 несут на себе эффект выбора масштаба. Однако если λ_1 зависит от масштаба измерения x_1, \dots, x_m одновременно для всех переменных, то λ_m/λ_1 отражает разницу в масштабах. Можно показать, что

- 1) $\lambda_1 \rightarrow 0$ не влечет $\lambda_m/\lambda_1 \rightarrow \infty$ и не влечет $|X'X| \rightarrow 0$;
- 2) $\lambda_m/\lambda_1 \rightarrow \infty$ не влечет $\lambda_1 \rightarrow 0$ и не влечет $|X'X| \rightarrow 0$;
- 3) $|X'X| \rightarrow 0$ не влечет $\lambda_1 \rightarrow 0$ и не влечет $\lambda_m/\lambda_1 \rightarrow \infty$.

г. Максимальная парная сопряженность x_1, \dots, x_m . Большую пользу при анализе регрессии приносит рассмотрение матрицы сопряженности

$$R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{bmatrix}, \quad (6.5)$$

где $r_{ij} = \cos(x_i, x_j)$. В качестве показателя мультиколлинеарности может выступить величина

$$\max_{i, j} |r_{ij}|, \quad i \neq j. \quad (6.6)$$

Интуитивно понятно, чем больше значение (6.6), тем сильнее мультиколлинеарность (см. теорему 6.1). Однако (6.6) выражает только парную коллинеарность, т. е. мультиколлинеарность второго порядка¹. Например, три переменные x_1, x_2, x_3 могут быть коллинеарны, но парно не сопряжены. На рис. 6.2 показаны три коллинеарных вектора, принадлежащие плоскости (по терминологии аналитической геометрии векторы x_1, x_2, x_3 компланарны), в то же время матрица сопряженности для этого случая равна:

$$\begin{bmatrix} 1 & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1 \end{bmatrix}$$

Максимальный коэффициент сопряженности не несет на себе эффекта масштаба. Он отражает только степень коллинеарности независимых переменных. Верны следующие соотношения:

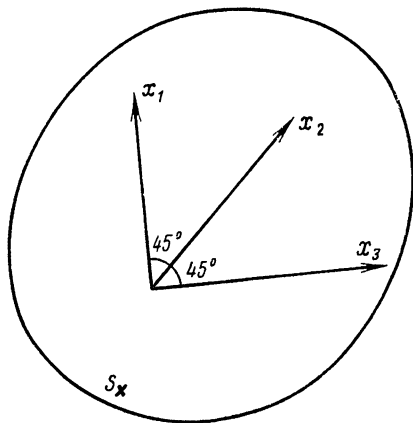


Рис. 6.2. Строгая мультиколлинеарность при попарно линейно-независимых векторах, $m=2$

- 1) $|X'X| \rightarrow 0, \lambda_1 \rightarrow 0, \lambda_m/\lambda_1 \rightarrow \infty$ не влекут обязательно $\max |r_{ij}| \rightarrow 1$;
- 2) $\max |r_{ij}| \rightarrow 1$ влечет $\lambda_1 \rightarrow 0$ и не влечет обязательно $|X'X| \rightarrow 0, \lambda_m/\lambda_1 \rightarrow \infty$.

д. Максимальная сопряженность x_1, \dots, x_m . Предыдущая мера мультиколлинеарности имеет существенный недостаток: она ориентирована только на парную коллинеарность. Поэтому даже если $|r_{ij}| < 1, i \neq j$, возможна строгая мультиколлинеарность.

От этого недостатка свободна другая мера, которую мы сейчас и рассмотрим. Зафиксируем независимую переменную x_i и найдем косинус угла R_i , который составляет эта переменная, т. е. вектор $x_i \in R^n$ с подпространством S_i , натянутым на остальное множество не-

¹Под порядком мультиколлинеарности понимаем число векторов, входящих в уравнение (6.1) с ненулевым коэффициентом.

зависимых переменных $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m$. В качестве меры мультиколлинеарности регрессионной задачи рассмотрим

$$\max_i |R_i|. \quad (6.7)$$

Формально R_i^2 есть коэффициент детерминации в регрессии x_i на $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m$. Для нахождения R_i^2 не обязательно вычислять регрессию x_i на остальные переменные. Укажем более простой способ одновременного нахождения всех R_i^2 , $i = 1, \dots, m$. Не теряя общности найдем, например, R_1^2 . Будем считать $\|x_i\| = 1$; это ограничение, очевидно, не повлияет на величину R_i^2 . Разобьем матрицу X на две подматрицы $X = [x_1 X_2]$, где $x_1^{n \times 1}$ — первый вектор-столбец матрицы X , $X_2^{n \times (m-1)}$ — остальные вектор-столбцы матрицы X . Тогда

$$X'X = \begin{bmatrix} x_1' x_1 & x_1' X_2 \\ X_2' x_1 & X_2' X_2 \end{bmatrix}.$$

Далее по формуле (П.3)

$$\begin{aligned} (X'X)_{11}^{-1} &= \frac{|X_2' X_2|}{|X'X|} = \\ &= \frac{|X_2' X_2|}{|X_2' X_2| |x_1' x_1 - x_1' X_2 (X_2' X_2)^{-1} X_2' x_1|} = \\ &= \frac{1}{x_1' x_1 - x_1' X_2 (X_2' X_2)^{-1} X_2' x_1} = \frac{1}{1 - R_1^2}. \end{aligned} \quad (6.8)$$

Откуда в общем случае

$$R_i^2 = 1 - \frac{1}{R_{ii}^{-1}}, \quad i = 1, \dots, m, \quad (6.9)$$

где R_{ii}^{-1} — (i, i) -элемент матрицы, обратной к сопряженной (6.5).

С помощью соотношения (6.9) можно, например, выяснить, скажется ли коллинеарность x_1 и x_2 на оценке МНК третьего параметра a_3 в регрессии ($m = 3$)

$$y_t = \alpha_1 + \alpha_2 x_{t2} + \alpha_3 x_{t3} + \varepsilon_t, \quad t = 1, \dots, n,$$

где x_1 и x_2 сильно сопряжены, а x_1 и x_3 , x_2 и x_3 — нет. Заметим, что

$$R_1^2 = R^2(x_1; x_2, x_3) \geq R^2(x_1, x_2) = r_{12}^2 \approx 1;$$

$$R_2^2 = R^2(x_2; x_1, x_3) \geq R^2(x_1, x_2) = r_{12}^2 \approx 1.$$

Однако

$$R_3^2 = R^2(x_3; x_1, x_2) \ll 1.$$

Для простоты приведем x_1, x_2, x_3 к одному масштабу, т. е. нормируем их. Тогда по формуле (6.8)

$$(X'X)_{ii}^{-1} = \frac{1}{1-R_i^2}, \quad i=1, 2, 3;$$

поэтому первые два параметра будут иметь большую дисперсию, а третий — нет. Другими словами, мультиколлинеарность не влияет на точность оценивания параметров, соответствующие независимые переменные которых не порождают мультиколлинеарность.

Мера (6.7) является хорошей мерой мультиколлинеарности. Она не связана с масштабом измерения независимых переменных, отражает их внутреннюю сопряженность и хорошо интерпретируется.

Между введенными мерами мультиколлинеарности существуют следующие соотношения:

- 1) $\max |R_i| \rightarrow 1$ влечет $\lambda_1 \rightarrow 0$;
- 2) $\max |r_{ij}| \rightarrow 1$ влечет $\max |R_i| \rightarrow 1$.

Зависимости между рассмотренными мерами мультиколлинеарности $a - d$ показаны на рис. 6.3. Как видно из этого

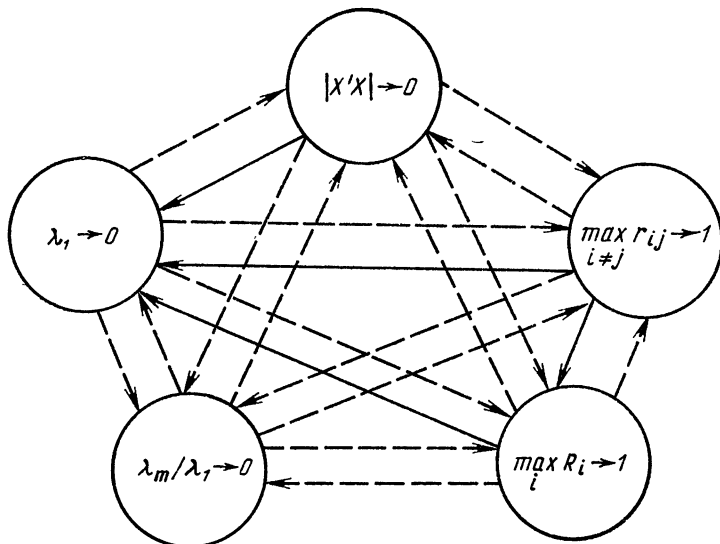


Рис. 6.3. Зависимость между различными мерами мультиколлинеарности

рисунка, между мерами $a - \delta$ мало зависимостей. Все они отражают определенную сторону мультиколлинеарности. На практике целесообразно пользоваться по меньшей мере λ_1 , λ_m/λ_1 , и $\max |R_i|$ как величинами, отражающими степень мультиколлинеарности независимых переменных.

Можно показать, что при увеличении степени мультиколлинеарности по каждой из рассмотренных мер, за исключением меры v , точность оценивания параметров методом наименьших квадратов убывает. Предельные случаи соответствуют строгой мультиколлинеарности.

Остановимся еще на одном вопросе. Как узнать, какие переменные порождают мультиколлинеарность? Как было отмечено, наилучшее приближение левой части (6.1) к нулю наблюдается, если за v взять характеристический вектор, отвечающий λ_1 . Если в (6.1) входят не все векторы, то соответствующие координаты в векторе v будут близки к нулю. Остальные переменные и порождают мультиколлинеарность.

6.2. Строгая мультиколлинеарность

Напомним, что под этим понимается случай $\text{rang } X = r < m$. Тогда обычная формула для нахождения оценки МНК неприменима, так как матрица $X'X$, также имеющая ранг $r < m$, необратима в обычном смысле слова, а регрессия становится неидентифицируемой [24]. Однако если под оценкой МНК понимать по-прежнему те значения a_1, \dots, a_m , которые обращают сумму квадратов отклонений в минимум, то в случае строгой мультиколлинеарности существует целое линейное многообразие оценок МНК:

$$a = \{\alpha \in R^m : X'X\alpha = X'y\}. \quad (6.10)$$

Размерность a равна $m - r$ (см. задачу 3 упражнения 1.1). Как найти семейство (6.10)? Для этого воспользуемся понятием обобщенной обратной матрицы или g -обратной матрицы. Пусть A — матрица порядка $n \times m$, $\text{rang } A = r$; g -обратной или обобщенной обратной матрицей к матрице A назовем такую матрицу A^- порядка $m \times n$, что $AA^-A = A$ [58, с. 39].

Основные свойства обобщенных обратных матриц:

1. Для любого $y \in R^n$, для которого система $Ax = y$ совместна, $x = A^-y$ является ее решением.

2. $H = A^-A$ — идемпотентная матрица.

3. Для любой матрицы существует хотя бы одна обобщенная обратная матрица, которая не обязательно единственна.

4. Пусть Λ — диагональная матрица $m \times m$, последние r диагональных элементов которой равны нулю, а первые $m - r$ — ненулевые. Обозначим через Λ^+ такую диагональную матрицу $m \times m$, что $\Lambda_{ii}^+ = 1/\Lambda_{ii}$ для $i = 1, \dots, m - r$ и $\Lambda_{ii}^+ = 0$ для $i = m - r + 1, \dots, m$. Тогда Λ^+ является обобщенной обратной матрицей к матрице Λ .

5. Если A — квадратная и симметричная матрица $m \times m$, то A^- может быть построена следующим образом. Обозначим через P ортогональную матрицу $m \times m$, сводящую A к диагональной, т. е. $P'AP = \Lambda$, где Λ_{ii} — характеристическое число матрицы A . Тогда матрица $A^- = P\Lambda^+P'$ является обобщенной обратной матрицей к матрице A . Эта матрица называется обратной матрицей Мура — Пенроуза [56, с. 40] и обозначается A^+ .

Как следует из свойства 1, оценка

$$a = (X'X)^+ X'y \quad (6.11)$$

является одним из членов семейства оценок МНК. Более подробно об оценках МНК (6.11) см. [4].

На практике мультиколлинеарность в строгом смысле не встречается. Как правило, независимые переменные являются результатом измерения и поэтому содержат ошибки. Это ведет к тому, что даже если истинные значения x_1, \dots, x_m линейно зависимы, то теоретически матрица $X'X$ невырождена, хотя и плохо определена.

6.3. Смещенные оценки

Если в классической линейной регрессии предположения $A - E$ выполняются, то оценка МНК

$$a = (X'X)^{-1}X'y \quad (6.12)$$

является эффективной в классе несмещенных оценок, линейных по y (теорема Гаусса—Маркова, параграф 1.5). Если к тому же предположить, что отклонения модели регрессии имеют нормальное распределение, то оценка МНК оказывается эффективной в классе всех несмещенных оценок (линейных и нелинейных). Однако даже если отклонения нормальны, то в некоторых ситуациях оценки МНК становятся нестабильными. Это происходит при сильной

сопряженности независимых переменных, т. е. при мультиколлинеарности. Как было показано в параграфе 6.1, при усилении мультиколлинеарности точность оценки МНК падает. В частности, мультиколлинеарность ведет к тому, что координаты вектора оценки принимают очень большие значения. Это утверждение следует из равенства

$$\begin{aligned} E \mathbf{a}' \mathbf{a} &= E (\boldsymbol{\alpha} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon})' (\boldsymbol{\alpha} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}) = \\ &= \boldsymbol{\alpha}'\boldsymbol{\alpha} + \sigma^2 \operatorname{tr} (\mathbf{X}'\mathbf{X})^{-1} \rightarrow \infty \end{aligned}$$

при $\lambda_{\min} (\mathbf{X}'\mathbf{X}) \rightarrow 0$. При сильной мультиколлинеарности оценка МНК становится настолько неудовлетворительной, что даже знаки некоторых координат α_i часто не соответствуют истинным.

Таким образом, задачу оценивания можно сформулировать следующим образом: найти оценку параметров регрессии, которая была бы устойчивой даже при сильной сопряженности независимых переменных, т. е. такую оценку, точность которой не падала бы до нуля при усилении мультиколлинеарности¹.

Ограничимся классом линейных оценок как наиболее простым. Следующим ограничением, связанным с оценкой МНК, является условие несмещенности оценки. Для того чтобы оценка была устойчивой по отношению к мультиколлинеарности, необходимо отказаться от этого условия, т. е. рассматривать и смещенные оценки. Итак, допустим σ^2 фиксировано. В качестве критерия оценки (функции риска) выберем среднюю сумму квадратов ошибок (ССКО) $L_a(\boldsymbol{\alpha}) = E(\mathbf{d} - \boldsymbol{\alpha})'(\mathbf{d} - \boldsymbol{\alpha})$, где \mathbf{d} — некоторая оценка параметра $\boldsymbol{\alpha} \in R^m$. Обозначим через M класс линейных оценок, каждая из которого имеет ограниченную ССКО для всех \mathbf{X} (в том числе для тех матриц, для которых $\operatorname{rang} \mathbf{X} < m$) и всех $\boldsymbol{\alpha}$ из некоторого априорного множества $\Theta \subset R^m$. Ясно, что несмещенная линейная оценка не принадлежит M . Действительно, пусть \mathbf{b} — линейная несмещенная оценка и $\mathbf{b} \in M$. Тогда $\operatorname{cov}(\mathbf{b}) \geq \operatorname{cov}(\mathbf{a})$, где \mathbf{a} — оценка МНК и

$$L_b(\boldsymbol{\alpha}) \geq L_a(\boldsymbol{\alpha}) = \sigma^2 \sum_{i=1}^m \frac{1}{\lambda_i} \rightarrow \infty; \lambda_1 = \lambda_{\min} (\mathbf{X}'\mathbf{X}) \rightarrow 0,$$

¹Под точностью несмещенной оценки можно, например, понимать величину, обратную к дисперсии или $1/\operatorname{tr} \operatorname{cov}(\mathbf{a})$ в многомерном случае.

т. е. ССКО оценки \mathbf{b} не ограничена, $\mathbf{b} \in M$ — возникает противоречие,

Итак, для того чтобы построить оценки, которые были бы хороши и в случае мультиколлинеарности, необходимо отказаться от условия несмещенности. Таким образом, приходим к более широкому классу линейных оценок (смещенных и несмещенных). Отказ от несмещенности имеет положительные и отрицательные стороны: 1) положительным является то, что возможно найти смещенную оценку, которая является устойчивой относительно сильной сопряженности независимых переменных, т. е. в случае плохо обусловленной матрицы $\mathbf{X}'\mathbf{X}$; 2) отрицательным фактом перехода в класс линейных смещенных оценок является то, что в этом классе нельзя найти оптимальной оценки в смысле минимальной матрицы средних квадратов отклонений. Аналогичная ситуация имела место в параграфе 1.4, где для каждого $\alpha_0 \in R^1$ существует смещенная оценка \mathbf{d} , для которой $E_{\alpha_0}(\mathbf{d} - \alpha_0)^2 = 0$. Отсутствие оптимальной оценки в классе всех линейных оценок приводит к тому, что число оценок будет велико.

Покажем, что класс линейных оценок с ограниченной ССКО (т.е. класс M) имеет смысл рассматривать только для ограниченных Θ . Обозначим $\Theta_r = \{\alpha \in R^m : \|\alpha\| \leq r\}$, $r > 0$. В частности, при $r = \infty$ $\Theta_r = R^m$, в остальных случаях Θ_r — ограниченное множество.

Теорема 6.1. *Класс M не пуст тогда и только тогда, когда r — конечное число.*

Доказательство. *Достаточность.* Пусть r — конечное число, т. е. Θ_r — ограниченное множество. Пусть \mathbf{d} — линейная оценка параметра α , т. е. $\mathbf{d} = \mathbf{C}\mathbf{y}$, где \mathbf{C} — детерминированная матрица $m \times n$. Найдем ССКО оценки \mathbf{d} . Имеем

$$L_d(\alpha) = E(\mathbf{d} - \alpha)'(\mathbf{d} - \alpha) = \alpha'(\mathbf{C}\mathbf{X} - \mathbf{I}_m)'(\mathbf{C}\mathbf{X} - \mathbf{I}_m)\alpha + \sigma^2 \text{tr } \mathbf{C}\mathbf{C}'. \quad (6.13)$$

Полагаем $\mathbf{C} \equiv \mathbf{0}$, т. е. $\mathbf{d} \equiv \mathbf{0}$. Тогда $L_0(\alpha) = \alpha'\alpha \leq r^2$, $\alpha \in \Theta_r$, т. е. $\mathbf{d} = \mathbf{0} \in M$ и класс M не пуст.

Необходимость. Пусть M — непустое множество, т. е. некоторая линейная оценка $\mathbf{d} \in M$. Было показано, что \mathbf{d} не может быть несмещенной оценкой, т. е. $\mathbf{C}\mathbf{X} - \mathbf{I}_m \neq \mathbf{0}$. Отсюда следует, что матрица $(\mathbf{C}\mathbf{X} - \mathbf{I}_m)'(\mathbf{C}\mathbf{X} - \mathbf{I}_m)$ имеет хотя бы одно ненулевое характеристическое число τ , $\tau > 0$. Пусть \mathbf{v} — характеристический вектор матрицы

$(CX - I_m)'(CX - I_m)$, отвечающий τ . Тогда, если $r = \infty$, то

$$L_d(v\lambda) = \lambda^2 v'(CX - I_m)'(CX - I_m)v + \sigma^2 \text{tr} CC' \geq \lambda^2 \tau,$$

так как $\|v\| = 1$ и при $\lambda \rightarrow \infty$ $L_d(v\lambda) \rightarrow \infty$, т. е. оценка d имеет неограниченную ССКО — противоречие.

Из теоремы 6.1 следует, что не существует линейной оценки для $\Theta = R^m$, которая имела бы ССКО меньшую, чем \hat{y} оценки МНК.

Итак, займемся отысканием оптимальных оценок в классе M , где $\Theta = \{\|\alpha\| \leq r\}$. Обозначим этот класс M_r .

Т е о р е м а 6.2. *В классе M_r не существует оптимальной оценки в смысле минимума ССКО.*

Д о к а з а т е л ь с т в о. Преобразуем выражение (6.13):

$$L_d(\alpha) = \alpha' X' C' C X \alpha - \alpha' X' C' \alpha - \alpha' C X \alpha + \alpha' \alpha + \sigma^2 \text{tr} CC'.$$

Но $\alpha' X' C' \alpha = \alpha' C X \alpha$, поэтому

$$L(\alpha) = L_d(\alpha) = \alpha' X' C' C X \alpha - 2\alpha' C X \alpha + \sigma^2 \text{tr} CC'. \quad (6.14)$$

Найдем $\min L(\alpha)$ по C при фиксированном α и σ^2 . Продифференцируем выражение (6.14) по C и приравняем производную к нулю. Применим следующие формулы: $\frac{\partial \text{tr} CC'}{\partial C} = 2C$, $\frac{\partial \alpha' C' C \alpha}{\partial C} = 2C \alpha \alpha'$, которые можно доказать, используя приложение П.2. Принимая во внимание формулу (П.8), получим

$$\frac{\partial L}{\partial C} = 2C X \alpha \alpha' X' - 2\alpha \alpha' X' + 2\sigma^2 C = 0,$$

или

$$C (X \alpha \alpha' X' + \sigma^2 I_n) = \alpha \alpha' X',$$

откуда

$$C = \alpha \alpha' X' (X \alpha \alpha' X' + \sigma^2 I_n)^{-1}. \quad (6.15)$$

В последнем выражении обращение матрицы корректно, так как матрица $X \alpha \alpha' X' + \sigma^2 I_n$ положительно определена. Далее воспользуемся следующим элементарным фактом. Для любого $b \in R^n$

$$b'(bb' + \sigma^2 I_n)^{-1} = \frac{1}{\sigma^2 + b'b} b. \quad (6.16)$$

Равенство (6.16) легко проверяется умножением обеих частей на матрицу $bb' + \sigma^2 I_n$.

Используя (6.16), выражение (6.15) перепишем следующим образом:

$$C^* = \frac{\alpha\alpha' X'}{\sigma^2 + \alpha' X' X \alpha}. \quad (6.17)$$

Таким образом, «оценка», минимизирующая ССКО, зависит от α и σ^2 и равна:

$$d^* = C^* y = \frac{\alpha\alpha' X' y}{\sigma^2 + \alpha' X' X \alpha}. \quad (6.18)$$

Для окончательного доказательства теоремы предположим, что d — оптимальная оценка в классе M_r , т. е. для всех $b \in M_r$ $L_d(\alpha) \leq L_b(\alpha)$, $\alpha \in \Theta_r$. Выберем некоторое $\alpha_0 \in \Theta_r$ и построим оценку

$$d_0 = C_0 y = \frac{\alpha_0 \alpha_0' X' y}{\sigma^2 + \alpha_0' X' X \alpha_0}.$$

Тогда $L_{d_0}(\alpha_0) \leq L_d(\alpha_0)$, причем равенство будет наблюдаться, если $d = d_0$. Теперь, если взять $\alpha_1 \neq \alpha_0$, то соответствующая оценка d_1 будет иметь меньшую ССКО, чем оценка d — противоречие; теорема доказана.

Доказанные теоремы подсказывают нам, что, во-первых, нельзя найти устойчивую относительно мультиколлинеарности оценку для всех $\alpha \in R^m$, т. е. априорное множество параметров необходимо ограничить; во-вторых, отсутствие эффективной оценки в смысле минимальной ССКО приводит к существованию большого числа несравнимых оценок, оптимальных каждый раз в некотором заранее определенном смысле. В следующих параграфах рассмотрим некоторые оценки из класса M_r .

Имеется еще один обходный путь трудностей, связанных с несравнимостью некоторых функций риска, а именно ССКО. Рассмотрим минимаксные оценки (см. параграф 1.4). Начнем с простейшего случая $m = 1$. Итак, имеется регрессия $y_t = \alpha x_t + \varepsilon_t$, $t = 1, \dots, n$. Требуется найти линейную оценку, минимизирующую максимальное значение ССКО для $\alpha \in \Theta_r$. Пусть $c = (c_1, \dots, c_n)$ и $d = \sum_t c_t y_t$ — линейная оценка α .

Найдем

$$\min_c \max_{\|\alpha\| < r} L_d(\alpha) = \min_c \max_{|\alpha| < r} [\alpha^2 (\sum c_t x_t - 1)^2 + \sigma^2 \sum c_t^2].$$

Пусть вектор c фиксирован, тогда

$$\max_{|\alpha| \leq r} [\alpha^2 (\sum_i c_i x_i - 1)^2 + \sigma^2 \sum_i c_i^2] = r^2 \left(\sum_i c_i x_i - 1 \right)^2 + \sigma^2 \sum_i c_i^2.$$

Теперь будем минимизировать функцию $L(c) = r^2 (\sum_i c_i x_i - 1)^2 + \sigma^2 \sum_i c_i^2$. Найдем ее производную по c_j и приравняем ее к нулю. Получим

$$\frac{\partial L}{\partial c_j} = 2r^2 (\sum_i c_i x_i - 1) x_j + 2\sigma^2 c_j = 0, \quad j = 1, \dots, n. \quad (6.19)$$

Умножим каждое из предыдущих уравнений на x_j и просуммируем по j от 1 до n . Получим

$$r^2 (\sum_i c_i x_i - 1) \sum_j x_j^2 + \sigma^2 \sum_j c_j x_j = 0,$$

откуда

$$\sum_j c_j x_j = \frac{r^2 \sum_j x_j^2}{\sigma^2 + \sum_j x_j^2}.$$

Подставляя найденное значение в j -е уравнение (6.19), окончательно найдем

$$c_j = \frac{x_j}{\sum_i x_i^2 + \sigma^2/r^2}, \quad j = 1, \dots, n.$$

Соответствующая оценка равна:

$$d = \frac{\sum_i x_i y_i}{\sum_i x_i^2 + \sigma^2/r^2}. \quad (6.20)$$

Минимаксная оценка (6.20) является ридж-оценкой (см. параграф 6.4). При $r \rightarrow \infty$ эта оценка переходит в оценку МНК. Для конечного r оценка (6.20) устойчива относительно мультиколлинеарности. В случае $m = 1$ мультиколлинеарность означает близость $\sum_i x_i^2$ к нулю. Можно проверить, что при $\sum_i x_i^2 \rightarrow 0$ $E(d - \alpha)^2$ не стремится к ∞ , тогда как $E(a - \alpha)^2 = \sigma^2(a) \rightarrow \infty$, где a — оценка МНК. Вычисления минимаксной оценки для $m > 1$ весьма затруднительны.

Как же поступить при наличии мультиколлинеарности? На практике часто идут по следующему пути. Как было показано, наличие мультиколлинеарности ведет к большим дисперсиям некоторых координат вектора оценки МНК. Будем считать, что отклонения регрессии гомоскедастич-

ны и нормально распределены. Тогда, проверяя гипотезы $H_i : \alpha_i = 0$, мы можем в случае их принятия отбросить соответствующие независимые переменные и пересчитать регрессию заново. Назовем такую процедуру отсеивания автоматической.

На рис. 6.4 показана оценка МНК для случая $m = 2$ — вектор \overline{OA} ; S_y — характеристический эллипс. Оценка автоматического отсева переменных характеризуется тем, что характеристический эллипс касается оси α_1 (при отбросе

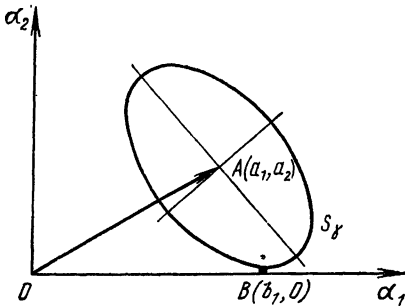


Рис. 6.4. Геометрия оценки автоматического отсева переменных

переменной x_2). Вектор этой оценки есть \overline{OB} . Оценка $(b_1, 0)$ есть оценка МНК при условии, что вторая координата равна нулю.

Автоматический отсев переменных предполагает равенство нулю некоторых координат оценки. При этом, вероятно, длина оценки уменьшается, и мы не будем получать больших по абсолютной величине значений оценок.

Однако подобная процедура имеет и недостатки. Во-первых, для гарантии того, что употребляемые статистические критерии достаточно эффективны, мы обязаны предположить, что отклонения имеют нормальное распределение или близкое к нему. Во-вторых, процедура приравнивания нулю некоторых координат оценки — весьма грубая. У нас мало уверенности в том, что истинное значение параметра в точности окажется равным нулю. Например, часто в правую часть регрессии входят переменные, которые тесно «коррелируют» между собой. И тем не менее мы не хотим ни одну из переменных выбросить из анализа, поскольку они часто имеют большой физический смысл. Разумеется, одновременно хорошо оценить параметры при этих переменных в силу мультиколлинеарности не удастся. Однако нас вполне устроит более или менее удовлетворительная оценка. В случае же схемы автоматического отсева одна из переменных была бы выброшена из анализа, т. е. нашей оценкой при выброшенной переменной был бы нуль!

Далее рассмотрены оценки, которые «смягчают» оценку МНК, не прибегая к экстраординарным мерам — считать

оценку некоторой координаты неизвестного вектора нулем, как это делается в схеме автоматического отсева переменных. Все описанные оценки уменьшают длину оценки МНК, таким образом являясь более устойчивым.

Мультиколлинеарность не сказывается на точности оценивания σ^2 . Другими словами, обычная несмещенная оценка

$$s^2 = \frac{1}{n-m} (\mathbf{y} - \mathbf{X}\mathbf{a})' (\mathbf{y} - \mathbf{X}\mathbf{a}),$$

использующая оценку МНК, является вполне удовлетворительной.

Действительно, $s^2 = \frac{1}{n-m} \mathbf{e}' \mathbf{A} \mathbf{e}$, где $\mathbf{A} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ — идемпотентная матрица $n \times n$. Найдем дисперсию оценки s^2 , т. е. $E(s^2 - \sigma^2)^2 = E s^4 - \sigma^4$. Очевидно

$$E(n-m)^2 s^4 = E \mathbf{e}' \mathbf{A} \mathbf{e} \mathbf{e}' \mathbf{A} \mathbf{e} = E \sum_{i, j, k, l} \varepsilon_i \varepsilon_j \varepsilon_k \varepsilon_l A_{ij} A_{kl}. \quad (6.21)$$

Будем предполагать $\varepsilon_1, \dots, \varepsilon_n$ независимыми симметрично распределенными, т. е. $E\varepsilon_i^3 = 0$, имеющими одинаковый четвертый момент $E\varepsilon_i^4 = \nu$, $i = 1, \dots, n$. Тогда слагаемые суммы не равны нулю, только если индексы i, j, k, l удовлетворяют одному из следующих условий:

- 1) $i = j = k = l$ — сумма (6.21) равна $E \sum_i \varepsilon_i^4 A_{ii}^2$;
- 2) $i = j \neq k = l$ — сумма (6.21) равна $E \sum_i \sum_k \varepsilon_i^2 \varepsilon_k^2 A_{ii} A_{kk}$;
- 3) $i = k \neq j = l$ — сумма (6.21) равна $E \sum_{i \neq j} \varepsilon_i^2 \varepsilon_j^2 A_{ij}^2$;
- 4) $i = l \neq j = k$ — сумма (6.21) равна $E \sum_{i \neq j} \varepsilon_i^2 \varepsilon_j^2 A_{ij} A_{ji}$.

Сумма (6.21) переписывается следующим образом:

$$E(n-m)^2 s^4 = \nu \sum_i A_{ii}^2 + \sigma^2 \sum_{i \neq k} A_{ii} A_{kk} + 2\sigma^4 \sum_{i \neq j} A_{ij}^2.$$

Обозначим $\sum_i A_{ii}^2 = p$. В силу идемпотентности $\sum_{i,j} A_{ij}^2 = \sum_i A_{ii} = n - m$, откуда

$$\begin{aligned} \left(\sum_i A_{ii} \right)^2 &= \sum_i A_{ii}^2 + \sum_{i \neq j} A_{ii} A_{jj} = (n-m)^2; \quad \sum_{i \neq k} A_{ii} A_{kk} = \\ &= (n-m)^2 - p; \quad \sum_{i \neq j} A_{ij}^2 = n - m - p. \end{aligned}$$

Поэтому

$$E s^4 = \frac{\nu p}{(n-m)^2} + \sigma^4 \left(1 + \frac{2(n-m) - 3p}{(n-m)^2} \right);$$

$$E (s^2 - \sigma^2)^2 = \frac{\nu p}{(n-m)^2} + \sigma^4 \frac{2(n-m) - 3p}{(n-m)^2}.$$

Но $n - m - p \geq 0$, т. е. $p \leq n - m$, поэтому

$$E (s^2 - \sigma^2)^2 \leq \frac{\nu + 2\sigma^4}{n-m}.$$

Таким образом, даже если $\lambda_{\min}(\mathbf{X}'\mathbf{X}) \rightarrow 0$, дисперсия оценки s^2 не стремится к бесконечности, как это наблюдается в оценке МНК.

У п р а ж н е н и я 6.3

1. Докажите, что оценка МНК является эффективной в классе линейных несмещенных оценок с априорным множеством $\Theta_r = \{\alpha \in R^m : \|\alpha\| \leq r\}$, $r > 0$.

2. Допустим, класс оценок M не пуст. Верно ли тогда, что Θ — ограниченное множество?

3. Докажите, что если A — матрица $m \times n$ и $A \neq 0$, то $\lambda_{\max}(A'A) > 0$.

4. Для каких r и σ^2 минимаксная оценка (6.20) имеет ССКО меньшую, чем оценка МНК?

6.4. Ридж-оценки

В предыдущих параграфах показано, что оценка МНК имеет большую дисперсию в случае мультиколлинеарности и при усилении мультиколлинеарности становится неустойчивой. Если же выйти за рамки несмещенных оценок, то оказывается можно построить более устойчивые оценки, с меньшим квадратом ошибки. К таким оценкам прежде всего относятся так называемые ридж-оценки (ridge — гребень, хребет), впервые введенные А. Гоэрлом. Такое название оценкам он дал при анализе регрессии, представляющей собой поверхность второго порядка [130]. В общем виде ридж-оценка вектора параметров α линейной регрессионной модели записывается так:

$$a(K) = (\mathbf{X}'\mathbf{X} + K)^{-1}\mathbf{X}'\mathbf{y}, \quad (6.22)$$

где K — некая неотрицательно определенная матрица $m \times m$. Суть оценки (6.22) ясна: добавление к матрице плана $\mathbf{X}'\mathbf{X}$ неотрицательно определенной матрицы делает ее лучше определенной, а оценки — более устойчивыми.

Часто матрицу K выбирают диагональной, причем ее диагональные элементы пропорциональны диагональным элементам исходной матрицы плана $X'X$, т. е.

$$K_{ii} = k (X'X)_{ii}, \quad K_{ij} = 0 \quad (i \neq j), \quad (6.23)$$

где $k \geq 0$. Еще более простой способ — прибавление к диагональным элементам матрицы $X'X$ некоторого неотрицательного числа, т. е. $K = kI_m$. Можно показать, что случай (6.23) сводится к последнему. Действительно, обозначим

$$D = \begin{bmatrix} (X'X)_{11} & & & \\ & & & 0 \\ & & \ddots & \\ & 0 & & \\ & & & (X'X)_{mm} \end{bmatrix},$$

$$Z = XD^{-1/2}, \quad \beta = D^{1/2} \alpha. \quad (6.24)$$

Тогда модель (1.2) сведется к следующей модели:

$$y = Z\beta + \varepsilon, \quad (6.25)$$

причем матрица $Z'Z = D^{-1/2}X'XD^{-1/2}$ является «корреляционной матрицей» независимых переменных x_1, \dots, x_m .

