

В.И. ДЕНИСОВ, Д.В. ЛИЦИЦИН

**МЕТОДЫ ПОСТРОЕНИЯ
МНОГОФАКТОРНЫХ МОДЕЛЕЙ
ПО НЕОДНОРОДНЫМ,
НЕГАУССОВСКИМ,
ЗАВИСИМЫМ НАБЛЮДЕНИЯМ**

УДК 519.242.7
Д 332

Рецензенты: *Г.С. Лбов*, д-р техн. наук, проф.;
В.И. Хабаров, д-р техн. наук, проф.

*Работа выполнена в рамках аналитической ведомственной
целевой программы
«Развитие научного потенциала высшей школы (2006–2008 годы)»,
проект РНП-2.1.2.43*

Денисов, В.И.

Д 332 Методы построения многофакторных моделей по неоднородным, негауссовским, зависимым наблюдениям : монография / В.И. Денисов, Д.В. Лисицин. – Новосибирск : Изд-во НГТУ, 2008. – 360 с. (Серия «Монографии НГТУ»).

ISBN 978-5-7782-0920-6

В монографии излагаются методы оценивания неизвестных параметров, оптимального планирования эксперимента и выбора структуры для моделей сложных многофакторных объектов в условиях неоднородности, негауссовости, зависимости наблюдений.

Рассматриваются методы оценивания, основанные на модели байесовского точечного засорения, методы оценивания для многомерных моделей с количественными и разнотипными откликами и ковариационной матрицей в виде кронекерова произведения нескольких матриц, методы планирования эксперимента при оценивании параметров моделей количественных и разнотипных откликов по неоднородным негауссовским наблюдениям, методы выбора структуры однооткликowych и многооткликowych гауссовских и негауссовских моделей.

Предназначена для научных работников, аспирантов и студентов старших курсов вузов, специализирующихся в области прикладной статистики.

УДК 519.242.7

ISBN 978-5-7782-0920-6

© Денисов В.И., Лисицин Д.В., 2008
© Новосибирский государственный
технический университет, 2008

Ministry of Education and Science of Russian Federation

NOVOSIBIRSK STATE TECHNICAL UNIVERSITY

V.I. DENISOV, D.V. LISITSIN

**METHODS OF BUILDING
MULTIFACTORIAL MODELS
BASED ON HETEROGENEOUS,
NON-GAUSSIAN,
DEPENDENT OBSERVATIONS**

NOVOSIBIRSK
2008

UDC 519.242.7
D 332

Reviewers:
Prof. G.S. Lboy, D. Sc. (Eng.)
Prof. V.I. Khabarov, D. Sc. (Eng.)

*Work is carried out within the framework of the departmental analytical target program for 2006–2008
Development of scientific potential of higher school
(2006–2008), project PHII-2.1.2.43*

Denisov, V.I.

D 332 Methods of building multifactorial models based on heterogeneous, non-Gaussian, dependent observations : monograph / V.I. Denisov, D.V. Lisitsin. – Novosibirsk : NSTU, 2008. – 360 pp. – (NSTU Monographs series)

ISBN 978-5-7782-0920-6

The monograph is devoted to methods of estimation of unknown parameters, optimal designing an experiment and structure selection for models of complex multifactorial objects based on heterogeneous, non-Gaussian, dependent observations.

The estimation methods based on the model of bayesian dot contamination, estimation methods for multivariate models with quantitative or mixed quantitative and qualitative responses and a covariance matrix presented as Kronecker product of several matrices, methods of designing experiments for estimation of parameters of models with quantitative or mixed quantitative and qualitative responses based on heterogeneous, non-Gaussian observations, methods of structure selection for Gaussian and non-Gaussian single- and multiresponse models are considered.

The monograph is intended for researchers, postgraduate and graduate students majoring in applied statistics.

UDC 519.242.7

ISBN 978-5-7782-0920-6

© Denisov V.I., Lisitsin D.V., 2008
© Novosibirsk state
technical university, 2008

ПРЕДИСЛОВИЕ

Многие исследования в различных областях знаний (физике, химии, биологии, технике, экономике и т. д.) опираются на статистические данные, отражающие некоторые стороны изучаемых объектов и явлений. На основе собранной статистики строятся математические модели, служащие для описания и прогнозирования поведения объектов изучения.

Среди основных этапов построения математической модели по статистическим данным – оценивание неизвестных параметров, планирование эксперимента и выбор структуры модели.

В основе методов, используемых для реализации данных этапов, лежат некоторые предположения о свойствах изучаемых объектов, в частности, о свойствах ошибок наблюдений. Так, метод наименьших квадратов (МНК) – наиболее часто используемый в практике статистических исследований метод оценивания параметров – обладает рядом положительных свойств при независимых гомоскедастичных (имеющих одинаковую дисперсию) ошибках, распределенных по нормальному закону.

При зависимости ошибок наблюдений переходят к использованию многомерных методов. Например, может использоваться обобщенный МНК (ОМНК), требующий, однако, знания ковариационной матрицы наблюдений. На практике значение ковариационной матрицы неизвестно и для ее определения может вводиться некоторая экономная параметризация, позволяющая оценить параметры методом максимального правдоподобия (ММП) в условиях многомерного нормального распределения ошибок. Используется достаточно большое число ковариационных структур, но при этом усложняются процедуры оценивания параметров. К сравнительно простым процедурам поиска оценок приводит мультипликативная ковариационная структура в виде

кронекерова произведения нескольких положительно определенных матриц.

В то же время предположение о нормальности ошибок часто является слишком жестким. Например, в наборе данных могут присутствовать грубые ошибки (выбросы), возникающие вследствие нарушения условий эксперимента, неправильного измерения, засорения данных и т. п. В таких случаях более адекватны распределения с тяжелыми хвостами, для которых оценки по МНК могут быть неустойчивыми. С другой стороны, в некоторых ситуациях более адекватным будет предположение об ошибках наблюдений, имеющих хвосты более легкие, чем у нормального распределения, оценки по МНК при этом также теряют некоторые свои положительные свойства.

На практике данные часто являются неоднородными (разнораспределенными). К неоднородности приводят выбросы в наборе данных, другой пример – гетероскедастичность ошибок наблюдений. В общем случае от наблюдения к наблюдению могут меняться любые параметры модели и даже сам вид распределения.

Возможно более полный учет статистических свойств наблюдений способствует повышению качества моделей, однако на практике далеко не всегда имеется полная информация.

Для решения этой проблемы разработаны различные подходы, приводящие к устойчивым методам оценивания. Сюда относят робастные, адаптивные и непараметрические методы. Непараметрические оценки строятся в предположении неизвестного распределения ошибок, лишь удовлетворяющего ряду ограничений. При робастном оценивании разрабатываются процедуры, оптимальные в окрестности параметрической модели. Один из перспективных подходов, приводящий к робастным решениям и разработанный А.М. Шурыгиным, основан на модели байесовского точечного засорения данных. При адаптивном оценивании конкретный вид статистической процедуры выбирается на основе оценки какой-либо характеристики неизвестной функции распределения наблюдений.

Процедуры выбора структуры модели, основанные на использовании МНК- и ОМНК-оценок, также оказываются чувствительными к нарушению предположений классической регрессионной модели. Поэтому при выборе структуры однооткликowej нестрогой нормальной модели применяются устойчивые модификации традиционных проце-

дур (пошаговая регрессия, информационный критерий Акаике, критерий Меллоуса).

В классических регрессионных моделях, изучаемых в рамках планирования эксперимента, ошибки наблюдений обычно предполагаются одинаково распределенными, могут использоваться условия нормальности распределения ошибок и их гетероскедастичности.

Несмотря на широкую разработанность методов устойчивого оценивания параметров, вопросы построения моделей многофакторных объектов по неоднородным, негауссовским, зависимым наблюдениям требуют дальнейшей разработки и исследования. Указанным вопросам и посвящена данная монография.

Книга состоит из пяти глав. Глава 1 носит обзорный характер и содержит описание рассматриваемых моделей, методов оценивания параметров и проверки гипотез. Глава 2 посвящена развитию методов устойчивого оценивания, основанных главным образом на модели байесовского точечного засорения. В главе 3 развиваются методы оценивания параметров многомерных моделей с количественными и разнотипными откликами и мультипликативной ковариационной структурой ошибок количественных откликов, в том числе по данным с пропусками. Глава 4 посвящена развитию методов оптимального планирования эксперимента при ММП-оценивании и робастном оценивании параметров моделей количественных откликов по неоднородным негауссовским наблюдениям и при ММП-оценивании параметров моделей разнотипных откликов. В главе 5 развиваются методы выбора структуры однооткликowych и многооткликowych гауссовских и негауссовских моделей.

Ряд результатов, отраженных в монографии, получен при поддержке гранта Президента РФ (№ МК-3376.2005.9, 2005–2006 гг.), Министерства образования и науки РФ (ведомственная программа «Развитие научного потенциала высшей школы», коды проектов 4574, 2005 г., РНП-2.1.2.43, 2006–2007 гг.), гранта администрации Новосибирской области (договор № ФГМ-4-05, 2005 г.).

Авторы выражают благодарность д-ру техн. наук, проф. А.А. Попову, канд. техн. наук К.В. Гаврилову, аспиранту Д.С. Форманчук, в соавторстве с которыми получен ряд результатов, изложенных в монографии.

ОГЛАВЛЕНИЕ

Предисловие	7
Глава 1. Введение в методы оценивания параметров	10
1.1. Однооткликовая модель	10
1.2. Многооткликовая модель	19
1.3. Модель с качественными и разнотипными факторами	26
1.4. Оценивание по данным с пропусками	29
1.5. <i>M</i> -оценки	34
1.6. Устойчивое оценивание одномерных моделей	37
1.7. Устойчивое оценивание многомерных моделей	46
1.8. Оценивание при наличии неоднородных данных	49
Глава 2. Устойчивое <i>M</i>-оценивание	52
2.1. Оценивание однородной модели	52
2.2. Оценивание неоднородной модели	59
2.2.1. Показатели качества	59
2.2.2. Оптимизация с ограничениями на матрицы регрессоров	64
2.2.3. Поэлементная оптимизация	69
2.2.4. Оценивание при ограничениях на параметры	78
2.3. Многомерная нормальная модель	80
2.3.1. Оценивание	80
2.3.2. Исследование	84
2.4. Двусторонняя экспоненциальная модель	86
2.4.1. Адаптивное оценивание	86
2.4.2. Оценивание при байесовском точечном засорении	91
2.4.3. Минимаксный подход	97
2.4.4. Исследование	100

Глава 3. Оценивание при мультипликативной ковариационной структуре ошибок	115
3.1. Оценивание параметров регрессионной модели с эллиптическим распределением ошибок.....	115
3.1.1. Модель	115
3.1.2. Оценивание параметров	120
3.1.3. Частные случаи эллиптического распределения	127
3.1.4. Прикладные аспекты разработанного подхода	132
3.1.5. Исследование.....	134
3.2. Оценивание параметров модели при наличии разнотипных откликов и пропусков в данных.....	142
3.2.1. Модель	142
3.2.2. Оценивание параметров	146
3.2.3. Исследование.....	155
Глава 4. Оптимальное планирование эксперимента	161
4.1. Теория планирования регрессионного эксперимента	161
4.1.1. Задачи оптимального планирования эксперимента	161
4.1.2. Критерии оптимальности планов эксперимента	163
4.1.3. Алгоритмы численного построения оптимальных планов	170
4.2. Планирование эксперимента при ММП-оценивании параметров по неоднородным наблюдениям.....	180
4.2.1. Планирование эксперимента при независимых ошибках наблюдений	180
4.2.2. Планирование эксперимента при зависимых ошибках наблюдений	186
4.2.3. Исследование свойств планов эксперимента.....	189
4.3. Планирование эксперимента при робастном оценивании параметров по неоднородным наблюдениям.....	202
4.3.1. Показатели качества M -оценок	202
4.3.2. Постановка задачи планирования эксперимента.....	206
4.3.3. Условия оптимальности планов эксперимента	212
4.3.4. Алгоритмы построения оптимальных планов	217
4.3.5. Исследование свойств планов эксперимента.....	219
4.4. Планирование эксперимента при оценивании параметров модели с разнотипными откликами и пропусками в данных	222
4.4.1. Постановка задачи планирования эксперимента.....	222

4.4.2. Планирование эксперимента с учетом появления пропусков	227
4.4.3. Свойства инвариантности планов эксперимента	231
4.4.4. Исследование свойств планов эксперимента при возникновении пропусков	237
Глава 5. Выбор структуры модели	243
5.1. Задачи и критерии выбора структуры.....	243
5.1.1. Задачи выбора структуры	243
5.1.2. Критерии, использующие разбиение выборки на две части	251
5.1.3. Критерии типа скользящего контроля.....	264
5.1.4. Критерии, не использующие экзаменационную выборку.....	274
5.1.5. Выбор структуры при наличии качественных и разнотипных факторов	280
5.1.6. Исследование критериев выбора структуры.....	282
5.2. Алгоритмы выбора структуры	288
5.2.1. Алгоритмы решения традиционной задачи выбора структуры	288
5.2.2. Алгоритмы решения обобщенной задачи выбора структуры	298
5.2.3. Исследование алгоритмов выбора структуры	308
5.3. Выбор структуры при негауссовских и зависимых ошибках	317
5.3.1. Робастные критерии выбора структуры	317
5.3.2. Исследование робастных критериев выбора структуры....	326
5.3.3. Выбор структуры при зависимых ошибках	334
5.3.4. Алгоритм последовательного уточнения оптимальной структуры	339
Заключение.....	343
Библиографический список	344

НАУЧНОЕ ИЗДАНИЕ

Владимир Иванович Денисов
Даниил Валерьевич Лисицин

**МЕТОДЫ ПОСТРОЕНИЯ
МНОГОФАКТОРНЫХ МОДЕЛЕЙ
ПО НЕОДНОРОДНЫМ, НЕГАУССОВСКИМ,
ЗАВИСИМЫМ НАБЛЮДЕНИЯМ**

Монография

Редактор *Н.А. Лукашова*
Выпускающий редактор *И.П. Брованова*
Технический редактор *Н.В. Гаврилова*
Художественный редактор *А.В. Ладыжская*
Корректор *И.Е. Семенова*
Компьютерная верстка и рисунки *С.И. Ткачева*

Подписано в печать 17.04.08
Формат 60 × 90 1/16. Бумага офсетная
Уч.-изд. л. 22,5. Печ. л. 22,5
Тираж 3000 экз. (1-й з-д – 100 экз.)
Изд. № 509. Заказ № 584

Издательство Новосибирского государственного
технического университета
630092, г. Новосибирск, пр. К. Маркса, 20
Тел. (383) 346-31-87
E-mail: office@publish.nstu.ru

Отпечатано в типографии
Новосибирского государственного технического университета
630092, г. Новосибирск, пр. К. Маркса, 20

ГЛАВА 1

ВВЕДЕНИЕ В МЕТОДЫ ОЦЕНИВАНИЯ ПАРАМЕТРОВ

В главе описываются типы многофакторных моделей, рассматриваются задачи оценивания неизвестных параметров и проверки гипотез для одно- и многооткликковых моделей, а также подходы к моделированию по неоднородным, негауссовским, зависимым данным.

1.1. ОДНООТКЛИКОВАЯ МОДЕЛЬ

Пусть количественные показатели, поведение которых интересует исследователя, зависят от ряда других показателей и стоит задача моделирования этой зависимости. Первую группу показателей называют откликами, выходными или зависимыми переменными системы, вторую – факторами, входными или независимыми переменными.

Предположим, что моделируется один отклик, вид зависимости известен с точностью до некоторых параметров и является линейным относительно них. Многофакторную регрессионную модель запишем в виде

$$y = f^T(x)\theta + e = \sum_{j=1}^m f_j(x)\theta_j + e, \quad (1.1)$$

где y – значение зависимой переменной; $f^T(x) = f_1(x), \dots, f_m(x)$ – заданная вектор-функция от неслучайной векторной входной переменной x ; T – знак транспонирования; $\theta = (\theta_1, \dots, \theta_m)^T$ – вектор неизвест-

ных параметров, которые необходимо определить по результатам экспериментов (измерений); e – ошибка. Функции $f_1(x), \dots, f_m(x)$ называют регрессорами.

Пусть проведено N измерений отклика при некоторых значениях переменной x и ошибки измерений являются некоррелированными случайными величинами с нулевыми математическими ожиданиями и одинаковыми дисперсиями σ^2 .

Введем обозначения: Y – N -мерный вектор всех измерений отклика; X – матрица размерности $N \times m$, строками которой являются значения регрессоров в соответствующих точках факторного пространства; E – N -мерный вектор ошибок измерений отклика. Тогда справедлива запись

$$Y = X\theta + E.$$

Для оценивания вектора параметров θ часто пользуются методом наименьших квадратов [2, 3, 18, 19, 96, 106], который требует минимизации по θ суммы квадратов отклонений вычисленных и измеренных значений. В результате МНК-оценки доставляют минимальное значение функции вида

$$Q(\theta) = Y - X\theta^T Y - X\theta.$$

Последняя задача сводится к решению системы линейных уравнений

$$X^T X \hat{\theta} = X^T Y, \quad (1.2)$$

где $\hat{\theta}$ – оценка θ . Данная система называется системой нормальных уравнений.

Если матрица $X^T X$ – невырожденная, то система (1.2) имеет единственное решение, которое можно записать как

$$\hat{\theta} = X^T X^{-1} X^T Y. \quad (1.3)$$

МНК-оценка θ является наилучшей линейной несмещенной оценкой, т. е. является несмещенной ($\mathbf{E}\hat{\theta} = \theta$, где \mathbf{E} – символ математического ожидания) и обладает наименьшей дисперсией среди всех линейных несмещенных оценок.

Величина $Y - X\hat{\theta}^T Y - X\hat{\theta}$ называется остаточной суммой квадратов и может быть представлена в виде

$$Y - X\hat{\theta}^T Y - X\hat{\theta} = Y^T Y - Y^T X\hat{\theta}.$$

В случае если дисперсия наблюдений неизвестна, ее также можно оценить. МНК-оценка $\hat{\sigma}^2$ имеет вид

$$\hat{\sigma}^2 = \frac{1}{N - m} Y - X\hat{\theta}^T Y - X\hat{\theta}$$

и является несмещенной.

Другим методом, используемым для оценивания параметров, является метод максимального правдоподобия [9, 18, 19]. ММП-оценки доставляют максимальное значение функции правдоподобия $g(\Phi|Y)$, которая представляет собой плотность распределения вектора наблюдений, рассматриваемую как функцию параметров:

$$g(\Phi|Y) = g(Y|\Phi),$$

где $g(Y|\Phi)$ – плотность распределения вектора наблюдений, Φ – вектор всех подлежащих оцениванию параметров. На практике часто вместо функции $g(\Phi|Y)$ используется ее логарифм и задача оценивания формулируется как поиск точки максимума функции

$$Q(\Phi) = \ln g(\Phi|Y).$$

Предположим, что ошибки наблюдений модели (1.1) распределены по нормальному закону, тогда максимум функции правдоподобия соответствует МНК-оценке $\hat{\theta}$ и оценке $\hat{\sigma}^2$ вида

$$\hat{\sigma}^2 = \frac{1}{N} Y - X\hat{\theta}^T Y - X\hat{\theta}. \quad (1.4)$$

Если функция правдоподобия дифференцируема и ограничена сверху, то ММП-оценку $\hat{\Phi}$ можно найти, решая уравнение (или систему уравнений) правдоподобия:

$$\left. \frac{\partial Q(\Phi)}{\partial \Phi} \right|_{\Phi=\hat{\Phi}} = 0. \quad (1.5)$$

Если существует несколько решений системы (1.5), то различают ММП-оценку в строгом смысле как значение $\hat{\Phi}$, максимизирующее функцию правдоподобия $g(\Phi|Y)$, и ММП-оценку в широком смысле как любое решение системы (1.5). Уравнение (1.5) часто бывает нелинейным относительно параметров, поэтому для поиска оценок необходимо использовать какие-либо численные методы.

При выполнении некоторых условий регулярности ММП-оценка $\hat{\Phi}$ является состоятельной, асимптотически несмещенной и асимптотически эффективной.

Для использования ММП необходимо знать распределение наблюдений, однако на практике оно не всегда известно. Тем не менее можно использовать ММП, зафиксировав некоторое распределение, независимо от того, распределены ли в соответствии с ним ошибки наблюдений. Полученные в результате оценки называются квазиправдоподобными [81].

Перейдем к рассмотрению ситуации, когда ошибки измерений являются коррелированными и гетероскедастичными, т. е. имеющими разные дисперсии. В частности, предположим, что вектор ошибок E имеет ковариационную матрицу вида

$$\text{cov } E = \sigma^2 V, \quad (1.6)$$

где V – известная положительно определенная матрица размерности $N \times N$. Произведем следующее преобразование исходной модели:

$$P^{-1}Y = P^{-1}X\theta + P^{-1}E, \quad (1.7)$$

где P – невырожденная $N \times N$ -матрица, для которой справедливо $V = PP^T$. Ошибки наблюдений в рамках модели (1.7) являются некор-

релированными и гомоскедастичными, поэтому наилучшей линейной несмещенной оценкой для θ будет МНК-оценка, полученная в рамках данной модели. Эта оценка имеет вид

$$\hat{\theta} = X^T V^{-1} X^{-1} X^T V^{-1} Y \quad (1.8)$$

и называется оценкой по обобщенному МНК, оценкой Гаусса–Маркова, а в эконометрической литературе – оценкой Эйткена [18]. Соответственно несмещенной оценкой параметра σ^2 будет

$$\hat{\sigma}^2 = \frac{1}{N-m} Y - X\hat{\theta}^T V^{-1} Y - X\hat{\theta} .$$

Заметим, что оценка (1.3) по обычному МНК останется в этом случае несмещенной, но уже не будет наилучшей в классе линейных оценок.

Рассмотрим модель (1.1), (1.6) с ограничениями на параметры вида [89, 187]:

$$R\theta = d , \quad (1.9)$$

где R – известная $u \times m$ -матрица полного строчного ранга, d – известный u -мерный вектор.

ОМНК-оценки параметров модели $\hat{\theta}_R$, удовлетворяющие (1.9), находят по методу неопределенных множителей Лагранжа, согласно которому оценки параметров доставляют минимальное значение функции Лагранжа [19, 32]:

$$Y - X\theta^T V^{-1} Y - X\theta + 2\gamma^T (R\theta - d) ,$$

где γ – u -мерный вектор неизвестных множителей Лагранжа.

Для единственности решения матрица $\begin{bmatrix} P^{-1}X^T & \vdots & R^T \end{bmatrix}^T$ должна иметь полный столбцовый ранг; данное условие заменяет требование

полноты ранга матрицы X для рассмотренного выше случая модели без ограничений.

Продифференцируем функцию Лагранжа по параметрам θ и γ и приравняем производную нулю, после сокращения константного сомножителя получим систему уравнений, которой удовлетворяют оценки:

$$\begin{cases} X^T V^{-1} X \hat{\theta}_R + R^T \hat{\gamma}_R = X^T V^{-1} Y, \\ R \hat{\theta}_R = d, \end{cases} \quad (1.10)$$

где $\hat{\gamma}_R$ – вектор значений множителей Лагранжа.

Введем обозначение

$$\begin{bmatrix} X^T V^{-1} X & R^T \\ R & 0 \end{bmatrix}^{-1} = \begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix}, \quad (1.11)$$

где 0 – нулевая ($u \times u$)-матрица, тогда справедливо

$$\hat{\theta}_R = C_1 X^T V^{-1} Y + C_2 d, \quad (1.12)$$

$$\hat{\gamma}_R = C_3 X^T V^{-1} Y + C_4 d. \quad (1.13)$$

Получим выражение для $\hat{\theta}_R$ в предположении, что матрица X имеет полный столбцовый ранг. Из первого уравнения системы (1.10) выразим вектор $\hat{\theta}_R$:

$$\hat{\theta}_R = \hat{\theta} - H^{-1} R^T \hat{\gamma}_R, \quad (1.14)$$

где $\hat{\theta} = H^{-1} X^T V^{-1} Y$, $H = X^T V^{-1} X$.

Подставим полученное выражение во второе уравнение системы (1.10), после чего выразим из него $\hat{\gamma}_R$:

$$\hat{\gamma}_R = R H^{-1} R^T \hat{\theta} - d.$$

Подставляя полученное выражение в (1.14), получаем

$$\hat{\theta}_R = \hat{\theta} + H^{-1}R^T (RH^{-1}R^T)^{-1} d - R\hat{\theta} . \quad (1.15)$$

Аналог остаточной суммы квадратов, соответствующий оценке $\hat{\theta}_R$, может быть записан в виде

$$(Y - X\hat{\theta}_R)^T V^{-1} (Y - X\hat{\theta}_R) = Y^T V^{-1} Y - Y^T V^{-1} X \hat{\theta}_R - d^T \hat{\gamma}_R .$$

Оценка параметра σ^2 имеет вид

$$\hat{\sigma}_R^2 = \frac{1}{N - m + u} Y - X\hat{\theta}_R^T V^{-1} Y - X\hat{\theta}_R .$$

На практике значение ковариационной матрицы V обычно неизвестно, поэтому использовать ОМНК невозможно. Один из методов, применяемых в такой ситуации, основан на использовании экономной параметризации ковариационной матрицы

$$V = V(\phi) , \quad (1.16)$$

позволяющей относительно просто оценить параметры по ММП в условиях многомерного нормального распределения ошибок [3]. Рассмотрим его.

Функция правдоподобия в данном случае имеет вид

$$g(\theta, \sigma^2, \phi | Y) = (2\pi\sigma^2)^{-N/2} |V(\phi)|^{-1/2} \exp \left\{ -\frac{Y - X\theta^T V^{-1}(\phi) Y - X\theta}{2\sigma^2} \right\} ,$$

где $|\cdot|$ – определитель матрицы.

При условии полноты ранга матрицы X система уравнений правдоподобия, учитывающая ограничения (1.9), после ряда преобразова-

ний примет вид системы (1.15) с $H = X^T V^{-1}(\hat{\phi}) X$, $\hat{\theta} = H^{-1} X^T V^{-1}(\hat{\phi}) Y$, дополненной соотношениями

$$\hat{\sigma}_R^2 = \frac{1}{N} Y - X \hat{\theta}_R^T V^{-1}(\hat{\phi}) Y - X \hat{\theta}_R,$$

$$\hat{\sigma}_R^2 \operatorname{tr} V(\hat{\phi}) D_i = Y - X \hat{\theta}_R^T D_i Y - X \hat{\theta}_R, \quad i = 1, \dots, \dim \phi,$$

где tr – операция взятия следа матрицы, $D_i = \left. \frac{\partial V^{-1}(\phi)}{\partial \phi_i} \right|_{\phi = \hat{\phi}}$, ϕ_i – i -й

элемент вектора ковариационных параметров, \dim – размерность вектора.

Решение полученной системы можно осуществлять методом покомпонентного оценивания [41], который выгодно применять, когда фиксация некоторых элементов множества оцениваемых параметров существенно упрощает оптимизацию функции качества относительно других элементов. Согласно данному методу множество параметров разбивается на ряд подмножеств. Решение находится итеративно, при этом итерация включает несколько этапов, каждый из которых состоит в нахождении очередного значения оценок для одного подмножества параметров из соответствующей подсистемы уравнений правдоподобия при фиксированных значениях параметров остальных подмножеств.

Если в рассматриваемом случае выделить три подмножества оцениваемых параметров:

$$\theta_i, i = 1, 2, \dots, m, \quad \sigma^2, \quad \phi_i, i = 1, 2, \dots, \dim \phi,$$

то первые два этапа каждой итерации являются неитеративными, причем первый этап состоит в применении ОМНК при фиксированных значениях параметров из второго и третьего подмножеств. Сложность третьего этапа зависит от вида используемой параметризации (1.16), в частности, решение соответствующей системы уравнений может представлять собой итеративный процесс. Как следствие, процедура покомпонентного оценивания будет «дважды итеративным» процессом: каждая «внешняя» итерация покомпонентного оценивания включает в себя «внутренний» итеративный процесс.

На практике часто используются простые типы параметризаций, при которых ковариационные матрицы строятся исходя из авторегрессионного представления последовательности ошибок, наличия случайных эффектов и т. д. [150]. К простым вычислительным схемам приводит использование мультипликативной параметризации ковариационной матрицы в виде кронекерова произведения нескольких положительно определенных матриц [108, 166, 175]. Подобная параметризация возникает при описании многооткликовых моделей (см. [175] и п. 1.2), в моделях пространственных и пространственно-временных случайных полей с разделимыми ковариационными матрицами, которые часто используются в метеорологии, экологии, океанографии, при обработке изображений (см. [108, 166]). Методы оценивания регрессионных моделей с такой ковариационной структурой подробно освещены в гл. 3.

Анализ рассмотренных моделей, помимо оценивания неизвестных параметров, может включать в себя проверку гипотез относительно параметров, имеющих вид линейных уравнений. Рассмотрим метод проверки линейных гипотез в модели (1.1), (1.6). Предположим, что ошибки наблюдений подчиняются многомерному нормальному распределению. Пусть проверяемая гипотеза имеет вид

$$K\theta = c, \quad (1.17)$$

где K – известная $v \times m$ -матрица, c – известный v -мерный вектор. Рассмотрим случай, когда предполагается, что матрицы R и $\begin{bmatrix} R^T & K^T \end{bmatrix}^T$ могут не иметь полного строчного ранга. Статистика для проверки гипотезы имеет вид [32]

$$F = \frac{h/f_1}{g/f_2}, \quad (1.18)$$

где $g = \min_{R\theta=d} Y - X\theta^T V^{-1} Y - X\theta$; h – увеличение g , вызываемое наложением на параметры ограничений (1.17), т. е. $g + h = \min_{\substack{R\theta=d \\ K\theta=c}} Y - X\theta^T V^{-1} Y - X\theta$, $f_1 = \text{rg} \begin{bmatrix} R^T & K^T \end{bmatrix} - \text{rg} R$, $f_2 = N - m + \text{rg} R$; rg – ранг матрицы. Данная статистика имеет F -распределение с f_1 и f_2 степенями свободы.

1.2. МНОГООТКЛИКОВАЯ МОДЕЛЬ

В многооткликовом случае имеется большое разнообразие возможных типов линейных моделей, которые применимы для описания поведения объекта исследования. Мы рассмотрим традиционную многооткликовую (многомерную) модель, модель псевдонезависимых регрессий и модель с общими параметрами, оставляя в стороне такие типы, как системы одновременных уравнений [3, 31] и модели кривых роста (модели неполного строчного ранга) [15, 189].

Пусть строится многофакторная регрессионная зависимость для l -мерного отклика. Предположим, что Y – матрица измерений откликов размерности $N \times l$; X – матрица размерности $N \times m$, имеющая тот же смысл, что и для модели (1.1); E – матрица ошибок измерений откликов размерности $N \times l$; Θ – матрица неизвестных параметров размерности $m \times l$. Тогда традиционная многооткликовая (многомерная) модель имеет вид [37, 96, 189]

$$Y = X\Theta + E. \quad (1.19)$$

Примем предположение, что строки матрицы E имеют нулевой вектор математического ожидания и положительно определенную ковариационную матрицу Σ размерности $l \times l$, а элементы различных строк матрицы E – некоррелированы. Это влечет коррелированность откликов и некоррелированность их различных измерений.

Характерной особенностью традиционной многооткликовой модели является то, что состав регрессоров в уравнениях регрессии всех откликов один и тот же.

Модель (1.19) можно записать в виде однооткликовой модели

$$\tilde{Y} = I_l \otimes X \theta + \tilde{E}, \quad (1.20)$$

где $\tilde{Y} = \begin{bmatrix} Y_1^T & \dots & Y_l^T \end{bmatrix}^T$, I_l – единичная матрица размерности $l \times l$, $\theta = \begin{bmatrix} \Theta_1^T & \dots & \Theta_l^T \end{bmatrix}^T$, $\tilde{E} = \begin{bmatrix} E_1^T & \dots & E_l^T \end{bmatrix}^T$, Y_i , Θ_i , E_i – i -е столбцы матриц Y , Θ , E соответственно.

Знаком \otimes обозначено кронекерово произведение, которое определяется следующей формулой:

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1k}B \\ \dots & \dots & \dots \\ a_{n1}B & \dots & a_{nk}B \end{bmatrix},$$

где A – матрица размерности $n \times k$, a_{ij} – i -й элемент j -го столбца матрицы A .

Ковариационная матрица вектора \tilde{E} имеет вид

$$\text{cov } \tilde{E} = \Sigma \otimes I_N. \quad (1.21)$$

Модель (1.20) можно преобразовать в однооткликую модель с некоррелированными случайными ошибками наблюдений, воспользовавшись преобразованием (1.7):

$$(S^{-1} \otimes I_N) \tilde{Y} = (S^{-1} \otimes I_N) I_l \otimes X \theta + (S^{-1} \otimes I_N) \tilde{E}, \quad (1.22)$$

где S – невырожденная $l \times l$ -матрица, для которой справедливо $\Sigma = SS^T$.

Для оценивания вектора параметров θ естественно воспользоваться ОМНК, который приводит к наилучшей линейной несмещенной оценке вида (1.8) с $V = \Sigma \otimes I_N$, $\sigma^2 = 1$, однако ОМНК-оценка равна обычной МНК-оценке [18, 96]. МНК-оценка матрицы Θ записывается в виде

$$\hat{\Theta} = (X^T X)^{-1} X^T Y. \quad (1.23)$$

Заметим, что для вычисления оценки (1.23) необходимо обратить матрицу размерности $m \times m$, в то время как непосредственное вычисление оценки вектора θ в рамках модели (1.20)–(1.21) требует обращения матрицы размерности $(lm) \times (lm)$.

Многооткликowym аналогом остаточных сумм квадратов является матрица остаточных сумм квадратов-произведений вида

$$Y - X \hat{\Theta} \quad Y - X \hat{\Theta} = Y^T Y - Y^T X \hat{\Theta}.$$

Одно из преимуществ МНК-оценки (1.23) – независимость от ковариационной матрицы Σ , поскольку последняя обычно неизвестна. Если необходимо определить значение ковариационной матрицы, то можно воспользоваться МНК-оценкой вида

$$\hat{\Sigma} = \frac{1}{N-m} Y - X\hat{\Theta}^T Y - X\hat{\Theta} . \quad (1.24)$$

Рассмотрим модель с ограничениями на параметры вида [89]

$$R\Theta = U , \quad (1.25)$$

где R – известная $u \times m$ -матрица полного строчного ранга, U – известная $u \times l$ -матрица. Система (1.25) может быть представлена также в виде

$$I_l \otimes R \theta = \left[U_1^T \mid \dots \mid U_l^T \right]^T , \quad (1.26)$$

где U_i – i -й столбец матрицы U .

Оценки параметров модели $\hat{\Theta}_R$, удовлетворяющие (1.25), находятся методом неопределенных множителей Лагранжа. Для единственности решения предположим, что матрица $\left[X^T \mid R^T \right]^T$ имеет полный столбцовый ранг. Тогда оценки $\hat{\Theta}_R$ можно найти, решив систему уравнений

$$\begin{bmatrix} X^T X & R^T \\ R & 0 \end{bmatrix} \begin{bmatrix} \hat{\Theta}_R \\ \hat{L}_R \end{bmatrix} = \begin{bmatrix} X^T Y \\ U \end{bmatrix} , \quad (1.27)$$

где 0 – нулевая $u \times u$ -матрица, \hat{L}_R – $u \times l$ -матрица значений множителей Лагранжа. Также справедливо представление

$$\hat{\Theta}_R = C_1 X^T Y + C_2 U , \quad (1.28)$$

$$\hat{L}_R = C_3 X^T Y + C_4 U , \quad (1.29)$$

где матрицы C_i определяются из (1.11) с $V = I_N$.

В предположении, что матрица X имеет полный столбцовый ранг, справедливо выражение

$$\hat{\Theta}_R = \hat{\Theta} + X^T X^{-1} R^T \left[R X^T X^{-1} R^T \right]^{-1} U - R \hat{\Theta} .$$

Матрицу остаточных сумм квадратов-произведений, соответствующую оценке $\hat{\Theta}_R$, можно представить следующим образом:

$$Y - X \hat{\Theta}_R^T \quad Y - X \hat{\Theta}_R = Y^T Y - Y^T X \hat{\Theta}_R - U^T \hat{L}_R . \quad (1.30)$$

Оценка матрицы Σ имеет вид

$$\hat{\Sigma}_R = \frac{1}{N - m + u} Y - X \hat{\Theta}_R^T \quad Y - X \hat{\Theta}_R . \quad (1.31)$$

Рассмотрим механизм проверки многомерных гипотез. Пусть гипотеза записана в виде

$$K \Theta = T , \quad (1.32)$$

где K – известная $v \times m$ -матрица, T – известная $v \times l$ -матрица.

Предположим, что строки E распределены по многомерному нормальному закону. Введем обозначения: G – матрица остаточных сумм квадратов-произведений для модели без ограничения (1.32); H – изменение матрицы G , вызванное введением системы ограничений (1.32).

Для проверки гипотез используются различные статистики. Наиболее часто встречаются при этом следующие статистики [6, 89, 96, 189]:

- Λ -статистика Уилкса (статистика максимального правдоподобия) $\frac{|G|}{|G + H|}$;

- статистика следа Лоули–Хотеллинга $\text{tr } G^{-1} H$;

- статистика максимального собственного числа Роя $\lambda_{\max} G^{-1} H$,

где λ_{\max} – максимальное собственное число матрицы;

- статистика следа Пиллаи $\text{tr } G + H^{-1} H$.

В общем случае трудно определить распределения указанных статистик, поэтому на практике пользуются аппроксимациями распределений.

Приведем аппроксимацию Рао распределения статистики Уилкса для случая, когда имеется модель (1.19) с ограничениями (1.25) [89]. При этом не будем предполагать линейную независимость уравнений в системе (1.25) и в ней же, но дополненной уравнениями (1.32). Введем в рассмотрение величину

$$\left(\frac{1}{\Lambda^{1/\gamma}} - 1 \right) \frac{f_2}{f_1}, \quad (1.33)$$

где Λ – значение статистики Уилкса, $f_1 = l\alpha$, $f_2 = \beta\gamma - 2\delta$,

$$\alpha = \text{rg} \begin{bmatrix} R^T & | & K^T \end{bmatrix} - \text{rg} R, \quad \beta = N - m + \text{rg} \begin{bmatrix} R^T & | & K^T \end{bmatrix} - \text{rg} R - \frac{l + \alpha + 1}{2},$$

$$\gamma = \left[\frac{l^2 \alpha^2 - 4}{l^2 + \alpha^2 - 5} \right]^{1/2}, \quad \delta = \frac{l\alpha - 2}{4}. \text{ Величина (1.33) имеет } F\text{-распре-$$

деление со степенями свободы f_1 и f_2 .

Есть несколько частных случаев, в которых Λ -статистика Уилкса приводится к F -статистике с помощью некоторых преобразований [6, 96, 189]. В одном из таких случаев, когда $v=1$ и предполагается линейная независимость уравнений в системе (1.25) и в ней же, дополненной уравнениями (1.32), величина

$$\frac{N - m + u - l + 1}{l} \frac{1 - \Lambda}{\Lambda} \quad (1.34)$$

имеет F -распределение со степенями свободы l и $N - m + u - l + 1$.

В том случае, если состав регрессоров в уравнениях разных откликов различен, приходим к модели псевдонезависимых регрессий [18, 31, 189, 194]. Запишем для i -го отклика

$$Y_i = X_i \theta_i + E_i,$$

где Y_i – вектор значений i -го отклика, X_i – $N \times m_i$ -матрица регрессоров i -го отклика, θ_i – вектор размерности m_i параметров i -го отклика, E_i – вектор ошибок i -го отклика.

Обозначим через X матрицу, содержащую значения всех регрессоров, которые встречаются в уравнениях всех откликов. Тогда модель псевдонезависимых регрессий можно записать в виде (1.19) с априорно обнуленными элементами матрицы Θ , при регрессорах, отсутствующих в уравнениях соответствующих откликов.

Модель псевдонезависимых регрессий может быть представлена также в виде однооткликowej с коррелированными наблюдениями:

$$\begin{bmatrix} Y_1^T & \dots & Y_l^T \end{bmatrix}^T = \tilde{X}\theta + \begin{bmatrix} E_1^T & \dots & E_l^T \end{bmatrix}^T, \quad (1.35)$$

где $\tilde{X} = \text{diag } X_1, \dots, X_l$, $\theta = \begin{bmatrix} \theta_1^T & \theta_2^T & \dots & \theta_l^T \end{bmatrix}^T$, а ковариационная матрица вектора ошибок задается выражением (1.21).

МНК-оценка вектора θ в модели (1.35) является несмещенной, однако она, в отличие от случая традиционной многооткликowej модели, не наилучшая линейная несмещенная оценка. Чтобы получить наилучшую линейную несмещенную оценку для вектора θ , необходимо воспользоваться ОМНК. Заметим, что для вычисления оценки вектора θ необходимо обратить матрицу размерности $m_1 + m_2 + \dots + m_l$.

В ряде случаев оценки по МНК и ОМНК совпадают [18], в частности, это справедливо, когда матрица Σ имеет диагональный вид (отклики не коррелированы) и когда $X_1 = X_2 = \dots = X_l$ (случай традиционной многооткликowej модели).

На практике часто матрица Σ неизвестна и оценка для вектора θ по ОМНК не может быть получена. В этом случае возможно оценить θ по методу Зельнера, в котором неизвестное значение матрицы Σ заменяется некоторой ее состоятельной оценкой. В частности, при использовании в качестве оценки Σ МНК-оценки (1.24), полученной в рамках традиционной многооткликowej модели (1.19), которая не учитывает неодинаковость множеств регрессоров, описывающих различные отклики, получают оценку Зельнера без ограничений [189]. Для оценки Зельнера с ограничениями используется оценка вида [31, 189]:

$$\hat{\Sigma}_{ij} = \frac{1}{\sqrt{(N - m_i)(N - m_j)}} (Y_i - X_i \hat{\theta}_i)^T (Y_j - X_j \hat{\theta}_j).$$

В [18] рассматривается использование оценки вида

$$\hat{\Sigma}_{ij} = \frac{1}{N} (Y_i - X_i \hat{\theta}_i)^T (Y_j - X_j \hat{\theta}_j). \quad (1.36)$$

В условиях неизвестной матрицы Σ более точной для вектора θ будет итеративная оценка Зельнера [18], эквивалентная применению ММП к модели (1.35) с параметризацией ковариационной матрицы вида $V = \Sigma \otimes I_N$, $\sigma^2 = 1$.

Рассмотрим механизм проверки гипотез в предположении, что строки матрицы E распределены по многомерному нормальному закону. Используя возможность преобразования (1.7) многооткликовой модели в однооткликовую модель с некоррелированными наблюдениями, проверку гипотезы вида (1.17) при ограничениях на параметры вида (1.9) можно осуществить с помощью статистики (1.18). В условиях неизвестной ковариационной матрицы ошибок необходимо использовать ее оценку, при этом статистика (1.18) будет иметь F -распределение лишь асимптотически [194].

В традиционной многооткликовой модели и модели псевдонезависимых регрессий отклики были параметризованы независимо. Отказавшись от этого предположения, мы приходим к многооткликовой модели с общими параметрами вида [114]

$$y_{ij} = f_i^T(x_j)\theta + e_{ij}, \quad i=1, \dots, l, \quad j=1, \dots, N, \quad (1.37)$$

где i – номер отклика, j – номер наблюдения, $f_i(x)$ – вектор регрессоров i -го отклика, или в матричных обозначениях:

$$Y_i = X_i \theta + E_i, \quad i=1, \dots, l. \quad (1.38)$$

Заметим, что данную модель можно представить в виде модели псевдонезависимых регрессий с ограничениями на параметры, которые приравнивают друг другу параметры, стоящие при разных откликах. Однако с точки зрения вычислений для оценивания общих параметров выгоднее воспользоваться представлением (1.38). Заметим, что, как и модель псевдонезависимых регрессий, данную модель можно записать в виде (1.19) с некоторыми линейными ограничениями-равенствами на

параметры. Тем не менее, избегая формулировок, использующих ограничения на параметры, модель (1.38) удобно считать самым общим представлением многооткликковой модели, а традиционную многооткликковую модель и модель псевдонезависимых регрессий – ее частными случаями.

Для получения оценок общих параметров многооткликковую модель необходимо записать в виде однооткликковой с коррелированными наблюдениями:

$$\tilde{Y} = X\theta + \tilde{E}, \quad (1.39)$$

$$\text{где } \tilde{Y} = \begin{bmatrix} Y_1^T & \dots & Y_l^T \end{bmatrix}^T, \quad X = \begin{bmatrix} X_1^T & X_2^T & \dots & X_l^T \end{bmatrix}^T,$$

$$\tilde{E} = \begin{bmatrix} E_1^T & \dots & E_l^T \end{bmatrix}^T,$$

или с некоррелированными наблюдениями:

$$(S^{-1} \otimes I_N)\tilde{Y} = (S^{-1} \otimes I_N)X\theta + (S^{-1} \otimes I_N)\tilde{E}. \quad (1.40)$$

ОМНК-оценка вектора общих параметров, построенная в рамках модели (1.39), (или МНК-оценка в рамках модели (1.40)) – наилучшая линейная несмещенная оценка. В том случае, если матрица Σ неизвестна, при оценивании параметров и проверке линейных гипотез следует использовать подходы, применяемые для псевдонезависимых регрессий.

1.3. МОДЕЛЬ С КАЧЕСТВЕННЫМИ И РАЗНОТИПНЫМИ ФАКТОРАМИ

До сих пор мы не обращали внимание на природу факторов, от которых зависят отклики. Однако рассмотренные выше методы непосредственно можно использовать лишь в случае, когда факторы измерены в количественной шкале. Этот случай изучает регрессионный анализ. Модели с качественными факторами рассматриваются в дисперсионном анализе [6, 32, 122, 187], а модели с разнотипными факторами – в ковариационном анализе [116, 122].

Рассмотрим пример однооткликowej модели с качественными факторами. Пусть имеется два фактора с J и K уровнями варьирования соответственно. Запишем модель в виде

$$y_{jk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + e_{jk},$$

где $j=1, \dots, J$, $k=1, \dots, K$, μ – аддитивная постоянная, α_j – главный эффект j -го уровня первого фактора, β_k – главный эффект k -го уровня второго фактора, $(\alpha\beta)_{jk}$ – взаимодействие j -го уровня первого фактора и k -го уровня второго фактора. Данная модель содержит неизвестные параметры μ , α_j , β_k , $(\alpha\beta)_{jk}$, $j=1, \dots, J$, $k=1, \dots, K$, требующие оценки.

Чтобы использовать качественные факторы, их необходимо закодировать набором вспомогательных (фиктивных) переменных, каждая из которых может принимать только два возможных значения: 1 или 0. При этом число вспомогательных переменных определяется числом уровней варьирования качественных факторов. Так, если некоторый фактор x_i имеет v уровней варьирования $s_{i1}, s_{i2}, \dots, s_{iv}$, то в модель вводится v вспомогательных переменных $x_{i1}, x_{i2}, \dots, x_{iv}$. Вспомогательная переменная x_{ij} принимает значение 1 в случае $x_i = s_{ij}$ и 0 – в случае $x_i \neq s_{ij}$.

Рассмотрим многооткликową модель вида

$$Y = X\Theta + Z\Delta + E, \quad (1.41)$$

где Y – $N \times l$ -матрица N измерений l откликов системы; X – $N \times t$ -матрица N значений t фиктивных переменных, связанных с качественными факторами; Z – $N \times m$ -матрица N значений m регрессоров, выражаемых в количественной шкале; Θ и Δ – $t \times l$ - и $m \times l$ -матрицы неизвестных параметров; E – $N \times l$ -матрица неизвестных значений l -мерного вектора ошибок наблюдений.

Система уравнений (1.41) представляет собой модель многомерно-го ковариационного анализа. Слагаемое $Z\Delta$ задает регрессионную часть модели, а слагаемое $X\Theta$ – дисперсионную.

Можно рассмотреть общую модель, включающую взаимодействия количественных и качественных факторов [31, 91]. В дальнейшем будем предполагать, что эти взаимодействия включены в дисперсионную часть.

Многомерную ковариационную модель (1.41) можно представить традиционной многооткликовой моделью специального вида. Как и в случае регрессионных моделей, зависимость откликов от качественных и разнотипных факторов может строиться также в виде моделей псевдонезависимых регрессий и моделей с общими параметрами, если априорно наложить некоторые ограничения на элементы матриц Θ и Δ .

Отличительная черта модели (1.41) – неполнота ранга матрицы X . Для получения единственного решения при оценивании по МНК используют обобщенное обращение матриц [19, 32, 187], приведение исходной модели к регрессионной модели полного ранга [28, 31, 91, 106], введение идентифицирующих ограничений [106, 122].

Отметим, что при неполном ранге матрицы X оценки параметров уже не будут несмещенными. В связи с этим особый интерес вызывают некоторые линейные комбинации параметров, которые называются функциями, допускающими оценку (ФДО).

В модели (1.1) параметрическая функция $c^T\theta$, где c – вектор постоянных коэффициентов, называется функцией, допускающей оценку, если для нее существует такой вектор a , что $\mathbf{E} a^T Y = c^T \theta$.

Для модели (1.1) с ограничениями (1.9) функция $c^T\theta$ называется функцией, допускающей оценку, если существует линейная функция $a^T Y + d$ такая, что $\mathbf{E} a^T Y + d = c^T \theta$ при $R\theta = d$.

Исходную модель (1.41) можно привести к модели полного ранга, воспользовавшись факторизацией матрицы X вида $X = X_0 X_1$ [73, 91]. Получим

$$Y = X_0 X_1 \Theta + Z \Delta + E = X_0 \tilde{\Theta} + Z \Delta + E, \quad (1.42)$$

при этом матрица X_0 имеет полный столбцовый ранг, матрица X_1 имеет полный строчный ранг, матрица $\tilde{\Theta} = X_1 \Theta$ содержит значения системы линейно независимых ФДО.

Ограничения

$$H \begin{bmatrix} \Theta^T & \Delta^T \end{bmatrix}^T = 0,$$

где H – матрица размерности $u \times (t+m)$, называются идентифицирующими, если выполняется следующее условие [122]: для каждой возможной матрицы параметров $\begin{bmatrix} \Theta^T & \Delta^T \end{bmatrix}^T$ модели (1.41) существует и единственна матрица $\begin{bmatrix} \Theta_H^T & \Delta_H^T \end{bmatrix}^T$, удовлетворяющая идентифицирующим ограничениям и условию $X\Theta + Z\Delta = X\Theta_H + Z\Delta_H$, при этом $u \geq m + t - \text{rg } X \begin{bmatrix} X \\ Z \end{bmatrix}$.

1.4. ОЦЕНИВАНИЕ ПО ДАННЫМ С ПРОПУСКАМИ

В практике регрессионного моделирования встречаются ситуации, когда отклики наблюдаются с пропусками. Если наблюдения откликов в наборе данных зависимы, для их обработки требуются специальные процедуры. Один из подходов состоит в применении такой модификации метода максимального правдоподобия, что функция правдоподобия строится на основе маргинальных распределений наблюдений, реально присутствующих в наборе данных [78]. Для вычисления оценок в данном случае могут потребоваться алгоритмы, отличные от использовавшихся при полных наборах данных. Однако последние часто обладают положительными свойствами, которые желательно сохранить и в рассматриваемом случае.

Решить возникающую задачу можно, предварительно восстановив пропущенные значения и обработав затем получающийся полный набор данных. Различные методы восстановления пропущенных значений описаны в [78], локальные алгоритмы, относящиеся к семейству алгоритмов ZET, – в работе [36].

К указанному подходу близки EM-алгоритмы, в которых, однако, восстанавливаются функции пропущенных данных, входящие в функцию правдоподобия полных данных [78, 133]. EM-алгоритмы исполь-

зуются для поиска ММП-оценок параметров широкого круга моделей, в том числе одномерных и многомерных распределений и регрессионных моделей [78, 156, 158–160, 168]. Рассмотрим ряд модификаций EM-алгоритмов.

При отсутствии пропущенных наблюдений ММП-оценки доставляют максимальное значение логарифму функции правдоподобия

$$Q(\Phi) = \ln g(Y|\Phi) .$$

При наличии пропущенных наблюдений, механизм возникновения которых можно игнорировать [78], для получения ММП-оценок следует максимизировать функцию

$$L(\Phi) = \ln g(Y_{obs}|\Phi) , \quad (1.43)$$

где $g(Y_{obs}|\Phi)$ – плотность распределения присутствующих наблюдений, Y_{obs} – совокупность значений присутствующих наблюдений.

Введем функцию

$$Q(\Phi|\tilde{\Phi}) = \mathbf{E} \ln g(Y|\Phi | Y_{obs}, \tilde{\Phi}) . \quad (1.44)$$

EM-алгоритм представляет собой итерационный процесс, каждая s -я итерация которого состоит из двух шагов: шага E (expectation) и шага M (maximization). На шаге E определяется функция $Q(\Phi|\Phi^{(s-1)})$

аргумента Φ . На шаге M находится значение $\Phi^{(s)}$ как решение задачи $\max_{\Phi} Q(\Phi|\Phi^{(s-1)})$.

Каждая итерация EM-алгоритма увеличивает значение функции $L(\Phi)$, и EM-алгоритм сходится в общем случае к стационарному значению функции $L(\Phi)$. Обычно рекомендуется выполнять EM-алгоритм несколько раз с разными начальными значениями параметров.

В том случае, если решение на шаге M невозможно получить в явном виде (требуется итерационный процесс), может использоваться обобщенный EM-алгоритм – GEM-алгоритм (generalized EM), в кото-

ром шаг М состоит в поиске значения $\Phi^{(s)}$, удовлетворяющего условию

$$Q \Phi^{(s)} | \Phi^{(s-1)} \geq Q \Phi^{(s-1)} | \Phi^{(s-1)} .$$

Разновидностью GEM-алгоритма является ЕСМ-алгоритм [168], в котором шаг М, являющийся слишком сложным в рамках EM-алгоритма, заменяется последовательностью CM-шагов условной (ограниченной) максимизации (conditional/constrained maximization). Каждый из CM-шагов максимизирует функцию $Q \Phi | \Phi^{(s-1)}$ по Φ при фиксированных значениях некоторых функций от Φ . Последовательность CM-шагов при этом должна обеспечивать максимизацию функции L во всем параметрическом пространстве.

Опишем алгоритм более подробно. Пусть $h_t(\Phi)$, $t=1, \dots, T$ – множество T предварительно выбранных вектор-функций. На s -й итерации, после проведения шага E, ЕСМ-алгоритм производит T CM-шагов, которые определяются следующим образом. Для каждого $t=1, \dots, T$ находится значение $\Phi^{(s-1+t/T)}$, максимизирующее функцию $Q(\Phi | \Phi^{(s-1)})$ по Φ при ограничении

$$h_t(\Phi) = h_t \Phi^{(s-1+(t-1)/T)} .$$

В простейшем случае множество параметров разбивается на несколько подмножеств, и максимизация на некотором CM-шаге производится по одному из подмножеств при фиксированных значениях остальных параметров. Подобную стратегию удобно применять, если оптимизация по отдельным подмножествам параметров более проста, чем оптимизация по всем параметрам. Указанный подход аналогичен описанному в п. 1.1 методу покомпонентного оценивания.

Преимущество ЕСМ-алгоритма – в наличии на CM-шаге либо аналитического решения, либо задачи меньшей размерности (часто одномерной). ЕСМ-алгоритм требует, как правило, большего числа итераций по сравнению с EM-алгоритмом, однако по времени может оказаться быстрее.

Модификация EM-алгоритма, предложенная в [159], называется ЕСМЕ-алгоритмом (ЕСМ either). В этом случае некоторые (или все) CM-шаги ЕСМ-алгоритма, которые называются здесь CMQ-шагами, заменяются CM-шагами условной максимизации исходной функции $L(\Phi)$ (CML-шагами), причем перед последовательностью CMQ-шагов должен быть выполнен шаг E.

Ускорение сходимости достигается здесь, во-первых, за счет замены функции $Q(\Phi | \Phi^{(s-1)})$, которая является текущей аппроксимацией функции $L(\Phi | Y_{obs})$, исходной функцией $L(\Phi | Y_{obs})$; во-вторых, за счет возможности выбора численных методов, наиболее эффективных в каждом конкретном случае. ЕСМЕ-алгоритм почти всегда требует меньшего числа итераций, чем EM- и ЕСМ-алгоритмы, и может оказаться быстрее их по времени.

Для ускорения сходимости предназначен и PX-EM-алгоритм (parameter expansion) [162]. В PX-EM-алгоритме вводится расширенная модель полных данных, которая имеет больший набор параметров, но приводит к первоначальной модели наблюдаемых данных с первоначальными параметрами, определяемыми по расширенному набору параметров через некоторую функцию редукции. PX-EM-алгоритм итеративно максимизирует ожидаемый логарифм правдоподобия полных данных расширенной модели, при этом скорость сходимости у PX-EM-алгоритма не меньше, чем у EM-алгоритма. Рассмотрим PX-EM-алгоритм подробнее.

Модель $g(Y | \Phi)$ полных данных заменяется расширенной моделью $g_*(Y | \tilde{\Phi})$ с вектором параметров $\tilde{\Phi} = \Phi_*^T, \gamma^T$, где Φ_* играет ту же роль в $g_*(Y | \tilde{\Phi})$, что Φ в $g(Y | \Phi)$; γ – дополнительный параметр, значение которого в исходной модели фиксировано и равно γ_0 .

При построении расширенной модели необходимо соблюдать два условия:

1) не должна изменяться модель присутствующих данных в том смысле, что для всех $\tilde{\Phi}$ существует функция редукции r такая, что плотность распределения $Y_{obs} | \tilde{\Phi}$ есть $g[Y_{obs} | \Phi = r(\tilde{\Phi})]$;

2) не должна изменяться при $\gamma = \gamma_0$ модель полных данных в том смысле, что для всех Φ выполняется $g_* \left[Y | \tilde{\Phi} = \Phi_*^T, \gamma_0^T \right] = g \left[Y | \Phi = \Phi_* \right]$.

Эти условия определяют, что из $\Phi_1 \neq \Phi_2$ следует $\tilde{\Phi}_1 \neq \tilde{\Phi}_2$ и что для всех Φ существует хотя бы одно значение $\tilde{\Phi}$ такое, что плотность распределения $Y_{obs} | \tilde{\Phi}$ равна $g \left[Y_{obs} | \Phi = r(\tilde{\Phi}) \right]$.

Пусть $\Phi^{(0)}$ – начальное приближение вектора Φ , тогда s -я итерация PX-EM-алгоритма состоит из следующих шагов:

- шаг PX-E, на котором определяется функция

$$Q_* \tilde{\Phi} | \Phi^{(s-1)} = \mathbf{E} \left\{ \ln g_* \left[Y | \tilde{\Phi} \right] \left| Y_{obs}, \begin{bmatrix} \Phi^{(s-1)} \\ \gamma_0 \end{bmatrix} \right. \right\};$$

- шаг PX-M, на котором находится

$$\tilde{\Phi}^{(s)} = \arg \max_{\tilde{\Phi}} Q_* \tilde{\Phi} | \Phi^{(s-1)},$$

после чего производится редукция

$$\Phi^{(s)} = r \left[\tilde{\Phi}^{(s)} \right].$$

Источник более быстрой сходимости PX-EM-алгоритма состоит в следующем. Пусть $\Phi^{(s-1)}$ – текущее приближение к Φ при $\gamma^{(s-1)} = \gamma_0$, а $\Phi_{EM}^{(s)}$ и $\Phi_{PX}^{(s)}$ – модификации, произведенные EM- и PX-EM-алгоритмами соответственно. Тогда справедливо

$$\Phi_{PX}^{(s)} \approx \Phi_{EM}^{(s)} + B \left(\gamma^{(s)} - \gamma_0 \right), \quad (1.45)$$

где $B = V_{\Phi\gamma} V_{\gamma\gamma}^{-1}$,

$$V = \begin{pmatrix} V_{\Phi\Phi} & V_{\Phi\gamma} \\ V_{\gamma\Phi} & V_{\gamma\gamma} \end{pmatrix} = \left\{ \mathbf{E} \left[- \frac{\partial}{\partial \tilde{\Phi} \partial \tilde{\Phi}^T} \ln g_*(Y | \tilde{\Phi}) \Big| Y_{obs}, \tilde{\Phi} \right] \Big|_{\tilde{\Phi} = (\hat{\Phi}, \gamma_0)} \right\}^{-1}.$$

Выражение (1.45) в явном виде определяет корректировку, производимую РХ-ЕМ-алгоритмом посредством извлечения дополнительной информации, находящейся в оценке полных данных. Шаги Е-ЕМ- и РХ-ЕМ-алгоритмов эффективно оценивают отсутствующие данные при неправильной модели с $\Phi^{(s)} \neq \hat{\Phi}$. И если на шаге М-ЕМ-алгоритма этот эффект игнорируется, то на шаге М-РХ-ЕМ-алгоритма используется дополнительный параметр: для коррекции очередного приближения ЕМ-алгоритма используется разность между очередным приближением для параметра γ и его ММП-оценкой γ_0 (являющейся также истинным значением) $\gamma^{(s)} - \gamma_0$ с множителем B , являющимся коэффициентом регрессии Φ на γ .

1.5. М-ОЦЕНКИ

Рассмотрим понятие M -оценки (оценки типа максимального правдоподобия) параметров однооткликowej регрессионной модели (1.1).

M -оценка $\hat{\theta}$ вектора параметров θ находится путем минимизации функции вида [2, 18, 118, 120]

$$Q(\theta) = \sum_{i=1}^N \rho \left(\frac{r_i(\theta)}{\lambda} \right), \quad (1.46)$$

где $\rho(z)$ – функция потерь, $r_i(\theta) = y_i - f^T(x_i)\theta$ – остаток i -го наблюдения, $\lambda > 0$ – параметр масштаба.

Приведем примеры M -оценок:

- $\rho(z) = z^2$ – МНК-оценка;
- $\rho(z) = |z|$ – оценка по методу наименьших модулей;
- $\rho(z) = |z|^\nu$ – L_ν -оценка, где $\nu > 0$ – фиксированное значение;

- $\rho(z) = \begin{cases} -(1-\gamma)z, & z < 0 \\ \gamma z, & z \geq 0 \end{cases}$ – оценка γ -й квантильной регрессии.

Частным случаем M -оценки является ММП-оценка.

При соответствующем выборе функции потерь получающиеся M -оценки могут быть устойчивыми. Например, при наличии выбросов в наборе данных МНК-оценка неустойчива, а оценка по методу наименьших модулей, наоборот, устойчива.

При альтернативном определении M -оценка находится как решение системы m уравнений

$$\sum_{i=1}^N \psi \left(\frac{r_i(\hat{\theta})}{\lambda} \right) f_j(x_i) = 0, \quad j = 1, \dots, m, \quad (1.47)$$

где функция ψ называется оценочной [125].

Если функция $\rho(z)$ дифференцируема, то, приравняв нулю частные производные функции (1.46) по элементам вектора θ в точке $\theta = \hat{\theta}$, получим систему оценочных уравнений, эквивалентную (1.47) при $\psi(z) = \rho'(z)$.

Предположим выполненным условие

$$\mathbf{E}\psi = 0, \quad (1.48)$$

где математическое ожидание берется по плотности модельного (принятого в модели) распределения ошибок наблюдений. Данное условие влечет, в частности, асимптотическую несмещенность оценки $\hat{\theta}$ [107].

При некоторых условиях регулярности оценка $\hat{\theta}$ является состоятельной и асимптотически нормальной [9, 107].

В данной постановке задачи параметр λ предполагается известным. Если же это не так, то параметр λ приходится оценивать, например, используя некоторое уравнение:

$$\sum_{i=1}^N \chi(r_i(\hat{\theta}), \hat{\lambda}) = 0. \quad (1.49)$$

При этом функции $\psi(z)$ и $\chi(z)$ не обязательно должны быть связаны с одной и той же функцией потерь.

Так, на практике часто предпочитают в качестве устойчивой оценки λ для нормального распределения использовать медиану абсолютных отклонений (MAD-оценку)

$$\hat{\lambda} = \frac{\text{med}_{i=1, \dots, N} |r_i(\hat{\theta})|}{0,67449}, \quad (1.50)$$

где med – медиана соответствующей последовательности.

Для поиска значений M -оценок можно применять общие методы оптимизации: метод Ньютона, метод сопряженных градиентов и т. д. Опишем один специальный метод минимизации – итеративный МНК (ИМНК), основанный на итеративном использовании ОМНК [18, 118].

Перепишем левую часть системы уравнений (1.47) в следующем виде:

$$\sum_{i=1}^N \left[\psi \left(\frac{r_i(\hat{\theta})}{\lambda} \right) \frac{\lambda}{r_i(\hat{\theta})} \right] f_j(x_i) \frac{r_i(\hat{\theta})}{\lambda} = \sum_{i=1}^N \frac{1}{\lambda} w_{r_i(\hat{\theta}), \lambda} f_j(x_i) r_i(\hat{\theta}),$$

где w – весовая функция вида $w_{r_i(\hat{\theta}), \lambda} = \psi \left(\frac{r_i(\hat{\theta})}{\lambda} \right) \frac{\lambda}{r_i(\hat{\theta})}$.

Переходя к матричным обозначениям и опуская константный множитель $\frac{1}{\lambda}$, мы можем переписать полученную систему следующим образом:

$$X^T W(\hat{\theta}) Y - X \hat{\theta} = 0,$$

где $W(\hat{\theta})$ – диагональная весовая матрица, i -й диагональный элемент которой равен $w_{r_i(\hat{\theta}), \lambda}$. В результате имеем

$$\hat{\theta} = \left[X^T W(\hat{\theta}) X \right]^{-1} X^T W(\hat{\theta}) Y. \quad (1.51)$$

Система уравнений (1.51) соответствует схеме ОМНК с ковариационной матрицей, зависящей от оценок $\hat{\theta}$. Для ее решения воспользу-

емя методом последовательных приближений, на $(s+1)$ -й итерации которого производится вычисление

$$\hat{\theta}^{(s+1)} = \left[X^T W(\hat{\theta}^{(s)}) X \right]^{-1} X^T W(\hat{\theta}^{(s)}) Y,$$

составляющее суть ИМНК.

Веса наблюдений на каждой итерации ИМНК «оцениваются» на основе оценки вектора параметров, полученной из предыдущей итерации. Окончательный вес, полученный на последней итерации, указывает, принадлежит ли данное наблюдение к выбросам.

Укажем достаточные условия сходимости ИМНК [81]:

- 1) функция потерь $\rho(z)$ четна, т. е. $\rho(z) = \rho(-z)$;
- 2) весовая функция $w(z, \lambda)$ является невозрастающей при $z \geq 0$.

В случае одновременного оценивания вектора параметров θ и параметра масштаба λ следует дополнить каждую итерацию ИМНК уточнением или вычислением оценки $\hat{\lambda}$, например, в соответствии с формулами (1.49), (1.50).

Основным недостатком применения ИМНК является то, что получающееся решение соответствует лишь локальному минимуму функции (1.46). Если функция потерь – выпуклая, то функция (1.46) имеет единственный минимум и в качестве начального приближения $\hat{\theta}$ можно взять МНК-оценку. При невыпуклой функции потерь может быть несколько локальных минимумов, соответствующих нескольким решениям системы (1.47) или (1.47), (1.49) (поэтому определения M -оценки через формулы (1.46) и (1.47) не эквивалентны). В силу этого выбор начального приближения здесь особенно важен.

1.6. УСТОЙЧИВОЕ ОЦЕНИВАНИЕ ОДНОМЕРНЫХ МОДЕЛЕЙ

Часто используемые на практике МНК-оценки имеют ряд положительных свойств. Они являются наилучшими линейными несмещенными оценками. В условиях нормальных ошибок МНК-оценки вектора θ совпадают с ММП-оценками. Однако при нарушении предположения нормальности МНК-оценки становятся неустойчивыми.

Заметим, что нормальное распределение имеет легкие хвосты – основная масса распределения сосредоточена на интервале $(-3\sigma, 3\sigma)$. На практике в наборе данных могут присутствовать грубые ошибки (выбросы), возникающие вследствие нарушения условия эксперимента, неправильного измерения, засорения данных и т. п. В таких случаях целесообразно перейти к использованию распределений с тяжелыми хвостами.

С другой стороны, особенности данных могут быть таковы, что более адекватным окажется предположение об ошибках наблюдений, имеющих хвосты более легкие, чем у нормального распределения.

В качестве устойчивой оценки при тяжелых хвостах может использоваться L_ν -оценка с $\nu < 2$, которая при $\nu = 1$ является оценкой по методу наименьших модулей; при наличии легких хвостов может оказаться полезной L_ν -оценка с $\nu \geq 2$, которая при $\nu = 2$ является МНК-оценкой, а при $\nu = \infty$ – чебышевской [2, 18]; оценки квантильной регрессии могут использоваться при наличии асимметричного распределения ошибок.

Среди различных подходов к оцениванию прежде всего упомянем три: непараметрический, робастный и адаптивный.

Непараметрические методы строятся безотносительно к конкретному виду распределения ошибок наблюдений, которое должно удовлетворять лишь некоторым весьма общим предположениям. Часто непараметрические оценки оказываются устойчивыми. Среди непараметрических упомянем ранговые [117] и знаковые методы [8, 29].

При **робастном оценивании** используются процедуры, устойчивые в окрестности параметрической модели [99]. В наиболее важном (и наиболее изученном) случае рассматриваются отклонения распределения наблюдений от постулируемого модельного («истинного») распределения. Для характеристики робастной процедуры необходимо указать модельное распределение, вид его окрестности и критерий качества (устойчивости) решения.