Ридж-оценкой (6.22) с матрицей $K = kI_m$ для модели (6.25) является

$$\begin{aligned} b(k) &= (Z'Z + kI)^{-1}Z'y = [D^{-1/2}X'XD^{-1/2} + \\ &+ kI]^{-1}D^{-1/2}X'y = D^{-1/2}(X'X + K)^{-1}X'y, \quad K_{ii} = \\ &= k(X'X)_{ii}. \end{aligned}$$

Поэтому

$$a(k) = D^{-1/2}b(k) \quad (6.26)$$

есть ридж-оценка (6.22) исходной модели (1.2) с выбором K по правилу (6.23). Итак, вместо (6.24) можно рассмотреть приведенную регрессионную модель (6.25), у которой ридж-оценка имеет более простую структуру

$$b(k) = (Z'Z + kI)^{-1}Z'y. \quad (6.27)$$

Кратко остановимся на свойствах этой оценки:

- 1) $b(0)$ — оценка МНК;
- 2) $b(k)$ — является линейным преобразованием оценки МНК:

$$b(k) = Bb; \quad B = [I + k(X'X)^{-1}]^{-1}, \quad (6.28)$$

где \mathbf{b} — оценка МНК уравнения (6.25). Поскольку $\mathbf{E}\mathbf{b} = \boldsymbol{\beta}$, а $\mathbf{B} \neq \mathbf{I}_m$, то ридж-оценка является смещенной;

3) в классе оценок с фиксированной длиной ридж-оценка (6.27) минимизирует сумму квадратов отклонений. Для доказательства достаточно показать, что (6.27) является

решением следующей задачи:

$$(\mathbf{y} - \mathbf{Z}\mathbf{d})' (\mathbf{y} - \mathbf{Z}\mathbf{d}) \Rightarrow \min, \\ \mathbf{d}'\mathbf{d} = c.$$

Построим функцию Лагранжа

$$\Phi(\mathbf{d}, k) = (\mathbf{y} - \mathbf{Z}\mathbf{d})' (\mathbf{y} - \mathbf{Z}\mathbf{d}) + k(\mathbf{d}'\mathbf{d} - c).$$

Необходимым условием минимума функции Φ является

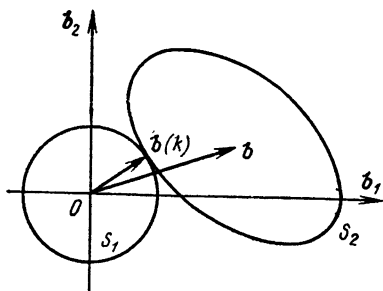


Рис. 6.5. Ридж-оценка как оценка с данной длиной и минимизирующая сумму квадратов отклонений, $m=2$

$$\partial\Phi/\partial\mathbf{d} = -2\mathbf{Z}'\mathbf{y} + 2\mathbf{Z}'\mathbf{Z}\mathbf{d} + 2k\mathbf{d} = 0,$$

откуда $(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_m)\mathbf{d} = \mathbf{Z}'\mathbf{y}$
и оценка (6.27)

$$\mathbf{d} = \mathbf{b}(k) = (\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_m)^{-1}\mathbf{Z}'\mathbf{y}.$$

На рис. 6.5 эта оценка есть точка соприкосновения окружности S_1 с радиусом \sqrt{c} и эллипса $S_2 = \{\mathbf{d} : (\mathbf{y} - \mathbf{Z}\mathbf{d})' (\mathbf{y} - \mathbf{Z}\mathbf{d}) = \text{const}\}$.

Можно показать, что верно и обратное утверждение: ридж-оценка имеет минимальную длину в классе оценок с данным значением суммы квадратов.

В случае мультиколлинеарности средняя сумма квадратов ошибок оценки МНК оказывается высокой. Что можно сказать об оценке (6.27)?

Обозначим $\mathbf{S} = (\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_m)^{-1}\mathbf{Z}'\mathbf{Z} - \mathbf{I}_m$. По определению

$$\begin{aligned} L(\boldsymbol{\beta}) &= \mathbf{E}(\mathbf{b}(k) - \boldsymbol{\beta})' (\mathbf{b}(k) - \boldsymbol{\beta}) = \mathbf{E}[\mathbf{S}\boldsymbol{\beta} + \\ &+ (\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}]' \times [\mathbf{S}\boldsymbol{\beta} + (\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}] = \\ &= \boldsymbol{\beta}'\mathbf{S}'\mathbf{S}\boldsymbol{\beta} + \mathbf{E}\boldsymbol{\beta}'\mathbf{S}'(\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon} + \mathbf{E}\boldsymbol{\varepsilon}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \\ &+ k\mathbf{I})^{-1}\mathbf{S}\boldsymbol{\beta} + \mathbf{E}\boldsymbol{\varepsilon}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-2}\mathbf{Z}'\boldsymbol{\varepsilon}. \end{aligned}$$

Второе и третье слагаемые в последнем выражении обращаются в нуль в силу равенства $E\varepsilon = 0$. К последнему слагаемому применим формулу (П.13). Тогда

$$L(\beta) = \beta' S' S \beta + \sigma^2 \operatorname{tr} Z (Z' Z + kI)^{-2} Z'.$$

Преобразуем матрицу S' следующим образом:

$$S' = Z' Z (Z' Z + kI)^{-1} - I = (Z' Z + kI)^{-1} [Z' Z - \\ - Z' Z + kI] = -k (Z' Z + kI)^{-1};$$

аналогично преобразуется матрица S . Таким образом, $S'S = k^2 (Z' Z + kI)^{-2}$. Далее, легко видеть, что характеристические векторы матриц $Z'Z$, $Z'Z + kI$, $(Z'Z + kI)^{-1}$, $(Z'Z + kI)^{-2}$ совпадают. Если i -е характеристическое число матрицы $Z'Z$ обозначить λ_i , то i -м х. ч. матрицы $Z'Z + kI$ будет $\lambda_i + k$, а матрицы $(Z'Z + kI)^{-2} = (\lambda_i + k)^{-2}$.

Л е м м а 6.1. Пусть A , B — симметричные матрицы с совпадающими х. в. и характеристическими числами, равными соответственно $\lambda_1, \dots, \lambda_m$; μ_1, \dots, μ_m . Тогда $\operatorname{tr} AB = \sum_i \lambda_i \mu_i$.

Д о к а з а т е л ь с т в о. Покажем, что i -е х. ч. матрицы AB есть $\lambda_i \mu_i$. Действительно, пусть P — матрица, составленная из х. в. матрицы A , такова, что $P'AP = \Lambda_A$; тогда $P'BP = \Lambda_B$ и $AB = P\Lambda_A P' P\Lambda_B P' = P\Lambda_A \Lambda_B P'$, откуда и следует утверждение леммы.

Используя лемму, находим

$$\operatorname{tr} Z (Z' Z + kI)^{-2} Z' = \operatorname{tr} (Z' Z + kI)^{-2} Z' Z = \sum_i \frac{\lambda_i}{(\lambda_i + k)^2}.$$

Далее, пусть P — ортогональная матрица, столбцы которой составлены из х. в. матрицы $Z'Z$; тогда $P'Z'ZP = \Lambda$, $\Lambda_{ii} = \lambda_i$ и

$$(Z' Z + kI)^{-2} = P' \Lambda^* P; \Lambda_{ii} = 1/(\lambda_i + k)^2.$$

Обозначим

$$\gamma = P\beta; W = ZP; y = W\gamma + \varepsilon, \quad (6.29)$$

тогда

$$k^2 \beta' (Z' Z + kI)^{-2} \beta = k^2 \sum_{i=1}^m \frac{\gamma_i^2}{(\lambda_i + k)^2}.$$

Окончательно

$$L(k) = L_b(k) = k^2 \sum_{i=1}^m \frac{\gamma_i^2}{(\lambda_i + k)^2} + \sigma^2 \sum_{i=1}^m \frac{\lambda_i}{(\lambda_i + k)^2} =$$

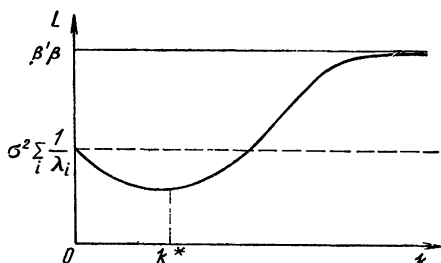
$$= \sum_{i=1}^m \frac{k^2 \gamma_i^2 + \sigma^2 \lambda_i}{(\lambda_i + k)^2}. \quad (6.30)$$

Исследуем среднюю сумму квадратов ошибок (6.30) ридж-оценки как функцию k . Ясно, что $L(0) = \sigma^2 \sum 1/\lambda_i = \sigma^2 \text{tr}(\mathbf{Z}'\mathbf{Z})^{-1}$ есть средняя сумма квадратов ошибок оценки МНК. Найдем первую и вторую производные $L(k)$:

$$L'(k) = 2 \sum_{i=1}^m \frac{\lambda_i \gamma_i^2 k - \sigma^2 \lambda_i}{(\lambda_i + k)^3};$$

$$L''(k) = 2 \sum_{i=1}^m \frac{3\sigma^2 \lambda_i + \gamma_i^2 \lambda_i^2 - 2\lambda_i \gamma_i^2 k}{(\lambda_i + k)^4}.$$

Как видим, $L'(0) < 0$, т. е. в точке $k = 0$ функция $L(k)$ убывает, а при $k \rightarrow \infty$ $L(k)$ имеет асимптоту (рис. 6.6):



$$\lim_{k \rightarrow \infty} L(k) = \sum_{i=1}^m \gamma_i^2 =$$

$$= \beta' \mathbf{P}' \mathbf{P} \beta = \beta' \beta.$$

Легко видеть, что для всех $0 < k < \sigma^2 / \max_i \gamma_i^2$ $L'(k)$ меньше

нуля. Это значит, для всех таких k $L(k) < L(0)$. Далее, можно найти оптимальное значение k : $L(k^*) =$

Рис. 6.6. График среднего квадрата ошибки для случая $\beta' \beta > \sigma^2 \sum_{i=1}^m \frac{1}{\lambda_i}$

$= \min_{k > 0} L(k)$. Для приближенного вычисления k^* можно воспользоваться методом Ньютона—Рафсона. Аппроксимируем $L(k)$ в окрестности точки $k = 0$ параболой, т. е.

$$L(k) \approx L(0) + kL'(0) + \frac{1}{2} k^2 L''(0). \quad (6.31)$$

Значение k , обращающее (6.31) в минимум, равно:

$$k^* = -\frac{L'(0)}{L''(0)} = \sigma^2 \frac{\sum 1/\lambda_i^2}{\sum (3\sigma^2 + \gamma_i^2 \lambda_i)/\lambda_i^3}. \quad (6.32)$$

Для нахождения k^* необходимо знание искоемых параметров $\sigma^2, \gamma_1, \dots, \gamma_m$, поэтому непосредственное применение (6.32) невозможно. Однако можно предложить следующую итеративную процедуру:

1) оцениваем исходное уравнение обычным МНК, получаем оценку \mathbf{a} . С помощью преобразований (6.24) и (6.29) переводим ее в оценку \mathbf{g} параметра $\boldsymbol{\gamma}$:

2) находим значение k^* по формуле (6.32), где σ^2 и γ_i заменяются на их оценки s^2 и g_i ;

3) находим на основе k^* ридж-оценку $\mathbf{b}(k^*)$ по формуле (6.27) и возвращаемся на второй шаг. Вычисления продолжаются до тех пор, пока результаты соседних итераций не совпадут.

В табл. 6.1 приводятся ридж-оценки регрессии-примера для разных значений k .

Таблица 6.1

k	Сумма квадратов отклонений (Q)	α_1	α_2	α_3	α_4
10^{-5}	68,6	0,395	0,229	3,77	-17,0
10^{-4}	69,0	0,386	0,226	3,97	-16,3
10^{-3}	91,4	0,336	0,232	5,12	-12,3
10^{-2}	437,9	0,226	0,303	5,89	8,77
10^{-1}	1793,1	0,148	0,393	4,53	36,5
1	21956,0	0,109	0,341	3,41	35,6

На рис. 6.7 показан график изменения суммы квадратов отклонений в зависимости от k . При $k \leq 0,001$ сумма квадратов отклонений растет медленно, далее рост резко увеличивается. Удовлетворительным ридж-оценкам отвечают $k \leq 0,001$.

Применение формулы (6.32) для оптимального k^* привело к сходимости процесса 1)–3) на втором шаге. Значение k^* для регрессии-примера на первом шаге, т. е. с применением оценки МНК, равно $k^* = 0,000029$, $L(k^*) = 36401 < 47332 = L(0)$. Ридж-оценка привела к уменьшению средней суммы квадратов ошибок на 23%. Второй шаг дал практически то же значение k^* , и ридж-оценка оказалась равной $\mathbf{a}(0,000029) = (0,392; 0,228; 3,82; -16,9)'$.

При исследовании ридж-оценок как критерия качества оценки мы рассматривали среднюю сумму квадра-

тов ошибок. Этот критерий имеет один недостаток: веса оценок разных координат совпадают. Более оправданным критерием для несмещенных оценок является сравнение матриц ковариаций оценок.

Для смещенных оценок матрицу ковариаций необходимо заменить на матрицу средних квадратов ошибок, т. е.

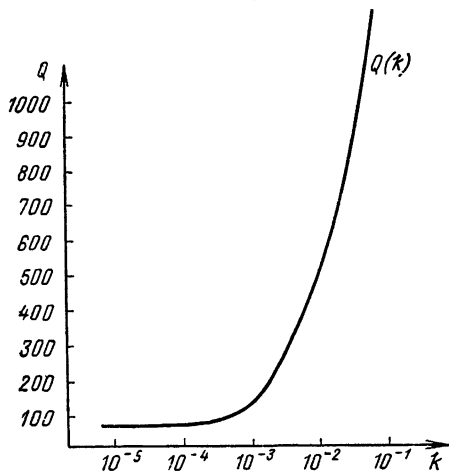


Рис. 6.7. Сумма квадратов отклонений как функция k регрессии-примера

$E(d_j - \alpha)(d_j - \alpha)'$, где d — оценка вектора α . Так, для некоторого k ридж-оценка лучше оценки МНК, если разность между матрицей ковариаций оценки МНК и матрицей средних квадратов ошибок ридж-оценки положительно определена. В [191] показано, что при специальном выборе k ридж-оценка будет лучше оценки МНК и в смысле критерия матрицы средних квадратов ошибок.

Вернемся к проблеме выбора k . А. Гозрл, Р. Кеннард и К. Болдвин [134] предлагают k брать равным $ms^2/g'g$, где g — оценка МНК ортогонального уравнения регрессии (6.29). Этот метод имеет существенный недостаток: при сильной мультиколлинеарности оценка МНК является неустойчивой, т. е. значение $g'g$ велико, что ведет к малому k . Это в свою очередь приведет к тому, что ридж-оценка будет мало отличаться от оценки МНК. В действительности же должно происходить обратное.

Г. Макдональд и Д. Галарню [163] предлагают два варианта выбора k . Предположим, что независимые переменные представлены в стандартизованном виде, т. е. $Z'Z$ — корреляционная матрица. Как и прежде, b — оценка МНК. Легко видеть, что

$$\begin{aligned} E b' b &= E [\beta + (X' X)^{-1} X' \varepsilon]' [\beta + (X' X)^{-1} X' \varepsilon] = \\ &= \beta' \beta + \sigma^2 \text{tr} (X' X)^{-1}, \end{aligned} \quad (6.33)$$

поэтому в качестве оценки $\beta'\beta$ можно взять $l = \mathbf{b}'\mathbf{b} - s^2 \text{tr}(\mathbf{X}'\mathbf{X})^{-1}$.

По первому варианту k выбирается так, чтобы $\|\mathbf{b}(k)\|^2 = l$ для $l > 0$, в противном случае $k = 0$. По второму варианту для $l > 0$ k выбирается так же, как и по первому, а при $l \leq 0$ значение k полагается бесконечно большим, т. е. $\mathbf{b}(k) = \mathbf{0}$. Корректность выбора k в указанных правилах следует из того, что $\|\mathbf{b}(k)\|^2$ — убывающая функция k .

Оба правила направлены на уменьшение длины вектора оценки. Существенным недостатком обоих правил является отсутствие теоретической аргументации; при наличии мультиколлинеарности оценка \mathbf{b} будет не наилучшей, поэтому оценка l также является не весьма удовлетворительной. Свойства оценок МНК, оценок, построенных по описанным правилам, исследовались методом Монте-Карло. При наличии мультиколлинеарности в большинстве случаев ридж-оценки оказываются лучше оценок МНК.

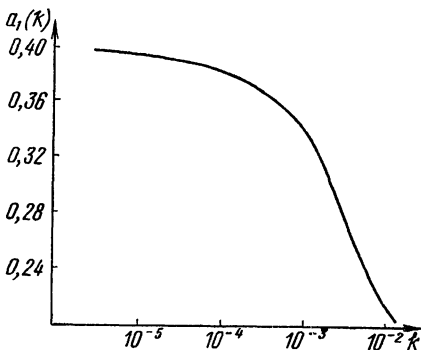


Рис. 6.8. Ридж-след первой координаты $a_1(k)$ — ридж-оценки a_1

А. Гоэрл и Р. Кеннард [131] предлагают строить так называемый «ридж-след» для каждой координаты $b_i(k)$, а k выбирать на основе визуального анализа ридж-следа. Конкретных рекомендаций по выбору k^* они не дают. На рис. 6.8 приведен ридж-след для $a_1(k)$ регрессии-примера. Аргументированно выбрать k на основе этого графика затруднительно. Мы предлагаем исследовать не ридж-след оценок, а ридж-след суммы квадратов отклонений. При этом значение k может быть выбрано следующим образом. При наличии мультиколлинеарности на отрезке $(-\infty, k^*)$ рост суммы квадратов отклонений не очень большой, а на отрезке (k^*, ∞) — значительно выше, полагаем, $k = k^*$.

Остановимся на геометрии ридж-оценки. Рассмотрим для простоты случай $m = 2$. Ридж-оценку будем искать для ортогонализованной модели (6.29). Ридж-оценкой в этом случае являются

$$g_1 = g_1(k) = \frac{d_1}{\lambda_1 + k}; \quad g_2 = g_2(k) = \frac{d_2}{\lambda_2 + k}, \quad (6.34)$$

где $\mathbf{d} = \mathbf{W}'\mathbf{y}$.

При изменении k от 0 до ∞ ридж-оценка $\mathbf{g} = (g_1(k), g_2(k))$ определяет некую кривую. Установим вид этой кривой. Выразим g_2 через g_1 , исключив тем самым k :

$$g_2 = \frac{d_2 g_1}{(\lambda_2 - \lambda_1) g_1 + d_1} = \frac{d_2}{\lambda_2 - \lambda_1} \left(1 - \frac{d_1}{(\lambda_2 - \lambda_1) g_1 + d_1} \right). \quad (6.35)$$

Как следует из (6.35), графиком g_2 от g_1 является гипербола (рис. 6.9). Асимптотой гиперболы является $d_2/(\lambda_2 - \lambda_1)$. Не теряя общности, можно считать $\lambda_2 > \lambda_1$ и $d_1 > 0$, тогда $0 < g_1 < d_1/\lambda_1$. При переходе к модели (6.25) необходимо сделать обратное ортогональное преобразование, т. е. перейти к системе координат (β_1, β_2) .

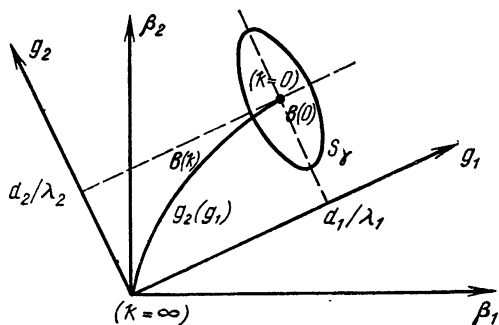


Рис. 6.9. Зависимости $g_2(g_1)$ и $\mathbf{b}(k)$ при $0 < k < \infty$, $\mathbf{b}(\infty) = 0$, S_y — уровень суммы квадратов отклонений, $\mathbf{b}(0) = \mathbf{b}$ — оценка МНК уравнения (6.25)

До сих пор мы рассматривали простейший вариант ридж-оценки, когда в качестве матрицы \mathbf{K} в выражении (6.22) бралась матрица, пропорциональная матрице \mathbf{D} . Рассмотрим более общий случай, когда \mathbf{K} — диагональная матрица, т. е.

$$\mathbf{K} = \begin{bmatrix} k_1 & & & & \\ & \ddots & & & \\ & & 0 & & \\ & & & \ddots & \\ 0 & & & & k_m \end{bmatrix}, \quad k_i \geq 0.$$

Ридж-оценкой для ортогонализованной модели (6.29) является

$$\mathbf{g}(\mathbf{K}) = (\mathbf{W}'\mathbf{W} + \mathbf{K})^{-1}\mathbf{W}'\mathbf{y} = (\mathbf{\Lambda} + \mathbf{K})^{-1}\mathbf{W}'\mathbf{y}.$$

Аналогично оценке (6.27) может быть найден средний квадрат ошибки

$$L(k_1, \dots, k_m) = \sum_{i=1}^m \frac{\sigma^2 \lambda_i + \gamma_i^2 k_i^2}{(\lambda_i + k_i)^2}.$$

Необходимым условием минимума $L(k_1, \dots, k_m)$ является равенство первых производных этой функции нулю, т. е.

$$\frac{\partial L}{\partial k_i} = 2 \frac{k_i \gamma_i^2 - \sigma^2}{(\lambda_i + k_i)^3} = 0, \quad i = 1, \dots, m,$$

откуда

$$k_i = \sigma^2 / \gamma_i^2, \quad i = 1, \dots, m. \quad (6.36)$$

Значения (6.36) дают глобальный минимум $L(k_1, \dots, k_m)$.

Найдем гессиан $L(k_1, \dots, k_m)$:

$$\frac{\partial^2 L}{\partial k_i^2} = 2 \frac{\gamma_i^2 \lambda_i + 3\sigma^2 - 2k_i \gamma_i^2}{(\lambda_i + k_i)^4}, \quad i = 1, \dots, m;$$

$$\frac{\partial L}{\partial k_i \partial k_j} = 0, \quad i \neq j.$$

Таким образом, гессиан L представляет собой диагональную матрицу, которая будет положительно определена в R^m :

$$k_i \leq \frac{\gamma_i^2 \lambda_i + 3\sigma^2}{2\gamma_i^2} = \frac{1}{2} \lambda_i + \frac{3}{2} \frac{\sigma^2}{\gamma_i^2}. \quad (6.37)$$

Поскольку $k_i = \sigma^2 / \gamma_i^2 < 1/2 \lambda_i + 3/2 \sigma^2 / \gamma_i^2$, то решение (6.36) дает локальный минимум на множестве (6.37). Этот минимум будет глобальным, так как для $k_i > \sigma^2 / \gamma_i^2$ функции $2(k_i \gamma_i^2 - \sigma^2) / (\lambda_i + k_i)^3$ — положительные.

На основе (6.36) А. Гозрл и Р. Кеннард [132] предложили итеративную процедуру оценивания: с помощью МНК строится оценка s^2 , находятся значения k_i по формуле (6.36), затем — ридж-оценка, следующие значения k_i и т. д. В. Хеммерл [126] нашел условия сходимости этого процесса, а также дал аналитическую форму предельной оценки.

Обозначим $e_i = \frac{s^2}{\lambda_i \gamma_i^2}$, $i = 1, \dots, m$.

где $g = (g_1, \dots, g_m)'$ — оценка МНК ортогонализованной регрессии (6.29)

Теорема 6.3 [126]. Если $e_i \leq 1/4$, то последовательность k_i^1, k_i^2, \dots , имеет предел, равный

$$k_i^* = \frac{1 - 2e_i - \sqrt{1 - 4e_i}}{2e_i}, \quad i = 1, \dots, m. \quad (6.38)$$

Если $e_i > 1/4$, то $k_i^r \rightarrow \infty$, $r \rightarrow \infty$ ($k_i^* = \infty$).

На основе этой теоремы с учетом (6.36) может быть легко найден аналитический вид предельной оценки:

$$g_i^* = \begin{cases} g_i / (1 + k_i^*), & e_i \leq 1/4; \\ 0, & e_i > 1/4. \end{cases} \quad (6.39)$$

Найдем предельную ридж-оценку для регрессии-примера, используя теорему Хеммерла. Все результаты показаны в табл. 6.2. В данной регрессии $e_4 = 0,503 > 1/4$, поэтому $k_4^* = \infty$ и $g_4^* = 0$. Для $i = 1, 2, 3$ значения k_i^* находятся по формуле (6.38), значение вектора g^* — по формуле (6.39). Последняя строка таблицы есть ридж-оценка исходного уравнения регрессии. Как видим, ридж-оценка сильно отличается от оценки МНК.

Таблица 6.2

Оценка	α_1	α_2	α_3	α_4
$g(0)$	93,9	340,9	315,3	292,9
e	0,00018	0,0047	0,037	0,503
k^*	0,00018	0,0047	0,04	∞
g^*	93,9	393,4	303,3	0
$a(k^*)$	0,279	-0,253	3,50	-56,3

У п р а ж н е н и я 6. 4

1. Докажите, что матрицы A и A^k , где A — симметричная матрица, $k \neq 0$ — целое число (в том числе отрицательное), имеют одинаковые характеристические векторы.

2. Докажите, что длина ридж-оценки (6.27) является убывающей функцией k .

3. Докажите, что $Q(k) = (y - Xa(k))' (y - Xa(k))$; где $a(k)$ — ридж-оценка (6.26), является возрастающей функцией k .

6.5. Редуцированные оценки

Как было показано в предыдущих параграфах, мультиколлинеарность ведет к увеличению длины оценки. Естественно поэтому вместо обычных оценок, т. е. оценок МНК, рассматривать «укороченные», редуцированные оценки, т. е. оценки с меньшей длиной.

В. Джеймс и Ч. Стейн [139] показали, что в случае нормальной случайной выборки в классе всех оценок может быть найдена оценка, которая лучше обычной средней в смысле наименьшего среднего квадрата ошибки. В отличие от средней оценка Джеймса—Стейна является смещенной оценкой. Для регрессии результат Джеймса—Стейна может быть сформулирован следующим образом [181]. Предположим, отклонения регрессии $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ независимы и одинаково распределены $N(0, \sigma^2)$. Для простоты будем считать матрицу плана регрессии $y = W\gamma + \varepsilon$ единичной, т. е. $W'W = I$. Сумму квадратов отклонений, отвечающую оценке МНК, обозначим $Q(g)$; средняя сумма квадратов ошибок оценки МНК тогда равна: $L(g) = E(g - \gamma)'(g - \gamma) = m\sigma^2$.

Т е о р е м а Джеймса—Стейна [181] 6.4. Для $m \geq 3$ оценка Джеймса—Стейна

$$g_{JS} = \left(1 - \frac{cQ(g)}{\|g\|^2}\right) g$$

для всех $0 < c < 2 \frac{m-2}{n-m+2}$ имеет среднюю сумму квадратов ошибок меньше $m\sigma^2$ при любых $\alpha \in R^m$. Наименьшее значение $L(g_{JS})$ достигается при $c^* = (m-2)/(n-m+2)$.

Очевидно, что при $c \neq 0$ для оценки Джеймса—Стейна коэффициент редукции

$$\lambda_{JS} = 1 - \frac{cQ(g)}{\|g\|^2} < 1,$$

поэтому оценку g_{JS} и называют редуцированной (shrunken estimator). В оценке Джеймса—Стейна каждая координата оценки МНК уменьшается в одинаковое число раз. Для оптимального $c^* = (m-2)/(n-m+2)$ $L(c^*) = 2n\sigma^2/(n-m+2) \approx 2\sigma^2$ для больших n .

Отсюда следует, что для больших n оценка Джеймса—Стейна лучше оценки МНК приблизительно в $m/2$ раз. При больших m эффект будет значительным. Для $m = 1$ и 2 , как по-

казано в [183], оценка МНК будет наилучшей в классе всех оценок (смещенных и несмещенных). В [181] приведены формулировки теорем более общих, чем теорема Джеймса—Стейна.

Оценку МНК возможно подправить и для некоторой части вектора. Так, допустим, $m = p + q$, причем p переменных x_1, \dots, x_p считаем предпочтительными и для них не будем исправлять оценку МНК. Разобьем вектор оценки МНК на два подвектора $\mathbf{g} = (\mathbf{g}^1, \mathbf{g}^2)$; первый размерности $p \times 1$, второй — $q \times 1$. Положим $\mathbf{g}_j^2 = \left(1 - \frac{cQ(\mathbf{g})}{\|\mathbf{g}\|^2}\right) \mathbf{g}^2$, где $0 < c < 2(q - 2)/(n - m + 2)$, тогда оценка $\mathbf{g}_{JS} = \left[\frac{\mathbf{g}^1}{\mathbf{g}^2}\right]$ имеет средний квадрат ошибки, меньший, чем средний квадрат ошибки оценки МНК для всех α , σ^2 .

Теорему Джеймса—Стейна нетрудно применить для случая $\mathbf{W}'\mathbf{W} \neq \mathbf{I}$. Действительно, пусть модель задана в исходном виде $\mathbf{y} = \mathbf{X}\alpha + \varepsilon$. Нормируя независимые переменные (6.24), придем к модели (6.25).

Подходящим поворотом осей координат трансформируем модель (6.25) в ортогональную (6.29), при этом $\mathbf{W}'\mathbf{W} = \Lambda$ — диагональная матрица. Далее положим $\mathbf{W}_0 = \mathbf{W}\Lambda^{-1/2}$, тогда $\mathbf{W}_0'\mathbf{W}_0 = \mathbf{I}$ и

$$\mathbf{y} = \mathbf{W}_0\boldsymbol{\gamma}_0 + \varepsilon, \quad (6.40)$$

где $\boldsymbol{\gamma}_0 = \Lambda^{1/2} \boldsymbol{\gamma}$. Находя оценку Джеймса—Стейна для модели (6.40) и делая обратные преобразования, найдем соответствующую оценку для исходной модели. Можно показать, что оценкой Джеймса—Стейна для модели (6.25) является редуцированная оценка $\mathbf{b}_{JS} = \lambda \mathbf{b}$, где \mathbf{b} — оценка МНК регрессии (6.25);

$$\lambda = 1 - \frac{cQ(\mathbf{b})}{\mathbf{b}'\mathbf{Z}'\mathbf{Z}\mathbf{b}}; \quad (6.41)$$

$Q(\mathbf{b})$ — минимальная сумма квадратов уравнения (6.25).

В общем виде редуцированную оценку можно записать в виде $\lambda \mathbf{a}$, где λ — коэффициент редукции $0 < \lambda < 1$. Допуская некоторую вольность, будем считать \mathbf{d} *стохастической редуцированной оценкой*, если λ — стохастический коэффициент, в противном случае $\mathbf{d} = \lambda \mathbf{a}$ — *детерминированная редуцированная оценка*. В последнем случае легко найти математическое ожидание, матрицу ковариаций,

среднюю сумму квадратов ошибок и матрицу средних квадратов ошибки:

$$\begin{aligned} E\mathbf{d} &= \lambda\alpha; \quad \text{cov}(\mathbf{d}) = \lambda^2 \text{cov}(\mathbf{a}) = \sigma^2\lambda^2 (\mathbf{X}'\mathbf{X})^{-1}; \\ L(\mathbf{d}) &= \lambda^2\sigma^2 \text{tr}(\mathbf{X}'\mathbf{X})^{-1} + (1 - \lambda)^2\alpha'\alpha; \\ E(\mathbf{d} - \alpha)(\mathbf{d} - \alpha)' &= \lambda^2\sigma^2 (\mathbf{X}'\mathbf{X})^{-1} + \alpha\alpha'(1 - \lambda)^2. \end{aligned} \quad (6.42)$$

Если \mathbf{d} — стохастическая редуцированная оценка, математическое ожидание, матрицу ковариаций и т. д. найти весьма сложно, так как $\lambda = \lambda(\mathbf{y})$ теперь имеет свое распределение.

Легко показать, что для любого α существует такое $0 < \lambda < 1$, что $L(\lambda\alpha) < L(\mathbf{a})$.

Оценка Джеймса—Стейна является представителем стохастической редуцированной оценки.

Обобщенной редуцированной оценкой Л. Мейер и Т. Уиллке [161] называют оценку, которая является линейным преобразованием оценки МНК, т. е. $\mathbf{d} = \mathbf{C}\mathbf{a}$, где \mathbf{C} — невырожденная матрица $m \times m$. Если \mathbf{C} — детерминированная матрица, говорим, что \mathbf{d} — детерминированная редуцированная оценка, в противном случае — стохастическая. Математическое ожидание, матрица ковариаций, средняя сумма квадратов ошибок и матрица средних квадратов ошибок обобщенной редуцированной оценки $\mathbf{d} = \mathbf{C}\mathbf{a}$ равны соответственно:

$$\begin{aligned} E\mathbf{d} &= \mathbf{C}\alpha; \quad \text{cov}(\mathbf{d}) = \mathbf{C} \text{cov}(\mathbf{a})\mathbf{C}' = \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'; \\ L(\mathbf{d}) &= \sigma^2 \text{tr} \mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C} + \alpha'(\mathbf{C} - \mathbf{I}_m)'(\mathbf{C} - \mathbf{I}_m)\alpha; \\ E(\mathbf{d} - \alpha)(\mathbf{d} - \alpha)' &= \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}' + (\mathbf{C} - \mathbf{I}_m)\alpha\alpha'(\mathbf{C} - \mathbf{I}_m). \end{aligned} \quad (6.43)$$

При использовании обобщенной редуцированной оценки, естественно, возникает вопрос о выборе матрицы преобразования \mathbf{C} . Мейер и Уиллке поступают следующим образом. Как известно, при заданном значении суммы квадратов отклонений ридж-оценка минимизирует длину оценки. Аналогичный путь можно предложить для получения матрицы \mathbf{C} . Нетрудно проверить, что

$$\begin{aligned} Q(\mathbf{d}) &= (\mathbf{y} - \mathbf{X}\mathbf{d})'(\mathbf{y} - \mathbf{X}\mathbf{d}) = (\mathbf{y} - \mathbf{X}\mathbf{C}\mathbf{a})'(\mathbf{y} - \mathbf{X}\mathbf{C}\mathbf{a}) = \\ &= (\mathbf{y} - \mathbf{X}\mathbf{a})'(\mathbf{y} - \mathbf{X}\mathbf{a}) + \mathbf{a}'(\mathbf{C} - \mathbf{I})'\mathbf{X}'\mathbf{X}(\mathbf{C} - \mathbf{I})\mathbf{a} = \\ &= Q(\mathbf{a}) + \mathbf{a}'(\mathbf{C} - \mathbf{I})'\mathbf{X}'\mathbf{X}(\mathbf{C} - \mathbf{I})\mathbf{a}. \end{aligned} \quad (6.44)$$

Очевидно, обобщенной детерминированной редуцированной оценкой $\mathbf{d} = \mathbf{C}\mathbf{a}$, имеющей данную сумму квадратов

отклонений, с минимальной длиной является ридж-оценка, для которой

$$C = (X'X + kI)^{-1}X'X = (I_m + k(X'X)^{-1})^{-1}.$$

Как показали Мейер и Уиллке [161], обобщенной редуцированной оценкой с данной суммой квадратов отклонений (6.44) и минимальным следом $\text{tr cov}(\mathbf{d}) = \sigma^2 \text{tr } C(X'X)^{-1}C'$ является оценка

$$\mathbf{d} = \delta \mathbf{a}\mathbf{a}'(\mathbf{I} + \delta \mathbf{a}\mathbf{a}')^{-1}\mathbf{a}, \quad (6.45)$$

где $\delta > 0$ определяется данным значением $Q(\mathbf{d})$. Оценка (6.45) может быть преобразована следующим образом. Матрицы $\mathbf{a}\mathbf{a}'$ и $\mathbf{I} + \delta \mathbf{a}\mathbf{a}'$ имеют одинаковые характеристические векторы, поэтому $\mathbf{a}\mathbf{a}' = \mathbf{P}\Lambda_1\mathbf{P}'$; $\mathbf{I} + \delta \mathbf{a}\mathbf{a}' = \mathbf{P}\Lambda_2\mathbf{P}'$, где Λ_1 и Λ_2 — диагональные матрицы, причем

$$\Lambda_1 = \begin{bmatrix} \mathbf{a}'\mathbf{a} & & & \\ & & & \\ & & & \\ & & & 0 \end{bmatrix}; \quad \Lambda_2 = \begin{bmatrix} 1 + \delta \mathbf{a}'\mathbf{a} & & & \\ & 1 & & 0 \\ & & \ddots & \\ & & & 1 \\ & 0 & & & \\ & & & & & 1 \end{bmatrix}.$$

Тогда с учетом этого

$$\begin{aligned} \mathbf{d} &= \delta \mathbf{a}\mathbf{a}'(\mathbf{I} + \delta \mathbf{a}\mathbf{a}')^{-1}\mathbf{a} = \delta \mathbf{P}\Lambda_1\mathbf{P}'\mathbf{P}\Lambda_2^{-1}\mathbf{P}'\mathbf{a} = \\ &= \delta \mathbf{P}\Lambda_1\Lambda_2^{-1}\mathbf{P}'\mathbf{a} = \frac{\delta}{1 + \delta \mathbf{a}'\mathbf{a}} \mathbf{P}\Lambda_1\mathbf{P}'\mathbf{a} = \frac{\delta}{1 + \delta \mathbf{a}'\mathbf{a}} \mathbf{a}\mathbf{a}'\mathbf{a} = \\ &= \frac{\delta \mathbf{a}'\mathbf{a}}{1 + \delta \mathbf{a}'\mathbf{a}} \mathbf{a} = \lambda \mathbf{a}, \quad \lambda = \frac{\delta \mathbf{a}'\mathbf{a}}{1 + \delta \mathbf{a}'\mathbf{a}}. \end{aligned} \quad (6.46)$$

Как видим, оценка (6.46) есть не что иное, как стохастическая редуцированная оценка, коэффициент редукиции $0 < \lambda < 1$. Напомним, что δ зависит от выбранного значения суммы квадратов отклонений.