Простейшим видом окрестности, используемой в теории робастности, является множество засоренных распределений с функцией

$$\tilde{G}(z) = (1 - \alpha)G(z) + \alpha H(z),$$

где α – известный уровень засорения (доля аномальных наблюдений), $0 < \alpha < 1$; $G(z)$ – известная функция модельного распределения;

$H(z)$ – неизвестная функция засоряющего распределения из некоторого множества функций. Таким образом, предполагается, что основная часть наблюдений имеет одно, например, нормальное, распределение ошибок, а ряд наблюдений имеет другое (засоряющее) распределение.

Можно выделить два основных подхода к получению робастных оценок: минимаксный подход Хьюбера [81, 107, 120] и локальный подход Хампеля [99, 123].

Основными количественными мерами робастности в подходе Хампеля являются функция влияния и пороговая точка – локальный и глобальный показатели робастности соответственно. Функция влияния рассматривается в п. 2.1, она связана со смещением, вызываемым добавлением в очень большую выборку одного наблюдения. Пороговая точка характеризует максимально возможное отклонение от модельного распределения, при котором смещение оценки остается ограниченным [123].

В табл. 1.1 представлены функции $\rho(z)$ и $\psi(z)$ ряда робастных M -оценок в модели засоренной нормальной однооткликковой регрессии [2, 18, 35, 99, 110, 118, 120]. Все представленные функции имеют один или несколько неочисляемых положительных параметров – a, b, c .

Оценка Хьюбера, являющаяся решением в рамках и минимаксного, и локального подходов, устойчива только при симметричности засоряющего распределения. Устойчивыми к асимметричному засорению оказываются M -оценки с функциями потерь $\rho(z)$, при $|z| \rightarrow \infty$ растущими медленнее функции $|z|$ [2]; это справедливо, например, для функций потерь, имеющих горизонтальную асимптоту. Среди таких оценок – предложенная Л.Д. Мешалкиным оценка экспоненциально-взвешенной регрессии (ЭВ-оценка). Заметим, что в подходе Мешалкина используется оценка σ^2 вида [2]:

$$\hat{\sigma}^2 = (1+c) \left\{ \sum_{i=1}^N \exp \left[-\frac{c}{2\hat{\sigma}^2} r_i^2(\hat{\theta}) \right] \right\}^{-1} \sum_{i=1}^N \exp \left[-\frac{c}{2\hat{\sigma}^2} r_i^2(\hat{\theta}) \right] r_i^2(\hat{\theta}).$$

К робастным решениям приводит и подход Шурыгина [125, 126], развитию этого подхода посвящена основная часть гл. 2. В нем рассматривается байесовское точечное засорение модельной плотности распределения. Особенностью данного подхода является наличие ре-

шений, устойчивых к асимметричному засорению, в то время как в рамках подходов Хьюбера и Хампеля получить решения с этим свойством не удастся. Устойчивая к асимметричному засорению оценка Мешалкина оптимальна с точки зрения подхода Шурыгина.

В работах [50–52] показывается робастность (в том числе с точки зрения функции влияния) ММП-оценок параметров распределений по группированным данным.

При *адаптивном оценивании* конкретный вид статистической процедуры выбирается на основе оценки какой-либо характеристики неизвестной функции распределения наблюдений [110].

Например, Р.В. Хогг предложил следующую процедуру оценивания вектора параметров регрессии θ [35, 145]:

1) находится достаточно устойчивая предварительная оценка вектора θ , на которую не слишком влияют большие выбросы, например, оценка по методу наименьших модулей или оценка Хьюбера;

2) вычисляются остатки модели, построенной при помощи предварительной оценки, и определяется тяжесть хвостов распределения остатков, например, посредством некоторой оценки коэффициента эксцесса;

3) строится окончательная оценка вектора θ , использующая информацию о тяжести хвостов распределения остатков.

Помимо тяжести хвостов, можно использовать характеристику асимметрии распределения (см., например, [172]).

Таблица 1.1

Робастные оценки

1. Название	2. $\rho(z)$	3. $\psi(z)$
Хьюбера	$\begin{cases} z^2/2, & z < c \\ c z - c^2/2, & z \geq c \end{cases}$	$\begin{cases} z, & z < c \\ c \operatorname{sign}(z), & z \geq c \end{cases}$
Эндрюса	$\begin{cases} c \left(1 - \cos \frac{z}{c} \right), & z < \pi c \\ 2c, & z \geq \pi c \end{cases}$	$\begin{cases} \sin \frac{z}{c}, & z < \pi c \\ 0, & z \geq \pi c \end{cases}$

Окончание табл. 1.1

1. Название	2. $\rho(z)$	3. $\psi(z)$
Рамсея	$1 - (1 + c z) \exp[-c z] / c^2$	$cz \exp[-c z]$
Тьюки	$\begin{cases} z^2/2, & z < c \\ c^2/2, & z \geq c \end{cases}$	$\begin{cases} z, & z < c \\ 0, & z \geq c \end{cases}$
Биквадратная Тьюки	$\begin{cases} \frac{z^6}{6c^4} - \frac{z^4}{2c^2} + \frac{z^2}{2}, & z < c \\ \frac{c^2}{6}, & z \geq c \end{cases}$	$\begin{cases} z \left[1 - \left(\frac{z}{c} \right)^2 \right]^2, & z < c \\ 0, & z \geq c \end{cases}$
Меррилла–Швеппе	$\begin{cases} z^2/2, & z < c \\ \sqrt{c^3 z }/2, & z \geq c \end{cases}$	$\begin{cases} z, & z < c \\ \sqrt{\frac{c^3}{ z }} \frac{\text{sign}(z)}{4}, & z \geq c \end{cases}$
Мешалкина	$-\sigma^{-\frac{c}{1+c}} \exp[-cz^2/2]$	$cz\sigma^{-\frac{c}{1+c}} \exp[-cz^2/2]$
Энскомба	$\begin{cases} z^2/2, & z < b \\ b z - b^2/2, & b \leq z < c \\ bc - b^2/2, & z \geq c \end{cases}$	$\begin{cases} z, & z < b \\ b \text{sign}(z), & b \leq z < c \\ 0, & z \geq c \end{cases}$
Хампеля	$\begin{cases} z^2/2, & z < a \\ a z - a^2/2, & a \leq z < b \\ \frac{ac z }{c-b} - \frac{a(z^2+b^2)}{2(c-b)} - \frac{a^2}{2}, & b \leq z < c \\ \frac{a}{2}(c+b-a), & z \geq c \end{cases}$	$\begin{cases} z, & z < a \\ a \text{sign}(z), & a \leq z < b \\ \frac{az}{c-b} \left(\frac{c}{ z } - 1 \right), & b \leq z < c \\ 0, & z \geq c \end{cases}$

Можно выбрать одну оценку из конечного набора. Другой подход состоит в построении квазиправдоподобных оценок при использовании некоторого параметризованного по форме семейства распределений. Благодаря наличию оцениваемого параметра формы имеется возможность для методов оценивания адаптироваться к свойствам ошибок измерений.

Среди таких семейств назовем двустороннее экспоненциальное (обобщенное гауссовское) распределение с плотностью

$$g(z, \lambda, \nu) = \frac{1}{2\lambda\Gamma(1+1/\nu)} \exp \left[- \left(\frac{|z|}{\lambda} \right)^\nu \right], \quad (1.52)$$

где $\lambda > 0$ – параметр масштаба, $\nu > 0$ – параметр формы, Γ – гамма-функция [7, 16, 152].

Частными случаями двустороннего экспоненциального распределения являются нормальное (при $\nu = 2$), Лапласа (при $\nu = 1$) и равномерное (при $\nu = \infty$). При $\nu < 2$ распределение имеет хвосты, более тяжелые, чем у нормального распределения, а при $\nu > 2$ – более легкие. Заметим, что двустороннее экспоненциальное распределение часто можно считать адекватным для описания ошибок измерений, встречающихся на практике [84].

ММП-оценка вектора параметров θ при двустороннем экспоненциальном распределении называется адаптивной L_ν -оценкой [84, 97], параметр ν при этом является уже нефиксированным, его значение оценивается по ММП. Одно из преимуществ адаптивных L_ν -оценок – возможность адаптироваться к распределениям ошибок наблюдений как с тяжелыми, так и с легкими хвостами. Подробно данный подход рассмотрен в п. 2.4.

Часто при оценивании параметров регрессионных моделей используется также распределение Стьюдента [156], приводящее к оценкам, относимым одновременно к адаптивным и робастным (см. п. 1.7).

Минимаксному подходу Хьюбера аналогичен подход к оцениванию, в основе которого лежит принцип оптимальности на классе [121]. Вместо окрестности некоторого модельного распределения здесь рассматривается класс распределений, возникающих из анализа априорной информации о характере ошибок.

Плотность распределения, которой соответствует наименьшая в классе информация Фишера о параметре сдвига μ

$$\int_Z \left(\frac{\partial}{\partial \mu} \ln g(z, \mu) \right)^2 g(z, \mu) dz,$$

где Z – область определения исследуемой случайной величины, называется наименее благоприятной. В соответствии с принципом оптимальности на классе для оценивания параметра сдвига (или параметров регрессии) необходимо использовать ММП-оценку при наименее благоприятной плотности ошибок наблюдений.

Для ошибок наблюдений в [121] рассмотрен ряд подклассов класса симметричных непрерывных унимодальных распределений, имеющих конечную информацию Фишера, и приведены соответствующие наименее благоприятные плотности.

Рассмотрим наименее благоприятные плотности распределения для классов, которые будут использоваться в дальнейшем.

1. В классе *засоренных нормальных распределений* наименее благоприятная плотность имеет вид [121]

$$g(z, \sigma, \alpha) = \begin{cases} \frac{1-\alpha}{\sqrt{2\pi}\sigma} \exp\left[-\frac{z^2}{2\sigma^2}\right], & \left|\frac{z}{\sigma}\right| \leq c, \\ \frac{1-\alpha}{\sqrt{2\pi}\sigma} \exp\left[-c\frac{|z|}{\sigma} + \frac{c^2}{2}\right], & \left|\frac{z}{\sigma}\right| > c, \end{cases}$$

где величина c определяется из уравнения

$$\Phi(c) = \frac{1}{2(1-\alpha)} - \frac{1}{\sqrt{2\pi}c} \exp\left(-\frac{c^2}{2}\right),$$

$$\Phi(c) = \frac{1}{\sqrt{2\pi}} \int_0^c \exp\left(-\frac{z^2}{2}\right) dz.$$

2. В классе *засоренных лапласовских распределений* наименее благоприятная плотность также является лапласовской с параметром масштаба $\frac{\lambda}{1-\alpha}$ [121].

3. В классе *засоренных двусторонних экспоненциальных распределений* наименее благоприятная плотность имеет вид [152]

$$g(z, \lambda, \nu, \alpha) = \begin{cases} \frac{1-\alpha}{2\lambda\Gamma(1+1/\nu)} \exp\left[-\frac{|z|^\nu}{\lambda^\nu}\right], & \left|\frac{z}{\lambda}\right| \leq c, \\ \frac{1-\alpha}{2\lambda\Gamma(1+1/\nu)} \exp\left[-\nu c^\nu \left(\frac{|z|}{\lambda c} + \frac{1-\nu}{\nu}\right)\right], & \left|\frac{z}{\lambda}\right| > c, \end{cases} \quad (1.53)$$

где величина c определяется из уравнения

$$\gamma\left(\frac{1}{\nu}, c^\nu\right) + \frac{e^{-c^\nu}}{c^{\nu-1}} = \frac{\Gamma(1/\nu)}{1-\alpha},$$

$\gamma(a, x) = \int_0^x z^{a-1} e^{-z} dz$ – неполная гамма-функция, $\nu > 1$.

4. Класс *невыврожденных распределений* определяется ограничением на значение плотности распределения при $z=0$ вида

$$g(0) \geq \frac{1}{2s} > 0.$$

Наименее благоприятная плотность для него является лапласовской с параметром масштаба s [121].

5. Класс *распределений с ограниченной дисперсией* определяется условием

$$\sigma^2 \leq \Delta.$$

Наименее благоприятная плотность для него является нормальной с дисперсией Δ [121].

6. Класс *финитных распределений* определяется ограничением на значения плотности распределения вида

$$\int_{-\lambda}^{\lambda} g(z) dz = 1.$$

Наименее благоприятная плотность для него имеет вид [121]

$$g(z, \lambda) = \begin{cases} \frac{1}{\lambda} \cos^2\left(\frac{\pi z}{2\lambda}\right), & |z| \leq \lambda, \\ 0, & |z| > \lambda. \end{cases}$$

7. Класс *приближенных финитных распределений* определяется ограничением на значения плотности распределения вида [13]

$$\int_{-\lambda}^{\lambda} g(z) dz \geq 1 - \nu,$$

где параметр ν характеризует степень отступления от финитности. Наименее благоприятная плотность для него имеет вид

$$g(z, \lambda, \nu) = \begin{cases} \frac{\beta_2}{\lambda} \cos^2 \frac{\beta_1 z}{\lambda}, & |z| \leq \lambda, \\ \frac{\beta_3}{\lambda} \exp\left(-\frac{\beta_4 |z|}{\lambda}\right), & |z| \geq \lambda, \end{cases}$$

где константы β_1, \dots, β_4 – неотрицательные числа, которые зависят от ν и определяются из уравнений

$$1 + \beta_1 \operatorname{tg} \beta_1 - \frac{1}{\nu} \cos^2 \beta_1 = 0, \quad \beta_4 = 2\beta_1 \operatorname{tg} \beta_1, \quad \beta_3 = \frac{\nu \beta_4}{2} \exp \beta_4,$$

$$\beta_2 = \frac{\beta_4}{2 + \beta_4}.$$

1.7. УСТОЙЧИВОЕ ОЦЕНИВАНИЕ МНОГОМЕРНЫХ МОДЕЛЕЙ

Если наблюдения представляют собой реализации зависимых случайных величин, необходимо использовать многомерные модели. При оценивании регрессии применяется большое количество различных одномерных модельных распределений ошибок наблюдений; при использовании многомерных вероятностных моделей спектр модельных распределений не так широк – в основном это нормальное распределение, а также распределение Стьюдента.

Плотность распределения Стьюдента имеет вид

$$g(z, \Sigma, \nu) = \frac{\nu^{\nu/2} \Gamma(\nu + l)/2}{\pi^{l/2} \Gamma(\nu/2)} |\Sigma|^{-1/2} \nu + q^{-(\nu+l)/2},$$

где z – l -мерный аргумент, соответствующий ошибке; Σ – псевдоковариационная матрица; $\nu > 0$ – параметр формы; $q = z^T \Sigma^{-1} z$. Предельным случаем распределения Стьюдента при $\nu \rightarrow \infty$ является нормальное распределение, при значении $\nu = 1$ получаем распределение Коши.

Методы ММП-оценивания регрессионной модели при стьюдентовских ошибках наблюдений рассматриваются в [156]. ММП-оценки при стьюдентовском распределении ошибок наблюдений с фиксированным значением параметра формы являются робастными. Оценивая, наряду с остальными, параметр формы, приходим к адаптивным оценкам параметра сдвига (уравнения регрессии), практически устойчивым и относимым к типу адаптивных робастных (см. [149, 163] для одномерного случая). ММП-оценивание многомерных моделей, основанное на использовании распределения Стьюдента, подробно рассматривается в гл. 3.

Нормальное и стьюдентовское распределения – это частные случаи эллиптических распределений с плотностью вида

$$g(z, \Sigma, \nu) = |\Sigma|^{-1/2} \tilde{g}(q, \nu),$$

где $\tilde{g}(q, \nu)$ – функция, обладающая свойствами плотности распределения (за исключением условия нормировки). Методы робастного оце-

нивания параметров эллиптического распределения рассмотрены в [120].

Разновидностью эллиптического распределения является N/I-распределение (Normal/Independent) [134, 149]: если распределение вектора Z – многомерное нормальное, а τ – положительная случайная величина, распределение которой часто имеет параметр формы $\nu > 0$, то вектор

$$Y = \frac{1}{\sqrt{\tau}} Z$$

имеет N/I-распределение. Рассматривая различные распределения переменной τ , можно получать различные негауссовские распределения вектора Y . Например, распределение Стьюдента является N/I-распределением при величинах τ , имеющих гамма-распределение $G(\alpha, \beta)$ с плотностью

$$g_G(\tau | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} e^{-\beta\tau}, \quad (1.54)$$

при $\alpha = \frac{\nu}{2}$, $\beta = \frac{\nu}{2}$.

Р.А. Маронна в [167] предложил оценивать параметр сдвига μ и псевдоковариационную матрицу Σ эллиптического распределения из системы уравнений

$$\sum_{i=1}^N w_1(q_i) Y_i - \hat{\mu} = 0, \quad \hat{\Sigma} = \frac{1}{N} \sum_{i=1}^N w_2(q_i) Y_i - \hat{\mu} Y_i - \hat{\mu}^T,$$

где $w_1(q)$, $w_2(q)$ – весовые функции; $q_i = Y_i - \hat{\mu}^T \hat{\Sigma}^{-1} Y_i - \hat{\mu}$; Y_i – вектор i -го наблюдения размерности l . Возможно находить оценки минимизацией функции

$$Q = \sum_{i=1}^N \rho(q_i) + \frac{1}{2} N \ln |\hat{\Sigma}|,$$

приводящей при дифференцируемой функции потерь ρ к оценкам Маронны с $w_1(q) = w_2(q) = 2\rho'(q)$ [153].

При должном выборе функции потерь можно обеспечить хорошие свойства локальной робастности (с точки зрения функции влияния) и хорошую точность введенных оценок. Однако заметим, что глобальные свойства оценок (с точки зрения пороговой точки) резко ухудшаются при увеличении l , а при ограниченной функции потерь функция Q не является ограниченной снизу: $Q \rightarrow -\infty$ при $|\hat{\Sigma}| \rightarrow 0$ [154]. Одновременно хорошими локальными и глобальными свойствами робастности обладают введенные в [154] ограниченные M -оценки (СМ-оценки).

Л.Д. Мешалкин предложил семейство ЭВ-оценок параметров модели псевдонезависимых регрессий (1.35) при нормальном распределении, находимых из системы уравнений [2]

$$\hat{\theta} = \left[\sum_{i=1}^N X_i^T \hat{\Sigma}^{-1} X_i \exp\left(-c \frac{q_i}{2}\right) \right]^{-1} \sum_{i=1}^N X_i^T \hat{\Sigma}^{-1} Y_i \exp\left(-c \frac{q_i}{2}\right),$$

$$\hat{\Sigma} = (1+c) \left[\sum_{i=1}^N \exp\left(-c \frac{q_i}{2}\right) \right]^{-1} \sum_{i=1}^N Y_i - X_i \hat{\theta} \quad Y_i - X_i \hat{\theta}^T \exp\left(-c \frac{q_i}{2}\right),$$

где $X_i = \text{diag} \left[f_1^T(x_i), \dots, f_l^T(x_i) \right]$ – матрица регрессоров в i -м наблюдении, c – параметр оценки, $q_i = Y_i - X_i \hat{\theta}^T \hat{\Sigma}^{-1} Y_i - X_i \hat{\theta}$, Y_i – вектор откликов в i -м наблюдении. Оценки являются асимптотически несмещенными при незасоренном нормальном распределении и доставляют минимум функции

$$Q = \sum_{i=1}^N \rho(\theta, \Sigma),$$

где функция потерь $\rho(\theta, \Sigma) = \rho(q, \Sigma)$ имеет вид [124]

$$\rho(q, \Sigma) = -|\Sigma|^{-\frac{c}{2(1+c)}} \exp\left\{-\frac{c}{2} q\right\}. \quad (1.55)$$

Отмечал Л.Д. Мешалкин и ухудшение свойств ЭВ-оценок при увеличении l .

Ряд оптимальных решений задачи робастного оценивания параметров нормального распределения предложил А.М. Шурыгин [125]. В частности, как и в одномерном случае, оценка Мешалкина оказывается оптимальной с точки зрения подхода Шурыгина. А.М. Шурыгин указал асимптотически несмещенную оценку ковариационной матрицы l -мерного нормального распределения

$$\hat{\Sigma} = \left[\beta(l) \sum_{i=1}^N w(q_i) \right]^{-1} \sum_{i=1}^N w(q_i) Y_i - \hat{\mu} \quad Y_i - \hat{\mu}^T, \quad (1.56)$$

$$\beta(l) = \frac{2}{l} \int_0^{\infty} s^{l/2} e^{-s} w(2s) ds \Big/ \int_0^{\infty} s^{l/2-1} e^{-s} w(2s) ds \quad (1.57)$$

с некоторой весовой функцией $w(q)$.

1.8. ОЦЕНИВАНИЕ ПРИ НАЛИЧИИ НЕОДНОРОДНЫХ ДАННЫХ

Модель неоднородных (засоренных) данных – одна из основных в теории робастности. Однако здесь в цель наших исследований не входит оценивание засоряющего распределения (считается, что неоднородность вызывает лишь небольшие отклонения от постулируемой модели [120]), и информация о нем не сопоставима с информацией о модельном распределении [44]. Рассмотрим моделирование по неоднородным данным в более общей постановке.

Ситуации, в которых наблюдения откликов неоднородны (разнораспределенны), часто встречаются на практике. В [79] говорится: «Наличие различных неоднородностей – типичное явление для любого реального эксперимента». В [100] отмечается, что однородность исследуемой совокупности – одно из «практически наиболее трудно выполнимых требований регрессионного анализа».

Причина неоднородности данных – в воздействии на отклики некоторых дополнительных «мешающих» факторов (сопутствующих пере-

менных, источников неоднородности) [3, 79]. Источники неоднородности могут быть детерминированными или стохастическими, дискретными или непрерывными, наблюдаемыми или ненаблюдаемыми [3, 79, 101]. В ряде случаев источники неоднородности рассматриваются как основные факторы [79] и наоборот.

Дать определение понятия «однородности» в общем случае трудно [43, 44, 100, 107]. В [100] говорится, что «... нельзя дать определения понятия однородности вообще, а можно лишь говорить об *относительной*, условной однородности объектов совокупности в связи с конкретными условиями и целями исследования».

Например, в регрессионных моделях всегда подразумевается неоднородность данных, поскольку значение математического ожидания (параметра сдвига) меняется от наблюдения к наблюдению. И называть регрессионные данные неоднородными на основании этого свойства нецелесообразно [107].

В [3] используется понятие «неоднородности (в регрессионном смысле)», под которой подразумевается изменение значений параметров функции линейной регрессии под действием источников неоднородности. Параметры могут меняться и в зависимости от основных факторов, выступающих в данном случае в качестве источников неоднородности [101]. Подобные модели называют моделями с переменной структурой [3, 46, 95, 100, 101], при этом могут рассматриваться модели с параметрами, меняющимися как непрерывно, так и дискретно (источники неоднородности непрерывные или дискретные).

В случае дискретного изменения параметров выделяют два типа моделей в зависимости от того, производится переход от одной частной регрессии к другой детерминированно на основе значений наблюдаемых переменных или стохастически (в зависимости от неизвестных вероятностей) [95].

В последнем случае приходим к смеси регрессий, в которой, наряду с параметрами частных регрессионных моделей, требуется оценить вероятности каждой из них; модель имеет вид

$$g(y|x, \Phi) = \sum_{j=1}^J \pi_j g_j(y|x, \Phi),$$

где g – плотность распределения отклика, y – значение отклика, x – значение входного фактора, Φ – вектор параметров модели, π_j – ве-

роятность j -й частной модели, g_j – плотность, соответствующая j -й частной модели, J – количество частных моделей.

В [87, 88, 148] рассматривается обобщение данной модели, когда вероятности зависят от входных факторов и некоторых параметров:

$$g_{y|x,\Phi} = \sum_{j=1}^J \pi_j(x, \Phi_1) g_{j,y|x,\Phi_2}, \quad (1.58)$$

где $\Phi = \Phi_1^T, \Phi_2^T$. В [87, 88] рассматривается однооткликовая нормальная регрессия, считается, что неоднородность обусловливается воздействием ненаблюдаемого (неконтролируемого) дискретного случайного фактора. В [148] рассмотрен случай многомерной модели.

До сих пор речь шла об уравнении регрессии (параметре сдвига). Неоднородность может проявляться и в изменении дисперсии (параметра масштаба). Такую модель называют регрессией с гетероскедастичными ошибками. Дисперсия часто моделируется как функция основных факторов или математического ожидания отклика [3]. В общем случае изменяться от наблюдения к наблюдению могут любые основные и/или мешающие параметры.

Примером модели с изменяющейся дисперсией служит модель N/I-распределения, рассмотренная в п. 1.7; дисперсия зависит от ненаблюдаемого случайного источника неоднородности, маргинальное распределение вектора $Y = \frac{1}{\sqrt{\tau}}Z$ – бесконечная смесь:

$$g_{Y|\Phi} = \int_{\Gamma} g_{Y|\tau,\Phi} g_{\tau} d\tau,$$

где Γ – область определения случайной величины τ .

ГЛАВА 2

УСТОЙЧИВОЕ M -ОЦЕНИВАНИЕ

В главе развивается теория устойчивого M -оценивания параметров модели [58, 59, 64, 68, 75–77]. Основная часть результатов посвящена подходу, базирующемуся на модели байесовского точечного засорения и приводящему к робастным оценкам. Особое внимание уделяется случаю неоднородных данных, когда наблюдения могут быть разбиты на группы так, что ошибки наблюдений в одной группе имеют одно и то же модельное распределение. Но реальное распределение группы является засоренным, и требуются устойчивые оценки. Полученные результаты применены к построению робастных оценок многомерной нормальной и двусторонней экспоненциальной моделей.

2.1. ОЦЕНИВАНИЕ ОДНОРОДНОЙ МОДЕЛИ

Одно из важнейших свойств ММП – асимптотическая эффективность получаемых оценок. Но нередко ММП-оценки оказываются слишком чувствительными к отклонению реального закона распределения от постулируемого модельного. При замене ММП-оценки произвольной M -оценкой теряется эффективность, но одновременно может быть обеспечена устойчивость.

В ряде подходов к устойчивому оцениванию конструируются решения, оптимальные по некоторому показателю качества, в том числе отражающему устойчивость. Среди таких подходов – подходы Хампеля и Шурыгина.

Рассмотрим задачу оценивания одномерного параметра θ по наблюдениям

$$z_i, \quad i = 1, \dots, N,$$

имеющим модельное распределение с плотностью $g(z, \theta)$.

M -оценку $\hat{\theta}$ параметра θ находим, решая уравнение

$$\sum_{i=1}^N \psi(z_i, \hat{\theta}) = 0,$$

где оценочная функция ψ удовлетворяет условию асимптотической несмещенности (1.48). Предположим, что решение уравнения является состоятельной оценкой.

При некоторых условиях регулярности M -оценка является асимптотически нормальной; в частности, величина $\sqrt{N}(\hat{\theta} - \theta)$ асимптотически нормальна с нулевым математическим ожиданием и дисперсией

$$V(\psi) = \frac{1}{d^2} \mathbf{E} \psi^2(z, \theta),$$

где $d(\psi) = \int_Z \psi(z, \theta) \frac{\partial}{\partial \theta} g(z, \theta) dz = - \int_Z g(z, \theta) \frac{\partial}{\partial \theta} \psi(z, \theta) dz$ [9, 81, 120, 123]. Величина $V(\psi)$ используется в качестве показателя неточности оценивания [9].

Для конструирования робастных оценок необходимо выбрать показатель качества, учитывающий устойчивость оценки.

Одним из инструментов, используемых в теории робастности (в частности, в подходе Хампеля), является *функция влияния* [99], которая имеет вид

$$IF(z) = \lim_{t \rightarrow 0} \frac{\theta[(1-t)G + t\Delta_z] - \theta[G]}{t},$$

где Δ_z – функция единичного скачка в точке z ; $\theta[H]$ – функционал от распределения, для M -оценок определяемый из равенства

$$\int_Z \psi(z, \theta[H]) dH(z) = 0.$$

При некоторых условиях регулярности для M -оценок справедливо [99]:

$$IF(z, \psi) = \frac{\psi(z, \theta)}{d(\psi)}.$$

Функция влияния является основой для ряда показателей устойчивости. В частности, *чувствительностью к большой ошибке* называется показатель [99]

$$\gamma^* = \sup_z |IF(z, \psi)|.$$

Если он конечен, то оценка называется B -робастной.

В основе подхода Шурыгина лежит модель нарушения предположений о распределении ошибок наблюдений в виде байесовского точечного засорения [125].

Рассмотрим засоренное распределение наблюдений с плотностью

$$(1 - \alpha)g(z, \theta) + \alpha\delta(z - \xi),$$

где α – уровень засорения; δ – функция Дирака; ξ – случайная величина, имеющая распределение с плотностью $s(z, \theta)$.

Засорение модельной плотности приводит к асимптотическому смещению оценки, приближенное значение которого при малом уровне засорения и некоторых условиях регулярности в рассматриваемой модели имеет вид [107]

$$b(\xi, \psi) = \alpha IF(\xi, \psi).$$

Рассмотрим показатель качества оценок вида

$$\Upsilon_s(\psi) = \mathbf{E}_s IF^2(\xi, \psi), \quad (2.1)$$

где \mathbf{E}_s – математическое ожидание по плотности $s(z, \theta)$. Показатель вида (2.1) рассматривается в [125], но вводится другим путем.

Оценочная функция, доставляющая минимальное значение функционалу (2.1), имеет вид

$$\psi(z, \theta) = c \left[\frac{\partial}{\partial \theta} \ln g(z, \theta) + \beta \right] \frac{g(z, \theta)}{s(z, \theta)}, \quad (2.2)$$

где $\beta = \beta(\theta)$ определяется из условия асимптотической несмещенности (1.48); $c = c(\theta)$ – ненулевой множитель, с точностью до которого определяется оценочная функция [125].

Конкретизируя функцию $s(z, \theta)$, можно получать различные решения. Например, при использовании плотности $s(z, \theta)$, пропорциональной $g^{1-\delta}(z, \theta)$, где δ – параметр, $0 \leq \delta \leq 1$, получаем оценочную функцию вида [64]

$$\psi(z, \theta) = c \left[\frac{\partial}{\partial \theta} \ln g(z, \theta) + \beta \right] g^{\delta}(z, \theta).$$

В случае нормального распределения данные оценки совпадают с оценками Мешалкина (ЭВ-оценками) [1, 2]. Оценки с параметром $\delta = 1/2$ называются *радикальными* [125]. В общем случае, подчеркивая связь с радикальными, будем называть данные оценки *обобщенными радикальными*.

Показатель (2.1) зависит от плотности $s(z, \theta)$, которая не известна на практике. Один из путей решения этой проблемы – использование максиминной формулировки [125]

$$s_* = \arg \max_{s \in S} \min_{\psi} \Upsilon_s(\psi).$$

Если множество S совпадает с множеством модельных плотностей, то соответствующая оптимальная оценка называется *стойкой* [125].

Другой путь – использование показателя неустойчивости оценки вида [125, 126]

$$W(\psi) = \int_Z IF^2(z, \psi) dz = \frac{1}{d^2(\psi)} \int_Z \psi^2(z, \theta) dz. \quad (2.3)$$

Функционал (2.3) эквивалентен (2.1) в условиях равномерного распределения засоряющего импульса ξ (см. [126]), или, говоря иначе, в условиях *максимально неопределенного точечного засорения* [70]. Действительно, максимально неопределенное распределение есть равномерное, поскольку равномерное распределение имеет максимальную энтропию [81]. Величину (2.3) можно также рассматривать как квадрат нормы функции влияния в пространстве L_2 [125, 126]. Функционал (2.3) назван в [125] *неустойчивостью* оценки, но вводится другим путем.

Условием *устойчивости* оценки (и соответствующей оценочной функции) является выполнение требования $W(\psi) < \infty$ [125]. Данному условию удовлетворяют оценки с функцией потерь, имеющей горизонтальную асимптоту, т. е. оценки, устойчивые при асимметричном засорении, а, например, оценка Хьюбера и медиана, неустойчивые при асимметричном засорении, ему не удовлетворяют.

Минимизация функционала (2.3) на множестве устойчивых оценочных функций приводит к *оценке максимальной устойчивости* (ОМУ) [125] с оценочной функцией

$$\psi_{\text{ОМУ}}(z, \theta) = c \left[\frac{\partial}{\partial \theta} \ln g(z, \theta) + \beta \right] g(z, \theta),$$

где c и β имеют тот же смысл, что и в (2.2).

Устойчивостью оценки называется относительная характеристика [125]

$$\text{stb } \psi = \frac{W(\psi_{\text{ОМУ}})}{W(\psi)},$$

введенная аналогично *эффективности*

$$\text{eff } \psi = \frac{V(\psi_{\text{ММП}})}{V(\psi)},$$

где $\psi_{\text{ММП}}(z, \theta) = c \frac{\partial}{\partial \theta} \ln g(z, \theta)$ – оценочная функция ММП-оценки.

Заметим, что обобщенные радикальные оценки при $\delta = 1$ совпадают с ОМУ, а в случае $\delta = 0$ – с ММП-оценкой. По этой причине функционал, оптимизируемый обобщенными радикальными оценками, можно считать промежуточным, неявно отражающим компромисс между критериями неточности $V(\psi)$ и неустойчивости $W(\psi)$ оценки (см. также [125], где обсуждается данный функционал при $\delta = 1/2$).

Можно рассматривать и задачу явной оптимизации обоих критериев. Так, *условно оптимальная* оценка [125] с оценочной функцией

$$\psi(z, \theta) = c \left[\frac{\partial}{\partial \theta} \ln g(z, \theta) + \beta \right] \left[1 + \frac{\gamma}{g(z, \theta)} \right]^{-1}, \quad \gamma \geq 0, \quad (2.4)$$

где c и β имеют тот же смысл, что и в (2.2), обеспечивает во множестве устойчивых оценок максимальную устойчивость при фиксированной эффективности и одновременно максимальную эффективность при фиксированной устойчивости. Параметр γ в (2.4) позволяет управлять устойчивостью и точностью оценок. Например, при $\gamma = 0$ получаем ММП-оценку, при $\lambda = \frac{V(\psi_{\text{ММП}})}{W(\psi_{\text{ОМУ}})}$ – *компромиссную оценку* [125], при $\gamma = \infty$ – ОМУ.

Условно оптимальную оценку, удовлетворяющую условию

$$\text{eff } \psi = \text{stb } \psi,$$

назовем *равнооптимальной* [70]. Обычно не существует простых аналитических выражений для равнооптимальной оценочной функции. Тем не менее она, по-видимому, обладает наиболее привлекательными свойствами с точки зрения рассмотренного подхода. В качестве альтернативы равнооптимальной оценке может использоваться радикальная оценка, которая во многих практически значимых частных случаях тоже в равной степени обеспечивает эффективность и устойчивость. Часто при незначительно меньших величинах эффективности и устойчивости радикальная оценка, благодаря относительно простому аналитическому виду, оказывается наиболее удобной для практического применения.

Назовем *обобщенной компромиссной* оценкой с оценочной функцией, оптимизирующей функционал

$$\frac{1-\delta}{\text{eff } \psi} + \frac{\delta}{\text{stb } \psi}, \quad 0 \leq \delta \leq 1$$

(см. [64]). Ее частным случаем при $\delta=1/2$ является компромиссная оценка.

Теорема 2.1. Оценочная функция обобщенной компромиссной оценки имеет вид (2.4) при $\gamma = \frac{\delta}{1-\delta} \frac{V(\psi_{\text{ММП}})}{W(\psi_{\text{ОМУ}})}$.

Доказательство. Оценочная функция доставляет минимальное значение функционалу

$$\int_Z \left\{ \psi^2(z, \theta) \left[\frac{1-\delta}{V(\psi_{\text{ММП}})} g(z, \theta) + \frac{\delta}{W(\psi_{\text{ОМУ}})} \right] + \zeta_1 \psi(z, \theta) g(z, \theta) + \zeta_2 \frac{\partial}{\partial \theta} \psi(z, \theta) g(z, \theta) \right\} dz, \quad (2.5)$$

в котором вторым и третьим слагаемыми с множителями Лагранжа ζ_1 и ζ_2 учитываются соответственно условие асимптотической несмещенности и условие нормировки оценочной функции

$$d(\psi) = 1.$$

В соответствии с [125] необходимым условием экстремума функционала (2.5) является выполнение равенства

$$2\psi(z, \theta) \left[\frac{1-\delta}{V(\psi_{\text{ММП}})} g(z, \theta) + \frac{\delta}{W(\psi_{\text{ОМУ}})} \right] + \zeta_1 g(z, \theta) - \zeta_2 \frac{\partial}{\partial \theta} g(z, \theta) = 0,$$

откуда

$$\psi(z, \theta) = \frac{\zeta_2 V(\psi_{\text{ММП}})}{2(1-\delta)} \left[\frac{\partial}{\partial \theta} \ln g(z, \theta) - \frac{\zeta_1}{\zeta_2} \right] \left[1 + \frac{\delta}{1-\delta} \frac{V(\psi_{\text{ММП}})}{W(\psi_{\text{ОМУ}})} \frac{1}{g(z, \theta)} \right]^{-1},$$

что с точностью до обозначений совпадает с доказываемым результатом.

Теорема 2.1 – это обобщение на обобщенную компромиссную оценку соответствующего результата из [125] для компромиссной оценки.

Положительным свойством обобщенной компромиссной оценки можно назвать большее удобство работы с параметром δ по сравнению с параметром γ условно оптимальной оценки.

На практике оцениваемый параметр часто является векторным. Например, вместе с параметром сдвига оцениваются мешающие параметры (параметры масштаба и формы). Конструирование векторных оценочных функций – более сложная задача. Возможный путь ее решения – составить вектор из оценочных функций, полученных для каждого параметра в отдельности.

2.2. ОЦЕНИВАНИЕ НЕОДНОРОДНОЙ МОДЕЛИ

2.2.1. ПОКАЗАТЕЛИ КАЧЕСТВА

Пусть вектор, составленный из независимых наблюдений за откликом (откликами) исследуемой системы, можно разделить на p подвекторов (групп наблюдений) так, что ошибки наблюдений одной группы будут одинаково распределенными. Рассмотрим регрессионную модель вида

$$y_{ij} = f^T(x_{ij})\theta + e_{ij}, \quad i = 1, \dots, p, \quad j = 1, \dots, n_i, \quad (2.6)$$

где y_{ij} – j -е наблюдение в i -й группе, x_{ij} – вектор входных переменных, e_{ij} – ненаблюдаемая ошибка, n_i – количество наблюдений в i -й группе.

Ошибки наблюдений i -й группы имеют модельное распределение с функцией $G_i(z)$, плотностью $g_i(z)$, нулевым параметром сдвига и мешающими параметрами: параметром масштаба и, возможно, параметром формы.

M -оценку $\hat{\theta}$ вектора параметров θ определим минимизацией функции вида

$$Q(\theta) = \sum_{i=1}^p \sum_{j=1}^{n_i} \rho_i [r_{ij}(\theta)], \quad (2.7)$$

где $\rho_i[z]$ – функция потерь i -й группы; $r_{ij}(\theta) = y_{ij} - f^T(x_{ij})\theta$ – остаток, соответствующий j -му наблюдению в i -й группе.

Если функции $\rho_i(z)$ дифференцируемы, то, приравняв нулю частные производные функции (2.7) по элементам вектора параметров при его значении, равном $\hat{\theta}$, получим систему оценочных уравнений

$$-\sum_{i=1}^p \sum_{j=1}^{n_i} f(x_{ij}) \psi_i [r_{ij}(\hat{\theta})] = 0, \quad (2.8)$$

где $\psi_i[z] = \rho'_i[z]$ – i -й элемент векторной оценочной функции ψ .

Введем уравнение

$$\sum_{i=1}^p \sum_{j=1}^{n_i} f(x_{ij}) \int_{Z_{ij}} \psi_i(y - f^T(x_{ij})t) dH_{ij}(y - f^T(x_{ij})\theta) = 0, \quad (2.9)$$

где Z_{ij} – область определения j -го наблюдения i -й группы.

Уравнением (2.9) при $t = \theta$ и $H_{ij} = G_i$ задается условие асимптотической несмещенности оценки $\hat{\theta}$ [107]. Будем считать выполненным условие

$$E\psi_i = 0, \quad i = 1, \dots, p, \quad (2.10)$$

которое влечет выполнение условия асимптотической несмещенности оценки.

Предположим, что решение уравнения (2.8) – состоятельная оценка. При некоторых условиях регулярности [9] асимптотический раз-

брос оценки $\hat{\theta}$, являющийся показателем неточности оценивания, можно представить в виде

$$V = D^{-1} \left[\sum_{i=1}^p \mathbf{E} \psi_i^2 X_i^T X_i \right] D^{-1}, \quad (2.11)$$

где $D = \sum_{i=1}^p \mathbf{E} \psi_i' X_i^T X_i$, X_i – $n_i \times m$ -матрица, составленная из векторов-строк $f^T(x_{ij})$ значений регрессоров i -й группы.

Распространим на случай неоднородной модели (2.6) подход Шурьгина, в котором рассматривается байесовское точечное засорение модельного распределения. Используем *неоднородное байесовское точечное засорение*.

Рассмотрим три случая: в первом засоренным является распределение ошибки только одного наблюдения, во втором засоренными являются распределения ошибок всех наблюдений одной группы, в третьем засоренными являются распределения ошибок всех наблюдений.

Функцию засоренного распределения ошибки j -го наблюдения в i -й группе определим в виде смеси

$$(1 - \alpha_i)G_i(z) + \alpha_i \Delta_{\xi_{ij}}(z),$$

где α_i – уровень засорения, причем в третьем случае $\alpha_i = \alpha$, $i = 1, \dots, p$; $\Delta_{\xi}(z) = \Delta(z - \xi)$ – функция единичного скачка; ξ_{ij} – случайная величина, имеющая распределение с плотностью $s_i(z)$. Все величины ξ_{ij} независимы.

Заметим, что можно вводить различные уровни засорения α_{ij} для каждого j -го наблюдения i -й группы, $j = 1, \dots, n_i$, тогда во втором случае должно выполняться условие $\alpha_{ij} = \alpha_i$, в третьем – $\alpha_{ij} = \alpha$.

При засоренном распределении оценка $\hat{\theta}$ будет асимптотически смещена; при малых уровнях засорения и некоторых условиях регулярности приближенное значение асимптотического смещения в пер-

вом, втором и третьем случаях соответственно определяется формулами

$$B_{ij}(\xi_{ij}) = \alpha_i D^{-1} f(x_{ij}) \psi_i(\xi_{ij}), \quad B_i(\Xi_i) = \alpha_i D^{-1} \sum_{j=1}^{n_i} f(x_{ij}) \psi_i(\xi_{ij}),$$

$$B(\Xi) = \alpha D^{-1} \sum_{i=1}^p \sum_{j=1}^{n_i} f(x_{ij}) \psi_i(\xi_{ij}),$$

где Ξ_i – вектор, составленный из элементов ξ_{ij} , $j=1, \dots, n_i$; Ξ – вектор, составленный из подвекторов – элементов Ξ_i , $i=1, \dots, p$ (см. [107]).

Показателем качества оценок выберем в первом, втором и третьем случаях соответственно величины

$$U_s = \sum_{i=1}^p \sum_{j=1}^{n_i} \alpha_i^{-2} w_i \mathbf{E}_{s_i} B_{ij} B_{ij}^T, \quad (2.12)$$

$$\bar{U}_s = \sum_{i=1}^p \alpha_i^{-2} w_i \mathbf{E}_{s_i} B_i B_i^T, \quad \tilde{U}_s = \alpha^{-2} \mathbf{E}_s B B^T,$$

где w_i – положительный весовой коэффициент i -й группы, \mathbf{E}_{s_i} и \mathbf{E}_s – математическое ожидание по плотности $s_i(z)$ и плотности вектора Ξ соответственно. В первом случае для показателя (2.12) имеем

$$U_s = D^{-1} \left[\sum_{i=1}^p w_i \mathbf{E}_{s_i} \psi_i^2 X_i^T X_i \right] D^{-1}. \quad (2.13)$$

Во втором и третьем случаях будем предполагать выполненным условие $\mathbf{E}_{s_i} \psi_i = 0$, тогда получим

$$\bar{U}_s = U_s, \quad \tilde{U}_s = D^{-1} \left[\sum_{i=1}^p \mathbf{E}_{s_i} \psi_i^2 X_i^T X_i \right] D^{-1}.$$

Как и в однородном случае, свяжем полученные показатели качества оценок с функцией влияния Хампеля. Определим функцию влияния для неоднородных наблюдений в первом, втором и третьем случаях соответственно как

$$IF_{ij}(\xi_{ij}) = \lim_{\beta \rightarrow 0} \frac{1}{\beta} \theta[\tilde{G}_{ij,\beta}] - \theta[G] , \quad IF_i(\Xi_i) = \lim_{\beta \rightarrow 0} \frac{1}{\beta} \theta[\tilde{G}_{i,\beta}] - \theta[G] ,$$

$$IF(\Xi) = \lim_{\beta \rightarrow 0} \frac{1}{\beta} \theta[\tilde{G}_\beta] - \theta[G] ,$$

где функционал $\theta[H]$ определяется из уравнения (2.9), в котором $t = \theta[H]$;

$$H = \prod_{i=1}^p \prod_{j=1}^{n_i} H_{ij} \quad y_{ij} - f^T(x_{ij})\theta \quad (2.14)$$

есть функция распределения вектора наблюдений; G определяется формулой (2.14) при $H_{ij} = G_{ij}$ для всех i, j ; $\tilde{G}_{ij,\beta}$ определяется формулой (2.14) при $H_{i'j'} = G_{i'j'}$ для $(i', j') \neq (i, j)$, $H_{ij} = (1 - \beta)G_{ij} + \beta\Delta_{\xi_{ij}}$; $\tilde{G}_{i,\beta}$ определяется формулой (2.14) при $H_{i'j} = G_{i'j}$ для $i' \neq i$ и всех j , $H_{ij} = (1 - \beta)G_{ij} + \beta\Delta_{\xi_{ij}}$ для всех j ; \tilde{G}_β определяется формулой (2.14) при $H_{ij} = (1 - \beta)G_{ij} + \beta\Delta_{\xi_{ij}}$ для всех i, j .

Рассуждая, аналогично изложенному в [99, 120], получим

$$IF_{ij}(\xi_{ij}) = -D^{-1} f(x_{ij})\psi_i(\xi_{ij}), \quad IF_i(\Xi_i) = -D^{-1} \sum_{j=1}^{n_i} f(x_{ij})\psi_i(\xi_{ij}),$$

$$IF(\Xi) = -D^{-1} \sum_{i=1}^p \sum_{j=1}^{n_i} f(x_{ij})\psi_i(\xi_{ij})$$

(см. [147] для первого случая). Следовательно, справедливо

$$U_s = \sum_{i=1}^p \sum_{j=1}^{n_i} w_i \mathbf{E}_{s_i} \mathbf{IF}_{ij} \mathbf{IF}_{ij}^T = \sum_{i=1}^p w_i \mathbf{E}_{s_i} \mathbf{IF}_i \mathbf{IF}_i^T, \quad \tilde{U}_s = \mathbf{E}_s \mathbf{IF} (\mathbf{IF})^T,$$

в результате показатели U_s и \tilde{U}_s , как и функция влияния, являются локальными. Поскольку показатель \tilde{U}_s равен U_s при единичных весовых коэффициентах, в дальнейшем будем рассматривать показатель (2.13) как более общий.

2.2.2. ОПТИМИЗАЦИЯ С ОГРАНИЧЕНИЯМИ НА МАТРИЦЫ РЕГРЕССОРОВ

Предположим, что для i -й группы выполнено условие

$$X_i^T X_i = k_i A, \quad (2.15)$$

где k_i – положительная константа, A – симметричная положительно определенная матрица соответствующей размерности. Показатель (2.13) в этом случае имеет вид $u_s A^{-1}$, где u_s – скалярная величина:

$$u_s = \frac{1}{d^2} \sum_{i=1}^p k_i w_i \mathbf{E}_{s_i} \psi_i^2, \quad (2.16)$$

$$d = \sum_{i=1}^p k_i \mathbf{E} \psi_i'.$$

Найдем оценочную функцию, которая доставляет минимальное значение функционалу $u_s[\psi]$.

Поскольку оценочная функция определена с точностью до множителя, не зависящего от i , введем условие нормировки вида

$$d = 1. \quad (2.17)$$

В результате необходимо оптимизировать функционал

$$\sum_{i=1}^p \int_{Z_i} \left[k_i w_i \psi_i^2(z) s_i(z) + \zeta_{1i} \psi_i(z) g_i(z) + \zeta_{2i} k_i \psi_i'(z) g_i(z) \right] dz, \quad (2.18)$$

где Z_i – область определения ошибок i -й группы, условия (2.10) и (2.17) учитываются соответственно вторым и третьим слагаемыми с множителями Лагранжа ζ_{1i} , $i=1, \dots, p$, и ζ_2 . Вид функционала показывает, что его оптимизация может быть проведена независимо по каждому элементу функции ψ .

Оценочная функция ψ_i с точностью до несущественного сомножителя совпадает с оценочной функцией i -й группы для (общего) параметра сдвига неоднородной модели; соответственно, оптимизируемый ею функционал эквивалентен функционалу, оптимизируемому в задаче поиска оптимальной оценочной функции параметра сдвига. Поэтому возможна замена исходной задачи задачей поиска оптимальной оценочной функции параметра сдвига, для решения которой можно воспользоваться аналогичным результатом из [125]. Решение имеет вид

$$\psi_i(z) = c [\ln g_i(z)]' + \beta_i \frac{g_i(z)}{w_i s_i(z)}, \quad (2.19)$$

где c – множитель, с точностью до которого определяется оценочная функция; β_i определяется из условия (2.10).

Конкретизируя функции $s_i(z)$ и коэффициенты w_i , можно получить различные решения.

Выбор

$$w_i s_i(z) = g_i^{1-\delta_i}(z),$$

где δ_i – параметр, $0 \leq \delta_i \leq 1$, приводит к обобщенным радикальным оценкам с i -м элементом оценочной функции вида

$$\psi_i(z) = c [\ln g_i(z)]' + \beta_i g_i^{\delta_i}(z). \quad (2.20)$$

При $\delta_i = 0$, $i = 1, \dots, p$, получаем показатель (2.13), совпадающий с показателем неточности V (2.11), величину u_s в данном случае обозначим u_0 , т. е.

$$u_0 = \frac{1}{d^2} \sum_{i=1}^p k_i \mathbf{E} \psi_i^2.$$

При выборе $\delta_i = 1$, $i = 1, \dots, p$, соответствующем максимально неопределенному точечному засорению, показатель (2.13) является аналогом показателя (2.3), величину u_s в данном случае обозначим u_1 , т. е.

$$u_1 = \frac{1}{d^2} \sum_{i=1}^p k_i \mathbf{E} \left[\frac{\psi_i^2}{g_i} \right].$$

При значениях параметров $0 < \delta_i < 1$, $i = 1, \dots, p$, показатель (2.13) является промежуточным, неявно отражающим компромисс между неточностью и неустойчивостью оценки.

Выбор

$$s_i(z) = \frac{1}{g_{\delta_i}^0} g_i^{1-\delta_i}(z), \quad w_i = 1,$$

где $g_{\delta_i}^0$ – константа, нормирующая плотность $s_i(z)$, приводит к решению

$$\psi_i(z) = c [\ln g_i(z)]' + \beta_i g_i^{\delta_i}(z) g_{\delta_i}^0. \quad (2.21)$$

Соответствующую оценку назовем *нормированной обобщенной радикальной*.

Рассмотрим явную задачу одновременной оптимизации показателей неточности u_0 и неустойчивости u_1 . При этом будем рассматривать устойчивые оценки, для которых все величины $\mathbf{E} \frac{\psi_i^2}{g_i}$ конечны.

Найдем вид оценочных функций условно оптимальной оценки при ограничениях (2.10) и (2.17) путем минимизации функционала

$$\sum_{i=1}^p \int_{Z_i} \left[k_i \psi_i^2(z) g_i(z) + \zeta_{1i} \psi_i(z) g_i(z) + \zeta_{2i} k_i \psi_i'(z) g_i(z) + \zeta_{3i} k_i \psi_i^2(z) \right] dz,$$

согласно которому оптимизируется показатель неточности. Условия (2.10) и (2.17) учитываются аналогично (2.18), последнее слагаемое с множителем Лагранжа ζ_3 отвечает фиксированной неустойчивости.

Проводить оптимизацию функционала можно независимо по каждому элементу $\psi_i(z)$, аналогично оптимизации (2.18), тогда, воспользовавшись аналогичным результатом из [125], получим решение вида

$$\psi_i(z) = c [\ln g_i(z)]' + \beta_i \frac{1}{1 + \gamma / g_i(z)}, \quad (2.22)$$

где у c и β_i тот же смысл, что и в формуле (2.19); параметр $\gamma > 0$ определяется выбранным уровнем неустойчивости.

Оптимизация показателя неустойчивости при фиксированной неточности приводит к решению того же вида, а параметр γ определяется выбранным уровнем неточности.

Введем показатель эффективности оценки

$$\text{eff} = \frac{u_0^*}{u_0},$$

где u_0^* – минимальное по оценочной функции значение функционала $u_0(\psi)$, и показатель устойчивости

$$\text{stb} = \frac{u_1^*}{u_1},$$

где u_1^* – минимальное по оценочной функции значение функционала $u_1(\psi)$.

Тогда обобщенная компромиссная оценка имеет оценочные функции, доставляющие минимальное значение функционалу

$$\sum_{i=1}^p \int_{Z_i} \left\{ k_i \psi_i^2(z) \left[\frac{1-\delta}{u_0^*} g_i(z) + \frac{\delta}{u_1^*} \right] + \zeta_{1i} \psi_i(z) g_i(z) + \zeta_{2i} k_i \psi_i'(z) g_i(z) \right\} dz,$$

в котором условия (2.10) и (2.17) учитываются аналогично (2.18).

Проводить оптимизацию функционала можно независимо по каждому элементу $\psi_i(z)$, аналогично (2.18). В соответствии с теоремой 2.1 решение имеет вид (2.22) со значением γ вида

$$\gamma = \frac{\delta u_0^*}{(1-\delta)u_1^*}. \quad (2.23)$$

При получении оценочных функций использовалось условие (2.15), сильно сужающее область их применимости. В ряде случаев оно выполняется: в однопараметрическом случае (например, при оценивании параметра сдвига) или при проведении повторных наблюдений в соответствии с фиксированным планом эксперимента в условиях различных значений мешающих параметров.

Получение решения в общем случае слишком сложно, поэтому здесь целесообразно использовать полученные выше решения, несмотря на их неоптимальность. Например, можно использовать оценки с оценочными функциями (2.20) и (2.21) или обобщенные компромиссные, построенные в условиях однородных наблюдений (см. теорему 2.1).

Оценивание параметров регрессии по однородным данным – еще один (тривиальный) частный случай, когда выполняется условие (2.15). Однако, согласно модели неоднородного байесовского точечного засорения, различным наблюдениям здесь соответствуют различные засоряющие импульсы.

На практике значения мешающих параметров распределения ошибок являются неизвестными, поэтому необходимо их оценивать одновременно с параметрами уравнения регрессии. Если мешающие пара-

метры различны для разных групп, то можно воспользоваться оценками, задаваемыми как решение системы оценочных уравнений

$$\sum_{j=1}^{n_i} \chi_i r_{ij} = 0, \quad i=1, \dots, p,$$

где $\chi_i(z)$ – оценочная функция i -й группы, удовлетворяющая условию асимптотической несмещенности оценок. При этом оценочные уравнения каждого из мешающих параметров, получаемые как в однородном случае, добавляются к системе (2.8).

2.2.3. ПОЭЛЕМЕНТНАЯ ОПТИМИЗАЦИЯ

Вначале рассмотрим следующую общую модель. Пусть независимые одномерные наблюдения

$$z_i, \quad i=1, \dots, N,$$

имеют распределения с функциями $G_i(z, \phi)$ и плотностями $g_i(z, \phi)$.

Состоятельную M -оценку $\hat{\phi}$ вектора параметров модели ϕ определим как решение системы оценочных уравнений

$$\sum_{i=1}^N \psi_i(z_i, \hat{\phi}) = 0, \quad (2.24)$$

где $\psi_i(z, \hat{\phi})$ – векторная оценочная функция, удовлетворяющая условию [9]

$$\mathbf{E}\psi_i(z, \phi) = 0, \quad i=1, \dots, N. \quad (2.25)$$

Рассмотрим два варианта неоднородного байесовского точечного засорения: в первом засоренным является распределение только одного наблюдения, во втором засоренными являются распределения всех наблюдений.

Функцию засоренного распределения i -го наблюдения определим в виде смеси

$$(1 - \alpha_i)G_i(z, \phi) + \alpha_i \Delta_{\xi_i}(z),$$

где α_i – уровень засорения, причем во втором варианте положим $\alpha_i = \alpha$, $i = 1, \dots, N$; независимые случайные величины ξ_i , $i = 1, \dots, N$, имеют распределения с плотностями $s_i(z)$.

Построим критерий качества оценок в предположении, что оценивается только один j -й элемент вектора ϕ решением j -го уравнения системы оценочных уравнений.

Приближенное значение асимптотического смещения оценки $\hat{\phi}_j$ (при засоренном распределении) при малых уровнях засорения и некоторых условиях регулярности в первом и во втором вариантах соответственно определяется формулами

$$B_{ij}(\xi_i, \phi) = -\frac{\alpha_i}{d_j(\phi)} \psi_{ij}(\xi_i, \phi), \quad B_j(\Xi, \phi) = -\alpha \sum_{i=1}^N \frac{1}{d_j(\phi)} \psi_{ij}(\xi_i, \phi),$$

где $\psi_{ij}(z, \phi)$ – j -й элемент вектора $\psi_i(z, \phi)$; $d_j(\phi) = \sum_{i=1}^N \mathbf{E} \frac{\partial}{\partial \phi_j} \psi_{ij}(z, \phi)$;

Ξ – вектор, составленный из элементов ξ_i , $i = 1, \dots, N$ (см. [107]).

Показателями качества оценок выберем в первом и во втором вариантах соответственно величины

$$U_{s,j} = \sum_{i=1}^N \frac{w_i}{\alpha_i^2} \mathbf{E}_{s_i} B_{ij}^2(\xi_i, \phi), \quad \tilde{U}_{s,j} = \frac{1}{\alpha^2} \mathbf{E}_s B_j^2(\Xi, \phi),$$

где w_i – положительный вес i -го наблюдения.

В первом варианте имеем

$$U_{s,j} = \frac{1}{d_j^2(\phi)} \sum_{i=1}^N w_i \mathbf{E}_{s_i} \psi_{ij}^2(z, \phi).$$

Во втором варианте будем предполагать выполненным условие $\mathbf{E}_{s_i} \psi_i(z, \phi) = 0$, тогда получим

$$\tilde{U}_{s,j} = \frac{1}{d_j^2(\phi)} \sum_{i=1}^N \mathbf{E}_{s_i} \psi_{ij}^2(z, \phi).$$

Поскольку показатель $\tilde{U}_{s,j}$ равен $U_{s,j}$ при единичных весовых коэффициентах, будем рассматривать показатель $U_{s,j}$ как более общий.

Как и в п. 2.2.1, при некоторых ограничениях введенные показатели можно выразить через функцию влияния.

Найдем оценочные функции $\psi_{ij}(z, \phi)$, $i=1, \dots, N$, которые доставляют минимальное значение функционалу

$$U_{s,j} = U_{s,j}(\psi_{1j}, \dots, \psi_{Nj}).$$

Будем искать оценочные функции, удовлетворяющие условию (2.25). Кроме того, поскольку оценочная функция определена с точностью до ненулевого множителя, не зависящего от i , введем условие нормировки вида

$$d_j(\phi) = 1. \quad (2.26)$$

В результате необходимо оптимизировать функционал

$$\sum_{i=1}^N \int_{Z_i} \left[w_i \psi_{ij}^2(z, \phi) s_i(z) + \zeta_{1ij} \psi_{ij}(z, \phi) g_i(z, \phi) + \zeta_{2j} g_i(z, \phi) \frac{\partial}{\partial \phi_j} \psi_{ij}(z, \phi) \right] dz, \quad (2.27)$$

где Z_i – область определения i -го наблюдения. Условия (2.25) и (2.26) учитываются в (2.27) соответственно вторым и третьим слагаемыми с множителями Лагранжа ζ_{1ij} , $i=1, \dots, N$, и ζ_{2j} .

Вид функционала показывает, что его оптимизация может быть проведена независимо по каждой функции $\psi_{ij}(z, \phi)$, тогда можно воспользоваться аналогичным результатом из [125]. Решение имеет вид

$$\psi_{ij}(z, \phi) = c_j \left[\frac{\partial}{\partial \phi_j} \ln g_i(z, \phi) + \beta_{ij} \right] \frac{g_i(z, \phi)}{w_i s_i(z)}, \quad (2.28)$$

где c_j – множитель, с точностью до которого определяется оценочная функция; β_{ij} находится из условия (2.25).

При выборе

$$w_i s_i(z) = g_i^{1-\delta_i}(z, \phi),$$

где δ_i – параметр, $0 \leq \delta_i \leq 1$, приходим к обобщенным радикальным оценкам с оценочными функциями вида

$$\psi_{ij}(z, \phi) = c_j \left[\frac{\partial}{\partial \phi_j} \ln g_i(z, \phi) + \beta_{ij} \right] g_i^{\delta_i}(z, \phi).$$

При выполнении некоторых условий регулярности в случае $\delta_i = 0$, $i = 1, \dots, N$, получаем показатель $U_{s,j}$, совпадающий с асимптотическим разбросом оценки ϕ_j ; величину $U_{s,j}$ в данном случае обозначим $U_{0,j}$, т. е.

$$U_{0,j} = \frac{1}{d_j^2(\phi)} \sum_{i=1}^N \mathbf{E} \psi_{ij}^2(z, \phi).$$

При выборе $\delta_i = 1$, $i = 1, \dots, N$, $U_{s,j}$ является показателем неустойчивости оценки ϕ_j , аналогичным (2.3); величину $U_{s,j}$ в данном случае обозначим $U_{1,j}$, т. е.

$$U_{1,j} = \frac{1}{d_j^2(\phi)} \sum_{i=1}^N \mathbf{E} \left[\frac{\psi_{ij}^2(z, \phi)}{g_i(z, \phi)} \right].$$

Рассмотрим задачу одновременной оптимизации показателей неточности $U_{0,j}$ и неустойчивости $U_{1,j}$. Рассмотрим устойчивые оценки, для которых величины $\mathbf{E} \frac{\psi_{ij}^2(z, \phi)}{g_i(z, \phi)}$, $i = 1, \dots, N$, конечны.

Найдем вид оценочных функций условно оптимальной оценки ϕ_j при ограничениях (2.25) и (2.26) путем минимизации функционала

$$\sum_{i=1}^N \int_{Z_i} \left[\psi_{ij}^2(z, \phi) g_i(z, \phi) + \zeta_{1ij} \psi_{ij}(z, \phi) g_i(z, \phi) + \zeta_{2j} g_i(z, \phi) \frac{\partial}{\partial \phi_j} \psi_{ij}(z, \phi) + \zeta_{3j} \psi_{ij}^2(z, \phi) \right] dz,$$

согласно которому оптимизируется показатель неточности. Условия (2.25) и (2.26) учитываются аналогично (2.27), последнее слагаемое с множителем Лагранжа ζ_{3j} отвечает фиксированной неустойчивости.

Проводить оптимизацию функционала можно независимо по каждой функции $\psi_{ij}(z, \phi)$, тогда можно воспользоваться аналогичным результатом из [125]. Решение имеет вид:

$$\psi_{ij}(z, \phi) = c_j \left[\frac{\partial}{\partial \phi_j} \ln g_i(z, \phi) + \beta_{ij} \right] \frac{1}{1 + \gamma_j / g_i(z, \phi)}, \quad (2.29)$$

где c_j и β_{ij} имеют тот же смысл, что и в формуле (2.28); параметр $\gamma_j > 0$ определяется выбранным уровнем неустойчивости.

Оптимизация показателя неустойчивости при фиксированной неточности приводит к решению того же вида, а параметр γ_j определяется выбранным уровнем неточности.

Введем показатель эффективности оценки ϕ_j

$$\text{eff}_j = \frac{U_{0,j}^*}{U_{0,j}},$$

где $U_{0,j}^*$ – минимальное значение функционала $U_{0,j}(\psi_{1j}, \dots, \psi_{Nj})$, и показатель устойчивости

$$\text{stb}_j = \frac{U_{1,j}^*}{U_{1,j}},$$

где $U_{1,j}^*$ – минимальное значение функционала $U_{1,j}(\psi_{1j}, \dots, \psi_{Nj})$.

Обобщенной компромиссной оценке соответствуют оценочные функции, доставляющие минимальное значение показателю

$$\frac{1 - \delta_j}{\text{eff}_j} + \frac{\delta_j}{\text{stb}_j},$$

где $0 \leq \delta_j \leq 1$. Указанные оценочные функции доставляют минимальное значение функционалу

$$\sum_{i=1}^N \int_{Z_i} \left\{ \psi_{ij}^2(z, \phi) \left[\frac{1 - \delta_j}{U_{0,j}^*} g_i(z, \phi) + \frac{\delta_j}{U_{1,j}^*} \right] + \zeta_{1ij} \psi_{ij}(z, \phi) g_i(z, \phi) + \zeta_{2j} g_i(z, \phi) \frac{\partial}{\partial \phi_j} \psi_{ij}(z, \phi) \right\} dz,$$

в котором условия (2.25) и (2.26) учитываются аналогично (2.27). Оптимизация функционала проводится независимо по каждой функции $\psi_{ij}(z, \phi)$, поэтому воспользуемся аналогичным результатом из теоремы 2.1. Получаем решение вида (2.29) со значением γ_j вида

$$\gamma_j = \frac{\delta_j}{1 - \delta_j} \frac{U_{0,j}^*}{U_{1,j}^*}.$$

Применим полученные результаты к задаче оценивания параметров регрессионной модели (2.6) по неоднородным данным. Подход, основанный на поэлементной одномерной оптимизации, позволяет получить решения без использования ограничений на матрицы регрессоров групп. Однако в общем случае он не будет совпадать с оптимизацией исходных показателей качества.

M -оценку вектора θ будем находить как решение системы оценочных уравнений

$$\sum_{i=1}^p \sum_{k=1}^{n_i} \psi_{ik} \left[r_{ik}(\hat{\theta}) \right] = 0,$$

где ψ_{ik} – векторная оценочная функция.

Выберем $\phi_j = \theta_j$, тогда

$$\frac{\partial}{\partial \theta_j} \ln g_i(z) = -\frac{\partial}{\partial z} \ln g_i(z) f_j(x_{ik})$$

и для оценочных функций типа (2.28) и (2.29) с точностью до несущественного сомножителя имеем

$$\begin{aligned} \psi_{ikj}(z) &= \left[\frac{\partial}{\partial z} \ln g_i(z) + \beta_i \right] \frac{g_i(z)}{w_i s_i(z)} f_j(x_{ik}), \\ \psi_{ikj}(z) &= \left[\frac{\partial}{\partial z} \ln g_i(z) + \beta_{ij} \right] \frac{1}{1 + b_j / g_i(z)} f_j(x_{ik}) \end{aligned}$$

соответственно, где $\psi_{ikj}(z)$ – j -й элемент вектора $\psi_{ik}(z)$.

Вектор функций $\psi_{ik}(z)$, соответствующий формуле (2.28) при $c_j = c$, имеет вид

$$\psi_{ik}(z) = \bar{\psi}_i(z) f(x_{ik}), \quad (2.30)$$

где

$$\bar{\psi}_i(z) = c \left[\frac{\partial}{\partial z} \ln g_i(z) + \beta_i \right] \frac{g_i(z)}{w_i s_i(z)},$$

что соответствует представлению оценки, полученному в п. 2.2.2.

Условно оптимальные и обобщенные компромиссные оценки получены исходя из условий, отличных от использовавшихся в п. 2.2.2, однако и они в частных случаях могут быть представлены в виде (2.30), который использовался в п. 2.2.1 и 2.2.2.

При выборе константного параметра оценочной функции условно оптимальной оценки ($\gamma_j = \gamma$, $j=1, \dots, m$, в формуле (2.29)) и $c_j = c$ вектор $\psi_{ik}(z)$ можно представить в виде (2.30) с

$$\bar{\psi}_i(z) = c \left[\frac{\partial}{\partial z} \ln g_i(z) + \beta_i \right] \frac{1}{1 + \gamma/g_i(z)},$$

что соответствует решению, полученному в п. 2.2.2. Выбор константного параметра оценочной функции в общем случае приведет к тому, что разным элементам вектора оценочной функции будут соответствовать разные значения одномерных показателей неточности $U_{0,j}$ и неустойчивости $U_{1,j}$. Тем не менее, поскольку конечной целью выбора параметров оценочной функции является управление значениями показателей неточности и неустойчивости, соответствующих всему вектору параметров θ , такое решение целесообразно.

Обобщенная компромиссная оценка может быть представлена в виде (2.30), если возможно выбрать значения параметров δ_j так, чтобы выполнялось

$$\gamma = \frac{\delta_j U_{0,j}^*}{1 - \delta_j U_{1,j}^*},$$

хотя к решению, полученному в п. 2.2.2, этот выбор не приводит.

Другим примером применения полученных результатов служит задача робастного оценивания негауссовской гетероскедастичной регрессии.

Обычно рассматривается гетероскедастичная регрессионная модель с нормальными ошибками. В работе [192] рассматривается стьюдентовская модель с оцениваемым константным параметром формы, что приводит к адаптивным робастным ММП-оценкам. Для других распределений, например, для двустороннего экспоненциального, можно использовать робастные M -оценки.

Рассмотрим следующую модель гетероскедастичной регрессии. Пусть независимые наблюдения за откликом исследуемой системы имеют распределения с плотностями

$$g_i(z_i, \phi) = g(y_i - \mu_i, \lambda_i), \quad i = 1, \dots, N,$$

где y_i – i -е наблюдение отклика, μ_i – параметр сдвига, λ_i – параметр масштаба.

Предположим, что параметры сдвига и масштаба зависят от входных переменных:

$$\begin{aligned} \mu_i &= X_i \theta, \\ \lambda_i &= \lambda(x_i, \Lambda), \end{aligned} \quad (2.31)$$

где X_i – вектор-строка регрессоров, θ – вектор неизвестных параметров модели параметра сдвига, $\lambda(x, \Lambda)$ – непрерывная положительная функция, x_i – значение вектора входных переменных, Λ – вектор неизвестных параметров модели параметра масштаба.

M -оценивание вектора параметров θ можно производить с использованием естественных модификаций подходов, рассмотренных выше. Рассмотрим оценивание вектора параметров Λ .

M -оценку $\hat{\Lambda}$ вектора Λ будем находить как решение системы оценочных уравнений (2.24) с оценочной функцией

$$\psi_i(z_i, \phi) = \chi_i[r_i(\theta), x_i, \Lambda],$$

где $r_i(\theta) = y_i - X_i \theta$.

Найдем вид оптимальных оценочных функций $\chi_i[r_i(\theta), x_i, \Lambda]$. Выбирая $\phi_j = \Lambda_j$, где Λ_j – j -й элемент вектора Λ , получим вид оценочной функции (2.28) с точностью до несущественного множителя:

$$\chi_{ij}[r_i(\theta), x_i, \Lambda] = \left\{ \frac{\partial}{\partial \lambda_i} \ln g[r_i(\theta), \lambda_i] \frac{\partial}{\partial \Lambda_j} \lambda(x_i, \Lambda) + \beta_{ij} \right\} \frac{g[r_i(\theta), \lambda_i]}{w_i s_i[r_i(\theta)]}$$

$$= \left\{ \frac{\partial}{\partial \lambda_i} \ln g[r_i(\theta), \lambda_i] + \beta_i \right\} \frac{g[r_i(\theta), \lambda_i]}{w_i s_i[r_i(\theta)]} \frac{\partial}{\partial \Lambda_j} \lambda(x_i, \Lambda) = \chi_i^\lambda[r_i(\theta), \lambda_i] \frac{\partial}{\partial \Lambda_j} \lambda(x_i, \Lambda),$$

где $\chi_{ij}[r_i(\theta), x_i, \Lambda]$ – j -й элемент вектора $\chi_i[r_i(\theta), x_i, \Lambda]$; $\chi_i^\lambda[r_i(\theta), \lambda]$ – оценочная функция, соответствующая задаче оценивания константного параметра масштаба; β_i находится из условия (2.25) и не зависит от вида параметризации (2.31).

2.2.4. ОЦЕНИВАНИЕ ПРИ ОГРАНИЧЕНИЯХ НА ПАРАМЕТРЫ

Перейдем к рассмотрению ситуации, когда на параметры уравнения регрессии (2.6) наложена система линейных ограничений-равенств вида (1.9).

Найдем вид M -оценок $\hat{\theta}_R$ вектора параметров регрессии θ . Условную минимизацию функции (2.7) осуществим методом неопределенных множителей Лагранжа. Для этого составим функцию Лагранжа вида

$$\tilde{Q}(\theta, \gamma) = Q(\theta) + \gamma^T (R\theta - d),$$

где γ – вектор неизвестных множителей Лагранжа. Продифференцируем функцию Лагранжа по элементам вектора аргументов при его значении, равном $\hat{\theta}_R, \hat{\gamma}_R$, где $\hat{\gamma}_R$ – вектор значений множителей Лагранжа, и приравняем производную нулю, получим систему

$$-\sum_{i=1}^p \sum_{j=1}^{n_i} f(x_{ij}) \psi_i[r_{ij}(\hat{\theta}_R)] + R^T \hat{\gamma}_R = 0, \quad (2.32)$$

$$R\hat{\theta}_R = d. \quad (2.33)$$

Перепишем систему (2.32) в виде

$$\sum_{i=1}^p \sum_{j=1}^{n_i} w[r_{ij}(\hat{\theta}_R)] f(x_{ij}) r_{ij}(\hat{\theta}_R) - R^T \hat{\gamma}_R = 0,$$

где $w[r_{ij}(\hat{\theta}_R)] = \frac{\Psi_i[r_{ij}(\hat{\theta}_R)]}{r_{ij}(\hat{\theta}_R)}$ – весовая функция. Отсюда справедливо

$$\hat{\theta}_R = \hat{\theta} - H^{-1}R^T \hat{\gamma}_R, \quad (2.34)$$

где

$$\hat{\theta} = H^{-1} \sum_{i=1}^p \sum_{j=1}^{n_i} w[r_{ij}(\hat{\theta}_R)] f(x_{ij}) y_{ij}, \quad H = \sum_{i=1}^p \sum_{j=1}^{n_i} w[r_{ij}(\hat{\theta}_R)] f(x_{ij}) f^T(x_{ij}),$$

H – невырожденная матрица. Для получения решения система (2.34) дополняется системой (2.33).

Формула (2.34) аналогична формуле (1.14), за исключением независимости матрицы V в (1.14) от оценки вектора $\hat{\theta}_R$. Подставляя (2.34) в (2.33), выражая из него $\hat{\gamma}_R$ и подставляя последнее выражение в (2.34), получим аналог формулы (1.15):

$$\hat{\theta}_R = \hat{\theta} + H^{-1}R^T \left[R H^{-1}R^T \right]^{-1} [d - R\hat{\theta}]. \quad (2.35)$$

Применяя метод последовательных приближений, будем решать систему (2.35) по итерационной схеме:

$$\hat{\theta}_R^{(s+1)} = \hat{\theta}^{(s+1)} + \left[H^{(s)} \right]^{-1} R^T \left\{ R \left[H^{(s)} \right]^{-1} R^T \right\}^{-1} [d - R\hat{\theta}^{(s+1)}], \quad (2.36)$$

$$H^{(s)} = \sum_{i=1}^p \sum_{j=1}^{n_i} w[r_{ij}(\hat{\theta}_R^{(s)})] f(x_{ij}) f^T(x_{ij}),$$

$$\hat{\theta}^{(s+1)} = \left[H^{(s)} \right]^{-1} \sum_{i=1}^p \sum_{j=1}^{n_i} w[r_{ij}(\hat{\theta}_R^{(s)})] f(x_{ij}) y_{ij}.$$

Формула (2.36) служит основой для применения в рассматриваемом случае итеративного МНК.

2.3. МНОГОМЕРНАЯ НОРМАЛЬНАЯ МОДЕЛЬ

2.3.1. ОЦЕНИВАНИЕ

Перейдем к рассмотрению M -оценивания регрессии в ситуации, когда вектор, составленный из наблюдений за одним или несколькими откликами системы, можно разделить на p n -мерных подвекторов (групп наблюдений) так, что ошибки наблюдений в одной группе являются коррелированными, а ошибки наблюдений в разных группах – независимыми.

Подобная модель описывает многооткликтовую регрессию, рассмотренную в п. 1.2, при этом p – количество наблюдений, а n – количество откликов. Другая ситуация, приводящая к рассматриваемой модели, возникает при экспериментах в виде серий, наблюдения в которых коррелированы во времени, например, p измерений n -точечных траекторий.

Рассмотрим модель вида (2.6) с ограничениями на параметры вида (1.9) и векторами ошибок наблюдений групп, имеющими засоренное многомерное нормальное распределение с ковариационной матрицей Σ .

Введем обозначения: X_i – $n \times m$ -матрица, составленная из строк $f^T(x_{ij})$, $j=1, \dots, n$; Y_i – n -мерный вектор, составленный из элементов y_{ij} , $j=1, \dots, n$; $q_i = q_i(\theta, \Sigma) = Y_i - X_i \theta^T \Sigma^{-1} (Y_i - X_i \theta)$.

M -оценку вектора параметров регрессии θ будем искать, минимизируя функцию

$$Q(\theta) = \sum_{i=1}^p \rho(q_i)$$

при ограничениях (1.9).

Условную минимизацию указанной функции осуществим методом неопределенных множителей Лагранжа. Составим функцию Лагранжа

$$\tilde{Q}(\theta, \gamma) = Q(\theta) + 2\gamma^T (R\theta - d),$$

продифференцируем ее по элементам вектора аргументов при его значении, равном $\hat{\theta}_R, \hat{\gamma}_R$, и приравняем производную нулю, после сокращения константного множителя получим систему

$$-\sum_{i=1}^p \rho' \left[q_i \hat{\theta}_R, \Sigma \right] X_i^T \Sigma^{-1} Y_i - X_i \hat{\theta}_R + R^T \hat{\gamma}_R = 0,$$

дополненную системой (2.33).

Применяя рассуждения, аналогичные использованным в п. 2.2.4, приходим для оценивания вектора $\hat{\theta}_R$ к уравнению (2.35) с

$$H = \sum_{i=1}^p w_1 \left[q_i \hat{\theta}_R, \Sigma \right] X_i^T \Sigma^{-1} X_i, \quad \hat{\theta} = H^{-1} \sum_{i=1}^p w_1 \left[q_i \hat{\theta}_R, \Sigma \right] X_i^T \Sigma^{-1} Y_i,$$

весовая функция при этом определяется формулой $w_1[q] = \rho'(q)$.

Если матрица Σ неизвестна, то для ее оценивания можно использовать формулу

$$\hat{\Sigma} = \left[\beta(n) \sum_{i=1}^p w_2 \left[q_i \hat{\theta}, \hat{\Sigma} \right] \right]^{-1} \sum_{i=1}^p w_2 \left[q_i \hat{\theta}, \hat{\Sigma} \right] Y_i - X_i \hat{\theta}_R \quad Y_i - X_i \hat{\theta}_R^T,$$

обобщающую формулу (1.56) на случай оценивания параметров регрессионной модели, при этом множитель $\beta(n)$ вычисляется по формуле (1.57), $w_2 q$ – весовая функция.

Оценки параметров θ и Σ можно вычислять итеративным МНК; на каждой итерации следует, применяя метод последовательных приближений, использовать веса $w_1 q_i$ и $w_2 q_i$, вычисленные по величинам q_i , которые получены на предыдущей итерации.

Для построения конкретных оценок можно воспользоваться функциями потерь из табл. 1.1, в этом случае весовые функции системы оценочных уравнений для параметров θ и Σ имеют вид

$$w_1(q) = w_2(q) = \frac{\psi(\sqrt{q})}{\sqrt{q}}.$$

Однако заметим, что некоторые из функций потерь приводят к неробастным оценкам ковариационной матрицы.

Для оценки Мешалкина весовые функции имеют вид

$$w_1(q) = w_2(q) = \exp\left[-\frac{c}{2}q\right],$$

а множитель $\beta(n)$ не зависит от n и равен $\beta(n) = \frac{1}{1+c}$ (см. п. 1.7 и [125]).

В [125] введены устойчивые оценки параметров многомерного нормального распределения, которые можно применить и для оценивания параметров регрессии (2.6), (1.9):

- стойкая оценка соответствует выбору

$$w_1(q) = \exp\left[-\frac{q}{2(n+1)}\right], \quad w_2(q) = \exp\left[-\frac{q}{n+2}\right], \quad \beta(n) = \frac{n+2}{n+4};$$

- компромиссная оценка соответствует выбору

$$w_1(q) = \left\{1 + \frac{1}{2^{n/2+1}} \exp\left[\frac{q}{2}\right]\right\}^{-1}, \quad w_2(q) = \left\{1 + \frac{1}{2^{n/2+2}} \exp\left[\frac{q}{2}\right]\right\}^{-1};$$

• радикальная оценка и ОМУ являются частными случаями оценок Мешалкина при значении параметра c , равном $1/2$ и 1 соответственно.

Рассмотрим неоднородную модель без ограничений на параметры. Предположим, что n -мерные наблюдения разделены на p групп и векторы ошибок измерений i -й группы имеют засоренное многомерное нормальное распределение с нулевым математическим ожиданием и ковариационной матрицей Σ_i .

Распространим ЭВ-оценку параметров регрессии, введенную Л.Д. Мешалкиным (см. п. 1.7 и [2]), на случай неоднородных наблюдений.

Оценочное уравнение имеет вид

$$\sum_{i=1}^p \sum_{j=1}^{n_i} (2\pi)^{-c_i n/2} |\Sigma_i|^{-c_i/2} \exp \left[-\frac{c_i}{2} Y_{ij} - X_{ij} \hat{\theta}^T \Sigma_i^{-1} Y_{ij} - X_{ij} \hat{\theta} \right] \times \\ \times X_{ij}^T \Sigma_i^{-1} Y_{ij} - X_{ij} \hat{\theta} = 0,$$

где n_i – количество наблюдений в i -й группе, c_i – скалярный параметр оценочной функции i -й группы, Y_{ij} – вектор j -го наблюдения в i -й группе, X_{ij} – $n \times m$ -матрица регрессоров j -го наблюдения в i -й группе. Для параметров Σ_i следует использовать соответствующие оценки Мешалкина, определяемые отдельно для каждой группы. Полученную оценку можно отнести к типу обобщенных радикальных.

Как показывает практика, система оценочных может иметь несколько решений, одно из которых целесообразно выбирать, ориентируясь на значения функции

$$Q = - \sum_{i=1}^p \sum_{j=1}^{n_i} \frac{1}{c_i} (2\pi)^{-c_i n/2} |\Sigma_i|^{-c_i/2} \exp \left[-\frac{c_i}{2} Y_{ij} - X_{ij} \hat{\theta}^T \Sigma_i^{-1} Y_{ij} - X_{ij} \hat{\theta} \right].$$

При этом локальный минимум функции Q будет соответствовать лишь оценке вектора параметров θ (получаемая путем оптимизации функции Q оценка матрицы Σ_i не будет удовлетворять условию асимптотической несмещенности).

В то же время функция Q не ограничена снизу: $Q \rightarrow -\infty$ при $|\hat{\Sigma}_i| \rightarrow 0$ хотя бы для одной i -й группы (если в этой группе имеется хотя бы один нулевой остаток), что вообще характерно для ограниченных функций потерь (см. п. 1.7).

Решение, соответствующее глобальному минимуму функции Q , лежит на краю параметрического пространства. Данная проблема, как справедливо замечено в [154], напоминает проблему неограниченности функции правдоподобия при ММП-оценивании параметров смеси нормальных распределений с неравными ковариационными матрицами (см. [94, 169] и п. 3.2). Как показывает практика, решение возникающей проблемы может быть также аналогичным: необходимо вы-

бирать другое решение системы оценочных уравнений, лежащее внутри параметрического пространства и не приводящее к малому (минимальному) количеству наблюдений с ненулевыми весами. При практических вычислениях целесообразно ограничивать снизу детерминанты оценок ковариационных матриц и количество наблюдений с существенно ненулевыми весами (см. также [169]). При увеличении количества откликов может потребоваться уменьшить значения параметров оценочной функции, как это происходит в стойкой оценке.

2.3.2. ИССЛЕДОВАНИЕ

Для проверки работоспособности изложенных в п. 2.3.1 подходов к устойчивому оцениванию параметров при засоренном многомерном нормальном распределении ошибок было проведено исследование. Использовался метод статистических испытаний (метод Монте-Карло), основанный на многократном генерировании выборки с последующим нахождением по ней оценок параметров.

Данные генерировались в соответствии с линейной двухоткликковой моделью

$$y_1 = 20 + 6x_1 + 4,6x_2 + e_1, \quad y_2 = 21 + 7x_1 + 0,5x_2 + e_2,$$

где y_i – значение i -го отклика, x_i – значение i -го фактора ($x_i \in [-1, 1]$). При оценивании на параметры накладывались ограничения

$$\theta_{20} - \theta_{10} = 1, \quad \theta_{21} - \theta_{11} = 1,$$

где θ_{kl} – параметр при факторе x_l в модели k -го отклика ($x_0 = 1$).

Ошибки наблюдений имели засоренное двумерное нормальное распределение с коэффициентом корреляции 0,9. Значения дисперсий откликов основного распределения: 0,116740 – для первого отклика, 0,098225 – для второго отклика. Засоряющие наблюдения имели независимые элементы вектора ошибок, представлявшие собой модули нормальных случайных величин с дисперсией, в 100 раз превышающей дисперсию соответствующего элемента основного распределения. Уровень засорения $\alpha = 0,2$.

Кроме робастных оценок, нацеленных на засоренное нормальное распределение, в исследованиях использовались МНК-оценка, квази-правдоподобная (КП) оценка (ММП-оценка при нормальном распределении) и адаптивная робастная ММП-оценка, полученная исходя из распределения Стьюдента.

Мерой, отражающей полезность разработанных подходов, служила величина

$$\tau = \theta - \hat{\theta}_R^T \theta - \hat{\theta}_R . \quad (2.37)$$

Средние значения величины τ , полученные по 100 экспериментам, отражены в табл. 2.1.

Таблица 2.1

Результаты исследования

Оценка	Значения параметров	Значение τ
Хьюбера	1,5	0,1643019
Эндрюса	2,1	0,07259289
Рамсея	0,3	0,09184191
Биквадратная Тьюки	6	0,01998339
Тьюки	1,5	0,05599482
Меррилла–Швеппе	1,5	0,0229515
Энскомба	$b = 1,5, c = 3$	0,01985761
Хампеля	$a = 1,7, b = 3,4, c = 8,5$	0,04369974
Мешалкина	1/9	0,05518168
Стойкая	–	0,0200843
Компромиссная	–	0,02920227
Радикальная	–	0,02641671
КП	–	0,7253716
МНК	–	0,9238528
Адаптивная робастная	–	0,02829866

Результаты исследований свидетельствуют о работоспособности предложенных процедур оценивания и о преимуществе используемых оценок перед МНК-оценкой и квазиравдоподобной оценкой, полученной исходя из нормального распределения.

Ряд робастных оценок по качеству превосходит адаптивную робастную оценку, полученную исходя из распределения Стьюдента. Оценка Хьюбера, не имеющая горизонтальной асимптоты, показывает результат, худший по сравнению с другими робастными оценками, включая и оценку Меррилла–Швеппе, также обладающую указанным свойством.

Исследование выявляет высокое качество модификаций введенных А.М. Шурыгиным оценок, которые не имеют неопределяемых параметров.

2.4. ДВУСТОРОННЯЯ ЭКСПОНЕНЦИАЛЬНАЯ МОДЕЛЬ

2.4.1. АДАПТИВНОЕ ОЦЕНИВАНИЕ

Рассмотрим регрессионную модель вида (1.1) с однородными ошибками наблюдений, имеющими двустороннее экспоненциальное распределение (1.52).

В предположении, что параметр формы известен, ММП-оценки вектора параметров θ находятся минимизацией функции [18]:

$$Q(\theta) = \sum_{i=1}^N |r_i(\theta)|^V. \quad (2.38)$$

Оценки вектора θ , доставляющие в соответствии с (2.38) минимум вектору регрессионных остатков в метрике Гельдера (L_V -метрике), носят название L_V -оценок [18, 97] и включают как частные случаи МНК-оценки, оценки по методу наименьших модулей, чебышевские (минимаксные) оценки (см. п. 1.6). Заметим также, что функция (2.38) не зависит от λ , по этой причине оценки вектора θ обладают свойством инвариантности к параметру масштаба.

Если требуется оценить параметр масштаба, то его ММП-оценка вычисляется по формуле

$$\hat{\lambda} = \left[\frac{\nu}{N} \sum_{i=1}^N |r_i(\hat{\theta})|^\nu \right]^{1/\nu}, \quad (2.39)$$

где $\hat{\theta}$, $\hat{\lambda}$ – ММП-оценки соответствующих параметров (см. [97, 170]).

При адаптивном оценивании, наряду с параметрами уравнения регрессии и параметром масштаба, оценивается параметр формы [97]. Минимизируемая при адаптивном оценивании функция имеет вид

$$Q(\theta, \lambda, \nu) = \frac{1}{\lambda^\nu} \sum_{i=1}^N |r_i(\theta)|^\nu + N \ln \left[\lambda \Gamma \left(1 + \frac{1}{\nu} \right) \right]. \quad (2.40)$$

Для оценивания параметра формы система оценочных уравнений дополняется уравнением

$$\sum_{i=1}^N \left(\frac{|r_i(\hat{\theta})|}{\hat{\lambda}} \right)^{\hat{\nu}} \ln \frac{|r_i(\hat{\theta})|}{\hat{\lambda}} - \frac{N}{\hat{\nu}^2} \text{DG} \left(1 + \frac{1}{\hat{\nu}} \right) = 0, \quad (2.41)$$

где $\hat{\nu}$ – ММП-оценка параметра формы, DG – логарифмическая производная гамма-функции.

Если подставить в (2.40) выражение для оценки параметра масштаба (2.39), то можно получить, вместо (2.41), уравнение правдоподобия, не зависящее от параметра масштаба (см. [97]).

При построении вычислительных алгоритмов оценивания удобно использовать покомпонентный метод с выделением следующих подмножеств оцениваемых параметров:

$$\theta_i, i = 1, \dots, m, \quad \lambda, \quad \nu.$$

При этом этап уточнения $\hat{\lambda}$ не является итеративным и заключается в вычислении по формуле (2.39). Для решения нелинейного уравнения на этапе уточнения $\hat{\nu}$ можно применить итерационные методы, например, метод Ньютона.

Рассмотрим особенности этапа уточнения $\hat{\theta}$. Здесь удобно использовать ИМНК, при этом, однако, возникает ряд трудностей.

Весовая функция имеет вид [18]

$$w(z, \nu) = |z|^{\nu-2}.$$

При $\nu > 2$ сходимость ИМНК не гарантируется (см. п. 1.5). Для обеспечения сходимости очередное полученное на текущей s -й итерации ИМНК значение оценки, обозначенные как $\tilde{\theta}_0^{(s)}$, модифицируют следующим образом [18]:

$$\tilde{\theta}^{(s)} = (1 - \kappa)\tilde{\theta}^{(s-1)} + \kappa\tilde{\theta}_0^{(s)}, \quad (2.42)$$

где $\tilde{\theta}^{(s)}$ – окончательное значение оценки на s -й итерации, $\kappa = \frac{1}{\nu-1}$.

В случае $\nu < 2$ весовая функция в окрестности нуля является неограниченно возрастающей, и при реальных вычислениях ее необходимо ограничивать, например, заменяя исходную функцию потерь (2.38) в окрестности нуля на квадратичную [81].

Наряду с ММП предложены другие методы оценивания параметра формы, при которых в системе уравнений правдоподобия соответствующее уравнение заменяется некоторым другим оценочным уравнением. Необходимость в разработке этих методов была продиктована как стремлением уменьшить трудоемкость вычислений, так и низким качеством ММП-оценки при малом числе наблюдений [170].

Один из подходов – использование оценочного уравнения, приравнивающего теоретическое и выборочное значения коэффициента эксцесса [84]

$$\beta(4) = \frac{\mu_4}{\sigma^4},$$

где μ_q – абсолютный момент распределения порядка q , σ – корень квадратный из дисперсии. Оценочное уравнение имеет вид

$$\beta_{\hat{\nu}}(4) = \hat{\beta}(4), \quad (2.43)$$

где $\beta_\nu(4)$ – коэффициент эксцесса двустороннего экспоненциального распределения с параметром формы ν :

$$\beta_\nu(4) = \frac{\Gamma(5/\nu) \Gamma(1/\nu)}{[\Gamma(3/\nu)]^2},$$

$\hat{\beta}(4)$ – выборочное значение коэффициента эксцесса:

$$\hat{\beta}(4) = \frac{1}{N} \sum_{i=1}^N r_i^4(\hat{\theta}) \left[\frac{1}{N} \sum_{i=1}^N r_i^2(\hat{\theta}) \right]^{-2}.$$

При вычислении значения оценки параметра формы вместо поиска корня уравнения (2.43) в качестве его приближенного значения можно использовать аппроксимирующее выражение [84]:

$$\hat{\nu} = \frac{1,46}{\ln \left[\hat{\beta}(4) - 2/9 - 10,7/\hat{\beta}^7(4) \right] - 0,289}.$$

В [97] предлагается для оценивания параметра формы при $1 \leq \nu \leq 2$ использовать уравнение

$$\beta_{\hat{\nu}}(1) = \hat{\beta}(1),$$

где $\beta_\nu(1) = \frac{\mu_1}{\sigma}$ – отношение соответствующих величин для двустороннего экспоненциального распределения с параметром формы ν , равное

$$\beta_\nu(1) = \frac{\Gamma(2/\nu)}{[\Gamma(1/\nu) \Gamma(3/\nu)]^{1/2}},$$

$\hat{\beta}(1)$ – выборочное значение показателя $\beta_\nu(1)$:

$$\hat{\beta}(1) = \frac{1}{N} \sum_{i=1}^N |r_i(\hat{\theta})| \left[\frac{1}{N} \sum_{i=1}^N r_i^2(\hat{\theta}) \right]^{-1/2}.$$

Такое же уравнение используется в [170], где по результатам моделирования делается вывод, что данный подход приводит к оценкам параметра формы, которые кажутся менее смещенными и более эффективными, чем ММП-оценки.

Перейдем к рассмотрению неоднородного случая. Рассмотрим регрессионную модель вида (2.6). Пусть ошибки наблюдений i -й группы имеют двустороннее экспоненциальное распределение с нулевым параметром сдвига, параметром масштаба λ_i и параметром формы ν_i .

При ММП-оценивании неизвестных параметров θ , $\lambda_1, \dots, \lambda_p$, ν_1, \dots, ν_p минимизируемая функция имеет вид

$$Q = \sum_{i=1}^p \left\{ \frac{1}{\lambda_i^{\nu_i}} \sum_{j=1}^{n_i} |r_{ij}(\theta)|^{\nu_i} + n_i \ln \left[\lambda_i \Gamma \left(1 + \frac{1}{\nu_i} \right) \right] \right\}.$$

Поиск оценок можно осуществлять с использованием покомпонентного оценивания, при этом множество оцениваемых параметров разделяется на подмножества $\{\theta\}$, $\{\lambda_i\}$, $\{\nu_i\}$, $i=1, \dots, p$. Оценки вектора θ определяются формулой (2.35) с весовой функцией

$$w(z; \lambda_i, \nu_i) = \frac{\nu_i}{\lambda_i^{\nu_i}} |z|^{\nu_i - 2}$$

для наблюдений i -й группы и могут быть вычислены посредством ИМНК. В случае, когда выполняется неравенство $\nu_i > 2$ хотя бы для одного i , необходима коррекция решения вида (2.42) на очередной итерации ИМНК, при этом

$$\kappa = \frac{1}{\max_i \nu_i - 1}.$$

При $\nu_i < 2$ требуется ограничивать рост весовой функции в окрестности нуля.

2.4.2. ОЦЕНИВАНИЕ ПРИ БАЙЕСОВСКОМ ТОЧЕЧНОМ ЗАСОРЕНИИ

Рассмотрим адаптивное робастное L_ν -оценивание неоднородной модели в предположении, что i -я группа имеет двустороннее экспоненциальное распределение с нулевым параметром сдвига, параметром масштаба λ_i , параметром формы $\nu_i > 1$ и выполнено (2.15).

Согласно (2.22) условно оптимальная оценка $\hat{\theta}$ имеет оценочную функцию (с точностью до несущественного множителя)

$$\begin{aligned} \psi_i(r_{ij}(\hat{\theta}), \lambda_i, \nu_i) &= \\ &= \frac{\nu_i |r_{ij}(\hat{\theta})|^{\nu_i-1} \text{sign}[r_{ij}(\hat{\theta})]}{\lambda_i^{\nu_i}} \left[1 + 2\gamma\lambda_i\Gamma\left(1 + \frac{1}{\nu_i}\right) \exp\left(\frac{|r_{ij}(\hat{\theta})|^{\nu_i}}{\lambda_i^{\nu_i}}\right) \right]^{-1}. \end{aligned}$$

Для определения величины γ , соответствующей обобщенной компромиссной оценке, в формулу (2.23) подставляются значения u_0^* и u_1^* :

$$u_0^* = \left[\sum_{i=1}^p k_i \frac{\nu_i^2}{\lambda_i^2} \frac{\Gamma(2-1/\nu_i)}{\Gamma(1/\nu_i)} \right]^{-1}, \quad u_1^* = \left[\sum_{i=1}^p \frac{k_i}{2^{3-1/\nu_i}} \frac{\nu_i^3}{\lambda_i^3} \frac{\Gamma(2-1/\nu_i)}{\Gamma^2(1/\nu_i)} \right]^{-1}.$$

Получим обобщенные радикальные оценки $\hat{\theta}$, $\hat{\lambda}_i$, $\hat{\nu}_i$, $i=1, \dots, p$. В случае $\nu_i=2$ данные оценки эквивалентны оценкам Мешалкина (ЭВ-оценкам) [2], в связи с чем они могут быть названы экспоненциально-взвешенными L_ν -оценками (ЭВ- L_ν -оценками). С точностью до несущественных множителей соответствующие оценочные функции имеют следующий вид:

- для $\hat{\theta}$

$$\psi_i[r_{ij}(\hat{\theta}), \lambda_i, \nu_i] = \left\{ \frac{\nu_i}{\lambda_i^{\nu_i}} |r_{ij}(\hat{\theta})|^{\nu_i-1} \text{sign}[r_{ij}(\hat{\theta})] \right\} a_{\delta_i} \exp\left\{ -\delta_i \frac{|r_{ij}(\hat{\theta})|^{\nu_i}}{\lambda_i^{\nu_i}} \right\},$$

где $a_{\delta_i} = \frac{1}{[2\lambda_i \Gamma(1+1/\nu_i)]^{\delta_i}}$;

- для $\hat{\lambda}_i$

$$\chi_i^\lambda [r_{ij}(\theta), \hat{\lambda}_i, \nu_i] = \left[\frac{\nu_i}{\hat{\lambda}_i^{\nu_i+1}} |r_{ij}(\theta)|^{\nu_i} - \frac{1}{\hat{\lambda}_i} + \beta_i^\lambda \right] \frac{\exp -\delta_i |r_{ij}(\theta)|^{\nu_i} / \hat{\lambda}_i^{\nu_i}}{[2\hat{\lambda}_i \Gamma(1+1/\nu_i)]^{\delta_i}},$$

где $\beta_i^\lambda = \frac{\delta_i}{1+\delta_i} \frac{1}{\hat{\lambda}_i}$;

- для $\hat{\nu}_i$

$$\chi_i^\nu [r_{ij}(\theta), \lambda_i, \hat{\nu}_i] = \left[-\frac{|r_{ij}(\theta)|^{\hat{\nu}_i}}{\lambda_i^{\hat{\nu}_i}} \ln \left(\frac{|r_{ij}(\theta)|}{\lambda_i} \right) + \frac{DG(1+1/\hat{\nu}_i)}{\hat{\nu}_i^2} + \beta_i^\nu \right] \times$$

$$\times \frac{\exp -\delta_i |r_{ij}(\theta)|^{\hat{\nu}_i} / \lambda_i^{\hat{\nu}_i}}{2\lambda_i \Gamma(1+1/\hat{\nu}_i)^{\delta_i}},$$

где $\beta_i^\nu = -\frac{\ln(1+\delta_i) + \delta_i DG(1+1/\hat{\nu}_i)}{\hat{\nu}_i^2(1+\delta_i)}$.

Нормированные обобщенные радикальные оценки $\hat{\theta}$, $\hat{\lambda}_i$, $\hat{\nu}_i$, $i=1, \dots, p$, задаются теми же соотношениями с

$$a_{\delta_i} = \frac{1}{(1-\delta_i)^{1/\nu_i}}.$$

Данные оценки могут быть названы *нормированными ЭВ- L_ν -оценками* и *НЭВ- L_ν -оценками*.

Найдем вид *стойких* L_ν -оценок. Рассмотрим оценки с оценочными функциями вида (2.19) в случае, если функция $s_i(z)$ является плотно-

стью двустороннего экспоненциального распределения с нулевым параметром сдвига, параметром масштаба $\lambda_i \omega_i$ и параметром формы ν_i , $\omega_i = 1$. Легко увидеть, что данные оценки являются НЭВ-оценками с

$$\delta_i = 1 - \frac{1}{\omega_i^{\nu_i}}.$$

Значения параметров ω_i или δ_i определим, максимизируя показатель u_s (2.16) для параметров уравнения регрессии или аналогичные ему показатели для мешающих параметров. Обозначим оптимальные значения параметров ω_i , δ_i через ω_i^* , δ_i^* . В результате, стойкая оценочная функция и наборы параметров ω_i^* или δ_i^* являются решением задачи максиминной оптимизации показателя u_s или его аналогов.

Для стойкой оценки $\hat{\theta}$ имеем

$$(\delta_1^*, \dots, \delta_p^*) = \arg \max_{\delta_1, \dots, \delta_p} \left[\sum_{i=1}^p k_i \frac{\nu_i^2 \Gamma(2 - 1/\nu_i)}{\lambda_i^2 \Gamma(1/\nu_i) (1 + \delta_i)^{2-1/\nu_i} (1 - \delta_i)^{1/\nu_i}} \right]^{-1},$$

откуда следует

$$\delta_i^* = \arg \max_{\delta_i} (1 + \delta_i)^{2-1/\nu_i} (1 - \delta_i)^{1/\nu_i}.$$

Решением является

$$\delta_i^* = 1 - \frac{1}{\nu_i}, \quad \omega_i^* = \nu_i^{1/\nu_i}.$$

Для стойкой оценки $\hat{\lambda}_i$ получаем

$$\delta_i^* = \arg \max_{\delta_i} \frac{\lambda_i^2}{\nu_i} (1 + \delta_i)^{2+1/\nu_i} (1 - \delta_i)^{1/\nu_i}.$$

Решением является

$$\delta_i^* = \frac{\nu_i}{1 + \nu_i}, \quad \omega_i^* = (1 + \nu_i)^{1/\nu_i}.$$

Для стойкой оценки \hat{v}_i имеем

$$\delta_i^* = \arg \max_{\delta_i} (1 + \delta_i)^{2+1/v_i} (1 - \delta_i)^{1/v_i} \left\{ \left(1 + \frac{1}{v_i} \right) \text{DG}' \left(1 + \frac{1}{v_i} \right) + \left[\text{DG} \left(1 + \frac{1}{v_i} \right) - \ln(1 + \delta_i) \right]^2 + 2 \left[\text{DG} \left(1 + \frac{1}{v_i} \right) - \ln(1 + \delta_i) \right] \right\}^{-1}.$$

Значение δ_i^* является корнем уравнения

$$(\delta_i - 1) \left[\text{DG} \left(1 + \frac{1}{v_i} \right) - \ln(1 + \delta_i) + 1 \right] + \frac{\delta_i - v_i + \delta_i v_i}{v_i} \left\{ \left(1 + \frac{1}{v_i} \right) \text{DG}' \left(1 + \frac{1}{v_i} \right) + \left[\text{DG} \left(1 + \frac{1}{v_i} \right) - \ln(1 + \delta_i) \right]^2 + 2 \left[\text{DG} \left(1 + \frac{1}{v_i} \right) - \ln(1 + \delta_i) \right] \right\} = 0.$$

Зависимости значений δ^* и ω^* от значения параметра формы v для стойких оценок $\hat{\theta}$, $\hat{\lambda}$ и \hat{v} представлены на рис. 2.1 и 2.2.

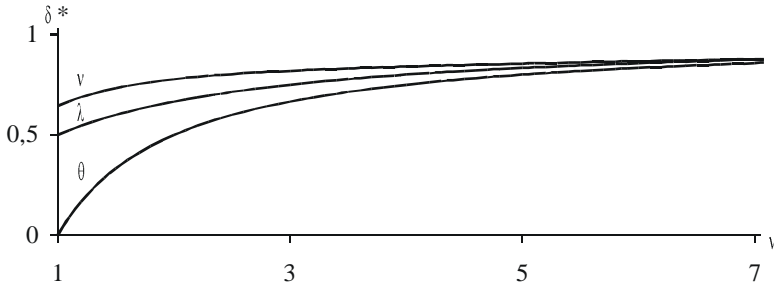
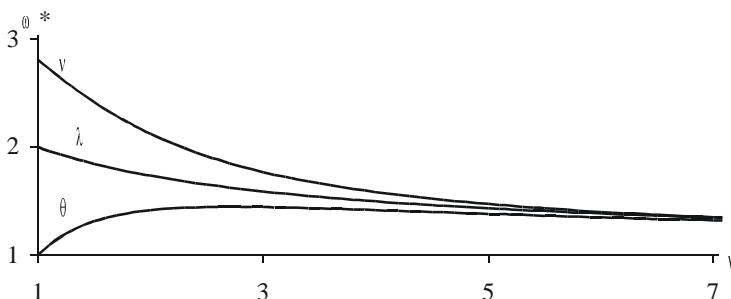


Рис. 2.1. Зависимости δ^* от v

Рис. 2.2. Зависимости ω^* от ν

Заметим, что при $\nu_i < 2$ производные полученных в данном пункте оценочных функций являются неограниченными и вопросы оптимальности оценочных функций для этого случая требуют отдельного рассмотрения.

Поиск значений оценок осуществляется, как и для ММП, методом покомпонентного оценивания. При использовании на этапе уточнения оценок вектора параметров θ ИМНК в случае $\nu_i > 2$ необходимо использовать модификацию типа (2.42), а в случае $\nu_i < 2$ — ограничивать значения весовой функции в окрестности нуля.

Хотя решения в теории робастности могут формулироваться в виде оптимальной оценочной функции, часто требуется знание вида функции потерь, например, при выборе одного решения оценочного уравнения из нескольких. Рассмотрим этот вопрос для ЭВ- L_ν -оценок.

Если наблюдения являются однородными (имеется только одна группа наблюдений), а параметр формы не оценивается, то ЭВ- L_ν -оценки параметров θ и λ соответствуют локальному минимуму функции

$$Q(\theta, \lambda) = -\frac{\lambda^{\delta^2/(1+\delta)}}{\delta} \sum_{i=1}^N g^\delta[r_i(\theta), \lambda, \nu] \quad (2.44)$$

(индекс группы опущен).

Минимизация функции (2.44) по параметру формы приводит к оценке, не удовлетворяющей условию асимптотической несмещенности. Однако, получив ряд решений исходной системы оценочных

уравнений, мы можем выбрать из них то, которому соответствует меньшее значение функции (2.44).

Рассмотрим неоднородный случай.

Теорема 2.2. Пусть ϕ – вектор всех оцениваемых параметров модели, тогда локальный минимум функции

$$Q(\phi) = -\sum_{i=1}^p \frac{1}{\delta_i} \sum_{j=1}^{n_i} g_i^{\delta_i} [r_{ij}(\theta), \lambda_i, v_i]$$

удовлетворяет системе оценочных уравнений

$$\sum_{i=1}^p \sum_{j=1}^{n_i} g_i^{\delta_i} [r_{ij}(\hat{\theta}), \hat{\lambda}_j, \hat{v}_j] \left. \frac{\partial}{\partial \phi} \ln g_i [r_{ij}(\theta), \lambda_i, v_i] \right|_{\phi=\hat{\phi}} = 0.$$

Заметим, что оценки для параметров λ_i и v_i в теореме 2.2 не удовлетворяют условию асимптотической несмещенности. Тем не менее полученная функция может использоваться для выбора одного из решений системы оценочных уравнений.

Рассмотрим оценивание гетероскедастичной регрессии, наблюдения которой имеют двустороннее экспоненциальное распределение с общим параметром формы v для всех наблюдений.

Предположим, что функция (2.31) имеет вид

$$\lambda(x_i, \Lambda) = \exp \tilde{X}_i \Lambda,$$

где \tilde{X}_i – вектор-строка регрессоров, и справедливо равенство

$$w_i s_i [r_i(\theta)] = g^{1-\delta} [r_i(\theta), \lambda_i, v].$$

Оценочная функция обобщенной радикальной оценки вектора параметров Λ с точностью до несущественного множителя имеет вид

$$\chi[r_i(\theta), x_i, \hat{\Lambda}, v] = \left[\frac{v |r_i(\theta)|^v}{\exp v \tilde{X}_i \hat{\Lambda}} - \frac{1}{1+\delta} \right] \frac{1}{\exp \delta \tilde{X}_i \hat{\Lambda}} \exp \left[-\delta \frac{|r_i(\theta)|^v}{\exp v \tilde{X}_i \hat{\Lambda}} \right] \tilde{X}_i^T.$$

При совместном оценивании всех параметров модели соответствующие им оценочные уравнения объединяются в общую систему.

Вид оценочных функций обобщенной радикальной оценки параметров θ и ν аналогичен полученному выше.

Оценки удобно вычислять методом покомпонентного оценивания. В рассматриваемом случае удобно выделить три подмножества: вектор θ , вектор Λ и скаляр ν .

На этапе уточнения значения оценки вектора Λ система оценочных уравнений часто имеет несколько решений, одно из которых можно выбрать исходя из минимизации вспомогательной целевой функции. Могут использоваться две вспомогательные целевые функции вида

$$Q_1 = -\frac{1}{\delta} \sum_{i=1}^N \lambda_i \delta^{2/(1+\delta)} g^\delta[r_i(\theta), \lambda_i, \nu], \quad Q_2 = -\frac{1}{\delta} \sum_{i=1}^N g^\delta[r_i(\theta), \lambda_i, \nu].$$

Локальные оптимумы обеих функций не соответствуют решениям исходной системы оценочных уравнений, однако часто бывают близки к ним. Как показывает практика, функция Q_1 предпочтительнее. Целесообразно использовать данные функции и при выборе начального приближения при итеративном поиске оценок.

Также возможны ситуации, когда решение, ближайшее к глобальному оптимуму, не является предпочтительным, поскольку при нем почти все наблюдения игнорируются, т. е. воспринимаются как аномальные. Такие ситуации сходны с обсуждавшимися в случае использования оценок Мешалкина при неоднородных данных. Как показывает практика, в этом случае возможно выбрать другое (удовлетворительное) решение системы оценочных уравнений, не ориентированное на малое (минимальное) количество наблюдений. При организации вычислений целесообразно ограничивать как область изменения параметра масштаба, так и количество игнорируемых наблюдений.

2.4.3. МИНИМАКСНЫЙ ПОДХОД

При однородных наблюдениях для конструирования оценок может быть использован минимаксный подход. Получим с его помощью наилучшую оценку $\hat{\theta}$ при наименее благоприятной плотности в классе засоренных двусторонних экспоненциальных распределений. Наименее благоприятная плотность имеет вид (1.53) при $\nu > 1$.

В рассматриваемом подходе параметры масштаба и формы считаются известными. При неизвестных значениях данных параметров их оценки необходимо искать совместно с оценкой $\hat{\theta}$, например, посредством ММП-оценивания при наименее благоприятной плотности [121].

При неоднородных наблюдениях можно использовать квазиравдоподобные оценки, получаемые по ММП для наименее благоприятной плотности; назовем данные оценки квазиминимаксными (КММ). Таким образом, квазиминимаксные L_v -оценки, получаемые по ММП в условиях, когда i -я группа наблюдений имеет распределение с плотностью $g(z; \lambda_i, v_i, \alpha_i)$ вида (1.53), задаются следующими оценочными функциями:

- для оценки $\hat{\theta}$

$$\begin{aligned} \Psi_i \left[r_{ij}(\hat{\theta}), \lambda_i, v_i \right] &= \frac{v_i}{\lambda_i^{v_i}} \left| r_{ij}(\hat{\theta}) \right|^{v_i-1} \text{sign} \left[r_{ij}(\hat{\theta}) \right] J \left(\frac{|r_{ij}(\hat{\theta})|}{\lambda_i} < c_i \right) + \\ &+ \frac{v_i}{\lambda_i} c_i^{v_i-1} \text{sign} \left[r_{ij}(\hat{\theta}) \right] J \left(\frac{|r_{ij}(\hat{\theta})|}{\lambda_i} \geq c_i \right), \end{aligned}$$

где $J(\cdot)$ – функция, равная единице, если значение аргумента – «истина», и равная нулю, если значение аргумента – «ложь»;

- для оценки $\hat{\lambda}_i$

$$\begin{aligned} \chi_i^\lambda \left[r_{ij}(\theta), \hat{\lambda}_i, v_i \right] &= \\ &= \frac{\beta_i^\lambda}{\hat{\lambda}_i} - \frac{v_i}{\hat{\lambda}_i^{v_i+1}} \left| r_{ij}(\theta) \right|^{v_i} J \left(\frac{|r_{ij}(\theta)|}{\hat{\lambda}_i} < c_i \right) - \frac{v_i c_i^{v_i-1}}{\hat{\lambda}_i^2} \left| r_{ij}(\theta) \right| J \left(\frac{|r_{ij}(\theta)|}{\hat{\lambda}_i} \geq c_i \right), \end{aligned}$$

где $\beta_i^\lambda = 1$;

- для оценки $\hat{\nu}_i$

$$\begin{aligned} \chi_i^y [r_{ij}(\theta), \lambda_i, \hat{\nu}_i] &= \frac{|r_{ij}|^{\hat{\nu}_i}}{\lambda_i^{\hat{\nu}_i}} \ln \left(\frac{|r_{ij}|}{\lambda_i} \right) J \left(\frac{|r_{ij}|}{\lambda_i} < c_i \right) + \\ &+ \left\{ c_i^{\hat{\nu}_i-1} [1 + \hat{\nu}_i \ln c_i] \frac{|r_{ij}|}{\lambda_i} - c_i^{\hat{\nu}_i} [1 + \hat{\nu}_i - 1 \ln c_i] \right\} J \left(\frac{|r_{ij}|}{\lambda_i} \geq c_i \right) - \beta_i^y, \end{aligned}$$

где $\beta_i^y = \frac{1}{\hat{\nu}_i^2} \text{DG} \left(1 + \frac{1}{\hat{\nu}_i} \right)$.

В теории робастности модельным распределением является незаурядное двустороннее экспоненциальное, оценки при таком распределении должны быть асимптотически несмещенными. Квазиминимаксные L_ν -оценки, удовлетворяющие данному условию, задаются теми же оценочными функциями с константами:

$$\beta_i^\lambda = \frac{\gamma [1 + 1/\nu_i, c_i^{\nu_i}] + c_i^{\nu_i-1} \Gamma [2/\nu_i, c_i^{\nu_i}]}{\Gamma [1 + 1/\nu_i]},$$

$$\begin{aligned} \beta_i^y &= \left\{ \frac{1}{\hat{\nu}_i} \gamma'_1 \left(1 + \frac{1}{\hat{\nu}_i}, c_i^{\hat{\nu}_i} \right) + c_i^{\hat{\nu}_i-1} [1 + \hat{\nu}_i \ln c_i] \Gamma \left(\frac{2}{\hat{\nu}_i}, c_i^{\hat{\nu}_i} \right) - \right. \\ &\left. - c_i^{\hat{\nu}_i} [1 + \hat{\nu}_i - 1 \ln c_i] \Gamma \left(\frac{1}{\hat{\nu}_i}, c_i^{\hat{\nu}_i} \right) \right\} \frac{1}{\Gamma [1/\hat{\nu}_i]}, \end{aligned}$$

где $\Gamma(a, x) = \Gamma(a) - \gamma(a, x)$, $\gamma'_1(a, x) = \frac{\partial}{\partial a} \gamma(a, x)$.

Заметим, что использование полученных оценочных функций приводит к оценкам мешающих параметров, не являющимся робастными (например, соответствующие функции влияния не ограничены).

Поиск значений оценок можно осуществлять методом покомпонентного оценивания, при использовании ИМНК необходимо использовать те же модификации, что и для ММП.

2.4.4. ИССЛЕДОВАНИЕ

Рассмотрим результаты исследования свойств адаптивных и адаптивных робастных оценок параметров линейной регрессионной модели в условиях неоднородности данных.

Исследовались правдоподобные оценки (ММП-ДСЭ), полученные в условиях двустороннего экспоненциального распределения ошибок, ЭВ- L_v -оценки (ЭВ), стойкие оценки и квазиминимаксные оценки, робастные в окрестности двустороннего экспоненциального распределения, формируемой путем его засорения.

Рассматривалась линейная регрессионная модель вида

$$y = \theta_1 x_1 x_2 x_3 + \theta_2 x_2^2 + \theta_3 x_1^2 x_3 + \theta_4 x_3^3 + \theta_5 x_1 x_2^2 + e$$

со значениями факторов x_i , $i=1, 2, 3$, изменяющимися на отрезке $[-1, 1]$, и с единичными значениями параметров.

Параметр ЭВ-оценок был выбран равным 0,5, уровень засорения для КММ-оценок – 0,05.

Набор данных состоял из двух разнораспределенных подвыборок. Одна из них была более информативная (в смысле информации Фишера для параметра сдвига) и содержала aN наблюдений, где a – доля наблюдений первой подвыборки, $[\cdot]$ – целая часть числа; ошибки имели двустороннее экспоненциальное распределение с нулевым математическим ожиданием, параметром формы 8 и дисперсией, равной 0,0025 % от мощности сигнала. Другая подвыборка была менее информативной и содержала $N - aN$ наблюдений, ошибки имели двустороннее экспоненциальное распределение с нулевым математическим ожиданием, параметром формы 1,5 и дисперсией, равной 0,25 % от мощности сигнала.

Мощность сигнала для i -й подвыборки определялась формулой

$$P_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \tilde{y}_{ij}^2 - \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \tilde{y}_{ij} \right)^2,$$

где \tilde{y}_{ij} – истинное (незашумленное) значение отклика.

Сравнивались оценки, определяемые с учетом имеющейся неоднородности данных и без ее учета. Мерой качества служила величина (2.37) при $\hat{\theta}_R = \hat{\theta}$. Значения меры качества оценок усреднялись по 300 экспериментам.

Результаты при $N = 216$ приведены в табл. 2.2, при $N = 1000$ – в табл. 2.3.

Заметим, что процедура КММ-оценивания может быть неустойчивой в условиях нарушения предположений, при которых оценки получены, в связи с этим в одном случае (для каждой таблицы) результат получить не удалось.

По результатам проведенных исследований можно сделать следующие выводы:

- наиболее качественными оценками в случае учета неоднородности, как правило, являются ММП-оценки (ошибки наблюдений подвыборок имели чистое двустороннее экспоненциальное распределение);
- при большом значении параметра a сохраняется довольно высокое качество робастных ЭВ-оценок и стойких оценок, не учитывающих неоднородность, при уменьшении параметра a неоднородность необходимо учитывать;
- учет неоднородности, как правило, позволяет существенно повысить качество ММП-оценок и КММ-оценок.

Целью второй части исследования было сравнение робастных и адаптивных робастных оценок между собой и с неробастными (адаптивными и неадаптивными) оценками, а также сравнение этих оценок при учете неоднородности данных и без ее учета.

Помимо названных выше в исследование были включены правдоподобные оценки при студентовском распределении ошибок (ММП-Ст), неадаптивные ММП-оценки и робастные оценки при модельном нормальном распределении (т. е. двустороннем экспоненциальном распределении с параметром формы $\nu = 2$ или распределении Стюдента с параметром формы $\nu = \infty$).

Таблица 2.2

Качество оценивания при $N = 216$

Значение a	Модель	Оценка			
		Стойкая	ЭВ	ММП-ДСЭ	КММ
1	Однородная	$4,764 \cdot 10^{-6}$	$3,534 \cdot 10^{-6}$	$3,038 \cdot 10^{-6}$	$3,396 \cdot 10^{-6}$
0,9	Неоднородная	$6,092 \cdot 10^{-6}$	$4,552 \cdot 10^{-6}$	$3,828 \cdot 10^{-6}$	$4,417 \cdot 10^{-6}$
	Однородная	$7,017 \cdot 10^{-6}$	$5,493 \cdot 10^{-6}$	$1,978 \cdot 10^{-6}$	–
0,8	Неоднородная	$9,311 \cdot 10^{-6}$	$6,598 \cdot 10^{-6}$	$5,248 \cdot 10^{-6}$	$5,768 \cdot 10^{-6}$
	Однородная	$1,233 \cdot 10^{-5}$	$1,155 \cdot 10^{-5}$	$3,238 \cdot 10^{-5}$	$1,790 \cdot 10^{-5}$
0,7	Неоднородная	$1,082 \cdot 10^{-5}$	$6,812 \cdot 10^{-6}$	$5,593 \cdot 10^{-6}$	$6,071 \cdot 10^{-6}$
	Однородная	$1,354 \cdot 10^{-5}$	$2,734 \cdot 10^{-5}$	$3,459 \cdot 10^{-5}$	$8,304 \cdot 10^{-5}$
0,6	Неоднородная	$1,459 \cdot 10^{-5}$	$9,909 \cdot 10^{-6}$	$7,672 \cdot 10^{-6}$	$8,978 \cdot 10^{-6}$
	Однородная	$3,670 \cdot 10^{-5}$	$9,353 \cdot 10^{-5}$	$5,385 \cdot 10^{-5}$	$1,735 \cdot 10^{-4}$
0,5	Неоднородная	$2,106 \cdot 10^{-5}$	$1,508 \cdot 10^{-5}$	$1,250 \cdot 10^{-5}$	$1,369 \cdot 10^{-5}$
	Однородная	$8,617 \cdot 10^{-5}$	$1,466 \cdot 10^{-4}$	$6,444 \cdot 10^{-5}$	$1,325 \cdot 10^{-4}$
0,4	Неоднородная	$3,152 \cdot 10^{-5}$	$2,364 \cdot 10^{-5}$	$1,823 \cdot 10^{-5}$	$1,819 \cdot 10^{-5}$
	Однородная	$1,894 \cdot 10^{-4}$	$1,362 \cdot 10^{-4}$	$8,928 \cdot 10^{-5}$	$1,255 \cdot 10^{-4}$
0,3	Неоднородная	$6,042 \cdot 10^{-5}$	$5,990 \cdot 10^{-5}$	$4,229 \cdot 10^{-5}$	$4,480 \cdot 10^{-5}$
	Однородная	$2,014 \cdot 10^{-4}$	$1,548 \cdot 10^{-4}$	$1,354 \cdot 10^{-4}$	$2,823 \cdot 10^{-4}$
0,2	Неоднородная	$7,857 \cdot 10^{-5}$	$7,101 \cdot 10^{-5}$	$4,980 \cdot 10^{-5}$	$5,360 \cdot 10^{-5}$
	Однородная	$1,797 \cdot 10^{-4}$	$1,927 \cdot 10^{-4}$	$1,677 \cdot 10^{-4}$	$3,040 \cdot 10^{-4}$
0,1	Неоднородная	$1,663 \cdot 10^{-4}$	$1,520 \cdot 10^{-4}$	$1,271 \cdot 10^{-4}$	$1,266 \cdot 10^{-4}$
	Однородная	$3,096 \cdot 10^{-4}$	$3,674 \cdot 10^{-4}$	$3,014 \cdot 10^{-4}$	$4,397 \cdot 10^{-4}$
0	Однородная	$6,198 \cdot 10^{-4}$	$7,427 \cdot 10^{-4}$	$5,222 \cdot 10^{-4}$	$5,537 \cdot 10^{-4}$

В каждом из рассмотренных случаев набор данных состоял из двух групп по 196 наблюдений в каждой.

Случай 1. Ошибки наблюдений имели двустороннее экспоненциальное распределение с нулевым математическим ожиданием и дисперсией, равной 0,002830667. Для ошибок наблюдений первой группы использовалось указанное распределение с параметром формы 4; для

Таблица 2.3

Качество оценивания при $N = 1000$

Значение a	Модель	Оценка			
		Стойкая	ЭВ	ММП-ДСЭ	КММ
1	Однородная	$7,612 \cdot 10^{-7}$	$5,933 \cdot 10^{-7}$	$4,864 \cdot 10^{-7}$	$5,986 \cdot 10^{-7}$
0,9	Неоднородная	$1,007 \cdot 10^{-6}$	$7,978 \cdot 10^{-7}$	$6,567 \cdot 10^{-7}$	$7,959 \cdot 10^{-7}$
	Однородная	$1,146 \cdot 10^{-6}$	$9,874 \cdot 10^{-7}$	$5,174 \cdot 10^{-6}$	–
0,8	Неоднородная	$1,176 \cdot 10^{-6}$	$8,855 \cdot 10^{-7}$	$7,115 \cdot 10^{-7}$	$9,049 \cdot 10^{-7}$
	Однородная	$1,408 \cdot 10^{-6}$	$1,325 \cdot 10^{-6}$	$6,827 \cdot 10^{-6}$	$5,424 \cdot 10^{-5}$
0,7	Неоднородная	$1,497 \cdot 10^{-6}$	$1,148 \cdot 10^{-6}$	$9,185 \cdot 10^{-7}$	$1,141 \cdot 10^{-6}$
	Однородная	$2,344 \cdot 10^{-6}$	$3,232 \cdot 10^{-6}$	$9,965 \cdot 10^{-6}$	$1,329 \cdot 10^{-5}$
0,6	Неоднородная	$1,987 \cdot 10^{-6}$	$1,466 \cdot 10^{-6}$	$1,216 \cdot 10^{-6}$	$1,473 \cdot 10^{-6}$
	Однородная	$3,646 \cdot 10^{-6}$	$8,614 \cdot 10^{-6}$	$1,147 \cdot 10^{-5}$	$3,422 \cdot 10^{-5}$
0,5	Неоднородная	$2,665 \cdot 10^{-6}$	$1,979 \cdot 10^{-6}$	$1,649 \cdot 10^{-6}$	$2,000 \cdot 10^{-6}$
	Однородная	$1,112 \cdot 10^{-5}$	$1,482 \cdot 10^{-5}$	$1,554 \cdot 10^{-5}$	$4,058 \cdot 10^{-5}$
0,4	Неоднородная	$4,614 \cdot 10^{-6}$	$3,547 \cdot 10^{-6}$	$2,771 \cdot 10^{-6}$	$3,765 \cdot 10^{-6}$
	Однородная	$2,047 \cdot 10^{-5}$	$4,141 \cdot 10^{-5}$	$2,190 \cdot 10^{-5}$	$4,465 \cdot 10^{-5}$
0,3	Неоднородная	$8,481 \cdot 10^{-6}$	$7,601 \cdot 10^{-6}$	$5,351 \cdot 10^{-6}$	$6,462 \cdot 10^{-6}$
	Однородная	$4,892 \cdot 10^{-5}$	$3,327 \cdot 10^{-5}$	$2,756 \cdot 10^{-5}$	$6,220 \cdot 10^{-5}$
0,2	Неоднородная	$1,376 \cdot 10^{-5}$	$1,576 \cdot 10^{-5}$	$1,066 \cdot 10^{-5}$	$1,128 \cdot 10^{-5}$
	Однородная	$3,296 \cdot 10^{-5}$	$4,209 \cdot 10^{-5}$	$3,297 \cdot 10^{-5}$	$5,140 \cdot 10^{-5}$
0,1	Неоднородная	$2,065 \cdot 10^{-5}$	$1,947 \cdot 10^{-5}$	$1,722 \cdot 10^{-5}$	$1,748 \cdot 10^{-5}$
	Однородная	$4,767 \cdot 10^{-5}$	$5,796 \cdot 10^{-5}$	$4,698 \cdot 10^{-5}$	$1,249 \cdot 10^{-4}$
0	Однородная	$1,139 \cdot 10^{-4}$	$1,211 \cdot 10^{-4}$	$1,098 \cdot 10^{-4}$	$1,092 \cdot 10^{-4}$

ошибок наблюдений второй группы – с параметром формы 6. Значения меры качества оценок приведены в табл. 2.4.

При адаптивном оценивании наиболее качественными оказались ЭВ-оценки с параметром оценочных функций $\delta = 0,0005$ при учете неоднородности данных и $\delta = 0,1$ в случае, когда неоднородность не учитывалась. При неадаптивном оценивании наиболее качественными

оказались ММП-оценки, к ним близки ЭВ-оценки с параметром $\delta = 0,0005$.

Таким образом, в рассмотренной ситуации адаптивные робастные оценки оказывались качественнее робастных. Учитывать неоднородность данных нет необходимости из-за малого различия распределений подвыборок.

Таблица 2.4

Качество оценивания в случае 1

Оценка	Неоднородная модель		Однородная модель	
	Адаптивная	Неадаптивная	Адаптивная	Неадаптивная
ММП-Ст	0,0002927	0,0002913	0,0002898	0,0002884
ММП-ДСЭ	0,0001988		0,0001962	
КММ	0,0002234	0,0003470	0,0002216	0,0003442
Стойкая	0,0003214	0,0004415	0,0002978	0,0004385
ЭВ ($\delta = 0,0005$)	0,0001983	0,0002914	0,0001961	0,0002885
ЭВ ($\delta = 0,1$)	0,0001988	0,0003185	0,0001954	0,0003154
ЭВ ($\delta = 0,2$)	0,0002009	0,0003483	0,0002001	0,0003449
ЭВ ($\delta = 0,3$)	0,0002114	0,0003807	0,0002090	0,0003770
ЭВ ($\delta = 0,4$)	0,0002249	0,0004156	0,0002221	0,0004115
ЭВ ($\delta = 0,5$)	0,0002453	0,0004532	0,0002380	0,0004485
ЭВ ($\delta = 0,6$)	0,0002733	0,0004933	0,0002568	0,0004879
ЭВ ($\delta = 0,7$)	0,0002984	0,0005360	0,0002779	0,0005298
ЭВ ($\delta = 0,8$)	0,0003305	0,0005813	0,0003017	0,0005740
ЭВ ($\delta = 0,9$)	0,0003659	0,0006292	0,0003281	0,0006206
ЭВ ($\delta = 0,9995$)	0,0003995	0,0006794	0,0003549	0,0006693

Примечание. Здесь и в табл. 2.5–2.9 жирным шрифтом выделено наименьшее значение в каждом столбце

Случай 2. Ошибки наблюдений обеих групп имели распределение Стьюдента с нулевым математическим ожиданием, параметром формы 2 и единичным масштабом. Значения меры качества оценок приведены в табл. 2.5.

Таблица 2.5

Качество оценивания в случае 2

Оценка	Неоднородная модель		Однородная модель	
	Адаптивная	Неадаптивная	Адаптивная	Неадаптивная
ММП-Ст	0,1944718	0,8029257	0,1865314	1,3282991
ММП-ДСЭ	0,2232408		0,2209996	
КММ	0,2417554	0,2721566	0,9424666	0,2696956
Стойкая	0,2071381	0,1895372	0,2046515	0,1883652
ЭВ($\delta = 0,0005$)	0,2232659	0,7941070	0,2211976	1,2011029
ЭВ($\delta = 0,1$)	0,2203452	0,2702175	0,2227583	0,2722443
ЭВ($\delta = 0,2$)	0,2254759	0,2217186	0,2308472	0,2221566
ЭВ($\delta = 0,3$)	0,2412269	0,2023131	0,2368548	0,2017766
ЭВ($\delta = 0,4$)	0,2437451	0,1976364	0,2465490	0,1923207
ЭВ($\delta = 0,5$)	0,2506930	0,1908866	0,2531394	0,1887525
ЭВ($\delta = 0,6$)	0,2643011	0,1915561	0,2618118	0,1886398
ЭВ($\delta = 0,7$)	0,2773443	0,1946180	0,2711765	0,1907702
ЭВ($\delta = 0,8$)	0,2860942	0,1997334	0,2812637	0,1946398
ЭВ($\delta = 0,9$)	0,2975221	0,2069676	0,2926992	0,2000716
ЭВ($\delta = 0,9995$)	0,3087591	0,2162953	0,3073907	0,2069757

В данном случае распределение ошибок наблюдений каждой группы имеет тяжелые хвосты и не имеет дисперсии; здесь возможно появление одиночных больших ошибок, которые могут восприниматься как засорение.

При адаптивном оценивании наиболее качественными оказались ММП-Ст-оценки, как раз и ориентированные на это распределение;

среди оценок, ориентированных на двустороннее экспоненциальное распределение ошибок наблюдений, наилучшими оказались стойкие оценки, получаемые как решение максиминной задачи и поэтому ориентированные на случаи наихудшего искажения модельного распределения. Стойкие оценки оказались наиболее качественными также при неадаптивном оценивании. Адаптивные робастные ММП-Ст-оценки оказались качественнее ММП-оценок при нормальном распределении. Сравнивая адаптивные и неадаптивные ЭВ-оценки при одном значении δ , можно заметить, что при $\delta \geq 0,2$ более качественными были неадаптивные оценки. Среди стойких оценок неадаптивные также были более качественными.

Поскольку в данном случае ошибки распределений подвыборок одинаковы, разбиение выборки на две подвыборки оказывалось нецелесообразным.

Случай 3. В первой группе 90 % наблюдений содержали ошибки, имеющие двустороннее экспоненциальное распределение с нулевым математическим ожиданием, параметром формы 4; остальные 10 % имели двустороннее экспоненциальное распределение с математическим ожиданием 1, параметром формы 4. Во второй группе 90 % наблюдений содержали ошибки, имеющие двустороннее экспоненциальное распределение с нулевым математическим ожиданием, параметром формы 4,5; остальные 10 % имели двустороннее экспоненциальное распределение с математическим ожиданием 1, параметром формы 4,5. Дисперсии всех распределений были равны 0,002830667. Значения меры качества оценок приведены в табл. 2.6.

В данном случае ошибки обеих подвыборок имеют слабо различающиеся модельные двусторонние экспоненциальные распределения с легкими хвостами и асимметричными засорениями.

При адаптивном оценивании наиболее качественными оказались ЭВ-оценки с параметром $\delta = 0,3$ при учете неоднородности данных и с параметром $\delta = 0,4$ без ее учета; данные оценки позволяют учесть асимметрию засорения, что характерно для этого случая. При неадаптивном оценивании наиболее качественными оказались ЭВ-оценки с параметром $\delta = 0,2$.

В рассмотренном случае адаптивные оценки чаще оказывались качественнее неадаптивных, причем наилучшие результаты давало использование именно адаптивных оценок. Из-за близости распределе-

ний подвыборок оценки, полученные с учетом неоднородности и без него, по качеству различаются слабо.

Таблица 2.6

Качество оценивания в случае 3

Оценка	Неоднородная модель		Однородная модель	
	Адаптивная	Неадаптивная	Адаптивная	Неадаптивная
ММП-Ст	0,0444101	0,0510844	0,0446039	0,0510773
ММП-ДСЭ	0,0014493		0,0014483	
КММ	0,0028326	0,0063474	0,0112833	0,0061259
Стойкая	0,0004326	0,0005448	0,0003984	0,0005402
ЭВ($\delta = 0,0005$)	0,0014495	0,0509674	0,0014449	0,0509553
ЭВ($\delta = 0,1$)	0,0010206	0,0005891	0,0010353	0,0064802
ЭВ($\delta = 0,2$)	0,0004935	0,0004362	0,0006632	0,0004322
ЭВ($\delta = 0,3$)	0,0003014	0,0004733	0,0004216	0,0004687
ЭВ($\delta = 0,4$)	0,0003240	0,0005133	0,0003092	0,0005079
ЭВ($\delta = 0,5$)	0,0003518	0,0005562	0,0003318	0,0005499
ЭВ($\delta = 0,6$)	0,0003840	0,0006021	0,0003574	0,0005946
ЭВ($\delta = 0,7$)	0,0004362	0,0006510	0,0003903	0,0006420
ЭВ($\delta = 0,8$)	0,0004880	0,0007031	0,0004309	0,0006920
ЭВ($\delta = 0,9$)	0,0005550	0,0007581	0,0004696	0,0007445
ЭВ($\delta = 0,9995$)	0,0006984	0,0008160	0,0005151	0,0007991

Случай 4. В первой группе 90 % наблюдений содержали ошибки, имеющие двустороннее экспоненциальное распределение с нулевым математическим ожиданием, параметром формы 1,5; остальные 10 % имели двустороннее экспоненциальное распределение с математическим ожиданием 1, параметром формы 4. Во второй группе 90 % наблюдений содержали ошибки, имеющие двустороннее экспоненциальное распределение с нулевым математическим ожиданием, параметром формы 1,6; остальные 10 % имели двустороннее экспоненциальное

ное распределение с математическим ожиданием 1, параметром формы 4,5. Дисперсии всех распределений были равны 0,002830667. Значения меры качества оценок приведены в табл. 2.7.

Таблица 2.7

Качество оценивания в случае 4

Оценка	Неоднородная модель		Однородная модель	
	Адаптивная	Неадаптивная	Адаптивная	Неадаптивная
ММП-Ст	0,0442586	0,0511594	0,0447128	0,0512023
ММП-ДСЭ	0,0007796		0,0007788	
КММ	0,0008262	0,0063629	0,0008997	0,0061455
Стойкая	0,0004129	0,0003651	0,0003944	0,0003646
ЭВ($\delta = 0,0005$)	0,0007776	0,0510425	0,0007768	0,0510803
ЭВ($\delta = 0,1$)	0,0044185	0,0005197	0,0007346	0,0005679
ЭВ($\delta = 0,2$)	0,0006708	0,0003359	0,0008951	0,0003377
ЭВ($\delta = 0,3$)	0,0003850	0,0003411	0,0015091	0,0003419
ЭВ($\delta = 0,4$)	0,0004056	0,0003504	0,0003949	0,0003503
ЭВ($\delta = 0,5$)	0,0004409	0,0003629	0,0004268	0,0003620
ЭВ($\delta = 0,6$)	0,0004741	0,0003780	0,0004641	0,0003761
ЭВ($\delta = 0,7$)	0,0005247	0,0003953	0,0005016	0,0003924
ЭВ($\delta = 0,8$)	0,0005700	0,0004146	0,0005407	0,0004102
ЭВ($\delta = 0,9$)	0,0007199	0,0004357	0,0005787	0,0004295
ЭВ($\delta = 0,9995$)	0,0008454	0,0004589	0,0006225	0,0004501

В данном случае ошибки обеих подвыборок имеют слабо различающиеся двусторонние экспоненциальные распределения с тяжелыми хвостами и асимметричными засорениями.