Мейер и Уиллке минимизировали след ковариационной матрицы редуцированной оценки. Можно выбрать и другие критерии. Например, в качестве критерия можно взять среднюю сумму квадратов ошибки. Найдем для данных α и σ^2 значение λ , минимизирующее $L(\mathbf{d}) = L(\lambda\alpha)$. Имеем (см. (6.43))

$$dL/d\lambda = 2\lambda\sigma^2 \text{tr}(X'X)^{-1} - 2(1 - \lambda)\alpha'\alpha = 0,$$

¹Мейер и Уиллке при преобразовании допустили ошибку; их ошибочная оценка равна $\delta(\mathbf{a}'\mathbf{a} + \delta(\mathbf{a}'\mathbf{a})^2)/(1 + \delta\mathbf{a}'\mathbf{a})\mathbf{a}$. При доказательстве (6.46) можно воспользоваться формулой (П.4).

откуда

$$\lambda^* = \frac{\alpha' \alpha}{\alpha' \alpha + \sigma^2 \text{tr} (X' X)^{-1}}. \quad (6.47)$$

Очевидно, для данных α и σ^2

$$L(a_{JS}) = L(\lambda^* a) < L(a) = \sigma^2 \text{tr} (X' X)^{-1}.$$

Использовать коэффициент редукции (6.47) на практике невозможно, поскольку он содержит неизвестные параметры. Однако вместо α и σ^2 можно подставить их оценки, например можно положить

$$\lambda^* = \frac{a' a}{a' a + s^2 \text{tr} (X' X)^{-1}},$$

что приведет к стохастической редуцированной оценке $a_{JS} = \lambda^* a$. Оценку a_{JS} можно использовать для нахождения следующего значения λ^* , по которому найдем новое значение a_{JS} , и т. д.

Между ридж-оценками и редуцированными оценками существует тесная взаимосвязь: ридж-оценки являются частным случаем редуцированных оценок [113]. Действительно, сведем исходную модель к ортогональной (6.29). Для нее обобщенной ридж-оценкой является

$$g(K) = (W'W + K)^{-1} W'y = (\Lambda + K)^{-1} W'y,$$

где K — диагональная матрица. Обозначим $W'y = d$, тогда

$$g_i(K) = d_i / (\lambda_i + k_i), \quad i = 1, \dots, m. \quad (6.48)$$

Обобщенной редуцированной оценкой для (6.29) является $g_r = Lg$, где L будем считать диагональной матрицей. Тогда $(g_r)_i = L_i g_i = L_i d_i / \lambda_i$. Если в качестве L_i взять $\lambda_i / (\lambda_i + k_i)$, то приходим к ридж-оценке (6.48).

У п р а ж н е н и е 6.5

1. Докажите, что для некоторой области $\alpha \in R^m$ детерминированная редуцированная оценка будет иметь меньшую ССКО, чем оценка МНК.

6.6. Оценка метода главных компонент

Эта оценка впервые была предложена М. Кендэллом [145]¹.

Как и прежде, вместо исходной модели регрессии рассмотрим преобразованную ортогональную модель

$$y = W\gamma + \varepsilon, \quad (6.49)$$

где $\Lambda = W'W$ — диагональная матрица с элементами на диагонали, равными характеристическим числам матрицы $Z'Z$ (см. уравнения (6.29)), причем $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m$. Оценками МНК для (6.49) являются $g_i = d_i/\lambda_i$, $i = 1, \dots, m$. Если x_1, \dots, x_m сильно коррелируют, то для некоторого r λ_{m-r} близко к нулю, а значения g_i , $i = 1, \dots, m-r$ будут большими по абсолютной величине. Дисперсии оценок, соответствующих малым значениям λ_i , т. е. σ^2/λ_i , также будут велики; таким образом, оценки МНК при малых λ_i будут неудовлетворительными. Идея оценки метода главных компонент (principal components estimator) заключается в том, чтобы для малых значений λ_i соответствующую координату оценки считать равной нулю; более точно оценка главных компонент есть

$$g_{pi}(r) = \begin{cases} 0, & i \leq m-r; \\ g_i, & i > m-r. \end{cases}$$

Оценку $g_p(r)$ можно рассматривать как частный случай обобщенной редуцированной оценки, матрица редукции которой диагональна, диагональные элементы равны 0 или 1. Производя обратное преобразование над регрессией (6.49), получим оценку главных компонент для моделей (6.25) и (1.2):

$$b_p(r) = P g_p(r); \quad a_p(r) = D^{-1} b_p(r).$$

Оценка метода главных компонент имеет много общего с оценкой автоматического отсева переменных. Оба метода предлагают считать оценкой некоторой координаты нуль, если первоначальная оценка оказалась неудовлетворительной. Разница заключается в том, что отсев переменных производится для исходной модели, а оценка метода главных компонент «работает» с преобразованной моделью (6.49). Легко показать, что отсев переменных на основе t -статистики не инвариантен относительно линейных преобразова-

¹Более подробно см. [116, 125, 160].

ний модели. Далее, при отсеке переменных основным показателем является t -статистика оценки МНК, а в методе главных компонент — величина характеристического числа матрицы λ_i .

Можно проверить, что оценка метода главных компонент $\mathbf{b}_p(r)$ для модели $\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ выражается в виде линейной комбинации характеристических векторов $\mathbf{p}_1, \dots, \mathbf{p}_m$ матрицы $\mathbf{Z}'\mathbf{Z}$ следующим образом:

$$\mathbf{b}_p(r) = \sum_{i=m-r+1}^m d_i \mathbf{p}_i / \lambda_i. \quad (6.50)$$

На основе (6.50) нетрудно определить основные статистические характеристики оценки $\mathbf{b}_p(r)$:

$$E\mathbf{b}_p(r) = \boldsymbol{\beta} - \sum_{i=1}^{m-r} (\mathbf{p}_i' \boldsymbol{\beta}) \mathbf{p}_i; \quad (6.51)$$

$$\text{cov}(\mathbf{b}_p(r)) = \sigma^2 \sum_{i=m-r+1}^m \mathbf{p}_i \mathbf{p}_i' / \lambda_i; \quad (6.52)$$

$$L(\mathbf{b}_p(r)) = E(\mathbf{b}_p(r) - \boldsymbol{\beta})' (\mathbf{b}_p(r) - \boldsymbol{\beta}) = \sigma^2 \sum_{i=m-r+1}^m 1/\lambda_i + \\ + \sum_{i=1}^{m-r} (\mathbf{p}_i' \boldsymbol{\beta})^2. \quad (6.53)$$

Заметим, что ССКО оценки МНК равна:

$$L(\mathbf{b}_p(m)) = L(\mathbf{b}) = \sigma^2 \sum_{i=1}^m 1/\lambda_i. \quad (6.54)$$

Учитывая (6.53), легко сделать следующий вывод: отбрасывая в (6.54) первые $m - r$ членов, мы выигрываем в первой сумме квадратов ошибок (6.53) и проигрываем во второй. Чем меньше значения $\lambda_1, \dots, \lambda_{m-r}$, тем шире область $\boldsymbol{\beta} \in R^m$, где ССКО оценки метода главных компонент меньше ССКО оценки МНК. Помимо значений вектора неизвестных параметров $\boldsymbol{\beta}$ для $L(\mathbf{b}_p(r))$, имеет значение направление \mathbf{p}_i и $\boldsymbol{\beta}$.

На рис. 6.10 показаны оценки метода главных компонент $\mathbf{g}_p(0)$, $\mathbf{g}_p(1)$ и $\mathbf{g}_p(2)$. В системе координат (γ_1, γ_2) эти оценки относятся к модели (6.49), в системе (β_1, β_2) — к модели (6.25).

Как следует из параграфа 6.2, обобщенная обратная матрица к $Z'Z$ равна:

$$(Z'Z)_r^+ = P\Lambda_r^+ P', \quad (6.56)$$

где

$$\Lambda_r^+ = \begin{bmatrix} 0 & & & & 0 \\ & \ddots & & & \\ & & 1/\lambda_{m-r+1} & & \\ & & & \ddots & \\ 0 & & & & 1/\lambda_m \end{bmatrix}; \quad (6.57)$$

(i, j) -й элемент матрицы $(Z'Z)_r^+$ равен

$$\sum_{k=m-r+1}^m P_{ik} P_{jk} / \lambda_k.$$

Далее Марквардт вводит понятие обратной матрицы с дробным рангом. Определим сначала обратную матрицу к диагональной матрице Λ с дробным рангом r . Итак, пусть $0 \leq r \leq m$, обозначим $e = [r]$ — целая часть r , $\theta = r - [r]$ — дробная часть. Зададим Λ_r^+ для любого действительного r следующим образом:

$$\Lambda_r^+ = \begin{bmatrix} 0 & & & & 0 \\ & \ddots & & & \\ & & \theta/\lambda_{m-e} & & \\ & & & 1/\lambda_{m-e+1} & \\ & & & & \ddots \\ 0 & & & & & 1/\lambda_m \end{bmatrix}. \quad (6.58)$$

Разница в (6.57) и (6.58) заключается в том, что мы вводим в (6.58) член $1/\lambda_{m-e}$ с коэффициентом, равным дробной части r . Под обратной матрицей к $Z'Z$ с приписываемым дробным рангом r будем понимать (6.56), где Λ_r^+ задается (6.58). Очевидно, в этом случае

$$((Z'Z)_r^+)_{ij} = \sum_{k=m-e+1}^m P_{ik} P_{jk} / \lambda_k + \theta P_{i, m-e} P_{j, m-e} / \lambda_{m-e}.$$

Оценка Марквардта есть функция выбранного ранга r и равна:

$$b_M(r) = (Z'Z)_r^+ Z' y. \quad (6.59)$$

Оценкой α для исходной модели (1.2) является

$$a_r^+ = a_M(r) = D^{-1/2} b_M(r). \quad (6.60)$$

Между оценкой Марквардта и оценкой метода главных компонент существует тесная взаимосвязь.

Теорема 6.5. Если r — целое, то оценки совпадают, в противном случае первая оценка выражается через вторую следующим образом:

$$a_M(r) = (1 - \theta)a_P(e) + \theta a_P(e + 1).$$

Доказательство. Сначала проведем его для ортогонализованной модели $y = W\gamma + \varepsilon$, где $W'W = \Lambda$, т. е. докажем

$$g_M(r) = (1 - \theta)g_P(e) + \theta g_P(e + 1). \quad (6.61)$$

По определению

$$g_M(r) = \begin{cases} 0, & i < m - e; \\ \theta d_{m-e}/\lambda_{m-e}, & i = m - e; \\ d_i/\lambda_i, & i > m - e. \end{cases}$$

Покоординатно расписывая правую часть (6.61), придем к выражению, которое и необходимо было доказать. Поскольку оценки совпали для ортогонализованной модели, то они совпадут и для исходной модели (1.2).

Как следует из доказанной теоремы, оценки Марквардта лежат на отрезке, соединяющем оценки главных компонент ближайших рангов e и $e + 1$. На рис. 6.10 оценки Марквардта лежат на отрезке NP .

Приведем основные свойства оценки Марквардта.

1. Квадрат длины оценки (6.61) есть возрастающая функция r , т. е. $\|a_r^{\dagger}\|^2$ растет с ростом r .

2. Оценка (6.60) есть линейная функция оценки МНК:

$$a_r^{\dagger} = (X'X)_r^{\dagger} X'X\alpha.$$

3. Оценка (6.60) смещена:

$$Ea_r^{\dagger} = (X'X)_r^{\dagger} X'X\alpha \neq \alpha;$$

матрица ковариаций оценки Марквардта равна:

$$\text{cov}(a_r^{\dagger}) = \sigma^2 (X'X)_r^{\dagger} X'X (X'X)_r^{\dagger}.$$

4. Аналогично ридж-оценке можно найти среднюю сумму квадратов ошибок оценки (6.60):

$$L(r) = L_{a_r^{\dagger}}(r) = E(a_r^{\dagger} - \alpha)'(a_r^{\dagger} - \alpha) = \alpha' S' S \alpha + \\ + \sigma^2 \text{tr} X (X'X)_r^{\dagger} (X'X)_r^{\dagger} X';$$

где

Поэтому если условие теоремы выполнено, то $L(r) < L(0)$.

Доказанная теорема показывает, что в некоторой области (β, σ^2) оценки Марквардта будут лучше оценок МНК. Если некоторые характеристические числа матрицы $Z'Z$ близки к нулю, то эта область будет значительной; другими словами, при наличии мультиколлинеарности оценки Марквардта, весьма вероятно, будут лучше оценок МНК. Для задания области (β, σ^2) можно привлечь некоторые априорные знания о возможных значениях β и σ^2 . Так, вполне возможно, что имеет смысл оценивать β только в некотором интервале $c_i \leq \beta_i \leq d_i$, $i = 1, \dots, m$. Далее, можно предположить, что $\sigma \geq \sigma_0$ — точности «измерения» y, x_1, \dots, x_m , поэтому можно считать

$$\beta' \beta / \sigma^2 \leq M = \sum_i \max(c_i^2, d_i^2) / \sigma_0^2.$$

Теперь, если найдено такое r , что $\sum_{i=1}^{m-r} 1/\lambda_i \geq M$, то можно утверждать, что, если наши априорные предположения верны, оценка $b_M(r)$ будет лучше оценки МНК. Оценка для $\beta' \beta$ может быть получена также на основе равенства

$$E b' b = \beta' \beta + \sigma^2 \sum_i 1/\lambda_i,$$

откуда

$$\beta' \beta = E b' b - \sigma^2 \sum_{i=1}^m 1/\lambda_i \leq E b' b. \quad (6.62)$$

В качестве оценки $\beta' \beta$ можно взять просто $b' b$. Если $\lambda_1 \approx 0$, то неравенство (6.62) даст приемлемую область для $\beta' \beta$. В качестве оценки σ^2 может быть выбрана s^2 , рассчитанная на основе МНК.

Найдем оценку Марквардта для регрессии-примера. В табл. 6.3 приводятся х. ч и х. в. матрицы $Z'Z$, при этом $\lambda_1 = 0,000144$; $\lambda_4 = 3,983$ и $|Z'Z| = 1,45 \cdot 10^{-8}$, $\lambda_4/\lambda_1 = 27660$. Полученные характеристики матрицы $Z'Z$ говорят о наличии мультиколлинеарности. Далее, $\lambda_1/4 = 3,6 \cdot 10^{-5}$, поэтому можно предположить, что ранг $Z'Z$ равен 3, т. е. считать λ_1 равным нулю.

Таблица 6.3

Характеристические векторы					Характеристические числа
p_1	0,1546	-0,8273	-0,1586	0,5162	0,000144
p_2	-0,7127	0,00516	0,7014	0,00617	0,00171
p_3	-0,4671	0,2543	-0,4826	0,6959	0,0148
p_4	0,500	0,5008	0,4999	0,4992	3,983

Используя формулу (6.56), находим

$$\begin{aligned}
 & (\mathbf{Z}' \mathbf{Z})_{\frac{1}{3}} = \\
 & = \begin{bmatrix} 316,4 & -12,44 & -272,9 & -30,83 \\ -12,44 & 5,715 & -8,51 & 15,5 \\ -272,9 & -8,51 & 308,4 & -26,66 \\ -30,83 & 15,5 & -26,66 & 42,77 \end{bmatrix}.
 \end{aligned}$$

Оценка Марквардта равна:

$$\mathbf{b}_M(3) = (568,8; 63,77; 206,6; -131,6)';$$

оценка исходной модели:

$$\mathbf{a}_M(3) = (0,430; 0,155; 4,90; -34,0)'.$$

Оценка Марквардта с приписанным рангом 3 значительно отличается от оценки МНК.

Подойдем к проблеме смещенного оценивания регрессии с более общих позиций [127]. Как и прежде, остановимся на ортогонализованной модели $\mathbf{y} = \mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$, где по определению $\mathbf{W}'\mathbf{W} = \mathbf{\Lambda}$ — диагональная матрица с диагональными элементами $\lambda_1 \leq \dots \leq \lambda_m$. Оценка МНК для этой модели равна $\mathbf{g} = \mathbf{\Lambda}^{-1}\mathbf{d}$, где $\mathbf{d} = \mathbf{W}'\mathbf{y}$, т. е. $g_i = d_i/\lambda_i$, $i = 1, \dots, m$.

Для ортогонализованной модели рассмотрим систему m оценок метода главных компонент $\mathbf{g}_p(1), \mathbf{g}_p(2), \dots, \mathbf{g}_p(m)$. Эта система оценок почти наверное линейно-независима в R^m , поэтому любая оценка $\mathbf{e} \in R^m$ представима в виде линейной комбинации оценок метода главных компонент, а именно:

$$\mathbf{e} = \sum_{j=1}^m c_j \mathbf{g}_p(j). \quad (6.63)$$

Зависимость (6.63) может быть переписана следующим образом:

$$\mathbf{e} = \mathbf{F}\mathbf{g}, \quad (6.64)$$

где \mathbf{F} — диагональная матрица, причем

$$F_{ii} = f_i = \sum_{j=m-i+1}^m c_j, \quad (6.65)$$

а \mathbf{g} — оценка МНК. Для оценки (6.64) легко найти среднюю сумму квадратов ошибок

$$L(\mathbf{e}) = \sigma^2 \sum_{i=1}^{m_1} f_i^2 / \lambda_i + \sum_{i=1}^m \gamma_i^2 (f_i - 1)^2 \quad (6.66)$$

и среднюю взвешенную сумму квадратов ошибок

$$L_{\Lambda} = \sigma^2 \sum_{i=1}^m f_i^2 + \sum_{i=1}^m \gamma_i^2 \lambda_i (f_i - 1)^2. \quad (6.67)$$

Очевидно, оценка Марквардта представима как в виде (6.63), так и в виде (6.67). Применяя теорему 6.4, находим, что

$$c_j = \begin{cases} 0, & \text{если } j < e; \\ 1 - \theta, & \text{если } j = e; \\ \theta, & \text{если } j = e + 1; \\ 0, & \text{если } j > e + 1. \end{cases}$$

В обозначениях матрицы \mathbf{F} коэффициенты равны:

$$f_j = \begin{cases} 0, & \text{если } j < m - e; \\ \theta, & \text{если } j = m - e; \\ 1, & \text{если } j > m - e. \end{cases}$$

Функция (6.66) для оценки $g_M(r)$ переписется следующим образом:

$$L(g_M(r)) = \sigma^2 \sum_{i=m-e+1}^m \frac{1}{\lambda_i} + \sigma^2 \frac{\theta^2}{\lambda_{m-e}} + \gamma_{m-e}^2 (1 - \theta)^2 + \sum_{i=1}^{m-e-1} \gamma_i^2. \quad (6.68)$$

Выражение (6.68) может быть использовано для нахождения оптимального ранга r . Допустим, целая часть $[r] = e$ известна и необходимо найти только дробную часть

θ . Минимизируя (6.68) относительно θ , найдем оптимальное значение

$$\theta^* = \tau_{m-e}^{21} (1 + \tau_{m-e}^2), \quad (6.69)$$

где $\tau_i^2 = \gamma_i^2 \lambda_i / \sigma^2$, $i = 1, \dots, m$. После того как дробная часть найдена, займемся отысканием оптимальной целой части $e = [r]$. Для этого подставим θ^* в (6.68). После несложных преобразований получим

$$L = L(\mathbf{g}) - \sigma^2 \left[\sum_{i=1}^{m-e} \frac{1 - \tau_i^2}{\lambda_i} + \frac{1}{\lambda_{m-e} (1 - \tau_{m-e}^2)} \right], \quad (6.70)$$

где $L(\mathbf{g})$ — средняя сумма квадратов для оценки МНК \mathbf{g} . Как видно из (6.70), уменьшение средней суммы квадратов ошибок оценки Марквардта зависит от значений τ_i^2 . Если $\tau_i^2 < 1$, $i = 1, \dots, m - e$, выражение в квадратных скобках (6.70) будет положительным, значит, $\mathbf{g}_M(r)$ лучше оценки МНК. Однако гарантий того, что $\tau_i^2 < 1$, у нас нет, поскольку порядок возрастания $\lambda_1, \lambda_2, \dots, \lambda_m$ не совпадает, вообще говоря, с порядком возрастания $\tau_1^2, \tau_2^2, \dots, \tau_m^2$. Если же х. ч. расположить в порядке возрастания $\tau_1^2, \tau_2^2, \dots, \tau_m^2$, то оптимальный целый ранг e будет равен числу τ_i^2 , меньших 1 (если, конечно, таковые существуют). Другими словами, для номеров j : $\tau_j^2 < 1$ будем считать $\lambda_j = 0$. Этот метод предложен в [127], назовем его *модифицированным методом Марквардта*.

Разумеется, в изложенном виде модифицированный метод Марквардта неприменим, поскольку значения σ^2 , γ нам неизвестны, а задача как раз заключается в их оценивании. Однако можно взять их оценки, например s^2 и \mathbf{g} . Что касается первой оценки, то она, как было показано, является удовлетворительной даже при наличии мультиколлинеарности. Таким образом, оцененные значения τ_i^2 равны: $\tau_i^2 = g_i^2 \lambda_i / s_i^2$, $i = 1, \dots, m$. Но $\tau_i = |g_i| / (s_i / \lambda_i) = |g_i| / s_i$, где s_i — стандартная ошибка оценки g_i , т. е. τ_i есть t -статистика i -го параметра. Итак, в модифицированном методе Марквардта ранжирование производится не по значениям λ_i , а по значениям t -статистик, причем критической величиной t -статистики является 1.

Допустим, значение e выбрано, тогда, найдя значение θ^* по формуле (6.69), можно найти новый вектор $\mathbf{g}_M(e + \theta^*)$, на основе которого получим следующую поправку к рангу e , и т. д. Будет ли сходиться этот процесс? Положительный ответ дается следующей леммой.

Л е м м а 6.2 [127]. Пусть p_0, p_1, \dots — последовательность чисел, таких, что $p_{i+1} = p_i^2 / (p_i + P)$, $i = 0, 1, \dots$. Обозначим $p^* = \lim p_i$, тогда

1) если $P > 1/4$, то $p^* = 0$;

2) если $P \leq 1/4$, то

а) если $p_0 > p''$, то $p^* = p'$;

б) если $p_0 < p''$, то $p^* = 0$;

в) если $p_0 = p''$, то $p^* = p''$,

где $p' = 1/2 + \sqrt{1/4 - P}$, $p'' = 1/2 - \sqrt{1/4 - P}$.

Используя эту лемму, можно найти предельное значение θ^* . Пусть θ_k^* — значение θ^* на k -й итерации, тогда $(g_M^k)_{m-e} = \theta_k^* g_{m-e}$, поэтому

$$\theta_{k+1}^* = \frac{\theta_k^{*2} g_{m-e}^2 \lambda_{m-e} / s^2}{1 + \theta_k^{*2} g_{m-e}^2 \lambda_{m-e} / s^2} = \frac{\theta_k^{*2}}{\theta_k^{*2} + P},$$

где $P = s^2 / g_{m-e}^2 \lambda_{m-e}$, $p_0 = 1 / (1 + P)$. Таким образом, предельное значение

$$\theta^* = \begin{cases} 0, & \text{если } t_{m-e} = |g_{m-e}| \lambda_{m-e} / s < 2; \\ \frac{1}{2} + \sqrt{\frac{1}{4} - P} & \text{— в противном случае,} \end{cases}$$

где t_{m-e} — статистика для $(m - e)$ -го параметра. Случай $\theta^* = 0$ соответствует оценке метода главных компонент. На основе предельного значения θ^* можно пересчитать значения τ_i^2 , которые могут привести и к другому выбору нулевых координат. Эти итерации будут продолжаться до тех пор, пока $t_{m-e} < 2$.

6.8. Оценка Хокинса [124]

Эта оценка по идее близка к оценке метода главных компонент. Суть ее заключается в следующем. Объединим вектор зависимой переменной y и матрицу независимых переменных Z в одну матрицу $n \times (m + 1)$, т. е. положим $T = [yZ]$. Пусть P — ортогональная матрица порядка $(m + 1) \times (m + 1)$, сводящая $T'T$ к диагональной, а $\lambda_1, \dots, \lambda_{m+1}$ — характеристические числа матрицы $T'T$. Разобьем матрицу следующим образом: $P = \begin{bmatrix} P_0 \\ P_1 \end{bmatrix}$, где P_0 — вектор-строка размерности $(m + 1)$, P_1 — матрица $m \times (m + 1)$, тогда

$$T'T = \begin{bmatrix} y'y & y'X \\ X'y & X'X \end{bmatrix} = \begin{bmatrix} P_0 \\ P_1 \end{bmatrix} \Lambda \begin{bmatrix} P_0' & P_1' \end{bmatrix} = \begin{bmatrix} P_0 \Lambda P_0' & P_0 \Lambda P_1' \\ P_1 \Lambda P_0' & P_1 \Lambda P_1' \end{bmatrix}. \quad (6.71)$$

Очевидно;

$$(T' T)^{-1} = P \Lambda^{-1} P' = \begin{bmatrix} p_0 \Lambda^{-1} p'_0 & p_0 \Lambda^{-1} p'_1 \\ p_1 \Lambda^{-1} p'_0 & p_1 \Lambda^{-1} p'_1 \end{bmatrix}. \quad (6.72)$$

Выразим оценку МНК через элементы матриц P и Λ . Для этого рассмотрим аналог формулы Фробениуса (П. 2). Пусть имеется блочная симметричная матрица

$$\begin{bmatrix} A & B \\ B' & D \end{bmatrix}. \quad (6.73)$$

Обратная к ней будет также симметричной, обозначим ее

$$\begin{bmatrix} C & E \\ E' & F \end{bmatrix}.$$

По определению

$$\begin{bmatrix} A & B \\ B' & D \end{bmatrix} \begin{bmatrix} C & E \\ E' & F \end{bmatrix} = \begin{bmatrix} AC + BE' & AE + BF \\ B' C + DE' & B' E + DF \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix},$$

откуда

$$-D^{-1} B' = E' C. \quad (6.74)$$

Пусть матрицы (6.71) и (6.73) совпадают, в частности $D = X'X$, $B' = Xu$. В обозначениях (6.73) оценка МНК равна $D^{-1}B'$, откуда с применением (6.72) и (6.74) следует

$$b = -E' C^{-1} = -P_1 \Lambda^{-1} p'_0 / p_0 \Lambda^{-1} p'_0. \quad (6.75)$$

Обозначим i -й вектор-столбец матрицы P_1 через p_i ; тогда (6.75) перепишется:

$$b = -\frac{1}{\sum_{i=1}^{m+1} p_{0i}^2 / \lambda_i} \sum_{i=1}^{m+1} p_i p_{0i} / \lambda_i.$$

Ясно, что если $\lambda_i \approx 0$, то оценка МНК будет неустойчивой. Поэтому, как и в оценке метода главных компонент, члены, содержащие $\lambda_i \approx 0$, аннулируем. Итак, приходим к модифицированной оценке МНК — оценке Хокинса¹:

$$b_L(s) = -\frac{1}{\sum_{i=s+1}^{m+1} p_{0i}^2 / \lambda_i} \sum_{i=s+1}^{m+1} p_i p_{0i} / \lambda_i. \quad (6.76)$$

Существует несколько причин ограниченности применения оценки (6.76). Во-первых; неясно, чем эта оценка лучше оценки метода главных компонент; во-вторых; не было предложений по вы-

¹ В иностранной литературе ее часто называют latent root — оценка характеристического корня, этим объясняется нижний индекс у оценки.

бору числа x . ч.; которые можно считать равными нулю; т. е. выбору s ; в-третьих; невозможно найти аналитический вид для математического ожидания оценки Хокинса; матрицы ковариации; средней суммы квадратов ошибок.

6.9. Сравнение оценок методом статистических испытаний

Рассмотрим результаты исследования, проведенного Дж. Вебстером, Р. Гунстом и Р. Мазоном [117], по выявлению предпочтительной оценки из 5 оценок: МНК, оценки метода главных компонент, ридж-оценки, редуцированной оценки и оценки Хокинса. Сравнение проводилось на основе метода статистических испытаний (метода Монте-Карло). При этом варьировались три характеристики: 1) ориентация неизвестного вектора β по отношению к характеристическим векторам матрицы плана $Z'Z$; 2) относительная длина вектора β , т. е. $\beta'\beta/\sigma^2$; 3) степень мультиколлинеарности. Во всех расчетах значение n было равно 30, $m = 10$. Матрица Z выбиралась двумя способами: в первом способе столбцы матрицы были почти ортогональны (отсутствие мультиколлинеарности), во втором — первый столбец представлял собой с небольшими отклонениями линейную комбинацию второго, третьего и четвертого (наличие мультиколлинеарности). Длина «истинного» вектора β во всех расчетах была равна 1, при этом σ^2 принимало четыре значения, так что $\rho = \beta'\beta/\sigma^2 : 0,04; 1; 100; 10000$. Вектор β принимал три значения: $\beta = p_{10}$, $\beta = (p_1 + p_2 + p_9 + p_{10})/2$, $\beta = p_1$; косинус угла между β и p_1 поэтому был равен 0; 0,5 и 1 соответственно. Таким образом, имелось $2 \cdot 4 \cdot 3 = 24$ варианта. Для каждого варианта проводилось 100 испытаний, на основе которых вычислялась средняя сумма квадратов ошибок, деленная на σ^2 , т. е. $L(d)/\sigma^2$. В ридж-оценке значение k выбиралось равным $ms^2/b'b$; представителем редуцированной оценки была выбрана оценка Джеймса—Стейна. В «почти ортогональной» модели для оценки метода главных компонент считалось $r = m$, в случае мультиколлинеарности $r = m - 1$, т. е. λ_1 равно нулю.

Очевидно, в почти ортогональной модели оценка МНК совпадает с оценкой главных компонент и оценкой Хокинса. Во всех вариантах наилучшей оценкой в «почти ортогональной» модели оказалась ридж-оценка. Наиболее заметное преимущество эта оценка (по сравнению с оценками

b , $b_p(m)$, $b_L(m+1)$) имела в случае небольших значений $\rho = \beta'\beta/\sigma^2$; для предельного $\rho = 10000$ все оценки привели приблизительно к одной и той же ССКО. Оценка Джеймса—Стейна была немного хуже ридж-оценки, но опять же намного лучше оценок МНК для $\rho < 10000$. Полученный результат можно было предсказать из теоретических соображений: для ортогональных моделей оценка Джеймса—Стейна имеет ССКО меньшую, чем оценка МНК, для всех β . Ридж-оценка есть частный случай редуцированной оценки.

В случае мультиколлинеарности положение несколько изменилось. Во-первых, оценка Хокинса во всех расчетах имела сходные свойства с оценкой метода главных компонент, поэтому будем говорить только о последней. Во-вторых, оценка Джеймса—Стейна была лишь немногим лучше оценки МНК во всех ситуациях. Оценка метода главных компонент и оценка Хокинса имели наименьшую ССКО в случае $\rho < 10000$. В то же время в наихудшем для них варианте $\rho = 10000$ и $\beta_1 = p_1$ отношение их ССКО к ССКО оценки МНК было приблизительно равно 40.

Ридж-оценка давала не такое низкое значение ССКО для $\rho < 10000$, как оценки b_p и b_L , зато для экстремальной ситуации $\rho = 10000$ ССКО этой оценки была не намного выше оценки МНК. Это позволяет сделать вывод: *предпочтительнее пользоваться ридж-оценкой; если мультиколлинеарности нет, ее ССКО практически совпадает с ССКО оценки МНК, в случае мультиколлинеарности, исключая экстремальные ситуации, она значительно лучше оценки МНК.* Есть еще одно обстоятельство, которое заставляет смотреть на предпочтение оценок b_p и b_L в случае $\rho < 10000$ с определенной осторожностью. Известно, что в приведенном эксперименте существовала только одна приближенная линейная зависимость ($z_1 \approx z_2 + z_3 + z_4$), поэтому значение $r = m - 1$ было известно. На практике неизвестно, сколько зависимостей существует между столбцами матрицы Z . Допуская ошибки в определении r , мы тем самым можем резко повысить ССКО для оценки метода главных компонент и оценки Хокинса. К тому же авторы исследования использовали не лучший вариант выбора k (см. параграф 6.5).

Р. Гунст, Дж. Вебстер и Р. Мазон [117] сравнили также эффективность оценивания отдельных параметров. В качестве первого параметра был выбран β_1 — параметр при переменной, которая «завязана» в мультиколлинеарность, и

β_7 — при переменной, которая не «завязана» в мультиколлинеарность. Сравнение проводилось по величине

$$\sum_{j=1}^{100} (d_j - \beta_i)^2 / 100, \quad i = 1, 2, \dots, 7,$$

где d_j — одна из пяти рассматриваемых оценок. Оказалось, что при оценивании β_1 ситуация остается такой же, как и при использовании в качестве критерия ССКО. При оценивании β_7 все пять методов мало отличаются друг от друга. Полученный вывод весьма ценен: задаваясь целью хорошо оценить параметры, «завязанные» в мультиколлинеарность, мы тем самым оцениванием остальные параметры не хуже, чем методом наименьших квадратов.

Можно прийти к выводу, что наиболее предпочтительна ридж-оценка.

Исследование, проведенное в [134], специально посвящено сравнению оценки МНК и ридж-оценок. Авторы рассмотрели три регрессии. Значения n и m в них были соответственно равны 13 и 4, 13 и 10, 50 и 17. В экспериментах варьировались значения $\beta' \beta$ и $\beta' \beta / \sigma^2$. Для первых двух регрессий было рассмотрено 342 варианта, для третьей — 180. Для каждого варианта моделировались случайные отклонения ε_i , распределенные по нормальному закону $N(0, \sigma^2 \mathbf{I})$. Значение k в ридж-оценке вычислялось по формуле $m\sigma^2/b'b$. Выводы, к которым пришли авторы, сводятся к следующему:

1) ридж-оценка была лучше оценки МНК в более чем 50% всех вариантов;

2) число вариантов, в которых ридж-оценка лучше оценки МНК, увеличивалось с ростом m ;

3) процент предпочтения ридж-оценки увеличивался также при увеличении разброса спектра матрицы $Z'Z$, т. е. при усилении мультиколлинеарности;

4) процент предпочтения ридж-оценки возрастал также с ростом величины $\beta' \beta / \sigma^2$.

Выявлением эффективных оценок методом Монте-Карло занимались А. Демпстер, М. Шатзофф и Н. Вермут [94]. В первой группе проведенных ими экспериментов мультиколлинеарность отсутствовала (32 варианта), во второй группе — мультиколлинеарность присутствовала (128 вариантов). Было исследовано 57 методов оценивания: ридж-оценки, оценки метода главных компонент, редуцированные оценки, оценки метода автоматического отсева пере-

менных и многие другие. В качестве критерия выбиралась средняя сумма квадратов ошибок. Не вдаваясь в подробности проведения экспериментов, сразу перейдем к основным выводам, полученным в ходе исследования:

1) обычный МНК в некоторых случаях оказался хуже тривиальной оценки, когда все координаты считались равными нулю. МНК занял одно из последних мест;

2) наиболее эффективной оказалась ридж-оценка;

3) редуцированные оценки, представителем которых была модифицированная оценка Джеймса—Стейна, также зарекомендовали себя относительно хорошо.

Как видим, авторы исследований по сравнению эффективностей методов оценок параметров регрессий в условиях мультиколлинеарности, рассмотренных в этой главе, приходят к общему выводу: ридж-оценка является одной из наиболее эффективных.