При адаптивном оценивании наиболее качественными оказались ЭВ-оценки с параметром $\delta = 0,3$ при учете неоднородности данных и стойкие – без ее учета. При неадаптивном оценивании наиболее качественными оказались ЭВ-оценки с параметром $\delta = 0,2$.

В рассмотренном случае при использовании адаптивных оценок часто идентифицировалось распределение, близкое к лапласовскому; по-видимому, в связи с этим неадаптивные оценки оказывались качественнее адаптивных. Неоднородность данных учитывать нет необходимости из-за малого различия распределений подвыборок.

Случай 5. В первой группе наблюдения содержали ошибки, имеющие двустороннее экспоненциальное распределение с нулевым математическим ожиданием, параметром формы 1,5 и дисперсией 0,184622027; во второй группе наблюдения содержали ошибки, имеющие двустороннее экспоненциальное распределение с нулевым математическим ожиданием, параметром формы 6 и дисперсией 1,273699768. Значения меры качества оценок приведены в табл. 2.8.

Таблица 2.8

Качество оценивания в случае 5

Оценка	Неоднородная модель		Однородная модель	
	Адаптивная	Неадаптивная	Адаптивная	Неадаптивная
ММП-Ст	0,0422086	0,0422863	0,0802122	0,0801849
ММП-ДСЭ	0,0354133		0,0725412	
КММ	0,0383405	0,0431395	0,0885538	0,0869074
Стойкая	0,0434346	0,0480957	0,0807817	0,1056355
ЭВ($\delta = 0,0005$)	0,0354120	0,0422814	0,0725564	0,0801186
ЭВ($\delta = 0,1$)	0,0359559	0,0420684	0,0790449	0,0810781
ЭВ($\delta = 0,2$)	0,0375647	0,0430060	0,0895909	0,0830861
ЭВ($\delta = 0,3$)	0,0401043	0,0446439	0,1009002	0,0865178
ЭВ($\delta = 0,4$)	0,0433185	0,0468148	0,1125321	0,0919349
ЭВ($\delta = 0,5$)	0,0473970	0,0494891	0,1199301	0,1001816
ЭВ($\delta = 0,6$)	0,0516206	0,0527011	0,1270128	0,1121909
ЭВ($\delta = 0,7$)	0,0564727	0,0565636	0,1295449	0,1281803
ЭВ($\delta = 0,8$)	0,0616128	0,0613909	0,1318020	0,1478965
ЭВ($\delta = 0,9$)	0,0663414	0,0671958	0,1336704	0,1679945
ЭВ($\delta = 0,9995$)	0,0718500	0,0737068	0,1339442	0,1861538

В этом случае первая подвыборка имеет распределение с тяжелыми хвостами, вторая – с легкими хвостами и в четыре раза большим параметром масштаба.

При адаптивном оценивании с учетом неоднородности данных наиболее качественными оказались ЭВ-оценки с параметром оценочных функций $\delta = 0,0005$. В случае, когда неоднородность данных не учитывалась, наиболее качественными были ММП-ДСЭ-оценки. При неадаптивном оценивании наиболее качественными оказались ЭВ-оценки с параметром $\delta = 0,1$ при учете неоднородности; с параметром $\delta = 0,0005$ – без ее учета.

Поскольку распределения ошибок в подвыборках сильно различаются, учет неоднородности данных оказался необходим. При учете неоднородности данных в большинстве ситуаций адаптивные оценки были качественнее неадаптивных.

Случай 6. В первой группе наблюдения содержали ошибки, имеющие двустороннее экспоненциальное распределение с нулевым математическим ожиданием, параметром формы 6 и дисперсией 0,019901558; во второй группе наблюдения содержали ошибки, имеющие двустороннее экспоненциальное распределение с нулевым математическим ожиданием, параметром формы 1,5 и дисперсией 0,738488111. Обе подвыборки содержали 5 % ошибок в виде модуля случайной величины, имеющей двустороннее экспоненциальное распределение с нулевым математическим ожиданием, параметром формы 3 и дисперсией 0,373282173. Значения меры качества оценок приведены в табл. 2.9.

В данном случае основная часть первой подвыборки имеет распределение с легкими хвостами, основная часть второй подвыборки – с тяжелыми хвостами и в четыре раза большим масштабом, обе подвыборки имеют асимметричное засорение.

При адаптивном оценивании с учетом неоднородности данных наиболее качественными оказались ЭВ-оценки с параметром оценочных функций $\delta = 0,5$; без учета неоднородности – стойкие оценки.

При неадаптивном оценивании с учетом неоднородности наиболее качественными оказались стойкие оценки. В случае, когда неоднородность данных не учитывалась, наилучшими были ЭВ-оценки с параметром $\delta = 0,2$.

В целом адаптивные оценки оказывались качественнее неадаптивных (за исключением КММ-оценок). Неоднородность данных требуется учитывать, так как свойства основных частей выборок сильно различаются.

Таблица 2.9

Качество оценивания в случае 6

Оценка	Неоднородная модель		Однородная модель	
	Адаптивная	Неадаптивная	Адаптивная	Неадаптивная
ММП-Ст	0,0410743	0,0555421	0,0411186	0,0427937
ММП-ДСЭ	0,0328788		0,0299777	
КММ	0,0270322	0,0265953	0,0536710	0,0351349
Стойкая	0,0129843	0,0176166	0,0294485	0,0378852
ЭВ($\delta = 0,0005$)	0,0328594	0,0553878	0,0299950	0,0427438
ЭВ($\delta = 0,1$)	0,0283145	0,0301155	0,0316594	0,0364543
ЭВ($\delta = 0,2$)	0,0184200	0,0212909	0,0333199	0,0347572
ЭВ($\delta = 0,3$)	0,0131144	0,0187306	0,0342432	0,0356810
ЭВ($\delta = 0,4$)	0,0115279	0,0180458	0,0357474	0,0374209
ЭВ($\delta = 0,5$)	0,0112193	0,0181294	0,0369012	0,0384316
ЭВ($\delta = 0,6$)	0,0114812	0,0186227	0,0376734	0,0377279
ЭВ($\delta = 0,7$)	0,0119443	0,0193863	0,0351662	0,0367498
ЭВ($\delta = 0,8$)	0,0124530	0,0203584	0,0341006	0,0363256
ЭВ($\delta = 0,9$)	0,0132151	0,0215067	0,0351542	0,0362248
ЭВ($\delta = 0,9995$)	0,0142030	0,0228030	0,0332967	0,0363039

На основании проведенного исследования можно сделать следующие выводы:

- использование адаптивных робастных оценок по сравнению с робастными оценками оказывается более предпочтительным, если

распределение ошибок хотя бы одной группы наблюдений имеет легкие хвосты;

- учет разделения набора данных на неоднородные группы может улучшить качество получаемых оценок, особенно в том случае, если эти группы наблюдений будут иметь распределения ошибок с существенно различающимися характеристиками (причем не только параметра масштаба, но и параметра формы);

- в ситуациях, когда набор данных имеет распределение ошибок с легкими хвостами либо может быть разделен на группы, в которых ошибки хорошо описываются двусторонними экспоненциальными распределениями, более предпочтительно для практики использование ММП-ДСЭ-оценок;

- в случаях, когда ошибки наблюдений могут иметь распределение в виде смеси, в том числе иметь засорения, как симметричные, так и асимметричные, более эффективно на практике использование более грубых оценок, чем ММП-ДСЭ-оценки, таких как ЭВ- и стойкие оценки;

- процедура вычисления КММ-оценок может быть неустойчивой в условиях нарушения предположений, при которых получены оценки, часто данные оценки имели низкое качество и поэтому пользоваться ими не рекомендуется;

- чем сильнее искажено модельное распределение, тем бóльшим должен быть параметр δ ЭВ-оценки; при недостатке априорной информации об ошибках, при сильных искажениях предпочтительнее может оказаться использование стойкой оценки, не имеющей, в отличие от ЭВ-оценки, дополнительного параметра оценочной функции.

Третья часть исследований посвящена исследованию подходов к оцениванию гетероскедастичной регрессии.

Использовались 20 выборок по 200 наблюдений. Распределение наблюдений было смесью двусторонних экспоненциальных распределений с параметром сдвига, линейно зависящим от скалярной входной переменной:

$$\mu_i = \theta_1 + \theta_2 x_i,$$

где $\theta_1 = 20$, $\theta_2 = 10$, $x_i \in [-1, 1]$, $i = 1, \dots, 200$, и параметром формы $\nu = 4$.

Основное распределение выборки имело параметр масштаба, зависящий от той же входной переменной:

$$\lambda_i = \exp(\Lambda_1 + \Lambda_2 x_i),$$

где $\Lambda_1 = 1$, $\Lambda_2 = 1,5$. Засоряющее распределение выборки имело константный параметр масштаба, равный $\exp(2,75)$. Вероятности компонентов смеси равнялись 0,9 и 0,1 для основного и засоряющего распределений соответственно.

Использовались обобщенные радикальные оценки с параметром $\delta = 0,5$. При оценивании значение параметра формы предполагалось лежащим в интервале 1,0000001...20, значения параметра масштаба – в интервале $\exp(-3)$... $\exp(3)$.

Кроме того, по тем же выборкам оценивались параметры гомоскедастичной регрессии при модели параметра масштаба вида $\lambda_i = \Lambda_1$.

В табл. 2.10 приведены усредненные значения оценок параметров моделей, полученных с учетом гетероскедастичности и без ее учета. При регрессионном моделировании в первую очередь интересуются, как правило, моделью параметра сдвига, поэтому в табл. 2.10 приведено также среднее значение показателя (2.37).

Таблица 2.10

Результаты оценивания гетероскедастичной регрессии

Показатель качества оценивания	Регрессия	
	гетероскедастичная	гомоскедастичная
$\hat{\theta}_1$	20,01642765	20,21404845
$\hat{\theta}_2$	9,9838866	10,2113881
$\hat{\Lambda}_1$	0,9860541	1,681125
$\hat{\Lambda}_2$	1,535854	–
\hat{v}	4,6407913	1,0000001
τ	0,080368904	0,578177481

Результаты говорят о высоком качестве оценок, полученных с учетом гетероскедастичности, и об их преимуществе перед оценками, полученными без учета гетероскедастичности. Фактически, в гомоскедастичной модели не реализовалось свойство адаптивности оценок, поскольку оценка параметра формы всегда была равна значению левой границы допустимой области. Причиной этого является настолько сильное различие значений параметра масштаба в разных точках факторного пространства, что учесть его лишь путем робастного оценивания гомоскедастичной регрессии не удастся.

ГЛАВА 3

ОЦЕНИВАНИЕ ПРИ МУЛЬТИПЛИКАТИВНОЙ КОВАРИАЦИОННОЙ СТРУКТУРЕ ОШИБОК

Глава посвящена методам оценивания неоднородных регрессионных моделей в случае, когда ошибки групп наблюдений являются зависимыми и имеют многомерное эллиптическое распределение с псевдоковариационной матрицей, параметризованной в виде кронекера произведения нескольких положительно определенных матриц [22, 23, 57, 60, 61, 65]. Отдельно рассматривается ситуация, когда наряду с количественными откликами имеются отклики качественной природы, а наблюдения могут содержать пропуски.

3.1. ОЦЕНИВАНИЕ ПАРАМЕТРОВ РЕГРЕССИОННОЙ МОДЕЛИ С ЭЛЛИПТИЧЕСКИМ РАСПРЕДЕЛЕНИЕМ ОШИБОК

3.1.1. МОДЕЛЬ

Пусть вектор, составленный из наблюдений за одним или несколькими откликами системы, можно разделить на p подвекторов (групп наблюдений) так, что ошибки наблюдений в одной группе являются коррелированными, а ошибки наблюдений из разных групп – независимыми.

Рассмотрим линейную по параметрам многофакторную регрессионную модель вида (2.6) с ограничениями на параметры вида (1.9).

Обозначим E_i вектор ошибок наблюдений i -й группы и предположим, что он имеет эллиптическое распределение с плотностью вида

$$g(E_i, V_i, n_i, v_{\gamma(i)}) = C(n_i, v_{\gamma(i)}) |V_i|^{-1/2} \tilde{g}(q_i, n_i, v_{\gamma(i)}),$$

где V_i – симметричная положительно определенная псевдоковариационная матрица; n_i – размерность вектора E_i ; $v_{\gamma(i)}$ – параметр формы распределения, при этом функция $\gamma(i)$ ставит в соответствие номеру группы номер параметра формы и всего имеется n_v различных векторных параметров формы v_j , $j=1, \dots, n_v$ (см. подобную модель в [177]); $C(n_i, v_{\gamma(i)})$ – нормирующая константа; $q_i = E_i^T V_i^{-1} E_i$; $\tilde{g}(q_i, n_i, v_{\gamma(i)})$ – функция, обладающая свойствами плотности распределения (за исключением условия нормировки).

Ковариационная матрица вектора E_i , если она существует, отличается от псевдоковариационной матрицы дополнительным скалярным множителем

$$\Delta(n_i, v_{\gamma(i)}) = \frac{\pi^{n_i/2}}{\Gamma(n_i/2 + 1)} C(n_i, v_{\gamma(i)}) \int_0^\infty z^{n_i+1} \tilde{g}(z^2, n_i, v_{\gamma(i)}) dz.$$

В табл. 3.1 приведены сведения о некоторых распределениях эллиптического типа, при этом n обозначает размерность случайного вектора, $v > 0$ – скалярный параметр формы.

Параметризуем псевдоковариационные матрицы векторов E_i посредством симметричных положительно определенных матриц U_j , $j=1, \dots, n_u$, размерности $k_j \geq 1$. Для псевдоковариационной матрицы i -й группы введем мультипликативную параметризацию в виде кронекера произведения параметров:

$$V_i = \bigotimes_{j \in J(V_i)} U_j, \quad i=1, \dots, p, \quad (3.1)$$

где $J(V_i)$ – упорядоченная последовательность индексов матриц U_j , посредством которых параметризована матрица V_i . При этом необходимо выполнение соотношения

$$n_i = \prod_{j \in J(V_i)} k_j.$$

Таблица 3.1

Многомерные распределения эллиптического типа

Распределение	$\tilde{g}(q, n, \nu)$	$C(n, \nu)$	$\Delta(n, \nu)$
Нормальное	$\exp -q/2$	$(2\pi)^{-n/2}$	1
Равномерное	1, при $q < 1$	$\frac{n\Gamma n/2}{2\pi^{n/2}}$	$1/(n+2)$
Стьюдента	$\nu + q^{-(n+\nu)/2}$	$\frac{\nu^{v/2}\Gamma(\nu+n)/2}{\pi^{n/2}\Gamma \nu/2}$	$\frac{\nu}{\nu-2}, \nu > 2$
Двустороннее экспоненциальное	$\exp -q^{v/2}$	$\frac{\nu\Gamma n/2}{2\pi^{n/2}\Gamma n/\nu}$	$\frac{\Gamma(n+2)/\nu}{n\Gamma(n/\nu)}$
Обобщенное Симпсона	$1 - q^{v/2}$, при $q < 1$	$\frac{n(n+\nu)\Gamma n/2}{2\pi^{n/2}\nu}$	$\frac{n+\nu}{(n+2)(n+\nu+2)}$

Обозначим m_i – количество элементов в последовательности $J(V_i)$, V_{ij} – j -й матричный параметр-сомножитель в матрице V_i , n_{ij} – размерность матрицы V_{ij} .

Поскольку ковариационная матрица вектора E_i отличается от псевдоковариационной матрицы только скалярным множителем $\Delta n_i, \nu_{\gamma(i)}$, она также мультипликативно параметризована.

Введенная параметризация псевдоковариационных матриц может быть использована при структурированности вектора наблюдений. В частности, данная параметризация соответствует такой структурированности вектора наблюдений, которая может быть отражена введением мультииндекса наблюдений $(s_{i1}, s_{i2}, \dots, s_{im_i})$, где $s_{ij} = 1, \dots, n_{ij}$. При этом элементы вектора E_i упорядочены посредством мультииндекса таким образом, что быстрее изменяется индекс, стоящий правее. Тогда ковариация двух элементов вектора E_i имеет вид

$$\begin{aligned} & \text{cov } E_{(\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{im_i})}, E_{(\beta_{i1}, \beta_{i2}, \dots, \beta_{im_i})} = \\ & = \Delta(n_i, v_{\gamma(i)}) V_{i1} \alpha_{i1}, \beta_{i1} V_{i2} \alpha_{i2}, \beta_{i2} \dots V_{im_i} \alpha_{im_i}, \beta_{im_i}. \end{aligned}$$

В результате каждый матричный параметр псевдоковариационной матрицы является с точностью до скалярного множителя кросс-ковариационной матрицей некоторых подвекторов вектора E_i . Таким образом, наложенная на ковариационную матрицу структура отражает наличие определенной стабильной кросс-ковариационной связи между отдельными подвекторами вектора ошибок.

На практике вид ковариационной структуры ошибок определяется спецификой конкретной прикладной задачи. Приведем примеры спецификации структуры ковариационных матриц (см. также [108, 166, 175]).

Для многооткликовой модели в условиях повторных измерений при $p = 1$ может быть введена следующая параметризация:

$$V = V_1 = \delta \Sigma \otimes \Omega \otimes \Lambda, \quad (3.2)$$

где δ – положительная скалярная величина; Σ , Ω и Λ – положительно определенные симметричные матрицы размерности $p_1 \times p_1$, $p_2 \times p_2$ и $p_3 \times p_3$ соответственно; p_1 – количество откликов; p_2 – количество повторных наблюдений; p_3 – количество точек факторного пространства. В получаемом при этом мультииндексе наблюдений $(s_{11}, s_{12}, s_{13}, s_{14})$, где $s_{11} = 1$, $s_{12} = 1, \dots, p_1$, $s_{13} = 1, \dots, p_2$, $s_{14} = 1, \dots, p_3$, второй элемент есть номер отклика, третий – номер повторного наблюдения, четвертый – номер точки факторного пространства.

Матричный параметр Σ с точностью до скалярного множителя интерпретируется как кросс-ковариационная матрица векторов откликов, Ω – как кросс-ковариационная матрица векторов повторных наблюдений, Λ – как кросс-ковариационная матрица векторов наблюдений в различных точках факторного пространства. Скаляр δ может быть проинтерпретирован при единичных матрицах Σ , Ω , Λ как величина дисперсии элементов вектора ошибок.

Модифицируем данный пример. Предположим, что в каждой точке факторного пространства проведено различное число повторных измерений. Кроме того, наблюдения в различных точках факторного пространства будем считать независимыми. Соответствующая параметризация псевдоковариационных матриц имеет вид

$$V_i = \delta \Sigma \otimes \Omega_i, \quad i = 1, \dots, p_3,$$

где Ω_i – положительно определенная симметричная матрица размерности $p_{2i} \times p_{2i}$; p_{2i} – число повторов в i -й точке факторного пространства.

Данную ковариационную структуру получаем из общей структуры (3.1) при $p = p_3$, $m_i = 3$, $n_{i1} = 1$, $n_{i2} = p_1$, $n_{i3} = p_{2i}$. В мультииндексе наблюдений (s_{i1}, s_{i2}, s_{i3}) , где $s_{i1} = 1$, $s_{i2} = 1, \dots, p_1$, $s_{i3} = 1, \dots, p_{2i}$, $i = 1, \dots, p_3$, второй элемент есть номер отклика, третий – номер повторного наблюдения в i -й точке факторного пространства.

Матричный параметр Σ с точностью до скалярного множителя интерпретируется как кросс-ковариационная матрица векторов откликов, измеренных в одной точке факторного пространства; Ω_i – как кросс-ковариационная матрица векторов повторных наблюдений в i -й точке факторного пространства.

В случае если данные не содержат повторных наблюдений, положив $p_{2i} = 1$, получим параметризацию

$$V_i = \delta \omega_i \Sigma,$$

где ω_i – скалярная величина, обеспечивающая различие ковариационных матриц откликов в различных точках факторного пространства.

Упрощая далее, предположим, что ковариационные матрицы откликов в различных точках факторного пространства равны и $\delta=1$; получим многооткликовую регрессионную модель, например, (1.19) с (псевдо)ковариационной матрицей вида

$$V_i = \Sigma.$$

3.1.2. ОЦЕНИВАНИЕ ПАРАМЕТРОВ

Введенная модель имеет следующие неизвестные параметры, которые требуют оценивания по данным: вектор параметров регрессии θ , матричные и скалярные параметры U_j , $j=1, \dots, n_u$, векторы параметров формы распределения v_j , $j=1, \dots, n_v$.

Для оценивания всех указанных параметров воспользуемся ММП. Минимизируемая функция при этом имеет вид

$$Q = - \sum_{j=1}^{n_v} \sum_{i \in J(v_j)} \ln \left\{ C_{n_i, v_j} \left| \bigotimes_{k \in J(V_i)} U_k \right|^{-1/2} \tilde{g}(q_i, n_i, v_j) \right\},$$

где $J(v_j)$ – множество номеров групп, для распределения которых v_j является параметром формы.

Поиск условного минимума будем производить методом неопределенных множителей Лагранжа. Соответствующая система оценочных уравнений включает в себя:

1) получаемую рассуждениями, аналогичными использованным в п. 1.1, систему (1.15), в которой

$$H = \sum_{i=1}^P w(\hat{q}_i, n_i, \hat{v}_{\gamma(i)}) X_i^T \hat{V}_i^{-1} X_i, \quad \hat{\theta} = H^{-1} \sum_{i=1}^P w(\hat{q}_i, n_i, \hat{v}_{\gamma(i)}) X_i^T \hat{V}_i^{-1} Y_i,$$

где $\hat{q}_i = \hat{E}_i^T \hat{V}_i^{-1} \hat{E}_i$; $\hat{E}_i = Y_i - X_i \hat{\theta}_R$ – вектор остатков i -й группы; матрица X_i и вектор Y_i аналогичны использованным в п. 2.3.1 (с точно-

стью до размерностей групп, различных в рассматриваемом случае); весовая функция имеет вид

$$w(q_i, n_i, v_{\gamma(i)}) = -\frac{2}{\tilde{g}(q_i, n_i, v_{\gamma(i)})} \frac{\partial \tilde{g}(q_i, n_i, v_{\gamma(i)})}{\partial q_i};$$

2) для каждого $j=1, \dots, n_u$ матричное уравнение

$$\hat{U}_j = \frac{k_j}{N_j} \sum_{i \in J(U_j)} w(\hat{q}_i, n_i, \hat{v}_{\gamma(i)}) \hat{E}_{i[j]}^T \hat{V}_{i[j]}^{-1} \hat{E}_{i[j]}, \quad (3.3)$$

где $N_j = \sum_{i \in J(U_j)} n_i$; $J(U_j)$ – множество номеров групп наблюдений, в

число параметров псевдоковариационных матриц которых входит матрица U_j ; $\hat{E}_{i[j]}$ – матрица размерности $[n_i/k_j] \times k_j$, t -й столбец которой содержит в исходном порядке элементы вектора остатков \hat{E}_i , для которых соответствующий матрице U_j элемент мультииндекса равен t ; $\hat{V}_{i[j]}$ – матрица \hat{V}_i с исключенным сомножителем \hat{U}_j ;

3) для каждого $j=1, 2, \dots, n_v$ векторное уравнение

$$\sum_{i \in J(v_j)} \left\{ \frac{1}{C(n_i, \hat{v}_j)} \frac{dC(n_i, \hat{v}_j)}{d\hat{v}_j} + \frac{1}{\tilde{g}(\hat{q}_i, n_i, \hat{v}_j)} \frac{\partial \tilde{g}(\hat{q}_i, n_i, \hat{v}_j)}{\partial \hat{v}_j} \right\} = 0. \quad (3.4)$$

Проанализируем полученную систему уравнений правдоподобия. Перепишем матричные уравнения (3.3) в виде

$$\sum_{i \in J(U_j)} w(\hat{q}_i, n_i, \hat{v}_{\gamma(i)}) \hat{U}_j^{-1} \hat{E}_{i[j]}^T \hat{V}_{i[j]}^{-1} \hat{E}_{i[j]} = \frac{N_j}{k_j} I_{k_j}, \quad j=1, \dots, n_u,$$

и применим к его обеим частям операцию взятия следа матрицы; получим уравнение

$$\sum_{i \in J(U_j)} w(\hat{q}_i, n_i, \hat{v}_{\gamma(i)}) \text{tr} \hat{U}_j^{-1} \hat{E}_{i[j]}^T \hat{V}_{i[j]}^{-1} \hat{E}_{i[j]} = N_j, \quad j=1, \dots, n_u.$$

Стоящие в левой части этого равенства величины $\text{tr } \hat{U}_j^{-1} \hat{E}_{i[j]}^T \hat{V}_{i[j]}^{-1} \hat{E}_{i[j]}$ равны \hat{q}_i , в чем можно убедиться, расписав поэлементно матричные произведения и операцию взятия следа. В результате получим систему

$$\sum_{i \in J(U_j)} w(\hat{q}_i, n_i, \hat{v}_{\gamma(i)}) \hat{q}_i = N_j, \quad j = 1, \dots, n_u,$$

или

$$\sum_{i \in J(U_j)} \left[w(\hat{q}_i, n_i, \hat{v}_{\gamma(i)}) \hat{q}_i - n_i \right] = 0, \quad j = 1, \dots, n_u. \quad (3.5)$$

Таким образом, ММП-оценки параметров должны удовлетворять системе ограничений (3.5).

Обозначим $z_i = w(\hat{q}_i, n_i, \hat{v}_{\gamma(i)}) \hat{q}_i - n_i$, тогда данная система уравнений относительно новых переменных z_i является однородной системой линейных уравнений с матрицей системы A , элемент A_{ji} которой равен единице, если параметр U_j присутствует среди сомножителей матрицы V_i , и нулю – в противном случае.

Если ранг матрицы системы равен p , то система имеет единственное тривиальное решение. Отсюда для всех $i = 1, \dots, p$ справедливо равенство

$$w(\hat{q}_i, n_i, \hat{v}_{\gamma(i)}) \hat{q}_i - n_i = 0. \quad (3.6)$$

Разрешим его относительно неизвестного \hat{q}_i :

$$\hat{q}_i = \hat{q}(n_i, \hat{v}_{\gamma(i)}). \quad (3.7)$$

Подставим (3.7) в уравнение (3.4), получим

$$\sum_{i \in J(v_j)} \left\{ \frac{1}{C(n_i, \hat{v}_j)} \frac{dC(n_i, \hat{v}_j)}{d\hat{v}_j} + \left(\frac{1}{g(q_i, n_i, \hat{v}_j)} \frac{\partial g(q_i, n_i, \hat{v}_j)}{\partial \hat{v}_j} \right) \right\}_{q_i = \hat{q}(n_i, \hat{v}_j)} = 0. \quad (3.8)$$

Уравнение (3.8) не зависит от данных, поэтому величины \hat{v}_j могут быть вычислены априорно – $\hat{v}_j = \hat{v}_j(n_1, n_2, \dots, n_p)$ – и названы априорно идентифицируемыми. В связи с этим может быть удобно вводить для каждой группы свой параметр формы v_i , $i = 1, \dots, p$, и определять априорное значение его оценки с использованием уравнения

$$\frac{1}{C(n_i, \hat{v}_i)} \frac{dC(n_i, \hat{v}_i)}{d\hat{v}_i} + \left(\frac{1}{g(q_i, n_i, \hat{v}_i)} \frac{\partial g(q_i, n_i, \hat{v}_i)}{\partial \hat{v}_i} \right) \Bigg|_{q_i = \hat{q}(n_i, \hat{v}_i)} = 0. \quad (3.9)$$

Подставив априорные значения величин \hat{v}_j в уравнение (3.7), получим априорные значения $\hat{q}_i = \hat{q}_i(n_1, n_2, \dots, n_p)$; в результате можно говорить об априорной идентифицируемости квадратичных форм \hat{q}_i .

Подставим априорные значения величин \hat{q}_i , \hat{v}_j в весовую функцию $w(\hat{q}_i, n_i, \hat{v}_{\gamma(i)})$, тогда значения весовой функции также могут быть вычислены априорно: $w_i = w_i(n_1, n_2, \dots, n_p)$.

Заметим, что для возможности проведения указанных преобразований решения уравнений (3.7) – (3.9) должны существовать.

Полученный результат позволяет свести процедуру оценивания параметров в рассматриваемом случае к ММП-оцениванию в условиях нормального распределения ошибок с ковариационной матрицей i -й группы, равной $\frac{1}{w_i} V_i$.

Более того, ММП-оценки вектора параметров θ исходной модели можно получить в рамках модели нормального распределения ошибок с ковариационной матрицей i -й группы \tilde{V}_i , параметризованной как

$$\tilde{V}_i = \bigotimes_{j \in J(V_i)} \tilde{U}_j,$$

где \tilde{U}_j – матрицы тех же размерностей, что U_j , причем справедливо равенство

$$\tilde{V}_i = \frac{1}{w_i} V_i. \quad (3.10)$$

Существование такой модели обосновывается следующей теоремой.

Теорема 3.1. При выполнении условия

$$\tilde{U}_j = U_j \prod_{k=1}^p w_k^{-A_{kj}^-}, \quad (3.11)$$

где $p \times n_u$ -матрица A^- ранга p является левой слабо обобщенной обратной к матрице A [32], выполняется (3.10).

Доказательство. Покажем, что при выполнении (3.11) справедливо (3.10):

$$\begin{aligned} \tilde{V}_i &= \bigotimes_{j \in J(V_i)} \tilde{U}_j = \bigotimes_{j \in J(V_i)} \left\{ U_j \prod_{k=1}^p w_k^{-A_{kj}^-} \right\} = \\ &= \left\{ \prod_{j \in J(V_i)} \prod_{k=1}^p w_k^{-A_{kj}^-} \right\} \bigotimes_{j \in J(V_i)} U_j = \left\{ \prod_{j=1}^{n_u} \prod_{k=1}^p w_k^{-A_{kj}^- A_{ji}} \right\} V_i = \\ &= \left\{ \prod_{k=1}^p \left(w_k \right)^{-\sum_{j=1}^{n_u} A_{kj}^- A_{ji}} \right\} V_i = \left\{ \prod_{k=1}^p w_k^{-I_{ki}} \right\} V_i = \frac{1}{w_i} V_i, \end{aligned}$$

где предпоследнее равенство имеет место в силу свойства левой слабо обобщенной обратной матрицы $A^- A = I_p$.

Заметим, что в общем случае при $n_u > p$ матрица A^- определяется неоднозначно, что влечет неоднозначность преобразования (3.11).

Перейдем к случаю, когда ранг матрицы системы (3.5) меньше p . Систему (3.5) можно (после перенумерации групп и матричных параметров, если это необходимо) заменить эквивалентной с матрицей системы, представленной в форме

$$\begin{bmatrix} I_{r_a} & 0 & 0 \\ 0 & I_{r_c} & F \\ 0 & 0 & 0 \end{bmatrix}.$$

В результате среди неизвестных системы имеется $p - r_a - r_c$ свободных переменных, r_a переменных, имеющих нулевые значения, и r_c переменных, которые можно представить в виде линейных комбинаций свободных переменных.

Поскольку первые r_a компонент вектора неизвестных имеют нулевое значение, для соответствующих групп наблюдений справедливо представление (3.6). Для данных групп может быть удобно ввести априорно идентифицируемые параметры формы, тогда их априорные значения подставляем в функцию правдоподобия, упрощая тем самым процесс оценивания.

Компоненты вектора неизвестных с номерами $r_a + 1 \dots r_a + r_c$ можно представить в виде линейных комбинаций свободных переменных

$$z_i = - \sum_{j=r_a+r_c+1}^p F_{i-r_a, j-r_a-r_c} z_j,$$

следовательно, справедливо ограничение на оцениваемые параметры вида

$$w(\hat{q}_i, n_i, \hat{v}_{\gamma(i)}) \hat{q}_i - n_i = -\beta_i, \quad (3.12)$$

где $\beta_i = \sum_{j=r_a+r_c+1}^p F_{i-r_a, j-r_a-r_c} w(\hat{q}_j, n_j, \hat{v}_{\gamma(j)}) \hat{q}_j - n_j$.

В результате проведенного анализа можно выделить случай полной априорной идентифицируемости, когда все величины \hat{q}_i , w_i и \hat{v}_j априорно идентифицируемы ($r_a = p$); случай частичной априорной идентифицируемости, когда только часть указанных величин априорно идентифицируема ($0 < r_a < p$), и случай априорной неидентифицируемости, когда среди указанных величин нет априорно идентифицируемых ($r_a = 0$).

Особенностью случая полной априорной идентифицируемости является необходимость выполнения соотношения $n_u \geq p$. В результате при увеличении числа групп для сохранения свойства полной априорной идентифицируемости должно увеличиваться количество ковариационных параметров. В этих условиях ММП может уже не обладать положительными асимптотическими свойствами [42, 49].

Решение системы оценочных уравнений в общем случае можно осуществлять методом покомпонентного оценивания. Удобно выделить подмножества оцениваемых параметров

$$\theta, U_1, \dots, U_{n_u}, v_1, \dots, v_{n_v}.$$

На этапе уточнения оценки параметра θ можно пользоваться ИМНК с весовой функцией $w(q_i, n_i, v_{\gamma(i)})$. При поиске оценок параметров U_1, \dots, U_{n_u} применяется аналогичный подход, который сводится к итеративному пересчету на соответствующем этапе оценок \hat{U}_j (по формуле (3.3)) с использованием в весовой функции $w(\hat{q}_i, n_i, \hat{v}_{\gamma(i)})$ оценок ее аргументов, вычисленных на предыдущей итерации. Для вычисления значений параметров формы на очередной итерации следует воспользоваться каким-либо численным методом поиска решения системы уравнений.

При данном подходе требуется положительная определенность приближений оценок матричных параметров ковариационной матрицы, вычисляемых на каждой итерации.

Теорема 3.2. Необходимыми условиями положительной определенности очередного приближения оценки U_j , $j=1, \dots, n_u$, полученного на любой итерации алгоритма покомпонентного оценивания, являются

$$\sum_{i \in J(U_j)} \min k_{j, n_i} / k_j \geq k_j, \quad j=1, \dots, n_u.$$

Доказательство следует из вида уравнения (3.3) и свойств ранга матрицы.

Например, для ковариационной структуры (3.2) условия имеют вид

$$P_1 \leq P_2 P_3, \quad P_2 \leq P_1 P_3, \quad P_3 \leq P_1 P_2.$$

Рассмотрим ситуацию, когда значения весовой функции для всех групп оказываются одинаковыми: $w(q_i, n_i, v_{\gamma(i)}) = w$. Так, весовая функция может являться константной; если в случае полной априорной идентифицируемости каждая группа имеет свой параметр формы, то весовая функция зависит (если зависит) только от аргумента n_i ,

в результате для обеспечения равенства весов все группы должны иметь одинаковую размерность.

Обозначим $\hat{\theta}$, \hat{U}_j – ММП-оценки параметров в условиях нормальных ошибок, данному случаю соответствует константная весовая функция $w = 1$.

Выделим ситуацию, когда множество матричных параметров U_j , $j = 1, \dots, n_u$, разбито на два подмножества J_1 и J_2 так, что оценки параметров из подмножества J_1 пересчитываются на итерации покомпонентного оценивания первыми и в каждой матрице V_i присутствует только один матричный параметр из подмножества J_1 . В этом случае на очередной s -й итерации покомпонентного оценивания справедливы равенства

$$\hat{\theta}^{(s)} = \hat{\theta}^{(s)},$$

$$\hat{U}_{j_1}^{(s)} = w \hat{U}_{j_1}^{(s)}, \quad \hat{U}_{j_2}^{(s)} = \hat{U}_{j_2}^{(s)}, \quad j_1 \in J_1, \quad j_2 \in J_2.$$

Таким образом, полученные в результате работы алгоритма оценки $\hat{\theta}$, U_j равны ММП-оценкам в условиях нормальных ошибок с ковариационной матрицей i -й группы, равной V_i (за исключением матричных параметров из подмножества J_1 , увеличенных по сравнению с соответствующими оценками в нормальном случае в w раз).

3.1.3. ЧАСТНЫЕ СЛУЧАИ ЭЛЛИПТИЧЕСКОГО РАСПРЕДЕЛЕНИЯ

Получить аналитический вид априорных значений можно для широкого круга распределений. Рассмотрим многомерные распределения эллиптического типа, указанные в табл. 3.1.

Для *нормального распределения*

$$w(q_i, n_i) = 1$$

и все этапы одной итерации покомпонентной процедуры являются неитеративными.

Рассмотрим результаты для *распределения Стьюдента*. Поскольку

$$\frac{\partial \tilde{g}(q_i, n_i, v_{\gamma(i)})}{\partial q_i} = - \frac{n_i + v_{\gamma(i)}}{2 v_{\gamma(i)} + q_i} \frac{(n_i + v_{\gamma(i)})/2 + 1}{(n_i + v_{\gamma(i)})/2 + 1},$$

для весовой функции справедливо представление

$$w(q_i, n_i, v_{\gamma(i)}) = \frac{n_i + v_{\gamma(i)}}{v_{\gamma(i)} + q_i}.$$

Уравнение (3.6) в этом случае имеет вид

$$\hat{q}_i = n_i \frac{\hat{v}_{\gamma(i)} + \hat{q}_i}{n_i + \hat{v}_{\gamma(i)}}.$$

В результате получим

$$\hat{q}_i = n_i \text{ и } w_i = 1.$$

В данном случае величины \hat{q}_i и w_i имеют указанные априорные значения независимо от априорной идентифицируемости соответствующего параметра формы и его значения, причем весовая функция является константной.

Уравнение (3.9) приводится к виду

$$\text{DG} \left(\frac{n_i + \hat{v}_i}{2} \right) - \ln(n_i + \hat{v}_i) + \ln(\hat{v}_i) - \text{DG} \left(\frac{\hat{v}_i}{2} \right) = 0.$$

Решением уравнения является

$$\hat{v}_i = \infty.$$

Таким образом, ММП-оценки в этом случае совпадают с ММП-оценками, полученными в условиях нормального распределения.

Рассмотрим оценивание при ошибках, имеющих многомерное распределение Стьюдента.

В случае априорной неидентифицируемости для вычисления ММП-оценок следует использовать исходную систему уравнений правдоподобия. Получаемый алгоритм назовем адаптивным. Им мож-

но пользоваться и в остальных случаях, игнорируя возможности учета априорной идентифицируемости.

В случае полной априорной идентифицируемости, согласно полученным результатам, следует воспользоваться ММП-оцениванием в рамках нормальной модели.

В случае частичной априорной идентифицируемости будем использовать алгоритм, который назовем частично адаптивным. Рассмотрим его особенности.

Априорно идентифицируемые параметры формы можно исключить из числа оцениваемых параметров, подставив их значения в функцию правдоподобия. Если же представление (3.6) справедливо при априорно неидентифицируемом параметре формы, то в уравнение для его вычисления можно подставить априорное значение \hat{q}_i . Аналогично при вычислении весовой функции можно использовать априорные значения w_i для соответствующих групп наблюдений.

Уравнения (3.12) в рассматриваемом случае имеют вид

$$\frac{n_i + \hat{v}_{\gamma(i)}}{\hat{v}_{\gamma(i)} + \hat{q}_i} \hat{q}_i - n_i = -\beta_i, \quad i = r_a + 1, \dots, r_a + r_c.$$

Их можно использовать, например, для дополнительной проверки точности получаемого решения. При этом возможные интерпретации данных уравнений можно получить, выразив из них какую-либо из вычисляемых величин, например,

$$\hat{v}_{\gamma(i)} = \frac{\hat{q}_i \beta_i}{n_i - \hat{q}_i - \beta_i}, \quad \hat{q}_i = \frac{\hat{v}_{\gamma(i)}(n_i - \beta_i)}{\hat{v}_{\gamma(i)} + \beta_i}, \quad w_i = 1 + \frac{\beta_i}{\hat{v}_{\gamma(i)}}.$$

Рассмотрим результаты для *двустороннего экспоненциального распределения*. Получаем

$$\frac{\partial \tilde{g}(q_i, n_i, v_{\gamma(i)})}{\partial q_i} = -\frac{v_{\gamma(i)}}{2} \exp -q_i^{v_{\gamma(i)}/2} q_i^{v_{\gamma(i)}/2-1},$$

$$w(q_i, n_i, v_{\gamma(i)}) = v_{\gamma(i)} q_i^{v_{\gamma(i)}/2-1}.$$

Уравнение (3.6) в этом случае имеет вид

$$\hat{q}_i = \frac{n_i}{\hat{v}_{\gamma(i)} \hat{q}_i^{\hat{v}_{\gamma(i)}/2-1}},$$

его решение есть

$$\hat{q}_i = \left(\frac{n_i}{\hat{v}_{\gamma(i)}} \right)^{2/\hat{v}_{\gamma(i)}}.$$

Следовательно,

$$w(n_i, v_{\gamma(i)}) = n_i \left(\frac{v_{\gamma(i)}}{n_i} \right)^{2/v_{\gamma(i)}}.$$

Уравнение (3.9) приводится к виду

$$\frac{n_i}{\hat{v}_i^2} \left[\text{DG} \left(\frac{n_i}{\hat{v}_i} \right) - \ln \left(\frac{n_i}{\hat{v}_i} \right) \right] + \frac{1}{\hat{v}_i} = 0.$$

Решением уравнения является

$$\hat{v}_i = \infty.$$

Для функции $w(n_i, v_i)$ получим

$$w(n_i) = \lim_{v_i \rightarrow \infty} w(n_i, v_i) = n_i.$$

Весовая функция в данном случае является константной только для групп одинаковой размерности.

Поскольку при $v_i \rightarrow \infty$ двустороннее экспоненциальное распределение стремится к равномерному, ММП-оценки в данном случае совпадают с ММП-оценками, полученными в условиях равномерного распределения.

Для *обобщенного распределения Симпсона* при $q_i < 1$ получаем

$$\frac{\partial \tilde{g}(q_i, n_i, v_{\gamma(i)})}{\partial q_i} = -\frac{v_{\gamma(i)}}{2} q_i^{v_{\gamma(i)}/2-1},$$

$$w(q_i, n_i, v_{\gamma(i)}) = \frac{v_{\gamma(i)} q_i^{v_{\gamma(i)}/2-1}}{1 - q_i^{v_{\gamma(i)}/2}}.$$

Уравнение (3.6) в этом случае имеет вид

$$\hat{q}_i = n_i \frac{1 - \hat{q}_i^{\hat{v}_{\gamma(i)}/2}}{\hat{v}_{\gamma(i)} \hat{q}_i^{\hat{v}_{\gamma(i)}/2-1}},$$

его решение есть

$$\hat{q}_i = \left[\frac{n_i}{n_i + \hat{v}_{\gamma(i)}} \right]^{2/\hat{v}_{\gamma(i)}}.$$

Следовательно,

$$w(n_i, v_{\gamma(i)}) = n_i \left[\frac{n_i + v_{\gamma(i)}}{n_i} \right]^{2/v_{\gamma(i)}}.$$

Уравнение (3.9) приводится к виду

$$\frac{1}{\hat{v}_i(n_i + \hat{v}_i)} - \frac{1}{\hat{v}_i^2} \ln \left(\frac{n_i + \hat{v}_i}{n_i} \right) = 0.$$

Его решением является

$$\hat{v}_i = \infty.$$

Для функции $w(n_i, v_i)$ получим

$$w(n_i) = \lim_{v_i \rightarrow \infty} w(n_i, v_i) = n_i.$$

Весовая функция в данном случае является константной только для групп одинаковой размерности.

Поскольку при $v_i \rightarrow \infty$ обобщенное распределение Симпсона стремится к равномерному, ММП-оценки в данном случае совпадают с ММП-оценками, полученными в условиях равномерного распределения.

3.1.4. ПРИКЛАДНЫЕ АСПЕКТЫ РАЗРАБОТАННОГО ПОДХОДА

Согласно результатам, полученным как аналитически (изложены выше), так и эмпирически (путем моделирования), в ряде случаев можно установить соотношения ММП-оценок $\hat{\theta}$, \hat{U}_j , \hat{V}_i и ММП-оценок $\hat{\theta}$, \hat{U}_j , \hat{V}_i , полученных, исходя из нормальности ошибок. Рассмотрим возможные соотношения:

$$\hat{\theta} = \hat{\theta}; \quad (3.13)$$

$$\hat{U}_{j_1} = w\hat{U}_{j_1}, \quad \hat{U}_{j_2} = \hat{U}_{j_2}, \quad j_1 \in J_1, \quad j_2 \in J_2; \quad (3.14)$$

$$\hat{V}_i = w_i\hat{V}_i, \quad i = 1, \dots, p. \quad (3.15)$$

При этом выполнение соотношения (3.14) будем рассматривать применительно к ситуации, когда все веса групп одинаковы и возможно разбиение множества параметров ковариационной матрицы на подмножества J_1 и J_2 так, как описано выше. Соотношения (3.13) – (3.15) справедливы:

- в случае полной априорной идентифицируемости;
- в иных случаях, если используются константная весовая функция или веса, равные размерностям групп.

Параметризацию ковариационной матрицы, приводящую к случаю полной априорной идентифицируемости, удобно использовать на практике, поскольку в данном случае на основании соотношений (3.13) – (3.15) можно воспользоваться оцениванием нормальной модели. Соотношения (3.14), (3.15) позволяют легко вычислить оценки ковариационных матриц или их параметры при известных значениях весов групп. При этом стремиться к выполнению условий разбиения па-

раметров ковариационной матрицы на подмножества J_1 и J_2 следует, если эти параметры не выступают в роли мешающих.

С другой стороны, на практике фактическое распределение ошибок неизвестно, следовательно, неизвестен вид весовой функции. В результате, применяя алгоритм оценивания в условиях нормальных ошибок наблюдений, можно получить, помимо оценок параметров регрессионных уравнений, лишь оценки корреляционных матриц и отношений дисперсий ошибок.

Однако следует заметить, что используемая параметризация псевдоковариационных матриц в любом случае позволяет оценить их параметры только с точностью до скалярного сомножителя. Следовательно, в рассматриваемом случае для получения оценок параметров при произвольном распределении эллиптического типа достаточно воспользоваться покомпонентным алгоритмом ММП-оценивания в условиях нормальных ошибок наблюдений, значение же весовой функции потребуется лишь при вычислении по полученным оценкам параметров оценок ковариационных или псевдоковариационных матриц, определяемых однозначно.

В случае использования параметризации ковариационной матрицы, не приводящей к случаю полной априорной идентифицируемости, можно воспользоваться оцениванием в предположении нормальности ошибок. При равных весах групп или при весах, равных размерностям групп, сохраняются соотношения (3.13) – (3.15), но полученные оценки уже не будут ММП-оценками, и возможно, что найденная адаптивным или частично адаптивным алгоритмом ММП-оценка окажется лучше.

Возможность свести в некоторых случаях оценивание параметров при произвольном распределении ошибок эллиптического типа к оцениванию нормальной модели позволяет говорить об определенной устойчивости в данных случаях ММП-оценок вектора параметров θ , полученных для нормального распределения, в условиях неизвестного эллиптического распределения ошибок.

В результате может быть целесообразным специально конструировать мультипликативную ковариационную структуру, приводящую к случаю полной априорной идентифицируемости. Для этого достаточно в ковариационную матрицу каждой группы добавить по скалярному параметру. Заметим, что такая модификация ковариационной структуры практически не увеличивает вычислительную сложность алгоритма

по сравнению с исходной (при известном распределении ошибок), так как не требует дополнительных вычислений функций от данных, одновременно освобождая от необходимости решать нелинейные уравнения для оценивания параметров формы. Учитывая специфику каждого конкретного случая, можно получить и более экономную параметризацию.

3.1.5. ИССЛЕДОВАНИЕ

Для проверки работоспособности рассмотренных подходов к оцениванию параметров было проведено методом статистических испытаний экспериментальное исследование. Данные генерировались в соответствии с моделью

$$y = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3 + e, \quad (3.16)$$

где $\theta_1 = 10$, $\theta_2 = -5$, $\theta_3 = 2$, $\theta_4 = 0,2$, $x_i \in [-1, 1]$.

При оценивании на параметры было наложено ограничение

$$\theta_2 + \theta_3 = -3. \quad (3.17)$$

Мерой, отражавшей полезность рассмотренных подходов, являлась величина τ , определяемая формулой (2.37). Кроме того, подходы сравнивались по вычислительным затратам, мерой которых служила величина t – время в секундах, затрачиваемое на поиск решения.

Рассмотрим результаты исследования случая полной априорной идентифицируемости.

Выборка состояла из 27 групп по четыре наблюдения в каждой ($p = 27$, $n_i = 4$, $i = 1, \dots, 27$). В качестве ошибок использовались четырехмерные случайные векторы, имевшие распределение Стьюдента с параметром формы, равным единице (распределение Коши). Псевдоковариационные матрицы групп имели мультипликативные структуры вида

$$V_i = \rho U_0 U_i,$$

где $\rho = 0,0484$, $U_1 = 0,1$, $U_2 = \dots = U_{27} = 1$,

$$U_0 = \begin{bmatrix} 1 & \sqrt{0,081} & \sqrt{0,081} & \sqrt{0,081} \\ \sqrt{0,081} & 0,1 & 0,09 & 0,09 \\ \sqrt{0,081} & 0,09 & 0,1 & 0,09 \\ \sqrt{0,081} & 0,09 & 0,09 & 0,1 \end{bmatrix}.$$

Сравнивались три оценки: МНК-оценка, ММП-оценка нормальной модели и адаптивная ММП-оценка (с оцениванием одного общего параметра формы). Значения оценок параметров формы предполагались лежащими в интервале $0,1 \dots 100$.

В табл. 3.2 приведены средние значения величин τ и t , полученные по 100 испытаниям. Результаты подтверждают, что последние две оценки имеют близкие показатели качества, при этом последняя оценка вычислялась на порядок медленнее.

Таблица 3.2

Сравнение оценок в случае полной априорной идентифицируемости

Оценка	τ	t
МНК	0,0138320	0,003
ММП для нормальной модели	$3,79346 \cdot 10^{-6}$	0,209
Адаптивная ММП	$3,80986 \cdot 10^{-6}$	1,7879

Рассмотрим результаты исследования случая частичной априорной идентифицируемости.

Структура выборки была такой же, как и в предыдущем случае, ошибки имели распределение Коши. Использовались мультипликативные структуры псевдоковариационных матриц групп вида

$$V_i = \rho U_0 U_i, \quad i = 1, \dots, 9,$$

$$V_i = V_{i+6} = V_{i+12} = \rho U_0 U_i, \quad i = 10, \dots, 15,$$

параметры U_0, \dots, U_{15} и величина ρ имели те же значения, что и в предыдущем случае. Таким образом, первым девяти группам соответствуют априорно идентифицируемые величины q_i и w_i .

Сравнивались четыре оценки: МНК-оценка; ММП-оценка нормальной модели (в данном случае сведение к нормальному случаю некорректно); ММП-оценка студентовской модели, полученная адаптивным алгоритмом (с оцениванием двух общих параметров формы: первый – для первых девяти групп, второй – для остальных), и ММП-оценка студентовской модели, полученная частично адаптивным алгоритмом (с априорно идентифицируемым общим параметром формы у первых девяти групп и с оцениваемым общим параметром формы у остальных групп). Значения оценок параметров формы предполагались лежащими в интервале $0,1 \dots 100$.

В табл. 3.3 приведены средние значения показателей τ и t , полученные по 100 испытаниям.

Таблица 3.3

Сравнение оценок в случае частичной априорной идентифицируемости

Оценка	τ	t
МНК	0,0138320	0,003
ММП для нормальной модели	$5,46181 \cdot 10^{-6}$	0,1688
ММП для студентовской модели – адаптивный алгоритм	$4,1005 \cdot 10^{-6}$	1,8273
ММП для студентовской модели – частично адаптивный алгоритм	$3,52546 \cdot 10^{-6}$	0,659

Результаты показывают, что МНК имеет худшее качество, ММП-оценка нормальной модели проигрывает по качеству ММП-оценкам студентовской модели. Оценка студентовской модели, полученная по частично адаптивному алгоритму, имела самое высокое качество, при этом частично адаптивный алгоритм работал почти в три раза быстрее адаптивного.

Рассмотрим результаты исследования свойств оценок, получаемых целенаправленным переходом к случаю полной априорной идентифицируемости.

Рассматривались модели с известным эллиптическим распределением ошибок групп размерности 4, общими псевдоковариационной матрицей Σ и параметром формы ν .

Традиционный подход предполагает оценивание по ММП параметров θ , Σ и ν . Заменяя исходную модель псевдоковариационных матриц групп моделью

$$V_1 = \Sigma, \quad V_i = \lambda_{i-1} \Sigma, \quad i = 2, \dots, p,$$

где λ_i , $i = 1, \dots, p-1$ – скалярные параметры, приходим к случаю полной априорной идентифицируемости.

Введем нормальную модель с ковариационными матрицами групп

$$\tilde{V}_1 = \tilde{\Sigma}, \quad \tilde{V}_i = \tilde{\lambda}_{i-1} \tilde{\Sigma}, \quad i = 2, \dots, p, \quad (3.18)$$

где $\tilde{\Sigma}$ – матричный параметр, $\tilde{\lambda}_i$, $i = 1, \dots, p-1$ – скалярные параметры. В рамках данной модели требуется оценить параметры θ , $\tilde{\Sigma}$, $\tilde{\lambda}_i$, $i = 1, \dots, p-1$. При этом между ковариационными параметрами, согласно (3.11), существуют следующие соотношения:

$$\tilde{\Sigma} = \frac{1}{w_1} \Sigma, \quad \tilde{\lambda}_i = \frac{w_1}{w_{i+1}} \lambda_i, \quad i = 1, \dots, p-1.$$

Данные генерировались в соответствии с уравнением (3.16) при $\theta_1 = 10$, $\theta_2 = 5$, $\theta_3 = 2$, $\theta_4 = 0,5$.

Векторы ошибок наблюдений групп определялись по формуле

$$E_i = \frac{1}{\sqrt{\xi_i}} Z_i,$$

где Z_i – многомерный нормальный вектор с нулевым математическим ожиданием и ковариационной матрицей Σ ; ξ_i – положительная случайная величина, имеющая распределение со скалярным параметром формы $\nu > 0$.

Использовались следующие распределения величин ξ_i : распределение Парето с плотностью $g(z) = \frac{\nu}{z^{\nu+1}}$, где $z \in 1, \infty$; распределение с

плотностью $g(z) = \nu z^{\nu-1}$, где $z \in [0,1]$, приводящее к слэш-распределению векторов E_i [149]; распределение $G\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$, приводящее к студентовскому распределению векторов E_i (см. п. 1.7).

Псевдоковариационная матрица групп имела мультипликативную структуру вида

$$\Sigma = U_1 \otimes U_2,$$

где $U_1 = U_2 = \sqrt{0,073125} \begin{bmatrix} 10 & 0 \\ 0 & 1 \end{bmatrix}$.

Сравнивались адаптивная робастная оценка, полученная как квази-правдоподобная, исходя из распределения Стьюдента с общим параметром формы для всех групп, и ММП-оценка в рамках нормальной модели с ковариационной структурой типа (3.18), приводящей к случаю полной априорной идентифицируемости (ПАИ); кроме того, вычислялась квазиправдоподобная оценка в рамках нормальной модели, а для первых двух типов распределений также использовалась ММП-оценка в рамках взвешенной нормальной модели с ковариационной матрицей группы $\frac{1}{w_i} \Sigma$ при известных «истинных» значениях весов

$w_i = \xi_i$. При оценивании параметра формы распределения Стьюдента предполагалось, что его значение лежит в интервале 0,01 – 300.

В табл. 3.4 – 3.11 даны средние значения показателей τ и t , полученные по 100 испытаниям при 64 группах наблюдений. Результаты исследования показывают, что нельзя однозначно говорить о преимуществе оценивания в рамках случая полной априорной идентифицируемости: с точки зрения величины τ при величинах ξ , имеющих распределение Парето с малым значением параметра формы, данная оценка имеет преимущество перед адаптивной робастной, при других распределениях имеет преимущество адаптивная робастная оценка. Нельзя также однозначно сказать о преимуществах одной из оценок относительно другой с точки зрения вычислительных затрат.

Таблица 3.4

Сравнение оценок при ξ , имеющих распределение Парето с $\nu = 0,05$

Модель ошибок	τ	t
Нормальная	$1,73832 \cdot 10^{-6}$	0,0175
Стьюдента	$1,19282 \cdot 10^{-17}$	1,282
Нормальная (ПАИ)	$1,46849 \cdot 10^{-18}$	0,3138

Таблица 3.5

Сравнение оценок при ξ , имеющих распределение Парето с $\nu = 0,2$

Модель ошибок	τ	t
Взвешенная нормальная	$4,86128 \cdot 10^{-9}$	0,0085
Нормальная	$6,5872 \cdot 10^{-6}$	0,01
Стьюдента	$6,95766 \cdot 10^{-9}$	2,4063
Нормальная (ПАИ)	$6,5579 \cdot 10^{-9}$	1,0082

Таблица 3.6

Сравнение оценок при ξ , имеющих распределение Парето с $\nu = 0,5$

Модель ошибок	τ	t
Взвешенная нормальная	$1,09661 \cdot 10^{-6}$	0,0088
Нормальная	$1,33319 \cdot 10^{-5}$	0,0098
Стьюдента	$3,53308 \cdot 10^{-6}$	0,5085
Нормальная (ПАИ)	$2,04014 \cdot 10^{-6}$	0,7762

Таблица 3.7

Сравнение оценок при ξ , имеющих распределение Парето с $\nu = 2$

Модель ошибок	τ	t
Взвешенная нормальная	$2,17555 \cdot 10^{-5}$	0,0094
Нормальная	$2,66286 \cdot 10^{-5}$	0,0102
Стьюдента	$2,64045 \cdot 10^{-5}$	1,8408
Нормальная (ПАИ)	$3,99744 \cdot 10^{-5}$	0,8268

Таблица 3.8

Сравнение оценок при слэш-распределении с $\nu = 1$

Модель ошибок	τ	t
Взвешенная нормальная	$8,12867 \cdot 10^{-5}$	0,0096
Нормальная	$2,85122 \cdot 10^{-4}$	0,0154
Стьюдента	$9,71 \cdot 10^{-5}$	0,1842
Нормальная (ПАИ)	$1,44219 \cdot 10^{-4}$	0,792

Таблица 3.9

Сравнение оценок при слэш-распределении с $\nu = 2$

Модель ошибок	τ	t
Взвешенная нормальная	$6,02601 \cdot 10^{-5}$	0,0095
Нормальная	$7,85518 \cdot 10^{-5}$	0,0097
Стьюдента	$6,96738 \cdot 10^{-5}$	1,5501
Нормальная (ПАИ)	$1,05277 \cdot 10^{-4}$	0,6771

Таблица 3.10

Сравнение оценок при распределении Стьюдента с $\nu = 1$

Модель ошибок	τ	t
Нормальная	$6,06386 \cdot 10^{-2}$	0,0479
Стьюдента	$6,03908 \cdot 10^{-5}$	0,217
Нормальная (ПАИ)	$8,36902 \cdot 10^{-5}$	1,1644

Таблица 3.11

Сравнение оценок при распределении Стьюдента с $\nu = 4$

Модель ошибок	τ	t
Нормальная	$7,29894 \cdot 10^{-5}$	0,0096
Стьюдента	$4,67445 \cdot 10^{-5}$	0,1928
Нормальная (ПАИ)	$7,33456 \cdot 10^{-5}$	0,7223

Поскольку ММП-оценки в рамках случая полной априорной идентифицируемости могут не обладать положительными асимптотическими свойствами, были изучены их свойства при увеличении числа групп. Уравнение регрессии и ковариационная структура имели тот же вид, что и в предыдущей части исследования. Использовалось распределение Парето величин ξ с $\nu = 0,2$. В табл. 3.12 дана информация, дополняющая представленную в табл. 3.5. В табл. 3.13 приведены средние значения модуля разности оценок и истинных значений параметров.

Результаты свидетельствуют о сохранении в рассматриваемом случае преимущества оценок, полученных в рамках случая полной априорной идентифицируемости, перед остальными и об улучшении их качества при увеличении числа групп.

Таблица 3.12

Средние значения величины τ при различном количестве групп

Число групп	Модель ошибок		
	Нормальная	Стьюдента	Нормальная (ПАИ)
32	$1,43347 \cdot 10^{-5}$	$1,48209 \cdot 10^{-7}$	$1,29681 \cdot 10^{-7}$
64	$6,5872 \cdot 10^{-6}$	$6,95766 \cdot 10^{-9}$	$6,5579 \cdot 10^{-9}$
128	$3,23537 \cdot 10^{-6}$	$2,86494 \cdot 10^{-10}$	$1,83508 \cdot 10^{-10}$
256	$1,39047 \cdot 10^{-6}$	$4,98363 \cdot 10^{-11}$	$1,12509 \cdot 10^{-11}$
512	$7,27778 \cdot 10^{-7}$	$1,18580 \cdot 10^{-11}$	$1,45072 \cdot 10^{-13}$

Таблица 3.13

Средние значения модуля разности оценок и истинных значений параметров

Число групп	$ \theta_0 - \hat{\theta}_0 $	$ \theta_1 - \hat{\theta}_1 $	$ \theta_2 - \hat{\theta}_2 $	$ \theta_3 - \hat{\theta}_3 $
32	$2,77802 \cdot 10^{-4}$	$1,05871 \cdot 10^{-4}$	$9,61727 \cdot 10^{-5}$	$8,98589 \cdot 10^{-5}$
64	$7,31354 \cdot 10^{-5}$	$6,14043 \cdot 10^{-5}$	$9,7639 \cdot 10^{-7}$	$5,84338 \cdot 10^{-5}$
128	$4,30342 \cdot 10^{-6}$	$6,22987 \cdot 10^{-6}$	$8,58169 \cdot 10^{-6}$	$3,96843 \cdot 10^{-6}$
256	$3,48599 \cdot 10^{-7}$	$2,24538 \cdot 10^{-6}$	$1,11092 \cdot 10^{-6}$	$9,38727 \cdot 10^{-7}$
512	$2,28378 \cdot 10^{-7}$	$9,24875 \cdot 10^{-8}$	$3,71855 \cdot 10^{-7}$	$8,34974 \cdot 10^{-8}$

3.2. ОЦЕНИВАНИЕ ПАРАМЕТРОВ МОДЕЛИ ПРИ НАЛИЧИИ РАЗНОТИПНЫХ ОТКЛИКОВ И ПРОПУСКОВ В ДАННЫХ

3.2.1. МОДЕЛЬ

Рассмотрим задачу оценивания параметров регрессионных моделей с количественными и качественными откликами при наличии пропусков в данных.

Постановка задачи моделирования зависимости вектора разнотипных откликов от вектора разнотипных факторов сделана в [48]; в [48, 111] предложено решение на основе логико-вероятностных моделей (см. также [47]).

Рассмотрим модель вероятностного типа, которая развивает методы, разработанные в [157, 161, 176] для общей модели сдвига (см. также [78]).

Пусть проведено N независимых наблюдений за откликами системы. В t -м наблюдении при фиксированном значении вектора входных переменных x_t измеряются значения векторов Y_t и Z_t количественных и качественных откликов соответственно.

Обозначим n_z размерность вектора Z_t ; I_s – количество возможных значений его s -го элемента. Тогда качественные отклики задают таблицу сопряженности с n_z входами, имеющую

$$n_c = \prod_{j=1}^{n_z} I_j$$

ячеек. Обозначим z_t номер ячейки, в которую попало t -е наблюдение; таким образом, z_t – скалярная дискретная величина, принимающая значения из множества $1, \dots, n_c$. В результате вместо вектора Z_t может использоваться скаляр z_t .

Распределение переменной z_t зададим в виде

$$\mathbf{P} \ z_t = j \mid x_t = \pi_j(x_t, \theta), \quad j = 1, \dots, n_c, \quad (3.19)$$

где \mathbf{P} – вероятность; π_j – некоторая функция, удовлетворяющая ограничениям

$$0 \leq \pi_j(x_t, \vartheta) \leq 1, \quad \pi_1(x_t, \vartheta) + \dots + \pi_{n_c}(x_t, \vartheta) = 1;$$

ϑ – вектор параметров модели качественных откликов. Часто для описания зависимости вероятностей от входных переменных используется логистическая регрессия (логит-модель) [146] вида

$$\pi_j(x_t, \vartheta) = \frac{e^{\eta_j x_t, \vartheta}}{\sum_{k=1}^{n_c} e^{\eta_k x_t, \vartheta}}, \quad j=1, \dots, n_c, \quad (3.20)$$

где $\eta_j x_t, \vartheta = X_{ij}^z \vartheta$, X_{ij}^z – вектор-строка значений регрессоров. На значения вектора ϑ может быть наложена система в общем случае нелинейных ограничений.

Вектор Y_t представим состоящим из n_t подвекторов (групп элементов) Y_{ti} размерности n_{ti} , $i=1, \dots, n_t$, так, что условная плотность распределения вектора Y_t в j -й ячейке имеет вид

$$g_{tj}(Y_t | x_t, \phi) = \prod_{i=1}^{n_t} g_{tij}(Y_{ti} | x_t, \phi),$$

где ϕ – совокупность параметров модели количественных откликов, $g_{tij}(Y_{ti} | x_t, \phi)$ – условная плотность распределения вектора Y_{ti} в j -й ячейке.

Из-за зависимости количественных и качественных откликов наблюдения количественных откликов неоднородны (разнораспределенные).

Частным случаем рассматриваемой модели является модель (1.58), которая может интерпретироваться как модель с разнотипными откликами, если источник неоднородности данных представить как полностью ненаблюдаемый качественный отклик.

С другой стороны, маргинальное распределение Y_t в рассматриваемой модели имеет плотность того же вида, что и (1.58):

$$g(Y_t | x_t, \vartheta, \phi) = \sum_{j=1}^{n_c} \pi_j(x_t, \vartheta) g_{tj}(Y_t | x_t, \phi).$$

Рассмотрим описание количественных откликов с помощью многомерной модели N/I -распределения (см. п. 1.7). Введем вспомогательную ненаблюдаемую положительную случайную величину τ_{ti} . Тогда распределение вектора Y_{ti} в j -й ячейке при фиксированном значении τ_{ti} является нормальным с математическим ожиданием $X_{tij}\theta$ и ковариационной матрицей $\frac{1}{\tau_{ti}}\Omega_{tij}$, где X_{tij} – матрица значений регрессоров, описывающих i -ю группу t -го наблюдения в j -й ячейке; θ – вектор неизвестных параметров регрессии; Ω_{tij} – симметричная положительно определенная матрица. На параметры уравнения регрессии при этом может быть наложена система линейных ограничений равенств вида (1.9).

Предположим, что величина τ_{ti} в j -й ячейке имеет гамма-распределение $G_{\alpha, \beta}$ с плотностью (1.54) и

$$\alpha = \frac{v_{\gamma(t,i,j)}}{2}, \quad \beta = \frac{v_{\gamma(t,i,j)}\lambda_{tij}}{2},$$

где $v_{\gamma(t,i,j)}$ – параметр формы распределения, при этом функция $\gamma(t,i,j)$ ставит в соответствие номерам наблюдения, группы и ячейки номер параметра формы и всего имеется n_{γ} различных параметров формы v_r , $r=1, \dots, n_{\gamma}$; λ_{tij} – скалярная положительная величина.

Тогда распределение вектора Y_{ti} в j -й ячейке является студентовским с плотностью вида

$$g_{tij}(Y_{ti} | x_t, \phi) = \frac{v_{\gamma(t,i,j)}^{v_{\gamma(t,i,j)}/2} \Gamma\left[\frac{v_{\gamma(t,i,j)} + n_{ti}}{2}\right]}{\pi^{n_{ti}/2} \Gamma\left[v_{\gamma(t,i,j)}/2\right]} |V_{tij}|^{-1/2} \left[v_{\gamma(t,i,j)} + q_{tij}\right]^{-v_{\gamma(t,i,j)} + n_{ti}/2},$$

где V_{tij} – симметричная положительно определенная псевдоковариационная матрица вида

$$V_{tij} = \lambda_{tij}\Omega_{tij};$$

$q_{tij} = Y_{ti} - X_{tij}\theta^T V_{tij}^{-1} Y_{ti} - X_{tij}\theta$. Предельным случаем распределения Стьюдента при $v_{\gamma(t,i,j)} \rightarrow \infty$ является нормальное распределение, также допустимое в модели.

Параметризуем матрицы Ω_{tij} посредством кронекерова произведения ряда симметричных положительно определенных матриц $U_{1,u}$, $u = 1, \dots, h_1$, размера $k_u \geq 1$:

$$\Omega_{tij} = \bigotimes_{u \in J(\Omega_{tij})} U_{1,u},$$

где $J(\Omega_{tij})$ – упорядоченная последовательность индексов матриц $U_{1,u}$, посредством которых параметризована матрица Ω_{tij} .

Скалярные величины λ_{tij} параметризуем посредством произведения положительных параметров $U_{2,u}$, $u = 1, \dots, h_2$:

$$\lambda_{tij} = \prod_{u \in J(\lambda_{tij})} U_{2,u},$$

где $J(\lambda_{tij})$ – множество индексов параметров $U_{2,u}$, являющихся сомножителями в λ_{tij} .

Векторы Z_t и Y_t могут содержать пропуски; в результате реально известны лишь их части, которые обозначим $Z_{t,obs}$ и $Y_{t,obs}$ соответственно. При наличии пропусков в качественных откликах неизвестно, в какой ячейке находится наблюдение, наблюдаемая часть качественных откликов позволяет провести лишь частичную классификацию наблюдения.

Введем обозначения: $Y_{ti,obs}$, $Y_{ti,mis}$ – векторы присутствующих и отсутствующих элементов i -й группы t -го наблюдения; $X_{tij,obs}$, $X_{tij,mis}$ – подматрицы матрицы X_{tij} , соответствующие $Y_{ti,obs}$, $Y_{ti,mis}$; $n_{ti,obs}$ – размер вектора $Y_{ti,obs}$; $\Omega_{tij,obs} = \Omega_{tij,obs,obs}$, $\Omega_{tij,obs,mis}$, $\Omega_{tij,mis,obs}$, $\Omega_{tij,mis} = \Omega_{tij,mis,mis}$ – подматрицы матрицы Ω_{tij} , соответ-

ствующие векторам $Y_{ti,obs}$, $Y_{ti,mis}$, так что псевдоковариационная матрица вектора $\begin{bmatrix} Y_{ti,obs}^T & Y_{ti,mis}^T \end{bmatrix}^T$ в j -й ячейке имеет вид

$$\lambda_{tij} \begin{bmatrix} \Omega_{tij,obs} & \Omega_{tij,obs,mis} \\ \Omega_{tij,mis,obs} & \Omega_{tij,mis} \end{bmatrix};$$

J_t – множество значений, которые может принимать переменная z_t при условии, что наблюдался вектор $Z_{t,obs}$; $\Phi = \vartheta, \phi$ – совокупность неизвестных параметров модели.

Совместное распределение векторов $Y_{t,obs}$, $Z_{t,obs}$ характеризуется плотностью

$$g(Y_{t,obs}, Z_{t,obs} | x_t, \Phi) = \sum_{j \in J_t} \pi_j(x_t, \vartheta) g_{tj}(Y_{t,obs} | x_t, \phi). \quad (3.21)$$

Заметим, что с помощью рассматриваемых моделей можно описывать многомодальные и асимметричные распределения ошибок количественных наблюдений. Полученные результаты могут быть использованы и при частичной классификации наблюдений, являющейся следствием не только пропусков, но и особенностей процедуры сбора данных.

3.2.2. ОЦЕНИВАНИЕ ПАРАМЕТРОВ

Для оценивания вектора параметров модели разнотипных откликов при наличии пропущенных наблюдений

$$\Phi = \vartheta^T, \phi^T$$

используем метод максимального правдоподобия.

Полученные в п. 3.1 алгоритмы вычисления ММП-оценок параметров регрессионной модели с мультипликативной ковариационной

структурой характеризуются достаточно простыми вычислительными формулами для ковариационных параметров. Для ММП-оценивания по данным с пропусками при построении функции правдоподобия по маргинальным распределениям ковариационная структура, основанная на кронекеровом произведении, нарушается, вследствие чего указанные вычислительные схемы не могут быть использованы. Поэтому для поиска оценок разрабатываются EM-алгоритмы (см. п. 1.4), сохраняющие простоту вычислительных формул.

В основе EM-алгоритмов вычисления ММП-оценок параметров модели количественных откликов при стьюдентовском распределении ошибок [156, 158, 160] лежит N/I -представление данного распределения (см. п. 1.7), а пропущенными данными, помимо пропущенных значений откликов (если имеются пропуски), считаются значения ненаблюдаемой величины τ , имеющей распределение $G\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$.

В РХ-EM модификации [162] при некоторых частных видах параметризации псевдоковариационной матрицы используется расширенная модель, в которой величины τ имеют распределение с плотностью $G\left(\frac{\nu}{2}, \frac{\nu}{2\gamma}\right)$, где γ – дополнительный скалярный параметр, $\gamma_0 = 1$. При мультипликативной ковариационной структуре общего вида функция редукции не может быть построена. Однако во введенной модели, по аналогии с данным алгоритмом, скалярные ковариационные параметры разделены на два подмножества, параметры одного из которых являются параметрами плотности распределения случайной величины τ . Использование такого представления может ускорить работу EM-алгоритмов.

Для получения ММП-оценок следует максимизировать функцию (1.43), имеющую вид

$$L \Phi = \ln g(Y_{obs}, Z_{obs} | X, \Phi) = \sum_{t=1}^N \ln \sum_{j \in J_t} \pi_j x_t \vartheta g_{tj} Y_{t,obs} | x_t, \phi,$$

где $X, Y_{obs}, Z_{obs} = x_t, Y_{t,obs}, Z_{t,obs} \quad \sum_{t=1}^N$.

Для конструирования EM-алгоритмов получим вид функции (1.44):

$$\begin{aligned}
 Q \Phi | \tilde{\Phi} &= \mathbf{E} \ln g(Y, z | X, \Phi | X, Y_{obs}, Z_{obs}, \tilde{\Phi}) = \\
 &= \sum_{t=1}^N \sum_{i=1}^{n_t} \sum_{j \in J_t} \tilde{\pi}_{tj} \mathbf{E} \ln g_{tij}(Y_{ti} | x_t, \phi | x_t, Y_{ti,obs}, z_t = j, \tilde{\Phi}) + \\
 &+ \sum_{t=1}^N \sum_{j \in J_t} \tilde{\pi}_{tj} \ln \pi_j(x_t, \vartheta) = Q_1(\phi | \tilde{\Phi}) + Q_2(\vartheta | \tilde{\Phi}),
 \end{aligned}$$

где $Y, z = Y_t, z_t \prod_{t=1}^N, \tilde{\pi}_{tj} = \mathbf{P}(z_t = j | x_t, Y_{t,obs}, Z_{t,obs}, \tilde{\Phi})$ – апостериорная вероятность:

$$\tilde{\pi}_{tj} = \begin{cases} \frac{\pi_j(x_t, \tilde{\vartheta}) g_{tj}(Y_{t,obs} | x_t, \tilde{\Phi})}{\sum_{k \in J_t} \pi_k(x_t, \tilde{\vartheta}) g_{tk}(Y_{t,obs} | x_t, \tilde{\Phi})}, & j \in J_t \\ 0, & j \notin J_t \end{cases},$$

$g_{tj}(Y_{t,obs} | x_t, \phi) = \prod_{i=1}^{n_t} g_{tij}(Y_{ti,obs} | x_t, \phi)$, $g_{tij}(Y_{ti,obs} | x_t, \phi)$ – плотность

распределения $Y_{ti,obs}$ в j -й ячейке.

Характерная черта рассматриваемой модели – многоэкстремальность функции правдоподобия. Поскольку итерации EM-алгоритма в данном случае сходятся обычно к локальному максимуму, необходим тщательный выбор начального приближения.

В [130] для задачи оценивания модели смеси многомерных нормальных распределений рассмотрен ряд стратегий выбора начального приближения, одна из которых состоит в следующем (см. также [151]). Многократно получают грубые приближения (путем выполнения EM-алгоритма из случайного начального приближения с остановом по достижении точности меньшей, чем требуется) с последующим отбором наилучшего из них как начального приближения. После получения на-

чального приближения EM-алгоритм выполняется с остановом по достижении требуемой точности. Такой процесс многократно повторяется с отбором наилучшего решения. Как показывает практика, данную стратегию можно рекомендовать также при моделировании разнотипных откликов по данным с пропусками.

Осторожность в выборе корня уравнения правдоподобия должна быть проявлена в том случае, если функция правдоподобия не ограничена. В частности, это имеет место для смеси нормальных распределений с неравными ковариационными матрицами, когда решение лежит на краю параметрического пространства [94, 155, 169]. Внимание здесь может быть направлено на локальные максимумы внутри параметрического пространства. При обычных условиях регулярности там существует последовательность корней уравнения правдоподобия, которые являются состоятельными, асимптотически эффективными и асимптотически нормально распределенными; с вероятностью, стремящейся к единице, эти корни соответствуют локальным максимумам внутри параметрического пространства [155, 169]. На практике целесообразно ограничить снизу значения оценок дисперсий (детерминантов ковариационных матриц) и вероятностей для каждого компонента смеси [169]. Аналогичные рекомендации могут быть даны и в рассматриваемом общем случае.

Задачи оптимизации функций $Q_1 \phi | \tilde{\Phi}$ и $Q_2 \vartheta | \tilde{\Phi}$ можно рассматривать независимо одну от другой.

Функция $Q_2 \vartheta | \tilde{\Phi}$ для логит-модели имеет вид

$$Q_2 \vartheta | \tilde{\Phi} = \sum_{t=1}^N \sum_{j \in J_t} \tilde{\pi}_{tj} \left[X_{tj}^z \vartheta - \ln \sum_{k=1}^{n_c} \exp X_{tk}^z \vartheta \right].$$

Соответствующая подсистема уравнений правдоподобия на s -й итерации EM-алгоритма имеет вид

$$\sum_{t=1}^N \sum_{j=1}^{n_c} X_{tj}^z \begin{bmatrix} \tilde{\pi}_{tj} - \pi_j \\ x_t, \vartheta^{(s)} \end{bmatrix} = 0.$$

Запишем функцию $Q_1 \phi | \tilde{\Phi}$ с точностью до константных слагаемых:

$$\begin{aligned}
 Q_1 \phi | \tilde{\Phi} = & - \sum_{(t,i,j) \in J_S} \tilde{\pi}_{tj} \left[\frac{1}{2} \ln |\Omega_{tij}| + \frac{1}{2} \text{tr} \Omega_{tij}^{-1} \tilde{w}_{tij} S_{tij} + \tilde{C}_{tij} \right] - \\
 & - \sum_{(t,i,j) \in J_N} \tilde{\pi}_{tj} \left[\frac{1}{2} \ln |V_{tij}| + \frac{1}{2} \text{tr} V_{tij}^{-1} S_{tij} + \tilde{\lambda}_{tij} \tilde{C}_{tij} \right] + \\
 & + \sum_{r=1}^{n_v} \sum_{(t,i,j) \in J(v_r)} \tilde{\pi}_{tj} \left[\frac{v_r}{2} \ln \frac{v_r \lambda_{tij}}{2} + \left(\frac{v_r}{2} - 1 \right) \tilde{u}_{tij} - \frac{v_r \lambda_{tij} \tilde{w}_{tij}}{2} - \ln \Gamma \left(\frac{v_r}{2} \right) \right],
 \end{aligned}$$

где J_S , J_N – множества троек индексов наблюдений-групп-ячеек, соответствующих распределениям Стьюдента и Гаусса; $\tilde{w}_{tij} = \mathbf{E} \tau_{ti} | Y_{ti,obs}, z_t = j, \tilde{\Phi}$, $S_{tij} = \tilde{Y}_{tij} - X_{tij} \theta$, $\tilde{Y}_{tij} = \tilde{Y}_{tij} - X_{tij} \theta^T$, $\tilde{Y}_{tij} = \mathbf{E} Y_{ti} | Y_{ti,obs}, z_t = j, \tilde{\Phi}$; $\tilde{Y}_{tij,obs} = Y_{ti,obs}$, $\tilde{Y}_{tij,mis} = X_{tij,mis} \tilde{\theta} + \tilde{\Omega}_{tij,mis,obs} \tilde{\Omega}_{tij,obs}^{-1} Y_{ti,obs} - X_{tij,obs} \tilde{\theta}$; \tilde{C}_{tij} – квадратная матрица размера n_{ti} , α, β -й элемент которой равен элементу матрицы $\tilde{\Omega}_{tij,mis} - \tilde{\Omega}_{tij,mis,obs} \tilde{\Omega}_{tij,obs}^{-1} \tilde{\Omega}_{tij,obs,mis}$, соответствующему элементам $Y_{ti\alpha}$ и $Y_{ti\beta}$, если они одновременно отсутствуют, и нулю – в остальных случаях; $J(v_r)$ – множество троек индексов наблюдений-групп-ячеек, для распределения которых v_r является параметром формы, $\tilde{u}_{tij} = \mathbf{E} \ln \tau_{ti} | Y_{ti,obs}, z_t = j, \tilde{\Phi}$.

Можно показать, что условное распределение $\tau_{ti} | Y_{ti,obs}, z_t = j, \tilde{\Phi}$ есть гамма-распределение $G \left[\tilde{v}_{\gamma(t,i,j)} + n_{ti,obs} / 2, \tilde{\lambda}_{tij} \tilde{v}_{\gamma(t,i,j)} + \right.$

$+ \tilde{q}_{tij,obs} / 2$], где $\tilde{q}_{tij,obs} = Y_{ti,obs} - X_{tij,obs} \tilde{\theta}^T \tilde{V}_{tij,obs}^{-1} Y_{ti,obs} - X_{tij,obs} \tilde{\theta}$, тогда имеем

$$\tilde{w}_{tij} = \frac{\tilde{v}_{\gamma t,i,j} + n_{ti,obs}}{\tilde{\lambda}_{tij} \tilde{v}_{\gamma t,i,j} + \tilde{q}_{tij,obs}},$$

$$\tilde{u}_{tij} = \text{DG} \left(\frac{\tilde{v}_{\gamma t,i,j} + n_{ti,obs}}{2} \right) - \ln \frac{\tilde{\lambda}_{tij} \tilde{v}_{\gamma t,i,j} + \tilde{q}_{tij,obs}}{2}.$$

Поиск условного максимума функции $Q_1 \phi | \tilde{\Phi}$ будем производить методом неопределенных множителей Лагранжа. Обозначим через $\hat{\phi}$ значение аргумента, соответствующее максимуму. Подсистему оценочных уравнений, соответствующую вектору θ , после ряда преобразований, аналогичных используемым в п. 1.1 при выводе формулы (1.15), можно представить в виде

$$\hat{\theta} = \hat{\theta}_0 + H^{-1} R^T \left[R H^{-1} R^T \right]^{-1} \left[d - R \hat{\theta}_0 \right],$$

$$H = \sum_{(t,i,j) \in J_S} \tilde{\pi}_{ij} \tilde{w}_{tij} X_{tij}^T \hat{\Omega}_{tij}^{-1} X_{tij} + \sum_{(t,i,j) \in J_N} \tilde{\pi}_{ij} X_{tij}^T \hat{V}_{tij}^{-1} X_{tij},$$

$$\hat{\theta}_0 = H^{-1} \left\{ \sum_{(t,i,j) \in J_S} \tilde{\pi}_{ij} \tilde{w}_{tij} X_{tij}^T \hat{\Omega}_{tij}^{-1} \tilde{Y}_{tij} + \sum_{(t,i,j) \in J_N} \tilde{\pi}_{ij} X_{tij}^T \hat{V}_{tij}^{-1} \tilde{Y}_{tij} \right\}.$$

Система оценочных уравнений включает дополнительно следующие уравнения:

$$\left[\hat{U}_{1,u} \right]_{\alpha\beta} = \frac{k_u}{\sum_{(t,i,j) \in J(U_{1,u})} \tilde{\pi}_{ij} n_{ti}} \left\{ \sum_{(t,i,j) \in J(U_{1,u}) \cap J_S} \tilde{\pi}_{ij} \text{tr} \hat{\Omega}_{tij}^{-1} [u, \alpha, \beta] \times \right.$$

$$\begin{aligned}
& \times \left[\tilde{w}_{tij} \widehat{S}_{tij} + \tilde{C}_{tij} \right] + \sum_{(t,i,j) \in J(U_{1,u}) \cap J_N} \tilde{\pi}_{tj} \operatorname{tr} \widehat{V}_{tij, [u, \alpha, \beta]}^{-1} \left[\widehat{S}_{tij} + \tilde{\lambda}_{tij} \tilde{C}_{tij} \right] \Big\} , \\
& \alpha, \beta = 1, \dots, k_u, \quad u = 1, \dots, h_1, \\
& \widehat{U}_{2,u} = \left[\sum_{(t,i,j) \in J(U_{2,u}) \cap J_S} \tilde{\pi}_{tj} \widehat{V}_{\gamma t,i,j} - \sum_{(t,i,j) \in J(U_{2,u}) \cap J_N} \tilde{\pi}_{tj} n_{ti} \right] \times \\
& \times \left\{ \sum_{(t,i,j) \in J(U_{2,u}) \cap J_S} \tilde{\pi}_{tj} \widehat{V}_{\gamma t,i,j} \tilde{w}_{tij} \prod_{\substack{k \in J \lambda_{tij} \\ k \neq u}} \widehat{U}_{2,k}^{(s)} - \frac{1}{[\widehat{U}_{2,u}]^2} \times \right. \\
& \left. \times \sum_{(t,i,j) \in J(U_{2,u}) \cap J_N} \tilde{\pi}_{tj} \prod_{\substack{k \in J \lambda_{tij} \\ k \neq u}} \frac{1}{\widehat{U}_{2,k}} \operatorname{tr} \widehat{\Omega}_{tij}^{-1} \left[\widehat{S}_{tij} + \tilde{\lambda}_{tij} \tilde{C}_{tij} \right] \right\}^{-1}, \quad (3.22) \\
& u = 1, \dots, h_2,
\end{aligned}$$

$$\ln \frac{\widehat{v}_r}{2} - \operatorname{DG} \left(\frac{\widehat{v}_r}{2} \right) + \Delta_r = 0, \quad r = 1, \dots, n_v,$$

$$\Delta_r = 1 + \left[\sum_{(t,i,j) \in J(v_r)} \tilde{\pi}_{tj} \right]^{-1} \sum_{(t,i,j) \in J(v_r)} \tilde{\pi}_{tj} \tilde{u}_{tij} - \widehat{\lambda}_{tij} \tilde{w}_{tij} + \ln \widehat{\lambda}_{tij},$$

где $J(U_{k,u})$ – множество троек индексов наблюдений-групп-ячеек, в число параметров псевдоковариационных матриц которых входит па-

параметр $U_{k,u}$; $\widehat{\Omega}_{tij[u,\alpha,\beta]}$, $\widehat{V}_{tij[u,\alpha,\beta]}$ – матрицы, получаемые из $\widehat{\Omega}_{tij}$, \widehat{V}_{tij} соответственно путем замены сомножителя $\widehat{U}_{1,u}$ матрицей той же размерности, α, β -й элемент которой равен единице, а остальные элементы – нулю; $\widehat{S}_{tij} = \widehat{Y}_{tij} - X_{tij}\widehat{\theta}$ $\widetilde{Y}_{tij} - X_{tij}\widehat{\theta}^T$.

Для поиска значений оценок неизвестных параметров можно использовать частный случай ЕСМ-алгоритма, соответствующий покомпонентному оцениванию. В рассматриваемом случае удобно разделить множество неизвестных параметров на следующие подмножества:

$$\theta, U_{1,1}, \dots, U_{1,h_1}, U_{2,1}, \dots, U_{2,h_2}, v_1, \dots, v_{n_v}.$$

Тогда итерационными являются только шаги максимизации по параметрам формы, а также параметрам $U_{2,u}$, если для них не пусто множество $J U_{2,u} \cap J_N$.

В [158 – 160] рассматривается ЕСМЕ-алгоритм для оценивания параметров распределения Стьюдента, в котором СМQ-шаг оценивания параметров формы заменяется СМL-шагом. Применим данный подход к оцениванию параметров формы для регрессионной модели в случае, если имеются только количественные отклики. При этом индекс ячейки будем опускать.

Уравнение, решаемое на СМL-шаге уточнения оценки параметра v_r , имеет вид

$$\frac{1}{2} \sum_{(t,i) \in J(v_r)} \left\{ \text{DG} \left(\frac{n_{ti,obs} + \widehat{v}_r}{2} \right) - \ln \frac{\widehat{v}_r + \widehat{q}_{ti,obs}}{2} - \frac{n_{ti,obs} + \widehat{v}_r}{\widehat{v}_r + \widehat{q}_{ti,obs}} \right\} +$$

$$+ \frac{\mu v_r}{2} \left\{ 1 + \ln \frac{\widehat{v}_r}{2} - \text{DG} \left(\frac{\widehat{v}_r}{2} \right) \right\} = 0,$$

где $\widehat{q}_{ti,obs} = Y_{ti,obs} - X_{ti,obs}\widehat{\theta}^T \widehat{V}_{ti,obs}^{-1} Y_{ti,obs} - X_{ti,obs}\widehat{\theta}$, $\mu(v_r)$ – мощность множества $J(v_r)$. Для поиска корня данного уравнения методом

Ньютона требуется вторая производная функции $L(\Phi)$ по параметру v_r при его значении, равном \hat{v}_r ; производная имеет вид

$$\begin{aligned} & \frac{1}{2} \sum_{(t,i) \in J(v_r)} \left\{ \frac{1}{2} \text{DG}' \left(\frac{n_{ti,obs} + \hat{v}_r}{2} \right) - \frac{1}{\hat{v}_r + n_{ti,obs}} \right\} + \\ & + \frac{\mu v_r}{2} \left\{ \frac{1}{\hat{v}_r} - \frac{1}{2} \text{DG}' \left(\frac{\hat{v}_r}{2} \right) \right\} + \frac{1}{2} \sum_{(t,i) \in J(v_r)} \frac{n_{ti,obs} - \hat{q}_{ti,obs}}{\hat{v}_r + n_{ti,obs}} \frac{2}{\hat{v}_r + \hat{q}_{ti,obs}}. \end{aligned}$$

Если данное выражение не является отрицательным, то при использовании метода Ньютона последнюю сумму в [158] рекомендуется отбрасывать.

Вычисление по формуле (3.22) при наличии групп с нормально распределенными ошибками может приводить к отрицательным значениям на некоторых итерациях. В результате целесообразно второе подмножество формировать из ковариационных параметров студентовских групп, возможно также небольшое число нормальных групп. В последнем случае при наличии на очередной итерации отрицательного значения необходимо модифицировать алгоритм оценивания.

При оценивании параметров регрессионной модели в случае, если имеются только количественные отклики, для решения данной проблемы можно использовать СМЛ-шаг, приводящий к итеративному процессу решения уравнения (индекс ячейки опущен):

$$\begin{aligned} \hat{U}_{2,u} = & \sum_{(t,i) \in J} U_{2,u} \left[\frac{\hat{v}_{\gamma t,i} + n_{ti,obs}}{\hat{v}_{\gamma t,i} + \hat{q}_{ti,obs}} \times \right. \\ & \left. \times \prod_{\substack{k \in J \setminus \lambda_i \\ k \neq u}} \frac{1}{\hat{U}_{2,k}} \hat{E}_{t,i,obs}^T \hat{\Omega}_{t,i,obs}^{-1} \hat{E}_{t,i,obs} \right] \left[\sum_{(t,i) \in J} U_{2,u} n_{ti,obs} \right]^{-1}, \end{aligned}$$

где $\hat{E}_{t,i,obs} = Y_{ti,obs} - X_{ti,obs} \hat{\theta}$, например, на основе метода последовательных приближений.

Заметим, что рассмотренные способы построения алгоритмов оценивания параметров регрессионной модели с количественными откликами, ошибки которых имеют стьюдентовское распределение, можно применять и при отсутствии пропущенных наблюдений. В этом случае пропусками будут только значения латентной переменной τ . В отличие от EM-алгоритмов, непосредственная максимизация функции $L(\Phi)$ с использованием покомпонентного оценивания приводит к алгоритму, имеющему внутренние итеративные процессы на всех этапах одной итерации.

3.2.3. ИССЛЕДОВАНИЕ

Для определения работоспособности предложенных подходов проведено экспериментальное исследование.

Рассмотрим оценивание модели с разнотипными откликами на модельном примере. Выборка состояла из 600 наблюдений за четырехмерным вектором количественных откликов и одномерным качественным откликом с тремя возможными значениями.

Математическое ожидание каждого из количественных откликов в j -й ячейке вычислялось по формуле

$$\theta_{1j} + \theta_{2j}x_1 + \theta_{3j}x_2 + \theta_{4j}x_3,$$

где $\theta_{11} = 10$, $\theta_{21} = 5$, $\theta_{31} = 2$, $\theta_{41} = 0,2$, $\theta_{12} = 15$, $\theta_{22} = 3$, $\theta_{32} = 2$, $\theta_{42} = 0,2$, $\theta_{13} = 20$, $\theta_{23} = 5$, $\theta_{33} = 3$, $\theta_{43} = 0,2$, $x_i \in [-1, 1]$.

Наблюдения во всех ячейках имели распределение Стьюдента с общим параметром формы, равным 5, и мультипликативной структурой общей псевдоковариационной матрицы вида

$$V = U_1 \otimes U_2,$$

$$\text{где } U_1 = \begin{bmatrix} 0,0256 & 0 \\ 0 & 0,0256 \end{bmatrix}, \quad U_2 = \begin{bmatrix} 1 & 0,5 \\ 0,5 & 1 \end{bmatrix}.$$

В результате псевдоковариационная матрица была равна

$$V = \begin{bmatrix} 0,0256 & 0,0128 & 0 & 0 \\ 0,0128 & 0,0256 & 0 & 0 \\ 0 & 0 & 0,0256 & 0,0128 \\ 0 & 0 & 0,0128 & 0,0256 \end{bmatrix}.$$

Пропуски в наблюдениях количественных откликов генерировались случайно, при этом вероятность появления пропуска равнялась 0,2.

Качественный отклик был ненаблюдаемым и генерировался в соответствии с логит-моделью при

$$\eta_j x_t, \vartheta = \vartheta_{1j} + \vartheta_{2j}x_1, \quad j=1,2, \quad \eta_3 x_t, \vartheta = 0,$$

где $\vartheta_{11} = -20$, $\vartheta_{21} = 40$, $\vartheta_{12} = 0$, $\vartheta_{22} = -20$.

При $x_1 > 0,5$ максимальна вероятность попадания наблюдения в первую ячейку, при $x_1 < 0$ – во вторую, при $0 < x_1 < 0,5$ – в третью. Фрагмент графика зависимости вероятностей от переменной x_1 показан на рис. 3.1.

Критерием останова алгоритма было относительное изменение функции $L \Phi$ на соседних итерациях, меньшее величины ε , равной 10^{-7} . Для определения начального приближения использовалось лучшее из 20 грубых решений с $\varepsilon = 10^{-2}$, при этом каждое из них было получено за 9 – 17 итераций.

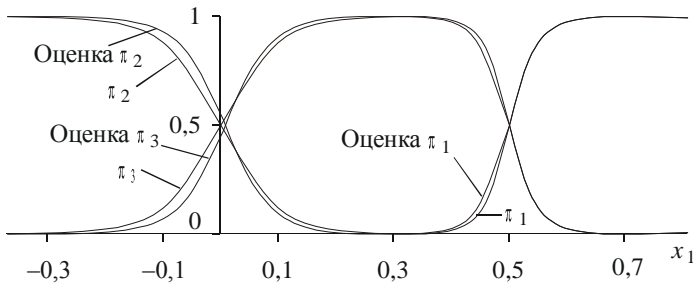


Рис. 3.1. Зависимости вероятностей от x_1

В основной части алгоритма было сделано 44 итерации. Получены следующие результаты (крышками обозначены оценки соответствующих величин):

$$\hat{\theta}_{11} = 9,907986, \quad \hat{\theta}_{21} = 5,111021, \quad \hat{\theta}_{31} = 2,026608, \quad \hat{\theta}_{41} = 0,2026251,$$

$$\hat{\theta}_{12} = 15,00301, \quad \hat{\theta}_{22} = 3,006231, \quad \hat{\theta}_{32} = 2,003477, \quad \hat{\theta}_{42} = 0,1886624,$$

$$\hat{\theta}_{13} = 20,00635, \quad \hat{\theta}_{23} = 5,001824, \quad \hat{\theta}_{33} = 3,018519, \quad \hat{\theta}_{43} = 0,2050659,$$

$$\hat{\theta}_{11} = -18,57852, \quad \hat{\theta}_{21} = 37,37919, \quad \hat{\theta}_{12} = 0,19018, \quad \hat{\theta}_{22} = -24,07816,$$

$$\hat{V} = \begin{bmatrix} 0,02693285 & 0,01360471 & 0,00114118 & 0,0005764495 \\ 0,01360471 & 0,02688295 & 0,0005764495 & 0,001139066 \\ 0,00114118 & 0,0005764495 & 0,02742213 & 0,01385186 \\ 0,0005764495 & 0,001139066 & 0,01385186 & 0,02737132 \end{bmatrix},$$

$$\hat{v} = 4,868008.$$

Оценки зависимостей вероятностей от переменной x_1 показаны на рис. 3.1. При классификации наблюдений из выборки по критерию максимума апостериорной вероятности не было ни одной ошибки. Результаты говорят о хорошей точности полученных оценок.

Рассмотрим исследование EM-алгоритмов оценивания параметров модели в случае, если имеются только количественные отклики, при наличии пропусков. Исследование проведено методом статистических испытаний.

Данные генерировались в соответствии с уравнением (3.16) при $\theta_1 = 10$, $\theta_2 = -5$, $\theta_3 = 2$, $\theta_4 = 0,2$, $x_i \in [-1, 1]$, распределение ошибок наблюдений – студентовское. При оценивании на параметры было наложено ограничение (3.17), а значение оценки параметра формы распределения ошибок наблюдений предполагалось лежащим в интервале $0,01 \dots 1000$.

Показателями, отражавшими полезность разработанных алгоритмов, являлись: t – время в секундах, затрачиваемое на поиск решения, и k – число итераций.

Выборка состояла из 64 восьмимерных наблюдений. Использовались мультипликативные структуры псевдоковариационной матрицы вида

$$V_i = \rho_i U_1 \otimes U_2,$$

где $i = 1, \dots, 64$, $k_1 = 4$, $k_2 = 2$,

$$U_1 = \sqrt{0,0557125} \begin{bmatrix} 10 & \sqrt{2,5} & \sqrt{2,5} & \sqrt{2,5} \\ \sqrt{2,5} & 1 & 0,5 & 0,5 \\ \sqrt{2,5} & 0,5 & 1 & 0,5 \\ \sqrt{2,5} & 0,5 & 0,5 & 1 \end{bmatrix},$$

$$U_2 = \sqrt{0,0557125} \begin{bmatrix} 10 & \sqrt{2,5} \\ \sqrt{2,5} & 1 \end{bmatrix},$$

при следующих мультипликативных параметризациях скаляра ρ_i :

- 1) $\rho_i = \delta_1$, $i = 1, \dots, 64$;
- 2) $\rho_i = \delta_{1+[(i-1)/16]}$, $i = 1, \dots, 64$, $[\cdot]$ – целая часть числа;
- 3) $\rho_i = \delta_{1+[(i-1)/16]} \sigma_{1+[(i-1)/16]}$, $i = 1, \dots, 64$;
- 4) $\rho_i = \delta_{1+[(i-1)/8]}$, $i = 1, \dots, 64$;
- 5) $\rho_i = \delta_{1+[(i-1)/8]} \sigma_{1+[(i-1)/8]}$, $i = 1, \dots, 64$;
- 6) $\rho_i = \delta_{1+[(i-1)/16]} \sigma_{1+[(i-1)/8]}$, $i = 1, \dots, 64$;
- 7) $\rho_i = \delta_{1+[(i-1)/4]} \sigma_{1+[(i-1)/8]}$, $\rho_{i+32} = \delta_{1+[(i-1)/4]} \sigma_{5+[(i-1)/8]}$, $i = 1, \dots, 32$;
- 8) $\rho_i = \delta_{1+[(i-1)/4]} \sigma_{1+[(i-1)/8]}$, $\rho_{i+16} = \delta_{1+[(i-1)/4]} \sigma_{3+[(i-1)/8]}$,
 $\rho_{i+32} = \delta_{1+[(i-1)/4]} \sigma_{5+[(i-1)/8]}$, $\rho_{i+48} = \delta_{1+[(i-1)/4]} \sigma_{7+[(i-1)/8]}$, $i = 1, \dots, 16$;
- 9) $\rho_i = \delta_1 \sigma_{1+[(i-1)/8]}$, $i = 1, \dots, 64$;

все скалярные параметры δ_j и σ_j имели единичные значения.

Пропуски в наборе данных генерировались случайно, при этом вероятность появления пропуска в каждом наблюдении равнялась 0,1.

Сравнивались четыре алгоритма. В первых двух алгоритмах – ЕСМ и ЕСМЕ – все ковариационные параметры принадлежали первому подмножеству, в последних двух алгоритмах – ЕСМ-G и ЕСМЕ-G – скалярные ковариационные параметры δ_j принадлежали второму подмножеству.

В качестве начальных значений в алгоритмах использовались оценка по методу наименьших квадратов для θ и единичные матрицы для ковариационных параметров. Критерием останова алгоритма было относительное изменение функции $L \Phi$ на соседних итерациях, меньшее 10^{-7} .

Рассмотрим результаты исследования случая, когда ошибки имели распределение Стьюдента с параметром формы $\nu = 8$, а начальное значение параметра формы было равно 1. В табл. 3.14 приведены средние значения показателей t и κ , полученные по 50 реализациям выборки, наименьшие значения показателей выделены.

Таблица 3.14

Показатели полезности алгоритмов при $\nu = 8$

Ковариационная структура	ЕСМ-алгоритм		ЕСМЕ-алгоритм		ЕСМ-G-алгоритм		ЕСМЕ-G-алгоритм	
	t	κ	t	κ	t	κ	t	κ
1	0,5442	31,72	0,3308	18,8	0,5498	33,08	0,1826	10,6
2	0,7824	48,36	0,3074	18,5	0,7686	49,44	0,3106	19,6
3	1,0476	48,36	0,4102	18,5	1,0204	48,92	0,2314	10,8
4	1,8106	106,84	0,3204	18,28	1,7482	107,64	0,6596	40,14
5	2,2388	106,84	0,3928	18,28	2,1394	106,34	0,2316	11,14
6	2,2256	106,84	0,3874	18,28	2,1442	107,24	0,286	13,96
7	3,2064	153,02	0,3754	17,54	3,0786	152,96	1,4542	71,9
8	2,9444	141,18	0,3802	17,88	2,8278	141,76	1,2162	60,46
9	2,2898	106,84	0,4008	18,28	2,23	107,76	0,3126	14,7

Результаты, приведенные в таблице, показывают существенное преимущество алгоритма ЕСМЕ перед ЕСМ. Преимущество ЕСМЕ-G проявлялось тех в случаях, когда каждое наблюдение из кластера наблюдений, имеющих определенный ковариационный параметр из вто-

рого подмножества, обязательно имело ковариационный параметр из первого подмножества, присутствующий только у наблюдений данного кластера, причем не обязательно у всех. Сравнение характеристик ЕСМЕ-алгоритма для ковариационной структуры 4 и ЕСМЕ-G-алгоритма для ковариационных структур 5, 6, 9 позволяет сделать вывод, что увеличение количества параметров, оцениваемых по гамма-распределению, уменьшает время работы алгоритма.

Результаты исследования при ошибках, имеющих распределение Стьюдента с параметром формы $\nu=1$ и начальным значением параметра формы, равным 1000, отражены табл. 3.15, аналогичной табл. 3.14. В этом случае ЕСМЕ несколько уступает ЕСМ, однако G-модификации всегда имеют преимущество перед стандартными алгоритмами, причем в ряде случаев это преимущество существенно.

Таблица 3.15

Показатели полезности алгоритмов при $\nu = 1$

Ковариационная структура	ЕСМ-алгоритм		ЕСМЕ-алгоритм		ЕСМ-G-алгоритм		ЕСМЕ-G-алгоритм	
	t	κ	t	κ	t	κ	t	κ
1	1,1528	66,66	1,2254	68,18	0,214	12,74	0,1932	10,82
4	1,2336	69,52	1,2796	69,7	0,2512	14,44	0,22	12,1
5	1,506	69,52	1,5496	69,7	0,3034	14,4	0,2612	11,86
6	1,5136	69,52	1,564	69,7	1,1284	53,36	1,1272	52,12
7	1,5	69,04	1,5422	68,94	1,1498	54,46	1,1596	53,58
9	1,5386	69,52	1,588	69,7	1,2538	58,38	1,2584	57,08

Проведенное исследование позволяет рекомендовать для использования ЕСМЕ-алгоритм и ЕСМЕ-G-алгоритм, в котором параметры, оцениваемые по гамма-распределению, дублируют параметры, оцениваемые по нормальному распределению, причем последние – наиболее гибкие, т. е. каждый из них присутствует у минимального количества наблюдений.

ГЛАВА 4

ОПТИМАЛЬНОЕ ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА

Развитие методов оценивания параметров по неоднородным, негауссовским, зависимым наблюдениям требует перехода к разработке соответствующих методов оптимального планирования эксперимента.

В этой главе дается описание основ теории оптимального планирования эксперимента; рассматриваются постановки задач планирования эксперимента при ММП-оценивании в условиях негауссовских неравномерно распределенных ошибок и результаты исследования зависимости планов от свойств ошибок наблюдений [25]; методы планирования эксперимента при робастном M -оценивании в этих же условиях [63]; планирование эксперимента для модели с разнотипными откликами, в том числе при появлении пропусков в данных [24, 66].

4.1. ТЕОРИЯ ПЛАНИРОВАНИЯ РЕГРЕССИОННОГО ЭКСПЕРИМЕНТА

4.1.1. ЗАДАЧИ ОПТИМАЛЬНОГО ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА

В предыдущих главах при рассмотрении методов оценивания точки x_1, x_2, \dots, x_N , в которых производились наблюдения, считались фиксированными. Теперь предположим, что мы можем по своему ус-

мотрению выбирать эти точки из компактной области \tilde{X} и проводить в них любое число измерений откликов.

Выбирать точки, в которых должны проводиться наблюдения, можно так, чтобы улучшить в каком-то смысле качество оценок. Решением этой задачи занимается теория оптимального планирования эксперимента.

На начальном этапе развития теории планирования регрессионного эксперимента оптимальные планы были нацелены на оптимизацию свойств наилучших линейных оценок для регрессионных (линейных и нелинейных по параметрам) моделей, ориентированных на описание механизма явления. При развитии методов планирования на более сложные модели, например, на динамические стохастические модели и модели в виде обыкновенных дифференциальных уравнений, стали рассматриваться ММП-оценки в условиях нормального распределения ошибок.

Имеются работы по планированию эксперимента при использовании робастных методов оценивания. В [115] в связи с задачей робастного оценивания рассматривается планирование эксперимента при использовании произвольной асимптотически несмещенной оценки параметров регрессии в условиях однородных негауссовских ошибок наблюдений, предлагается использовать классические методы планирования. В [193] рассматривается планирование эксперимента при использовании M -оценок в условиях однородности ошибок наблюдений и неправильной спецификации модели регрессии. В работах [173, 174] планирование эксперимента направлено на увеличение пороговой точки робастных оценок; отмечается, что соответствующие оптимальные планы часто отличаются от классических, направленных на улучшение точности оценок, в связи с этим предлагается компромиссная стратегия.

В последнее время активно развивается планирование эксперимента для нормальной гетероскедастичной регрессии (см., например, [129, 135 – 137, 141]). Планирование в условиях неоднородности, связанной с изменением уравнения регрессии, исследуется в [79].

В классической теории оптимального планирования эксперимента рассматриваются регрессионные модели с количественными откликами, большое число работ посвящено также модели с одномерным бинарным откликом (см., например, [86] и краткие обзоры в [144, 195]),

исследуются также модели с двумерным бинарным откликом [138, 144], мультиномиальные логистические модели [195].

При обработке спланированных экспериментов могут возникнуть серьезные проблемы, если наблюдения в части точек отсутствуют. В связи с этим изучаются свойства известных «стандартных» планов при возникновении пропусков [142, 165, 178, 191]. Рассматривается устойчивость (робастность) планов по ряду критериев, среди которых назовем сохранение оцениваемости параметров и высокой эффективности по сравнению с оптимальным планом, полученным без учета возможности появления пропусков. Изучена робастность ортогональных дробных факторных планов, сбалансированных неполных блочных планов, планов Юдена, латинских квадратов, ротатабельных планов второго порядка и др.

Далее, при рассмотрении основ теории оптимального планирования эксперимента (п. 4.1), будем акцентировать внимание на случае линейной многооткликowej модели (1.37) с известной ковариационной матрицей откликов $\Sigma(x)$, зависящей от точки, в которой производится измерение, а также, оговаривая это особо, на случае линейной гомоскедастичной однооткликowej модели (1.1). При этом для неизвестных параметров уравнений регрессии используются наилучшие линейные оценки.

4.1.2. КРИТЕРИИ ОПТИМАЛЬНОСТИ ПЛАНОВ ЭКСПЕРИМЕНТА

Вначале рассмотрим понятия плана эксперимента и информационной матрицы.

Планом ξ_N называется совокупность величин

$$x_1, r_1; x_2, r_2; \dots; x_n, r_n \quad ,$$

где $\sum_{j=1}^n r_j = N$; x_j – точка, в которой проводится r_j наблюдений; n –

количество различных точек в плане. Совокупность точек x_1, x_2, \dots, x_n называется спектром плана ξ_N .

Информационная матрица Фишера для модели (1.37), соответствующая плану ξ_N , есть $m \times m$ -матрица вида [113]:

$$M(\xi_N) = \sum_{j=1}^n r_j F^T(x_j) \Omega(x_j) F(x_j), \quad (4.1)$$

где $F(x_j)$ – $l \times m$ -матрица регрессоров в точке x_j , i -й строкой которой является вектор $f_i^T(x_j)$; $\Omega(x)$ – функция эффективности, в данном случае равная $\Sigma^{-1}(x)$. Информационная матрица связана с ковариационной матрицей оценок параметров $\hat{\theta}$ соотношением

$$\text{cov}(\hat{\theta}) = M^{-1}(\xi_N). \quad (4.2)$$

Дискретным нормированным планом ε_N называется совокупность величин

$$x_1, p_1; x_2, p_2; \dots; x_n, p_n,$$

где $\sum_{j=1}^n p_j = 1$, $p_j = r_j / N$. Для плана ε_N определяется *нормированная информационная матрица* вида

$$M(\varepsilon_N) = \sum_{j=1}^n p_j F^T(x_j) \Omega(x_j) F(x_j) = N^{-1} M(\xi_N). \quad (4.3)$$

Непрерывным нормированным планом ε называется совокупность величин

$$x_1, p_1; x_2, p_2; \dots; x_n, p_n, \quad (4.4)$$

где $\sum_{j=1}^n p_j = 1$, $p_j \geq 0$.

Непрерывный нормированный план в общем случае может соответствовать вероятностной мере $P_\varepsilon(x)$, заданной на области \tilde{X} . Эта вероятностная мера удовлетворяет условиям

$$\int_{\tilde{X}} dP_\varepsilon(x) = 1, \quad P_\varepsilon(x) \geq 0, \quad \forall x \in \tilde{X}.$$

Нормированная информационная матрица в этом случае имеет вид

$$M(\varepsilon) = \int_{\tilde{X}} F^T(x) \Omega(x) F(x) dP_\varepsilon(x).$$

Матрицу, обратную к соответствующей информационной, будем называть *дисперсионной* и обозначать $D(\varepsilon)$, т. е.

$$D(\varepsilon) = M^{-1}(\varepsilon), \quad D(\varepsilon_N) = M^{-1}(\varepsilon_N).$$

Укажем основные свойства информационной матрицы [113]:

- для любого плана ε информационная матрица $M(\varepsilon)$ симметрична и положительно полуопределена;

- матрица $M(\varepsilon)$ вырожденная, если спектр плана ε содержит меньше чем $\frac{m}{l}$ точек;

- множество матриц $M(\varepsilon)$, соответствующих всем возможным нормированным планам, является выпуклым; если функции $f_i(x)$ и $\Omega(x)$ непрерывные, а область \tilde{X} замкнутая, то множество информационных матриц замкнутое;

- для любого плана ε с заданной информационной матрицей всегда найдется некоторый план $\tilde{\varepsilon}$, спектр которого содержит не более чем $n_0 = \frac{m(m+1)}{2} + 1$ точек и справедливо $M(\varepsilon) = M(\tilde{\varepsilon})$. Таким образом, можно ограничиться рассмотрением планов, спектр которых содержит $n \leq n_0$ точек.

Чтобы улучшить точность оценивания вектора параметров θ , качество плана ε (или ε_N) в соответствии с (4.2) следует оценивать по

значению некоторого функционала Ψ от информационной или дисперсионной матрицы.

План ε^* называется Ψ -оптимальным, если он минимизирует значение некоторого функционала Ψ от информационной матрицы плана, т. е.

$$\varepsilon^* = \arg \min_{\varepsilon} \Psi M(\varepsilon) . \quad (4.5)$$

Приведем основные, наиболее известные, критерии оптимальности.

План ε^* называется D -оптимальным, если он максимизирует определитель информационной матрицы или минимизирует определитель дисперсионной матрицы, т. е.

$$\varepsilon^* = \arg \max_{\varepsilon} |M(\varepsilon)|, \quad \varepsilon^* = \arg \min_{\varepsilon} |D(\varepsilon)| .$$

План ε^* называется A -оптимальным, если его дисперсионная матрица имеет наименьший след, т. е.

$$\varepsilon^* = \arg \min_{\varepsilon} \text{tr} D(\varepsilon) .$$

План ε^* называется E -оптимальным, если он минимизирует (максимизирует) максимальное (минимальное) собственное значение дисперсионной (информационной) матрицы, т. е.

$$\varepsilon^* = \arg \max_{\varepsilon} \min_i \mu_i M(\varepsilon) , \quad \varepsilon^* = \arg \min_{\varepsilon} \max_i \mu_i D(\varepsilon) ,$$

где μ_i – собственное значение матрицы $M(\varepsilon)$ или $D(\varepsilon)$.

План ε^* называется Φ_r -оптимальным, если

$$\varepsilon^* = \arg \min_{\varepsilon} \Phi_r(\varepsilon) = \arg \min_{\varepsilon} \left[\frac{1}{m} \text{tr} D^r(\varepsilon) \right]^{1/r} .$$

Класс функционалов $\Phi_r(\varepsilon)$ включает только что введенные критерии оптимальности: $\Phi_0(\varepsilon)$ — соответствует критерию D -оптимальности, $\Phi_1(\varepsilon)$ — критерию A -оптимальности, $\Phi_\infty(\varepsilon)$ — критерию E -оптимальности. В практических ситуациях для $0 < r < \infty$ можно использовать вместо $\Phi_r(\varepsilon)$ его эквивалент

$$\tilde{\Phi}_r(\varepsilon) = \frac{1}{r} \operatorname{tr} D^r(\varepsilon).$$

План ε^* называется Λ -оптимальным, если

$$\varepsilon^* = \arg \min_{\varepsilon} \sum_{i=1}^m \mu_i D(\varepsilon) - \bar{\mu} D(\varepsilon)^2,$$

где $\bar{\mu}$ — среднее значение собственных чисел.

План ε^* называется MV -оптимальным, если

$$\varepsilon^* = \arg \min_{\varepsilon} \max_i D_{ii}(\varepsilon),$$

где $D_{ii}(\varepsilon)$ — диагональный элемент дисперсионной матрицы.

Рассмотренные критерии относятся к группе критериев точности оценок параметров. Ряд других критериев, например, относящихся к группе критериев точности оценки функции отклика, рассмотрен в [15, 19, 82, 113].

Рассматриваемая до сих пор ситуация характеризовалась тем, что информационная матрица не зависела от оцениваемых параметров, следовательно, не зависел от них и оптимальный план. По этой причине можно построить оптимальный план еще до проведения экспериментов без какой-либо априорной информации о неизвестных параметрах этого уравнения.

Для более сложных моделей информационная матрица может зависеть от вектора оцениваемых параметров модели $\Phi: M(\varepsilon | \Phi)$. Такая ситуация характерна, например, для нелинейных регрессионных моделей [19, 21, 28, 86] и для гетероскедастичной регрессии [129, 141].

Зависимость информационной и дисперсионной матриц от неизвестных параметров приводит к тому, что априорное решение задачи планирования в общем невозможно, так как оптимальный план ε^* зависит от истинного значения параметра Φ , который при планировании неизвестен. Планирование эксперимента в такой ситуации возможно только при наличии некоторой априорной информации о параметрах. В зависимости от уровня априорной информации можно выбрать ту или иную стратегию проведения эксперимента и выбора оптимального плана [19, 21, 28].

Последовательное планирование. Суть стратегии последовательного планирования заключается в том, что вся совокупность экспериментов делится на небольшие «порции», состоящие как минимум из одного наблюдения. Зная оценки $\hat{\Phi}_N$ после проведения N наблюдений, планируют $(N + 1)$ -й эксперимент, оптимизирующий свойства $M_{\varepsilon_{N+1} | \hat{\Phi}_N}$, а затем результаты этого эксперимента используются для уточнения $\hat{\Phi}_N$, т. е. для получения оценок $\hat{\Phi}_{N+1}$, и т. д. Для реализации стратегии последовательного планирования необходимо иметь априорные, хотя бы приблизительные сведения о значении вектора оценок – $\hat{\Phi}_0$.

На каждом этапе планирования в общем случае можно проводить по несколько измерений. При этом рассчитывается сразу группа точек $x_{N+1}, x_{N+2}, x_{N+3}, \dots$, что позволяет уменьшить число обращений к ЭВМ.

Локально оптимальное планирование. В этом случае предполагается, что экспериментатору известны значения оценок $\hat{\Phi}$, достаточно близкие к истинному значению вектора параметров. Допускается также, что в окрестности истинного значения вектора параметров характеристики оптимального плана мало чувствительны к изменениям значений параметра. Таким образом, задача построения локально оптимальных планов принципиально ничем не отличается от классической задачи планирования эксперимента для линейных по параметрам моделей.

Минимаксное планирование. Пусть об искомых параметрах Φ имеются априорные сведения, а именно известна область Ω локализа-

ции истинного значения вектора параметров. Оптимальный план ε^* минимизирует максимум $\Psi \left[M(\varepsilon^* | \Phi) \right]$ относительно Φ , т. е.

$$\max_{\Phi \in \Omega} \Psi \left[M(\varepsilon^* | \Phi) \right] \leq \max_{\Phi \in \Omega} \Psi M(\varepsilon | \Phi)$$

для всех ε .

Решить эту задачу в таком виде чрезвычайно сложно. Используются различные аппроксимации, сводящие исходную непрерывную минимаксную задачу к дискретной минимаксной вида

$$\varepsilon^* = \arg \min_{\varepsilon} \max_{1 \leq i \leq r} \Psi \left[M(\varepsilon | \Phi^{(i)}) \right],$$

где $\Phi^{(i)} \in \Omega_r$, $i = 1, \dots, r$, $\Omega_r \subset \Omega$. Решать задачу в таком виде уже значительно проще. Применяя принцип минимакса, экспериментатор как бы стремится минимизировать максимально возможные потери.

Байесовское планирование. Предположим теперь, что информация о параметрах модели задана в виде априорного распределения $W(\Phi)$ на Ω . Тогда в качестве оптимального плана можно рассматривать план ε^* , который минимизирует средние потери

$$p(W, \varepsilon) = \int_{\Omega} \Psi M(\varepsilon | \Phi) dW(\Phi).$$

В этом случае непрерывное априорное распределение $W(\Phi)$ на Ω также можно аппроксимировать дискретной вероятностной мерой $w(\Phi^{(k)})$ на Ω и решать задачу

$$\varepsilon^* = \arg \min_{\varepsilon} \sum_{k=1}^K w(\Phi^{(k)}) \Psi \left[M(\varepsilon | \Phi^{(k)}) \right].$$

4.1.3. АЛГОРИТМЫ ЧИСЛЕННОГО ПОСТРОЕНИЯ ОПТИМАЛЬНЫХ ПЛАНОВ

В основе численных процедур синтеза оптимальных планов лежит теорема оптимальности для вогнутого (выпуклого) относительно информационной матрицы функционала, впервые сформулированная и доказанная В.В. Федоровым [112]. Запишем ее для случая многооткликковой модели.

Теорема 4.1. Необходимым и достаточным условием Ψ -оптимальности плана ε^* в задаче (4.5) является выполнение равенства

$$\min_{x \in \tilde{X}} \varphi(x, \varepsilon^*) = \text{tr} M(\varepsilon^*) \frac{\partial \Psi[M(\varepsilon^*)]}{\partial M(\varepsilon)},$$

где $\varphi(x, \varepsilon) = \text{tr} M(x) \frac{\partial \Psi M(\varepsilon)}{\partial M(\varepsilon)}$.

В случае решения задачи максимизации функционала Ψ знак \min в формулировке теоремы должен быть заменен на \max .

Часто приходится иметь дело с функционалами, зависящими от дисперсионной матрицы, т. е. с функционалами вида $\Psi D(\varepsilon) = \Psi [M^{-1}(\varepsilon)]$.

Теорема 4.2. Необходимым и достаточным условием Ψ -оптимальности плана ε^* является выполнение равенства

$$\max_{x \in \tilde{X}} \varphi(x, \varepsilon^*) = \text{tr} D(\varepsilon^*) \frac{\partial \Psi[D(\varepsilon^*)]}{\partial D(\varepsilon)},$$

где $\varphi(x, \varepsilon) = \text{tr} M(x) D(\varepsilon) \frac{\partial \Psi D(\varepsilon)}{\partial D(\varepsilon)} D(\varepsilon)$.

С помощью приведенных теорем можно получить необходимые и достаточные условия для конкретных критериев оптимальности, яв-

ляющихся вогнутыми (выпуклыми). Приведем условия оптимальности для ряда критериев:

- для D -оптимальности

$$\max_{x \in \tilde{X}} \operatorname{tr} M(x) M^{-1}(\varepsilon^*) = m,$$

где m – количество оцениваемых параметров модели;

- для A -оптимальности

$$\max_{x \in \tilde{X}} \operatorname{tr} M(x) M^{-2}(\varepsilon^*) = \operatorname{tr} M^{-1}(\varepsilon^*);$$

- для Φ_r -оптимальности

$$\max_{x \in \tilde{X}} \operatorname{tr} M(x) M^{-r-1}(\varepsilon^*) = \operatorname{tr} M^{-r}(\varepsilon^*).$$

Необходимые и достаточные условия для некоторых других критериев оптимальности приведены в [19, 28].

Рассмотрим условия оптимальности в дискретной минимаксной задаче. Назовем план ε^* $\tilde{\Psi}$ -оптимальным, если

$$\varepsilon^* = \arg \min_{\varepsilon} \tilde{\Psi}(\varepsilon) = \arg \min_{\varepsilon} \max_{1 \leq i \leq r} \Psi_i M(\varepsilon),$$

где $\Psi_i M(\varepsilon)$, $i = 1, \dots, r$, – выпуклые по координатам p_1, \dots, p_n , непрерывно дифференцируемые функции, зависящие от информационной матрицы $M(\varepsilon)$.

Для MV -оптимального планирования

$$\Psi_i M(\varepsilon) = D_{ii}(\varepsilon),$$

а для задачи минимаксного планирования при зависимости информационной матрицы от параметров

$$\Psi_i M(\varepsilon) = \Psi \left[M(\varepsilon | \Phi^{(i)}) \right], \quad \Phi^{(i)} \in \Omega_r.$$

Имеет место следующая теорема [26, 28].

Теорема 4.3. План ε^* $\tilde{\Psi}$ -оптимален тогда и только тогда, когда существуют такие $\zeta_i^* \geq 0$, $\sum_{i \in I[M(\varepsilon^*)]} \zeta_i^* = 1$, $i \in I[M(\varepsilon^*)]$, что

$$\min_{x \in \tilde{X}} \sum_{i \in I[M(\varepsilon^*)]} \zeta_i^* \varphi_i(x, \varepsilon^*) = \sum_{i \in I[M(\varepsilon^*)]} \zeta_i^* \operatorname{tr} M(\varepsilon^*) \frac{\partial \Psi_i[M(\varepsilon^*)]}{\partial M(\varepsilon)}.$$

Здесь множество

$$I[M(\varepsilon^*)] = \{i \mid \Psi_i[M(\varepsilon^*)] = \tilde{\Psi}(\varepsilon^*), 1 \leq i \leq r\},$$

$$\varphi_i(x, \varepsilon) = \operatorname{tr} M(x) \frac{\partial \Psi_i M(\varepsilon)}{\partial M(\varepsilon)}.$$

Данная теорема получена в работе [26] применительно к однооткликowym регрессионным моделям. Она остается справедливой и для многооткликовой модели [93].

В применении к критерию MV -оптимальности получаем следующую теорему [26].

Теорема 4.4. План ε^* минимизирует максимальную дисперсию оценок тогда и только тогда, когда существуют такие $\zeta_i^* \geq 0$,

$$\sum_{i \in I[M(\varepsilon^*)]} \zeta_i^* = 1, \quad i \in I[M(\varepsilon^*)], \quad \text{что}$$

$$\min_{x \in \tilde{X}} \sum_{i \in I[M(\varepsilon^*)]} \zeta_i^* \varphi_i(x, \varepsilon^*) = - \max_{1 \leq i \leq m} D_{ii}(\varepsilon^*),$$

где $\varphi_i(x, \varepsilon) = \left[D^{-1}(\varepsilon) M(x) D^{-1}(\varepsilon) \right]_{ii}$.

Обратимся к алгоритмам синтеза оптимальных планов. В соответствии со спецификой задач различают алгоритмы построения непрерывных планов и алгоритмы построения дискретных планов. Рассмотрим алгоритмы построения непрерывных планов.

Основная идея, используемая при построении Ψ -оптимальных планов, заключается в следующем [19]. Строится итерационный процесс, на каждом шаге которого перераспределяется на множестве \tilde{X} вероятностная мера таким образом, чтобы получить увеличение (уменьшение) функционала. При этом для выпуклых функционалов при соблюдении условий оптимальности достигается глобальный экстремум. Для функционалов общего вида может быть найден локальный экстремум.

Алгоритмы построения Ψ -оптимальных планов можно разбить на две группы: алгоритмы последовательного типа (двойственные алгоритмы) и алгоритмы прямого типа, т. е. использующие прямые методы.

Эффективность алгоритмов последовательного типа определяется заменой в вычислительном процессе трудоемких операций (таких как, например, обращение матрицы, вычисление определителя и т. п.) на более простые операции типа вычисления квадратичных форм [19]. Такая замена оправдана, поскольку для реализации соответствующих алгоритмов требуется меньшая емкость оперативной памяти ЭВМ. Кроме того, экстремальная задача большой размерности распадается на ряд экстремальных задач относительно малой размерности.

Приведем основную итерационную процедуру, основанную на двойственном подходе для случая минимизации функционала $\Psi M(\varepsilon)$ [113].

1. Выбираются невырожденный начальный план ε^0 и константы $\gamma > 1$ и $\delta > 0$, полагается $s = 0$.

2. Отыскивается точка x^s :

$$x^s = \arg \min_{x \in \tilde{X}} \varphi(x, \varepsilon^s).$$

3. Составляется план

$$\varepsilon^{s+1} = (1 - \alpha^s) \varepsilon^s + \alpha^s \varepsilon(x^s),$$

где $0 < \alpha^s < 1$; $\varepsilon(x^s)$ – план, состоящий из одной точки x^s .

4. Величина $\Psi[M(\varepsilon^{s+1})]$ сравнивается с величиной $\Psi[M(\varepsilon^s)]$:

а) если

$$\Psi[M(\varepsilon^{s+1})] \geq \Psi[M(\varepsilon^s)],$$

то величина α^s уменьшается в γ раз и повторяются шаги 3 и 4;

б) если имеет место обратное неравенство, то s заменяется на $s + 1$ и осуществляется переход на шаг 2.

Выполнение алгоритма прекращается, когда впервые выполнено условие

$$-\varphi(x^s, \varepsilon^s) + \text{tr} M(\varepsilon^s) \frac{\partial \Psi[M(\varepsilon^s)]}{\partial M(\varepsilon)} \leq \delta. \quad (4.6)$$

В основу алгоритмов прямого типа положены методы нелинейного программирования [19, 28]. При этом решается непосредственно прямая задача минимизации (максимизации) функционала от информационной матрицы. Из алгоритмов, реализующих прямые методы поиска, следует отметить алгоритмы, основанные на методах случайного поиска, методах проектирования градиента, методах возможных направлений и методах штрафных функций. Основное внимание в поисковых процедурах все же обычно уделяют градиентным методам. В работе [20] предложен достаточно эффективный комбинированный алгоритм построения Ψ -оптимальных планов, основанный на прямых градиентных методах. Алгоритмы, сочетающие прямой и двойственный подходы, называют прямыми двойственными или комбинированными.

Рассмотрим основные этапы прямого двойственного алгоритма построения Ψ -оптимальных планов [19].

1. Выбор произвольного невырожденного начального плана ε^s .

2. Градиентный спуск по координатам точек x плана с получением плана $\bar{\varepsilon}^s$.

3. Градиентный спуск по координатам p весов плана с получением плана $\bar{\varepsilon}^s$.

4. Поиск некоторого множества ε_x точек, соответствующих $\min_{x \in \tilde{X}} \varphi(x, \bar{\varepsilon}^s)$.

5. При выполнении для всех точек ε_x достаточных условий оптимальности типа (4.6) – прекращение вычислений, в противном случае – составление плана

$$\varepsilon^{s+1} = 1 \cdot \bar{\varepsilon}^s + 0 \cdot \varepsilon_x$$

и переход к этапу 2.

Если в рассмотренной схеме задействовать только этапы 1 – 3, то получим прямой алгоритм построения Ψ -оптимальных планов.

Опишем прямой двойственный алгоритм построения $\tilde{\Psi}$ -оптимальных планов [26, 28].

1. Выбирается произвольный начальный невырожденный план ε^s .

2. Производится наискорейший спуск по координатам x с получением плана $\bar{\varepsilon}^s$.

3. Выполняется наискорейший спуск по координатам p с получением плана $\bar{\varepsilon}^s$ в стационарной точке и определяются величины ζ_i^s из решения задачи поиска направления наискорейшего спуска для плана $\bar{\varepsilon}^s$.

4. Отыскивается некоторое число точек, соответствующих

$$\min_{x \in \tilde{X}} \sum_i \zeta_i^s \varphi_i(x, \bar{\varepsilon}^s)$$

(обозначим эти точки через ε_x).

5. Составляется план

$$\varepsilon^{s+1} = 1 \cdot \bar{\varepsilon}^s + 0 \cdot \varepsilon_x.$$

б. Производится переход к этапу 2. Расчеты прекращаются, если

$$-\min_{x \in \tilde{X}} \sum_i \zeta_i^s \varphi_i(x, \bar{\varepsilon}^s) + \sum_i \zeta_i^s \operatorname{tr} M(\bar{\varepsilon}^s) \frac{\partial \Psi_i[M(\bar{\varepsilon}^s)]}{\partial M(\varepsilon)} \leq \delta,$$

где δ – заранее заданная точность.

Перейдем к рассмотрению алгоритмов построения дискретных планов. Здесь можно выделить методы трех типов [93]: прямые методы, основанные на применении алгоритмов нелинейной оптимизации, методы замены точек и методы последовательного типа. Ограничимся рассмотрением методов замены точек.

Пусть область планирования \tilde{X} представляет собой фиксированное множество точек x . Число точек x , входящих в \tilde{X} , равно N_x . Необходимо выбрать план ε_N^* с $N < N_x$, доставляющий минимум функционалу $\Psi M(\varepsilon_N)$. Задача может быть решена путем перераспределения дискретной вероятностной меры P среди точек множества \tilde{X} .

Начнем рассмотрение с алгоритмов построения D -оптимальных планов для однооткликовых моделей с единичной функцией эффективности.

Рассмотрим *алгоритм Федорова* [113]. Пусть имеется план ε_N и план $\tilde{\varepsilon}_N$ получен перенесением одного измерения из точки x_{j_-} в точку x_{j_+} . Определители информационных матриц планов связаны соотношением

$$|M(\tilde{\varepsilon}_N)| = |M(\varepsilon_N)| \left[1 + \Delta_{x_{j_-}, x_{j_+}} \right],$$

где

$$\Delta_{x_{j_-}, x_{j_+}} = \frac{1}{N} \left[d_{x_{j_+}, \varepsilon_N} - d_{x_{j_-}, \varepsilon_N} \right] - \frac{1}{N^2} \left[d_{x_{j_+}, \varepsilon_N} d_{x_{j_-}, \varepsilon_N} - d^2_{x_{j_+}, x_{j_-}, \varepsilon_N} \right],$$

$$d(x, \varepsilon) = f^T(x) M^{-1}(\varepsilon) f(x), \quad d(x, \tilde{x}, \varepsilon) = f^T(x) M^{-1}(\varepsilon) f(\tilde{x}).$$

Следовательно, максимального увеличения определителя при обмене двух точек можно достигнуть, если выбирать эти точки, исходя из максимизации величины $\Delta(x_j, x)$. В результате алгоритм состоит из следующих действий.

1. Выбирается невырожденный начальный план ε_N^0 и малая константа $\delta > 0$, полагается $s = 0$.

2. Выбирается пара точек: $x_{j_-}^s$, принадлежащая плану ε_N^s , и $x_{j_+}^s$, принадлежащая области \tilde{X} , по правилу

$$x_{j_-}^s, x_{j_+}^s = \arg \max_{x_{j_-} \in \varepsilon_N^s} \max_{x_{j_+} \in \tilde{X}} \Delta(x_{j_-}, x_{j_+}).$$

3. Величина $\Delta(x_{j_-}^s, x_{j_+}^s)$ сравнивается с δ . Если

$$\Delta(x_{j_-}^s, x_{j_+}^s) < \delta,$$

то вычисления прекращаются, в противном случае осуществляется переход на шаг 4.

4. Точка $x_{j_-}^s$ заменяется в плане на точку $x_{j_+}^s$. В результате получается новый план ε_N^{s+1} . Далее s заменяется на $s + 1$ и осуществляется переход на шаг 2.

Оптимизационная процедура, выполняемая на шаге 2, может оказаться слишком трудоемкой в вычислительном плане, поэтому ее можно заменить на поиск первой пары точек, обмен которых приводит к увеличению значения Δ .

Алгоритм Митчелла [171] строится исходя из следующего. Добавление в план ε_N $(N + 1)$ -й точки x_{N+1} приводит к плану ε_{N+1} с определителем информационной матрицы, равным

$$|\mathbf{M}(\varepsilon_{N+1})| = |\mathbf{M}(\varepsilon_N)| \left(\frac{N}{N+1} \right)^m \left[1 + \frac{1}{N} d(x_{N+1}, \varepsilon_N) \right].$$

В результате максимального увеличения $|\mathbf{M}(\varepsilon_{N+1})|$ можно достигнуть посредством выбора точки x_{N+1} , исходя из максимума величины $d(x_{N+1}, \varepsilon_N)$. С другой стороны, исключение из плана ε_{N+1} некоторой точки x_j приведет к плану $\tilde{\varepsilon}_N$, определитель информационной матрицы которого равен

$$|\mathbf{M}(\tilde{\varepsilon}_N)| = |\mathbf{M}(\varepsilon_{N+1})| \left(\frac{N+1}{N} \right)^m \left[1 - \frac{1}{N+1} d(x_j, \varepsilon_{N+1}) \right].$$

Как следствие минимальное уменьшение определителя будет достигаться при выборе точки x_j , исходя из минимума величины $d(x_j, \varepsilon_{N+1})$. В результате алгоритм состоит из следующих действий.

1. Выбирается невырожденный начальный план ε_N^0 и малая константа $\delta > 0$, $s = 0$.

2. Выбирается точка x^s , принадлежащая области \tilde{X} , по правилу

$$x^s = \arg \max_{x \in \tilde{X}} d(x, \varepsilon_N^s).$$

3. Точка x^s добавляется в план ε_N^s . В результате формируется план ε_{N+1}^s , состоящий из $N+1$ точек.

4. Выбирается точка x_j^s , принадлежащая плану ε_{N+1}^s , по правилу

$$x_j^s = \arg \min_{x \in \varepsilon_{N+1}^s} d(x, \varepsilon_{N+1}^s).$$

5. Точка x_j^s исключается из плана ε_{N+1}^s . В результате формируется план ε_N^{s+1} .

6. Если точка x^s , выбранная на шаге 2, совпадает с точкой x_j^s , выбранной на шаге 4, то вычисления прекращаются, в противном случае s заменяется на $s+1$ и осуществляется переход на шаг 2.

Рассмотрим *комби-градиентный алгоритм замены* [93], решающий задачу

$$\varepsilon_N^* = \arg \max_{\varepsilon_N} \Psi M(\varepsilon_N) .$$

Алгоритм состоит из следующих действий.

1. Выбирается невырожденный начальный план ε_N^0 , полагается $s = 0$.

2. Вычисляются величины

$$\varphi(x_j, \varepsilon_N) = \text{tr} M(x_j) \frac{\partial \Psi M(\varepsilon_N)}{\partial M(\varepsilon_N)}, \quad j = 1, \dots, N_x,$$

для плана ε_N^s , $i = 0$, $\varepsilon_{N,i}^s = \varepsilon_N^s$.

3. Выбирается точка x_{j_+} , принадлежащая области \tilde{X} , по правилу

$$x_{j_+} = \arg \max_{x \in \tilde{X}} \varphi(x, \varepsilon_N^s) .$$

4. Среди точек плана $\varepsilon_{N,i}^s$ выбирается точка x_{j_-} по правилу

$$x_{j_-} = \arg \min_{x \in \varepsilon_{N,i}^s} \varphi(x, \varepsilon_N^s) .$$

5. Точка x_{j_-} заменяется в плане $\varepsilon_{N,i}^s$ на точку x_{j_+} . В результате формируется план $\varepsilon_{N,i+1}^s$.