При применении ридж-оценок сталкиваются с трудностью выбора параметра регуляризации k . Как правило, наиболее аргументированные методы выбора этого параметра приводят к тому, что он становится стохастическим, поэтому аналитическое исследование статистических свойств оценки затруднительно. Метод Монте-Карло оказывает здесь неоценимую услугу.

Часть третья

НЕЛИНЕЙНАЯ РЕГРЕССИЯ

Глава 7

ЧИСЛЕННОЕ НАХОЖДЕНИЕ ОЦЕНКИ МНК

7.1. Основные определения. Постановка задачи

Основным предположением рассматриваемых до сих пор взаимосвязей было предположение линейности входящих параметров. Теперь мы откажемся от него и будем рассматривать нелинейные регрессии. По аналогии с (1.1) нелинейную регрессию можно записать в следующем виде:

$$y_t = \Psi(x_{t1}, \dots, x_{tk}; \alpha_1, \dots, \alpha_m) + \varepsilon_t, \quad t = 1, \dots, n \quad (7.1)$$

В уравнении (7.1) по-прежнему y_t — зависимая переменная, x_{t1}, \dots, x_{tk} — независимые переменные, отвечающие номеру наблюдения t , $(\alpha_1, \alpha_2, \dots, \alpha_m)' = \alpha$ — вектор неизвестных параметров, подлежащий оцениванию, ε_t — случайное отклонение. Функция ψ может иметь самый общий вид. Для нелинейной регрессии (7.1) так же, как и для ее прототипа (1.1), будем считать выполненными следующие предположения: множество априорных значений α есть все пространство R^m ; $E\varepsilon_t = 0$, $\text{cov}(\varepsilon) = \sigma^2 I_n$, x_{t1}, \dots, x_{tk} — детерминированы.

Вместо (7.1) удобнее пользоваться другой тождественной записью нелинейной регрессии:

$$y_t = f_t(\alpha_1, \alpha_2, \dots, \alpha_m) + \varepsilon_t, \quad (7.2)$$

где $f_t(\alpha_1, \dots, \alpha_m) = \Psi(x_{t1}, \dots, x_{tk}; \alpha_1, \dots, \alpha_m)$.

Число типов нелинейных регрессий, встречающихся в практике расчетов, так велико, что мы не будем даже пытаться описывать их. Заметим лишь, что часто можно встретить регрессии, линейные в логарифмах, для которых

$$\begin{aligned} f_t(\alpha_1, \dots, \alpha_m) &= \exp(\alpha_1 x_{t1} + \dots + \alpha_m x_{tm}) = \\ &= \exp(\alpha' x_t), \quad t = 1, \dots, n, \end{aligned} \quad (7.3)$$

где $x_1, x_2, \dots, x_n \in R^m$. В этот класс входят тренды

$$f_t(\alpha_1, \alpha_2) = e^{\alpha_1 + \alpha_2 t}; \quad f_t(\alpha_1, \alpha_2, \alpha_3) = e^{\alpha_1 + \alpha_2 t + \alpha_3 t^2}.$$

Класс (7.3) содержит также производственные функции Кобба—Дугласа

$$f_t(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = e^{\alpha_1 + \alpha_2 t} K_t^{\alpha_3} L_t^{\alpha_4}; \quad K_t, L_t > 0. \quad (7.4)$$

Продолжим пример с регрессией результатов химического эксперимента (параграф 1.1). Вместо линейной зависимости между переменными (1.5) теперь будет нелинейная. Допустим, зависимость между y и x_1, x_2, x_3 выражается линейной функцией (относительно параметров) в логарифмах:

$$y_t = \beta_4 x_{t1}^{\beta_1} x_{t2}^{\beta_2} x_{t3}^{\beta_3} + \varepsilon_t, \quad t = 1, \dots, 15$$

или в виде (7.3):

$$y_t = \exp(\alpha_1 z_{t1} + \alpha_2 z_{t2} + \alpha_3 z_{t3} + \alpha_4) + \varepsilon_t, \quad (7.5)$$

где $z_{ti} = \ln x_{ti}$, $\alpha_i = \beta_i$ ($i = 1, 2, 3$), $\alpha_4 = \ln \beta_4$. Значения β_i могут быть интерпретированы следующим образом: при изменении x_i на 1% (при остальных фиксированных независимых переменных) значение y увеличивается на $\beta_i\%$. Величины β_i называют эластичностями ($i = 1, 2, 3$).

Функции f_1, f_2, \dots, f_n будем считать непрерывными на R^m . Часто, кроме непрерывности, будем требовать дифференцируемость функций.

Основным методом оценивания в линейной регрессии является метод наименьших квадратов. Принцип минимизации суммы квадратов отклонений легко обобщается и на нелинейную регрессию. Под оценкой МНК нелинейной регрессии (7.2) будем понимать то значение вектора $a \in R^m$, для которого сумма квадратов отклонений

$$Q(a) = \sum_{t=1}^n (y_t - f_t(a))^2 \quad (7.6)$$

принимает минимальное значение, т. е. оценка МНК есть

$$a = \arg \inf_{a \in R^m} Q(a). \quad (7.7)$$

Нетрудно показать, что если отклонения нормальны, т. е. $\varepsilon \sim N(0, \sigma^2 I)$, то оценка МНК совпадает с оценкой метода максимального правдоподобия. Действительно, в случае нормальных отклонений $y_t \sim N(f_t(a), \sigma^2)$, y_1, y_2, \dots, y_n независимы между собой. Поэтому плотность выборки $y = (y_1, y_2, \dots, y_n) \in R^n$ равна:

$$p(y; a, \sigma^2) = (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp \left[-\frac{1}{2\sigma^2} \sum_{t=1}^n (y_t - f_t(a))^2 \right].$$

Максимизация плотности p , очевидно, соответствует минимизации суммы квадратов отклонений (7.6), т. е. оценки МНК и ММП совпадают.

В линейной регрессии минимизируется аналогичная сумма квадратов отклонений. Однако в линейном случае минимум Q находится сравнительно просто — как решение системы линейных нормальных уравнений. В нелинейном случае система нормальных уравнений $\partial Q/\partial \alpha = 0$ ничего не дает, так как теперь она нелинейна по α , и поэтому не-

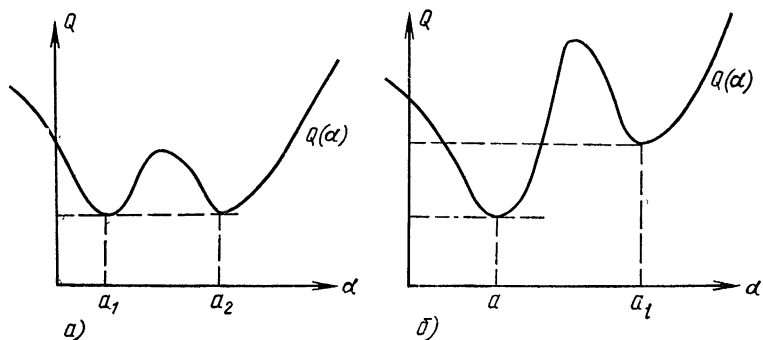


Рис. 7.1. Возможные графики суммы квадратов отклонений: а) оценка МНК неединственна, $Q(a_1) = Q(a_2) = \min$; б) a — истинная оценка, a_1 — ложная оценка, отвечающая локальному минимуму $Q(\alpha)$

посредственно ее решить нельзя. На практике чаще бывает удобнее найти минимум $Q(\alpha)$, чем решить соответствующую систему нормальных уравнений.

Более того, оценка МНК для нелинейной регрессии вообще может не существовать; таким образом, запись (7.7) в некоторых случаях является некорректной. Для примера рассмотрим нелинейную регрессию $y_t = e^{\alpha t} + \varepsilon_t$, $t = 1, \dots, n$, т. е. $f_t(\alpha) = e^{\alpha t}$, $\alpha \in R^1$. Пусть наблюдения таковы, что $y_t \equiv 0$, $t = 1, \dots, n$. Тогда соответствующая сумма квадратов отклонений равна:

$$Q(\alpha) = \sum_{t=1}^n e^{2\alpha t}. \quad (7.8)$$

Инфимум функции (7.8) равен нулю. Он не достигается ни в одной точке, так как $Q(\alpha) \rightarrow 0$ при $\alpha \rightarrow -\infty$, оценки МНК не существует. В [109] приведен практический пример, где оценка МНК не существует.

В нелинейной регрессии может существовать несколько оценок МНК, приводящих к одному и тому же значению суммы квадратов отклонений (рис. 7.1). По определению оценка МНК отвечает глобальному минимуму суммы квадратов отклонений. На практике случай, показанный на рис. 7.1, а, встречается редко. Как правило, имеем ситуацию, изображенную на рис. 7.1, б, где a_1 — ложное значение оценки МНК, отвечающее локальному минимуму

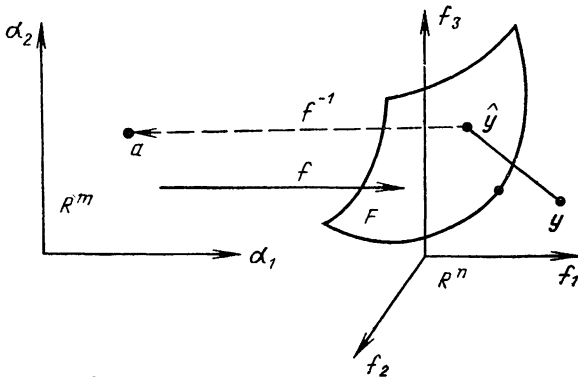


Рис. 7.2. Геометрия МНК нелинейной регрессии, $m=2$, $n=3$

$Q(\alpha)$, a — истинная оценка МНК: Распознать, какое из значений оценок является ложным, а какое — истинным, довольно затруднительно. Эта проблема подробно обсуждается в параграфе 7.6.

Остановимся на геометрическом смысле оценки МНК в нелинейной регрессии. Совокупность функций-регрессий $f_1(\alpha)$, $f_2(\alpha)$, ..., $f_n(\alpha)$ задает отображение из пространства R^m в пространство R^n . Это отображение будем обозначать $f(\alpha) = (f_1(\alpha), f_2(\alpha), \dots, f_n(\alpha))$: $R^m \rightarrow R^n$. При отображении R^m переходит в некоторое множество в пространстве R^n , т. е. в образ R^m . Обозначим это множество $F = f(R^m)$; оно представляет собой поверхность размерности m (рис. 7.2). Помимо F в R^n задана точка y , отвечающая выборке регрессии. Задача заключается в том, чтобы на поверхности F найти точку, наименее удаленную от y . При этом значение суммы квадратов отклонений (7.6) будет минимальным. Обозначим эту точку $\hat{y} \in F$, т. е.

$$\|y - \hat{y}\|^2 = \sum_{t=1}^n (y_t - \hat{y}_t)^2 = \min.$$

Для того чтобы найти оценку МНК, необходимо совершить обратную операцию: по заданному значению образа отображения \mathbf{f} , т. е. $\hat{\mathbf{y}}$, восстановить значение аргумента: $\mathbf{f}(\mathbf{a}) = \hat{\mathbf{y}}$; \mathbf{a} является оценкой МНК. Напомним, что в линейной регрессии поверхность F представляет собой линейное пространство размерности m , натянутое на векторы $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$.

Обратное отображение к \mathbf{f} , заданное на F , должно существовать, т. е. каждой точке $\mathbf{y} \in F$ должна соответствовать единственная точка $\mathbf{a} \in R^m$, для которой $\mathbf{f}(\mathbf{a}) = \hat{\mathbf{y}}$ (см. рис. 7.2). В противном случае оценка МНК не будет единственной. Нелинейную регрессию, в которой отображение \mathbf{f} взаимно-однозначно, будем называть идентифицируемой. Другими словами, регрессия (7.2) идентифицируема, если для любых $\alpha_1, \alpha_2 \in R^m$ из равенства $\mathbf{f}(\alpha_1) = \mathbf{f}(\alpha_2)$ следует $\alpha_1 = \alpha_2$. В дальнейшем все рассматриваемые регрессии будем считать идентифицируемыми¹.

Найдем условия, при которых регрессия (7.3) с функциями, линейными в логарифмах, идентифицируема. Пусть $\alpha_1, \alpha_2 \in R^m$ и $f_t(\alpha_1) = f_t(\alpha_2)$ для всех $t = 1, \dots, n$, т. е. $e^{\alpha_1' \mathbf{x}_t} = e^{\alpha_2' \mathbf{x}_t}$, или

$$\alpha_1' \mathbf{x}_t - \alpha_2' \mathbf{x}_t = 0; \quad (7.9)$$

$$(\alpha_1 - \alpha_2)' \mathbf{x}_t = 0, \quad t = 1, \dots, n.$$

Ясно, что (7.9) влечет равенство $\alpha_1 = \alpha_2$, только если система векторов $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ имеет ранг, равный m . Таким образом, регрессия, линейная в логарифмах, идентифицируема, если $\text{rang } \mathbf{X} = m$, где матрица \mathbf{X} имеет своими строками \mathbf{x}_t .

У п р а ж н е н и я 7. 1

1. Докажите, что сумма квадратов отклонений (7.8) является возрастающей выпуклой вниз функцией.

2. Установите условия идентифицируемости производственной функции Кобба—Дугласа (7.4) при $\alpha_3 + \alpha_4 = 1$.

3. Найдите условия идентифицируемости регрессии $y_t = \alpha_1 + \alpha_2 e^{\alpha_3 t} + \varepsilon_t$.

4. Найдите условия идентифицируемости регрессии $y_t = \alpha_1 / (1 + \alpha_2 e^{\alpha_3 t}) + \varepsilon_t$.

¹Более подробно о понятии идентифицируемости и о критериях идентифицируемости при наличии ограничений см. [24].

5. Найдите условия идентифицируемости регрессии $y_t = \sin \alpha t + \varepsilon_t$.

6. Докажите, что оценка МНК в регрессии $y_t = \alpha + \alpha^2 x_t + \varepsilon_t$, $-\infty < \alpha < +\infty$, существует ($x_t \neq 0$ хотя бы для одного $t = 1, \dots, n$).

7.2. Существование оценки МНК

Рассмотрим критерий, с помощью которого можно устанавливать существование оценок МНК. Прежде всего дадим определение ограниченности снизу функции на бесконечности. Пусть $\varphi = \varphi(x_1, x_2, \dots, x_m) = \varphi(x)$ — функция m аргументов, заданная на всем пространстве R^m . Она ограничена снизу на бесконечности числом B , если для любой последовательности векторов x_1, x_2, \dots , такой, что $\|x_h\| \rightarrow \infty$, найдется такое p , что $\varphi(x'_p) > B$. Данное определение может быть аргументировано следующим образом. Обозначим

$$g(r) = \inf_{\|x\| \geq r} \varphi(x). \quad (7.10)$$

Функция $g(r)$ — неубывающая функция действительного переменного $r \geq 0$, что позволяет нам говорить о ее пределе при $r \rightarrow \infty$ (даже если этот предел равен бесконечности). Тогда функция φ будет ограничена на бесконечности снизу любым числом $B < \lim_{r \rightarrow \infty} g(r)$.

Теорема 7.1. Пусть функция $\varphi(x)$ непрерывна и ограничена снизу числом B . Тогда, если найдется такая точка $x^0 \in R^m$, в которой $\varphi(x^0) \leq B$, то абсолютный минимум функции $\varphi(x)$ достигается.

Доказательство см. в параграфе 7.8.

Применим теорему 7.1 к вопросу о существовании оценки МНК в нелинейной регрессии (7.2). В качестве φ теперь будет выступать сумма квадратов отклонений (7.6). Пусть функция $Q(\alpha)$ ограничена снизу на бесконечности числом B . Тогда если найденное нами начальное приближение a^0 приводит к сумме квадратов отклонений $Q(a^0) = \sum (y_t - \hat{f}_t(a^0))^2 < B$, то оценка МНК существует. Так, на рис. 7.3 сумма квадратов отклонений имеет асимптоту $C = \lim g(r)$ при $\alpha \rightarrow +\infty$ и $Q(\alpha) \rightarrow +\infty$ при $\alpha \rightarrow -\infty$. Поскольку $B < C$, оценка МНК в этом случае существует. Как правило, начальное приближение a^0 найти нетрудно. Основная задача заключается в том, чтобы найти значение B или, еще лучше, $\lim g(r)$ для разных функций регрессий. Найдем оценку снизу для этого предела для регрессий, функции которых линейны в логарифмах. Итак, пусть за-

дана регрессия $y_t = f_t(\alpha) + \varepsilon_t$, $t = 1, \dots, n$, где $f_t(\alpha) = e^{\alpha' x_t}$, $t = 1, \dots, n$, $\alpha = (\alpha_1, \dots, \alpha_m)'$, $x_t = (x_{t1}, \dots, x_{tm})'$. Могут возникнуть два взаимоисключающих случая.

Случай А: векторы x_1, x_2, \dots, x_n разнонаправлены. Это означает, что для любого $\alpha \neq 0$, $\alpha \in R^m$ найдется вектор x_j из системы векторов x_1, \dots, x_n , что $\alpha' x_j > 0$. Докажем, что в случае А число В может быть выбрано любым. Действительно, пусть $\alpha_1, \alpha_2, \dots$ — такая последовательность точек из

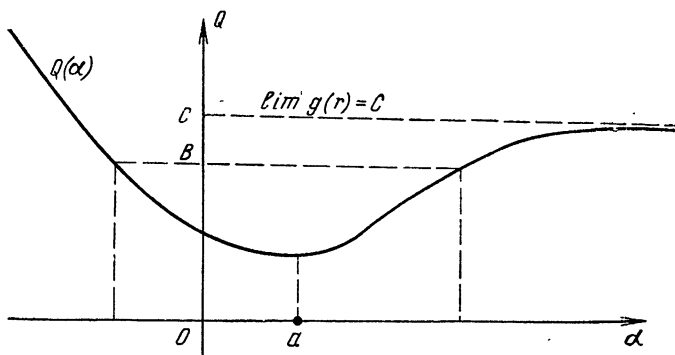


Рис. 7.3. График суммы квадратов отклонений, $m=1$

R^m , что $\|\alpha_k\| \rightarrow \infty$, $k \rightarrow \infty$. Рассмотрим нормированную систему векторов $\beta_k = \alpha_k / \|\alpha_k\|$. Так как β_k принадлежит компактному множеству, то существует сходящаяся подпоследовательность последовательности $\{\beta_k\}$. Обозначим эту подпоследовательность через $\{\beta_s\}$, $\beta_s \rightarrow \beta^*$, $s \rightarrow \infty$. В силу разнонаправленности векторов x_1, \dots, x_n найдется j , для которого $x_j' \beta^* > 0$, т. е. $\cos(x_j, \beta^*) = \gamma > 0$. Поскольку \cos есть функция непрерывная, то начиная с некоторого номера можно утверждать, что $\cos(x_j, \beta_s) \geq \gamma' > 0$, $\gamma' < \gamma$. Поэтому

$$\alpha_s' x_j = \|\alpha_s\| \|x_j\| \cos(\alpha_s, x_j) \geq \|\alpha_s\| \|x_j\| \gamma' \rightarrow \infty$$

при $s \rightarrow \infty$. В силу этого для любого наперед заданного числа найдется номер p , для которого сумма квадратов отклонений

$$\sum_t (y_t - e^{\alpha_p' x_t})^2 \geq (y_j - e^{\alpha_p' x_j})^2$$

будет больше выбранного числа. Итак, в случае разнонаправленных векторов $\lim g(r) = \infty$, $r \rightarrow \infty$.

Случай Б: векторы x_1, \dots, x_n равнонаправлены. Это означает, что найдется такой вектор $\alpha \neq 0, \alpha \in R^m$, что для всех $t = 1, \dots, n$ $\alpha'x_t \leq 0$. Здесь нам потребуется характеристика «степени линейной зависимости» векторов x_1, x_2, \dots, x_n . Обозначим p — такое наименьшее число, что p любых векторов из системы векторов x_1, x_2, \dots, x_n образуют базис пространства R^m . Минимальное значение p , очевидно, равно m .

Итак, пусть задана такая последовательность векторов $\alpha_1, \alpha_2, \dots$, что $\|\alpha_k\| \rightarrow \infty$. Определим $\delta_t = \overline{\lim}_k \cos(\alpha_k, x_t)$, $t = 1, \dots, n$.

Рассмотрим три возможности:

1) $\delta_t > 0$ для некоторого t . Тогда для некоторой подпоследовательности $\{\alpha_s\}$ $\cos(\alpha_s, x_t) \rightarrow \delta_t$. Поэтому

$$\alpha'_s x_t = \|\alpha_s\| \|x_t\| \cos(\alpha_s, x_t) \rightarrow +\infty, s \rightarrow \infty.$$

Таким образом, число B может быть выбрано сколь угодно большим;

2) $\delta_t < 0$ для некоторых t из $1, \dots, n$. Тогда

$$\alpha'_k x_t = \|\alpha_k\| \|x_t\| \cos(\alpha_k, x_t) \leq \|\alpha_k\| \|x_t\| \delta_t \rightarrow -\infty, k \rightarrow \infty.$$

Это приводит к тому, что $\exp(\alpha'_k x_t) \rightarrow 0$, и соответствующие слагаемые в сумме квадратов отклонений стремятся к y_t^2 ;

3) $\delta_t = 0$. Число номеров, для которых $\delta_t = 0$, обозначим n_1 . Докажем, что $n_1 \leq p - 1$. В силу $\delta_t = 0$ для каждого t найдется такая подпоследовательность $\{\alpha_s\}$, что $\cos(\alpha_s, x_t) \rightarrow 0$. Рассуждая, как в случае *A*, можно выбрать единую подпоследовательность α_s , для которой для всех t , таких, что $\delta_t = 0$, $\cos(\alpha_s, x_t) \rightarrow 0$. В силу замкнутости и ограниченности сферы $\|\alpha\| = 1$ найдется предельная точка α_* этой подпоследовательности. В этой точке $\cos(\alpha_*, x_t) = 0$, т. е. $\alpha_* \perp x_t$ ($\|\alpha_*\| = 1$). Теперь ясно, почему n_1 не может быть больше $p - 1$, так как в противном случае система векторов, дополненная α_* , была бы линейно-независимой, что противоречит определению p .

Таким образом, сумму квадратов отклонений $Q(\alpha)$ в случае *B* можно разложить на две суммы:

$$\begin{aligned} & \sum_{t:\delta_t < 0} (y_t - e^{\alpha'_k x_t})^2 + \sum_{t:\delta_t = 0} (y_t - e^{\alpha'_k x_t})^2 \geq \\ & \geq \sum_{t:\delta_t < 0} (y_t - e^{\alpha'_k x_t})^2 = \sum_{t:\delta_t < 0} y_t^2 \geq \sum_{t=1}^n y_t^2 - \sum_{t=n-p+2}^n y_t^2, \end{aligned}$$

где $y_{(1)}^2 \leq y_{(2)}^2 \leq \dots \leq y_{(n)}^2$ — ранжированный ряд. Окончательно

$$\lim_{r \rightarrow \infty} g(r) \geq \sum_{t=1}^n y_t^2 - \sum_{t=n-p+2}^n y_t^2 \quad (7.11)$$

и B может быть выбрано любым, меньшим числа в правой части (7.11).

Рассмотрим регрессию (7.5). Для всех $t = 1, \dots, 15$, $i = 1, 2, 3, 4$, $z_{ti} > 0$, поэтому для всех $i, j = 1, 2, 3, 4$ $z_i' z_j > 0$. Как следует из задачи 2 упражнения 7.2, векторы $\{z_i\}$ при этих условиях будут однонаправлены. Далее, легко показать, что для регрессии (7.5) $p = 4$ и $y_{(13)} = 206,77$; $y_{(14)} = 216,48$; $y_{(15)} = 221,45$ (см. табл. 1.1). Поэтому для регрессии (7.5)

$$g(\infty) \geq \sum_{t=1}^{11} y_t^2 + y_{13}^2 = 330211,7.$$

Для регрессий (7.3), линейных в логарифмах, начальное приближение находится просто. Прологарифмируем обе части регрессии; тогда формально можно записать

$$\ln y_t = \alpha_1 x_{t1} + \dots + \alpha_m x_{tm} + \xi_t. \quad (7.12)$$

Применяя к уравнению (7.12) метод наименьших квадратов, найдем a^0 .

Как правило, a^0 является хорошим приближением к искомой оценке МНК a . Оцененная таким способом регрессия (7.5) выглядит следующим образом:

$$\begin{aligned} \ln y_t = & 0,7465z_{t1} + 0,1529z_{t2} + 0,2025z_{t3} - 0,3381; & (7.13) \\ & (0,0784) \quad (0,262) \quad (0,0925) \quad (0,896) \\ Q(a^0) = & \sum_t (y_t - f_t(a^0))^2 = 66,36. \end{aligned}$$

Поскольку $330211,7 > 66,36$, то оценка МНК по теореме 7.1 регрессии (7.5) существует.

У п р а ж н е н и я 7.2

1. Будет ли функция $g(r)$ непрерывной?
2. Докажите, что если $x_i' x_j > 0$ для всех $i, j = 1; \dots; n$, то векторы $x_1; \dots; x_n$ будут однонаправлены.

3. Докажите, что если $x_{t1} > 0$ для всех $t = 1, 2, \dots, n$; то векторы x_1, x_2, \dots, x_n будут равнонаправлены.

4. Дайте геометрическую интерпретацию однонаправленности векторов.

5. Покажите, что $g(\infty) = \sum_{t=1}^n y_t^2$ для регрессии $f_t(\alpha) = e^{\alpha t}$.

7.3. Метод Ньютона—Гаусса и его модификации

Специфический вид суммы квадратов отклонений позволяет построить методы минимизации, более эффективные, чем общие методы¹. Основой является метод Ньютона—Гаусса.

Найдем первые и вторые производные суммы квадратов отклонений:

$$g_i = \frac{\partial Q}{\partial \alpha_i} = -2 \sum_t (y_t - f_t(\alpha)) \frac{\partial f_t}{\partial \alpha_i}, \quad i = 1, \dots, m; \quad (7.14)$$

$$H_{ij} = \frac{\partial^2 Q}{\partial \alpha_i \partial \alpha_j} = 2 \sum_t \frac{\partial f_t}{\partial \alpha_i} \cdot \frac{\partial f_t}{\partial \alpha_j} - 2 \sum_t (y_t - f_t(\alpha)) \frac{\partial^2 f_t}{\partial \alpha_i \partial \alpha_j},$$

$$i = 1, \dots, m, \quad j = 1, \dots, m. \quad (7.15)$$

Значения $q_i = -g_i$ образуют вектор антиградиента $\mathbf{q} = (q_1, \dots, q_m)'$, значения H_{ij} образуют гессиан, симметричную матрицу \mathbf{H} порядка $m \times m$ (естественно предполагаем, что функции f_t дважды непрерывно дифференцируемы). Обозначим через \mathbf{P} матрицу $n \times m$, (t, i) -й элемент которой равен $\partial f_t / \partial \alpha_i$. Эта матрица является матрицей производных отображения $f: R^m \rightarrow R^n$. В силу (7.15) гессиан суммы квадратов отклонений \mathbf{H} может быть разложен в сумму двух матриц: $\mathbf{H} = \mathbf{H}_1 - \mathbf{H}_2$, где $\mathbf{H}_1 = 2\mathbf{P}'\mathbf{P}$. Предположим, матрица \mathbf{H}_1 положительно определена. Для этого необходимо и достаточно предположить, что матрица \mathbf{P} имеет полный ранг m в любой точке $\alpha \in R^m$. Далее предположим:

а) нелинейная регрессия (7.2) имеет невысокий порядок нелинейности: вторые производные $\partial^2 f_t / \partial \alpha_i \partial \alpha_j$ принимают не очень большие значения;

б) гессиан \mathbf{H} рассматривается в достаточно малой окрестности минимизирующего вектора \mathbf{a} , для которого приближение $f_t(\mathbf{a})$ к y достаточно хорошо, т. е. величины $y_t - f_t(\mathbf{a})$ близки к нулю.

¹К общим методам минимизации функций многих переменных относятся такие методы, как градиентный, сопряженных градиентов, Ньютона и т. д. [15, 55, 57].

С учетом условий а) и б) приближенно можно считать $\mathbf{H}_2 \approx \mathbf{0}$, поэтому $\mathbf{H} \approx \mathbf{H}_1$.

Общий метод минимизации, метод Ньютона, предлагает двигаться из данного начального приближения \mathbf{a}^0 к следующему по правилу

$$\mathbf{a}^1 = \mathbf{a}^0 + \mathbf{H}^{-1} \mathbf{q}, \quad (7.16)$$

где \mathbf{q} — антиградиент, равный — \mathbf{g} (7.14), который в матричной форме записывается в виде $\mathbf{q} = 2\mathbf{P}'(\mathbf{y} - \mathbf{f}(\mathbf{a}^0))$. Окончательно (7.16) переписывается следующим образом:

$$\mathbf{a}^1 = \mathbf{a}^0 + 2\mathbf{H}_1^{-1} \mathbf{P}'(\mathbf{y} - \mathbf{f}(\mathbf{a}^0)) = \mathbf{a}^0 + (\mathbf{P}'\mathbf{P})^{-1} \mathbf{P}'(\mathbf{y} - \mathbf{f}(\mathbf{a}^0)).$$

Вычислив следующее значение приближения \mathbf{a}^1 , на его основе можно построить \mathbf{a}^2 и т. д. В общем виде метод Ньютона—Гаусса записывается следующим образом:

$$\mathbf{a}^{k+1} = \mathbf{a}^k + (\mathbf{P}'_k \mathbf{P}_k)^{-1} \mathbf{P}'_k (\mathbf{y} - \mathbf{f}(\mathbf{a}^k)), \quad k = 0, 1, \dots, \quad (7.17)$$

где
$$\mathbf{P}_k = \left. \frac{\partial \mathbf{f}}{\partial \boldsymbol{\alpha}} \right|_{\boldsymbol{\alpha} = \mathbf{a}^k}.$$

Формула (7.17) может быть получена также из других соображений. Пусть приближение \mathbf{a}^k известно; разложим функцию регрессии в окрестности \mathbf{a}^k в ряд Тейлора до линейных членов:

$$f_t(\boldsymbol{\alpha}) \approx f_t(\mathbf{a}^k) + \sum_{i=1}^m \frac{\partial f_t(\mathbf{a}^k)}{\partial \alpha_i} (\alpha_i - a_i^k), \quad t = 1, \dots, n.$$

В матричном виде это равенство может быть переписано как

$$\mathbf{f}(\boldsymbol{\alpha}) = \mathbf{f}(\mathbf{a}^k) + \mathbf{P}_k (\boldsymbol{\alpha} - \mathbf{a}^k),$$

поэтому регрессия (7.2) линейризуется

$$\mathbf{y} = \mathbf{f}(\mathbf{a}^k) + \mathbf{P}_k (\boldsymbol{\alpha} - \mathbf{a}^k) + \boldsymbol{\varepsilon} \quad (7.18)$$

или

$$\mathbf{y} - \mathbf{f}(\mathbf{a}^k) = \mathbf{P}_k (\boldsymbol{\alpha} - \mathbf{a}^k) + \boldsymbol{\varepsilon}.$$

Применяя МНК к линейризованной регрессии (7.18), найдем следующее значение вектора приближения:

$$\mathbf{a}^{k+1} - \mathbf{a}^k = (\mathbf{P}'_k \mathbf{P}_k)^{-1} \mathbf{P}'_k (\mathbf{y} - \mathbf{f}(\mathbf{a}^k)),$$

которое совпадает с ранее полученным (7.17).

Метод Ньютона—Гаусса является в некотором смысле интерполяцией градиентного метода и метода Ньютона. Действительно, как и в градиентном методе, здесь вычисляются только первые производные; таким образом, время,

затрачиваемое на одну итерацию, не намного больше, чем в градиентном методе (как правило, время, необходимое на обращение и перемножение матриц, намного меньше, чем на вычисление значения функций или их производных). Этот метод близок к методу Ньютона, так как по смыслу является его приближением. В нем отсутствует существенный недостаток метода Ньютона — если в последнем матрица вторых производных должна быть положительно определена, то в методе Ньютона—Гаусса матрица \mathbf{H}_1 по построению неотрицательно определена, а при условии, что \mathbf{P} имеет полный ранг, определена положительно.

Длина шага в методе Ньютона—Гаусса (7.17) равна единице. Метод будет более гибким, если длину шага сделать переменной. Таким образом может быть построен модифицированный метод Ньютона—Гаусса:

$$\mathbf{a}^{k+1} = \mathbf{a}^k + \lambda_k (\mathbf{P}'_k \mathbf{P}_k)^{-1} \mathbf{P}'_k (\mathbf{y} - \mathbf{f}(\mathbf{a}^k)), \quad \lambda_k > 0. \quad (7.19)$$

Значение коэффициента λ_k , определяющего длину шага, в оптимальном случае находится из условия минимизации $Q(\alpha)$ в направлении

$$\delta_k^{NG} = (\mathbf{P}'_k \mathbf{P}_k)^{-1} \mathbf{P}'_k (\mathbf{y} - \mathbf{f}(\mathbf{a}^k)). \quad (7.20)$$

Приведем теорему, с помощью которой можно легко получать условия сходимостей тех или иных методов.

Прежде всего заметим, что любой итерационный метод минимизации функций задается неким отображением R^m в себя. Действительно, при минимизации функции $\varphi(\mathbf{x})$ каждому приближению \mathbf{x}^k мы должны поставить в соответствии направление движения \mathbf{p}_k к следующему приближению \mathbf{x}^{k+1} . Таким образом, итерационный процесс задается функцией $\mathbf{p}(\mathbf{x})$, $\mathbf{x} \in R^m$. Так, в градиентном методе $\mathbf{p}(\mathbf{x}) = \mathbf{q}(\mathbf{x})$, а в методе Ньютона $\mathbf{p}(\mathbf{x}) = \mathbf{H}^{-1}(\mathbf{x}) \mathbf{q}(\mathbf{x})$ и т. д. По определению

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \lambda_k \mathbf{p}(\mathbf{x}^k) = \mathbf{x}^k + \lambda_k \mathbf{p}_k, \quad (7.21)$$

где λ_k выбрано тем или иным образом. В оптимальном случае

$$\varphi(\mathbf{x}^k + \lambda_k \mathbf{p}_k) = \min_{\lambda > 0} \varphi(\mathbf{x}^k + \lambda \mathbf{p}_k). \quad (7.22)$$

Теорема 7.2. Пусть функция $\varphi(\mathbf{x})$ непрерывна на R^m вместе со своими первыми и вторыми производными. Пусть $\varphi(\mathbf{x})$ ограничена снизу на бесконечности числом B (см. параграф 7.2) и начальное приближение итерационного процесса (7.21) \mathbf{x}^0 выбрано так, что $\varphi(\mathbf{x}^0) \leq B$. Пред-

положим, итерационный процесс, задаваемый отображением $p(x)$, удовлетворяет соотношению

$$\cos(p(x), q(x)) \geq \varepsilon > 0, x \in R^m. \quad (7.23)$$

Тогда, если λ_k выбирать в интервале

$$v \frac{2q'_k p_k}{M \|p_k\|^2} \leq \lambda_k \leq (1-v) \frac{2q'_k p_k}{M \|p_k\|^2}, \quad (7.24)$$

где v — любое число: $0 < v \leq 1/2$ и $M \geq M_0$ (M_0 задается равенством (7.25)), то для всех предельных точек x^* последовательности x^0, x^1, x^2, \dots имеем $d\varphi(x^*)/dx = 0$, $\varphi(x^*) = \text{const}$.

Доказательство теоремы приводится в параграфе 7.8¹. Сделаем некоторые замечания. Очевидно, в силу условий теоремы множество $S_0 = \{x \in R^m : \varphi(x^0) \leq B\}$ является замкнутым и ограниченным (см. теорему 7.1), и минимум функции φ достигается. В силу непрерывности гессиана функции и непрерывности максимального характеристического числа матрицы от своих элементов существует

$$M_0 = \sup_{x \in S_0} \lambda_{\max} H(x). \quad (7.25)$$

Поскольку мы не делали предположений о выпуклости функции, постольку нельзя доказать, что пределом последовательности итераций служит точка, обращающая φ в минимум. Единственное, что можно утверждать в этом случае, что градиент в предельных точках равен нулю.

Далее, из доказательства теоремы следует, что если λ_k выбирать из условия (7.22), то (7.23) достаточно для выполнения теоремы.

С помощью теоремы 7.2 легко проверяется сходимость различных методов. Для примера рассмотрим градиентный метод. Здесь $\cos(p, q) = \cos(q, q) = 1$, поэтому условие (7.23) автоматически выполняется. Если λ_k отыскивается по правилу (7.22), то сходимость градиентного метода к стационарным точкам следует из теоремы 7.2. Допустим теперь, что в градиентном методе $\lambda_k = 1$, $k = 0, 1, 2, \dots$ Найдем условия сходимости такого метода. Выражение (7.24) переписывается следующим образом:

$$v2 \frac{\|q_k\|^2}{M \|q_k\|^2} \leq 1 \leq (1-v) \frac{2 \|q_k\|^2}{M \|q_k\|^2},$$

¹Теорема 7.2 с оптимальным выбором λ_k (7.22) приводится в [55, с. 47].

или $v \leq M/2$, $v \leq 1 - M/2$. Ясно, что если $M_0 < 2$, то значение $0 < v < 1/2$ существует, неравенство (7.24) выполняется, и градиентный метод сходится.

Выясним условия сходимости итерационных методов минимизации суммы квадратов отклонений. Начнем с метода Ньютона—Гаусса (7.17). В [17] предлагаются условия сходимости метода Ньютона—Гаусса, однако они трудно проверяемы и завышены.

Как и ранее, будем считать, что матрица \mathbf{P} имеет полный ранг, таким образом, матрица $\mathbf{P}'\mathbf{P}$ положительно определена для всех $x \in S_0$. Обозначим

$$[\delta = 2 \inf_{x \in S_0} \lambda_{\min}(\mathbf{P}'(x)\mathbf{P}(x)) > 0.$$

В методе Ньютона—Гаусса $\mathbf{p}_k = \mathbf{H}_{1k}^{-1} \mathbf{q}_k$, $\lambda_k = 1$. Докажем сначала, что для этого метода выполняется условие (7.23). Обозначим

$$m_* = \sup_{x \in S_0} \lambda_{\max} \mathbf{H}_{1k}(x) = 2 \sup_{x \in S_0} \lambda_{\max} \mathbf{P}'(x)\mathbf{P}(x).$$

Тогда, как легко показать,

$$\begin{aligned} \frac{1}{m_*} \mathbf{q}'\mathbf{q} &\leq \mathbf{q}'\mathbf{H}_1^{-1} \mathbf{q} \leq \frac{1}{\delta} \mathbf{q}'\mathbf{q}; \\ \frac{1}{m_*^2} \mathbf{q}'\mathbf{q} &\leq \mathbf{q}'\mathbf{H}_1^{-2} \mathbf{q} \leq \frac{1}{\delta^2} \mathbf{q}'\mathbf{q}. \end{aligned}$$

Но

$$\begin{aligned} \cos(\mathbf{p}_k, \mathbf{q}_k) &= \cos(\mathbf{H}_{1k}^{-1} \mathbf{q}_k, \mathbf{q}_k) = \frac{\mathbf{q}'_k \mathbf{H}_{1k}^{-1} \mathbf{q}_k}{(\mathbf{q}'_k \mathbf{q}_k)^{1/2} (\mathbf{q}'_k \mathbf{H}_{1k}^{-2} \mathbf{q}_k)^{1/2}} \geq \\ &\geq \frac{\frac{1}{m_*} \mathbf{q}'_k \mathbf{q}_k}{(\mathbf{q}'_k \mathbf{q}_k)^{1/2} \frac{1}{\delta} (\mathbf{q}'_k \mathbf{q}_k)^{1/2}} = \frac{\delta}{m_*} > 0, \end{aligned}$$

т. е. в качестве ε можно взять отношение δ/m_* .

Так же, как в градиентном методе, для существования $0 < v \leq 1/2$ достаточно, чтобы

$$1 - \frac{M_0 \mathbf{q}'_k \mathbf{H}_{1k}^{-2} \mathbf{q}_k}{2 \mathbf{q}'_k \mathbf{H}_{1k}^{-1} \mathbf{q}_k} > 0. \quad (7.26)$$

Можно показать, что

$$\frac{\mathbf{q}'_k \mathbf{H}_{1k}^{-2} \mathbf{q}_k}{\mathbf{q}'_k \mathbf{H}_{1k}^{-1} \mathbf{q}_k} \leq \frac{1}{\delta},$$

поэтому условие (7.26) переписывается как

$$M_0 < 2\delta. \quad (7.27)$$

На основе (7.27) можно сделать вывод: если нелинейные регрессии «не очень нелинейны» (M_0 не велико), а сингулярность матрицы $P'P$ не очень велика (δ достаточно велико), то метод (7.17) сходится к стационарным точкам. Этот вывод полностью подтверждается на практике. В [87] предлагается другой критерий сходимости метода Ньютона — Гаусса:

$$KQ(\mathbf{a}) < \delta, \quad (7.28)$$

где \mathbf{a} — оценка МНК, минимизирующая $Q(\alpha)$; K — число, имеющее ту же природу, что и M_0 . Условие сходимости (7.28) нам кажется несколько искусственным, поскольку значение $Q(\mathbf{a})$ неизвестно, его-то и требуется найти.

Не представляет труда выяснение условий сходимости модифицированного метода Ньютона—Гаусса (7.19). Если $|P'P| \neq 0$ на S_0 , а λ_h минимизирует $Q(\alpha)$ вдоль направления (7.20), то метод (7.19) сходится к стационарным точкам. Сходимость в (7.19) будет выполняться, если λ_h выбрать на интервале

$$\left[v \frac{2q'_k H_{1k}^{-1} q_k}{M q'_k H_{1k}^{-2} q_k}; (1-v) \frac{2q'_k H_{1k}^{-1} q_k}{M q'_k H_{1k}^{-2} q_k} \right],$$

где v — некоторое выбранное заранее положительное число, $v \leq 1/2$; M — оценка сверху максимального х. ч. гессiana $Q(\alpha)$ на S_0 . В частности, если $v = 1/2$, то

$$\lambda_h = \frac{q'_k H_{1k}^{-1} q_k}{M q'_k H_{1k}^{-2} q_k}.$$

Х. Хартли [121] для выбора λ_h предложил следующую процедуру. Вдоль выбранного направления в методе Ньютона—Гаусса аппроксимируем сумму квадратов отклонений Q параболой. Для этого подсчитаем значение Q для $\lambda_h = 1/2$ и $\lambda_h = 1$ (метод Ньютона—Гаусса). Нам известно $Q(0) = Q(\mathbf{a}^k)$. Проведем через полученные три точки параболу $P(\lambda) = a\lambda^2 + b\lambda + c$. Найдем ее коэффициенты. Очевидно, $P(0) = Q(0) = c$, так как парабола проходит через точку $\lambda = 0$. Далее, $P(1) = Q(1) = a + b + Q(0)$

и $P(1/2) = Q(1/2) = a/4 + b/2 + Q(0)$. Отсюда легко найти a и b . Имеем

$$a = 2Q(1) - 4Q(1/2) + 2Q(0);$$

$$b = 4Q(1/2) - Q(1) - 3Q(0); \quad c = Q(0).$$

Как известно, парабола $P(\lambda)$ принимает минимальное значение при $\lambda = -b/2a$. Подставляя полученные значения для a, b, c , получим:

$$\begin{aligned} \lambda_h &= \frac{1}{4} \frac{Q(1) + 4Q(1/2) + 3Q(0)}{Q(1) - 2Q(1/2) + Q(0)} = \frac{1}{2} + \\ &+ \frac{1}{4} \frac{Q(0) - Q(1)}{Q(1) - 2Q(1/2) + Q(0)}. \end{aligned} \quad (7.29)$$

При приближении к оценке МНК, как правило, λ_h стремится к 1. Это и понятно, так как говорит о том, что вблизи минимизируемого вектора $Q(\alpha)$ хорошо аппроксимируется квадратичной функцией, для которой $\lambda_h = 1$ является оптимальным.

Выбор λ_h может быть осуществлен и по другому принципу. Выберем некоторое число $0,5 \leq s < 1$ и положим $\lambda_h = s^q$, где $q = 0, 1, \dots$ — такое максимальное число, что $Q(a^{k-1} + \lambda_h \delta_k^{NG}) < Q(a^{k-1})$. Замечено, что стратегия выбора λ_h не имеет решающего значения.

Рассмотрим теперь отыскание оценки МНК для функций регрессий, линейных в логарифмах $f_t(\alpha) = \exp(\alpha' x_t)$, $t = 1, \dots, n$. Прежде всего покажем, что в случае идентифицируемости на любом ограниченном множестве $S \subset R^m$ матрица производных этой функции имеет полный ранг. Действительно, легко проверить, что матрица производных $P = DX$, где D — диагональная матрица $n \times n$, (t, t) -й элемент которой равен $e^{\alpha' x_t}$, X — матрица порядка $n \times m$, в качестве строк имеет векторы x'_t . Из условия идентифицируемости следует, что $\text{rang } X = m$, поэтому $\lambda_{\min}(X'X) = \mu > 0$. Далее, из неравенства (П.12) следует

$$\begin{aligned} \lambda_{\min}(P'P) &= \lambda_{\min}(X'D^2X) \geq \lambda_{\min}(D^2) \lambda_{\min}(X'X) \geq \\ &\geq \min_{t, \alpha} e^{2\alpha' x_t} \cdot \mu = \delta/2. \end{aligned}$$

Очевидно, если S — ограниченное множество, то $\min_{t, \alpha} e^{2\alpha' x_t} > 0$ и $\delta > 0$.

1. При каких условиях $|\mathbf{H}_1| \neq 0$ в регрессиях а) (7.4), $\alpha_3 + \alpha_4 = 1$; б) $f_t(\alpha_1, \alpha_2, \alpha_3) = \alpha_1 + \alpha_2 e^{\alpha_3 t}$ — модифицированная экспонента; в) $f_t(\alpha_1, \alpha_2, \alpha_3) = \alpha_1 / (1 + \alpha_2 e^{-\alpha_3 t})$ — логистическая кривая.

7.4. Метод Левенберга—Марквардта

Суть этого широко используемого метода проста. При минимизации по методу Ньютона—Гаусса мы требовали невырожденность матрицы \mathbf{H}_1 . Иногда матрица \mathbf{H}_1 становится настолько плохо обусловленной, что практически обратить ее невозможно. К. Левенберг [152] предложил такие матрицы «подправлять» следующим образом. Вместо \mathbf{H}_1 рассмотрим матрицу $\mathbf{H}_1 + 2\mu \mathbf{I}_m$, $\mu > 0$. Тогда матрица $\mathbf{H}_1 + 2\mu \mathbf{I}_m = 2(\mathbf{P}'\mathbf{P} + \mu \mathbf{I}_m)$ всегда невырождена и обратима. Итерационный процесс минимизации суммы квадратов отклонений в методе Левенберга строится по формуле

$$\mathbf{a}^{k+1} = \mathbf{a}^k + (\mathbf{P}'_k \mathbf{P}_k + \mu_k \mathbf{I})^{-1} \mathbf{P}'_k (\mathbf{y} - \mathbf{f}(\mathbf{a}^k)), \quad k=0, 1, 2, \dots, \quad (7.30)$$

где $\mu_k > 0$. Поскольку считаем, что матрица \mathbf{H}_1 может быть вырождена, можно даже предположить $\delta = 0$, зато $\mu_k \geq \mu_0 > 0$ для всех $k=0, 1, 2, \dots$. Формула (7.30) имеет много общего с ридж-оценкой (6.27).

Рассмотрим условия сходимости процесса (7.30). Воспользуемся теоремой 7.2. Условие (7.23), очевидно, выполняется. В методе Левенберга $\lambda_k = 1$, что накладывает определенные ограничения на поправки μ_k . Можно доказать, что если $M_0 < 2\mu_0$, то метод Левенберга сходится (точнее, во всех предельных точках последовательности $\mathbf{a}^0, \mathbf{a}^1, \mathbf{a}^2, \dots$ градиент $Q(\alpha)$ равен нулю).

Обозначим δ_k^L поправку в методе Левенберга к предыдущему вектору параметров \mathbf{a}^k :

$$\delta_k^L = (\mathbf{P}'_k \mathbf{P}_k + \mu \mathbf{I})^{-1} \mathbf{P}'_k (\mathbf{y} - \mathbf{f}(\mathbf{a}^k)) = \frac{1}{2} (\mathbf{P}'_k \mathbf{P}_k + \mu \mathbf{I})^{-1} \mathbf{q}_k.$$

Свойства этой поправки при изменении параметра μ приводятся в теореме 7.3.

Поскольку (7.30) является ридж-оценкой линеаризованной регрессии (7.18), оценка (7.30) дает минимальную сумму квадратов отклонений регрессии (7.18) в классе оценок с фиксированной длиной (см. параграф 6.4).

Теорема 7.3.

а) длина вектора поправки $\|\delta_k^L\|$ является убывающей функцией μ , при увеличении μ от 0 до $+\infty$ $\|\delta_k^L\|$ монотонно убывает от $\|\delta_k^{NG}\|$ до 0;

б) $\cos(\mathbf{q}_k, \delta_k^L)$ является возрастающей функцией μ и при изменении μ от 0 до $+\infty$ изменяется от $\cos(\mathbf{q}_k, \delta_k^{NG})$ до 1;

в) $\cos(\delta_k^L, \delta_k^{NG})$ является убывающей функцией μ и при изменении μ от 0 до $+\infty$ изменяется от 1 до $\cos(\mathbf{q}_k, \delta_k^{NG})$.

Доказательство этой теоремы дано в параграфе 7.8.

Суть теоремы заключается в том, что, варьируя значением μ , можно изменять направление вектора поправки. При малых μ вектор поправки δ_k^L расположен ближе к вектору поправки метода Ньютона—Гаусса, при больших μ имеет направление, близкое к антиградиенту минимизируемой функции.

Д. Марквардт [158] предложил другую поправку для корректировки матрицы $\mathbf{P}'\mathbf{P}$:

$$\delta_k^M = (\mathbf{P}'_k \mathbf{P}_k + \mu_k \mathbf{D}_k)^{-1} \mathbf{P}'_k (y - f(\mathbf{a}^k)), \quad (7.31)$$

где $\mathbf{D}_k = \text{diag}(\mathbf{P}'_k \mathbf{P}_k)$. Марквардт корректирует диагональные элементы матрицы $\mathbf{P}'_k \mathbf{P}_k$ в зависимости от их величины. Метод Марквардта часто используется при минимизации суммы квадратов отклонений. Используя теорему 7.2, можно найти условия сходимости этого метода.

Остановимся теперь на стратегии выбора μ_k . Градиентный метод хорошо работает в начале итерационного процесса, т. е. когда начальное приближение находится на достаточном расстоянии от минимизирующего вектора. Наоборот, метод Ньютона—Гаусса, как правило, быстро сходится в случае, когда приближение лежит в непосредственной близости к оценке МНК. Поэтому алгоритм, который на первых итерациях работал бы как градиентный метод, а в конце итерационного процесса — как метод Ньютона—Гаусса, подобрал бы в себя лучшие характеристики обоих процессов. При надлежащем выборе μ можно добиться того, чтобы методы Левенберга и Марквардта обладали этими свойствами. Действительно, пусть в начале процесса минимизации значения μ будут относительно большими, а в конце процесса — малыми. Марквардт, в частности, предложил следующую процедуру:

1) на нулевом шаге полагаем $\mu_0 = 0,01$;

2) на $k + 1$ шаге полагаем $\mu^* = \mu_k/10$:

а) если $Q(\mu^*) \geq Q(a^k)$, то увеличиваем значение μ^* ($\mu^* = 10\mu^*$) и снова проверяем выполнение неравенства $Q(\mu^*) \geq Q(a^k)$. Таким способом увеличиваем μ^* до тех пор, пока новое значение суммы квадратов отклонений не будет меньше прежнего значения $Q(\alpha)$, и полагаем $\mu_{k+1} = \mu^*$;

б) если $Q(\mu^*) < Q(a^k)$, то полагаем $\mu_{k+1} = \mu^*$.

В [22] приведена программа на Алголе метода Марквардта с описанной процедурой выбора μ .

Предложенная процедура выбора μ имеет один недостаток. Если матрица $P'_k P_k$ плохо обусловлена, то в некоторых случаях для выполнения неравенства $Q(a^{k+1}) < Q(a^k)$ необходимо увеличивать μ до чрезвычайно больших значений. Хотя при увеличении μ направление вектора δ^L приближается к направлению вектора q , длина вектора δ^L стремится к нулю, что может стать препятствием для уменьшения $Q(\alpha)$. В таких ситуациях можно модифицировать описанную процедуру следующим образом. Вместо того чтобы увеличивать μ до слишком высоких значений, будем уменьшать длину шага вдоль направления, соответствующего пороговому значению $\bar{\mu}$. Таким образом, полагаем $a^{k+1} = a^k + \lambda \delta_k^L(\bar{\mu})$, где $\lambda > 0$ выбрано так, чтобы $Q(a^{k+1}) < Q(a^k)$. Поскольку угол между $\delta_k^L(\bar{\mu})$ и q острый, такое λ всегда существует. В качестве порогового μ мы предлагаем брать $\bar{\mu} = 1^1$.

Для примера рассмотрим оценивание регрессии (7.5). В качестве a^0 положим оценку МНК логарифмированной регрессии (7.13). В табл. 7.1 показана минимизация $Q(\alpha)$ методом Марквардта. Отметим медленную скорость сходимости, а также равномерное снижение значения μ . Таким образом, $a = (0,733; 0,199; 0,207; 0,615)^t$.

Остановимся теперь на сравнении методов Ньютона—Гаусса и Хартли и методов Левенберга—Марквардта. Как показывает практика расчетов, методы второй группы являются более «осторожными». Для регрессий, в которых нелинейности не очень велики (к таким, в частности, относятся регрессии, линейные в логарифмах), методы Ньютона—Гаусса и Хартли сходятся быстрее методов второй

¹Марквардт [158] (см. также [49]) предлагает несколько иной метод модификации этой процедуры.

Таблица 7.1

Номер итерации	a_1	a_2	a_3	a_4	Q	μ
0	0,7456	0,1529	0,2025	0,7131	66,36	10^{-2}
1	0,7451	0,1522	0,2092	0,7060	65,58	10^{-3}
2	0,7426	0,1550	0,2134	0,7000	65,42	10^{-4}
3	0,7409	0,1580	0,2140	0,6961	65,39	10^{-4}
22	0,7326	0,1982	0,2070	0,6156	65,20	10^{-6}
23	0,7326	0,1986	0,2070	0,6149	65,20	10^{-7}

группы. Для регрессий же с высокой степенью нелинейности более предпочтительными оказываются методы второй группы. При этом их скорость сходимости невысока, зато методы Ньютона—Гаусса и Хартли вовсе расходятся.

Иногда ни один из описанных методов не приводит к оценке МНК. В частности, сходимость часто отсутствует в регрессиях с функцией

$$f_i(\alpha_1, \alpha_2, \alpha_3) = \alpha_1(\alpha_2 K_i^{\alpha_3} + (1 - \alpha_2) L_i^{\alpha_3})^{1/\alpha_3}, \quad (7.32)$$

которая представляет собой производственную функцию с постоянной эластичностью замены [12].

Остановимся на вопросе окончания процесса счета. Практически счет может быть остановлен, когда выполняются следующие условия:

- а) поправка вектора $\mathbf{a}^{k+1} - \mathbf{a}^k$ мала;
- б) градиент $\partial Q / \partial \alpha$ близок к нулю.

Таким образом, при произвольно выбранных числах δ , κ процесс минимизации оканчивается, если выполняются оба условия:

$$\text{а) } |a_i^{k+1} - a_i^k| < \delta (|a_i^k| + 1), \quad i = 1, \dots, m;$$

$$\text{б) } \|\partial Q(\mathbf{a}^k) / \partial \alpha\| < \kappa.$$

На практике значения δ и κ выбирают равными 10^{-5} . Практика автора показывает, что часто вычисления приходится выполнять с двойной точностью. Как правило, это происходит с «плохо идентифицируемыми» регрессиями, для которых $|\mathbf{P}'\mathbf{P}| \approx 0$.

Большое значение для скорости сходимости процессов минимизации имеет выбор начального вектора параметров. В [123], например, предлагается следующая общая процедура

ра: из n наблюдений выберем m , для которых решим систему уравнений

$$y_{t_i} = f_{t_i}(\alpha_1, \alpha_2, \dots, \alpha_m). \quad (7.33)$$

Решение этой системы примем в качестве начального приближения процесса минимизации. Однако такой метод имеет существенный недостаток. Решение системы (7.33) требует привлечения нелинейных итерационных методов, если f_t не сводятся к линейным функциям некоторым преобразованием. Однако если функция регрессии сводится к линейной с помощью некоторого преобразования, то хорошее начальное приближение может быть получено другим путем (см. параграф 7.6). Если же $f_t(\alpha)$ «действительно» нелинейны, то задача решения системы (7.33) может оказаться сложнее исходной задачи минимизации $Q(\alpha)$.

В том случае, когда начальное приближение не может быть найдено какими-либо способами, можно предложить следующую процедуру. Часто даже если регрессии не сводятся к линейным с помощью некоторого преобразования, можно найти такую трансформацию регрессии, в которой основная часть параметров становится линейной. Так, если число нелинейных параметров после преобразования не больше двух, то, задаваясь сеткой для этих параметров, обычным МНК оцениваем остальные линейные параметры и в качестве начального приближения выберем то значение, которое приводит к минимальному Q . Этот метод, в частности, приемлем для нахождения начального приближения для функции (7.32).

У п р а ж н е н и я 7.4

1. Докажите, что условие (7.23) для метода Левенберга выполняется.
2. Докажите, что если $M_0 < 2\mu_0$, то метод Левенберга сходится.
3. Найдите условия сходимости метода Марквардта.
4. Какие пункты теоремы 7.3 остаются верными для поправки в методе Марквардта (7.31)?

7.5. Единственность оценки МНК

Обсудим вопросы, связанные с наличием у минимизируемой функции нескольких локальных минимумов.

Как только минимизируемая функция многих переменных $f(x)$ становится выпуклой вниз, то все методы минимизации будут сходиться. То же самое можно сказать и о спе-

циальных методах минимизации суммы квадратов отклонений: если в некоторой области $S \subset R^m$ функция $Q(\alpha)$ выпукла и $a \in S$, то a дает единственный глобальный минимум $Q(\alpha)$ на S . Сумма квадратов отклонений не может быть выпуклой на всем пространстве R^m и для всех наблюдений $y \in R^n$. Действительно, если хотя бы для одного $t = 1, \dots, n$, $i, j = 1, \dots, m$, $\alpha \in R^m$ $\partial^2 f_t(\alpha) / \partial \alpha_i \partial \alpha_j \neq 0$, то специальным выбором y_t можно добиться того, что гессиан H (см. уравнение 7.15) не будет положительно определен. Другими словами, если $f_t(\alpha)$ нелинейна, то найдутся наблюдения, для которых $Q(\alpha)$ не будет выпукла. В работе [22] доказывается более сильное утверждение: *найдется $y \in R^n$, для которого $Q(\alpha)$ будет иметь несколько локальных минимумов*. Если же считать, что y имеет, например, нормальное распределение, то вероятность наличия у $Q(\alpha)$ больше двух локальных минимумов больше нуля. Все это говорит о том, что минимизировать Q необходимо с большой осторожностью, чтобы не принять ложное значение оценки МНК за истинное. В работе [22] предлагается метод определения соответствия найденного вектора глобальному минимуму суммы квадратов отклонений. Этот критерий доказан для случая $m = 1$ и для его использования необходимы оценки некоторых характеристик поверхности $F = f(\alpha)$, $\alpha \in R^m$. Для каждой конкретной регрессии может быть предложен, однако, свой способ определения области S , на которой соответствующая сумма квадратов отклонений была бы выпукла. Так, рассмотрим следующий простой критерий для регрессии, линейной в логарифмах.

Легко видеть, что уравнение (7.15) для регрессии (7.3) переписывается следующим образом:

$$H_{ij} = 2 \sum_t e^{\alpha' x_t} x_{ti} x_{tj} (2e^{\alpha' x_t} - y_t), \quad i, j = 1, \dots, m.$$

В матричном виде гессиан $Q(\alpha)$ для регрессии (7.3) равен:

$$H = 2X'DX, \quad (7.34)$$

где D — диагональная матрица с элементом $e^{\alpha' x_t} (2e^{\alpha' x_t} - y_t)$ на главной диагонали. Ясно, что на множестве $S = \{\alpha \in R^m : e^{\alpha' x_t} > y_t/2, t = 1, \dots, n\}$ гессиан (7.34) будет положительно определен, а сумма квадратов отклонений выпукла вниз. Множество S представим в другом виде:

$$S = \{\alpha \in R^m : \alpha' x_t > \ln y_t - \ln 2, t = 1, \dots, n\}. \quad (7.35)$$

Очевидно, S есть выпуклый многогранник в пространстве R^m (рис. 7.4). Далее, легко видеть, если

$$Q(\alpha) < \frac{1}{4} \min_t y_t^2, \quad (7.36)$$

то $\alpha \in S$. Множество векторов, удовлетворяющих (7.36), обозначим S' . Тогда $S' \subset S$. Критерий проверки единственности оценки МНК для регрессии (7.3) прост. Допустим, один из методов минимизации суммы квадратов

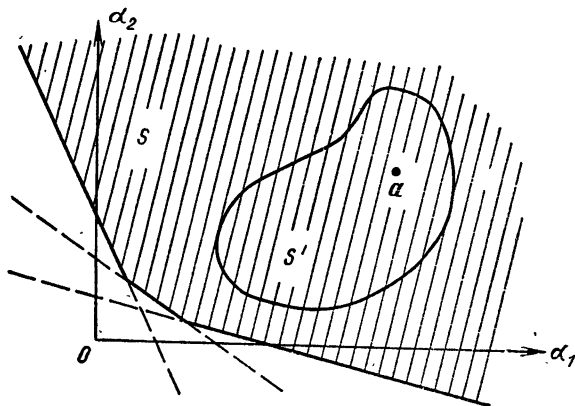


Рис. 7.4. Множество, на котором функция $Q(\alpha)$ регрессии $y_t = e^{\alpha_1 x_{t1} + \alpha_2 x_{t2}} + \varepsilon_t$ выпукла

отклонений привел нас к точке α , в которой $\partial Q(\alpha)/\partial \alpha = 0$, причем $\alpha \in S'$, т. е. $Q(\alpha) < \frac{1}{4} \min_t y_t^2$. Тогда можно утверждать, что на S' точка α отвечает глобальному минимуму. Далее, для всех $\alpha \in \bar{S}'$ $Q(\alpha) \geq \frac{1}{4} \min_t y_t^2$ и поэтому $Q(\alpha) > Q(\alpha)$. Таким образом, α — единственная оценка МНК. Оценка (7.36) завышена, и иногда доказательство единственности оценки МНК по критерию (7.36) «не проходит».

В предыдущем параграфе найден вектор, который обращает градиент $Q(\alpha)$ регрессии (7.5) в нуль. Является ли этот вектор оценкой МНК? Ответ положительный, поскольку $65,2 < \frac{1}{4} (140,28)^2 \approx 7711$.

Часто для того чтобы убедиться в том, что α отвечает глобальному минимуму $Q(\alpha)$, начинают процесс минимизации

ции с другого начального приближения. Если процесс сойдется к старому значению \mathbf{a} , есть уверенность в том, что $Q(\mathbf{a})$ — минимальное значение функции.

У п р а ж н е н и я 7.5

1. Докажите, что если $Q(\alpha)$ выпукла вниз на множестве $S \subset R^m$, то любая точка, приводящая к локальному минимуму $Q(\alpha)$ и лежащая в S , дает глобальный минимум.

2. Докажите, что если $\partial^2 f_t(\alpha) / \partial \alpha_i \partial \alpha_j \neq 0$, то специальным выбором y_1, \dots, y_n можно добиться того, что $Q(\alpha)$ не будет выпуклой вниз функцией.

7.6. Сведение нелинейной регрессии к линейной

Иногда некоторым преобразованием функцию регрессии $f_t(\alpha)$ можно свести к линейной относительно параметров. В дальнейшем регрессию (7.2) будем называть сводящейся к линейной (квазилинейной [148]), если существует такая функция действительного переменного g , что

$$g(f_t(\alpha)) = \alpha_1 f_{t1} + \alpha_2 f_{t2} + \dots + \alpha_m f_{tm}, \quad t = 1, \dots, n, \quad (7.37)$$

где f_{ti} — некоторые константы. Так, для регрессии (7.3) g — логарифмическая функция. Формально производя преобразование g над регрессией (7.2), получим редуцированную регрессию¹:

$$g(y_t) = \alpha_1 f_{t1} + \dots + \alpha_m f_{tm} + \xi_t. \quad (7.38)$$

Применяя обычный МНК к регрессии (7.38), получим начальное приближение для оценки МНК исходной регрессии

$$\mathbf{a}^0 = (\mathbf{f}'\mathbf{f})^{-1} \mathbf{f}' \mathbf{g}(\mathbf{y}), \quad (7.39)$$

где \mathbf{f} — матрица $n \times m$, $\mathbf{g}(\mathbf{y})$ — вектор $n \times 1$. Чем «менее аддитивна» функция g , тем более заметно будет различие между оценкой \mathbf{a}^0 (7.39) и оценкой МНК, минимизирующей $Q(\alpha)$ (7.6).

Различие между \mathbf{a}^0 и \mathbf{a} может быть уменьшено следующим способом. Произведем над обеими частями нелинейной регрессии (7.2) преобразование g . Разложим функцию g в ряд Тейлора до линейных членов в окрестности $f_t(\alpha)$. Тогда

$$g(y_t) = g(f_t(\alpha) + \varepsilon_t) = g(f_t(\alpha)) + g'(s_t) \varepsilon_t, \quad (7.40)$$

¹ «Формально»; поскольку $g(f_t(\alpha) + \varepsilon_t) \neq g(f_t(\alpha)) + g(\varepsilon_t)$.

где s_t лежит между $f_t(\alpha)$ и y_t . Допустим, $f_t(\alpha)$ хорошо аппроксимирует выборку y_t , тогда $f_t(\alpha) \approx y_t$, и (7.40) переписывается следующим образом:

$$g(y_t) = \alpha_1 f_{t1} + \dots + \alpha_m f_{tm} + \xi_t, \quad \xi_t = g'(y_t) \varepsilon_t, \\ t = 1, \dots, n. \quad (7.41)$$

Очевидно, $E\xi_t \approx 0$; $\sigma^2(\xi_t) \approx \sigma^2 \cdot (g'(y_t))^2$. Если $\{\varepsilon_t\}$ независимы, то таковыми будут и $\{\xi_t\}$. Регрессия (7.41) является гетероскедастичной, что ведет к оценке Эйткена

$$a^1 = (f' W^{-1} f)^{-1} f' W^{-1} g(y), \quad (7.42)$$

где W — диагональная матрица с элементом $(g'(y_t))^2$ на диагонали (см. параграф 2.1). Оценка Эйткена может быть найдена и без применения формулы (7.42). Для этого обе части уравнения (7.41) необходимо разделить на $|g'(y_t)|$ и применить МНК. Описанная процедура для некоторого частного случая предложена С. А. Айвазяном [1, с. 172—177]. В более общем случае она рассмотрена в [148]. Как показывает практика, оценка (7.42) является лучшим приближением к оценке МНК, чем (7.39).

Вернемся к регрессиям, линейным в логарифмах. Как следует из вышеизложенного, a^1 есть оценка МНК регрессии

$$y_t \ln y_t = \alpha_1 y_t x_{t1} + \dots + \alpha_m y_t x_{tm} + \varepsilon_t,$$

минимизирующая

$$Q_1(\alpha) = \sum_t (\ln y_t - \alpha_1 x_{t1} - \dots - \alpha_m x_{tm})^2 \cdot y_t^2. \quad (7.43)$$

Оценка a^0 минимизирует

$$Q_0(\alpha) = \sum_t (\ln y_t - \alpha_1 x_{t1} - \dots - \alpha_m x_{tm})^2. \quad (7.44)$$

Сравнивая (7.43) и (7.44), можно утверждать, что a^0 ориентируется по сравнению с a^1 больше на наблюдения с малыми значениями y_t , чем с большими. Для конкретности будем считать, что y_t — наблюдения во времени, причём $y_{t+1} > y_t$. Тогда регрессия $f_t(a^0)$ будет смещена в прошлое, так как наблюдения для малых t входят в (7.44) по сравнению с (7.43) с большим весом. Регрессия $f_t(a^0)$, как видно из рисунка, лучше аппроксимирует данные для малых значений t , тогда как $f_t(a^1) = f_t(\hat{a})$ дает равномерное приближение (рис. 7.5).

До сих пор предполагалось, что отклонения ε_t аддитивны в регрессии, линейной в логарифмах, т. е.

$$y_t = e^{\alpha' x_t} + \varepsilon_t. \quad (7.45)$$

В этом случае оценка логарифмированной регрессии будет иметь систематическое смещение в сторону малых y_t . Однако если регрессию специфицировать по-другому, то

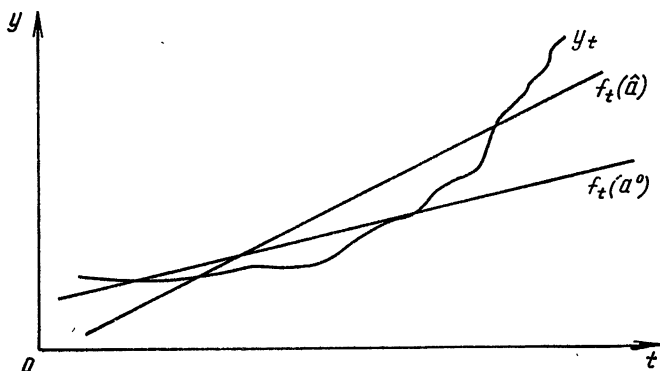


Рис. 7.5. Смещение регрессии в прошлое

оценка a^0 будет оптимальной. Это произойдет, если отклонения входят в модель мультипликативно:

$$y_t = e^{\alpha' x_t} + u_t \quad (7.46)$$

Тогда (7.46) эквивалентна линейной регрессии $\ln y_t = \alpha' x_t + u_t$, и оценка a^0 обладает всеми присущими ей оптимальными свойствами (по теореме 1.2 оценка a^1 будет хуже оценки a^0). Таким образом, вопрос о выборе оценок a^0 или a^1 упирается в выбор спецификации с аддитивной или мультипликативной ошибкой. Разумеется, спецификация регрессии остается за исследователем, однако при выборе (7.45) или (7.46) могут помочь следующие рассуждения. Допустим, исследователь выбрал модель с мультипликативной ошибкой (7.46). Тогда

$$e^{\alpha' x_t} + u_t = e^{\alpha' x_t} e^{u_t} \approx e^{\alpha' x_t} (1 + u_t);$$

$$E \left(\frac{y_t - e^{\alpha' x_t}}{e^{\alpha' x_t}} \right)^2 \approx E u_t^2 = \text{const.}$$

Последнее уравнение означает, что в записи (7.46) близкой к постоянной оказывается относительная ошибка, а в модели (7.45) по определению постоянной является абсолютная ошибка. Таким образом, если исследователь считает, что дисперсия отклонений не будет расти с ростом y_t , то необходимо выбрать модель с аддитивной ошибкой. Если предполагается, что дисперсия отклонений растет вместе с ростом y_t , предпочтительнее выбрать модель (7.46).

Вопросу выбора мультипликативной или аддитивной ошибки в статистической литературе посвящено несколько работ. В [112] рассматривается регрессия, линейная в логарифмах, одновременно с аддитивной и мультипликативной ошибками

$$y_t = e^{\alpha' x_t + \varepsilon_t} + \eta_t, \quad t = 1, \dots, n, \quad (7.47)$$

где $E\varepsilon_t = E\eta_t = 0$, $\sigma^2(\varepsilon_t) = \sigma_1^2$, $\sigma^2(\eta_t) = \sigma_2^2$, $E\varepsilon_{t_1}\eta_{t_2} = 0$, $t_1 = 1, \dots, n$; $t_2 = 1, \dots, n$. Оба отклонения считаются нормально распределенными. Критерий отношения правдоподобия (см. параграф 1.9) приводит к решению задачи о проверке гипотез $H_1: \sigma_1^2 = 0$ и $H_2: \sigma_2^2 = 0$. Однако описанный подход при выборе конкурирующих моделей весьма сложен с точки зрения вычислений. При выборе спецификаций регрессии (7.47) в [151] предлагается использовать λ -преобразование Бокса—Кокса [89]:

$$b^{(\lambda)} = \begin{cases} (b^\lambda - 1)/\lambda, & \lambda \neq 0; \\ \ln b, & \lambda = 0. \end{cases}$$

Применив λ -преобразование к нелинейной регрессии, получим

$$y_t^{(\lambda)} = [f_t(\alpha)]^{(\lambda)} + \delta_t, \quad t = 1, \dots, n,$$

где δ_t — отклонение; $E\delta_t = 0$; $\text{cov}(\delta) = \sigma^2 I_n$. Допустим, δ имеет нормальное распределение, тогда, применяя метод максимального правдоподобия, можно найти оценку для α и λ . Легко проверить, что если $\lambda = 0$, то исходная модель имеет мультипликативную ошибку, если $\lambda = 1$, то — аддитивную.