6. Сравниваются величины $\Psi \left[M \varepsilon_{N,i+1}^s \right]$ и $\Psi \left[M \varepsilon_{N,i}^s \right]$:

а) если

$$\Psi \left[M \varepsilon_{N,i+1}^s \right] > \Psi \left[M \varepsilon_{N,i}^s \right],$$

то i заменяется на $i+1$ и осуществляется переход на шаг 3, при этом точки x_{j_-} и x_{j_+} исключаются из рассмотрения;

б) в противном случае: если $i > 0$, то осуществляется переход на шаг 7; если $i = 0$, то осуществляется полный перебор возможных замен точек плана $\varepsilon_{N,0}^s$ на точки области \tilde{X} , если в результате перебора не получена ни одна удачная замена, то происходит останов, иначе – переход к этапу 7.

7. Полагается

$$\varepsilon_N^{s+1} = \varepsilon_{N,i}^s,$$

s заменяется на $s + 1$ и осуществляется переход на шаг 2.

Данный алгоритм позволяет на каждом шаге (в благоприятном случае) производить по несколько замен точек одновременно, что значительно повышает его быстродействие. При большом числе точек в области планирования \tilde{X} целесообразно ограничивать число замен на шагах 3–6 без пересчета величин $\varphi(x_j, \varepsilon_N)$ и число рассматриваемых точек на этапе полного перебора.

Описанная схема комби-градиентного алгоритма замены применяется также в случае многооткликowych моделей.

4.2. ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА ПРИ ММП-ОЦЕНИВАНИИ ПАРАМЕТРОВ ПО НЕОДНОРОДНЫМ НАБЛЮДЕНИЯМ

4.2.1. ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА ПРИ НЕЗАВИСИМЫХ ОШИБКАХ НАБЛЮДЕНИЙ

Пусть наблюдения за l откликами исследуемой системы описываются моделью вида (1.37). Предположим, что ошибки различных наблюдений независимы и, кроме того, независимы ошибки одновременных измерений откликов. Предположим также, что ошибки i -го

отклика в точке x имеют нулевой параметр сдвига и распределены по некоторому закону с плотностью

$$g_i(z, x) = g_i(z, \lambda_i(x), \nu_i(x)),$$

имеющей мешающие параметры: параметр масштаба $\lambda_i(x)$ и параметр формы $\nu_i(x)$, зависящие от точки факторного пространства.

В результате факторное пространство является неоднородным с точки зрения распределения ошибок наблюдений.

Показатель точности оценивания – информационная матрица. Нормированная информационная матрица, соответствующая непрерывному плану эксперимента ε (4.4), имеет вид

$$M(\varepsilon) = \sum_{i=1}^l \sum_{j=1}^n p_j \omega_i(x_j) f_i(x_j) f_i^T(x_j),$$

где $\omega_i(x) = \omega_i(\lambda_i(x), \nu_i(x))$ – функция эффективности i -го отклика, равная информации Фишера для параметра сдвига

$$\omega_i(x) = \mathbf{E} \left[\frac{\partial}{\partial z} \ln g_i(z, x) \right]^2.$$

Рассмотрим примеры различных вероятностных моделей ошибок наблюдений, которые могут быть полезны в практике регрессионного моделирования, и соответствующие им выражения для функции эффективности. При этом не будем подчеркивать зависимость параметров распределения от x .

Функция эффективности для нормального распределения имеет вид

$$\omega(\sigma^2) = \frac{1}{\sigma^2},$$

где σ^2 – дисперсия.

Функция эффективности для двустороннего экспоненциального распределения с плотностью (1.52) имеет вид

$$\omega(\lambda, \nu) = \frac{\nu^2 \Gamma(2 - 1/\nu)}{\lambda^2 \Gamma(1/\nu)},$$

где $\nu > 1$ (см. также [7]). Зависимость функции $\omega(1, \nu)$ от значения параметра формы приведена на рис. 4.1.

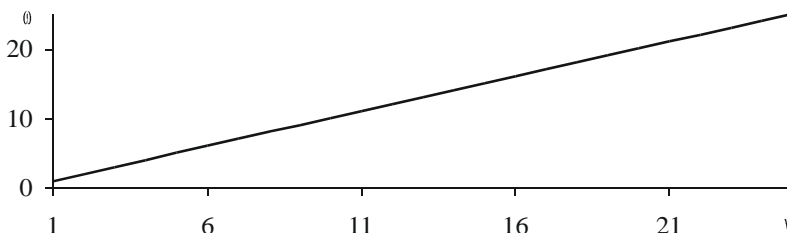


Рис. 4.1. Зависимость функции эффективности от значения параметра формы для двустороннего экспоненциального распределения при $\lambda = 1$

Вид функций эффективности, соответствующих наименее благоприятным плотностям распределения для перечисленных в п. 1.6 классов распределений, приведен в табл. 4.1 (см. также рис. 4.2 – 4.6).

Поскольку функции эффективности $\omega_i(x)$ зависят от функций $\lambda_i(x)$ и $\nu_i(x)$, для вычисления информационной матрицы данные функции должны быть известны.

Однако в [136] отмечается, что предположение об известной функции эффективности может быть нереалистичным в практических ситуациях, особенно если постулированная модель не имеет теоретического обоснования, и опасным, так как небольшая неточность в спецификации функции эффективности может привести к серьезным проблемам при оценивании параметров.

В результате решение задачи планирования эксперимента для линейной по параметрам модели при неоднородных наблюдениях усложняется по сравнению со случаем однородных наблюдений. Здесь могут быть использованы стратегии локально оптимального, последовательного, минимаксного, байесовского планирования [113, 135–137, 141].

Таблица 4.1

Функции эффективности для наименее благоприятных плотностей

Класс распределений	Функция эффективности	Рисунок
Засоренные нормальные	$\frac{1}{\sigma^2} \left\{ 1 + \left[\sqrt{2\pi} c \Phi(c) \right]^{-1} \exp \left[-\frac{c^2}{2} \right] \right\}^{-1}$ [121]	4.2, 4.3
Засоренные лапласовские	$\frac{1-\alpha^2}{\lambda^2}$ [121]	—
Засоренные двусторонние экспоненциальные	$\frac{v(v-1)}{\lambda^2} \frac{\gamma^{1-1/v}, c^v}{\gamma^{1/v}, c^v + e^{-c^v}/c^{v-1}}$	4.4, 4.5
Невырожденные	$\frac{1}{s^2}$ [121]	—
С ограниченной дисперсией	$\frac{1}{\Delta}$ [121]	—
Финитные	$\frac{\pi^2}{\lambda^2}$ [121]	—
Приближенные финитные	$\frac{1}{\lambda^2} \frac{4\beta_1^2}{1+1/(\beta_1 \operatorname{tg} \beta_1)}$ [120]	4.6

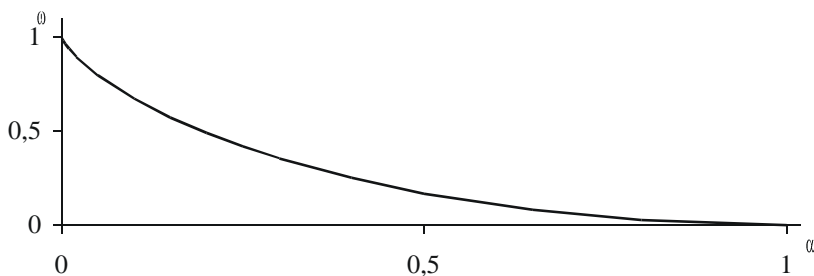


Рис. 4.2. Зависимость функции эффективности от значения уровня засорения для наименее благоприятной плотности в классе засоренных нормальных распределений при $\sigma = 1$

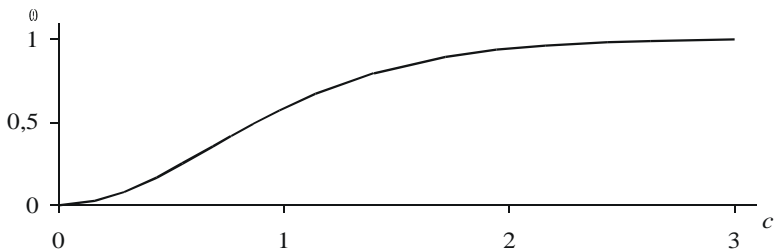


Рис. 4.3. Зависимость функции эффективности от значения параметра c для наименее благоприятной плотности в классе засоренных нормальных распределений при $\sigma = 1$

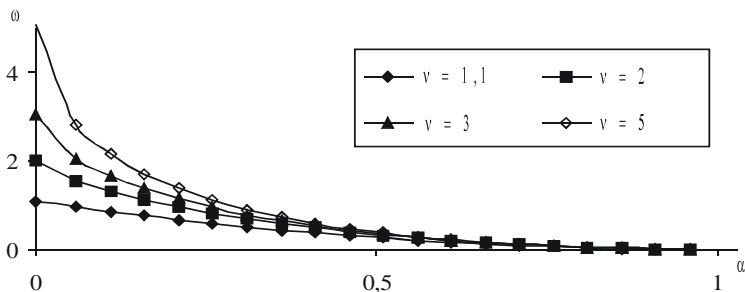


Рис. 4.4. Зависимость функции эффективности от значения уровня засорения для наименее благоприятной плотности в классе засоренных двусторонних экспоненциальных распределений при $\lambda = 1$

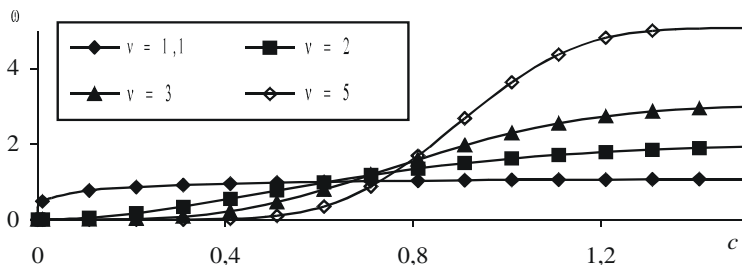


Рис. 4.5. Зависимость функции эффективности от значения параметра c для наименее благоприятной плотности в классе засоренных двусторонних экспоненциальных распределений при $\lambda = 1$

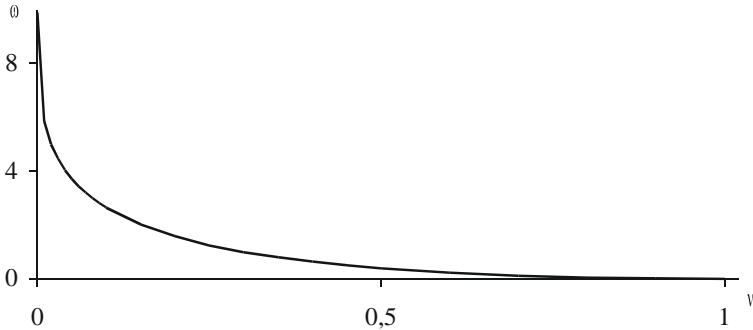


Рис. 4.6. Зависимость функции эффективности от значения параметра ν для наименее благоприятной плотности в классе приближенных финитных распределений при $\lambda = 1$

В том случае, если функция эффективности неизвестна, ее можно параметризовать и оценить [113].

Предположим, что вид функций $\lambda_i(x)$, $\nu_i(x)$ известен с точностью до векторных параметров $\tilde{\lambda}_i$ и $\tilde{\nu}_i$:

$$\lambda_i(x) = \lambda_i x, \tilde{\lambda}_i, \quad \nu_i(x) = \nu_i x, \tilde{\nu}_i .$$

Введенные параметры могут быть оценены по ММП совместно с вектором параметров регрессии θ . Планирование эксперимента в этом случае может быть направлено на повышение точности не только основных параметров θ , но и параметров функции эффективности [137].

В качестве примера рассмотрим *планирование эксперимента на составных областях*. Предположим, что для i -го отклика введены два разбиения области планирования \tilde{X} на области

$$\tilde{X}_j^{\lambda_i}, \quad j = 1, \dots, n_{\lambda_i},$$

и на области

$$\tilde{X}_k^{\nu_i}, \quad k = 1, \dots, n_{\nu_i},$$

в которых являются постоянными функции $\lambda_i(x)$ и $\nu_i(x)$ соответственно.

В этом случае функции $\lambda_i(x)$ и $v_i(x)$ определяются величинами λ_{ij} , $j=1, \dots, n_{\lambda_i}$, и v_{ik} , $k=1, \dots, n_{v_i}$, задающими значения функций в соответствующих областях:

$$\lambda_i(x) = \lambda_{ij}, \quad x \in \tilde{X}_j^{\lambda_i}, \quad v_i(x) = v_{ik}, \quad x \in \tilde{X}_k^{v_i}.$$

В том случае, если разбиения $\tilde{X}_j^{\lambda_i}, j=1, \dots, n_{\lambda_i}$ и $\tilde{X}_k^{v_i}, k=1, \dots, n_{v_i}$ совпадают при всех i (для $n_{\lambda_i} = n_{v_i}$), для оценивания неизвестных параметров можно применять подходы, рассмотренные в гл. 2, или их модификации.

На практике модель составных областей может быть использована, если в разных подобластях области планирования измерение откликов производится приборами, имеющими разные характеристики, либо если с помощью соответствующей функции эффективности можно аппроксимировать неизвестную функцию эффективности.

4.2.2. ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА ПРИ ЗАВИСИМЫХ ОШИБКАХ НАБЛЮДЕНИЙ

Пусть l -мерный вектор ошибок одновременных наблюдений за всеми откликами в некоторой точке x имеет эллиптическое распределение с плотностью вида

$$g(z, x) = g[z, V(x), v(x)] = C v(x) |V(x)|^{-1/2} \tilde{g} q(x, v(x)),$$

где $v(x)$ – значение скалярного параметра формы распределения в точке x ; $V(x)$ – значение псевдоковариационной матрицы в точке x ; $q(x) = z^T V^{-1}(x)z$; $\tilde{g} q, v$ – функция, обладающая свойствами плотности распределения (за исключением условия нормировки). При этом предполагаются известными функции $V(x)$ и $v(x)$, определяющие зависимость псевдоковариационной матрицы и параметра формы от точки x .

В качестве законов распределения ошибок наблюдений на практике часто используют частные случаи эллиптического распределения: многомерное нормальное распределение и многомерное распределение Стьюдента.

Информационная матрица для многооткликовой модели имеет вид (4.3) с матричной функцией эффективности, определяемой в общем случае как

$$\Omega_x = \mathbf{E} \left\{ \frac{\partial \ln g(z, x)}{\partial z} \frac{\partial \ln g(z, x)}{\partial z^T} \right\}.$$

Для произвольного эллиптического распределения при выполнении некоторых условий регулярности функция эффективности имеет вид [156]

$$\Omega_x = \Omega V(x), v(x) = \omega(x) V^{-1}(x), \quad (4.7)$$

где

$$\omega(x) = \omega V(x), v(x) = \frac{1}{l} \mathbf{E} q(x) w^2(x),$$

$$w(x) = - \frac{2}{\tilde{g} q(x), v(x)} \frac{\partial \tilde{g} q, v(x)}{\partial q} \Big|_{q=q(x)}.$$

Для нормального распределения

$$\omega(x) = 1,$$

для распределения Стьюдента [156]

$$\omega(x) = \omega v(x) = \frac{v(x) + l}{v(x) + l + 2}.$$

Вид зависимости функции ωv от значения параметра формы при различных значениях размерности l для распределения Стьюдента приведен на рис. 4.7.

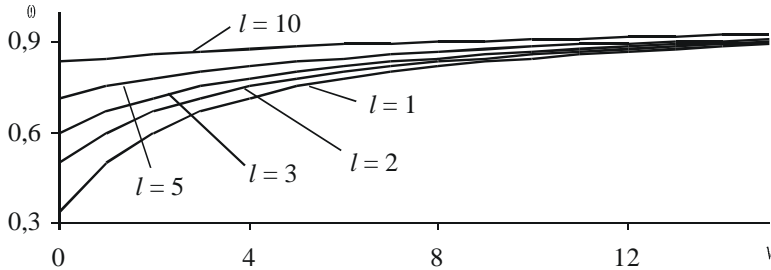


Рис. 4.7. Зависимость функции $\omega(v)$ от значения параметра формы для распределения Стьюдента различной размерности

Как и в случае независимых наблюдений, неполнота информации о функции эффективности усложняет задачу планирования эксперимента. При практическом применении различных стратегий планирования можно параметризовать каким-либо образом функции $V(x)$, $v(x)$:

$$V(x) = V(x, \tilde{V}), \quad v(x) = v(x, \tilde{v}),$$

где \tilde{V} и \tilde{v} – векторы параметров функций $V(x)$ и $v(x)$ соответственно, и оценивать их параметры по наблюдениям. Например, можно ввести разбиения области планирования \tilde{X} на области

$$\tilde{X}_j^V, \quad j = 1, \dots, n_V,$$

в которых является постоянной функция $V(x)$:

$$V(x) = V_j, \quad x \in \tilde{X}_j^V,$$

и на области

$$\tilde{X}_k^v, \quad k = 1, \dots, n_v,$$

в которых является постоянной функция $v(x)$:

$$v(x) = v_k, \quad x \in \tilde{X}_k^v.$$

В результате вновь приходим к задаче планирования на составных областях.

4.2.3. ИССЛЕДОВАНИЕ СВОЙСТВ ПЛАНОВ ЭКСПЕРИМЕНТА

Исследуем зависимость свойств планов эксперимента от мешающих параметров распределения ошибок наблюдений для ряда моделей.

Модель 1. Область планирования \tilde{X} является одномерной и разделена на две области с различным распределением ошибок: первая область \tilde{X}_1 определяется условием $|x| \in (1-s, 1]$, где $0 < s < 1$, вторая область \tilde{X}_2 – условием $|x| \in [0, 1-s]$. Геометрия области планирования отражена на рис. 4.8.

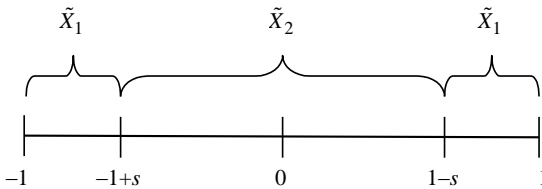


Рис. 4.8. Геометрия области

Рассмотрим многооткликową модель вида

$$y_{ij} = \theta_{0i} + x_j \theta_{1i} + e_{ij}, \quad i = 1, \dots, l, \quad (4.8)$$

где i – номер отклика, j – номер наблюдения. В этом случае размерность вектора параметров равна $m = 2l$.

Предположим, что наблюдения имеют следующие матричные функции эффективности:

- в области \tilde{X}_1

$$\Omega(x) = \Omega_1;$$

- в области \tilde{X}_2

$$\Omega(x) = \Omega_2.$$

Рассмотрим условия предпочтительности и оптимальности планов вида

$$\varepsilon_\alpha = \left\{ -\alpha, \frac{1}{2}; \alpha, \frac{1}{2} \right\}, \quad (4.9)$$

где $0 \leq \alpha \leq 1$ – параметр плана, по ряду критериев оптимальности. Информационная матрица имеет вид

$$M_\alpha = M(\varepsilon_\alpha) = \Omega(\alpha) \otimes \begin{pmatrix} 1 & 0 \\ 0 & \alpha^2 \end{pmatrix}.$$

D-оптимальность. Определитель матрицы M_α равен

$$|M_\alpha| = \left[\det \Omega(\alpha) \alpha^l \right]^2.$$

Сравнивая определители двух планов ε_α и ε_β , получим: планы имеют одинаковое качество – $|M_\alpha| = |M_\beta|$ – при

$$\left| \Omega(\alpha) \Omega^{-1}(\beta) \right| = \left[\frac{\beta}{\alpha} \right]^l; \quad (4.10)$$

если левая часть равенства (4.10) меньше (больше) правой части, то предпочтительнее план ε_β (ε_α).

Обозначим

$$\rho_{D,0} = \left[\frac{|\Omega_1|}{|\Omega_2|} \right]^{1/l}, \quad \rho_{D,1} = \frac{l}{\text{tr} \Omega_2 \Omega_1^{-1}}, \quad \rho_{D,2} = \frac{\text{tr} \Omega_1 \Omega_2^{-1}}{l}.$$

Два плана ε_1 и ε_{1-s} имеют одинаковое качество при выполнении

$$\rho_{D,0} = 1 - s;$$

если $\rho_{D,0} < 1 - s$, то предпочтительнее план ε_{1-s} ; если $\rho_{D,0} > 1 - s$, то предпочтительнее план ε_1 .

Необходимые и достаточные условия оптимальности планов ε_1 , ε_{1-s} соответственно имеют вид

$$\rho_{D,1} \geq \frac{1}{2} \left[1 + (1-s)^2 \right], \quad \rho_{D,2} \leq \frac{2}{\left[1 + 1/(1-s)^2 \right]}.$$

Области предпочтительности планов ε_1 , ε_{1-s} в координатах $(s, \rho_{D,0})$ и области оптимальности планов ε_1 , ε_{1-s} в координатах $(s, \rho_{D,1})$, $(s, \rho_{D,2})$ соответственно представлены на рис. 4.9. План ε_1 оптимален в области выше кривой «Граница 1» и предпочтителен в области выше кривой «Граница 2». План ε_{1-s} оптимален в области ниже кривой «Граница 3» и предпочтителен в области ниже кривой «Граница 2».

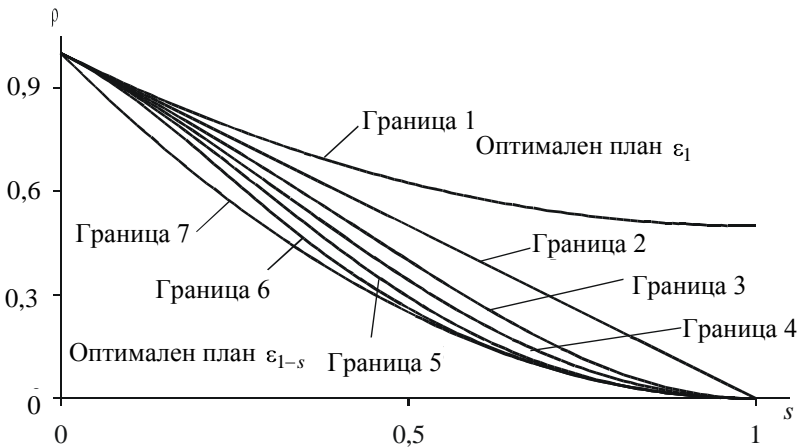


Рис. 4.9. Области оптимальности и предпочтительности планов ε_1 и ε_{1-s}

Φ_r -оптимальность. Для $0 < r < \infty$ справедливо

$$\frac{1}{r} \operatorname{tr} M_{\alpha}^{-r} = \frac{1}{r} \operatorname{tr} \Omega_{\alpha}^{-r} \left[1 + \frac{1}{\alpha^{2r}} \right].$$

Планы ε_α и ε_β имеют одинаковое качество при справедливости

$$\frac{\text{tr} \Omega^{-r}(\alpha)}{\text{tr} \Omega^{-r}(\beta)} = \frac{1+1/\beta^{2r}}{1+1/\alpha^{2r}}; \quad (4.11)$$

если левая часть равенства (4.11) меньше (больше) правой части, то предпочтительнее план ε_α (ε_β).

Обозначим

$$\rho_{r,0} = \left[\frac{\text{tr} \Omega_2^{-r}}{\text{tr} \Omega_1^{-r}} \right]^{1/r}, \quad \rho_{r,1} = \frac{\text{tr} \Omega_1^{-r}}{\text{tr} [\Omega_1^{-r-1} \Omega_2]}, \quad \rho_{r,2} = \frac{\text{tr} [\Omega_2^{-r-1} \Omega_1]}{\text{tr} \Omega_2^{-r}}.$$

Планы ε_1 и ε_{1-s} имеют одинаковое качество при выполнении равенства

$$\rho_{r,0} = \left[\frac{2}{1+1/(1-s)^{2r}} \right]^{1/r}; \quad (4.12)$$

если левая часть равенства (4.12) меньше (больше) правой части, то предпочтительнее план ε_{1-s} (ε_1).

Необходимые и достаточные условия оптимальности плана ε_1 имеют вид

$$\rho_{r,1} \geq \frac{1}{2} [1 + (1-s)^2].$$

План ε_{1-s} оптимальный, если выполняется неравенство

$$\rho_{r,2} \leq \frac{1+1/(1-s)^{2r}}{1+1/(1-s)^{2r+2}}.$$

Области предпочтительности планов ε_1 , ε_{1-s} в координатах $(s, \rho_{r,0})$ и области оптимальности планов ε_1 , ε_{1-s} в координатах $(s, \rho_{r,1})$, $(s, \rho_{r,2})$ соответственно представлены на рис. 4.9 для $r=1,2$. План ε_1 А-оптимален (Φ_1 -оптимален) в области выше кривой «Граница 1» и предпочтителен в области выше кривой «Граница 3», план ε_{1-s}

А-оптимален в области ниже кривой «Граница 5» и предпочтителен в области ниже кривой «Граница 3». План ε_1 Φ_2 -оптимален в области выше кривой «Граница 1» и предпочтителен в области выше кривой «Граница 4», план ε_{1-s} Φ_2 -оптимален в области ниже кривой «Граница 6» и предпочтителен в области ниже кривой «Граница 4».

MV-оптимальность. Для данного критерия справедливо

$$\max_{j=1,\dots,m} [\mathbf{M}_\alpha^{-1}]_{jj} = \frac{1}{\alpha^2} \max_{j=1,\dots,m} [\mathbf{\Omega}_\alpha^{-1}]_{jj}.$$

Планы ε_α и ε_β имеют одинаковое качество при справедливости

$$\frac{\max_{j=1,\dots,m} [\mathbf{\Omega}_\alpha^{-1}]_{jj}}{\max_{j=1,\dots,m} [\mathbf{\Omega}_\beta^{-1}]_{jj}} = \left[\frac{\alpha}{\beta} \right]^2; \quad (4.13)$$

если левая часть равенства (4.13) меньше (больше) правой части, то предпочтительнее план ε_α (ε_β).

Обозначим

$$\rho_{MV,0} = \frac{\max_{j=1,\dots,m} [\mathbf{\Omega}_2^{-1}]_{jj}}{\max_{j=1,\dots,m} [\mathbf{\Omega}_1^{-1}]_{jj}}, \quad \rho_{MV,1} = \frac{[\mathbf{\Omega}_1^{-1}]_{j_1 j_1}}{[\mathbf{\Omega}_1^{-1} \mathbf{\Omega}_2 \mathbf{\Omega}_1^{-1}]_{j_1 j_1}},$$

$$\rho_{MV,2} = \frac{[\mathbf{\Omega}_2^{-1} \mathbf{\Omega}_1 \mathbf{\Omega}_2^{-1}]_{j_2 j_2}}{[\mathbf{\Omega}_2^{-1}]_{j_2 j_2}},$$

где j_1 – номер максимального диагонального элемента матрицы $\mathbf{\Omega}_1^{-1}$ и этот номер единствен; j_2 – номер максимального диагонального элемента матрицы $\mathbf{\Omega}_2^{-1}$ и этот номер единствен.

Планы ε_1 и ε_{1-s} имеют одинаковое качество при выполнении

$$\rho_{MV,0} = (1-s)^2; \quad (4.14)$$

если левая часть равенства (4.14) меньше (больше) правой части, то предпочтительнее план ε_{1-s} (ε_1).

Необходимые и достаточные условия оптимальности плана ε_1 имеют вид

$$\rho_{MV,1} \geq (1-s)^2.$$

План ε_{1-s} оптимальный, если выполняется неравенство

$$\rho_{MV,2} \leq (1-s)^2.$$

Области предпочтительности планов ε_1 , ε_{1-s} в координатах $(s, \rho_{MV,0})$ и области оптимальности планов ε_1 , ε_{1-s} в координатах $(s, \rho_{MV,1})$, $(s, \rho_{MV,2})$ соответственно представлены на рис. 4.9. План ε_1 MV -оптимален и предпочтителен в области выше кривой «Граница 7», план ε_{1-s} MV -оптимален и предпочтителен в области ниже кривой «Граница 7».

Эффективность плана ε_1 в области оптимальности плана ε_{1-s} оценим следующими величинами:

- для D -оптимальности

$$\Delta_D = m \sqrt{\frac{|M_1|}{|M_{1-s}|}} = \frac{\rho_{D,0}}{1-s};$$

- для Φ_r -оптимальности

$$\Delta_r = \left[\frac{\text{tr } M_{1-s}^{-r}}{\text{tr } M_1^{-r}} \right]^{1/r} = \rho_{r,0} \left[\frac{1}{2} + \frac{1}{2(1-s)^{2r}} \right]^{1/r};$$

- для MV -оптимальности

$$\Delta_{MV} = \max_{j=1,\dots,m} [M_{1-s}^{-1}]_{jj} / \max_{j=1,\dots,m} [M_1^{-1}]_{jj} = \rho_{MV,0} / (1-s)^2.$$

На рис. 4.10 представлены зависимости $\Delta_D/\rho_{D,0}$, $\Delta_r/\rho_{r,0}$ (для $r=1$ и $r=2$), $\Delta_{MV}/\rho_{MV,0}$ от s .

Модель 2. Применим полученные для модели 1 результаты к изучению модели с зависимыми откликами, имеющими эллиптическое распределение. Пусть матрицы Ω_i , $i=1,2$, имеют вид

$$\Omega_i = \frac{\omega_i}{\sigma_i} V^{-1},$$

где $\omega_i = \omega(x)$, $x \in \tilde{X}_i$, а псевдоковариационная матрица имеет вид

$$V_i = \sigma_i V.$$

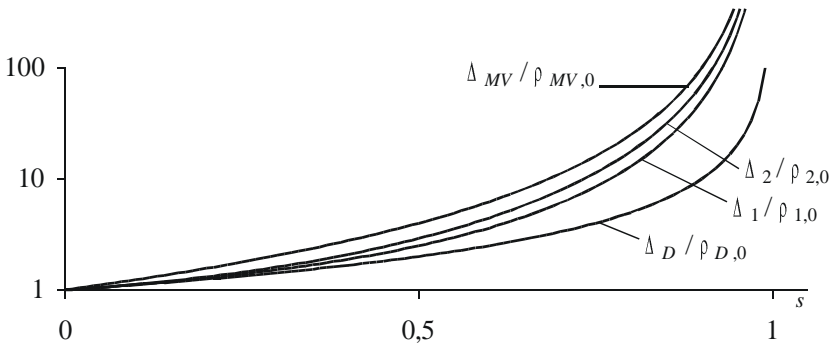


Рис. 4.10. Эффективность плана ε_1 в области оптимальности плана ε_{1-s}

В данном случае справедливо соотношение

$$\rho_0^D = \rho_1^D = \rho_2^D = \rho_0^r = \rho_1^r = \rho_2^r = \rho_0^{MV} = \rho_1^{MV} = \rho_2^{MV} = \rho = \frac{\omega_1 \sigma_2}{\omega_2 \sigma_1}.$$

В результате по рис. 4.9 можно изучать взаимное расположение областей оптимальности и предпочтительности планов ε_1 и $\varepsilon_{1-\delta}$ для всех критериев.

Для величины Δ_D справедливо неравенство

$$\Delta_* < \Delta_D \leq \Delta^*,$$

где $\Delta_* = \rho$, $\Delta^* = \sqrt{\rho(2-\rho)}$. Значения полученных границ в зависимости от величины ρ представлены на рис. 4.11.

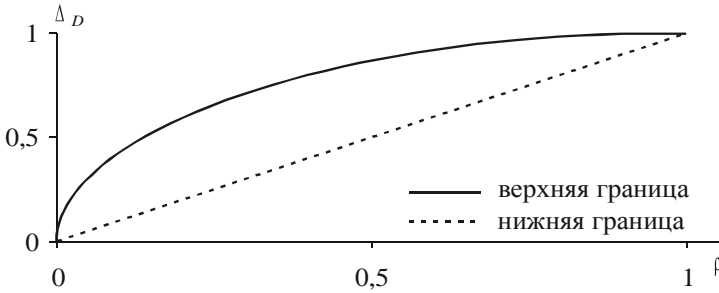


Рис. 4.11. Границы изменения Δ_D в зависимости от ρ

Рассмотрим ряд примеров применения полученных результатов при двумерном распределении Стьюдента.

Пусть параметры в областях однородности равны $\nu_1=1$, $\nu_2=4$, $\sigma_1=\sigma_2=1$. В этом случае $\omega_1=3/5$, $\omega_2=6/8$, $\rho=0,8$. Тогда области предпочтительности планов разграничены значением $s=0,2$, план ε_1 D -оптимален в области $s \geq 1 - \sqrt{0,6} \approx 0,23$, план ε_{1-s} D -оптимален в области $s \leq 1 - \sqrt{2/3} \approx 0,18$. Для границ изменения величины Δ_D имеем: $\Delta_* \approx 0,8$, $\Delta^* \approx 0,98$; например, при $s=0,1$ величина Δ_D приближенно равна 0,89.

Пусть параметры в областях однородности равны $\nu_1=1$, $\nu_2=\infty$, $\sigma_1=1$, $\sigma_2=1,5$ (во второй области распределение является нормальным). В этом случае $\omega_1=3/5$, $\omega_2=2/3$, $\rho=0,9$. Тогда план ε_1 D -оптимален в области $s \geq 1 - \sqrt{0,8} \approx 0,11$, план ε_{1-s} D -оптимален в области $s \leq 1 - \sqrt{9/11} \approx 0,1$. Для границ изменения величины Δ имеем: $\Delta_* \approx 0,9$, $\Delta^* \approx 1$; например, при $s=0,05$ величина Δ приближенно равна 0,95.

Пусть $s = 0,2$. Рассмотрим два плана ε_1 и $\varepsilon_{0,8}$ при условии, что параметры областей однородности равны $v_1 = 1$, $\sigma_1 = \sigma_2 = 1$. Определим, при каких значениях параметра формы второй области v_2 план $\varepsilon_{0,8}$ является D -оптимальным, а при каких – он предпочтительнее плана ε_1 .

В данном случае $\rho = \frac{3}{5\omega_2}$, поэтому план $\varepsilon_{0,8}$ предпочтительнее плана ε_1 при $\omega_2 > 0,75$, что соответствует $v_2 > 4$. План $\varepsilon_{0,8}$ D -оптимален при $\omega_2 \geq \frac{123}{160} \approx 0,77$, что соответствует $v_2 \geq 4\frac{24}{37} \approx 4,65$.

Рассмотрим пример применения полученных результатов при двустороннем экспоненциальном распределении ($l = 1$). Пусть в области \tilde{X}_1 параметр формы v имеет значение 1, что соответствует распределению Лапласа, а в области \tilde{X}_2 – значение 2, что соответствует нормальному распределению, значения параметра масштаба λ при этом предполагаются равными единице в обеих областях. В результате функция эффективности в области \tilde{X}_1 имеет значение 1, в области \tilde{X}_2 – значение 2, значение параметра ρ равно $1/2$.

При сравнении планов ε_1 и ε_{1-s} получим, что при $s < 1/2$ предпочтительнее использовать план ε_{1-s} , а при $s > 1/2$ – план ε_1 . В то же время план ε_{1-s} является D -оптимальным при $s \leq 1 - 1/\sqrt{3} \approx 0,42$, а план ε_1 не будет D -оптимальным ни при каких значениях s . Величина Δ_D лежит в границах $0,5 \leq \Delta < \sqrt{3}/2$; в частности, при $s = 0,1$ $\Delta_D = 5/9$, а при $s = 0,2$ $\Delta_D = 0,625$.

Модель 3. Рассмотрим двухоткликовую модель вида

$$y_{1j} = \theta_0 + x_j\theta_1 + e_{1j}, \quad y_{2j} = \theta_0 + x_j^2\theta_1 + e_{2j},$$

где j – номер наблюдения. Особенность модели – в наличии общего параметра θ_1 , стоящего при различных регрессорах в уравнениях различных откликов; подобные модели с общими параметрами рассматривались в [113, 114].

Предположим, что отклики независимы и имеют константные функции эффективности во всей области планирования $\tilde{X} = -1, 1$:

- первому отклику соответствует

$$\omega_1(x) = \omega_1;$$

- второму отклику соответствует

$$\omega_2(x) = \omega_2.$$

Будем изучать влияние значений функции эффективности откликов на свойства планов, принадлежащих семейству

$$\varepsilon_\delta = \left\{ -1, \frac{\delta}{2}; 0, 1-\delta; 1, \frac{\delta}{2} \right\},$$

где $0 \leq \delta \leq 1$ – параметр плана. Планы эксперимента при этом будем сравнивать с точки зрения критерия D -оптимальности.

Информационная матрица имеет вид

$$M(\varepsilon_\delta) = M_\delta = \begin{bmatrix} \omega_1 + \omega_2 & \omega_2 \delta \\ \omega_2 \delta & \omega_1 + \omega_2 \delta \end{bmatrix},$$

а ее определитель равен

$$|M_\delta| = \omega_1 + \omega_2 - 2\omega_2^2 \delta^2.$$

Обозначим $\rho = \frac{\omega_1}{\omega_2}$, тогда оптимальное значение параметра плана

есть

$$\delta^* = \min \left\{ 1, \frac{(\rho + 1)^2}{2} \right\}.$$

На рис. 4.12 представлена зависимость δ^* от параметра ρ . Так, при $\rho > \sqrt{2} - 1 \approx 0,41$ предпочтительнее план ε_1 ; при $\rho \rightarrow 0$ имеем $\delta \rightarrow 1/2$.

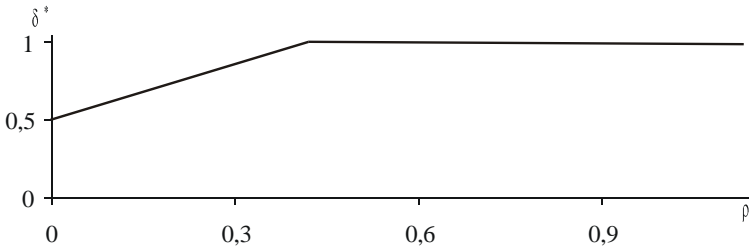


Рис. 4.12. Зависимость δ^* от ρ

Определим условия предпочтительности двух крайних (с точки зрения возможных значений δ^*) планов:

$$\varepsilon_1 = \left\{ -1, \frac{1}{2}; 1, \frac{1}{2} \right\}, \quad \varepsilon_{0,5} = \left\{ -1, \frac{1}{4}; 0, \frac{1}{2}; 1, \frac{1}{4} \right\}.$$

Информационные матрицы будут иметь вид

$$M_1 = \begin{bmatrix} \omega_1 + \omega_2 & \omega_2 \\ \omega_2 & \omega_1 + \omega_2 \end{bmatrix}, \quad M_{0,5} = \begin{bmatrix} \omega_1 + \omega_2 & \frac{\omega_2}{2} \\ \frac{\omega_2}{2} & \frac{\omega_1 + \omega_2}{2} \end{bmatrix},$$

а их определители равны

$$|M_1| = \omega_1^2 + 2\omega_1\omega_2, \quad |M_{0,5}| = \frac{\omega_1 + \omega_2}{2}^2 - \frac{\omega_2^2}{4}.$$

В результате, планы имеют одинаковое качество при $\rho = \gamma$, где $\gamma = \sqrt{6}/2 - 1 \approx 0,22$; если $\rho > \gamma$, то предпочтительнее план ε_1 ; если $\rho < \gamma$, то предпочтительнее план $\varepsilon_{0,5}$. Например, при $\omega_1 = 4$, $\omega_2 = 1$ и при $\omega_1 = 1$, $\omega_2 = 4$ предпочтительным является план ε_1 , а при $\omega_1 = 1$, $\omega_2 = 5$ – план $\varepsilon_{0,5}$.

Модель 4. Рассмотрим модель (4.8) при $l = 2$ с коррелированными ошибками наблюдений, распределенными по нормальному закону. Форма областей однородности области планирования \tilde{X} совпадает с использовавшейся в моделях 1 и 2, при этом в области \tilde{X}_1 ковариационная матрица ошибок является единичной, а в области \tilde{X}_2 диагональные элементы матрицы равны единице, а внедиагональные – γ . Изучим влияние степени коррелированности откликов на свойства планов эксперимента ε_1 и ε_{1-s} из семейства (4.9). Варьируемым параметром является γ , представляющий собой коэффициент корреляции. Планы эксперимента при этом будем сравнивать с точки зрения критерия D -оптимальности.

Матричные функции эффективности имеют следующий вид:

- в области \tilde{X}_1

$$\Omega_1 = I_2;$$

- в области \tilde{X}_2

$$\Omega_2 = \frac{1}{1-\gamma^2} \begin{bmatrix} 1 & -\gamma \\ -\gamma & 1 \end{bmatrix}.$$

Определитель информационной матрицы в этом случае равен

$$|M_\alpha| = \alpha^4$$

при $\alpha \in \tilde{X}_1$ и

$$|M_\alpha| = \frac{\alpha^4}{[1-\gamma^2]^2}$$

при $\alpha \in \tilde{X}_2$.

В результате план ε_{1-s} предпочтительнее плана ε_1 при выполнении условия

$$s < 1 - \sqrt{1 - \gamma^2}.$$

Необходимые и достаточные условия D -оптимальности планов ε_1 и ε_{1-s} имеют вид

$$s \geq 1 - \sqrt{1 - 2\gamma^2}, \quad 2 \left[1 + \frac{1}{(1-s)^2} \right] \leq 4$$

соответственно. Последнее условие не выполняется ни при каких значениях $s > 0$. Таким образом, план ε_{1-s} не является D -оптимальным ни при каких значениях δ и γ .

Полученные в результате анализа области оптимальности и предпочтительности планов изображены на рис. 4.13.

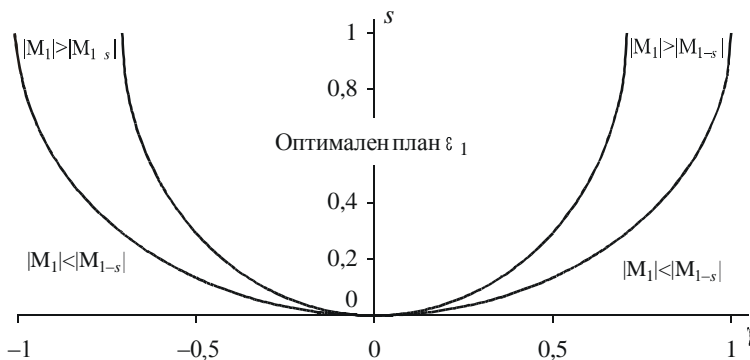


Рис. 4.13. Области оптимальности и предпочтительности планов ε_1 и ε_{1-s} для модели 4

Заметим, что максимальное возможное значение коэффициента корреляции, при котором план ε_1 является D -оптимальным, есть $\gamma_{\max} = 1/\sqrt{2} \approx 0,71$.

Проведенный анализ моделей 1–4 говорит о влиянии на оптимальность планов эксперимента вида зависимости значений мешающих параметров от точки факторного пространства и демонстрирует наличие выигрыша при планировании эксперимента от учета данной зависимости.

4.3. ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА ПРИ РОБАСТНОМ ОЦЕНИВАНИИ ПАРАМЕТРОВ ПО НЕОДНОРОДНЫМ НАБЛЮДЕНИЯМ

4.3.1. ПОКАЗАТЕЛИ КАЧЕСТВА M -ОЦЕНОК

Пусть в каждой j -й точке факторного пространства x_j , $j=1, \dots, n$, производится по r_j наблюдений за l откликами исследуемой системы.

Рассмотрим регрессионную модель вида (1.37). Предположим, что ошибки разных откликов в одном наблюдении имеют совместное распределение с плотностью $g(z, x)$, зависящей от точки x , и нулевой параметр сдвига.

M -оценку $\hat{\theta}$ вектора параметров θ определим путем минимизации функции вида

$$\sum_{j=1}^n \sum_{k=1}^{r_j} \rho(y_{jk} - F(x_j)\theta, x_j), \quad (4.15)$$

где $\rho(z, x)$ – функция потерь, y_{jk} – k -е измерение вектора откликов в j -й точке факторного пространства, матрица $F(x_j)$ определена в формуле (4.1).

Если функция $\rho(z, x)$ дифференцируема, то, приравняв нулю частные производные функции (4.15) по элементам вектора θ в точке $\theta = \hat{\theta}$, получим систему оценочных уравнений

$$-\sum_{j=1}^n \sum_{k=1}^{r_j} F^T(x_j) \psi(y_{jk} - F(x_j)\hat{\theta}, x_j) = 0,$$

где $\psi(z, x) = \frac{\partial \rho(z, x)}{\partial z}$ – векторная оценочная функция, удовлетворяющая условию асимптотической несмещенности оценки $\hat{\theta}$:

$$\mathbf{E}\psi(z, x) = 0.$$

Предположим, что решение системы оценочных уравнений дает состоятельную оценку. При симметричности матрицы $\mathbf{E} \frac{\partial}{\partial z^T} \psi(z, x)$ и некоторых условиях регулярности [9] асимптотический разброс оценки $\hat{\theta}$ можно представить в виде

$$U_0 = \frac{1}{N} D_0, \quad (4.16)$$

где $D_0 = M_1^{-1} M_{2,0} M_1^{-1}$, $M_1 = \sum_{j=1}^n p_j M_1(x_j)$, $M_{2,0} = \sum_{j=1}^n p_j M_{2,0}(x_j)$,

$$p_j = \frac{r_j}{N}, \quad N = \sum_{j=1}^n r_j, \quad M_1(x) = F^T(x) \Omega_1(x) F(x), \quad \Omega_1(x) = \mathbf{E} \frac{\partial}{\partial z^T} \psi(z, x),$$

$$M_{2,0}(x) = F^T(x) \Omega_{2,0}(x) F(x), \quad \Omega_{2,0}(x) = \mathbf{E} \left[\psi(z, x) \psi^T(z, x) \right].$$

В случае независимости ошибок разных откликов предположим, что ошибки i -го отклика имеют плотность распределения $g_i(z_i, x)$ и используется оценочная функция с i -м элементом вида $\psi_i(z_i, x)$. Матрицы $\Omega_1(x)$ и $\Omega_{2,0}(x)$ в этом случае являются диагональными.

В рамках теории робастного оценивания, наряду с показателем (4.16), служащим мерой неточности оценок, рассматриваются различные меры их неустойчивости. Например, в гл. 2 вводится мера качества (2.13) в модели неоднородного байесовского точечного засорения.

Распространим данный подход на случай неоднородных многомерных наблюдений. Рассматривая повторные наблюдения в одной точке как группу однородных наблюдений, вводя положительную весовую функцию $w(x)$ и проводя рассуждения, аналогичные используемым в п. 2.2.1, получим для показателя (2.13) многомерный аналог

$$\frac{1}{N} D_s, \quad \text{где}$$

$$D_s = M_1^{-1} M_{2,s} M_1^{-1}, \quad (4.17)$$

$$M_{2,s} = \sum_{j=1}^n p_j F^T(x_j) \Omega_{2,s}(x_j) F^T(x_j), \quad \Omega_{2,s}(x_j) = w(x_j) \mathbf{E}_s \left[\Psi(\xi_{jk}, x_j) \Psi^T(\xi_{jk}, x_j) \right],$$

\mathbf{E}_s – математическое ожидание по плотности распределения $s(z, x_j)$ засоряющего импульса ξ_{jk} . Как и в п. 2.2.1, полученный показатель качества оценок можно связать с функцией влияния Хампеля.

В случае независимых откликов будем пользоваться одномерной моделью, когда повторные наблюдения одного отклика в одной точке составляют отдельную группу однородных наблюдений и для каждого i -го отклика введены плотность распределения засоряющего импульса $\xi_{jki} - s_i(z_i, x)$ и положительная весовая функция $w_i(x)$. Тогда показатель D_s имеет вид (4.17) с диагональной матрицей $\Omega_{2,s}(x_j)$, i -й диагональный элемент которой имеет вид

$$\left[\Omega_{2,s}(x_j) \right]_{ii} = w_i(x_j) \mathbf{E}_{s_i} \left[\Psi_i^2(\xi_{jki}, x_j) \right],$$

где \mathbf{E}_{s_i} – математическое ожидание по плотности $s_i(z_i, x_j)$.

Отдельно рассмотрим частный случай, когда для многомерной модели имеет место

$$w(x)s(z, x) = g^{1-\beta}(z, x),$$

где β – скалярный параметр, $0 \leq \beta \leq 1$; для одномерной модели –

$$w_i(x)s_i(z_i, x) = g_i^{1-\beta_i}(z_i, x),$$

где $\beta = (\beta_1, \dots, \beta_l)$ – векторный параметр, $0 \leq \beta_i \leq 1$.

Введем матрицу

$$D_\beta = M_1^{-1} M_{2,\beta} M_1^{-1}, \quad (4.18)$$

где $M_{2,\beta} = \sum_{j=1}^n p_j M_{2,\beta}(x_j)$, $M_{2,\beta}(x) = F^T(x) \Omega_{2,\beta}(x) F(x)$, матрица

$\Omega_{2,\beta}(x)$ при зависимых откликах имеет вид

$$\Omega_{2,\beta}(x) = \mathbf{E} \left[\Psi(z, x) \Psi^T(z, x) g^{-\beta}(z, x) \right],$$

а при независимых откликах является диагональной с элементами

$$\left[\Omega_{2,\beta}(x) \right]_{ii} = \mathbf{E} \left[\psi_i^2(z_i, x) g_i^{-\beta_i}(z_i, x) \right].$$

При $\beta = 0$ ($\beta_i = 0$, $i = 1, \dots, l$) приходим к показателю неточности (4.16). При $\beta = 1$ ($\beta_i = 1$, $i = 1, \dots, l$) приходим к максимально неопределенному байесовскому точечному засорению и показателю неустойчивости $\frac{1}{N} D_1$, аналогичному (2.3). При значениях $0 < \beta < 1$ ($0 < \beta_i < 1$, $i = 1, \dots, l$) получаем характеристику, аналогичную приводившей в гл. 2 к обобщенным радикальным оценкам и неявно учитывающую неточность и неустойчивость оценки.

По аналогии с терминологией, используемой в классической теории планирования эксперимента, симметричные матрицы $\Omega_1(x)$, $\Omega_{2,s}(x)$, $\Omega_{2,\beta}(x)$ будем называть матрицами эффективности, а матрицы M_1 , $M_{2,s}$, $M_{2,\beta}$ – квазиинформационными матрицами.

Рассмотрим вид показателей качества для робастных оценок, введенных в гл. 2. Для ЭВ- L_v -оценок, рассмотренных в п. 2.4.2, при $v_i \geq 2$ диагональные элементы матриц эффективности имеют вид

$$\Omega_1(x)_{ii} = \frac{v_i^{\delta_i+2} \Gamma(2-1/v_i)}{2^{\delta_i} \lambda_i^{\delta_i+2} \Gamma^{\delta_i+1}(1/v_i) (\delta_i+1)^{2-1/v_i}}, \quad (4.19)$$

$$\left[\Omega_{2,\beta}(x) \right]_{ii} = \frac{v_i^{2\delta_i+2-\beta_i} \Gamma(2-1/v_i)}{2^{2\delta_i-\beta_i} \lambda_i^{2\delta_i+2-\beta_i} \Gamma^{2\delta_i+1-\beta_i}(1/v_i) (2\delta_i+1-\beta_i)^{2-1/v_i}}, \quad (4.20)$$

где $\lambda_i = \lambda_i(x)$, $v_i = v_i(x)$, $\delta_i = \delta_i(x)$.

Для ЭВ-оценок неоднородной многомерной (l -откликовой) нормальной модели, рассмотренных в п. 2.3.1, матрицы эффективности имеют вид

$$\Omega_1(x) = \frac{(2\pi)^{-cl/2} |\Sigma|^{-c/2}}{(c+1)^{1+l/2}} \Sigma^{-1}, \quad \Omega_{2,\beta}(x) = \frac{(2\pi)^{-l(2c-\beta)/2} |\Sigma|^{-(2c-\beta)/2}}{(2c+1-\beta)^{1+l/2}} \Sigma^{-1},$$

где $\Sigma = \Sigma(x)$, $c = c(x)$.

Параметр β показателя (4.18) может быть постоянным, а может зависеть от точки x : $\beta = \beta(x)$. Например, поскольку обобщенные радикальные оценки в одномерной модели получены путем оптимизации показателя, аналогичного (4.18) при $\beta = \delta$ (см. п. 2.2), может оказаться целесообразным положить при использовании данных оценок

$$\beta(x) = \delta(x).$$

В последнем случае при независимых откликах и некоторых ограничениях на гладкость функций $g_i(z_j, x)$ и $\psi_i(z_j, x)$ справедливо

$$\Omega_1(x) = \Omega_{2,\delta}(x),$$

откуда следует

$$M_1 = M_{2,\delta}$$

и

$$D_\delta = M_1^{-1} = M_{2,\delta}^{-1}.$$

Заметим, что при однородных наблюдениях в случае $\delta = 0$, соответствующем ММП-оцениванию, последнее равенство отражает известную связь асимптотической дисперсионной матрицы и информационной матрицы Фишера. Для ЭВ-оценок при многомерном нормальном распределении справедливы те же соотношения.

4.3.2. ПОСТАНОВКА ЗАДАЧИ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА

Целью планирования эксперимента при робастном оценивании будем считать оптимизацию некоторого скалярного показателя качества оценки за счет выбора точек факторного пространства, в которых должны проводиться наблюдения. При планировании эксперимента в классической ситуации, т. е. при планировании для наилучших линейных оценок или ММП-оценок, обращают внимание на их точность, поэтому показателем качества здесь является некоторый функционал от матрицы D_0 . Планируя эксперимент, можно повысить точность и

робастных оценок, при этом показателем качества вновь будет некоторый функционал от матрицы D_0 . С другой стороны, качество робастных оценок характеризуется не только точностью, но и устойчивостью, поэтому при планировании эксперимента в данной ситуации может быть целесообразно выбрать показателем качества функционал от матрицы (4.17) или ее частного случая – матрицы (4.18).

Матрицы (4.17) и (4.18) зависят от мешающих параметров распределения ошибок, которые на практике обычно неизвестны. По этой причине здесь необходимо использовать сложные стратегии, такие как локально оптимальное, последовательное, байесовское, минимаксное планирование эксперимента.

Рассмотрим локально оптимальное планирование эксперимента, при котором функции зависимости параметров распределения ошибок от точки факторного пространства заданы исследователем.

При планировании эксперимента в классической ситуации часто используется информационная матрица. В нашем случае (по аналогии с классическим), наряду с матрицей (4.17), будем использовать и матрицу

$$M_s = D_s^{-1} = M_1 M_{2,s}^{-1} M_1,$$

причем в дальнейшем индекс s будем опускать, т. е. $\Omega_2(x) = \Omega_{2,s}(x)$, $M_2 = M_{2,s}$, $M = M_s$, $D = D_s$. Аналогично будем использовать матрицу $M_\beta = D_\beta^{-1}$.

Заметим, что при равенстве функций эффективности $\Omega_1(x)$ и $\Omega_2(x)$, как в одной из рассмотренных выше ситуаций, вид матрицы M совпадает с видом информационной матрицы при ММП-оценивании по неоднородным наблюдениям (см. п. 4.2), особенность заключается лишь в ином способе вычисления матрицы эффективности. По этой причине в указанной ситуации может использоваться классическая теория планирования эксперимента при соответствующей «матрице эффективности».

В дальнейшем изучим случай неравных функций эффективности, причем получаемые результаты будут справедливы при любых функциях эффективности, обеспечивающих выполнение приведенных ниже свойств квазиинформационных матриц.

Дискретному плану эксперимента ε_N сопоставим значение

$$D(\varepsilon_N) = M^{-1}(\varepsilon_N)$$

матрицы (4.18), при этом фигурирующие в квазиинформационных матрицах величины p_j , $j=1, \dots, n$, являются весами точек x_j .

Непрерывному плану эксперимента ε со спектром, имеющим n точек, сопоставим значение

$$D(\varepsilon) = M^{-1}(\varepsilon),$$

также определяемое формулой (4.18), при этом фигурирующие в квазиинформационных матрицах величины p_j , $j=1, \dots, n$, удовлетворяют только ограничениям $p_1 + \dots + p_n = 1$, $p_j \geq 0$.

Если непрерывному плану ε соответствует вероятностная мера $P_\varepsilon(x)$, заданная на области \tilde{X} , то квазиинформационные матрицы определяются формулой

$$M_k(\varepsilon) = \int_{\tilde{X}} F^T(x) \Omega_k(x) F(x) dP_\varepsilon(x), \quad k=1, 2.$$

Докажем теорему о свойствах введенных матриц.

Теорема 4.5. Справедливы следующие высказывания.

1. Для любого плана ε матрицы $M_1(\varepsilon)$ и $M_2(\varepsilon)$ симметричны, матрица $M_2(\varepsilon)$ положительно полуопределенная.

2. Если матрицы $M_1(\varepsilon)$ и $M_2(\varepsilon)$ невырождены, то матрицы $M(\varepsilon)$ и $D(\varepsilon)$ существуют, матрицы $M_2(\varepsilon)$, $M(\varepsilon)$ и $D(\varepsilon)$ – положительно определены, в случае конечного числа точек в плане справедливо неравенство $n \geq \frac{m}{l}$.

3. Множества квазиинформационных матриц $M_k(\varepsilon)$, $\varepsilon \in \Xi$, где $k=1, 2$, Ξ – множество всех непрерывных планов, выпуклы, а если компактны множества $M_k[\varepsilon(x)]$, $x \in \tilde{X}$, где $k=1, 2$, $\varepsilon(x)$ – одноточечный план, то компактны и множества квазиинформационных матриц.

4. Для любого плана ε с заданными квазиинформационными матрицами $M_1(\varepsilon)$ и $M_2(\varepsilon)$ всегда найдется некоторый план $\tilde{\varepsilon}$, содержащий не более чем $n_0 = m(m+1) + 1$ точек, такой, что $M_k(\varepsilon) = M_k(\tilde{\varepsilon})$, $k = 1, 2$.

5. Для любых двух планов ε и $\tilde{\varepsilon}$, для которых существуют матрицы $M(\varepsilon)$ и $M(\tilde{\varepsilon})$, справедливо $(1 - \alpha)M(\varepsilon) + \alpha M(\tilde{\varepsilon}) \geq M$ $(1 - \alpha)\varepsilon + \alpha\tilde{\varepsilon}$, где неравенство понимается в смысле положительной полуопределенности разности матриц, стоящих слева и справа.

Доказательство. 1. Симметричность матриц $M_1(\varepsilon)$ и $M_2(\varepsilon)$ следует из их определения и симметричности матриц эффективности. Докажем положительную полуопределенность матрицы $M_2(\varepsilon)$ для случая зависимых откликов, для независимых откликов доказательство проводится аналогично. Рассмотрим функцию $\chi(x) = a^T M_2(x)a$ для любого вектора a размерности m .

$$\begin{aligned} \chi(x) &= a^T M_2(x)a = a^T F^T(x)w(x)\mathbf{E}_s \left[\psi(z, x)\psi^T(z, x) \right] F(x)a = \\ &= w(x) \int_Z a^T F^T(x)\psi(z, x)\psi^T(z, x)F(x)as(z, x)dz = \\ &= w(x) \int_Z \left[a^T F^T(x)\psi(z, x) \right]^2 s(z, x)dz \geq 0, \end{aligned}$$

где Z – область определения ошибок. В результате матрица $M_2(x)$ положительно полуопределена для любого x . Положительная полуопределенность матрицы $M_2(\varepsilon)$ следует тогда из ее представления в виде выпуклой линейной комбинации положительно полуопределенных матриц $M_2(x)$.

2. Из невырожденности матриц $M_1(\varepsilon)$ и $M_2(\varepsilon)$ следует существование матриц $M_1^{-1}(\varepsilon)$ и $M_2^{-1}(\varepsilon)$, в результате существуют матрицы $M(\varepsilon)$ и $D(\varepsilon)$. Невырожденная матрица $M_2(\varepsilon)$ является положительно определенной (см. свойство 1), в результате по свойствам положительно определенных матриц матрицы $M(\varepsilon)$ и $D(\varepsilon)$ являются положи-

тельно определенными. Если число точек в плане конечно и равно n , то матрицы $M_1(\varepsilon)$ и $M_2(\varepsilon)$ представляют собой суммы n матриц, каждая из которых имеет ранг не более l . Если $n < \frac{m}{l}$, то ранги матриц $M_1(\varepsilon)$ и $M_2(\varepsilon)$ меньше m и они вырождены.

3. Покажем выпуклость множеств квазиинформационных матриц. Составим план $\varepsilon = (1-\alpha)\varepsilon_1 + \alpha\varepsilon_2$. Тогда

$$\begin{aligned} (1-\alpha)M_k(\varepsilon_1) + \alpha M_k(\varepsilon_2) &= (1-\alpha) \int_{\tilde{X}} F^T(x)\Omega_k(x)F(x)dP_{\varepsilon_1}(x) + \\ &+ \alpha \int_{\tilde{X}} F^T(x)\Omega_k(x)F(x)dP_{\varepsilon_2}(x) = \int_{\tilde{X}} F^T(x)\Omega_k(x)F(x)dP_{\varepsilon}(x) = \\ &= M_k(1-\alpha)\varepsilon_1 + \alpha\varepsilon_2, \end{aligned}$$

где $k=1,2$. Множества квазиинформационных матриц $M_k(\varepsilon), \varepsilon \in \Xi$, $k=1,2$, являются выпуклыми оболочками соответствующих множеств $M_k[\varepsilon(x)], x \in \tilde{X}$, $k=1,2$, и если последние множества компактны, то компактны и первые.

4. Поскольку квазиинформационные матрицы симметричны, каждая из них полностью описывается $\frac{m(m+1)}{2}$ элементами, т. е. квазиинформационным матрицам $M_1(\varepsilon)$ и $M_2(\varepsilon)$ можно сопоставить вектор размерности $m(m+1)$. Из свойства 3 следует, что множество векторов, определяющих квазиинформационные матрицы $M_1(\varepsilon)$ и $M_2(\varepsilon)$, является выпуклой оболочкой множества, которое состоит из векторов, соответствующих матрицам $M_1 \varepsilon(x)$ и $M_2 \varepsilon(x)$. Отсюда и из теоремы Каратеодори вытекает справедливость доказываемого.

Свойство 4 является важным с практической точки зрения. Поскольку матрицы $M(\varepsilon)$ и $D(\varepsilon)$ определяются только через матрицы $M_1(\varepsilon)$ и $M_2(\varepsilon)$, можно ограничиться рассмотрением планов, содержащих $n \leq n_0$ точек.

5. Составим линейную комбинацию матриц $M(\varepsilon)$, $M(\tilde{\varepsilon})$ и воспользуемся неравенством из [113, теорема 1.1.13]:

$$\begin{aligned} (1-\alpha)M(\varepsilon) + \alpha M(\tilde{\varepsilon}) &= (1-\alpha)M_1 M_2^{-1} M_1 + \alpha \tilde{M}_1 \tilde{M}_2^{-1} \tilde{M}_1 \geq \\ &\geq \left[(1-\alpha)M_1 + \alpha \tilde{M}_1 \right] \left[(1-\alpha)M_2 + \alpha \tilde{M}_2 \right]^{-1} \left[(1-\alpha)M_1 + \alpha \tilde{M}_1 \right] = \\ &= M_1 (1-\alpha)\varepsilon + \alpha \tilde{\varepsilon} M_2^{-1} (1-\alpha)\varepsilon + \alpha \tilde{\varepsilon} M_1 (1-\alpha)\varepsilon + \alpha \tilde{\varepsilon} = \\ &= M (1-\alpha)\varepsilon + \alpha \tilde{\varepsilon} , \end{aligned}$$

где M_k и \tilde{M}_k , $k=1,2$, – квазиинформационные матрицы планов ε и $\tilde{\varepsilon}$ соответственно.

Сформулируем задачу оптимального планирования эксперимента при робастном оценивании. План ε^* назовем Ψ -оптимальным, если он минимизирует значение некоторого функционала Ψ от матрицы $M(\varepsilon)$:

$$\varepsilon^* = \arg \min_{\varepsilon \in \Xi} \Psi M(\varepsilon) . \quad (4.21)$$

Отдельно сформулируем задачу для функционалов типа дискретного минимакса. Назовем план ε^* $\tilde{\Psi}$ -оптимальным, если

$$\varepsilon^* = \arg \min_{\varepsilon \in \Xi} \tilde{\Psi} \varepsilon = \arg \min_{\varepsilon \in \Xi} \max_{1 \leq i \leq r} \Psi_i M(\varepsilon) . \quad (4.22)$$

Заметим, что во множестве Ξ содержатся планы, имеющие вырожденные квазиинформационные матрицы. Будем считать, что такому плану соответствует бесконечно большое значение функционала. План, которому соответствуют невырожденные квазиинформационные матрицы, назовем невырожденным. В дальнейшем будем предполагать, что оптимальный план существует и является невырожденным.

Наряду с оптимизацией функционала от матрицы (4.18), учитывающей неточность и неустойчивость оценки неявно, может быть поставлена явная задача одновременной оптимизации функционалов от показателей неточности D_0 и неустойчивости D_1 . Данная задача от-

носится к многокритериальным, для получения Парето-оптимальных решений можно использовать различные принципы компромисса [28]. Так, может быть поставлена задача минимизации по плану ε значения функционала

$$\bar{\Psi}[M_1(\varepsilon), M_{2,0}(\varepsilon), M_{2,1}(\varepsilon)] = (1-b) \frac{\Psi_0[D_0(\varepsilon)]}{\Psi_0[D_0(\varepsilon_0^*)]} + b \frac{\Psi_1[D_1(\varepsilon)]}{\Psi_1[D_1(\varepsilon_1^*)]},$$

где Ψ_0, Ψ_1 – некоторые, в общем случае различные, функционалы; $0 < b < 1$; $\varepsilon_0^* = \arg \min_{\varepsilon \in \Xi} \Psi_0[D_0(\varepsilon)]$, $\varepsilon_1^* = \arg \min_{\varepsilon \in \Xi} \Psi_1[D_1(\varepsilon)]$. Заметим, что похожий принцип положен в основу построения робастных обобщенных компромиссных оценок в гл. 2. Другой компромиссный функционал может иметь тип дискретного минимакса:

$$\varepsilon^* = \arg \min_{\varepsilon \in \Xi} \widehat{\Psi}[M_1(\varepsilon), M_{2,0}(\varepsilon), M_{2,1}(\varepsilon)] = \arg \min_{\varepsilon \in \Xi} \max_{i=1,2} \frac{\Psi_i[D_i(\varepsilon)]}{\Psi_i[D_i(\varepsilon_i^*)]}.$$

В дальнейшем ограничимся изучением условий Ψ - и $\tilde{\Psi}$ -оптимальности; свойства и условия оптимальности планов в случае использования функционалов $\bar{\Psi}$ и $\widehat{\Psi}$ можно получить аналогичным образом.

4.3.3. УСЛОВИЯ ОПТИМАЛЬНОСТИ ПЛАНОВ ЭКСПЕРИМЕНТА

Предположим, что функции $\Psi[M]$ и $\Psi_i[M]$, $i=1, \dots, r$, непрерывно дифференцируемы.

Рассмотрим условия Ψ -оптимальности планов эксперимента. Введем обозначения

$$\begin{aligned} \varphi(\tilde{\varepsilon}, \varepsilon) = & \operatorname{tr} M_1 \tilde{\varepsilon} \left[M_2^{-1} M_1 \frac{\partial \Psi}{\partial M} + \frac{\partial \Psi}{\partial M} M_1 M_2^{-1} \right] - \\ & - \operatorname{tr} M_2 \tilde{\varepsilon} M_2^{-1} M_1 \frac{\partial \Psi}{\partial M} M_1 M_2^{-1}, \end{aligned}$$

$$\varphi(x, \varepsilon) = \varphi(\varepsilon(x), \varepsilon),$$

где $M = M(\varepsilon)$, $M_k = M_k(\varepsilon)$, $k = 1, 2$, план ε – невырожденный. Величина $\varphi(x_j, \varepsilon)$ является производной функционала $\Psi M(\varepsilon)$ по весу j -й точки.

Теорема 4.6. Для любого невырожденного плана ε справедливо тождество

$$\int_{\tilde{X}} \varphi(x, \varepsilon) dP_\varepsilon(x) = \text{tr} M(\varepsilon) \frac{\partial \Psi M(\varepsilon)}{\partial M(\varepsilon)}. \quad (4.23)$$

Доказательство. Распишем левую часть равенства (4.23):

$$\begin{aligned} \int_{\tilde{X}} \varphi(x, \varepsilon) dP_\varepsilon(x) &= \text{tr} \int_{\tilde{X}} M_1 x dP_\varepsilon(x) \left[M_2^{-1} M_1 \frac{\partial \Psi M}{\partial M} + \frac{\partial \Psi M}{\partial M} M_1 M_2^{-1} \right] - \\ &\quad - \text{tr} \int_{\tilde{X}} M_2 x dP_\varepsilon(x) M_2^{-1} M_1 \frac{\partial \Psi M}{\partial M} M_1 M_2^{-1}, \end{aligned}$$

где $M = M(\varepsilon)$, $M_k = M_k(\varepsilon)$, $k = 1, 2$. Значение первого интеграла в последнем выражении равно $M_1(\varepsilon)$, а второго – $M_2(\varepsilon)$. Переставляя сомножители аргументов операции взятия следа и сокращая подобные члены, получаем доказываемый результат.