У п р а ж н е н и я 7. 6.

1. Как найти оценку a^1 для квазилинейных регрессий:

а) $y_t = \alpha_1 e^{\alpha_2 t} K_t^{\alpha_3} L_t^{1-\alpha_3} + \varepsilon_t$; б) $y_t = \alpha_1 / (1 + \alpha_2 x_t) + \varepsilon_t$;

в) $y_t = \sqrt{\alpha_1 + \alpha_2 x_t} + \varepsilon_t$?

7.7. Доказательства

1. Доказательство теоремы 7.1. Рассмотрим множество

$$S = \{x \in R^m : \varphi(x) \leq B\},$$

замкнутое в силу непрерывности функции φ . Оно будет также ограничено, так как в противном случае найдется такая последовательность точек x^1, x^2, \dots из S , что $\|x^k\| \rightarrow \infty$. По условию ограниченности φ на бесконечности снизу найдется такая точка $x^p \in S$ в этой последовательности, что $\varphi(x^p) > B$ — противоречие с определением множества S . Таким образом, S — ограниченное замкнутое множество в R^m , значит, φ достигает на нем своего инфимума (глобального минимума).

2. Доказательство теоремы 7.2. Введем в рассмотрение функцию действительного переменного $\Phi_k(\lambda) = \varphi(x^k + \lambda p_k)$, $\lambda \geq 0$. Легко проверить, что $\Phi'_k = q'(x^k + \lambda p_k) p_k$; $\Phi''_k = p'_k H(x^k + \lambda p_k) p_k$. Разложим функцию $\Phi_k(\lambda)$ в ряд Тейлора до членов второго порядка в окрестности $\lambda = 0$:

$$\Phi_k(\lambda) = \Phi_k(0) - \lambda q'_k p_k + \frac{\lambda^2}{2} p'_k H(x^k + \lambda^* p_k) p_k,$$

где $\lambda^* \in [0, \lambda]$. Допустим, $q_k \neq 0$, в противном случае теорему можно считать доказанной. В силу того что $q'_k p_k \geq \varepsilon > 0$, найдется такое λ^{**} , что для всех $0 < \lambda < \lambda^{**}$ $\Phi(\lambda) < \Phi(0)$. Положим $\lambda^{***} = \sup \lambda^{**}$; где для всех $0 < \lambda < \lambda^{**}$ $\Phi(\lambda) < \Phi(0)$. Тогда $\Phi(\lambda^{***}) = \Phi(0)$ и для всех $0 \leq \lambda \leq \lambda^{***}$ $\Phi(\lambda) < \Phi(0)$, т. е. $\varphi(x^k + \lambda p_k) \leq \varphi(x^k)$. Действительно, неравенство $\Phi(\lambda^{***}) > \Phi(0)$ невозможно в силу непрерывности Φ и определения λ^{***} . Неравенство $\Phi(\lambda^{***}) < \Phi(0)$ также неверно в силу определения λ^{***} . По построению S_0 для всех $0 \leq \lambda \leq \lambda^{***}$ $x^k + \lambda p_k \in S_0$. Далее, по определению M_0 для всех $M \geq M_0$

$$\Phi(\lambda) - \Phi(0) \leq -\lambda q'_k p_k + \frac{M}{2} \lambda^2 p'_k p_k = P_k(\lambda).$$

Очевидно, при всех $0 < \lambda < 2q'_k p_k / M p'_k p_k$ $P(\lambda) < 0$. Легко показать, что $\lambda^{***} \geq 2q'_k p_k / M p'_k p_k$. Пусть $v \in (0, 1/2)$; выберем λ_k так, чтобы

$$v \leq M p'_k p_k \lambda_k / 2 q'_k p_k \leq 1 - v.$$

В этом случае

$$P_k(\lambda_k) = P_k \left(v \frac{2q'_k p_k}{M p'_k p_k} \right) = -\frac{2}{M} \frac{(q'_k p_k)^2}{p'_k p_k} (v - v^2) < 0.$$

Принимая во внимание условие (7.23), последнее неравенство можно переписать следующим образом:

$$P_k(\lambda_k) \leq -\frac{2}{M} \varepsilon^2 (v - v^2) \|q_k\|^2 < 0,$$

или окончательно

$$\varphi(x^k + \lambda_k p_k) < F(x^k) - A \|q_k\|^2,$$

где $A = 2\varepsilon^2(v - v^2)/M > 0$, причем $x^{k+1} = x^k + \lambda_k p_k \in S_0$.

Последовательность x^0, x^1, \dots имеет хотя бы одну предельную точку $x^* \in S_0$. Докажем, что $\|g(x^*)\| = 0$. Допустим противное. Тогда можно найти такую подпоследовательность z^0, z^1, \dots , для которой $\|g(z^k)\| \geq \delta > 0$. Но тогда

$$\varphi(z^{k+1}) < \varphi(z^k) - kA\delta^2,$$

или

$$\varphi(z^{k+1}) < -kA\delta^2 + \varphi(x^0).$$

Переходя к пределу при $k \rightarrow \infty$, получим $\varphi(z^{k+1}) \rightarrow -\infty$, что противоречит ограниченности φ на S_0 . Последнее утверждение теоремы очевидно.

3. Д о к а з а т е л ь с т в о т е о р е м ы 7.3. Очевидно,

$$\|\delta_k^L\|^2 = \frac{1}{4} q'_k (P'_k P_k + \mu I)^{-2} q_k.$$

Используя формулу (П.5), легко показать, что

$$d(P'_k P_k + \mu I)^{-2}/d\mu = -2(P'_k P_k + \mu I)^{-3},$$

поэтому

$$d\|\delta_k^L\|^2/d\mu = -\frac{1}{2} q'_k (P'_k P_k + \mu I)^{-3} q_k.$$

Матрица $P'_k P_k + \mu I$ положительно определена, такой будет и матрица $(P'_k P_k + \mu I)^{-3}$. Таким образом, производная квадрата длины поправки метода Левенберга по μ отрицательна, что доказывает первую часть утверждения а). Вторая часть этого утверждения очевидна.

Докажем теперь, что $\cos(q_k, \delta_k^L)$ является возрастающей функцией μ . По определению

$$\cos(q_k, \delta_k^L) = \frac{q'_k (P'_k P_k + \mu I)^{-1} q'_k}{(q'_k q_k)^{1/2} [q'_k (P'_k P_k + \mu I)^{-2} q'_k]^{1/2}}.$$

Для доказательства достаточно показать, что функция

$$\frac{q'_k (P'_k P_k + \mu I)^{-1} q_k}{[q'_k (P'_k P_k + \mu I)^{-2} q_k]^{1/2}}$$

является возрастающей по μ . Найдем производную этой функции. Она равна:

$$\frac{q'_k (P'_k P_k + \mu I)^{-1} q_k q'_k (P'_k P_k + \mu I)^{-3} q_k - [q'_k (P'_k P_k + \mu I)^{-2} q_k]^2}{[q'_k (P'_k P_k + \mu I)^{-2} q_k]^{3/2}}. \quad (7.48)$$

Теперь заметим, что если A, B, C — квадратные симметричные матрицы, $C = AB$, то по неравенству Шварца

$$(x' C x)^2 = (x' A B x)^2 = [(A x)' (B x)]^2 \leq x' A^2 x \cdot x' B^2 x.$$

Применяя полученное выше неравенство для числителя (7.48), где

$A = (P'_k P_k + \mu I)^{-1/2}$; $B = (P'_k P_k + \mu I)^{-3/2}$; $C = (P'_k P_k + \mu I)^{-2}$, приходим к выводу о положительности производной (7.48) — утверждение б) доказано. Аналогично доказывается утверждение в).

Глава 8

СТАТИСТИЧЕСКИЕ СВОЙСТВА ОЦЕНКИ МНК

8.1. Непрерывность и асимптотические свойства оценки МНК

Исследование статистических свойств оценок МНК в нелинейной регрессии технически весьма сложно. При конечном объеме выборки для установления свойств оценки МНК необходимо знать конкретный вид функции регрессии.

Прежде чем переходить к рассмотрению статистических свойств оценки МНК, обратим внимание на следующий

факт: оценка МНК может не существовать с вероятностью, большей нуля. Для примера рассмотрим регрессию

$$y_t = e^{\alpha t} + \varepsilon_t, \quad (8.1)$$

в которой $\alpha \in \Theta = (-\infty, \infty)$. Если $y_t \leq 0$ для всех $t = 1, 2, \dots, n$, то оценка МНК не существует. Если же предположить, например, что $\varepsilon_t \sim N(0, \sigma^2)$, то $P\{y_t \leq 0, t = 1, \dots, n\} > 0$, и оценка МНК не существует для регрессии (8.1) с вероятностью, большей нуля. Естественно, говорить о непрерывности, состоятельности и других свойствах оценки невозможно, если с положительной вероятностью она не существует. Поэтому в дальнейшем будем предполагать, что Θ — компактное (т. е. замкнутое и ограниченное) множество в R^m . Практически это ограничение не будет жестким, поскольку всегда Θ может быть выбрано произвольных размеров; компактность Θ нам необходима только из теоретических соображений. По-прежнему будем считать $f_t(\alpha)$ непрерывными функциями на Θ , а регрессию — идентифицированной. Компактность Θ и непрерывность f влечет компактность образа F (см. параграф 7.1). Идентифицируемость регрессии означает существование обратного отображения f^{-1} на F . Нетрудно показать, что отображение f^{-1} будет также непрерывным (см.; например, [39, с. 95]).

Компактность Θ влечет существование оценки МНК $a = a_n(y)$ для любого $y \in R^n$. Оценка МНК по определению минимизирует сумму квадратов отклонений

$$Q_n(\alpha; y) = \sum_{t=1}^n (y_t - f_t(\alpha))^2. \quad (8.2)$$

В общем случае может существовать несколько оценок МНК. Р. Дженрич [140] показал, что в этом случае для каждого y можно выбрать $a_n(y)$ так, чтобы оценка МНК стала измеримой функцией. В статье [22] показано, что если оценка МНК единственна, то она непрерывна по y .

Т е о р е м а 8.1. Пусть вышеперечисленные условия выполняются. Если для любого $y \in R^n$ существует единственная оценка МНК, то она является непрерывной функцией y .

Теорема 8.1 означает, что малые изменения в наблюдениях приводят к малым изменениям оценки МНК.

Теперь перейдем к асимптотическим свойствам оценки МНК. Начнем с состоятельности. Под состоятельностью

в слабом смысле понимаем сходимость по вероятности оценки к истинному значению вектора параметров α_0 , т. е. говорим, что $a_n(y)$ — состоятельная оценка, если для любого $\varepsilon > 0$

$$P \{y \in R^n : \|a_n(y) - \alpha_0\| \leq \varepsilon\} \rightarrow 1, n \rightarrow \infty. \quad (8.3)$$

Оценка строго состоятельна, если сходимость к истинному значению вектора параметров происходит почти наверное, т. е. с вероятностью, равной 1. Можно показать, что строгая состоятельность влечет состоятельность в слабом смысле.

Теорема 8.2 [140]. Пусть Θ — компактное множество, $f_t(\alpha)$ — непрерывные функции регрессий на Θ . Допустим, отклонения регрессии $\varepsilon_1, \varepsilon_2, \dots$ независимы и одинаково распределены. Предположим также, что для любых $\alpha, \beta \in \Theta$ существует предел

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n [f_t(\alpha) - f_t(\beta)]^2 = \varphi(\alpha, \beta), \quad (8.4)$$

причем сходимость в (8.4) равномерная, и $\varphi(\alpha, \beta) = 0$ тогда и только тогда, когда $\alpha = \beta$. Тогда оценка МНК строго состоятельна, т. е. $P \{a_n(y) \rightarrow \alpha_0\} = 1$, где α_0 — истинное значение вектора параметров.

Доказательство этой теоремы приведено в параграфе 8.5.

Рассмотрим, что означает условие (8.4) для линейной регрессии $f_t(\alpha) = \alpha' x_t$. Имеем

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n [f_t(\alpha) - f_t(\beta)]^2 &= \frac{1}{n} \sum_{t=1}^n [(\alpha - \beta)' x_t]^2 = \\ &= (\alpha - \beta)' \frac{X_n' X_n}{n} (\alpha - \beta), \end{aligned}$$

где X_n — матрица $n \times m$, вектор-строками которой являются x_t' . Таким образом, (8.4) означает, что $\lim_{n \rightarrow \infty} \frac{1}{n} X_n' X_n = A$, $|A| \neq 0$, т. е. сильную регулярность матриц X_n (1.38).

Э. Маленво [156] также доказал состоятельность оценки МНК при условии (8.4), однако он не предполагал компактности Θ . Вместо этого он рассмотрел другую гипотезу: оценка МНК существует и единственна, причем существует действительное число $d \geq 0$ и компактное мно-

жество $K \subset \Theta$, содержащее истинное значение α_0 , такие, что, начиная с некоторого n_0 ,

$$\frac{1}{n} \sum_t [f_t(\alpha) - f_t(\alpha_0)]^2 \geq 4\sigma^2 + d, n \geq n_0 \quad (8.5)$$

для любого $\alpha \in K$.

Проанализируем условия состоятельности оценки МНК Дженрича и Маленво. Существование предела (8.4) соответствует регулярному поведению матриц X_n в линейной регрессии (1.38). Это условие, как правило, не выполняется для регрессий-трендов и поэтому является весьма ограничительным. Маленво сделал попытку отказаться от компактности Θ , однако, во-первых, он вынужден был предположить существование оценки МНК, а, во-вторых, для этого необходима выполнимость (8.5), что накладывает ограничение на σ^2 сверху.

Было найдено необходимое условие состоятельности оценки МНК в нелинейной регрессии. Можно показать, что если $\varepsilon_t \sim N(0, \sigma^2)$, то для состоятельности оценки МНК необходимо, чтобы

$$\sum_{t=1}^n [f_t(\alpha) - f_t(\alpha_0)]^2 \rightarrow \infty, n \rightarrow \infty,$$

при условии, что все остальные стандартные условия, налагаемые на нелинейную регрессию, выполнены. Приведенное условие является необходимым и достаточным для линейной регрессии. Докажем это. В линейной регрессии $f_t(\alpha) = \alpha' x_t$, $t = 1, \dots, n$, поэтому

$$\sum_{t=1}^n (\alpha' x_t - \alpha_0' x_t)^2 = (\alpha - \alpha_0)' X_n' X_n (\alpha - \alpha_0),$$

где X_n — матрица $n \times m$, вектор-строки которой суть x_t' . Докажем, что если для любого $v \in R^m$ $v' X_n' X_n v \rightarrow \infty$ при $n \rightarrow \infty$, то $\lambda_{\min}(X_n' X_n) \rightarrow \infty$, и выполнено условие Эйкера (1.39). Положим $\psi_n(v) = v' X_n' X_n v$, причем $\|v\| = 1$. Имеем $X_{n+1}' X_{n+1} = X_n' X_n + x_{n+1}' x_{n+1}$, поэтому $X_{n+1}' X_{n+1} \geq X_n' X_n$ и $\psi_{n+1}(v) \geq \psi_n(v)$ для любого $v \in S$ — единичной сфере в R^m . В силу компактности S и монотонности последовательности функций ψ_n сходимость $\psi_n(v) \rightarrow \infty$ будет равномерной на S , поэтому

$$\min_{v \in S} \psi_n(v) = \lambda_{\min} X_n' X_n \rightarrow \infty, n \rightarrow \infty,$$

что требовалось показать. Обратное очевидно.

Если ε_t независимы и одинаково распределены с плотностью, вторая производная которой не обращается в нуль вне некоторого интервала, и существуют α и α_0 , для которых $\sum_{t=1}^{\infty} [f_t(\alpha) - f_t(\alpha_0)]^2 < \infty$, то вообще состоятельной оценки не существует [44].

Перейдем к асимптотической нормальности оценки МНК. Помимо (8.4), предположим, что *вторые производные $\partial^2 f_t / \partial \alpha^2$ равномерно ограничены в совокупности*, т. е. найдется такое число M , что для всех $\alpha \in \Theta$

$$\left| \frac{\partial^2 f_t(\alpha)}{\partial \alpha_i \partial \alpha_j} \right| \leq M, \quad t = 1, 2, \dots; \quad i, j = 1, 2, \dots, m. \quad (8.6)$$

Далее предположим, что при $n \rightarrow \infty$

$$\frac{\max(\partial f_t / \partial \alpha_i)^2}{\sum_{t=1}^n (\partial f_t / \partial \alpha_i)^2} \rightarrow 0; \quad \frac{1}{n} \sum_{t=1}^n \frac{\partial f_t}{\partial \alpha_i} \frac{\partial f_t}{\partial \alpha_j} \rightarrow A_{ij}, \quad i, j = 1, \dots, m \quad (8.7)$$

равномерно по $\alpha \in \Theta$, где A_{ij} — матрица $m \times m$, причем $|A(\alpha_0)| \neq 0$.

Теорема 8.3. Пусть условия теоремы 8.2 выполняются, т. е. $a_n(\mathbf{y})$ — строго состоятельная оценка α_0 . Предположим также, что $a_n(\mathbf{y})$ является внутренней точкой Θ почти наверное. Тогда если условия (8.6) и (8.7) выполняются, то оценка МНК является асимптотически-нормальной:

$$\sqrt{n}(\mathbf{a}_n(\mathbf{y}) - \alpha_0) \overset{d}{\rightarrow} N(0, \sigma^2 \mathbf{A}^{-1}(\alpha_0)).$$

Доказательство теоремы приведено в параграфе 8.5.

Теорема 8.3 помогает найти приближенную матрицу ковариаций оценки МНК. Так, положим

$$\text{cov}(\mathbf{a}_n) \approx s^2 \left[\sum_{t=1}^n \left(\frac{\partial f_t(\mathbf{a}_n)}{\partial \alpha} \right) \left(\frac{\partial f_t(\mathbf{a}_n)}{\partial \alpha} \right)' \right]^{-1}, \quad (8.8)$$

где

$$s^2 = \frac{1}{n} \sum_t (\mathbf{y}_t - f_t(\mathbf{a}_n))^2$$

— состоятельная оценка σ^2 . Матрица (8.8) совпадает с матрицей ковариаций оценки МНК линейной регрессии (7.18), которая является линейризацией исходной регрессии (7.2) в точке \mathbf{a}_n ; здесь $\mathbf{P} = \mathbf{P}(\mathbf{a}_n) = \partial f(\mathbf{a}_n) / \partial \alpha$.

Упражнения 8.1

1. Рассмотрим нелинейную регрессию $f(t)(\alpha) = \alpha x_{t1} + \alpha^2 x_{t2}$, у которой матрица X_n , вектор-строки которой равны (x_{t1}, x_{t2}) , регулярна, т. е. $\lim X_n' X_n / n = B$, $|B| \neq 0$. Пусть $\Theta = \{\alpha \in R^1 : 0 \leq \alpha \leq \alpha^*\}$. Докажите, что условие Дженрича (8.4) для этой регрессии выполняется.

2. При каком условии оценка МНК для регрессии из задачи 1 будет асимптотически-нормальна?

3. Предположим $\Theta = \{\alpha \in R^1 : \alpha \geq 0\}$. Существует ли тогда оценка МНК для регрессии из задачи 1? Для каких σ^2 выполняется условие Маленво (8.5)? Будет ли оценка МНК состоятельной?

4. Выполняется ли условие Дженрича (8.4) для регрессии; линейной в логарифмах $f_t(\alpha) = \exp(\alpha' x_t)$, у которой $\lim X_n' X_n / n = B$, $|B| \neq 0$, $\alpha \in \Theta$ — компактное множество? Будет ли оценка МНК асимптотически-нормальной?

5. Пусть в нелинейной регрессии $y_t = \alpha + \alpha^2 x_t + \varepsilon_t$ существуют пределы $c_1 = \sum x_t / n$, $c_2 = \sum x_t^2 / n > 0$, $c_1 > 0$, $\alpha \in [0, b]$. Будут ли условия (8.4) и (8.7) для этой регрессии выполняться?

8.2. Оценка смещения МНК

В линейной регрессии оценка МНК является несмещенной оценкой. Это свойство не сохраняется для оценки МНК в нелинейной регрессии. Даже в простейшей нелинейной регрессии $y_t = \sqrt{\alpha} + \varepsilon_t$, $t = 1, \dots, n$, где y_1, \dots, y_n независимы и одинаково распределены по нормальному закону, оценка МНК $(\sum y_t / n)^2$ параметра α будет иметь смещение.

М. Бокс попытался [88] оценить величину смещения оценки МНК в нелинейной регрессии. Оценка МНК удовлетворяет матричному уравнению (система нормальных уравнений):

$$P(\mathbf{a})(\mathbf{y} - f(\mathbf{a})) = 0, \quad (8.9)$$

где $P(\mathbf{a})$ — матрица производных $n \times m$, \mathbf{a} — оценка МНК. Разложим функцию регрессии в ряд Тейлора до членов второго порядка

$$\begin{aligned} f_t(\alpha) &= f_t(\alpha_0) + (\alpha - \alpha_0)' \frac{\partial f_t}{\partial \alpha}(\alpha_0) + \\ &+ \frac{1}{2} (\alpha - \alpha_0)' H_t(\alpha_0) (\alpha - \alpha_0), \end{aligned}$$

где α_0 — истинное значение параметра, $H_t(\alpha_0)$ — матрица вторых производных f_t . Последнее выражение может

быть переписано в матричном виде

$$f(\alpha) = f(\alpha_0) + P(\alpha_0)(\alpha - \alpha_0) + \frac{1}{2} G(\alpha - \alpha_0), \quad (8.10)$$

где

$$G(\alpha_0) = \{H_1(\alpha - \alpha_0), H_2(\alpha - \alpha_0), \dots, H_n(\alpha - \alpha_0)\}'$$

— составная матрица $n \times m$. С учетом (8.10)

$$y - f(a) = \varepsilon - P(a)(a - \alpha_0) - \frac{1}{2} G(a)(a - \alpha_0). \quad (8.11)$$

Матрицу производных также разложим в ряд Тейлора до линейных членов

$$P(a) = P(\alpha_0) + G(\alpha_0). \quad (8.12)$$

Разность $a - \alpha_0$ есть функция наблюдений y_1, y_2, \dots, y_n и неизвестного параметра α_0 , т. е. функция $\varepsilon_1, \dots, \varepsilon_n$ и α_0 . Приближенно она может быть аппроксимирована квадратичной функцией, т. е.

$$a - \alpha_0 = A\varepsilon + q, \quad (8.13)$$

где

$$q = \{\varepsilon' B_1 \varepsilon, \varepsilon' B_2 \varepsilon, \dots, \varepsilon' B_m \varepsilon\}',$$

A — матрица $m \times n$, B_i — матрица $n \times n$. В формуле (8.13) отсутствует постоянный член, так как при $\sigma^2 = 0$ ($\varepsilon_1 = \varepsilon_2 = \dots = \varepsilon_n = 0$ п.н.) $a - \alpha_0 = 0$. Нас в дальнейшем будет интересовать математическое ожидание $E(a - \alpha_0) = E q$.

Подставим (8.12) и (8.11) в (8.9), получим

$$(P + G)'(\varepsilon - P(a - \alpha_0) - \frac{1}{2} G(a - \alpha_0)) = 0,$$

где P и G рассчитаны в точке $\alpha = \alpha_0$. Теперь подставим в полученное уравнение выражение (8.13):

$$(P + G)' \left(\varepsilon - PA\varepsilon - Pq - \frac{1}{2} GA\varepsilon - \frac{1}{2} Gq \right) = 0.$$

Приравняем к нулю члены при ε : $P' - P'PA = 0$, откуда $A = (P'P)^{-1}P'$. Теперь приравняем к нулю члены второго порядка

$$P' \left(-P'q - \frac{1}{2} JA\varepsilon \right) + J'(\varepsilon - PA\varepsilon) = 0, \quad (8.14)$$

где \mathbf{J} — матрица $n \times m$

$$\mathbf{J} = \{\mathbf{H}_1 \mathbf{A}\varepsilon, \mathbf{H}_2 \mathbf{A}\varepsilon, \dots, \mathbf{H}_n \mathbf{A}\varepsilon\}.$$

Возьмем математическое ожидание от обеих частей (8.14). Можно показать, что $\mathbf{E}[\mathbf{J}(\varepsilon - \mathbf{P}\mathbf{A}\varepsilon)] = 0$, поэтому

$$\mathbf{P}' \left(\mathbf{P}\mathbf{E}(\mathbf{q}) + \frac{1}{2} \mathbf{d} \right) = 0, \quad (8.15)$$

где

$$\begin{aligned} d_t &= \sigma^2 \operatorname{tr}(\mathbf{A}' \mathbf{H}_t \mathbf{A}) = \sigma^2 \operatorname{tr} \mathbf{P} (\mathbf{P}' \mathbf{P})^{-1} \mathbf{H}_t (\mathbf{P}' \mathbf{P})^{-1} \mathbf{P}' = \\ &= \sigma^2 \operatorname{tr} \mathbf{H}_t (\mathbf{P}' \mathbf{P})^{-1} \mathbf{P}' \mathbf{P} (\mathbf{P}' \mathbf{P})^{-1} = \sigma^2 \operatorname{tr} \mathbf{H}_t (\mathbf{P}' \mathbf{P})^{-1}. \end{aligned}$$

Выражая $\mathbf{E}(\mathbf{q})$ из уравнения (8.15), окончательно получаем

$$\mathbf{E}(\mathbf{q}) = -\frac{1}{2} (\mathbf{P}' \mathbf{P})^{-1} \mathbf{P}' \mathbf{d}. \quad (8.16)$$

В формуле (8.16) значения матрицы \mathbf{P} и вектора \mathbf{d} могут быть приближенно заменены их значениями в точке — оценке МНК.

М. Бокс проверял формулу (8.16) методом Монте-Карло. Формула (8.16) давала хорошее приближение к истинному смещению оценки МНК.

В качестве иллюстрации формулы (8.16) рассмотрим нелинейную регрессию

$$y_t = \sqrt{\alpha} + \varepsilon_t, \quad t = 1, \dots, n, \quad (8.17)$$

где $f_t(\alpha) = \sqrt{\alpha}$, $\Theta = \{\alpha : \alpha \geq 0\}$. Оценкой МНК для регрессии (8.17) является $a = \frac{1}{n^2} (\sum y_t)^2$. Найдем смещение оценки МНК непосредственно:

$$\mathbf{E}a = \frac{1}{n^2} \sum_t \mathbf{E}y_t^2 + \frac{1}{n^2} \sum_{t \neq j} \mathbf{E}(y_t y_j).$$

Но

$$\begin{aligned} \mathbf{E}y_t^2 &= \mathbf{E}(\sqrt{\alpha} + \varepsilon_t)^2 = \alpha + \sigma^2, \\ \mathbf{E}(y_t y_j) &= \mathbf{E}(\sqrt{\alpha} + \varepsilon_t)(\sqrt{\alpha} + \varepsilon_j) = \alpha, \quad t \neq j. \end{aligned}$$

Поэтому

$$\mathbf{E}a = \frac{1}{n^2} [n(\alpha + \sigma^2) + \alpha(n^2 - n)] = \alpha + \frac{\sigma^2}{n},$$

т. е. смещение равно $\mathbf{E}a - \alpha = \frac{\sigma^2}{n}$.

Найдем смещение оценки МНК, используя формулу (8.16). Для регрессии (8.17)

$$\frac{\partial f_t}{\partial \alpha} = \frac{1}{2} \cdot \frac{1}{\sqrt{\alpha}}; \quad \frac{\partial^2 f_t}{\partial \alpha^2} = -\frac{1}{4} \alpha^{-3/2},$$

поэтому $\mathbf{P}'\mathbf{P} = n/4 \alpha$ и $d_t = -\sigma^2/\sqrt{\alpha} n$, $t = 1, \dots, n$. Таким образом,

$$Eg = \frac{1}{2} \left[\frac{4\alpha}{n} \cdot \frac{\sigma^2}{2n} \cdot \frac{n}{\alpha} \right] = \frac{\sigma^2}{n},$$

что совпадает с истинным смещением.

У п р а ж н е н и я 8.2.

1. Оцените смещение МНК в регрессии $f_t(\alpha) = e^{\alpha t}$, $t = 1, \dots, n$ по формуле (8.16).

2. То же самое сделайте для регрессии $f_t(\alpha) = e^{\alpha x t}$.

3. Пусть ε_t распределены по нормальному закону $N(0, \sigma^2)$ и $y_t = \sqrt[3]{\alpha} + \varepsilon_t$. Найдите истинные смещения оценок МНК и сравните их со смещениями, полученными по формуле (8.16).

8.3. Проверка статистических гипотез и доверительное оценивание

Начнем с проверок статистических гипотез в нелинейной регрессии.

Имеется нелинейная регрессия (7.2), относительно которой выполнены стандартные предположения: Θ — априорное множество α компактно, оценка МНК $\mathbf{a}_n(\mathbf{y})$ единственна, $f_t(\alpha)$ непрерывны на Θ и т. д. В дополнение предположим, что $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ имеют нормальное распределение, т. е. $\varepsilon \sim N(0, \sigma^2 \mathbf{I}_n)$.

Выдвигается простая статистическая гипотеза

$$H_0: \alpha = \alpha_0, \quad (8.18)$$

где α_0 — фиксированная точка в Θ . Требуется построить критерий проверки гипотезы (8.18). Необходимо отметить, что поскольку мы не ограничиваемся специальным классом нелинейных регрессий, т. е. $f_t(\alpha)$, вообще говоря, могут быть любыми, то оптимальность любого критерия при фиксированном n установить невозможно. В лучшем случае говорим об асимптотической оптимальности критерия: асимптотическая несмещенность, асимптотически наиболее мощный критерий и т. д. Даже если класс функций регрессий известен, например функции, линейные в логарифмах,

исследование критериев для данного n — задача, технически весьма сложная.

Простейший путь проверки статистической гипотезы (8.18) состоит в следующем. Аппроксимируем регрессию (7.2) линейной:

$$f_t(\alpha) \approx f_t(\mathbf{a}) + \sum_{i=1}^m \frac{\partial f_t(\mathbf{a})}{\partial \alpha_i} (\alpha_i - a_i), \quad t=1, \dots, n, \quad (8.19)$$

где \mathbf{a} — оценка МНК. Уравнение (7.2) с учетом (8.19) в матричном виде может быть переписано следующим образом:

$$\mathbf{z} = \mathbf{P}\alpha + \varepsilon, \quad (8.20)$$

где $\mathbf{z} = \mathbf{y} - \mathbf{f}(\mathbf{a}) + \mathbf{P}\mathbf{a}$ — вектор $n \times 1$, \mathbf{P} — матрица производных $\partial f_t / \partial \alpha_i$, вычисленных в точке \mathbf{a} . Условия (8.4) и (8.6) гарантируют хорошую замену исходной нелинейной регрессии линеаризованной (8.20) при больших n . Таким образом, первый метод проверки статистических гипотез состоит в том, чтобы вместо исходной нелинейной модели (7.2) рассматривать линеаризованную модель (8.20). Проверка линейных гипотез для линейной регрессии подробно рассмотрена в параграфе 1.10.

Разумеется, подобный метод проверок гипотез является весьма грубым. Он будет тем точнее, чем «линейнее» будет исходная модель.

Более точный метод предложил А. Галлант [104, 108]. Он основан на критерии отношения правдоподобия (см. параграф 1.9). Отношение правдоподобия для гипотезы (8.18) равно:

$$\frac{\max_{\alpha_0, \sigma^2} p(\mathbf{y}; \alpha_0, \sigma^2)}{\max_{\alpha, \sigma^2} p(\mathbf{y}; \alpha, \sigma^2)} = (T(\mathbf{y}))^{-n/2},$$

где плотность равна:

$$p(\mathbf{y}; \alpha, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{f}(\alpha))' (\mathbf{y} - \mathbf{f}(\alpha)) \right].$$

Статистика критерия

$$T(\mathbf{y}) = \frac{\sum_t (y_t - f_t(\alpha_0))^2}{\sum_t (y_t - f_t(\mathbf{a}))^2}.$$

Критическим множеством проверки простой гипотезы (8.18) является

$$E_K = \{y \in R^n : T(y) > \varphi\},$$

где φ выбрано так, чтобы $P_{\alpha_0} \{T(y) > \varphi\} = \lambda$ — вероятности совершения ошибки первого рода. Галлантом доказано, что статистика $T(y)$ может быть разложена в сумму двух случайных величин $T(y) = x + c_n$, где nc_n по вероятности стремится к нулю при $n \rightarrow \infty$, а случайная величина x имеет определенное распределение, которое затабулировано для некоторых значений φ Галлантом в [104]. Распределение x весьма сложно, однако с применением ЭВМ оно может быть вычислено для любого φ . Таким образом, отождествляя $T(y)$ и x , задаваясь некоторым φ , мы можем найти соответствующий уровень значимости λ . Расчеты по методу Монте-Карло для регрессии $f_t(\alpha_1, \alpha_2) = \alpha_1 e^{\alpha_2 x_t}$ показали, что критерий отношения правдоподобия Галланта приводит к хорошим результатам. В следующей его статье [108] разбирается случай проверки сложной гипотезы $H : \alpha_i = \alpha_i^0, i = 1, 2, \dots, k < m$, где $\alpha^0 = (\alpha_1^0, \dots, \alpha_k^0)' \in R^k$ — фиксированный вектор. Применяя тот же метод, Галлант строит аппроксимацию распределения статистики критерия отношения правдоподобия, по которой для заданного значения φ может быть вычислено соответствующее значение λ .

Перейдем теперь к построению доверительных интервалов и областей для параметров нелинейной регрессии. Простейший способ доверительного оценивания в этом случае — вместо исходной регрессии рассматривать ее линейный аналог (8.20).

Матрица ковариаций оценки МНК, вычисленная на основе (8.20), приближенно равна $s^2 (P'P)^{-1}$. Стандартной ошибкой параметра a_i является $s_i = s \sqrt{(P'P)^{-1}_{ii}}$, $i = 1, \dots, m$, а 95%-ным доверительным интервалом для a_i будет

$$(a_i - t_{0,05} s_i; a_i + t_{0,05} s_i), \quad (8.21)$$

где $t_{0,05}$ — критическая точка t -распределения с $n - m$ степенями свободы, т. е. $P\{|t| \geq t_{0,05}\} = 0,05$. Можно проверять гипотезы о значимости параметров, т. е. $H_i : \alpha_i = 0$. Так, если $|a_i|/s_i > t_{0,05}$, то гипотезу $\alpha_i = 0$ отвергаем.

А. Галлант исследовал распределение a_i/s_i для регрессии

$$y_i = \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 e^{\alpha_4 x_{i3}} + \varepsilon_i \quad (8.22)$$

методом Монте-Карло [105]. Для данных значений x_{ti} , $t = 1, \dots, 30$; $i = 1, 2, 3, 4$; $\alpha = (0; 1; -1; -0,5)'$; $\sigma^2 = 0,001$, были смоделированы ε_t , распределенные по нормальному закону $N(0, \sigma^2)$. Затем вычислялись y_t , регрессия (8.22) оценивалась модифицированным методом Ньютона — Гаусса и вычислялось значение $\tilde{t}_i = a_i/s_i$, $i = 1, 2, 3, 4$. Такие эксперименты были проделаны 5 тыс. раз. Для заданных значений c были вычислены эмпирические вероятности $P(\tilde{t} \leq c)$ и теоретические, основанные на t -распределении. В табл. 8.1 приведены выдержки из табл. 1 [105], где $P(\tilde{t}_i \leq c)$ — эмпирические значения вероятностей.

Таблица 8.1

c	$P(t \leq c)$	$P(\tilde{t}_1 \leq c)$	$P(\tilde{t}_2 \leq c)$	$P(\tilde{t}_3 \leq c)$	$P(\tilde{t}_4 \leq c)$
0,0	0,5000	0,5152	0,4800	0,4974	0,5196
1,315	0,9000	0,9038	0,8914	0,8776	0,9004
1,706	0,9500	0,9552	0,9498	0,9314	0,9486
2,779	0,9950	0,9950	0,9940	0,9852	0,9936

Как видим, расхождения между $P(t \leq c)$ и $P(\tilde{t}_i \leq c)$ весьма малы, что указывает на то, что доверительные интервалы (8.21) для регрессии (8.22), вероятно, будут хорошими, и, в частности, $P(a_i - t_{0,05} s_i \leq a_i \leq a_i + t_{0,05} s_i) = 0,95$. Однако необходимо отметить следующее: во-первых, регрессия (8.22) является «не очень нелинейной»: три из четырех параметров — линейны, а, во-вторых, в эксперименте было взято малое значение σ^2 , при котором расхождение между моделью (8.22) и ее линейным аналогом будет невелико.

Можно предложить другую формулу для вычисления ковариационной матрицы оценки МНК в нелинейной регрессии. В линейной регрессии $2 \mathbf{X}'\mathbf{X} = \frac{\partial^2 Q}{\partial \alpha^2}$, поэтому

$$\text{cov}(\mathbf{a}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = \sigma^2 \left(\frac{1}{2} \cdot \frac{\partial^2 Q}{\partial \alpha^2} \right)^{-1}.$$

Эту же формулу используем в нелинейной регрессии, но теперь

$$\frac{1}{2} \cdot \frac{\partial^2 Q}{\partial \alpha^2}(\mathbf{a}) = \mathbf{P}'\mathbf{P} - \sum_{i=1}^n \mathbf{e}_i \mathbf{H}_i(\mathbf{a}), \quad (8.23)$$

где $\mathbf{H}_t(\mathbf{a})$ — матрица вторых производных $f_t(\alpha)$, вычисленная в точке \mathbf{a} . Стандартной ошибкой параметра a_i является

$$s \sqrt{\left(\frac{1}{2} \partial^2 Q(\mathbf{a}) / \partial \alpha^2\right)_{ii}^{-1}}, \quad i = 1, \dots, m.$$

Аналогично (8.21) могут быть построены доверительные интервалы. Как видно из формулы (8.23), матрица ковариаций, построенная на

основе гессiana суммы квадратов отклонений, учитывает нелинейность регрессии, которая отражается во вторых производных $\mathbf{H}_t(\mathbf{a})$. Если \mathbf{a} принадлежит внутренности априорного множества Θ , то $\partial^2 Q(\mathbf{a}) / \partial \alpha^2$ по крайней мере будет неотрицательно определена (в противном случае в окрестности $V(\mathbf{a})$ нашлась бы точка $\bar{\mathbf{a}}$, в которой $Q(\bar{\mathbf{a}}) < Q(\mathbf{a})$). Чем островеершинней будет поверхность $Q(\alpha)$ в окрестности точки \mathbf{a} , тем меньше будут дисперсии оценки МНК; чем поверхность $Q(\alpha)$ будет положе, тем дисперсии a_i будут больше. Чем ближе $\partial^2 Q(\mathbf{a}) / \partial \alpha^2$ к вырожденной матрице, тем сложнее «отделить» один параметр от другого. Так, на рис. 8.1 линии уровня S вытянуты в одном направлении и сжаты в другом, поэтому матрица $\partial^2 Q / \partial \alpha^2$ близка к вырожденной.

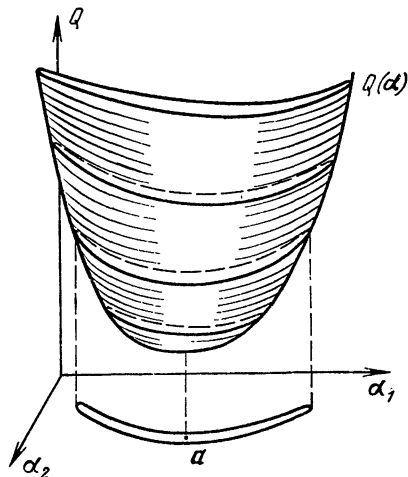


Рис. 8.1. Пример суммы квадратов отклонений, имеющей овражный характер

Перейдем к построению одновременных доверительных областей. Как и прежде, предполагаем, что ϵ_t распределены по нормальному закону. Ранее показано, что критерий отношения правдоподобия проверки простой гипотезы (8.18) приводит к множеству принятия гипотезы

$$E_H = \{\mathbf{y} \in R^n : Q(\alpha; \mathbf{y}) \leq \varphi Q(\mathbf{a}; \mathbf{y})\}, \quad \varphi \geq 1,$$

или

$$E_H = \{\mathbf{y} \in R^n : Q(\alpha; \mathbf{y}) - Q(\mathbf{a}; \mathbf{y}) \leq \varphi' Q(\mathbf{a}; \mathbf{y})\},$$

$$\varphi' \geq 0. \quad (8.24)$$

В параграфе 1.9 установлена связь между проверкой простой гипотезы и доверительным оцениванием. В частности, если имеется критерий проверки гипотезы, то по нему может быть построен метод доверительного оценивания. Используя (8.24), найдем соответствующее доверительное множество

$$D(y) = \{\alpha \in R^m : Q(\alpha; y) - Q(a; y) \leq \varphi' Q(a; y)\}. \quad (8.25)$$

Значение φ' должно быть выбрано таким образом, чтобы $\{P[D(y) \text{ накрывает истинное значение параметра}$

$$\alpha_0\} \geq 1 - \lambda. \quad (8.26)$$

В линейной регрессии множеству (8.25) соответствует множество (1.75), так как

$$Q(\alpha; y) - Q(a; y) = (\alpha - a)' X' X (\alpha - a).$$

В случае линейной регрессии

$$\varphi' = \frac{m}{n-m} F_\lambda(m, n-m), \quad (8.27)$$

где $F(m, n-m)$ обозначает f -распределение с m и $n-m$ степенями свободы, а $F_\lambda(m, n-m)$ — такая точка, что $P(f > F_\lambda(m, n-m)) = \lambda$.

Значение φ' (8.27) может быть выбрано различными методами и для нелинейной регрессии. Первый метод выбора φ' совпадает с методом выбора в линейной регрессии. Очевидно, в этом случае у нас нет уверенности, что неравенство (8.26) будет выполнено, можно лишь надеяться, что оно выполняется с достаточной точностью.

Второй метод выбора $\varphi' = 1 - \varphi$ основан на рассмотренной работе А. Галланта [104].

Третий метод нахождения φ' предложен Е. Билом [83]: φ' выбирается как в линейной регрессии (8.27), но с учетом поправки на нелинейность. Рассмотрим этот метод более подробно. Коэффициент нелинейности регрессии

$$y_t = f_t(\alpha) + \varepsilon_t, \quad t = 1, \dots, n, \quad (8.28)$$

который обозначим N_α , определяется следующим образом. Пусть a — оценка МНК; выберем в окрестности a k произвольных точек a_1, a_2, \dots, a_k , которым на образе $F = f(\alpha)$, $\alpha \in \Theta$ соответствуют точки $f(a_1), f(a_2), \dots, f(a_k)$. После линеаризации регрессии (8.28) превращается в регрессию (8.20), которая соответствует касательному линейному многообразию размерности $n-m$:

$$l(\alpha) = f(a) + P(\alpha - a), \quad \alpha \in \Theta, \quad l(\alpha) \in R^n. \quad (8.29)$$

Тогда сумма квадратов расстояний в точках $\mathbf{a}_1, \dots, \mathbf{a}_k$ будет характеризовать отклонение нелинейной регрессии (8.28) от ее линейного приближения (8.29):

$$\sum_{i=1}^k \|f(\mathbf{a}_i) - l(\mathbf{a}_i)\|^2. \quad (8.30)$$

Сумма (8.30) зависит, во-первых, от числа выбранных точек k , во-вторых, от расстояния точек $f(\mathbf{a}_i)$ от $f(\mathbf{a})$. Для нормировки суммы (8.30) введем величину $\sum_{i=1}^k \|f(\mathbf{a}_i) - f(\mathbf{a})\|^4$.

Окончательно коэффициентом нелинейности регрессии (8.28) по Билу называется число

$$N_\alpha = ms^2 \frac{\sum_{i=1}^k \|f(\mathbf{a}_i) - l(\mathbf{a}_i)\|^2}{\sum_{i=1}^k \|f(\mathbf{a}_i) - f(\mathbf{a})\|^4}, \quad (8.31)$$

где s^2 — оценка параметра σ^2 , рассчитанная по формуле

$$s^2 = \frac{\sum (y_t - f_t(\mathbf{a}))^2}{n - m}.$$

Коэффициент N_α отражает нелинейность регрессии относительно параметра α .

Теперь рассмотрим другой коэффициент, выступающий показателем внутренней нелинейности регрессии (8.28) и отражающий степень нелинейности образа F . При взаимно-однозначных отображениях Θ на себя, т. е. $\psi: \Theta \rightarrow \Theta$, образ $F = f(\Theta)$ остается неизменным, тогда как величина (8.31) меняется. В качестве показателя нелинейности F Бил называет минимальное значение N_α при всех преобразованиях параметрического множества Θ (репараметризация α) при условии, что точки $f(\mathbf{a}_1), \dots, f(\mathbf{a}_k)$ остаются неизменными: $N_f = \min_{\psi} N_\alpha$. Очевидно, для нахождения N_f необходимо из точек $f(\mathbf{a}_i)$ опустить перпендикуляр на пространство (8.29), сумма этих перпендикуляров составит числитель N_f :

$$N_f = ms^2 \frac{\sum_i \|e_i\|^2}{\sum_i \|f(\mathbf{a}_i) - f(\mathbf{a})\|^4},$$

где \mathbf{e}_i — перпендикуляр, опущенный из точки $f(\mathbf{a}_i)$ на линейное многообразие (8.29). Практически $\|\mathbf{e}_i\|^2$ может быть вычислен как сумма квадратов отклонений в регрессии $f(\mathbf{a}_i) = f(\mathbf{a}) + P'(\alpha - \mathbf{a}_i) + \xi$. Е. Билом была установлена тесная взаимосвязь коэффициента N_f с вероятностью множества $\{\alpha_0 \in D(\mathbf{y})\}$. В [83] предлагаются следующие окончательные рекомендации:

в случае $m = 1$ φ' брать равным

$$\varphi'' = \left(1 + \frac{n}{n-1} N_f\right) \varphi';$$

в случае $m > 1$ $\varphi'' = \left(1 + \frac{n(m+2)}{(n-m)m} N_f\right) \varphi'$, (8.32)

φ' вычисляется по формуле (8.27). Бил утверждает, что с большой степенью приближения

$$P\{Q(\alpha_0; \mathbf{y}) - Q(\alpha; \mathbf{y}) \leq \varphi'' Q(\mathbf{a}; \mathbf{y})\} \geq 1 - \lambda.$$

Построение доверительных областей (8.25) технически может оказаться весьма сложным. Дело может осложниться тем, что область $D(\mathbf{y})$ будет несвязной. Для облегчения построения области D можно предложить следующий способ. Разложим функцию $Q(\alpha)$ в точке $\alpha = \mathbf{a}$ в ряд Тейлора до членов второго порядка:

$$Q(\alpha; \mathbf{y}) - Q(\mathbf{a}; \mathbf{y}) \approx (\alpha - \mathbf{a})' \frac{\partial Q(\mathbf{a})}{\partial \alpha} + \\ + \frac{1}{2} (\alpha - \mathbf{a})' \frac{\partial^2 Q(\mathbf{a})}{\partial \alpha^2} (\alpha - \mathbf{a}).$$

Но если \mathbf{a} является внутренней точкой Θ , то $\partial Q(\mathbf{a})/\partial \alpha = 0$, поэтому вместо (8.25) можно найти его приближение

$$D'(\mathbf{y}) = \left\{ \alpha \in R^m : (\alpha - \mathbf{a})' \left[\frac{1}{2} \frac{\partial^2 Q(\mathbf{a})}{\partial \alpha^2} \right] \times \right. \\ \left. \times (\alpha - \mathbf{a}) \leq \varphi' Q(\mathbf{a}; \mathbf{y}) \right\}, \quad (8.33)$$

где $1/2 \partial^2 Q(\mathbf{a})/\partial \alpha^2$ рассчитывается по формуле (8.23). Область D' представляет собой эллипсоид в пространстве R^m . Строить эллипсоид D' не обязательно, достаточно определить положение его осей и их длины. Направление осей эллипсоида D' совпадает с характеристическими векторами матрицы $1/2 \partial^2 Q(\mathbf{a})/\partial \alpha^2$. Длина i -й полуоси эллипсоида D' равна $\sqrt{\varphi' Q(\mathbf{a}; \mathbf{y})/\lambda_i}$, где λ_i — характеристическое число матрицы $1/2 \partial^2 Q(\mathbf{a})/\partial \alpha^2$.

В некоторых случаях нелинейная регрессия после репараметризации превращается в линейную. Нелинейная регрессия (7.2) репараметризуема, если ее функция регрессии представима в виде

$$\hat{f}_t(\alpha) = g_1(\alpha) f_{t1} + g_2(\alpha) f_{t2} + \dots + g_m(\alpha) f_{tm}, \quad t=1, \dots, m,$$

где f_{ti} — константы, причем матрица, составленная из этих чисел, имеет ранг m ; $g_i(\alpha)$ — непрерывные взаимно-однозначные функции, отображающие R^m на R^m . Исходная нелинейная регрессия может быть репараметризована следующим образом:

$$y_t = \beta_1 f_{t1} + \beta_2 f_{t2} + \dots + \beta_m f_{tm} + \varepsilon_t,$$

где $\beta_i = g_i(\alpha)$ — новые параметры. Ясно, что оценка МНК нелинейной регрессии (7.2) равна $\mathbf{a} = \mathbf{g}^{-1}(\mathbf{b})$, где \mathbf{b} — оценка МНК соответствующей линейной регрессии, \mathbf{g}^{-1} — отображение, обратное к $\mathbf{g} = (g_1, \dots, g_m)$. Х. Хартли [122] справедливо замечает, что если D — доверительное множество с коэффициентом доверия $1 - \lambda$ для параметров β_1, \dots, β_m , то доверительное множество $\mathbf{g}^{-1}(D)$ для параметров исходной регрессии $\alpha_1, \alpha_2, \dots, \alpha_m$ имеет тот же коэффициент доверия.

Коротко остановимся на вопросе оценивания нелинейной регрессии в случае, когда ковариационная матрица отклонений имеет общий вид и известна с точностью до постоянного множителя. Допустим, в регрессии (7.2) $\text{cov}(\mathbf{e}) = \sigma^2 \mathbf{\Omega}$, где $\sigma^2 > 0$ — неизвестный параметр, а $\mathbf{\Omega}$ — известная весовая матрица $n \times n$, $|\mathbf{\Omega}| \neq 0$. Так же, как в случае линейной регрессии, обобщенная оценка МНК минимизирует взвешенную сумму квадратов отклонений $Q(\alpha) = (\mathbf{y} - \mathbf{f}(\alpha))' \mathbf{\Omega}^{-1} (\mathbf{y} - \mathbf{f}(\alpha))$. Пусть \mathbf{T} — такая невырожденная матрица $n \times n$, что $\mathbf{T}'\mathbf{T} = \mathbf{\Omega}$, тогда исходная регрессия преобразуется в новую нелинейную регрессию, у которой ковариационная матрица отклонений пропорциональна единичной. Действительно, положим $\mathbf{v} = \mathbf{T}^{-1} \mathbf{y}$, $\boldsymbol{\psi}(\alpha) = \mathbf{T}^{-1} \mathbf{f}(\alpha)$, тогда в нелинейной регрессии $\mathbf{v} = \boldsymbol{\psi}(\alpha) + \boldsymbol{\xi}$ $\text{cov}(\boldsymbol{\xi}) = \sigma^2 \mathbf{I}$. Таким образом, случай $\text{cov}(\mathbf{e}) = \sigma^2 \mathbf{\Omega}$ с известной $\mathbf{\Omega}$ практически не отличается от обычного предположения $\text{cov}(\mathbf{e}) = \sigma^2 \mathbf{I}$.

В литературе рассмотрен также случай, когда ε_t имеют стационарное распределение. Тогда $\mathbf{\Omega}$ имеет простую структуру и возможно ее оценивание. В простейшем случае отклонения имеют автокорреляцию первого порядка:

$\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t$. При некоторых условиях регулярности обобщенная оценка МНК также будет состоятельной и асимптотически-нормальной (более подробно см. [120, 178, 106]).

8.4. Псевдонезависимые нелинейные регрессии

Системой псевдонезависимых нелинейных регрессий называется совокупность нелинейных регрессий

$$y_i = f^i(\alpha) + \varepsilon_i, \quad i = 1, \dots, k, \quad (8.34)$$

где y_i — вектор зависимой переменной $n \times 1$; $f^i(\alpha)$ — векторная функция, отображающая R^m в R^n ; $\alpha \in R^m$ — общий вектор неизвестных параметров; ε_i — случайный вектор отклонений $n \times 1$.

Употребление термина «псевдонезависимые» объясняется следующими предположениями, накладываемыми на систему (8.34):

а) $E \varepsilon_{ti} \varepsilon_{tj} = \sigma^2 \omega_{ij}$; $E \varepsilon_{ti} \varepsilon_{\tau j} = 0$ ($\tau \neq t$), где $E \varepsilon_i = 0$. Другими словами, если через ε^t обозначить вектор отклонений, i -я координата которого соответствует i -му уравнению t -го наблюдения, то $\text{cov}(\varepsilon^t) = \sigma^2 \Omega$, $|\Omega| \neq 0$;

б) неизвестный вектор параметров является общим для всех уравнений системы (8.34). При рассмотрении линейных псевдонезависимых регрессий предполагалось, что вектор α дизъюнктивен (множества неизвестных параметров разных уравнений не пересекались друг с другом).

Система (8.34) легко может быть сведена к одной нелинейной регрессии. Действительно, обозначим

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad f(\alpha) = \begin{bmatrix} f^1(\alpha) \\ f^2(\alpha) \\ \vdots \\ f^n(\alpha) \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

— векторы порядка $nk \times 1$. Тогда (8.34) переписывается следующим образом: $y = f(\alpha) + \varepsilon$, где, как легко показать, $\text{cov}(\varepsilon) = \sigma_0^2 (\Omega \otimes I_m)$ (см. параграф 2.5). Обозначим далее y^t — вектор-столбец, i -я координата которого соответствует i -му уравнению системы (8.34), которая в свою очередь отвечает t -му наблюдению. Аналогично введем вектор-функцию $f_t(\alpha)$ и ε^t . Тогда (8.34) переписывается следующим образом: $y_t = f^t(\alpha) + \varepsilon_t$, $t = 1, 2, \dots, n$, причем $\text{cov}(\varepsilon_t) = \sigma^2 \Omega$. При известной матрице Ω взвешенный МНК

(см. параграф 2.1) приводит нас к минимизации следующего выражения:

$$Q(\alpha) = \sum_{t=1}^n (y_t - f^t(\alpha))' \Omega^{-1} (y_t - f^t(\alpha)). \quad (8.35)$$

Пусть T — такая невырожденная матрица $k \times k$, что $\Omega = TT'$ (см. параграф 2.5). Тогда, обозначая

$$v_t = T^{-1} y_t; \quad \varphi^t(\alpha) = T^{-1} f^t(\alpha); \quad \xi_t = T^{-1} \varepsilon_t, \quad t = 1, \dots, n,$$

приходим к регрессии

$$v = \varphi(\alpha) + \xi, \quad (8.36)$$

где v — вектор-столбец $nk \times 1$, составленный из v^t ; $\varphi(\alpha)$ — вектор-функция, отображающая R^m в R^{nk} , составленная из $\varphi^t(\alpha)$; ξ — вектор порядка $nk \times 1$, составленный аналогично v . Важно отметить, что $\text{cov}(\xi) = \sigma^2 I_{nk}$, и поэтому все методы, разработанные в гл. 7 и 8, применимы к регрессии (8.36). В частности, нетрудно показать, что если для каждого уравнения (8.34) имеют место условия теорем 8.1, 8.2 и 8.3, то оценка взвешенного МНК, минимизирующая (8.35) или сумму квадратов отклонений регрессии (8.36), будет непрерывной, состоятельной, асимптотически-нормальной.

До сих пор речь шла о ситуации, когда Ω известна. Этот случай имеет скорее теоретическую ценность, чем практическую.

Допустим, Ω — неизвестная матрица. В этом случае возможны два пути. Первый — оценить матрицу Ω на основе (8.34) и использовать ее при минимизации (8.35). Полученную оценку будем называть оценкой Zellнера¹. Процесс можно продолжить: на основе оценки Zellнера оценить матрицу Ω , найти новую оценку и т. д. Такие оценки так же, как и в линейной регрессии, будем называть итеративными оценками Zellнера. Если число итераций оценивания Ω равно 1, итеративная оценка превращается в оценку Zellнера. В условиях теоремы 8.2 можно доказать состоятельность итеративной оценки Zellнера. Накладывая на регрессию (8.34) условия, аналогичные условиям теоремы 8.3, можно показать, что итеративные оценки Zellнера асимптотически нормальны [48, с. 106].

¹ Иногда эту оценку называют оценкой по минимальному расстоянию [48].

Перейдем ко второму пути оценивания системы (8.34.) Предположим, что ε_t нормально распределены. Тогда минимизируя функцию плотности вектора y , найдем оценку метода максимального правдоподобия. Можно использовать эту оценку, даже если отклонения распределены не нормально. Такие оценки называются *оценками метода квазimaxимального правдоподобия* (МКМП). В [172] доказано, что в условиях регулярности типа (8.4) оценка МКМП является состоятельной оценкой α . Филлипс также показывает, что при некоторых условиях итеративная оценка Зеллера для больших n устойчива (т. е. сходится) и ее предел равен оценке МКМП. Барнетт [80] также устанавливает сходимость оценки МКМП, ее асимптотическую нормальность и эффективность.

8.5. Доказательства

1. Доказательство теоремы 8.2. Рассмотрим следующую последовательность случайных величин:

$$\begin{aligned} \frac{1}{n} Q_n(\alpha; y) &= \frac{1}{n} \sum_t (y_t - f_t(\alpha))^2 = \\ &= \frac{1}{n} \sum_t \varepsilon_t [f_t(\alpha_0) - f_t(\alpha)] + \frac{1}{n} \sum_t \varepsilon_t^2 + \\ &\quad + \frac{1}{n} \sum_t [f_t(\alpha_0) - f_t(\alpha)]^2. \end{aligned} \quad (8.37)$$

Существование предела (8.4) ведет к тому, что первое слагаемое правой части выражения (8.37) по закону больших чисел стремится к нулю с вероятностью, равной 1. В силу независимости и одинаковой распределенности $\varepsilon_1, \varepsilon_2, \dots$ предел второго слагаемого (8.37) равен σ^2 . Таким образом,

$$\lim_{n \rightarrow \infty} \frac{1}{n} Q_n(\alpha; y) = \sigma^2 + \varphi(\alpha, \alpha_0) \quad (8.38)$$

с вероятностью 1.

В силу компактности Θ для каждой последовательности наблюдений y_1, y_2, \dots существует предельная точка последовательности оценок МНК $a_{n_0}(y), a_{n_1}(y), \dots$, которую обозначим α^* . По определению оценки МНК

$$\frac{1}{n} Q_n(a_n(y); y) \leq \frac{1}{n} Q_n(\alpha_0; y). \quad (8.39)$$

Воспользуемся следующим результатом Уилкса [63, с. 116]. Пусть $X_n(\mathbf{c})$ — случайная величина, зависящая от параметра \mathbf{c} . Предположим, что равномерно по \mathbf{c} $X_n(\mathbf{c}) \rightarrow g(\mathbf{c})$ по вероятности при $n \rightarrow \infty$, где $g(\mathbf{c})$ — непрерывная функция. Тогда, если $\text{plim } Y_n = \mathbf{c}_0$, то $\text{plim } X_n(Y_n) = g(\mathbf{c}_0)$.

Так как сходимость (8.4) равномерна по α при фиксированном α_0 (таковой является и сходимость (8.38)), неравенство (8.39), как следует из предыдущего, будет верно и для предельной точки α^* :

$$\lim_{n \rightarrow \infty} \frac{1}{n} Q_n(\alpha^*; \mathbf{y}) \leq \lim_{n \rightarrow \infty} \frac{1}{n} Q_n(\alpha_0; \mathbf{y}),$$

или с учетом (8.38) $\sigma^2 + \varphi(\alpha^*, \alpha_0) \leq \sigma^2$, откуда $\varphi(\alpha^*, \alpha_0) \leq 0$, что влечет $\alpha_0 = \alpha^*$.

2. Доказательство теоремы 8.3 имеет много общего с доказательством асимптотической нормальности оценки максимального правдоподобия [63, с. 369—371].

Поскольку $\mathbf{a}_n(\mathbf{y})$ — внутренняя точка Θ п. н., то, разлагая в этой точке градиент суммы квадратов отклонений в ряд до линейных членов, получим

$$\frac{\partial Q_n(\alpha_0; \mathbf{y})}{\partial \alpha} = \frac{\partial Q_n(\alpha_n(\mathbf{y}); \mathbf{y})}{\partial \alpha} + (\alpha_0 - \mathbf{a}_n(\mathbf{y})) \frac{\partial^2 (Q_n(\gamma_n(\mathbf{y}); \mathbf{y}))}{\partial \alpha^2}, \quad (8.40)$$

где α_0 — истинное значение параметра; $\gamma_n(\mathbf{y})$ — случайная величина, удовлетворяющая неравенству

$$\|\mathbf{a}_n(\mathbf{y}) - \gamma_n(\mathbf{y})\| \leq \|\mathbf{a}_n(\mathbf{y}) - \alpha_0\|. \quad (8.41)$$

В силу того что $\mathbf{a}_n(\mathbf{y})$ — внутренняя точка, минимизирующая $Q_n(\alpha; \mathbf{y})$, градиент Q_n в этой точке равен нулю, т. е. первое слагаемое в правой части (8.40) равно нулю. Перепишем (8.40) следующим образом:

$$\begin{aligned} & \frac{1}{\sqrt{n}} \cdot \frac{\partial Q_n(\alpha_0; \mathbf{y})}{\partial \alpha} = \\ & = \sqrt{n}(\alpha_0 - \mathbf{a}_n(\mathbf{y})) \cdot \frac{1}{n} \cdot \frac{\partial^2 Q_n(\gamma_n(\mathbf{y}); \mathbf{y})}{\partial \alpha^2}. \end{aligned} \quad (8.42)$$

Легко показать, что случайный вектор

$$\frac{1}{\sqrt{n}} \cdot \frac{\partial Q_n(\alpha_0; y)}{\partial \alpha} = -\frac{2}{\sqrt{n}} \sum_t \varepsilon_t \frac{\partial f_t(\alpha_0)}{\partial \alpha}$$

имеет предельное распределение $N(0, 4 \sigma^2 \mathbf{A}(\alpha_0))$. Далее,

$$\begin{aligned} & \frac{1}{n} \cdot \frac{\partial^2 Q_n(\gamma_n(y); y)}{\partial \alpha^2} = \\ &= \frac{2}{n} \sum_t \left[\frac{\partial f_t(\gamma_n(y))}{\partial \alpha} \right] \left[\frac{\partial f_t(\gamma_n(y))}{\partial \alpha} \right]' - 2 \times \\ & \times \frac{1}{n} \sum_t (y_t - f_t(\gamma_n(y))) \frac{\partial^2 f_t(\gamma_n(y))}{\partial \alpha^2}. \end{aligned} \quad (8.43)$$

Воспользуемся опять результатом Уилкса [63, с. 116]. Из неравенства (8.41) следует, что $\text{plim } \gamma_n(y) = \alpha_0$. Поэтому в силу (8.7) и вышеизложенного результата первое слагаемое правой части выражения (8.43) сходится по вероятности к $2 \mathbf{A}(\alpha_0)$. Аналогично, используя (8.6), можно показать, что второе слагаемое (8.43) сходится по вероятности к нулю. Далее пользуемся следующим хорошо известным фактом (см., например, [58, с. 118]). Если ξ_n — случайный вектор $m \times 1$, \mathbf{A}_n — случайная матрица $m \times m$ и векторы $\mathbf{A}_n \xi_n$ имеют предельное распределение F , $\text{plim } \mathbf{A}_n = \mathbf{A}$ — детерминированная матрица, $|\mathbf{A}|_1^* \neq 0$, то вектор ξ_n имеет асимптотическое предельное распределение, совпадающее с распределением вектора $\mathbf{A}^{-1} \eta$, где вектор η имеет распределение F . Поэтому предельным распределением вектора $\sqrt{n} (\alpha_0 - \mathbf{a}_n(y))$ будет $N(0, \sigma^2 \mathbf{A}^{-1}(\alpha_0))$ — теорема доказана.

Приложение. Некоторые дополнительные формулы

П.1. *Разбиение матрицы на блоки. Формула Фробениуса.* Пусть A, B — прямоугольные матрицы одинакового порядка $m \times n$. Разобьем их одинаковым же образом на подматрицы

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}; \quad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

так что матрицы A_{11} и B_{11} имеют порядок $m_1 \times n_1$, матрицы A_{12} и B_{12} — $m_1 \times n_2$, матрицы A_{21} и B_{21} — $m_2 \times n_1$ и матрицы A_{22} и B_{22} — $m_2 \times n_2$ ($m_1 + m_2 = m$; $n_1 + n_2 = n$). Существуют простые формулы для блочного сложения и умножения матриц:

$$A + B = \begin{bmatrix} A_{11} + B_{11} & A_{12} + B_{12} \\ A_{21} + B_{21} & A_{22} + B_{22} \end{bmatrix};$$

$$AB = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}. \quad (\text{П.1})$$

При умножении блочных матриц можно руководствоваться правилом «строка на столбец».

Приведем формулы блочного обращения матрицы. Для простоты будем рассматривать только симметричные матрицы. Итак, пусть A — симметричная матрица, разбитая на блоки:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A'_{12} & A_{22} \end{bmatrix}.$$

Тогда

$$A^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A'_{12} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1}_{11} + FE^{-1}F' & -FE^{-1} \\ -E^{-1}F' & E^{-1} \end{bmatrix}, \quad (\text{П.2})$$

где $E = A_{22} - A'_{12}A^{-1}_{11}A_{12}$; $F = A^{-1}_{11}A_{12}$. Доказательство этого несложно, оно следует непосредственно из определения обратной матрицы.

Далее, существует формула для вычисления определителя блочной матрицы:

$$|A| = \begin{vmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{vmatrix} = |A_{11}| |A_{22} - A_{21}A^{-1}_{11}A_{12}|. \quad (\text{П.3})$$

Доказательство можно найти в [58, с. 44].

Нам понадобится следующая формула:

$$(A + BDB')^{-1} = A^{-1} - A^{-1}B(B'A^{-1}B + D^{-1})^{-1}B'A^{-1} \quad (\text{П.4})$$

где A и D — невырожденные матрицы $m \times m$ и $n \times n$ соответственно, B — прямоугольная матрица $m \times n$.

П.2. *Матричное дифференцирование.* Пусть A — прямоугольная матрица порядка $m \times n$, элементы которой являются функциями некоторой действительной переменной t , т. е. $a_{ij} = a_{ij}(t)$, $i = 1, \dots, m$, $j = 1, \dots, n$. По определению полагаем

$$\frac{dA}{dt} = \left[\frac{da_{ij}}{dt} \right].$$

Легко показать, что если $\mathbf{B}^{k \times m}$ не зависит от t , а $\mathbf{C}^{k \times m} = \mathbf{C}(t)$, то верны следующие формулы:

$$\frac{d\mathbf{B}\mathbf{A}}{dt} = \mathbf{B} \frac{d\mathbf{A}}{dt}; \quad \frac{d(\mathbf{C}\mathbf{A})}{dt} = \frac{d\mathbf{C}}{dt} \mathbf{A} + \mathbf{C} \frac{d\mathbf{A}}{dt}. \quad (\text{П.5})$$

Пусть $f = f(t_1, t_2, \dots, t_m)$ — функция m аргументов (дифференцируемая). Производной этой функции является вектор-столбец $m \times 1$

$$\frac{\partial f}{\partial t} = \left[\frac{\partial f}{\partial t_i} \right]_{i=1}^m.$$

Легко доказываются следующие формулы:

$$\frac{\partial \mathbf{x}' \mathbf{t}}{dt} = \mathbf{x}; \quad \frac{\partial \mathbf{t}' \mathbf{A} \mathbf{t}}{dt} = 2\mathbf{A} \mathbf{t}, \quad (\text{П.6})$$

где \mathbf{x} — постоянный вектор-столбец $m \times 1$; \mathbf{A} — симметричная матрица $m \times m$.

Теперь предположим, что f есть векторная функция или отображение, т. е. $f: R^m \rightarrow R^n$, тогда df/dt — прямоугольная матрица $n \times m$, (i, j) -й элемент которой по определению равен df_i/dt_j , $i = 1, \dots, n$, $j = 1, \dots, m$. В частности, если \mathbf{A} — постоянная матрица $k \times n$ и $\mathbf{y} = \mathbf{y}(\mathbf{x})$, $\mathbf{y}^{n \times 1}$, $\mathbf{x}^{m \times 1}$, то

$$\frac{\partial \mathbf{A} \mathbf{y}}{\partial \mathbf{x}} = \mathbf{A} \frac{\partial \mathbf{y}}{\partial \mathbf{x}}. \quad (\text{П.7})$$

Можно рассматривать дифференцирование по матрице. Пусть \mathbf{A} — матрица $m \times n$ и $f = f(\mathbf{A})$ — действительная функция матрицы. По определению $df/d\mathbf{A}$ есть матрица того же порядка, т. е. $m \times n$ и

$$\frac{\partial f}{\partial \mathbf{A}} = \left[\frac{\partial f}{\partial \mathbf{A}_{ij}} \right], \quad i = 1, \dots, m, \quad j = 1, \dots, n.$$

Легко проверить, например, что

$$\frac{\partial \mathbf{x}' \mathbf{A} \mathbf{y}}{\partial \mathbf{A}} = \mathbf{x} \mathbf{y}', \quad (\text{П.8})$$

де $\mathbf{x}^{m \times 1}$, $\mathbf{y}^{n \times 1}$.

Если \mathbf{S} — матрица $m \times m$ с положительным определителем, то нетрудно убедиться в том, что

$$\frac{\partial \ln |\mathbf{S}|}{\partial \mathbf{S}_{ij}} = (\mathbf{S}^{-1})_{ji}, \quad j, i = 1, \dots, m. \quad (\text{П.9})$$

П. 3. *Выпуклые функции и оптимизация.* Допустим, $f = f(\mathbf{x}) = f(x_1, \dots, x_m)$ есть дифференцируемая функция на R^m . Нас интересует глобальный минимум этой функции. Хорошо известно необходимое условие минимума функции: если \mathbf{x}^* — минимум (может быть и локальный) функции f , то

$$\frac{\partial f(\mathbf{x}^*)}{\partial \mathbf{x}} = 0. \quad (\text{П.10})$$

Строго выпуклой вниз функцией называется функция, для которой при любых $0 < \alpha < 1$:

$$f(\alpha x_1 + (1-\alpha)x_2) < \alpha f(x_1) + (1-\alpha)f(x_2)$$

для всех $x_1 \neq x_2 \in R^m$. Класс строго выпуклых вниз функций важен тем, что любой локальный минимум является и глобальным. Если f — дважды дифференцируемая функция, то *выпуклость вниз следует из положительной определенности матрицы вторых производных (гессиан функции)* $\frac{\partial^2 f}{\partial x^2}$. Исследование выпуклости функции можно свести к исследованию выпуклости функции одной переменной. Пусть $x_1, x_2 \in R^m$ — любые, определим новую функцию одной переменной $F(\alpha) = f(x_1 + \alpha x_2)$. Функция f является строго выпуклой вниз функцией тогда и только тогда, когда F — строго выпуклая вниз функция.