Теорема 4.7. Для того чтобы план ε^* был решением задачи (4.21), необходимо, а в случае выпуклости функционала Ψ на множестве Ξ и достаточно, чтобы выполнялось равенство

$$\min_{x \in \tilde{X}} \varphi(x, \varepsilon^*) = \text{tr} M(\varepsilon^*) \frac{\partial \Psi[M(\varepsilon^*)]}{\partial M(\varepsilon)}; \quad (4.24)$$

в точках оптимального плана выполняется

$$\varphi(x_j, \varepsilon^*) = \text{tr} M(\varepsilon^*) \frac{\partial \Psi[M(\varepsilon^*)]}{\partial M(\varepsilon)}, \quad j = 1, \dots, n. \quad (4.25)$$

Доказательство. Необходимость. Пусть план ε^* Ψ -оптимален. Рассмотрим план

$$\tilde{\varepsilon} = (1 - \alpha)\varepsilon^* + \alpha\varepsilon_0, \quad 0 \leq \alpha \leq 1.$$

Так как план ε^* минимизирует значение функционала Ψ , то для любого плана ε_0

$$\Psi[M(\tilde{\varepsilon})] \geq \Psi[M(\varepsilon^*)],$$

следовательно,

$$\left. \frac{\partial \Psi[M(\tilde{\varepsilon})]}{\partial \alpha} \right|_{\alpha=0+} = \varphi(\varepsilon_0, \varepsilon^*) - \text{tr} M(\varepsilon^*) \frac{\partial \Psi[M(\varepsilon^*)]}{\partial M(\varepsilon)} \geq 0.$$

Выберем $\varepsilon_0 = \varepsilon(x)$, тогда

$$\varphi(x, \varepsilon^*) - \text{tr} M(\varepsilon^*) \frac{\partial \Psi[M(\varepsilon^*)]}{\partial M(\varepsilon)} \geq 0.$$

Из этого неравенства и формулы (4.23), примененной для $\varepsilon = \varepsilon^*$, следует справедливость (4.24) и (4.25).

Достаточность. Предположим, что план ε^* , для которого выполняется (4.24), не является Ψ -оптимальным, т. е.

$$\Psi[M(\varepsilon^*)] > \Psi[M(\varepsilon^0)] = \min_{\varepsilon} \Psi[M(\varepsilon)].$$

Рассмотрим выпуклую линейную комбинацию планов ε^* и ε^0 — план $\tilde{\varepsilon} = (1 - \alpha)\varepsilon^* + \alpha\varepsilon^0$. Из выпуклости Ψ и определения ε^0 следует, что

$$\Psi[M(\tilde{\varepsilon})] \leq (1 - \alpha)\Psi[M(\varepsilon^*)] + \alpha\Psi[M(\varepsilon^0)],$$

$$\left. \frac{\partial \Psi[M(\tilde{\varepsilon})]}{\partial \alpha} \right|_{\alpha=0+} \leq \Psi[M(\varepsilon^0)] - \Psi[M(\varepsilon^*)] < 0,$$

тогда

$$\left. \frac{\partial \Psi[M(\tilde{\varepsilon})]}{\partial \alpha} \right|_{\alpha=0+} = \int_{\tilde{X}} \varphi(x, \varepsilon^*) dP_{\varepsilon^0}(x) - \text{tr} M(\varepsilon^*) \frac{\partial \Psi[M(\varepsilon^*)]}{\partial M(\varepsilon)} < 0. \quad (4.26)$$

С другой стороны, из (4.24) вытекает

$$\int_{\tilde{X}} \varphi(x, \varepsilon^*) dP_{\varepsilon^0}(x) - \operatorname{tr} M(\varepsilon^*) \frac{\partial \Psi[M(\varepsilon^*)]}{\partial M(\varepsilon)} \geq 0. \quad (4.27)$$

Сравнивая (4.26) и (4.27), приходим к противоречию, которое доказывает достаточность условия (4.24).

Для функционалов, зависящих от матрицы D , т. е. для функционалов вида $\Psi[D(\varepsilon)] = \Psi[M^{-1}(\varepsilon)]$, имеет место следующее утверждение.

Теорема 4.8. Для того чтобы план ε^* был решением задачи (4.21), необходимо, а в случае выпуклости функционала Ψ на множестве Ξ и достаточно, чтобы выполнялось равенство

$$\max_{x \in \tilde{X}} \tilde{\varphi}(x, \varepsilon^*) = \operatorname{tr} D(\varepsilon^*) \frac{\partial \Psi[D(\varepsilon^*)]}{\partial D(\varepsilon)},$$

где

$$\begin{aligned} & \tilde{\varphi}(x, \varepsilon) = \\ & = \operatorname{tr} \left\{ M_1(x) \left[D \frac{\partial \Psi}{\partial D} M_1^{-1} + M_1^{-1} \frac{\partial \Psi}{\partial D} D \right] - M_2(x) M_1^{-1} \frac{\partial \Psi}{\partial D} M_1^{-1} \right\}, \end{aligned}$$

$D = D(\varepsilon)$, $M_k = M_k(\varepsilon)$, $k = 1, 2$; в точках оптимального плана выполняется

$$\tilde{\varphi}(x_j, \varepsilon^*) = \operatorname{tr} D(\varepsilon^*) \frac{\partial \Psi[D(\varepsilon^*)]}{\partial D(\varepsilon)}, \quad j = 1, \dots, n.$$

Для иллюстрации результатов в табл. 4.2 приведены условия оптимальности планов для ряда критериев. В таблице μ_{\max} и q обозначают некрратное максимальное собственное число матрицы D и соответствующий ему нормированный собственный вектор; μ_i , $i = 1, \dots, m$ – собственные числа матрицы D , $D = D(\varepsilon^*)$, $M_k = M_k(\varepsilon^*)$, $k = 1, 2$.

Таблица 4.2

Условия оптимальности планов эксперимента

Критерий	Функционал	Условия оптимальности
D	$\ln D $	$\max_{x \in \tilde{X}} \operatorname{tr} 2M_1(x)M_1^{-1} - M_2(x)M_2^{-1} = m$
A	$\operatorname{tr} D$	$\max_{x \in \tilde{X}} \operatorname{tr} 2M_1(x)M_1^{-1}D - M_2(x)M_1^{-2} = \operatorname{tr} D$
E	μ_{\max}	$\max_{x \in \tilde{X}} \operatorname{tr} 2M_1(x)M_1^{-1}qq^T\mu_{\max} -$ $- M_2(x)M_1^{-1}qq^TM_1^{-1} = \mu_{\max}$
Φ_r	$r^{-1} \operatorname{tr} D^r$	$\max_{x \in \tilde{X}} \operatorname{tr} 2M_1(x)M_1^{-1}D^r -$ $- M_2(x)M_1^{-1}D^{r-1}M_1^{-1} = \operatorname{tr} D^r$
Λ	$\sum_{i=1}^m (\mu_i - \bar{\mu})^2$ $\bar{\mu} = \frac{1}{m} \sum_{i=1}^m \mu_i$	$\max_{x \in \tilde{X}} \operatorname{tr} \left\{ 4M_1(x) \left[M_1^{-1}D^2 - M_1^{-1}D \frac{\operatorname{tr} D}{m} \right] - \right.$ $\left. - 2M_2(x) \left[M_1^{-1}DM_1^{-1} - M_1^{-2} \frac{\operatorname{tr} D}{m} \right] \right\} =$ $= 2 \operatorname{tr} D^2 - \frac{2}{m} \operatorname{tr}^2 D$

Заметим, что соответствующие функционалы не выпуклы на множестве Ξ , поэтому приведенные условия являются лишь необходимыми.

Рассмотрим условия $\tilde{\Psi}$ -оптимальности планов эксперимента.

Теорема 4.9. Для того чтобы план ε^* , содержащий n точек, был решением задачи (4.22), необходимо, а в случае выпуклости Ψ_i по координатам $p = p_1, \dots, p_n$ и достаточно, чтобы существовали такие величины $\zeta_i^* \geq 0$, $i \in I[M(\varepsilon^*)]$, $\sum_{i \in I[M(\varepsilon^*)]} \zeta_i^* = 1$, что

$$\min_{x \in \tilde{X}} \sum_{i \in I[M(\varepsilon^*)]} \zeta_i^* \varphi_i(x, \varepsilon^*) = \sum_{i \in I[M(\varepsilon^*)]} \zeta_i^* \operatorname{tr} M(\varepsilon^*) \frac{\partial \Psi_i[M(\varepsilon^*)]}{\partial M(\varepsilon)};$$

в точках оптимального плана выполняется

$$\sum_{i \in I[M(\varepsilon^*)]} \zeta_i^* \varphi_i(x_j, \varepsilon^*) = \sum_{i \in I[M(\varepsilon^*)]} \zeta_i^* \operatorname{tr} M(\varepsilon^*) \frac{\partial \Psi_i[M(\varepsilon^*)]}{\partial M(\varepsilon)}, \quad j = 1, \dots, n,$$

где $I[M(\varepsilon^*)] = \{i \mid \Psi_i[M(\varepsilon^*)] = \tilde{\Psi}[\varepsilon^*], 1 \leq i \leq r\}$, функция $\varphi_i(x, \varepsilon)$ определяется аналогично $\varphi(x, \varepsilon)$ с заменой $\frac{\partial \Psi[M(\varepsilon)]}{\partial M(\varepsilon)}$ на $\frac{\partial \Psi_i[M(\varepsilon)]}{\partial M(\varepsilon)}$.

Доказательство теоремы 4.9 аналогично доказательству теоремы 4.3.

Применим данный результат к критерию MV -оптимальности, для которого $\Psi_i[M(\varepsilon)] = D_{ii}(\varepsilon)$, $i = 1, \dots, m$. В этом случае

$$\varphi_i(x, \varepsilon^*) = \left[M_1^{-1} M_2(x) M_1^{-1} - 2 M_1^{-1} M_1(x) D \right]_{ii},$$

$$\sum_{i \in I[M(\varepsilon^*)]} \zeta_i^* \operatorname{tr} M(\varepsilon^*) \frac{\partial \Psi_i[M(\varepsilon^*)]}{\partial M(\varepsilon)} = - \max_{1 \leq i \leq m} D_{ii}(\varepsilon^*),$$

где $M_1 = M_1(\varepsilon^*)$, $D = D(\varepsilon^*)$. Заметим, что получаемые условия являются лишь необходимыми.

4.3.4. АЛГОРИТМЫ ПОСТРОЕНИЯ ОПТИМАЛЬНЫХ ПЛАНОВ

Изученные свойства планов эксперимента, квазиинформационных матриц и полученные условия оптимальности позволяют применять в рассматриваемом случае алгоритмы построения оптимальных планов, используемые в классическом случае.

Обоснование возможности использования двойственного подхода к получению алгоритмов построения непрерывных Ψ -оптимальных планов основывается на доказательстве теоремы 4.7 и может быть проведено аналогично классическому случаю (см., например, [34, 113]). Обоснование применения в нашем случае алгоритма, комбинирующего прямой и двойственный подходы, может быть проведено аналогично используемому в [19].

Рассмотрим алгоритмы построения дискретных планов, доставляющих минимум функционалу $\Psi[M(\varepsilon_N)]$. Среди алгоритмов различного типа рассмотрим алгоритмы, осуществляющие улучшение плана путем замены точек и описанные в п. 4.1.3. В частности, здесь может быть применен комби-градиентный алгоритм замены.

Алгоритмы Федорова и Митчелла основаны на применении эффективных формул пересчета значений определителя при изменении плана. Поскольку квазиинформационные матрицы совпадают по структуре с информационной матрицей в классическом случае, для пересчета их определителей можно использовать формулы, однотипные с приведенными в п. 4.1.3. Рассмотрим получаемые модификации.

На итерациях алгоритма Федорова осуществляется замена в текущем плане ε_N точки $x_{j_-}^*$ на точку $x_{j_+}^*$ из области \tilde{X} , которые определяются по правилу

$$x_{j_-}^*, x_{j_+}^* = \arg \max_{x_{j_-}, x_{j_+}} \frac{\Delta_1^2(x_{j_-}, x_{j_+})}{\Delta_2(x_{j_-}, x_{j_+})},$$

$$\begin{aligned} \Delta_k(x_{j_-}, x_{j_+}) = & 1 + \frac{1}{N} \left[\Omega_k(x_{j_+}) d_k(x_{j_+}) - \Omega_k(x_{j_-}) d_k(x_{j_-}) \right] - \\ & - \frac{1}{N^2} \Omega_k(x_{j_+}) \Omega_k(x_{j_-}) \left[d_k(x_{j_+}) d_k(x_{j_-}) - d_k^2(x_{j_+}, x_{j_-}) \right], \end{aligned}$$

$$d_k(x) = F(x) M_k^{-1}(\varepsilon_N) F^T(x), \quad d_k(x, \tilde{x}) = F(x) M_k^{-1}(\varepsilon_N) F^T(\tilde{x}), \quad k=1, 2.$$

На итерациях алгоритма Митчелла чередуются шаги добавления в план ε_N точки $x_{j_+}^*$ из области \tilde{X} , в результате чего получается план ε_{N+1} , и удаления из плана ε_{N+1} точки $x_{j_-}^*$. Указанные точки определяются по следующим правилам:

$$x_{j_+}^* = \arg \max_x \frac{\Delta_1^2(x)}{\Delta_2(x)}, \quad \Delta_k(x) = 1 + \frac{1}{N} \Omega_k(x) d_k(x);$$

$$x_{j-}^* = \arg \max_x \frac{\tilde{\Delta}_1^2(x)}{\tilde{\Delta}_2(x)}, \quad \tilde{\Delta}_k(x) = 1 - \frac{1}{N+1} \Omega_k(x) F(x) M_k^{-1}(\varepsilon_{N+1}) F^T(x).$$

Для построения $\tilde{\Psi}$ -оптимальных планов эксперимента также можно применять модификации разработанных для классического случая алгоритмов [26–28, 93].

4.3.5. ИССЛЕДОВАНИЕ СВОЙСТВ ПЛАНОВ ЭКСПЕРИМЕНТА

Изучим зависимость свойств планов эксперимента от параметров показателя качества оценки (4.18) и оценочной функции. Планы эксперимента будем сравнивать с точки зрения критерия D -оптимальности.

Рассмотрим область планирования и уравнение регрессии, использовавшиеся в моделях 1, 2 в п. 4.2.3, при этом рассмотрим однооткликový случай.

Модель 1. Предположим, что ошибки наблюдений имеют одномерное нормальное распределение с дисперсией σ_1^2 в области \tilde{X}_1 и σ_2^2 в области \tilde{X}_2 . Используются ЭВ-оценки параметров с постоянным значением параметра оценочной функции.

Сравним между собой два плана ε_1 и ε_{1-s} из семейства (4.9). Обозначим

$$\rho = \frac{\sigma_2}{\sigma_1},$$

тогда планы имеют одинаковое качество при

$$s = 1 - \rho^{2+\beta}. \quad (4.28)$$

Если левая часть равенства (4.28) меньше (больше) правой, то предпочтительнее план ε_{1-s} (ε_1).

Зафиксируем значения β_1 и β_2 параметра β , тогда при $\beta_1 < \beta_2$ и $\rho < 1$ в области

$$1 - \rho^{2+\beta_1} < s < 1 - \rho^{2+\beta_2} \quad (4.29)$$

по критерию $|D_{\beta_1}|$ предпочтительнее план ε_1 , а по критерию $|D_{\beta_2}|$ — план ε_{1-s} .

Полученные области предпочтительности планов изображены на рис. 4.14. Заметим, что при увеличении значения β область предпочтительности плана ε_{1-s} увеличивается. В области, расположенной между линиями, выполняется неравенство (4.29).

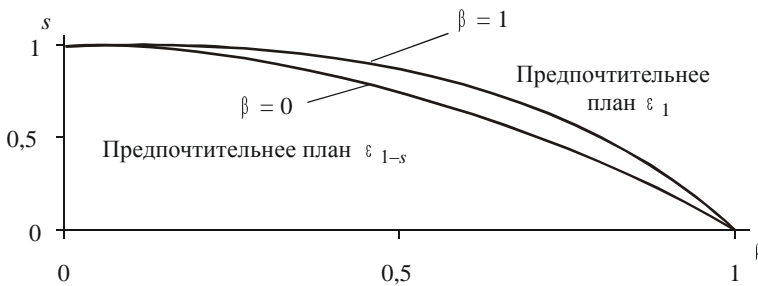


Рис. 4.14. Области предпочтительности планов ε_1 и ε_{1-s}

Степень предпочтительности плана ε относительно $\tilde{\varepsilon}$ будем измерять величиной

$$\tau_{\beta}(\varepsilon, \tilde{\varepsilon}) = \sqrt{\frac{|D_{\beta}(\tilde{\varepsilon})|}{|D_{\beta}(\varepsilon)|}}.$$

В нашем случае

$$\tau_{\beta}(\varepsilon_1, \varepsilon_{1-s}) = \frac{\rho^{2+\beta}}{1-s},$$

следовательно, при $\rho < 1$ данная величина уменьшается с увеличением значения β . Так, при $\rho = 0,5$, $s = 0,5$ имеем $\tau_0(\varepsilon_1, \varepsilon_{0,5}) = 0,5$, $\tau_1(\varepsilon_1, \varepsilon_{0,5}) = 0,25$.

Модель 2. Предположим, что ошибки наблюдений имеют двустороннее экспоненциальное распределение с единичным параметром

масштаба и параметрами формы $v_1 = 4$ в области \tilde{X}_1 и $v_2 = 5$ в области \tilde{X}_2 , $s = 0,2$. Используются ЭВ-оценки со значениями параметра δ_1 и δ_2 в областях \tilde{X}_1 и \tilde{X}_2 соответственно.

В табл. 4.3 приведены непрерывные планы ε^* , оптимальные по показателю $|D_0|$, при разных значениях параметров оценочной функции и показатель $\tau = \tau_0(\tilde{\varepsilon}, \varepsilon^*)$, определяющий эффективность в соответствующих случаях плана $\tilde{\varepsilon}$, оптимального для ММП-оценки (ЭВ-оценки при $\delta_1 = \delta_2 = 0$). Оптимальные планы найдены численно.

Таблица 4.3

Оптимальные планы

Параметр δ_2	Параметр δ_1		
	0	0,5	1
0	$\{-0,8, 0,5; 1, 0,5\}$ $\tau = 1$	$\{-0,8, 0,5; 0,8, 0,5\}$ $\tau = 0,907503$	$\{-0,8, 0,5; 0,8, 0,5\}$ $\tau = 0,782141$
0,5	$\{-1, 0,5; 1, 0,5\}$ $\tau = 0,905214$	$\{-1, 0,105426; -0,8, 0,395603; 1, 0,498971\}$ $\tau = 0,999738$	$\{-0,8, 0,5; 0,8, 0,5\}$ $\tau = 0,869606$
1	$\{-1, 0,5; 1, 0,5\}$ $\tau = 0,776862$	$\{-1, 0,5; 1, 0,5\}$ $\tau = 0,861196$	$\{-1, 0,28526; -0,8, 0,216257; 1, 0,498483\}$ $\tau = 0,998182$

Результаты показывают, что в случае $\delta_1 = \delta_2$ можно практически без ущерба использовать при ЭВ-оценивании планы, оптимальные для ММП-оценок. Такое свойство можно считать полезным, поскольку исследователи часто находят и сравнивают между собой несколько оценок, варьируя значения параметров оценочной функции. С другой стороны, в случае $\delta_1 \neq \delta_2$ потери от использования плана $\tilde{\varepsilon}$ при ЭВ-оценивании доходят до 22 %.

4.4. ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА ПРИ ОЦЕНИВАНИИ ПАРАМЕТРОВ МОДЕЛИ С РАЗНОТИПНЫМИ ОТКЛИКАМИ И ПРОПУСКАМИ В ДАННЫХ

В п. 3.2 рассматривалась задача оценивания параметров регрессионной модели с количественными и качественными откликами при наличии пропусков в данных. Перейдем к решению задачи оптимального планирования эксперимента для данной модели.

4.4.1. ПОСТАНОВКА ЗАДАЧИ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА

Будем описывать n_z -мерный вектор качественных откликов логистической регрессией вида (3.19), (3.20) с функцией $\eta_j(x_t, \vartheta)$ вида

$$\eta_j(x_t, \vartheta) = G_j(x_t)\vartheta, \quad j = 1, \dots, n_c,$$

где $G_j(x_t)$ – вектор-строка значений регрессоров модели качественных откликов.

Для n_y -мерного вектора Y_t в j -й ячейке используем распределение с параметром сдвига (уравнением регрессии) $F_j(x_t)\theta$ и плотностью $g_j(Y_t | x_t, \theta, \phi)$, где $F_j(x_t)$ – матрица значений регрессоров модели количественных откликов, θ – вектор параметров уравнения регрессии количественных откликов, ϕ – вектор прочих (мешающих) параметров модели количественных откликов.

Например, распределение вектора Y_t в j -й ячейке может быть эллиптическим с плотностью вида

$$g_j(Y_t | x_t, \theta, \phi) = |V_j|^{-1/2} \tilde{g}[q_j(x_t), v_j],$$

где V_j – псевдоковариационная матрица; $\tilde{g}(q, v)$ – функция, обладающая свойствами плотности распределения (за исключением усло-

вия нормировки); $q_j(x_t) = Y_t - F_j(x_t)\theta^T V_j^{-1} Y_t - F_j(x_t)\theta$; v_j – скалярный параметр формы; при этом не все значения V_j , v_j , $j = 1, \dots, n_c$, могут быть различны.

Введенная модель имеет множество параметров θ , ϕ и ϑ , которые объединим в вектор Φ . Пусть проведено N наблюдений. Для оценивания неизвестного значения вектора параметров Φ используется ММП.

Качество ММП-оценок вектора параметров Φ при некоторых условиях регулярности характеризуется (нормированной) информационной матрицей

$$M_{\Phi}(X | \Phi) = \frac{1}{N} \sum_{t=1}^N \mathbf{E} \left\{ \frac{\partial \ln g(Y_t, Z_t | x_t, \Phi)}{\partial \Phi} \frac{\partial \ln g(Y_t, Z_t | x_t, \Phi)}{\partial \Phi^T} \right\},$$

где X – совокупность значений точек факторного пространства, в которых проводились измерения; $g(Y_t, Z_t | x_t, \Phi)$ – плотность распределения откликов [9]; нижний индекс в обозначении информационной матрицы указывает множество параметров, которому она соответствует, а параметр Φ в ее аргументе – зависимость в общем случае от значения Φ .

Задачу оптимального планирования эксперимента формулируем следующим образом:

$$\varepsilon^* = \arg \min_{\varepsilon} \Psi M_{\Phi}(\varepsilon | \Phi),$$

где ε^* – оптимальный непрерывный план эксперимента; Ψ – некоторый функционал от информационной матрицы; $M_{\Phi}(\varepsilon | \Phi)$ – информационная матрица параметров θ , ϕ и ϑ , соответствующая непрерывному плану ε (4.4).

Информационная матрица $M_{\Phi}(\varepsilon | \Phi)$ имеет вид

$$M_{\Phi}(\varepsilon | \Phi) = \sum_{t=1}^n p_t M_{\Phi}(x_t | \Phi),$$

где $M_{\Phi}(x_t | \Phi)$ – информационная матрица однократного наблюдения в точке x_t , являющаяся блочно-диагональной с блоками $M_{\theta, \phi}(x_t | \Phi)$ и $M_{\vartheta}(x_t | \Phi)$.

Рассмотрим вид информационной матрицы при условии, что матрица $M_{\theta, \phi} x_t | \Phi$ является блочно-диагональной с блоками $M_{\theta} x_t | \phi, \vartheta$ и $M_{\phi} x_t | \phi, \vartheta$. Данное условие часто справедливо для используемых на практике распределений, например, для многомерных нормального и студентовского.

Матрица $M_{\theta} x_t | \phi, \vartheta$ имеет вид

$$M_{\theta} x_t | \phi, \vartheta = \sum_{j=1}^{n_c} \pi_j(x_t, \vartheta) F_j^T(x_t) \Omega_j F_j(x_t),$$

где $\Omega_j = \Omega_j(\phi) = \mathbf{E}_j \left\{ \frac{\partial \ln g_j(\zeta_{tj} | \phi)}{\partial \zeta_{tj}} \frac{\partial \ln g_j(\zeta_{tj} | \phi)}{\partial \zeta_{tj}^T} \right\}$ – матричная функция эффективности, \mathbf{E}_j – математическое ожидание по плотности $g_j Y_t | x_t, \theta, \phi$, $g_j(\zeta_{tj} | \phi) = g_j Y_t | x_t, \theta, \phi$, $\zeta_{tj} = Y_t - F_j(x_t)\theta$.

Вид функции эффективности для эллиптического распределения в рассматриваемых условиях аналогичен (4.7).

В том случае, если уравнения регрессии количественных откликов параметризованы независимо, т. е. если параметр сдвига j -й ячейки можно представить в виде $\tilde{F}_j(x_t)\theta_j$, информационная матрица $M_{\theta} x_t | \phi, \vartheta$ является блочно-диагональной с блоками $M_{\theta_j} x_t | \phi, \vartheta$ вида

$$M_{\theta_j} x_t | \phi, \vartheta = \pi_j(x_t, \vartheta) \tilde{F}_j^T(x_t) \Omega_j(x_t) \tilde{F}_j(x_t).$$

Информационная матрица $M_{\phi} x_t | \phi, \vartheta$ вектора параметров ϕ представима в виде

$$M_{\phi} x_t | \phi, \vartheta = \sum_{j=1}^{n_c} \pi_j(x_t, \vartheta) M_{\phi}^j \phi,$$

где $M_{\phi}^j \phi = \mathbf{E}_j \left\{ \frac{\partial \ln g_j(\zeta_{tj} | \phi)}{\partial \phi} \frac{\partial \ln g_j(\zeta_{tj} | \phi)}{\partial \phi^T} \right\}$. Вид матрицы $M_{\phi}^j \phi$

для многомерного распределения Стьюдента приведен в [156].

Информационная матрица $M_{\vartheta} \ x_t | \vartheta$ вектора параметров ϑ имеет вид

$$M_{\vartheta} \ x_t | \vartheta = \sum_{j=1}^{n_c} \pi_j(x_t, \vartheta) G_j^T(x_t) \left[G_j(x_t) - \sum_{s=1}^{n_c} \pi_s(x_t, \vartheta) G_s(x_t) \right].$$

В том случае, если уравнения $\eta_j \ x_t, \vartheta$ параметризованы независимо

$$\eta_j \ x_t, \vartheta = \bar{G}_j(x_t) \vartheta_j, \quad j=1, \dots, n_c - 1, \quad \eta_{n_c} \ x_t, \vartheta = 0,$$

где $\bar{G}_j(x_t)$ – вектор-строка значений регрессоров модели качественных откликов для j -й ячейки, информационная матрица $M_{\vartheta} \ x | \vartheta$ имеет блочную структуру

$$M_{\vartheta} \ x_t | \vartheta = M_{\vartheta_s \vartheta_u} \ x_t | \vartheta \quad \begin{matrix} n_c - 1 \\ s, u = 1 \end{matrix},$$

где

$$M_{\vartheta_s \vartheta_u} \ x_t | \vartheta = -\pi_s(x_t, \vartheta) \pi_u(x_t, \vartheta) \bar{G}_s^T(x_t) \bar{G}_u(x_t), \quad s \neq u,$$

$$M_{\vartheta_s \vartheta_s} \ x_t | \vartheta = \pi_s(x_t, \vartheta) [1 - \pi_s(x_t, \vartheta)] \bar{G}_s^T(x_t) \bar{G}_s(x_t).$$

При независимости качественных откликов используем представление

$$\pi_j(x_t, \vartheta) = \prod_{s=1}^{n_z} \pi_{si^s_j}(x_t, \vartheta), \quad j=1, \dots, n_c,$$

где $\pi_{si^s_j}(x_t, \vartheta) = \mathbf{P} \ Z_{ts} = i_j^s$, i_j^s – уровень s -го качественного отклика в j -й ячейке, Z_{ts} – s -й элемент вектора Z_t . Параметризуем $\pi_{su}(x_t, \vartheta)$ в виде логистической регрессии

$$\pi_{su}(x_t, \vartheta) = \frac{\exp[\tilde{G}_{su}(x_t)\vartheta]}{\sum_{w=1}^{I_s} \exp[\tilde{G}_{sw}(x_t)\vartheta]}, \quad u=1, \dots, I_s,$$

где $\tilde{G}_{su}(x_t)$ – вектор-строка значений регрессоров модели для u -го уровня s -го отклика. Функция $\pi_j(x_t, \vartheta)$ в этом случае представима логистической регрессией (3.20) с регрессорами

$$G_j(x_t) = \sum_{s=1}^{n_z} \tilde{G}_{si_j}(x_t).$$

Информационная матрица в данном случае имеет вид

$$M_{\vartheta} \quad x_t | \vartheta = \sum_{s=1}^{n_z} \sum_{u=1}^{I_s} \pi_{su}(x_t, \vartheta) \tilde{G}_{su}^T(x_t) \left[\tilde{G}_{su}(x_t) - \sum_{w=1}^{I_s} \pi_{sw}(x_t, \vartheta) \tilde{G}_{sw}(x_t) \right].$$

Если справедливо

$$\pi_{su}(x_t, \vartheta) = \pi_{su}(x_t, \vartheta_s) = \frac{\exp[\hat{G}_{su}(x_t)\vartheta_s]}{\sum_{w=1}^{I_s} \exp[\hat{G}_{sw}(x_t)\vartheta_s]},$$

где $\hat{G}_{su}(x)$ – вектор-строка значений регрессоров модели u -го уровня s -го отклика, то информационная матрица является блочно-диагональной с блоками вида

$$M_{\vartheta_s} \quad x_t | \vartheta_s = \sum_{u=1}^{I_s} \pi_{su}(x_t, \vartheta_s) \hat{G}_{su}^T(x_t) \left[\hat{G}_{su}(x_t) - \sum_{w=1}^{I_s} \pi_{sw}(x_t, \vartheta_s) \hat{G}_{sw}(x_t) \right].$$

Если справедливо

$$\pi_{su}(x_t, \vartheta) = \pi_{su}(x_t, \vartheta_s) = \frac{\exp[\check{G}_{su}(x_t)\vartheta_{su}]}{1 + \sum_{w=1}^{I_s-1} \exp[\check{G}_{sw}(x_t)\vartheta_{sw}]},$$

$$\pi_{sI_s}(x_t, \vartheta_s) = \left\{ 1 + \sum_{w=1}^{I_s-1} \exp \left[\tilde{G}_{sw}(x_t) \vartheta_{sw} \right] \right\}^{-1}, \quad u=1, \dots, I_s-1,$$

где \tilde{G}_{su} – вектор-строка значений регрессоров модели u -го уровня s -го отклика, то информационная матрица является блочно-диагональной с блоками $M_{\vartheta_s} x_t | \vartheta_s$, каждый из которых имеет блочную структуру

$$\left[M_{\vartheta_s} x_t | \vartheta_s \right]_{uw} = -\pi_{su}(x_t, \vartheta_s) \pi_{sw}(x_t, \vartheta_s) \tilde{G}_{su}^T(x_t) \tilde{G}_{sw}(x_t), \quad u \neq w;$$

$$\left[M_{\vartheta_s} x_t | \vartheta_s \right]_{uu} = \pi_{su}(x_t, \vartheta_s) \left(1 - \pi_{su}(x_t, \vartheta_s) \right) \tilde{G}_{su}^T(x_t) \tilde{G}_{su}(x_t).$$

Полученный вид информационной матрицы показывает, что она зависит от параметров модели ϕ и ϑ . По этой причине в данном случае необходимо использовать сложные стратегии планирования эксперимента – локально оптимальное, последовательное, минимаксное, байесовское. Далее рассмотрим локально оптимальное планирование. Последовательное и байесовское планирование для рассматриваемой модели исследовано в [11, 12].

4.4.2. ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА С УЧЕТОМ ПОЯВЛЕНИЯ ПРОПУСКОВ

При наличии пропусков в наборе данных неизвестные параметры оценивают методом максимального правдоподобия с использованием маргинальной плотности распределения совокупности реально наблюдавшихся откликов (см. п. 3.2).

В том случае, если имеются только количественные отклики и отсутствуют пропуски, информационная матрица, соответствующая вектору θ , определяется формулой

$$M_{\theta}(X | \Phi) = \sum_{t=1}^N F^T(x_t) \Omega F(x_t).$$

При наличии пропусков информационная матрица, соответствующая вектору θ , принимает вид

$$M_{\theta}(X | \Phi, R) = \sum_{t=1}^N F_t^T(x_t) \Omega_t F_t(x_t),$$

где R – структура пропусков набора данных; $F_t(x_t)$ – матрица регрессоров реально наблюдавшихся откликов в t -м измерении; Ω_t – матрица эффективности, построенная по маргинальному распределению реально наблюдавшихся откликов в t -м измерении.

Рассмотрим случай разнотипных откликов. Предположим, что пропуски возникают при измерении как количественных, так и качественных откликов. Плотность маргинального распределения наблюдаемых значений откликов имеет тогда вид (3.21).

В результате информационная матрица, соответствующая реальному набору данных, отличается от информационной матрицы, соответствующей гипотетическому полному набору данных, и зависит от структуры пропусков набора данных. Если наблюдения проводились в соответствии с оптимальным планом эксперимента, который построен без учета возможности возникновения пропусков, то качество оценок может не соответствовать планируемому. В этом случае задачу оптимального планирования эксперимента возможно решить с учетом указанной возможности.

Поскольку до проведения экспериментов не известно, в каких местах набора данных появятся пропуски, информационная матрица не может быть построена априорно. Поэтому в данном случае требуются специальные подходы к планированию эксперимента. Воспользуемся вероятностным подходом к учету пропусков [78].

В результате проведения одного эксперимента значение каждого отклика может быть либо зафиксировано, либо пропущено, поэтому всего имеется $u = 2^{n_y + n_z}$ возможных структур пропусков. Структуру пропусков t -го измерения вектора откликов можно отразить бинарным $(n_y + n_z)$ -мерным вектором R_t , элементы которого равны нулю или единице в зависимости от того, пропущены или нет соответствующие элементы в векторе откликов.

Будем считать вектор структуры пропусков случайным, тогда он может рассматриваться как дополнительный (мешающий) качественный отклик и может стоять задача построения его модели одновременно с построением модели основных откликов, включающая оптимизацию характеристик качества оценок параметров модели с использованием методов планирования эксперимента.

Предположим независимость вектора R_t от векторов Y_t , Z_t и введем модель распределения вектора R_t в виде набора вероятностей

$$\rho_i(x_t, \beta), \quad i = 1, \dots, u,$$

где β – вектор параметров. Совместная плотность распределения векторов R_t , $Y_{t,obs}$, $Z_{t,obs}$ представима в виде

$$g(R_t, Y_{t,obs}, Z_{t,obs} | x_t, \Phi, \beta) = \mathbf{P}(R_t | x_t, \beta) g(Y_{t,obs}, Z_{t,obs} | x_t, \Phi),$$

поэтому оценивать их можно независимо.

При справедливости некоторых условий регулярности [83] (нормированная) информационная матрица непрерывного плана ε , соответствующая всем параметрам, является блочно-диагональной с блоками $M_\Phi(\varepsilon | \Phi, \beta)$ и $M_\beta(\varepsilon | \beta)$. Информационная матрица $M_\beta(\varepsilon | \beta)$ при использовании логистической модели вероятностей $\rho_i(x, \beta)$ имеет вид, аналогичный виду информационной матрицы $M_\beta(\varepsilon | \beta)$ при отсутствии пропусков.

Информационная матрица $M_\Phi(\varepsilon | \Phi, \beta)$ представима в виде

$$M_\Phi(\varepsilon | \Phi, \beta) = \sum_{t=1}^n \sum_{i=1}^u p_t \rho_i(x_t, \beta) M_\Phi^i(x_t | \Phi),$$

где $M_\Phi^i(x_t | \Phi)$ – информационная матрица вектора параметров Φ в точке x_t при i -й структуре пропусков.

В случае только количественных откликов, матрица $M_{\Phi}^i x|\Phi$ имеет простой вид. Так, матрица $M_{\theta}^i x|\Phi$ имеет вид

$$M_{\theta}^i x|\Phi = F_{(i)}^T(x)\Omega_{(i)}F_{(i)}(x),$$

где $F_{(i)}(x)$ – подматрица матрицы $F(x)$, соответствующая наблюдаемым откликам при i -й структуре пропусков; $\Omega_{(i)}$ – матрица эффективности, построенная по маргинальному распределению вектора наблюдаемых откликов при i -й структуре пропусков. Соответствующий вектору θ диагональный блок информационной матрицы $M_{\Phi}(\varepsilon|\Phi, \beta)$ в данном случае может быть представлен в виде

$$M_{\theta}(\varepsilon|\Phi, \beta) = \sum_{t=1}^n p_t F^T(x_t) \bar{\Omega}(x_t) F(x_t), \quad \bar{\Omega}(x_t) = \sum_{i=1}^u \rho_i(x_t, \beta) \Omega_{[i]},$$

где $\Omega_{[i]}$ – $n_y \times n_y$ -матрица, полученная путем добавления в матрицу $\Omega_{(i)}$ нулевых строк и столбцов на места, соответствующие пропускам. В результате задача построения оптимального плана эксперимента для оценивания вектора θ при фиксированных остальных параметрах эквивалентна задаче поиска оптимального плана эксперимента при ММП-оценивании по неоднородным наблюдениям, рассмотренной в п. 4.2. Специфика заключается лишь в способе построения матрицы эффективности.

Для модели с разнотипными откликами матрица $M_{\Phi}^i x_t|\Phi$ определяется формулой

$$M_{\Phi}^i x_t|\Phi = \sum_{j=1}^{n_i} \int_{\Delta_i} \frac{\partial}{\partial \Phi} \ln \sum_{k \in J(i,j)} \pi_k x_t, \vartheta g_k Y_{obs,(i)} | x_t, \theta, \phi \times \\ \times \frac{\partial}{\partial \Phi^T} \sum_{k \in J(i,j)} \pi_k x_t, \vartheta g_k Y_{obs,(i)} | x_t, \theta, \phi dY_{obs,(i)},$$

где n_i – количество возможных значений вектора наблюдаемых качественных откликов в i -й структуре пропусков (при нулевом подвекто-

ре структур пропусков, соответствующем качественным откликам, $n_i = 1$); Δ_i – область определения вектора $Y_{obs,(i)}$; $J_{(i,j)}$ – множество возможных номеров ячеек при j -м возможном значении вектора наблюдаемых качественных откликов в i -й структуре пропусков; $Y_{obs,(i)}$ – вектор наблюдаемых количественных откликов в i -й структуре пропусков.

В данном случае информационная матрица $M_{\Phi}^i x_t | \Phi$ может быть вычислена аналитически при независимости количественных и качественных откликов или при отсутствии пропусков в наблюдениях качественных откликов. В общем случае невозможно получить ее значение аналитически.

В [87] изучался частный случай рассматриваемой модели, когда имеется единственный полностью ненаблюдаемый качественный отклик и полностью наблюдаемый одномерный нормальный количественный отклик. Для вычисления информационной матрицы предложено использовать численное интегрирование, однако отмечено, что этот подход обычно очень трудоемок.

С целью сокращения вычислений при последовательном планировании эксперимента для наблюдений с пропусками в качественных откликах можно использовать *наблюдаемую* информационную матрицу, имеющую вид [45, 140]

$$M_{\Phi}(x_t | Y_{t,obs}, Z_{t,obs}, \hat{\Phi}) = - \left. \frac{\partial^2 \ln g(Y_{t,obs}, Z_{t,obs} | x_t, \tilde{\Phi})}{\partial \tilde{\Phi} \partial \tilde{\Phi}^T} \right|_{\tilde{\Phi} = \hat{\Phi}}.$$

Возможно также использование наблюдаемой информационной матрицы, вычисленной условно только по наблюдениям качественных откликов и структурам пропусков (см. близкий пример в [140]).

4.4.3. СВОЙСТВА ИНВАРИАНТНОСТИ ПЛАНОВ ЭКСПЕРИМЕНТА

Рассмотрим ситуацию, когда пропуски возникают в наблюдениях количественных откликов, при этом количественные отклики либо независимы с качественными откликами, либо в качественных откли-

ках пропуски не возникают. Пусть для каждого наблюдения количественных откликов есть две структуры пропусков: в результате эксперимента значения откликов могут быть либо зафиксированы, либо пропущены, и вероятность получения результата в точке x_t равна $\rho(x_t)$.

В этом случае информационная матрица $M_\Phi \varepsilon | \Phi$ является блочно-диагональной с блоками, соответствующими параметрам θ , ϕ и ϑ . Заметим также, что при $\rho(x_t) = 1$ набор данных не содержит пропусков в количественных откликах.

Для классической регрессионной модели известно свойство инвариантности D -оптимальных планов эксперимента к невырожденному линейному преобразованию параметров θ [113]. Подобное свойство справедливо и для рассматриваемой модели. Однако здесь имеется ряд частных случаев, в которых можно выявить дополнительные свойства инвариантности D -оптимальных планов, позволяющие снизить требования к априорной информации о модели, упростить алгоритмы построения оптимальных планов, получить аналитические решения.

Модель 1. Пусть распределения вектора количественных откликов в разных ячейках различаются только параметром сдвига (уравнением регрессии), в частности, имеют общие мешающие параметры ϕ , тогда для информационной матрицы параметров ϕ справедливо представление

$$M_\Phi \varepsilon | \phi = M_\phi \phi \sum_{t=1}^n p_t \rho(x_t),$$

где $M_\phi \phi$ – информационная матрица параметров ϕ , соответствующая распределению в одной ячейке. В результате D -оптимальный план является инвариантным к значениям матрицы $M_\phi \phi$ и в алгоритмах построения оптимальных планов не требуется оперировать с данными матрицами, размер которых может быть большим.

В том случае, если функция $\rho(x_t)$ является константной $\rho(x_t) = \rho$, справедливо

$$M_\Phi \varepsilon | \phi = \rho M_\phi \phi .$$

Таким образом, информационная матрица $M_{\phi} \varepsilon | \phi$ не зависит от плана ε .

Модель 2. Пусть можно выделить K групп ячеек так, что распределения вектора количественных откликов в разных ячейках k -й группы различаются только параметром сдвига (уравнением регрессии), в частности, имеют общий вектор мешающих параметров $\phi_{(k)}$, элементы которого не являются параметрами плотностей распределений в других ячейках.

В этом случае информационная матрица вектора $\phi_{(1)}, \dots, \phi_{(K)}, \tilde{\phi}$ является блочно-диагональной с блоками

$$M_{\phi_{(k)}} \varepsilon | \phi_{(k)}, \vartheta = M_{\phi_{(k)}} \phi_{(k)} \sum_{t=1}^n p_t \rho(x_t) \sum_{j \in J(k)} \pi_j(x_t, \vartheta),$$

$$k = 1, \dots, K, M_{\tilde{\phi}} \varepsilon | \tilde{\phi}, \vartheta,$$

где $\tilde{\phi}$ – элементы вектора ϕ , не входящие в векторы $\phi_{(k)}$, $k = 1, \dots, K$; $M_{\phi_{(k)}} \phi_{(k)}$ – информационная матрица параметров $\phi_{(k)}$, соответствующая распределению в одной ячейке из k -й группы; $J(k)$ – множество номеров ячеек k -й группы.

Тогда при D -оптимальном планировании имеем

$$|M_{\Phi}(\varepsilon | \Phi)| = |M_{\theta}(\varepsilon | \phi, \vartheta)| |M_{\vartheta}(\varepsilon | \vartheta)| \times \\ \times \prod_{k=1}^K \left\{ \left| M_{\phi_{(k)}}(\phi_{(k)}) \left[\sum_{t=1}^n p_t \rho(x_t) \sum_{j \in J(k)} \pi_j(x_t, \vartheta) \right]^{d(k)} \right\} |M_{\tilde{\phi}}(\varepsilon | \tilde{\phi}, \vartheta)|,$$

где $d(k)$ – размерность вектора $\phi_{(k)}$.

Необходимые и достаточные условия D -оптимальности в данном случае имеют вид

$$\begin{aligned} & \max_{x \in \tilde{X}} \operatorname{tr} M_{\theta}^{-1}(\varepsilon | \phi, \vartheta) M_{\theta}(x | \phi, \vartheta) + \operatorname{tr} M_{\vartheta}^{-1}(\varepsilon | \vartheta) M_{\vartheta}(x | \vartheta) + \\ & + \rho(x) \sum_{k=1}^K \left\{ d_{(k)} \sum_{j \in J(k)} \pi_j(x, \vartheta) \left[\sum_{t=1}^n p_t \rho(x_t) \sum_{j \in J(k)} \pi_j(x_t, \vartheta) \right]^{-1} \right\} + \\ & + \operatorname{tr} M_{\tilde{\phi}}^{-1}(\varepsilon | \tilde{\phi}, \vartheta) M_{\tilde{\phi}}(x | \tilde{\phi}, \vartheta) = d_{\theta} + d_{\vartheta} + \sum_{k=1}^K d_{(k)} + d_{\tilde{\phi}}, \end{aligned}$$

где d_{θ} , d_{ϑ} , $d_{\tilde{\phi}}$ – размерности векторов θ , ϑ и $\tilde{\phi}$ соответственно.

В результате D -оптимальный план является инвариантным к значениям информационных матриц $M_{\phi_{(k)}} \phi_{(k)}$, $k = 1, \dots, K$.

Модель 3. Рассмотрим планирование эксперимента для параметров $\phi_{(k)}$, $k = 1, \dots, K$, в условиях модели 2.

Прежде всего заметим, что план, D -оптимальный для векторов $\phi_{(k)}$, $k = 1, \dots, K$, инвариантен к виду распределения количественных откликов и к значениям самих векторов $\phi_{(k)}$, но не инвариантен к их размерностям. Кроме того, план, сосредоточенный в точках, в которых максимальна вероятность получить наблюдение без пропусков в количественных откликах в k -й группе ячеек, является D -оптимальным для вектора $\phi_{(k)}$.

Пусть $d_{(k)} = d$, $k = 1, \dots, K$, тогда достаточными условиями D -оптимальности плана для параметров $\phi_{(k)}$, $k = 1, \dots, K$, является выполнение равенства

$$\sum_{t=1}^n p_t \rho(x_t) \sum_{j \in J(k)} \pi_j(x_t, \vartheta) = \frac{1}{K} \max_{x \in \tilde{X}} \rho(x) \sum_{i=1}^K \sum_{s \in J(i)} \pi_s(x, \vartheta), \quad k = 1, \dots, K. \quad (4.30)$$

Пусть для каждой группы имеются точки $x_{(k),s}$, наблюдения в которых с единичной вероятностью попадают в эту группу, и выполнено условие

$$\rho(x_{(k),s}) = \max_{x \in \tilde{X}_g} \rho(x),$$

где

$$\tilde{X}_g = \left\{ x \in \tilde{X} \left| \sum_{k=1}^K \sum_{j \in J(k)} \pi_j(x, \vartheta) > 0 \right. \right\}.$$

Тогда план эксперимента, спектр которого состоит из любого количества этих точек, причем суммарный вес точек для каждой группы одинаков, удовлетворяет (4.30), следовательно, является D -оптимальным.

Рассмотрим частный случай, когда для ячеек, входящих в группы, справедливо $\pi_j(x) = 1$ при $x \in \tilde{X}_j$ и $\pi_j(x) = 0$ при $x \notin \tilde{X}_j$, где \tilde{X}_j – некоторая подобласть области планирования и все подобласти являются непересекающимися, функция $\rho(x_t)$ – константная. В данном случае выполняются условия (4.30) для любого плана, удовлетворяющего системе равенств

$$\sum_{t: x_t \in \tilde{X}_{(k)}} p_t = \frac{1}{K}, \quad k = 1, \dots, K,$$

где $\tilde{X}_{(k)} = \bigcup_{j \in J(k)} \tilde{X}_j$. Таким образом, согласно D -оптимальному плану

во всех областях $\tilde{X}_{(k)}$ должно быть проведено одинаковое количество наблюдений.

Рассмотрим для этого же частного случая A -оптимальное планирование. Пусть распределение количественных откликов – одномерное нормальное, k -я группа имеет мешающий параметр – дисперсию $\sigma_{(k)}^2$. Оптимальный план удовлетворяет условию

$$\sum_{t: x_t \in \tilde{X}_{(k)}} p_t = \frac{\sigma_{(k)}^2}{\sum_{s=1}^K \sigma_{(s)}^2}, \quad k = 1, \dots, K.$$

В результате, в отличие от D -оптимального плана, при увеличении значения $\sigma_{(k)}^2$ увеличивается доля наблюдений, которые должны быть проведены в области $\tilde{X}_{(k)}$.

Модель 4. Пусть можно выделить K групп ячеек так, что распределения вектора количественных откликов в ячейках k -й группы являются эллиптическими, отличаются только параметром сдвига (уравнением регрессии), имеют общие псевдоковариационную матрицу группы $V_{(k)}$ и параметр формы группы $\nu_{(k)}$, уравнения регрессии имеют при этом общий вектор параметров $\theta_{(k)}$, элементы которого не используются для параметризации плотностей распределений в других ячейках.

В этом случае информационная матрица вектора $\theta_{(1), \dots, \theta_{(K)}, \tilde{\theta}}$ является блочно-диагональной с блоками

$$M_{\theta_{(k)}} \varepsilon | \phi_{(k)}, \vartheta = \omega_{(k)} \sum_{t=1}^n p_t \rho(x_t) \sum_{j \in J_{(k)}} \pi_j(x_t, \vartheta) F_j^T(x_t) V_{(k)}^{-1} F_j(x_t),$$

$$k = 1, \dots, K, \quad M_{\tilde{\theta}} \varepsilon | \phi, \vartheta,$$

где $\tilde{\theta}$ – элементы вектора θ , не входящие в векторы $\theta_{(k)}$, $k = 1, \dots, K$; $\omega_{(k)}$ – функция $V_{(k)}$ и $\nu_{(k)}$. В результате D -оптимальный план для параметров θ инвариантен к виду эллиптического распределения и значениям параметра формы $\nu_{(k)}$, если они не используются для параметризации плотностей распределений в ячейках, не входящих в группы. При $n_y = 1$ аналогично свойство справедливо для параметров $V_{(k)}$.

Комбинируя результаты, полученные для моделей 3 и 4, можно сформулировать соответствующие свойства инвариантности к значениям мешающих параметров и виду (эллиптического) распределения в случае одновременного планирования для параметров θ и ϕ при отсутствии ячеек, не входящих в группы.

4.4.4. ИССЛЕДОВАНИЕ СВОЙСТВ ПЛАНОВ ЭКСПЕРИМЕНТА ПРИ ВОЗНИКНОВЕНИИ ПРОПУСКОВ

Изучим зависимость свойств D - и A -оптимальных планов эксперимента от величины вероятности возникновения пропусков в качественных откликах на одном примере модели с разнотипными откликами.

Рассматривалась зависимость одномерного количественного отклика и одномерного качественного отклика с двумя возможными значениями от скалярной входной переменной x .

Наблюдения количественного отклика в обеих ячейках имели нормальное распределение с дисперсией, равной 0,01. Математическое ожидание количественного отклика в j -й ячейке имело вид

$$\theta_{1j} + \theta_{2j}x,$$

где $j=1, 2$, $\theta_{11}=20$, $\theta_{21}=-5$, $\theta_{12}=10$, $\theta_{22}=5$; $x \in [-1, 1]$.

Модель качественного отклика имела вид

$$\pi_1(x, \vartheta) = \frac{\exp \vartheta_0 + \vartheta_1 x}{1 + \exp \vartheta_0 + \vartheta_1 x}, \quad \pi_2(x, \vartheta) = \frac{1}{1 + \exp \vartheta_0 + \vartheta_1 x},$$

где $\vartheta_0 = -3$, $\vartheta_1 = 5$.

Зависимости $\pi_1(x, \vartheta)$ и $\pi_2(x, \vartheta)$ от x отражены на рис. 4.15. На рис. 4.16 показаны зависимости от x математических ожиданий количественного отклика в первой и во второй ячейках и безусловное математическое ожидание количественного отклика.

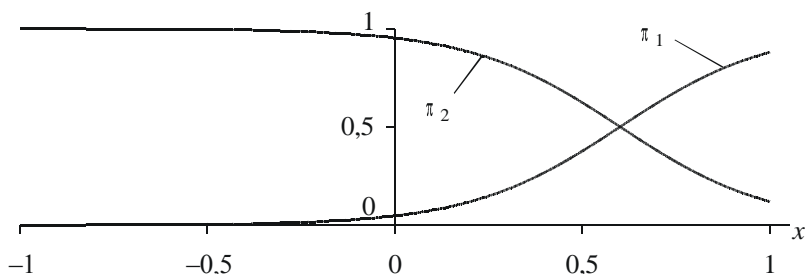


Рис. 4.15. Функции $\pi_1(x, \vartheta)$ и $\pi_2(x, \vartheta)$

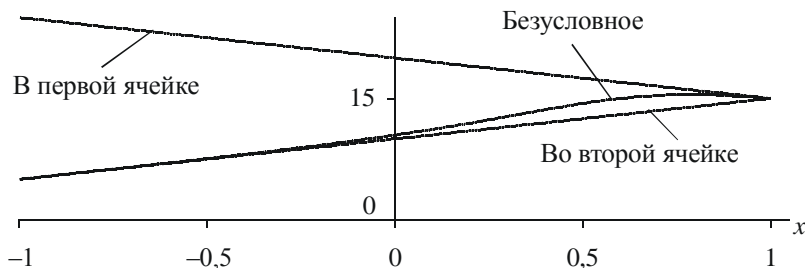


Рис. 4.16. Математические ожидания

Вероятность появления пропущенного значения в качественном отклике не зависела от входной переменной.

В табл. 4.4 приведены непрерывные локально оптимальные планы для параметров θ и ϑ , которые получены численно при различных значениях вероятности пропуска в качественном отклике, обозначенной κ , а также эффективность плана ε_0 , полученного без учета возникновения пропусков. Показатели эффективности для критериев D - и A -оптимальности вычислялись соответственно по формулам

$$\tau_D = \sqrt[6]{\frac{|M(\varepsilon_0)|}{|M(\varepsilon^*)|}}, \quad \tau_A = \frac{\text{tr} M^{-1}(\varepsilon^*)}{\text{tr} M^{-1}(\varepsilon_0)},$$

где ε^* — соответствующий оптимальный план.

Результаты, приведенные в табл. 4.4, показывают влияние значения вероятности появления пропусков на характеристики планов. Так, в важном для практики случае латентного качественного отклика ($\kappa = 1$) использование плана ε_0 нецелесообразно (при D -оптимальном планировании потери эффективности составляют около 58 %, а при A -оптимальном планировании данный план практически является вырожденным), велики потери и в другом важном для практики случае при $\kappa = 0,9$, когда имеются небольшие классифицированные подвыборки (см. [94]).

Исследуем свойства планов эксперимента для оценивания вектора параметров θ в условиях появления пропусков для моделей с количественными откликами. Планы эксперимента будем сравнивать с точки зрения критерия D -оптимальности.

Таблица 4.4

Оптимальные планы

κ	План		Эффективность	
	<i>D</i> -оптимальный	<i>A</i> -оптимальный	τ_D	τ_A
0	{− 1,000, 0,163; 0,373, 0,464; 1,000, 0,373}	{0,148, 0,608; 1,000, 0,392}	1	1
0,1	{− 1,000, 0,163; 0,365, 0,463; 0,964, 0,207; 1,000, 0,167}	{0,141, 0,622; 0,962, 0,378}	0,999	0,979
0,5	{− 1,000, 0,162; 0,352, 0,460; 0,946, 0,378}	{0,139, 0,627; 0,950, 0,373}	0,934	0,854
0,9	{− 1,000, 0,162; 0,349, 0,459; 0,941, 0,379}	{0,138, 0,629; 0,946, 0,371}	0,874	0,733
1	{− 1,000, 0,162; 0,348, 0,459; 0,939, 0,379}	{0,137, 0,630; 0,942, 0,370}	0,418	0,000

Рассмотрим область планирования и уравнение регрессии, использовавшиеся в моделях 1, 2 в п. 4.2.3.

Модель 1. Рассмотрим модель (4.8) в однооткликном случае. Предположим, что ошибки наблюдений являются независимыми нормально распределенными с дисперсией σ^2 . Всего имеется две структуры пропусков: в результате эксперимента значение отклика может быть либо зафиксировано, либо пропущено. Пусть вероятность получения результата имеет следующие значения: $\tilde{\pi}_1$ в области \tilde{X}_1 , $\tilde{\pi}_2$ в области \tilde{X}_2 . Тогда значение функции $\bar{\Omega}(x)$ равно $\frac{\tilde{\pi}_1}{\sigma^2}$ в области \tilde{X}_1 и $\frac{\tilde{\pi}_2}{\sigma^2}$ в области \tilde{X}_2 .

Для изучения данного случая можно воспользоваться результатами анализа модели 2 из п. 4.2.3, обозначив $\rho = \frac{\tilde{\pi}_1}{\tilde{\pi}_2}$. Рассмотрим ряд конкретных примеров.

Пусть $s = 0,1$, $\tilde{\pi}_2 = 0,9$, тогда при $\tilde{\pi}_1 < 0,8$ оптимален план ε_{1-s} , при $\tilde{\pi}_1 = 0,75$ получаем $\Delta_D \approx 0,92$ (показатель Δ_D определен в п. 4.2.3), а при $\tilde{\pi}_1 = 0,5$ – $\Delta_D \approx 0,62$. Пусть $\tilde{\pi}_1 = 0,7$, $\tilde{\pi}_2 = 0,9$, тогда при $s > 0,26$ оптимален план ε_1 , а при $s < 0,2$ оптимален план ε_{1-s} , величина Δ_D в последнем случае лежит в интервале $(0,77...0,98)$, вычисленном приближенно.

Определим потери информации при использовании плана ε_α вида (4.9) в условиях возникновения пропусков. Пусть вероятности получения наблюдений в точках α и $-\alpha$ одинаковы и равны π_α . Определитель информационной матрицы, вычисленной без учета возможности возникновения пропусков, равен

$$|M_\theta| = \frac{\alpha^2}{\sigma^4},$$

а с учетом пропусков –

$$|M_{\theta,m}| = \frac{\alpha^2 \pi_\alpha^2}{\sigma^4}.$$

Для сравнения полученных информационных матриц можно воспользоваться величиной

$$\Delta = \sqrt{\frac{|M_{\theta,m}|}{|M_\theta|}} = \pi_\alpha.$$

Модель 2. Рассмотрим двухоткликую модель вида (4.8). Ошибки наблюдений распределены по двумерному нормальному закону. Есть четыре структуры пропусков: первая соответствует наблюдению без пропусков, вторая – пропуску второго отклика, третья – пропуску первого отклика, четвертая – пропуску обоих откликов.

Изучим влияние коэффициента корреляции γ ошибок наблюдения откликов, величины s и векторов вероятностей структур пропусков в областях \tilde{X}_1 и \tilde{X}_2 , обозначаемых $\tilde{\pi}_1$ и $\tilde{\pi}_2$ соответственно, на свойства непрерывных оптимальных планов эксперимента.

Все оптимальные планы найдены численно и принадлежат семейству

$$\varepsilon_p = \left\{ -1, \frac{1-p}{2}; -1+s, \frac{p}{2}; 1-s, \frac{p}{2}; 1, \frac{1-p}{2} \right\},$$

где $p \in [0, 1]$. Заметим, что значению $p=0$ соответствует двухточечный «классический» план с точками $x_{1,2} = \pm 1$, а значению $p=1$ – двухточечный план с точками $x_{1,2} = \pm(1-s)$ на границе внутренней области.

Рассмотрим области оптимальности планов ε_0 и ε_1 для двух наборов вероятностей: первому соответствуют векторы вероятностей структур пропусков $\tilde{\pi}_1 = 0,4 \ 0,15 \ 0,15 \ 0,3$, $\tilde{\pi}_2 = 0,8 \ 0,1 \ 0,1 \ 0$, второму – $\tilde{\pi}_1 = 0,3 \ 0,2 \ 0,2 \ 0,3$, $\tilde{\pi}_2 = 0,9 \ 0,05 \ 0,05 \ 0$. На рис. 4.17 представлены области, в которых оптимальны планы: ε_0 оптимален в области выше верхней линии, а ε_1 – в области ниже нижней линии, в области между линиями оптимальны другие планы из семейства ε_p .

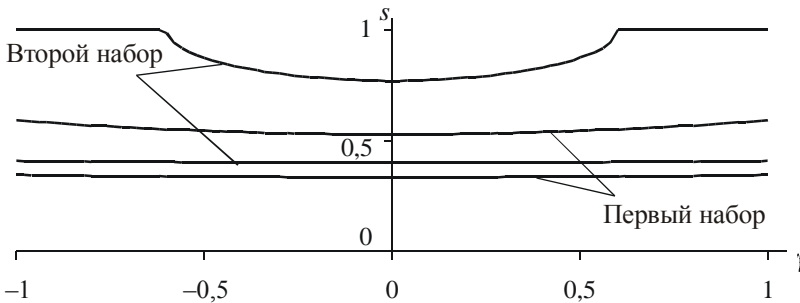


Рис. 4.17. Области оптимальности планов ε_0 и ε_1

На рис. 4.18 представлены оптимальные значения параметра p при $\tilde{\pi}_1 = 0,4 \ 0,2 \ 0,2 \ 0,2$, $\tilde{\pi}_2 = 0,8 \ 0,1 \ 0,1 \ 0$ и различных значениях s .

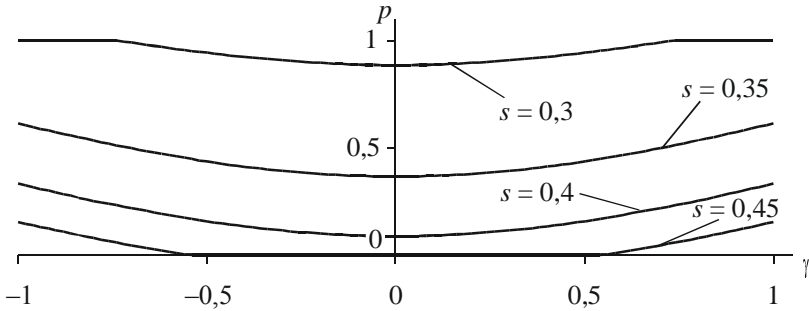


Рис. 4.18. Оптимальные значения параметра p

Пусть $\tilde{\pi}_1 = 0,4 \ 0,15 \ 0,15 \ 0,3$, $\tilde{\pi}_2 = 0,8 \ 0,1 \ 0,1 \ 0$, $\gamma = 0$. Тогда при $s \leq 0,335$ оптимальным является план ε_1 , а при $s \geq 0,77$ – план ε_0 . При $s = 0,1$ показатель эффективности плана ε_0 равен $\Delta_D \approx 0,67$, а при $s = 0,2$ – $\Delta_D \approx 0,76$. При тех же значениях $\tilde{\pi}_2$ и γ , при $\tilde{\pi}_1 = 0,5 \ 0,2 \ 0,2 \ 0,1$ и $s \leq 0,2$ оптимальным является план ε_1 , а при $s \geq 0,255$ – план ε_0 . Показатель эффективности плана ε_0 при $s = 0,1$ составляет $\Delta_D \approx 0,86$.

ГЛАВА 5

ВЫБОР СТРУКТУРЫ МОДЕЛИ

Глава посвящена развитию методов выбора структуры регрессионной модели [53–56, 59, 62, 67, 69, 71–74]. Сформулирована обобщенная задача выбора структуры модели, включающая оптимизацию состава регрессоров и системы линейных ограничений на параметры, рассмотрены критерии качества структур моделей, алгоритмы решения задачи выбора структуры и эффективные вычислительные схемы их реализации. Критерии и алгоритмы, изучаемые в п. 5.1 и 5.2 ориентированы на модели и методы оценивания, рассмотренные в п. 1.1–1.3. Для многооткликовых моделей при выборе структуры используется информация о гетероскедастичности и коррелированности ошибок одного наблюдения. В п. 5.3 рассматриваются подходы к выбору структуры при более общих предположениях о свойствах ошибок наблюдений.

5.1. ЗАДАЧИ И КРИТЕРИИ ВЫБОРА СТРУКТУРЫ

5.1.1. ЗАДАЧИ ВЫБОРА СТРУКТУРЫ

В условиях большой сложности или недостаточной изученности объекта исследования построение математической модели основывается на концепции «черного ящика», когда исследователь фиксирует отклики, описание или прогноз которых он желает получить, и воздействующие на них факторы, а функция, описывающая механизм воз-

действия факторов на отклики, неизвестна. Часто исследователь может выделить достаточно большое количество факторов, влияющих на отклики; число регрессоров линейной по параметрам регрессии, которые можно получить, используя, например, полиномиальную модель, будет еще большим. Однако не все регрессоры одинаково полезны для описания и прогнозирования значений откликов – какие-то регрессоры являются информативными, какие-то – нет. Часто регрессоры оказываются взаимозависимыми и в модель достаточно включить только часть из них. Кроме того, исключение из модели части взаимозависимых и неинформативных регрессоров способствует улучшению статистических свойств оценок параметров. По указанным причинам выбор структуры модели – один из важнейших этапов ее построения. Однако этот этап разработан и формализован недостаточно, субъективные суждения играют здесь еще существенную роль.

Будем рассматривать отдельно задачу выбора структуры однооткликowej модели вида (1.1) и общей многооткликowej модели (1.38), включающей в себя модель псевдонезависимых регрессий, и задачу выбора структуры традиционной многооткликowej модели в силу ее специфики.

Традиционно под выбором структуры модели понимается определение регрессоров, полезных для описания откликов, из некоторого полного набора $f_1(x), \dots, f_m(x)$, определенного исследователем. Задачу, состоящую в определении этих регрессоров, назовем *традиционной задачей выбора структуры модели*.

Сформулируем задачу выбора структуры как оптимизационную [109]:

$$s^* = \arg \min_{s \in \Omega} CR(s), \quad (5.1)$$

где s^* – оптимальная структура модели; s – структура модели; Ω – множество всех возможных структур, содержащих некоторые регрессоры из полного набора регрессоров; $CR(s)$ – критерий качества структуры. Выбор конкретного критерия качества может определяться целями построения модели.

Модель со структурой, доставляющей критерию качества минимальное значение, называют моделью оптимальной сложности [39]

или «наилучшим» уравнением регрессии [33]. Под сложностью при этом понимается количество регрессоров в модели.

Заметим, что не все подходы к выбору структуры могут быть сформулированы в виде задачи (5.1). Например, если при добавлении регрессоров в модель значение критерия качества монотонно уменьшается, то задача выбора структуры может быть модифицирована следующим образом: выбрать наиболее простую структуру s , критерий качества для которой удовлетворяет условию $CR(s) \leq k$, где k – пороговое значение [85].

Любая модель, полученная исключением из полной модели некоторых регрессоров, может рассматриваться как полная модель, на параметры которой наложена система исключаящих ограничений – линейных ограничений о равенстве нулю параметров, стоящих при исключенных регрессорах. Следовательно, структуры, составляющие множество всех возможных структур Ω (5.1), отличаются системой исключаящих ограничений, наложенных на параметры полной модели, содержащей все возможные регрессоры. В результате традиционную задачу выбора структуры можно сформулировать как задачу выбора оптимальной системы исключаящих ограничений на параметры модели.

Полное множество исключаящих ограничений, из которого выбирается оптимальное подмножество для общей многооткликowej модели (1.38), можно, записывая их одно под другим, представить в матричном виде как

$$I_m \theta = 0_m, \quad (5.2)$$

где I_m – единичная матрица размерности m ; m – количество регрессоров в полной модели; 0_m – нулевой m -мерный вектор.

Для традиционной многооткликowej модели полное множество исключаящих ограничений имеет вид

$$I_m \Theta = 0_{m \times l}, \quad (5.3)$$

где m – количество регрессоров в полной модели, описывающих один отклик; $0_{m \times l}$ – нулевая матрица размерности $m \times l$.

Пример. Пусть на отклик предположительно влияют три фактора x_1, x_2, x_3 и регрессия является линейной, полная модель при этом запишется в виде

$$y = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + e.$$

Тогда для моделирования отклика возможно выбрать одну из восьми структур, задающихся следующими наборами регрессоров:

$$\emptyset, x_1, x_2, x_3, x_1, x_2, x_1, x_3, x_2, x_3, x_1, x_2, x_3,$$

при этом первой структуре соответствует случай отсутствия зависимости отклика от факторов.

С другой стороны, можно получить те же структуры, накладывая на параметры полной модели исключаящие ограничения; соответственно получим

$$\theta_1 = 0, \theta_2 = 0, \theta_3 = 0, \theta_1 = 0, \theta_3 = 0, \theta_1 = 0, \theta_3 = 0, \theta_1 = 0, \theta_2 = 0,$$

$$\theta_3 = 0, \theta_2 = 0, \theta_1 = 0, \emptyset,$$

при этом полное множество ограничений, из которых выбираются ограничения, определяющие структуру, есть $\theta_1 = 0, \theta_2 = 0, \theta_3 = 0$, или в матричном виде

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

При построении модели исследователю может быть доступна априорная информация об объекте исследования, на основании которой возможно сформировать набор предположений в виде произвольных линейных ограничений-равенств на параметры модели.

Например, в виде линейного ограничения можно сформулировать предположение о прохождении графика зависимости через определенную точку; можно задать линейное ограничение, отражающее незначимость различий между двумя уровнями качественного фактора; по-

средством наложения системы линейных ограничений можно ввести в модель псевдонезависимых регрессий общий параметр.

Если ограничения соответствуют экспериментальным данным, то их учет делает модель более адекватной объекту исследования. Обычно проверку соответствия ограничений экспериментальным данным осуществляют с помощью механизма проверки гипотез. Другой путь использования данных предположений состоит в формировании оптимальной системы линейных ограничений на параметры путем решения задачи (5.1).

Введем множество ограничений, сформированных на основе априорной информации об объекте исследования. Для многооткликowych моделей общего вида (1.38) данное множество запишем в матричном виде как

$$C^T \theta = \xi, \quad (5.4)$$

где C^T – матрица размерности $w \times m$; ξ – вектор размерности w .