П.4. *Характеристические числа и векторы.* Для симметричных неотрицательно определенных матриц часто максимальное х. ч. обозначаем λ_{\max} , минимальное — λ_{\min} . В книге неоднократно используется следующий факт. Пусть A — симметричная матрица $m \times m$, B — прямоугольная матрица $m \times n$, тогда

$$\lambda_{\max}(B'AB) \leq \lambda_{\max}(A) \lambda_{\max}(B'B); \quad (\text{П.11})$$

$$\lambda_{\min}(B'AB) \geq \lambda_{\min}(A) \lambda_{\min}(B'B). \quad (\text{П.12})$$

Докажем сначала (П.11). Пусть $v \in R^m$, $\|v\| = 1$, тогда

$$\begin{aligned} v' B' A B v &= (Bv)' A (Bv) \leq \lambda_{\max}(A) \|Bv\|^2 = \\ &= \lambda_{\max}(A) v' B' B v \leq \lambda_{\max}(A) \cdot \lambda_{\max}(B'B), \end{aligned}$$

откуда и следует (П.11). Аналогично доказывается (П.12):

$$\begin{aligned} v' B' A B v &= (Bv)' A (Bv) \geq \lambda_{\min}(A) \|Bv\|^2 = \\ &= \lambda_{\min}(A) v' B' B v \geq \lambda_{\min}(A) \cdot \lambda_{\min}(B'B). \end{aligned}$$

П.5. *Случайные квадратичные формы.* Пусть d — случайный вектор, A — детерминированная симметричная матрица. Рассмотрим случайную квадратичную форму $d'Ad$. Предположим, $\text{cov}(d) = \sigma^2 I$, $E d = 0$ (I — единичная матрица), тогда

$$E d' A d = \sigma^2 \text{tr}(A). \quad (\text{П.13})$$

Действительно,

$$E d' A d = E \sum_{i,j} d_i d_j A_{ij} = \sum_{i,j} (E d_i d_j) A_{ij} = \sigma^2 \sum_i A_{ii} = \sigma^2 \text{tr}(A).$$

Допустим, случайный вектор d имеет многомерное нормальное распределение с нулевым математическим ожиданием; компоненты вектора считаем некоррелируемыми, имеющими единичную дисперсию. Коротко это может быть записано как $d \sim N(0, I)$. Тогда квадратичная форма $d'Ad$ имеет χ^2 -распределение с k степенями свободы тогда и только тогда, когда A — идемпотентная матрица ($A^2 = A$) и $\text{rank } A = \text{tr } A = k$:

$$\begin{aligned} d'Ad \sim \chi^2(k) \Leftrightarrow A \text{ — идемпотентная матрица, } \text{rank } A = \\ = \text{tr } A = k. \end{aligned} \quad (\text{П.14})$$

Далее, пусть \mathbf{B} — детерминированная матрица и $\mathbf{d} \sim N(\mu, \sigma^2 \mathbf{I})$. Тогда

$$\mathbf{BA} = \mathbf{0} \Rightarrow \mathbf{Bd} \text{ и } \mathbf{Ad}'\mathbf{d} \text{ независимы.} \quad (\text{П. 15})$$

Пусть по-прежнему $\mathbf{d} \sim N(\mathbf{0}, \mathbf{I})$, тогда две квадратичные формы $\mathbf{d}'\mathbf{Ad}$ и $\mathbf{d}'\mathbf{Cd}$ независимы тогда и только тогда, когда $\mathbf{AC} = \mathbf{0}$:

$$\mathbf{d}'\mathbf{Ad} \text{ и } \mathbf{d}'\mathbf{Cd} \text{ независимы} \Leftrightarrow \mathbf{AC} = \mathbf{0}. \quad (\text{П. 16})$$

Доказательство (П.13) — (П.16) можно найти в [115].

П.6. *Неравенства, связывающие элементы положительно определенной матрицы с ее максимальным характеристическим числом.* Для любой положительно определенной матрицы \mathbf{D} имеют место следующие неравенства:

$$\max_{i, j} \{ |D_{ij}| \} \leq \lambda_{\max}(\mathbf{D}) \leq \sum_{i, j} |D_{ij}|. \quad (\text{П. 17})$$

Пусть \mathbf{z} — вектор единичной длины. Тогда

$$\mathbf{z}'\mathbf{Dz} = \sum_{i, j} z_i z_j D_{ij} \leq \sum_{i, j} |z_i z_j D_{ij}| \leq \sum_{i, j} |D_{ij}|,$$

поскольку $\mathbf{z}'\mathbf{z} = 1$ влечет $|z_i| \leq 1$ для всех i . Таким образом, правая часть неравенства (П.17) доказана.

Для доказательства левой части неравенства (П.17) представим \mathbf{D} в виде произведения $\mathbf{T}'\mathbf{T}$, где \mathbf{T} — невырожденная матрица. Применяя неравенство Коши, получим $|D_{ij}| \leq \sqrt{D_{ii} \cdot D_{jj}}$ или $|D_{ij}| \leq \leq \max \{ D_{ii}, D_{jj} \}$. Далее легко видеть, что $\lambda_{\max}(\mathbf{D}) \geq D_{ii}$. С учетом этого неравенства получаем $\lambda_{\max}(\mathbf{D}) \geq \max_{i, j} \{ |D_{ij}| \}$.

Список использованной литературы

1. Айвазян С. А. Статистическое исследование зависимостей. М., Металлургия, 1968.
2. Айвазян С. А., Богдановский И. М. Методы статистического исследования парных зависимостей в схемах конъюнктного анализа и их применения. — Заводская лаборатория, 1974, т. 40, № 3.
3. Айвазян С. А., Розанов Ю. А. Некоторые замечания к асимптотически эффективным линейным оценкам коэффициентов регрессии. — Труды математического института им. В. А. Стеклова. М., Наука, 1964, т. 71.
4. Алберт А. Регрессия, псевдорегрессия и рекуррентное оценивание. М., Наука, 1977.
5. Андерсон Т. Введение в многомерный статистический анализ. М., Физматгиз, 1963.
6. Андерсон Т. Статистический анализ временных рядов. М., Мир, 1976.
7. Араманович И. Г. и др. Математический анализ. Дифференцирование и интегрирование. М., Гос. изд-во физ.-мат. лит., 1961.
8. Бард Я. Нелинейное оценивание параметров. М., Статистика, 1979.
9. Беллман Р. Введение в теорию матриц. М., Наука, 1969.
10. Бородкин Ф. М. Статистическая оценка связей экономических показателей, М., Статистика, 1968.
11. Брандт З. Статистические методы анализа наблюдений. М., Мир, 1975.
12. Браун М. Теория и измерение технического прогресса. М., Статистика, 1971.
13. Вапник В. Н. Восстановление зависимостей по эмпирическим данным. М., Наука, 1979.
14. Варыгин В. Н.; Казарян С. А.; Рафаелян Р. С. Выбор начального вектора при адаптивных методах построения моделей, нелинейных по параметрам. — Автоматика и телемеханика, 1978, № 3.
15. Васильев Ф. П. Лекции по методам решения экстремальных задач. М., изд-во МГУ, 1974.
16. Венецкий И. Г.; Венецкая В. И. Основные математико-статистические понятия и формулы в экономическом анализе. М., Статистика, 1980.
17. Гавурин М. К.; Фарфоровская Ю. Б. Об одном итеративном методе разыскания суммы квадратов отклонений. — Журнал вычислительной математики и мат. физики, 1966, т. 6, № 6.

18. Гнеденко Б. В. Курс теории вероятностей. М., Физматгиз, 1961.

19. Головач В. А.; Ерина А. М., Трофимов В. П. Критерии математической статистики в экономических исследованиях. М., Статистика, 1973.

20. Демиденко Е. З. Регрессионный анализ в случае стохастической матрицы независимых переменных.— В кн.: Математические методы в экономике и международных отношениях. М., изд. ИМЭМО АН СССР, 1974. вып. 3.

21. Демиденко Е. З. О некоторых вопросах применения регрессионного анализа к временным экономическим рядам.— В кн.: Математические методы в экономике и международных отношениях. М., изд. ИМЭМО АН СССР, 1974, вып. 3.

22. Демиденко Е. З. Нелинейная регрессия.— В кн.: Математические методы решения экономических задач. М., Наука, 1977.

23. Демиденко Е. З. Псевдонезависимые регрессии и их оценивание.— В кн.: Малоразмерные модели экономического роста. М.; изд. ИМЭМО АН СССР, 1978.

24. Демиденко Е. З. Идентификация линейных эконометрических моделей.— Экономика и математические методы, 1978, т. 14 вып. 6.

25. Демиденко Е. З. Линейная и нелинейная регрессия. Фортран IV. М., изд. ИМЭМО АН СССР, 1979.

26. Джонстон Дж. Эконометрические методы. М.; Статистика, 1980.

27. Демидович Б. П.; Марон И. А. Основы вычислительной математики. М.; Наука, 1970.

28. Дрейпер Н.; Смит Г. Прикладной регрессионный анализ. М., Статистика, 1973.

29. Дружинин Н. К. Математическая статистика в экономике. М., Статистика, 1971.

30. Дружинин Н. К. Логика оценки статистических гипотез. М., Статистика, 1973.

31. Епишин Ю. Г. Об оценках параметров регрессии по методу наименьших абсолютных отклонений.— Экономика и математические методы, 1974, т. 10, вып. 5.

32. Ершов А. А. Стабильные методы оценки параметров (обзор). — Автоматика и телемеханика, 1978, № 8.

33. Закс Ш. Теория статистических выводов. М.; Мир, 1975.

34. Иванов А. В. Состоятельность оценок нелинейной регрессии.— Теория вероятностей и математическая статистика. Киев, 1972, вып. 6.

35. Иванов А. В. Асимптотическое поведение оценок НК в случае нелинейной регрессии. — Теория вероятностей и математическая статистика. Киев, 1972, вып. 7.

36. Иванов А. В. Асимптотическое разложение для распределения оценки наименьших квадратов параметра нелинейной регрессии. — Теория вероятностей и ее применения, 1976, т. 21, № 3.

37. Кендэлл М. Дж. Стьюарт А. Статистические выводы и связи. М., Наука, 1973.
38. Колмаев В. А. Регрессионный анализ в случае схемы с двумя дисперсиями.— Экономика и математические методы, 1971, т. 7, вып. 1.
39. Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. М., Наука, 1972.
40. Колмогоров А. Н. Несмещенные оценки.— Изв. АН СССР. Сер. мат., 1950, № 4.
41. Кошеев В. А. Метод учета априорной информации в линейном оценивании параметров. М., Наука, 1978.
42. Крамер Г. Математические методы статистики. М., Наука, 1975.
43. Крастинь О. П. Методы анализа регрессий и корреляций при определении агроэкономических функций. Рига, 1970.
44. Кузнецов С. Е. Статистический анализ моделей динамики поведения планируемых экономических показателей.— В кн.: Прикладной многомерный статистический анализ. М., Наука, 1978.
45. Леман Э. Проверка статистических гипотез. М., Наука, 1979.
46. Лизер С. Эконометрические методы и задачи. М., Статистика, 1971.
47. Линник Ю. В. Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. М., Физматгиз, 1962.
48. Маленко Э. Статистические методы эконометрии. М., Статистика, 1975, вып. 1; 1976, вып. 2.
49. Меклер С. Г.; Павлов Н. В. Метод максимальной окрестности для оценки параметров нелинейной функции по способу наименьших квадратов.— Экономика и математические методы, 1974, т. 10, вып. 1.
50. Мироновский Л. А.; Юдович В. С. Об одном подходе к идентификации линейных стационарных объектов.— Автоматика и телемеханика, 1972, № 1.
51. Мудров В. И.; Кушко В. Л. Методы обработки измерений. М., Сов. радио, 1976.
52. Орлов А. И. Предельное распределение одной оценки базисных функций в регрессии.— В кн.: Прикладной многомерный статистический анализ. М., Наука, 1978.
53. Островский А. М. Решение уравнений и систем уравнений. М., ИЛ, 1963.
54. Перегудов В. Н. Метод наименьших квадратов и его применение в исследованиях. М., Статистика, 1965.
55. Полак Э. Численные методы оптимизации. М., Мир, 1974.
56. Поляк Б. Т.; Цыпкин Я. З. Помехоустойчивая идентификация.— Труды IV симпозиума ИФАК «Идентификация и оценка параметров систем». Тбилиси, 1976, ч. 1.
57. Пшеничный Б. Н., Данилин Ю. М. Численные методы в экстремальных задачах. М., Наука, 1975.

58. Рао С. Р. Линейные статистические методы и их применение. М.; Наука, 1968.
59. Романовский В. И. Математическая статистика. Ташкент, 1963, кн. 2.
60. Сирл С., Госман У. Матричная алгебра в экономике. М., Статистика, 1974.
61. Гейл Г. Экономические прогнозы и принятие решений. М., Статистика, 1971.
62. Смоляк С. А., Титаренко Б. П. Устойчивые методы оценивания. М., Статистика, 1980.
63. Уилкс С. Математическая статистика. М., Наука, 1967.
64. Хеннан Э. Многомерные временные ряды. М., Мир, 1974.
65. Холево А. С. Об оценках коэффициентов регрессии.— Теория вероятностей и ее применения, 1969, т. 14, № 1.
66. Холево А. С. Об асимптотической нормальности оценок коэффициентов регрессии.— Теория вероятностей и ее применения, 1971, т. 16, № 4.
67. Худсон Д. Статистика для физиков. М., Мир, 1970.
68. Форсайт Дж., Молер К. Численное решение систем линейных алгебраических уравнений. М., Мир., 1969.
69. Четыркин Е. М. Статистические методы прогнозирования. М., Статистика, 1977.
70. Adichie J. N. Estimation of regression parameters based on rank tests.— AMS, 1967, v. 38, № 4.
71. Aitken A. C. On least squares and linear combination of observation. — Proc. of the Royal Society, 1936, v. 55, № 1.
72. Anderson T. W. Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics.— JRSS, ser. B, 1977, v. 38, № 1.
73. Andrews D. F. A robust method for multiple linear regression. — Technometrics, 1974, v. 16, № 4.
74. Andrews D. F. et al. Robust estimates of location. Princeton U. P., 1972.
75. Anscombe F. I. Examination of residuals.— Proc. 4-th Berkley Symp. Math. Stat. Prob., 1961, v. 1, California Press.
76. Anscombe F. I., Tukey I. W. Analysis of residuals. — Technometrics, 1963, v. 5, № 1.
77. Armstrong R. D.; Frome E. L. A comparison of two algorithm for absolute deviation curve fitting.— JASA, 1976, v. 71, № 354.
78. Bard Y. Comparison on gradient methods for the solution of nonlinear parameter estimation problems.— SIAM Journal on Numerical Anal., 1970, v. 7, № 1.
79. Barham R. H.; Drane W. An algorithm for least squares estimation of nonlinear parameters when some of the parameters are linear.— Technometrics, 1972, v. 14, № 3.
80. Barnett W. A. Maximum likelihood and iterated Aitken estimation of nonlinear systems of equations. — JASA, 1976, v. 71, № 354.

81. Bartlett M. S. The fitting of straight lines if both variables are subject to error.— *Biometrics*, 1949, v. 5, № 2.
82. Beach C. M.; MacKinnon J. G. A maximum likelihood procedure for regression with autocorrelated errors.— *Econometrica*, 1978, v. 46, № 1.
83. Beale E. M. L. Confidence regions in nonlinear estimation.— *JRSS*, ser. B.; 1960, v. 22, № 1.
84. Beaton A. E.; Rubin D. B.; Barone J. L. The acceptability of regression solutions: another look at computational accuracy.— *JASA*, 1976, v. 71, № 353.
85. Bhattacharya P. K. Some properties of the least squares estimator in regression analysis when the predictor variables are stochastic.— *AMS*, 1963, v. 34, № 4.
86. Bickel P. I. Using residuals robustly I: tests for heteroskedasticity, nonlinearity.— *AS*, 1978, v. 6, № 2.
87. Brown K. M., Dennis J. E. Jr. Derivative-free analogues of the Levenberg — Marquardt and Gauss algorithm for nonlinear least squares approximation.— *Numerische Mathematik*, 1972, v. 18, № 2.
88. Box M. J. Bias in nonlinear estimation.— *JRSS*, ser B.; 1971, v. 33, № 2.
89. Box G. E. P., Cox D. R. An analysis of transformation.— *JRSS*, ser B, 1964, v. 26, № 2.
90. Casson M. C. Generalised errors in variables regression.— *Review of Economic Studies*, 1974, v., 41, № 127.
91. Chambers J. M. Fitting nonlinear models: numerical techniques.— *Biometrika*, 1973, v. 60, № 1.
92. Chipman J. S.; Rao M. M. The treatment of linear restrictions in regression analysis.— *Econometrica*, 1964, v. 32, № 1—2.
93. Daniel C., Wood F. S. Fitting equation to data. N.-Y. J. Willey, 1971.
94. Dempster A. P., Schatsoff M.; Wermuth N. A simulation study of alternatives to ordinary least squares.— *JASA*, 1977, v. 72, № 357.
95. Dennis J. E. Jr. Some computational techniques for the nonlinear least squares problem.— In: *Numerical solution of systems*. N. Y., 1973.
96. Dhrymes P. J. *Econometrics*. Harper and Row. N. Y., 1970.
97. Eicker F. Asymptotic normality and consistency of the least squares estimators for families of linear regressions.— *AMS*, 1963, v. 34, № 2.
98. Farebrother R. W. Further results on the mean square error of ridge regression.— *JRSS*, ser B, 1976, v. 38, № 3.
99. Farebrother R. W. The minimum mean square error linear estimator and ridge regression.— *Technometrics*, 1975, v. 17, № 1.

100. Farrar D. E.; Glauber R. R. Multicollinearity in regression analysis. The problem revisited.— *The Review of Economics and Statistics*, 1967, v. 49, № 1.

101. Feldstein M. Errors in variables: a consistent estimator with smaller MSE in finite sample.— *JASA*, 1974, v. 69, № 348.

102. Fletcher R., Grant J. A.; Heblen H. D. The calculation of linear lest L_p -approximations.— *Computer Journal*, 1971, v. 14, № 3.

103. Forsythe A. B. Robust estimation of straight line regression coefficients by minimising p -th power deviations.— *Technometrics*, 1972, v. 14, № 1.

104. Callant A. R. The Power of the likelihood ratio test of location in nonlinear regression models.— *JASA*, 1975, v. 70, № 349.

105. Gallant A. R. Nonlinear regression.— *The American Statistician*, 1975, v. 29, № 2.

106. Gallant A. R., Goebel J. J. Nonlinear regression with autocorrelated errors.— *JASA*, 1976, v. 71, № 356.

107. Gallant A. R. Seemingly unrelated nonlinear regressions.— *Journal of Econometrics*, 1975, v. 3, № 1.

108. Gallant A. R. Testing a subset of the parameters of a nonlinear regression model.— *JASA*, 1975, v. 70, № 352.

109. Gallant A. R. Testing a nonlinear regression specification: a nonregular case.— *JASA*, 1977, v. 72, № 359.

110. Gibson W. M., Jowett G. H. Three-group regression analysis, part I, simple regression analysis.— *Applied Statistics*, 1957, v. 6, № 1.

111. Glejser H. A new test for heteroskedasticity.— *JASA*, 1969, v. 64, № 325.

112. Goldfeld S. M.; Quandt R. E. Nonlinear methods in econometrics, Amsterdam, North-Holland, 1972.

113. Goldstein M.; Smith A. F. M. Ridge-type estimators for regression analysis.— *JRSS*, ser B, 1974, v. 36, № 2.

114. Gorman J. W., Tomnan R. J. Selection of variables for fitting equations to data.— *Technometrics*, 1966, v. 8, № 1.

115. Graybill F. Introduction to Linear Statistical Methods, N. Y. McGraw-Hill, 1961.

116. Greenberg E. Minimum variance properties of principal component regression.— *JASA*, 1975, v. 70, № 349.

117. Gunst R. F., Webster J. T., Mason R. L. A comparison of least squares and latent root regression estimators.— *Technometrics*, 1976, v. 18, № 1.

118. Halperin M. Fitting of straight lines and prediction when both variables are subject to error.— *JASA*, 1961, v. 56, № 295.

119. Halperin M. Confidence interval estimation in nonlinear regression.— *JRSS*, ser B, 1963, v. 25, № 2.

120. Hannan E. J. Non-linear series regression.— *Journal of Appl. Probab.*, 1971, v. 8, № 3.

121. Hartley H. O. The modified Gauss-Newton method for the fitting of non-linear regression function by least squares.— *Technometrics*, 1961, v. 3, № 2.
122. Hartley H.O. Exact confidence regions for the parameters in non-linear regression laws.— *Biometrika*, 1964, v. 51, № 384.
123. Hartley H. O.; Booker A. Nonlinear least squares estimation.— *AMS*, 1965, v. 36, № 2.
124. Hawkins D. M. On the investigation of alternative regressions by principal component analysis.— *Applied Statistics*, 1973, v. 22 № 3.
125. Hawkins D. M. Relations between ridge regression and eigenanalysis of the augmented correlation matrix.— *Technometrics*, 1975, v. 17, № 4.
126. Hemmerle W. J. An explicit solution for generalised ridge regression.— *Technometrics*, 1975, v. 17, № 3.
127. Hocking R. R.; Speed F. M.; Lynn M. J. A class of biased estimators in linear regression.— *Technometrics*, 1976, v. 18, № 4.
128. Hodges J. L. Jr. Lehman F. L. Estimates of location based on rank tests.— *AMS*, 1963, v. 34, № 3.
129. Hodges J. L. Jr, Lehman E. L. Some problems in minimax point estimation.— *AMS*, 1950, v. 21, № 2.
130. Hoerl A. E. Application of ridge analysis to regression problems.— *Chemical Engineering Progress*, 1962, v. 58, № 1.
131. Hoerl A. E.; Kennard R. W. Ridge regression: application to nonorthogonal problems.— *Technometrics*, 1970, v. 12, № 1.
132. Hoerl A. E.; Kennard R. W. Ridge regression: biased estimation for nonorthogonal problems.— *Technometrics*, 1970, v. 12, № 1.
133. Hoerl A. E., Kennard R. W. A note on a power generalised of ridge regressions.— *Technometrics*, 1975, v. 17, № 2.
134. Hoerl A. E.; Kennard R. W., Baldwin K. F. Ridge regression: some simulation.— *Communications in Statistics*; 1975, v. 4, № 2.
135. Hogan W. W. Norm minimisation and unbiasedness.— *Econometrica*, 1976, v. 44, № 3.
136. Huber P. J. Robust estimation of a location parameters.— *AMS*, 1964, v. 35, № 1.
137. Huber P. J. Robust regression: asymptotic, conjectures and Monte-Carlo.— *AS*, 1973, v. 1, № 5.
138. Jackel L. B. Robust estimation of location: symmetry and asymptotic contamination.— *AMS*, 1971, v. 42, № 4.
139. James W., Stein C. Estimation with quadratic loss.— *Proceedings of the 4-th Berkley simp. UCP*, 1961.
140. Jennrich R. I. Asymptotic properties of nonlinear least squares estimation.— *AMS*, 1969, v. 40, № 2.
141. Jurčhová J. Asymptotic linearity of a rank statistic in regression parameters.— *AMS*, 1969, v. 40, № 6.

142. K a k w a n i N. C. The unbiasedness of Zellner's seemingly unrelated regression equations estimators.— JASA, 1967, v. 62; № 317.

143. K a r t n i E.; W e i s s m a n I. A consistent estimator of the slope in regression model with errors in variables.— JASA; 1974, v. 69, № 345.

144. K e l l y J. E. An application of linear programming to curve fitting.— Journal of the Society for Industrial and Appl. Mathematics, 1958, v. 6 № 1.

145. K e n d a l l M. G. A course in multivariate analysis, London, Griffin, 1957.

146. K l o e k T. Note on consistent estimation of the variance of the disturbances in the linear model.— Econometrica, 1972, v. 40, № 5.

147. K m e n t a J.; G i l b e r t R. F. Small sample properties of alternative estimators of seemingly unrelated regressions. — JASA; 1968, v. 63, № 324. -

148. K u b i c e k M.; M a r e k M.; E c k e r t E. Quasilinearised regression.— Technometrics, 1971, v. 13, № 3.

149. L a w t o n W. H.; S y l v e s t r e E. A. Elimination of linear parameters in nonlinear regression.— Technometrics, 1971; v. 13, № 3.

150. L e a m e r E. E. Least-squares versus instrumental variables estimation in a simple error in variables model.— Econometrica, 1978, v. 46, № 4.

151. L e e c h D. Testing the error specification in nonlinear regression. — Econometrica, 1975, v. 43, № 4.

152. L e v e n b e r g K. A method for the solution of certain non-linear problems in least squares. — Quarterly of Applied Mathematics, 1944, v. 2, № 2.

153. L i e w C. K. Inequality constrained least squares estimation.— JASA, 1976, v. 71, № 355.

154. L o n g l e y J. W. An appraisal of least-squares programs for the electronic computer from the point of view the user. — JASA, 1967, v. 72, № 3.

155. M a d a n s k y A. The fitting of straight lines when both variables a subject to error.— JASA, 1959 v. 54, № 285.

156. M a l i n v a u d E. The consistency of nonlinear regression.— AMS, 1970, v. 41, № 3.

157. M a l l o w s C. L. Some comments on C_p .— Technometrics, 1973, v. 15, № 4.

158. M a r q u a r d t D. W. An algorithm for least squares estimation of nonlinear parameters. — Journal Society of Appl. Math.; 1963, v. 2, № 4.

159. M a r q u a r d t D. W. Generalised inverses, ridge regression, biased linear estimation and nonlinear estimation.— Technometrics, 1970, v. 12, № 3.

160. M a s s y W. F. Principal components regression in exploratory statistical research. — JASA, 1965, v. 60, № 2.

161. Mayer L. S.; Willke T. A. On biased estimation in linear models.— *Technometrics*, 1973, v. 15, № 3.

162. McCallum B. T. Relative asymptotic bias from errors of omission and measurement.— *Econometrica*, 1972, v. 40, № 4.

163. McDonald G. C.; Galarneau O. I. A Monte-Carlo evaluation of somme ridge-type estimators.— *JASA*, 1975, v. 70, № 350.

164. Mehta J. S.; Swamy P. A. Further evidence of the relative efficiencies of Zellner's seemingly unrelated regression equations.— *JASA*, 1976, v. 71, № 355.

165. Neuman J.; Goldstine H. Numerical inversion of matrix of high order.— *Bullityn of the American Mathematical Society*, 1947, v. 53, № 11.

166. Neyman J., Scott E. L. On certain methods of estimating the linear structural relations.— *AMS*, 1961, v. 22, № 3.

167. Obenchain R. L.; Vinod H. D. Ridge analysis following a preliminary test of a shrunken hypothesis.— *Technometrics*, 1975, v. 17, № 4.

168. Oberhoffer W., Kmenta J. A general procedure for obtaining maximum likelihood estimates in generalised regression models.— *Econometrica*, 1974, v. 42, № 3.

169. Osborne M. R. Some aspects of non-linear least squares calculation, In: *Numerical methods for non-linear optimisation*, ed. by Lootsma, NY., 1972.

170. Park R. E. Estimation with heteroscedastic error terms.— *Econometrica*, 1966, v. 34, № 4.

171. Pesaran M. H., Deaton A. S. Testing nonnested nonlinear regression models.— *Econometrica*, 1978, v. 46, № 3.

172. Phillips P. C. B. The iterated minimum distance estimator and the quasi-maximum likelihood estimator.— *Econometrica*, 1976, v. 44, № 3.

173. Ramsay J. O. A comparative study of several robust estimates of slope, intercept, and scale in linear regression.— *JASA*, 1977, v. 72, № 3.

174. Ramsey J. B. Nonlinear estimation and asymptotic approximation.— *Econometrica*, 1978, v. 46, № 4.

175. Revankar N. S. Some finite results in the context of two seemingly unrelated equations.— *JASA*, 1974, v. 69, № 345.

176. Rice J. R.; White J. S. Norms for smoothing and estimation.— *SIAM Review*, 1964, v. 6, № 3.

177. Richardson D. H., Wu D.-M. Least squares and grouping method estimators in the errors in variables model.— *JASA*, 1970, v. 65, № 3.

178. Robinson P. M. Non-linear regression for multiple time-series.— *Journal of Applied Probab.*, 1972, v. 9, № 4.

179. Rothenberg T. J. Efficient estimation with a priori restrictions, New Haven, 1973.

180. Rutenmiller H. C.; Bowers D. A. Estimation in a heteroskedastic regression model.— *JASA*, 1968, v. 63, № 322.

181. S c l o v e S. L. Improved estimators for coefficients in linear regression.— *JASA*, 1968, v. 63, № 322.

182. S p j o t v o l l E. Alternatives to plotting C_p in multiple regression.— *Biometrika*, 1977, v. 64, № 1.

183. S t e i n C. Inadmissibility of the usual estimator for the mean multivariate normal distribution.— *Proceedings of the 3-th Berkly Simp. on Math. Stat. and Probab*, 1956, v. 1, UCP, Berkly.

184. S t e i n C. Multiple regression. — In: *Contributions to probab. and statistics*, N. Y., 1960.

185. S t i g l e r S. M. Do robust estimators work with real data? — *Annals of Statistics*, 1977, v. 5, № 6.

186. S t o n e C. Consistent nonparametric regression.— *Annals of Statistics*, 1977, v. 5, № 4.

187. T a y l o r L. D. Estimation by minimising the sum of absolute errors — In: *Frontiers in econometrics*, ed. Zarembka, N. Y., 1974.

188. T a y l o r W. E. Small sample properties of a class of two stage Aitken estimators.— *Econometrica*, 1977, v. 42, № 2.

189. T a y l o r W. E. The heteroscedastic linear model: exact finite sample results.— *Econometrica*, 1978, v. 46, № 3.

190. T h e i l H., v a n I j z e r e n J. On the efficiency of Wald's method of fitting straight lines.— *Review of International Statistics. Inst.*, 1956, v. 24, № 1.

191. T h e o b a l d C. M. Generalisation of mean square error applied to ridge regression.— *JRSS, ser. B*, 1974, v. 36, № 1.

192. W a g n e r H. M. Linear programming techniques for regression analysis — *JASA*, 1959, v. 54, № 285.

193. W a l d A. The fitting of straight lines if both variables a subject to error.— *AMS*, 1940, v. 11, № 2.

194. W e b s t e r J. T., G u n s t R. F., M a s o n R. L. Latent root regression analysis.— *Technometrica*, 1974, v. 16, № 4.

195. W h i t e J. S. Norms for smoothing and estimation.— *SIAM Review*, 1964, v. 6, № 3.

196. W i c k e n s M. R. A note on the use of proxy variables.— *Econometrica*, 1972, v. 40, № 4.

197. Z e l l n e r A. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. — *JASA*, 1962, v. 57, № 299.

198. Z e l l n e r B. Estimators for seemingly unrelated regression equations: some exact finite sample results. — *JASA*, 1963, v. 67, № 255.

199. Z e l l n e r A.; H u a n g D. S. Further properties of efficient estimators for seemingly unrelated regression equations.— *International Economic Review*, 1962, т. 3, № 3.

ОГЛАВЛЕНИЕ

Предисловие	3
Часть первая. ЛИНЕЙНАЯ РЕГРЕССИЯ КАК БЕЗУСЛОВНОЕ МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ	5
Г л а в а 1. Классическая регрессия. Свойства оценки МНК	5
1.1. Основные предположения. Оценка МНК	5
1.2. Геометрия МНК	15
1.3. Обсуждение предпосылок классической регрессии	16
1.4. Методология статистического оценивания	20
1.5. Теорема Гаусса — Маркова	29
1.6. Коэффициент детерминации и его интерпретация	34
1.7. Состоятельность и асимптотическая нормальность оценки МНК	41
1.8. Свойства оценки МНК при нормальных отклонениях	51
1.9. Общие принципы проверки статистических гипотез и построения доверительных интервалов.	53
1.10. Проверка гипотез и доверительное оценивание в линейной регрессии	61
1.11. Доказательства	72
Г л а в а 2. Другие вопросы линейной регрессии	78
2.1. Взвешенный МНК. Оценка Эйткена	78
2.2. Прогноз по регрессии	87
2.3. Регрессия с ограничениями на параметры	89
2.4. Перебор и недобор факторов в регрессии	92
2.5. Псевдонезависимые регрессии	98
2.6. Вычислительные трудности МНК	108
Часть вторая. АЛЬТЕРНАТИВНЫЕ СХЕМЫ И МЕТОДЫ ОЦЕНИВАНИЯ	115
Г л а в а 3. Регрессия как условное математическое ожидание	115
3.1. Основные предположения	115
3.2. Свойства оценки МНК	116
3.3. Схема случайной выборки	125
3.4. Доказательства	130
Г л а в а 4. Ошибки в независимых переменных.	131
4.1. Постановка задачи. Оценка МНК	131
4.2. Ортогональная регрессия	136
4.3. Метод максимального правдоподобия	142
4.4. Метод группировки	147
4.5. Метод инструментальных переменных	153

4.6. Оценка Картни — Вайссмана	158
4.7. Сравнение оценок	160
4.8. Доказательства	162
Глава 5. Робастные оценки	167
5.1. Робастные оценки параметра положения	167
5.2. Простейшие методы робастного оценивания регрессии	172
5.3. L_ν -оценки	174
5.4. Оценки Хюбера, Андрюса и Рамсея	179
5.5. Сравнение оценок методом статистических испытаний	183
Глава 6. Мультиколлинеарность. Смещенные оценки	186
6.1. Мультиколлинеарность и ее измерение	186
6.2. Строгая мультиколлинеарность	195
6.3. Смещенные оценки	196
6.4. Ридж-оценки	204
6.5. Редуцированные оценки	215
6.6. Оценка метода главных компонент	220
6.7. Оценка Марквардта	222
6.8. Оценка Хокинса	230
6.9. Сравнение оценок методом статистических испытаний	232
Часть третья. НЕЛИНЕЙНАЯ РЕГРЕССИЯ	236
Глава 7. Численное нахождение оценки МНК	236
7.1. Основные определения. Постановка задачи	236
7.2. Существование оценки МНК	241
7.3. Метод Ньютона — Гаусса и его модификации	245
7.4. Метод Левенберга — Марквардта	252
7.5. Единственность оценки МНК	256
7.6. Сведение нелинейной регрессии к линейной	259
7.7. Доказательства	263
Глава 8. Статистические свойства оценки МНК	265
8.1. Непрерывность и асимптотические свойства оценки МНК	265
8.2. Оценка смещения МНК	270
8.3. Проверка статистических гипотез и доверительное оценивание	273
8.4. Псевдонезависимые нелинейные регрессии	282
8.5. Доказательства	284
Приложение. Некоторые дополнительные формулы	287
Список использованной литературы	291

Евгений Зямович Демиденко

ЛИНЕЙНАЯ И НЕЛИНЕЙНАЯ РЕГРЕССИИ

Рецензенты *Г. Г. Пирогов, В. А. Колемаев*

Зав. редакцией *Р. А. Казьмина*

Редактор *Л. Н. Вылегжанина*

Мл. редакторы

Н. М. Лопарева, О. Л. Борисова, М. В. Ульянова

Техн. редактор *Р. Н. Феоктистова*

Корректоры *Т. М. Васильева, Г. В. Хлопцева, З. С. Кандыба*

Худож. редактор *Э. А. Смирнов*

Переплет художника *Т. Н. Погореловой*

ИБ № 1051

Сдано в набор 25.08.80. Подписано в печать 17.04.81.
А08336. Формат 84×108¹/₃₂. Бум. кн.-журн. Гарнитура
«Литературная». Печать высокая. П. л. 9,5. Усл. п. л. 15,96
Уч.-изд. л. 15,94. Тираж 5000 экз. Заказ 2067 Цена 2 р. 60 к.
Издательство «Финансы и статистика»,
Москва, ул. Чернышевского, 7.

Московская типография № 4 Союзполиграфпрома
при Государственном комитете СССР
по делам издательств, полиграфии и книжной торговли,
129041, Москва, Б. Переяславская ул., 46

Д30

Демиденко Е. З.

Линейная и нелинейная регрессии. — М.: Финансы и статистика, 1981. — 302 с., ил.

В пер.: 2 р. 60 к.

Рассматриваются современные методы оценивания параметров в линейных и нелинейных регрессиях — оценивание в условиях ошибок измерения, устойчивое и др. Даются практические рекомендации нахождения оценок. Изучаются статистические свойства оценок, методы построения доверительных интервалов и проверки гипотез в линейных и нелинейных регрессиях.

Для специалистов в области моделирования социально-экономических процессов.

$\frac{10805-054}{Д010(01)-81}$ 15—81 (С) 1702060000

ББК 22172
517.8