Для традиционной многооткликовой модели данное множество ограничений имеет вид

$$C^T \Theta = \Xi, \quad (5.5)$$

где C^T и Ξ – матрицы размерности $w \times m$ и $w \times l$ соответственно.

Тогда вся совокупность имеющихся ограничений для моделей типа (1.38) запишется как

$$\begin{bmatrix} I_m \\ C^T \end{bmatrix} \theta = \begin{bmatrix} 0_m \\ \xi \end{bmatrix}, \quad (5.6)$$

а для традиционной многооткликовой модели – как

$$\begin{bmatrix} I_m \\ C^T \end{bmatrix} \Theta = \begin{bmatrix} 0_{m \times l} \\ \Xi \end{bmatrix}. \quad (5.7)$$

В общем случае (5.6) и (5.7) можно рассматривать как системы несовместных уравнений. Несовместность отражает тот факт, что набор предположений может быть противоречивым, например, содержащим конкурирующие предположения.

Продолжая наш пример, предположим, что исследователь включает в рассмотрение ограничения $\theta_1 = \theta_2$, $\theta_3 = 5$. Тогда множество (5.4) имеет вид

$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \end{bmatrix},$$

а множество (5.6) – вид

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 5 \end{bmatrix}.$$

Заметим, что ограничения $\theta_3 = 0$ из множества (5.2) и $\theta_3 = 5$ из множества (5.4) являются несовместными и не могут одновременно присутствовать в одной структуре.

Под *обобщенной задачей выбора структуры модели* будем понимать задачу формирования совместной системы ограничений на параметры, в наибольшей степени соответствующей экспериментальным данным. При этом данная система ограничений является подмножеством исходного множества ограничений (5.6) или (5.7). С другой стороны, можно сказать, что обобщенная задача выбора структуры модели подразумевает формирование множества (множеств) регрессоров, описывающих отклики, и формирование совместной системы ограничений на параметры, являющейся подмножеством множества (5.4) или (5.5).

Структуру модели в обобщенной задаче можно представить в виде системы линейных ограничений на параметры полной модели. Для моделей вида (1.38) будем рассматривать систему *и* линейных ограничений на параметры вида (1.9), являющуюся подмножеством (5.6), а для традиционной многооткликной модели – систему, являющуюся подмножеством (5.7) и имеющую вид (1.25).

Учитывая, что включению-исключению ограничения из множества (5.2) или (5.3) соответствует исключение-включение регрессора, мож-

но дать другое определение структуры: структура состоит из множества (множеств) регрессоров, описывающих отклики, и системы линейных ограничений на параметры при соответствующих регрессорах, являющейся подмножеством множества (5.4) или (5.5).

В обобщенной задаче выбора структуры понятие *сложности структуры* целесообразно связать с числом степеней свободы, используемых на оценивание параметров регрессии. Для традиционной многооткликковой модели в качестве значений сложности целесообразно использовать число степеней свободы, используемых на оценивание параметров уравнения одного отклика. С учетом сделанных выше предположений сложность определяется как $m - u$.

Заметим, что при построении модели достаточно отыскать систему ограничений с матрицей коэффициентов R , имеющей линейно независимые строки. При построении «физической» модели, когда ограничения несут смысловую нагрузку, после поиска такой системы ограничений можно отыскать другие ограничения из множества (5.6) или (5.7), которым оценки параметров удовлетворяют автоматически.

Рассмотрим понятие устойчивости применительно к задачам выбора структуры. В данном случае можно выделить три аспекта устойчивости.

Первый аспект связан с устойчивостью результатов моделирования к варьированию выборки [2, 17]. Если строятся две модели с одной структурой по двум различным частям выборки и результаты моделирования получаются различными, то выбранная структура не обеспечивает устойчивости результатов и в целом приводит к противоречивой модели. Один из способов обеспечения устойчивости в данном смысле – это применение критериев, использующих экзаменационную выборку. Данные критерии рассмотрены в п. 5.1.2 и 5.1.3.

Второй аспект связан с помехоустойчивостью критериев качества, т. е. с их способностью выбирать структуры, обеспечивающие удовлетворительные в каком-либо смысле результаты моделирования, в условиях увеличивающейся дисперсии ошибок [39].

Рассмотрим способность критериев выбирать структуры, обеспечивающие удовлетворительный прогноз значений откликов. Чтобы определить, удовлетворяет ли некоторая структура данному условию, вводится теоретический критерий прогнозирования, использующий

истинные (не содержащие ошибки) значения данных. Для классической однооткликовой регрессионной модели он имеет вид

$$\text{TSP}(Z) = \mathbf{E} \left\{ \left[\mathbf{E}(Y_Z) - X_Z \hat{\theta} \right]^T \left[\mathbf{E}(Y_Z) - X_Z \hat{\theta} \right] \right\},$$

где Z – индекс выборки; Y_Z , X_Z – вектор значений откликов и матрица значений регрессоров соответствующей выборки. Индекс Z может соответствовать выборке, по которой строилась модель, а может – и выборке, не использовавшейся при построении модели. В [39] критерий $\text{TSP}(Z)$ исследуется и доказывается, что при малой дисперсии ошибок его минимум соответствует истинной структуре, а при увеличении дисперсии – структурам уменьшающейся сложности. Данный эффект и выявляет помехоустойчивость критериев – ошибка восстановления математического ожидания отклика уменьшается за счет упрощения модели.

С другой стороны, реальный критерий, использующий данные с ошибками, – это лишь «оценка» для теоретического критерия, и при увеличении дисперсии он может переходить к модели меньшей сложности «быстрее» теоретического. В этом смысле теоретический критерий является наилучшим, а положительное свойство реальных критериев – как можно более «медленный» (но не медленнее, чем у теоретического критерия) переход к модели меньшей сложности.

Третий аспект устойчивости связан с устойчивостью процедур выбора модели к отклонению от классических предположений о свойствах ошибок наблюдений. Традиционно процедуры выбора структуры опираются на МНК-оценки и сами критерии строятся на основе остаточных сумм квадратов. При таком подходе не учитываются, например, негауссовость, гетероскедастичность, зависимость ошибок наблюдений.

Методы выбора структуры, разработанные для классических предположений, могут быть неустойчивыми в условиях их нарушения. Например, в работе [183] показывается неустойчивость методов выбора структуры при наличии выбросов в наборе данных. Если для оценивания параметров используется много различных устойчивых методов, то задача выбора структуры в данных условиях исследована в гораздо меньшей степени. Ряд подходов рассматривается в п. 5.3.

5.1.2. КРИТЕРИИ, ИСПОЛЬЗУЮЩИЕ РАЗБИЕНИЕ ВЫБОРКИ НА ДВЕ ЧАСТИ

Критерии, разработанные для выбора структуры, удобно разделить на две группы: критерии, использующие экзаменационную выборку, и критерии, не использующие экзаменационную выборку.

Критерии первой группы строятся на основе принципа, согласно которому параметры модели необходимо оценить по одной части выборки (обучающей), а качество структуры – по другой части выборки (экзаменационной, или проверочной).

Критерии первой группы в свою очередь удобно разделить на критерии, использующие разбиение выборки на две части, и критерии типа скользящего контроля.

Рассмотрим критерии, использующие разбиение выборки на две части. Это критерии регулярности, стабильности, непротиворечивости и вариативности, разработанные школой Ивахненко [39, 40, 109].

Рассмотрим однооткликтовую модель (1.1). Пусть исходная выборка $W: Y, X$ разделена на части $A: Y_A, X_A$ и $B: Y_B, X_B$.

Введем следующие обозначения: $\hat{\theta}_W, \hat{\theta}_A$ и $\hat{\theta}_B$ – МНК-оценки вектора параметров θ , вычисленные с использованием выборок W, A и B соответственно и удовлетворяющие системе ограничений, которая задает конкретную структуру.

Тогда критерии можно представить в следующем виде:

- симметричный и несимметричный критерии регулярности

$$\text{Reg} = (Y_A - X_A \hat{\theta}_B)^T (Y_A - X_A \hat{\theta}_B) + (Y_B - X_B \hat{\theta}_A)^T (Y_B - X_B \hat{\theta}_A), \quad (5.8)$$

$$\text{Reg}(B) = (Y_B - X_B \hat{\theta}_A)^T (Y_B - X_B \hat{\theta}_A); \quad (5.9)$$

- симметричный и несимметричный критерии стабильности

$$\text{Stab} = (Y - X \hat{\theta}_A)^T (Y - X \hat{\theta}_A) + (Y - X \hat{\theta}_B)^T (Y - X \hat{\theta}_B), \quad (5.10)$$

$$\text{Stab}(A) = (Y - X \hat{\theta}_A)^T (Y - X \hat{\theta}_A); \quad (5.11)$$

- симметричный и несимметричный критерии непротиворечивости (минимума смещения решений)

$$NC = (\hat{\theta}_A - \hat{\theta}_B)^T X^T X (\hat{\theta}_A - \hat{\theta}_B), \quad (5.12)$$

$$NC(B) = (\hat{\theta}_A - \hat{\theta}_B)^T X_B^T X_B (\hat{\theta}_A - \hat{\theta}_B); \quad (5.13)$$

- симметричный и несимметричный критерии вариативности (абсолютно помехоустойчивые критерии)

$$V = (\hat{\theta}_W - \hat{\theta}_A)^T X^T X (\hat{\theta}_B - \hat{\theta}_W), \quad (5.14)$$

$$V(B) = (\hat{\theta}_W - \hat{\theta}_A)^T X_B^T X_B (\hat{\theta}_B - \hat{\theta}_W). \quad (5.15)$$

Заметим, что любой симметричный критерий равен сумме двух несимметричных:

$$CR = CR(A) + CR(B).$$

Перечисленные критерии отражают следующие требования к структуре модели [39]: критерий регулярности – хорошие экстраполирующие свойства модели; критерий стабильности – хорошие аппроксимирующие и экстраполирующие свойства; критерий непротиворечивости – минимальное различие между оценками выхода моделей с оценками параметров, полученными по выборкам A и B ; критерий вариативности – максимальную согласованность оценок выхода модели при оценках параметров, полученных по выборкам W , A и B .

При рассмотрении многооткликовых моделей будем использовать представление вида (1.19), в том числе и для общей модели (1.38).

Для «однооткликовых» критериев (5.8) – (5.15) введем их матричные аналоги – матрицы размерности $l \times l$. Введем обозначения: $\hat{\Theta}_W, \hat{\Theta}_A, \hat{\Theta}_B$ – оценки матрицы параметров Θ , которые вычислены с использованием выборок W, A, B соответственно и удовлетворяют системе ограничений, задающей конкретную структуру. Тогда матричные аналоги критериев задаются теми же формулами, что и одномерные критерии, с заменой $\hat{\theta}_W, \hat{\theta}_A, \hat{\theta}_B$ на $\hat{\Theta}_W, \hat{\Theta}_A, \hat{\Theta}_B$.

В качестве критериев будем использовать свертки матричных аналогов. Предлагаются следующие свертки: определитель (D), след (A), максимальное собственное число (E), максимальный диагональный элемент (M), произведение диагональных элементов (P), след с взвешиванием $\text{tr} \Sigma^{-1} \text{MA}$, где MA – любой из введенных матричных аналогов критериев. В скобках указан префикс, добавляемый к обозначению критерия при упоминании конкретной свертки матричного аналога этого критерия. Последней свертке поставим в соответствие префиксы R или Z в зависимости от того, используется известное значение матрицы Σ или ее оценка при полной модели.

Матричные аналоги всех критериев, кроме критериев вариативности, по построению являются симметричными положительно полуопределенными матрицами, следовательно, для них значения всех сверток будут неотрицательными числами. Для формирования критериев вариативности будем применять R-, Z-, A-, M-, P-свертки, при этом в последних трех свертках необходимо использовать абсолютные значения элементов матриц.

Различные виды сверток можно рассматривать как различные принципы нахождения Парето-оптимальных решений в задачах векторной оптимизации [71, 104]. Пользуясь следом, мы придерживаемся принципа абсолютной уступки относительно критериев в виде собственных чисел матричных аналогов (кроме критериев вариативности). Определитель задает модель относительной уступки относительно тех же критериев, максимальное собственное число – модель минимакса. След, произведение диагональных элементов и максимальный диагональный элемент соответствуют принципам абсолютной уступки, относительной уступки и минимакса относительно критериев в виде диагональных элементов матричного аналога (в виде абсолютных значений диагональных элементов матриц для критериев вариативности).

Воспользовавшись возможностью преобразования рассматриваемых многооткликковых моделей в однооткликковые (для традиционной многооткликковой модели – (1.20), (1.22), для многооткликковой модели общего вида – (1.39), (1.40)), поиск оптимальной структуры можно проводить как сравнение структур однооткликковых моделей по «однооткликковым» критериям, представляющим собой свертки в виде следа матричного аналога или в виде следа матричного аналога с взвешиванием. При этом используются МНК-оценки параметров, а модели

(1.20), (1.39) интерпретируются соответственно как модели (1.22), (1.40), в которых априорно полагается $\Sigma = I_l$.

В рамках традиционной задачи выбора структуры известны некоторые соотношения критериев для однооткликowych моделей [39]. Введем подобные соотношения матричных аналогов критериев качества для традиционной многооткликовой модели в рамках обобщенной задачи выбора структуры.

Введем обозначения: $RSS(W)$, $RSS(A)$, $RSS(B)$ – матрицы остаточных сумм квадратов-произведений, вычисленные по выборке W , A , B соответственно с использованием оценок $\hat{\Theta}_W$, $\hat{\Theta}_A$, $\hat{\Theta}_B$; \hat{L}_W , \hat{L}_A , \hat{L}_B – значения матрицы множителей Лагранжа, полученные при оценивании параметров по выборкам W , A , B соответственно; C_i^B , где $i = 1, 2, 3, 4$, – матрицы из формул (1.28), (1.29), построенные по выборке B . Введем также обозначение

$$RSS_B(W) = (Y_B - X_B \hat{\Theta}_W)^T (Y_B - X_B \hat{\Theta}_W).$$

Приведем две вспомогательные формулы, следующие из определения матрицы (1.11):

$$X^T X C_1 = I_m - R^T C_3, \quad (5.16)$$

$$(X^T X C_2)^T = C_3 X^T X = -C_4 R. \quad (5.17)$$

Перейдем к выводу основных соотношений. Непосредственно из определения критериев следуют такие выражения:

$$\text{Stab}(A) = \text{Reg}(B) + RSS(A), \quad (5.18)$$

$$\text{Stab} = \text{Reg} + RSS(A) + RSS(B). \quad (5.19)$$

Рассмотрим матричный аналог критерия непротиворечивости:

$$NC(B) = \left[(Y_B - X_B \hat{\Theta}_B) - (Y_B - X_B \hat{\Theta}_A) \right]^T \left[(Y_B - X_B \hat{\Theta}_B) - (Y_B - X_B \hat{\Theta}_A) \right] =$$

$$\begin{aligned}
&= \text{RSS}(B) + \text{Reg}(B) - Y_B - X_B \hat{\Theta}_A^T Y_B - X_B \hat{\Theta}_B - \\
&\quad - Y_B - X_B \hat{\Theta}_B^T Y_B - X_B \hat{\Theta}_A . \quad (5.20)
\end{aligned}$$

С учетом (1.28) – (1.30), (5.16) и (5.17) осуществим следующие преобразования:

$$\begin{aligned}
&Y_B - X_B \hat{\Theta}_A^T Y_B - X_B \hat{\Theta}_B = Y_B^T Y_B - Y_B^T X_B \hat{\Theta}_B - \hat{\Theta}_A^T X_B^T Y_B + \\
&+ \hat{\Theta}_A^T X_B^T X_B \hat{\Theta}_B = Y_B^T Y_B - Y_B^T X_B \hat{\Theta}_B - \hat{\Theta}_A^T X_B^T Y_B + \hat{\Theta}_A^T X_B^T X_B \times \\
&\times C_1^B X_B^T Y_B + C_2^B U = Y_B^T Y_B - Y_B^T X_B \hat{\Theta}_B - \hat{\Theta}_A^T X_B^T Y_B + \hat{\Theta}_A^T I - R^T C_3^B \times \\
&\times X_B^T Y_B - \hat{\Theta}_A^T R^T C_4^B U = Y_B^T Y_B - Y_B^T X_B \hat{\Theta}_B - \hat{\Theta}_A^T X_B^T Y_B + \hat{\Theta}_A^T X_B^T Y_B - \\
&- \hat{\Theta}_A^T R^T C_3^B X_B^T Y_B + C_4^B U = Y_B^T Y_B - Y_B^T X_B \hat{\Theta}_B - U^T \hat{L}_B = \text{RSS}(B) . \quad (5.21)
\end{aligned}$$

Предпоследнее и последнее слагаемые в разложении (5.20) равны $-\text{RSS}(B)$ согласно (5.21). Тогда с учетом (5.18), (5.19) получаем следующие формулы:

$$\text{Reg}(B) = \text{NC}(B) + \text{RSS}(B), \quad (5.22)$$

$$\text{Stab}(A) = \text{NC}(B) + \text{RSS}(A) + \text{RSS}(B),$$

$$\text{Reg} = \text{NC} + \text{RSS}(A) + \text{RSS}(B),$$

$$\text{Stab} = \text{NC} + 2 \text{RSS}(A) + \text{RSS}(B) .$$

Преобразуем критерий вариативности с учетом формул для оценок параметров:

$$\begin{aligned}
 V(B) &= \left[Y_B - X_B \hat{\Theta}_A \quad Y_B - X_B \hat{\Theta}_W \right]^T \left[Y_B - X_B \hat{\Theta}_W \quad Y_B - X_B \hat{\Theta}_B \right] = \\
 &= Y_B - X_B \hat{\Theta}_A^T Y_B - X_B \hat{\Theta}_W^T Y_B - Y_B - X_B \hat{\Theta}_A^T Y_B - X_B \hat{\Theta}_B^T Y_B - \\
 &- Y_B - X_B \hat{\Theta}_W^T Y_B - X_B \hat{\Theta}_W^T Y_B + Y_B - X_B \hat{\Theta}_W^T Y_B - X_B \hat{\Theta}_B^T Y_B . \quad (5.23)
 \end{aligned}$$

По аналогии с (5.21) получим для первого слагаемого:

$$\begin{aligned}
 Y_B - X_B \hat{\Theta}_A^T Y_B - X_B \hat{\Theta}_W^T Y_B &= Y_B^T Y_B - Y_B^T X_B \hat{\Theta}_W - \hat{\Theta}_A^T X_B^T Y_B + \\
 + \hat{\Theta}_A^T X_B^T X_B \hat{\Theta}_W &= Y^T Y - Y_A^T Y_A - Y^T X - Y_A^T X_A \hat{\Theta}_W - \\
 - \hat{\Theta}_A^T X^T Y - X_A^T Y_A + \hat{\Theta}_A^T X^T X - X_A^T X_A \hat{\Theta}_W &= \\
 = Y^T Y - Y_A^T Y_A - Y^T X \hat{\Theta}_W + Y_A^T X_A \hat{\Theta}_W - \hat{\Theta}_A^T X^T Y + \\
 + \hat{\Theta}_A^T X_A^T Y_A + \hat{\Theta}_A^T X^T X \hat{\Theta}_W - \hat{\Theta}_A^T X_A^T X_A \hat{\Theta}_W &= \\
 = Y^T Y - Y_A^T Y_A - Y^T X \hat{\Theta}_W + Y_A^T X_A \hat{\Theta}_W - \hat{\Theta}_A^T X^T Y + \hat{\Theta}_A^T X_A^T Y_A + \hat{\Theta}_A^T X^T Y - \\
 - U^T \hat{L}_W - Y_A^T X_A \hat{\Theta}_W + \hat{L}_A^T U &= \text{RSS}(W) - \text{RSS}(A) .
 \end{aligned}$$

Рассмотрим другие слагаемые в разложении (5.23). Второе слагаемое согласно (5.21) равно $-\text{RSS}(B)$. Третье слагаемое по определению равно $-\text{RSS}_B(W)$. Последнее слагаемое согласно (5.21) равно $\text{RSS}(B)$.

В результате получаем:

$$V(B) = \text{RSS}(W) - \text{RSS}(A) - \text{RSS}_B(W) = \text{RSS}_A(W) - \text{RSS}(A),$$

$$V + \text{RSS}(A) + \text{RSS}(B) = \text{RSS}(W).$$

Воспользовавшись выведенными ранее формулами, получим:

$$V(B) + \text{Stab}(A) = \text{Reg}(B) + \text{RSS}_A(W),$$

$$V(B) + \text{Reg}(A) = \text{NC}(A) + \text{RSS}_A(W),$$

$$V + \text{Stab} = \text{Reg} + \text{RSS}(W), \quad (5.24)$$

$$V + \text{Reg} = \text{NC} + \text{RSS}(W).$$

Выведем из (5.24) еще одну формулу, полезную в дальнейшем:

$$\begin{aligned} \text{Stab}(A) - \text{RSS}(W) &= \text{Reg} - V - \text{Stab}(B) = \text{Reg}(B) - \text{RSS}(B) - V = \\ &= \text{NC}(B) - V. \end{aligned}$$

Для многооткликowej модели общего вида все полученные соотношения несправедливы, кроме (5.18), (5.19), так как при их выводе использовалось представление оценок неизвестных матриц в виде (1.28), (1.29). Однако справедливость соотношений сохраняется для сверток матричных аналогов критериев, связанных с преобразованием многооткликowej модели в однооткликowej, а именно: для следа в случае использования модели (1.39), для следа с взвешиванием в случае использования модели (1.40).

Каждый из введенных выше матричных аналогов критериев несет смысловую нагрузку желательных свойств модели. Помимо этого, его можно рассматривать как матрицу остаточных сумм квадратов-произведений некоторой вспомогательной модели с введенными на параметры дополнительными ограничениями специального вида, либо как одно из слагаемых такой матрицы, либо как сумму подобных матриц, соответствующих нескольким моделям [71].

Рассмотрим традиционную многооткликтовую модель. Обозначим через G матрицу остаточных сумм квадратов-произведений вспомогательной модели, через H – изменение G , вызванное введением в модель дополнительных ограничений; если моделей несколько, то у матриц появляется индекс.

В качестве вспомогательных моделей выступают исходная модель (1.19), аналогичные модели, описывающие отдельно выборки A и B , а также модель вида

$$Y = \begin{bmatrix} X_A & 0 \\ 0 & X_B \end{bmatrix} \begin{bmatrix} \Theta_1 \\ \Theta_2 \end{bmatrix} + E. \quad (5.25)$$

Заметим также, что параметры всех вспомогательных моделей удовлетворяют системе ограничений вида (1.25). Перечень моделей, дополнительных ограничений и интерпретация получаемых матриц даны в табл. 5.1.

В общей многооткликтовой модели (1.38) указанная интерпретация возможна для сверток, связанных с преобразованием многооткликтовой модели в однооткликтовую. При этом параметры вспомогательных моделей должны удовлетворять ограничениям вида (1.9). Однако для матричных аналогов критериев регулярности и стабильности справедлива интерпретация критериев как матриц остаточных сумм квадратов-произведений модели с дополнительными ограничениями.

Предположим, что ошибки наблюдений имеют многомерное нормальное распределение. В качестве критерия выбора структуры можно использовать взятый с противоположным знаком логарифм функции правдоподобия для вспомогательной модели со специальными ограничениями. Тогда критерий (с точностью до константных слагаемых и сомножителей) представим в виде

$$CR = \text{tr} \Sigma^{-1} G + H + N_S \ln \det \Sigma, \quad (5.26)$$

где N_S – число наблюдений, используемое во вспомогательной модели.

В случае нескольких вспомогательных моделей следует взять сумму критериев (5.26) (как интегральный критерий). Для построения

критериев типа непротиворечивости и вариативности можно использовать разность значений взятых с противоположным знаком логарифмов функции правдоподобия для вспомогательной модели со специальными ограничениями и без них.

Таблица 5.1

Интерпретация матричных аналогов критериев

Модель	Ограничения	Интерпретация
$Y_B = X_B \Theta + E_B$	$\Theta = \hat{\Theta}_A$	$\text{Reg}(B) = G + H$ $\text{NC}(B) = H$
(1.19)	$\Theta = \hat{\Theta}_A$	$\text{Stab}(A) = G + H$ $\text{NC}(B) - V = H$
$Y_A = X_A \Theta + E_A$	$\Theta = \hat{\Theta}_W$	$V(B) = H$
(5.25)	$\begin{bmatrix} \Theta_1 \\ \Theta_2 \end{bmatrix} = \begin{bmatrix} \hat{\Theta}_B \\ \hat{\Theta}_A \end{bmatrix}$	$\text{Reg} = G + H$ $\text{NC} = H$
Две модели $Y = X \Theta_i + E, \quad i = 1, 2$	$\Theta_1 = \hat{\Theta}_A,$ $\Theta_2 = \hat{\Theta}_B$	$\text{Stab} = G_1 + H_1 + G_2 + H_2$ $\text{NC} - 2V = H_1 + H_2$
(5.25)	$\begin{bmatrix} \Theta_1 \\ \Theta_2 \end{bmatrix} = \begin{bmatrix} \hat{\Theta}_W \\ \hat{\Theta}_W \end{bmatrix}$	$V = H$

Дадим интерпретацию D- и R-сверток с точки зрения рассмотренного подхода. Пусть матрица Σ известна, тогда второе слагаемое (5.26) для всех структур будет одинаковым. Отбрасывая его, получаем R-свертку. Пусть матрица Σ неизвестна. Подставим в выражение (5.26) значение ММП-оценки $\hat{\Sigma}$, равно

$$\hat{\Sigma} = \frac{1}{N_S} G + H .$$

Получим

$$CR = lN_S + N_S \ln \det \hat{\Sigma}.$$

После исключения константных слагаемых и сомножителей получим свертку в виде $\ln \det G + H$. Поскольку данная свертка не является линейной функцией матричного параметра, для симметричного критерия стабильности и критериев непротиворечивости и вариативности в результате данной интерпретации получаются представления, отличные от рассмотренных выше.

В [92] для однооткликowej модели введен нормированный критерий регулярности таким образом, что его математическое ожидание равно дисперсии ошибок наблюдений, умноженной на объем выборки $B - N_B$. Введем матричный аналог нормированного критерия регулярности для традиционной многооткликowej модели без ограничений на параметры и получим формулы его связи с матричными аналогами других критериев.

Вычислим математическое ожидание матричного аналога несимметричного критерия регулярности $\text{Reg}(B)$. Для этого представим матричный аналог в канонической форме [39]:

$$\text{Reg}(B) = Y^T D_R Y,$$

где $D_R = \begin{bmatrix} P^T P & -P^T \\ -P & I_{N_B} \end{bmatrix}$, $P = X_B X_A^T X_A^{-1} X_A^T$. Тогда можно записать

$$\mathbf{E}\text{Reg}(B) = U^T D_R U + \Sigma \text{tr} D_R = U_B - P U_A^T U_B - P U_A + \Sigma \text{tr} D_R,$$

где $U = \mathbf{E}Y$, $U_A = \mathbf{E}Y_A$, $U_B = \mathbf{E}Y_B$.

В предположении адекватности модели, когда используемая структура включает в себя истинную, первое слагаемое можно положить равным нулю, тогда

$$\mathbf{E}\text{Reg}(B) = \Sigma \text{tr} \left[I_{N_B} + X_B X_A^T X_A^{-1} X_B^T \right]$$

и величина $\frac{1}{N_B} \text{Reg}(B)$ является смещенной оценкой для Σ .

Введем в рассмотрение матрицу

$$\text{Nreg}(B) = Y_B - X_B \hat{\Theta}_A^T \left[I_{N_B} + X_B X_A^T X_A^{-1} X_B^T \right]^{-1} Y_B - X_B \hat{\Theta}_A,$$

являющуюся матричным аналогом нормированного критерия регулярности.

Осуществим преобразование матрицы $\text{Nreg}(B)$. Воспользуемся частным случаем формулы пересчета обратной матрицы [14]

$$I_n + AB^{-1} = I_n - A(I_r + BA^{-1}B),$$

где A , B – матрицы размерностей $n \times r$ и $r \times n$ соответственно и все записанные обратные матрицы существуют. Получим

$$\left[I_{N_B} + X_B X_A^T X_A^{-1} X_B^T \right]^{-1} = I_{N_B} - X_B X^T X^{-1} X_B^T.$$

Тогда справедливо

$$\text{Nreg}(B) = \text{Reg}(B) - X_B^T Y_B - X_B^T X_B \hat{\Theta}_A^T X^T X^{-1} X_B^T Y_B - X_B^T X_B \hat{\Theta}_A.$$

Преобразуем вычитаемое, которое обозначим Δ :

$$\begin{aligned} \Delta = & Y_B^T X_B X^T X^{-1} X_B^T Y_B - \hat{\Theta}_A^T X_B^T X_B X^T X^{-1} X_B^T Y_B - \\ & - Y_B^T X_B X^T X^{-1} X_B^T X_B \hat{\Theta}_A + \hat{\Theta}_A^T X_B^T X_B X^T X^{-1} X_B^T X_B \hat{\Theta}_A = \end{aligned}$$

$$\begin{aligned}
&= Y^T X - Y_A^T X_A \quad X^T X^{-1} \quad X^T Y - X_A^T Y_A - \\
&\quad - \hat{\Theta}_A^T \quad X^T X - X_A^T X_A \quad X^T X^{-1} \quad X^T Y - X_A^T Y_A - \\
&\quad - Y^T X - Y_A^T X_A \quad X^T X^{-1} \quad X^T X - X_A^T X_A \quad \hat{\Theta}_A + \\
&\quad + \hat{\Theta}_A^T \quad X^T X - X_A^T X_A \quad X^T X^{-1} \quad X^T X - X_A^T X_A \quad \hat{\Theta}_A .
\end{aligned}$$

Раскрывая скобки и приводя подобные слагаемые, получим

$$\begin{aligned}
\Delta &= \hat{\Theta}_W^T X^T Y - \hat{\Theta}_A^T X^T Y - Y^T X \hat{\Theta}_A + \hat{\Theta}_A^T X^T X \hat{\Theta}_A = \\
&= \hat{\Theta}_W^T X^T X \hat{\Theta}_W - \hat{\Theta}_A^T X^T X \hat{\Theta}_W - \hat{\Theta}_W^T X^T X \hat{\Theta}_A + \hat{\Theta}_A^T X^T X \hat{\Theta}_A = \\
&= \hat{\Theta}_W - \hat{\Theta}_A \quad X^T X \quad \hat{\Theta}_W - \hat{\Theta}_A .
\end{aligned}$$

В результате для матрицы $\text{Nreg}(B)$ можно записать

$$\text{Nreg}(B) = \text{Reg}(B) - \hat{\Theta}_W - \hat{\Theta}_A \quad X^T X \quad \hat{\Theta}_W - \hat{\Theta}_A .$$

Учитывая представление (5.22), получим

$$\text{Nreg}(B) = \text{RSS}(B) + \text{NC}(B) - \hat{\Theta}_W - \hat{\Theta}_A \quad X^T X \quad \hat{\Theta}_W - \hat{\Theta}_A .$$

Преобразуя величину $\text{NC}(B)$, получим

$$\begin{aligned}
\text{NC}(B) &= \hat{\Theta}_A - \hat{\Theta}_B \quad X_B^T X_B \quad \hat{\Theta}_A - \hat{\Theta}_B = \hat{\Theta}_A^T X_B^T X_B \hat{\Theta}_A + \hat{\Theta}_A^T X_B^T Y_B + \\
&\quad + Y_B^T X_B \hat{\Theta}_A + \hat{\Theta}_B^T X_B^T X_B \hat{\Theta}_B = \hat{\Theta}_A^T \quad X^T X - X_A^T X_A \quad \hat{\Theta}_A +
\end{aligned}$$

$$\begin{aligned}
& + \hat{\Theta}_A^T X^T Y - X_A^T Y_A + Y^T X - Y_A^T X_A \hat{\Theta}_A + \hat{\Theta}_B^T X_B^T X_B \hat{\Theta}_B = \\
& = \hat{\Theta}_A^T X^T X \hat{\Theta}_A - \hat{\Theta}_A^T X^T Y - Y^T X \hat{\Theta}_A + Y_A^T X_A \hat{\Theta}_A + Y_B^T X_B \hat{\Theta}_B.
\end{aligned}$$

Тогда

$$\begin{aligned}
\text{NC}(B) - \Delta & = Y_A^T X_A \hat{\Theta}_A + Y_B^T X_B \hat{\Theta}_B - Y^T X \hat{\Theta}_W = Y^T Y - Y_A^T Y_A - Y_B^T Y_B + \\
& + Y_A^T X_A \hat{\Theta}_A + Y_B^T X_B \hat{\Theta}_B - Y^T X \hat{\Theta}_W = \text{RSS}(W) - \text{RSS}(A) - \text{RSS}(B) = V.
\end{aligned}$$

В результате для $\text{Nreg}(B)$ справедливы следующие представления:

$$\text{Nreg}(B) = \text{RSS}(W) - \text{RSS}(A), \quad (5.27)$$

$$\text{Nreg}(B) = \text{RSS}(B) + V.$$

Согласно последней формуле матричный аналог нормированного критерия регулярности можно представить в виде суммы матрицы остаточных сумм квадратов-произведений по выборке B , являющейся мерой подгонки под выборку B , и матричного аналога симметричного критерия вариативности, являющегося мерой согласованности модели на выборках A , B , W .

Вычисление $\text{Nreg}(B)$ удобно производить по формуле (5.27), требующей лишь знания величин матриц остаточных сумм квадратов-произведений, полученных по выборкам W и A .

Найдем математическое ожидание матричного аналога нормированного критерия регулярности. Получим

$$\mathbf{E}\text{Nreg}(B) = \mathbf{E}\text{RSS}(W) - \mathbf{E}\text{RSS}(A) =$$

$$= U^T D_W U + \Sigma \text{tr} D_W - U^T D_A U - \Sigma \text{tr} D_A,$$

$$\text{где } D_W = I_N - X X^T X^{-1} X^T, \quad D_A = \begin{bmatrix} I_{N_A} - X_A X_A^T X_A^{-1} X_A^T & 0 \\ 0 & 0 \end{bmatrix},$$

N_A – объем выборки A .

В предположении адекватности модели первое и третье слагаемые можно положить равными нулю, тогда

$$\mathbf{E}N\text{reg}(B) = \Sigma \text{tr } D_W - D_A = \Sigma(N - p - N_A + p) = N_B \Sigma,$$

где p – количество регрессоров в рассматриваемой модели. В результате величина $\frac{1}{N_B} N\text{reg}(B)$ является несмещенной оценкой для Σ .

Для общей многооткликной модели полученные результаты справедливы для сверток, связанных с преобразованием многооткликной модели в однооткликную.

Укажем на связь нормированного критерия регулярности однооткликной модели с критерием Чоу [3], служащим для проверки гипотезы однородности двух выборок A и B (гипотеза предполагает равенство векторов параметров регрессии и дисперсий ошибок) в ситуации, когда объем выборки B не позволяет надежно оценить по ней параметры регрессии. Статистика для проверки гипотезы вычисляется по формуле

$$\frac{\text{rss}(W) - \text{rss}(A) / N_B}{\text{rss}(A) / (N_A - p)},$$

где $\text{rss}(W)$, $\text{rss}(A)$ – остаточные суммы квадратов соответствующих выборок. Легко увидеть, что в числителе статистики фигурирует величина нормированного критерия регулярности.

5.1.3. КРИТЕРИИ ТИПА СКОЛЬЗЯЩЕГО КОНТРОЛЯ

Критерии типа скользящего контроля (cross-validation) вычисляются по более сложным правилам.

Для однооткликной модели критерий скользящего контроля (скользящего экзамена, PRESS, усредненный критерий регулярности, leave-one-out cross-validation) [13, 105, 128, 188] вычисляется следующим образом: из выборки поочередно удаляется одно наблюдение, после чего вычисляются оценки параметров и квадрат ошибки прогноза в

удаленную точку; среднее значение квадрата ошибки прогноза по всем наблюдениям является величиной этого критерия. Критерий имеет вид

$$Sl = \frac{1}{N} \sum_{i=1}^N Y_i - X_i \hat{\theta}_{(i)} \quad Y_i - X_i \hat{\theta}_{(i)} \quad , \quad (5.28)$$

где Y_i – i -й элемент вектора Y ; X_i – i -я строка матрицы X ; $\hat{\theta}_{(i)}$ – МНК-оценка вектора параметров модели при исключенном i -м наблюдении.

Для эффективного вычисления критерия скользящего контроля используется формула

$$Sl = \frac{1}{N} \sum_{i=1}^N \frac{1}{(1 - w_i)^2} Y_i - X_i \hat{\theta}_{(i)} \quad Y_i - X_i \hat{\theta}_{(i)} \quad ,$$

где $w_i = X_i (X^T X)^{-1} X_i^T$.

Обобщенный критерий скользящего контроля (leave- n_2 -out cross-validation) [188] вычисляется следующим образом: из выборки n_1 раз удаляется по n_2 наблюдений; каждый раз, после того как n_2 наблюдений удалены из выборки, по оставшимся наблюдениям оцениваются параметры модели и делается прогноз в удаленные точки; среднее значение квадрата ошибки прогноза является величиной данного критерия:

$$GSl = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} Y_{[i]} - X_{[i]} \hat{\theta}_{[i]} \quad Y_{[i]} - X_{[i]} \hat{\theta}_{[i]} \quad , \quad (5.29)$$

где $Y_{[i]}$ и $X_{[i]}$ – n_2 элементов вектора Y и n_2 строк матрицы X , исключенных из выборки при i -м удалении; $\hat{\theta}_{[i]}$ – МНК-оценка параметров модели при исключенных $Y_{[i]}$, $X_{[i]}$.

Для эффективного вычисления обобщенного критерия скользящего контроля может использоваться формула [188]

$$\begin{aligned} \text{GSI} = & \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} Y_{[i]} - X_{[i]} \hat{\theta}^T I_{n_2} - \\ & - X_{[i]} (X^T X)^{-1} X_{[i]}^T \quad Y_{[i]} - X_{[i]} \hat{\theta} . \end{aligned} \quad (5.30)$$

Имеется $C_N^{n_2}$ различных способов разделения выборки объема N на части, состоящие из $N - n_2$ и n_2 наблюдений. Поскольку это число может быть слишком большим, вычисление GSI при $n_1 = C_N^{n_2}$ может оказаться очень трудоемким. По этой причине используется ряд вариантов вычисления данного критерия при меньшем значении n_1 [188]. Например, в варианте критерия с применением метода Монте-Карло наблюдения из набора данных удаляются случайно (с возвращением или без него).

Обычный критерий скользящего контроля можно представить как один из вариантов вычисления обобщенного критерия скользящего контроля при $n_2 = 1$.

В [188] показано, что критерий скользящего контроля не состоятельный в том смысле, что вероятность выбора наилучшей прогнозирующей модели не стремится к единице при увеличении числа наблюдений, обычно сложность наилучшей модели завышается. В то же время обобщенный критерий скользящего контроля при некоторых условиях, среди которых $\frac{n_2}{N} \rightarrow 1$ при $N \rightarrow \infty$, обладает указанным свойством.

Для многооткликowej модели введем матричные аналоги критериев типа скользящего контроля, заменяя в формулах (5.28), (5.29) векторы $\hat{\theta}_{(i)}$, $\hat{\theta}_{[i]}$ на матрицы $\hat{\Theta}_{(i)}$, $\hat{\Theta}_{[i]}$, имеющие аналогичный смысл (Y_i , $Y_{[i]}$ являются теперь соответственно вектор-строкой и матрицей).

При интерпретации сверток, связанных с преобразованиями (1.20), (1.22), (1.39), (1.40), для этих критериев имеются некоторые особенности. Осуществив переход к однооткликowej модели, мы получаем «однооткликową» выборку. Далее вычисляем соответствующие критерии, используя одновременное исключение нескольких из $N!$ наблю-

дений «однооткликковой» выборки. Критерий скользящего контроля можно представить тогда как один из способов вычисления обобщенного критерия скользящего контроля для однооткликковой модели.

В то же время можно вычислить критерий скользящего контроля для одномерной модели типа (1.20), (1.22), (1.39), (1.40), удаляя по одному из Nl наблюдений «одномерной» выборки. Такой критерий назовем одномерным критерием скользящего контроля – USI.

Так как при преобразованиях (1.20), (1.22), (1.39), (1.40) количество наблюдений увеличивается в l раз, то величины А-, R-, Z-сверток критериев типа скользящего контроля должны быть уменьшены в l раз.

Получим аналог формулы (5.30) для матричного аналога обобщенного критерия скользящего контроля в случае традиционной многооткликковой модели с системой ограничений на параметры (1.25). Подобная формула будет справедлива для критерия GSI в однооткликковом случае (включая модели (1.20), (1.22), (1.39), (1.40)) и для критерия USI при использовании эквивалентной однооткликковой модели.

Обозначим $\hat{L}_{[i]}$ значения множителей Лагранжа, полученные в результате решения системы уравнений (1.27) после удаления из выборки i -й группы наблюдений.

При выводе удобно использовать формулу для пересчета обратной матрицы [14]:

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B C^{-1} + DA^{-1}B^{-1} DA^{-1}, \quad (5.31)$$

где A , C , B , D – матрицы размерностей $n \times n$, $r \times r$, $n \times r$, $r \times n$ соответственно и все записанные обратные матрицы существуют.

Получаем для оценок:

$$\begin{aligned} \begin{bmatrix} \hat{\Theta}_{[i]} \\ \hat{L}_{[i]} \end{bmatrix} &= \begin{bmatrix} X^T X - X_{[i]}^T X_{[i]} & R \\ R^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} X^T Y - X_{[i]}^T Y_{[i]} \\ U \end{bmatrix} = \\ &= \left(\begin{bmatrix} X^T X & R^T \\ R & 0 \end{bmatrix} - \begin{bmatrix} X_{[i]}^T \\ 0 \end{bmatrix} \begin{bmatrix} X_{[i]} & 0 \end{bmatrix} \right)^{-1} \left(\begin{bmatrix} X^T Y \\ U \end{bmatrix} - \begin{bmatrix} X_{[i]}^T Y_{[i]} \\ 0 \end{bmatrix} \right). \quad (5.32) \end{aligned}$$

Преобразуем первый сомножитель (5.32) по формуле (5.31). В предположении, что матрица $I - X_{[i]}C_1X_{[i]}^T$, где I – единичная матрица соответствующей размерности, является невырожденной, получим

$$\begin{aligned}
 & \left(\begin{bmatrix} X^T X & R^T \\ R & 0 \end{bmatrix} - \begin{bmatrix} X_{[i]}^T \\ 0 \end{bmatrix} \begin{bmatrix} X_{[i]} & \vdots & 0 \end{bmatrix} \right)^{-1} = \begin{bmatrix} X^T X & R^T \\ R & 0 \end{bmatrix}^{-1} + \\
 & + \begin{bmatrix} X^T X & R^T \\ R & 0 \end{bmatrix}^{-1} \begin{bmatrix} X_{[i]}^T \\ 0 \end{bmatrix} \left\{ I - \begin{bmatrix} X_{[i]} & \vdots & 0 \end{bmatrix} \begin{bmatrix} X^T X & R^T \\ R & 0 \end{bmatrix}^{-1} \begin{bmatrix} X_{[i]}^T \\ 0 \end{bmatrix} \right\}^{-1} \times \\
 & \times \begin{bmatrix} X_{[i]} & \vdots & 0 \end{bmatrix} \begin{bmatrix} X^T X & R^T \\ R & 0 \end{bmatrix}^{-1} = \begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix} + \begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix} \begin{bmatrix} X_{[i]}^T \\ 0 \end{bmatrix} \times \\
 & \times \left\{ I - \begin{bmatrix} X_{[i]} & \vdots & 0 \end{bmatrix} \begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix} \begin{bmatrix} X_{[i]}^T \\ 0 \end{bmatrix} \right\}^{-1} \begin{bmatrix} X_{[i]} & \vdots & 0 \end{bmatrix} \begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix} = \\
 & = \begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix} + \begin{bmatrix} C_1 X_{[i]}^T \\ C_3 X_{[i]}^T \end{bmatrix} \left[I - X_{[i]} C_1 X_{[i]}^T \right]^{-1} \times \\
 & \times \begin{bmatrix} X_{[i]} C_1 & \vdots & X_{[i]} C_2 \end{bmatrix} = \begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix} + \\
 & + \begin{bmatrix} C_1 X_{[i]}^T \left[I - X_{[i]} C_1 X_{[i]}^T \right]^{-1} X_{[i]} C_1 & C_1 X_{[i]}^T \left[I - X_{[i]} C_1 X_{[i]}^T \right]^{-1} X_{[i]} C_2 \\ C_3 X_{[i]}^T \left[I - X_{[i]} C_1 X_{[i]}^T \right]^{-1} X_{[i]} C_1 & C_3 X_{[i]}^T \left[I - X_{[i]} C_1 X_{[i]}^T \right]^{-1} X_{[i]} C_2 \end{bmatrix}.
 \end{aligned}$$

Тогда для ошибки прогноза имеем следующее:

$$\begin{aligned}
 Y_{[i]} - X_{[i]}\hat{\Theta}_{[i]} &= Y_{[i]} - X_{[i]}\left(C_1 \parallel C_2 + \right. \\
 &+ \left. \left[C_1 X_{[i]}^T \left[I - X_{[i]} C_1 X_{[i]}^T \right]^{-1} X_{[i]} C_1 \parallel C_1 X_{[i]}^T \left[I - X_{[i]} C_1 X_{[i]}^T \right]^{-1} X_{[i]} C_2 \right] \right) \times \\
 &\times \left(\begin{bmatrix} X^T Y \\ U \end{bmatrix} - \begin{bmatrix} X_{[i]}^T Y_{[i]} \\ 0 \end{bmatrix} \right) = Y_{[i]} - X_{[i]} \left(\hat{\Theta}_W - C_1 X_{[i]}^T Y_{[i]} + C_1 X_{[i]}^T \times \right. \\
 &\times \left. \left[I - X_{[i]} C_1 X_{[i]}^T \right]^{-1} X_{[i]} \hat{\Theta}_W - C_1 X_{[i]}^T \left[I - X_{[i]} C_1 X_{[i]}^T \right]^{-1} X_{[i]} C_1 X_{[i]}^T Y_{[i]} \right) = \\
 &= Y_{[i]} - \left(I + X_{[i]} C_1 X_{[i]}^T \left[I - X_{[i]} C_1 X_{[i]}^T \right]^{-1} \right) X_{[i]} \hat{\Theta}_W + \\
 &+ X_{[i]} C_1 X_{[i]}^T \left(I + \left[I - X_{[i]} C_1 X_{[i]}^T \right]^{-1} X_{[i]} C_1 X_{[i]}^T \right) Y_{[i]} = \\
 &= \left[I - X_{[i]} C_1 X_{[i]}^T \right]^{-1} Y_{[i]} - X_{[i]} \hat{\Theta}_W .
 \end{aligned}$$

В результате получаем для матричного аналога

$$\text{GSI} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} Y_{[i]} - X_{[i]} \hat{\Theta}_W^T \left[I - X_{[i]} C_1 X_{[i]}^T \right]^{-2} Y_{[i]} - X_{[i]} \hat{\Theta}_W .$$

Для матричного аналога критерия скользящего контроля данная формула существенно упрощается, поскольку выражение в квадратных скобках является скаляром:

$$SI = \frac{1}{N} \sum_{i=1}^N \frac{1}{1 - X_i C_1 X_i^T} \frac{1}{2} (Y_i - X_i \hat{\Theta}_W)^T (Y_i - X_i \hat{\Theta}_W) .$$

Подобная формула будет справедлива для критерия скользящего контроля в однооткликовом случае, а значит, и для критерия USI при использовании однооткликowej модели (1.22).

Для традиционной многооткликowej модели в условиях полноты ранга матрицы X критерии USI и свертки матрицы SI, связанные с преобразованиями (1.20) и (1.22), эквивалентны.

Покажем эквивалентность критериев для случая преобразования (1.22). Обозначим через $Z_{(i)} = I_l \otimes X_i$ $l \times lm$ -матрицу, состоящую из l строк матрицы $I_l \otimes X$, соответствующих i -му наблюдению вектора откликов; s_j – j -ю строку матрицы S^{-1} , фигурирующей в формуле (1.22); $\hat{\theta}$ – оценку вектора параметров соответствующей модели.

Вначале заметим, что в рассматриваемом случае для матрицы C_1 из формулы (1.28) справедливо представление

$$C_1 = X^T X^{-1} - X^T X^{-1} R^T \left[R X^T X^{-1} R^T \right]^{-1} R X^T X^{-1} . \quad (5.33)$$

При помощи соотношения (5.33), а также учитывая представление (1.26), найдем вид матрицы C_1 , фигурирующей в системе уравнений (1.12) – (1.13), используемой для нахождения оценок параметров в модели (1.22) с ограничениями (1.26):

$$\left[Z^T (\Sigma^{-1} \otimes I_N) Z \right]^{-1} - \left[Z^T (\Sigma^{-1} \otimes I_N) Z \right]^{-1} \left[I_l \otimes R^T \right] \times$$

$$\begin{aligned}
& \times \left\{ I_l \otimes R \left[Z^T (\Sigma^{-1} \otimes I_N) Z \right]^{-1} \left[I_l \otimes R^T \right] \right\}^{-1} I_l \otimes R \left[Z^T (\Sigma^{-1} \otimes I_N) Z \right]^{-1} = \\
& = \Sigma \otimes X^T X^{-1} - \left[\Sigma \otimes X^T X^{-1} \right] \left[I_l \otimes R^T \right] \times \\
& \times \left\{ I_l \otimes R \left[\Sigma \otimes X^T X^{-1} \right] \left[I_l \otimes R^T \right] \right\}^{-1} I_l \otimes R \left[\Sigma \otimes X^T X^{-1} \right] = \\
& = \Sigma \otimes X^T X^{-1} - \left[\Sigma \otimes X^T X^{-1} \right] \left[I_l \otimes R^T \right] \left[\Sigma \otimes R X^T X^{-1} R^T \right]^{-1} \times \\
& \times I_l \otimes R \left[\Sigma \otimes X^T X^{-1} \right] = \\
& = \Sigma \otimes \left\{ X^T X^{-1} - X^T X^{-1} R^T \left[R X^T X^{-1} R^T \right]^{-1} R X^T X^{-1} \right\} = \\
& = \Sigma \otimes C_1,
\end{aligned}$$

где $Z = I_l \otimes X$; C_1 – матрица из формулы (1.28).

Рассмотрим слагаемые критерия USI, соответствующие i -му наблюдению вектора откликов:

$$\sum_{j=1}^l \frac{s_j Y_i^T - s_j Z_{(i)} \hat{\theta}^2}{\left[1 - s_j Z_{(i)} \Sigma \otimes C_1 Z_{(i)}^T s_j^T \right]^2}. \quad (5.34)$$

Преобразуем знаменатель, воспользовавшись свойствами кронекера произведения:

$$\begin{aligned} \left[1 - s_j Z_{(i)} \Sigma \otimes C_1 Z_{(i)}^T s_j^T \right]^2 &= \left[1 - s_j I_l \otimes X_i \Sigma \otimes C_1 I_l \otimes X_i^T s_j^T \right]^2 = \\ &= \left[1 - s_j \Sigma \otimes X_i C_1 X_i^T s_j^T \right]^2 = \left[1 - s_j \Sigma s_j^T X_i C_1 X_i^T \right]^2. \end{aligned}$$

Последнее равенство справедливо, так как величина $X_i C_1 X_i^T$ — скаляр. Величина $s_j \Sigma s_j^T$ является j -м диагональным элементом матрицы $S^{-1} \Sigma (S^T)^{-1} = I_l$, поэтому она равна 1. В результате знаменатель (5.34) не зависит от j и может быть вынесен за знак суммирования. Тогда рассмотрим сумму:

$$\begin{aligned} \sum_{j=1}^l s_j Y_i^T - s_j Z_{(i)} \hat{\theta} \quad ^2 &= \text{tr} \left[S^{-1} Y_i^T - Z_{(i)} \hat{\theta} \quad Y_i^T - Z_{(i)} \hat{\theta} \quad ^T S^T \quad ^{-1} \right] = \\ &= \text{tr} \left[\Sigma^{-1} Y_i^T - Z_{(i)} \hat{\theta} \quad Y_i^T - Z_{(i)} \hat{\theta} \quad ^T \right] = \\ &= \text{tr} \left[\Sigma^{-1} Y_i - X_i \hat{\Theta}_W \quad ^T Y_i - X_i \hat{\Theta}_W \right]. \end{aligned}$$

В результате получаем величину одномерного критерия скользящего контроля, совпадающую с величиной свертки матричного аналога критерия скользящего контроля, связанной с преобразованием (1.22):

$$\text{USI} = \frac{1}{Nl} \sum_{i=1}^N \frac{\text{tr} \Sigma^{-1} Y_i - X_i \hat{\Theta}_W \quad ^T Y_i - X_i \hat{\Theta}_W}{1 - X_i C_1 X_i^T \quad ^2} =$$

$$= \frac{1}{l} \operatorname{tr} \left[\Sigma^{-1} \frac{1}{N} \sum_{i=1}^N \frac{1}{1 - X_i C_1 X_i^T} Y_i - X_i \hat{\Theta}_W^T \quad Y_i - X_i \hat{\Theta}_W \right].$$

Совпадение одномерного критерия скользящего контроля со следом матричного аналога критерия скользящего контроля в случае преобразования (1.20) показывается как частный случай проведенных рассуждений при $\Sigma = I_l$.

Матричные аналоги критерия скользящего контроля и обобщенного критерия скользящего контроля можно представить как матрицы остаточных сумм квадратов-произведений некоторой вспомогательной модели с дополнительными ограничениями на параметры специального вида.

В случае критерия скользящего контроля для традиционной многооткликной регрессии получаем вспомогательную модель

$$Y = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & X_N \end{bmatrix} \begin{bmatrix} \Theta_1 \\ \Theta_2 \\ \dots \\ \Theta_N \end{bmatrix} + E, \quad R\Theta_i = U, \quad i = 1, \dots, N,$$

и дополнительные ограничения

$$\left[\Theta_1^T \mid \dots \mid \Theta_N^T \right]^T = \left[\hat{\Theta}_{(1)}^T \mid \dots \mid \hat{\Theta}_{(N)}^T \right]^T.$$

Тогда имеем $G + H = N$ SI.

Вычислим обобщенный критерий скользящего контроля так, чтобы каждое наблюдение было удалено из выборки ровно один раз; в том случае, если N не делится без остатка на n_2 , произведем $n_1 = [N/n_2] + 1$ удалений, причем на последнем шаге удалим меньшее количество наблюдений, равное $N - [N/n_2]n_2$, где $[\cdot]$ – операция взятия

целой части числа; в результате множитель $\frac{1}{n_1 n_2}$ в формуле критерия следует заменить на $\frac{1}{N}$. Тогда вспомогательная модель имеет вид

$$\begin{bmatrix} Y_{[1]} \\ Y_{[2]} \\ \dots \\ Y_{[n_1]} \end{bmatrix} = \begin{bmatrix} X_{[1]} & 0 & \dots & 0 \\ 0 & X_{[2]} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & X_{[n_1]} \end{bmatrix} \begin{bmatrix} \Theta_1 \\ \Theta_2 \\ \dots \\ \Theta_{n_1} \end{bmatrix} + \begin{bmatrix} E_{[1]} \\ E_{[2]} \\ \dots \\ E_{[n_1]} \end{bmatrix}, \quad R\Theta_i = U, \quad i=1, \dots, n_1,$$

где $E_{[i]}$ – ошибки наблюдений, исключенных при i -м удалении. Дополнительные ограничения имеют вид

$$\begin{bmatrix} \Theta_1^T & \dots & \Theta_{n_1}^T \end{bmatrix}^T = \begin{bmatrix} \hat{\Theta}_{[1]}^T & \dots & \hat{\Theta}_{[n_1]}^T \end{bmatrix}^T.$$

В этом случае имеем $G + H = N$ GSI.

Аналогичную интерпретацию можно провести в случае общей многооткликowej модели для матриц SI, GSI и для значений критерия USI.

5.1.4. КРИТЕРИИ, НЕ ИСПОЛЬЗУЮЩИЕ ЭКЗАМЕНАЦИОННУЮ ВЫБОРКУ

Для выбора однооткликowej регрессионной модели разработано большое количество критериев, не использующих экзаменационную выборку [2, 85, 102, 106]. Некоторые из них распространены на случай традиционной многооткликowej модели.

Остаточную сумму квадратов однооткликowej модели обозначим rss , матрицу остаточных сумм квадратов-произведений для многооткликowej модели – RSS ; p – число регрессоров в рассматриваемой модели (для традиционной многооткликowej модели – число регрессоров в уравнении одного отклика).

Критерий Меллоуса. При решении традиционной задачи выбора структуры однооткликowej модели критерий определяется формулой [164]

$$C_p = \frac{\text{RSS}}{\sigma^2} + 2p - N$$

и построен исходя из справедливости соотношения

$$\mathbf{E}C_p = \mathbf{E} \left\{ \frac{1}{\sigma^2} \sum_{i=1}^N \hat{Y}_i - \mathbf{E}Y_i \right\}^2, \quad (5.35)$$

где \hat{Y}_i – оценка значения отклика при исследуемой структуре в i -м наблюдении, Y_i – измеренное значение отклика в i -м наблюдении.

Для традиционной многооткликowej модели вводится матричный аналог критерия [164]

$$\text{MC} = \Sigma^{-1} \text{RSS} - (N - 2p)I_1.$$

Для выбора структуры предлагается использовать его свертки, например, в виде следа или максимального собственного числа.

В том случае, если значения σ^2 и Σ неизвестны, используются их оценки при полной модели. Критерий в виде следа матрицы MC будем обозначать RMC или ZMC в зависимости от того, известно значение Σ или оно оценивается.

Рассмотрим традиционную многооткликowej модель с ограничениями на параметры (1.25). Получим для нее матричный аналог критерия Меллоуса.

$$\text{Обозначим } \text{SSB} = \sum_{i=1}^N \mathbf{E} \hat{Y}_i - \mathbf{E}Y_i \quad \mathbf{E} \hat{Y}_i - \mathbf{E}Y_i \quad T.$$

Введем матрицу

$$\Delta = \mathbf{E} \left\{ \Sigma^{-1} \sum_{i=1}^N \hat{Y}_i - \mathbf{E}Y_i \quad T \quad \hat{Y}_i - \mathbf{E}Y_i \right\},$$

которая является матричным аналогом показателя, стоящего в правой части равенства (5.35). Преобразуем матрицу Δ :

$$\Delta = \Sigma^{-1} \left[\sum_{i=1}^N \text{cov} \hat{Y}_i + \sum_{i=1}^N \mathbf{E} \hat{Y}_i - \mathbf{E} Y_i \quad \mathbf{E} \hat{Y}_i - \mathbf{E} Y_i \right] = \Sigma^{-1} \sum_{i=1}^N \text{cov} \hat{Y}_i + \Sigma^{-1} \text{SSB} .$$

Ковариация двух элементов вектора \hat{Y}_i равна

$$\text{cov} \hat{Y}_{ij}, \hat{Y}_{ik} = \mathbf{E} \left[\hat{Y}_{ij} - \mathbf{E} \hat{Y}_{ij} \quad \hat{Y}_{ik} - \mathbf{E} \hat{Y}_{ik} \right] = \sum_{jk} X_i C_1 X_i^T .$$

Тогда для матрицы Δ справедливо

$$\Delta = \Sigma^{-1} \sum_{i=1}^N \sum X_i C_1 X_i^T + \Sigma^{-1} \text{SSB} = I_l \text{tr} X C_1 X^T + \Sigma^{-1} \text{SSB} .$$

Из определения матрицы (1.11) следует $C_3 R^T = I_u$. Воспользовавшись этим результатом, а также свойствами следа матрицы и формулой (5.16), получим

$$\text{tr} X C_1 X^T = \text{tr} X^T X C_1 = \text{tr} I_m - \text{tr} C_3 R^T = m - \text{tr} I_u = m - u . \quad (5.36)$$

В результате справедливо

$$\Delta = (m - u) I_l + \Sigma^{-1} \text{SSB} .$$

Введем матричный аналог критерия Меллоуса вида

$$\text{MC} = \Sigma^{-1} \text{RSS} + 2(m - u) - N I_l .$$

Рассмотрим величину **ERSS**:

$$\begin{aligned} \text{ERSS} &= \mathbf{E} \left[Y - X \hat{\Theta}_W \quad Y - X \hat{\Theta}_W \right] = \mathbf{E} \left[Y^T Y - Y^T X \hat{\Theta}_W - U^T \hat{L}_W \right] = \\ &= \mathbf{E} \left[Y^T I_N - X C_1 X^T Y \right] - \mathbf{E} \left[Y^T \right] X C_2 U - U^T C_3 X^T \mathbf{E} Y - U^T C_4 U = \end{aligned}$$

$$= \Sigma \operatorname{tr}(I_N - XC_1X^T) + \mathbf{E} \left[Y^T \right] I_N - XC_1X^T \mathbf{E} Y - \mathbf{E} \left[Y^T \right] XC_2U - \\ - U^T C_3X^T \mathbf{E} Y - U^T C_4U = \Sigma(N - m + u) + \text{SSB}.$$

Найдем математическое ожидание МС :

$$\mathbf{E}MC = \Sigma^{-1}\mathbf{E}RSS + 2(m - u) - N I_1 = (m - u)I_1 + \Sigma^{-1}\text{SSB}.$$

Таким образом, выполнено условие

$$\mathbf{E}MC = \Delta,$$

которое является многомерным аналогом соотношения (5.35), определяющего однооткликový критерий Меллоуса.

Для общей многооткликовой модели, воспользовавшись преобразованием (1.40), получим следующую модификацию критерия Меллоуса:

$$MC = \operatorname{tr} \Sigma^{-1}RSS + 2(m - u) - N I_1.$$

Критерий финальной ошибки прогнозирования (FPE), или статистика Ротмана. При решении традиционной задачи выбора структуры однооткликовой модели используется формула [103, 106, 184]

$$FPE = \operatorname{rss} \frac{N + p}{N - p}.$$

Для традиционной многооткликовой модели используется матрица

$$\text{MFPE} = \text{RSS} \frac{N + p}{N - p}, \quad (5.37)$$

а в качестве критерия – некоторая ее свертка, например, в [103] рассматривается детерминант матрицы (5.37).

Получим критерий FPE для традиционной многооткликовой модели с ограничениями на параметры.

Для одномерных моделей критерий FPE тесно связан с критерием MSEF. Введем матричный аналог критерия MSEF и преобразуем его по аналогии с [106]:

$$\begin{aligned}
 \text{MSEF} &= \frac{1}{N} \mathbf{E} \left\{ \sum_{i=1}^N \hat{Y}_i - Y_i \quad \hat{Y}_i - Y_i \right\} = \\
 &= \frac{1}{N} \sum_{i=1}^N \left\{ \text{cov } \hat{Y}_i - Y_i + \left[\mathbf{E} \hat{Y}_i - Y_i \right]^T \mathbf{E} \hat{Y}_i - Y_i \right\} = \\
 &= \frac{1}{N} \left\{ \sum_{i=1}^N \left[\Sigma + \text{cov } \hat{Y}_i \right] + \sum_{i=1}^N \left[\mathbf{E} \hat{Y}_i - Y_i \right]^T \mathbf{E} \hat{Y}_i - Y_i \right\} = \\
 &= \frac{1}{N} \left[\sum_{i=1}^N \Sigma + \text{tr} X C_1 X^T + \text{SSB} \right] = \frac{1}{N} \left[\Sigma N + \text{tr} X C_1 X^T + \text{SSB} \right] = \\
 &= \frac{1}{N} \left[\Sigma N + m - u + \text{SSB} \right].
 \end{aligned}$$

Последнее равенство справедливо на основании формулы (5.36). По аналогии с одномерным случаем (см. [102]), заменив ковариационную матрицу Σ ее оценкой (1.31), положив матрицу SSB равной нулевой матрице и опустив сомножитель-константу, получим формулу для матричного аналога критерия FPE вида

$$\text{MFPE} = \text{RSS} \frac{N + m - u}{N - m + u}. \quad (5.38)$$

В качестве критериев следует использовать свертки матрицы (5.38).

Для многооткликowej модели общего вида, пользуясь преобразованием (1.40), получим

$$\text{MFPE} = \text{tr } \Sigma^{-1} \text{RSS} \frac{Nl + m - u}{Nl - m + u}.$$

Средний квадрат ошибки (MSE). Для традиционной многооткликковой модели с ограничениями на параметры, согласно (1.31), матричный аналог критерия имеет вид

$$\text{MSE} = \text{RSS} \frac{1}{N - m + u}.$$

Критерии могут быть построены в виде какой-либо свертки матрицы MSE – оценки ковариационной матрицы Σ . Свертка, связанная с преобразованием (1.22), имеет вид $\frac{1}{l} \text{tr} \Sigma^{-1} \text{MSE}$; в этом случае справедлива интерпретация свертки MSE как оценки дисперсии ошибок наблюдений однооткликковой модели, полученной путем преобразования (1.22).

Для решения обобщенной задачи выбора структуры в случае общей многооткликковой модели исходя из преобразования (1.40) можно использовать критерий в виде

$$\text{MSE} = \text{tr} \Sigma^{-1} \text{RSS} \frac{1}{Nl - m + u}.$$

Информационные критерии для моделей без ограничений на параметры строятся исходя из представления [4]

$$\text{IC} = 2L + \varphi(N)\gamma,$$

где L – логарифм функции правдоподобия, взятый с противоположным знаком; γ – число независимо определяемых параметров: lp для традиционной многооткликковой модели и p для однооткликковой модели и многооткликковой модели общего вида, параметры оцениваются по ММП.

Вид функции $\varphi(N)$ определяет конкретный критерий [4, 103, 132, 143, 180, 186, 189]:

- критерий Акаике (AIC) с $\varphi(N) = 2$;
- критерий Шварца–Риссанена (BIC) с $\varphi(N) = \ln N$;
- критерий Хеннана–Куинна (HQ) с $\varphi(N) = 2 \ln \ln N$;
- KIC с $\varphi(N) = 3$.

5.1.5. ВЫБОР СТРУКТУРЫ ПРИ НАЛИЧИИ КАЧЕСТВЕННЫХ И РАЗНОТИПНЫХ ФАКТОРОВ

При наличии качественных или разнотипных факторов, как и в случае регрессионных моделей, многооткликową зависимость можно строить в виде традиционных многооткликowych моделей, моделей псевдонезависимых регрессий или моделей общего вида (1.38).

Отличительной чертой рассматриваемых моделей является неполнота столбцового ранга матрицы X . В п. 1.3 рассмотрены подходы к решению задачи оценивания параметров модели (1.41). Рассмотрим, как могут быть использованы при выборе структуры многооткликовой модели подходы, основанные на применении факторизации матрицы X и идентифицирующих ограничений.

Первый подход аналогичен подходу, применяемому в одномерном случае [90]. Пусть уравнение зависимости имеет вид (1.41), и для получения единственного решения при оценивании параметров производится преобразование модели вида (1.42). В результате получается модель регрессионного типа с параметрами, интерпретируемыми как ФДО.

Другой подход к выбору структуры модели с качественными и разнотипными факторами основан на решении обобщенной задачи выбора структуры. В этом случае легко сформировать линейные ограничения на параметры, несущие смысловую нагрузку. Все фиктивные переменные, стоящие в уравнении (1.41) при главных эффектах уровней одного качественного фактора или при взаимодействиях отдельных факторов, и соответствующие им идентифицирующие ограничения должны одновременно присутствовать или отсутствовать в модели (будем говорить, что они составляют один блок). В дисперсионном и ковариационном анализе часто используются легко интерпретируемые идентифицирующие ограничения типа «сумма главных эффектов уровней одного качественного фактора равна нулю».

Заметим, что оба подхода к построению моделей являются не альтернативными, а дополняющими друг друга. Рассмотрим для примера преобразование (1.42) в модели без взаимодействий. С целью сокращения количества перебираемых структур можно объединить в один блок фиктивные переменные, стоящие при всех ФДО, образованных эффектами уровней одного качественного фактора.

Можно сформировать предположения в виде ограничений, левые части которых являются произвольными ФДО. Последняя возможность способна помочь в ряде случаев избавиться от неоднозначности результатов выбора структуры, связанных с неоднозначностью преобразования к модели (1.42).

Пример. Предполагается, что на два отклика аддитивно влияют два фактора α , β , имеющие три уровня варьирования, и один фактор γ , изменяющийся на двух уровнях. Используются следующие истинные значения для двух откликов:

$$\mu = 10,12, \quad \alpha_1 = -6,0, \quad \alpha_2 = 3,0, \quad \alpha_3 = 3,0, \quad \beta_1 = 0,4,$$

$$\beta_2 = 0,4, \quad \beta_3 = 0,-8, \quad \gamma_1 = 0,0, \quad \gamma_2 = 0,0,$$

где μ – аддитивная постоянная; α_i , β_i , γ_i – эффекты i -х уровней соответствующих факторов.

Если неполнота ранга матрицы X является следствием только способа кодировки качественных факторов, то преобразование модели (1.41) приводит к модели полного ранга, например, со следующими ФДО в качестве параметров:

$$\psi_1 = \mu + \alpha_3 + \beta_3 + \gamma_2, \quad \psi_2 = \alpha_1 - \alpha_3, \quad \psi_3 = \alpha_2 - \alpha_3,$$

$$\psi_4 = \beta_1 - \beta_3, \quad \psi_5 = \beta_2 - \beta_3, \quad \psi_6 = \gamma_1 - \gamma_2.$$

В случае традиционной многооткликовой модели элиминирование избыточности истинной модели удалением нулевых параметров приведет к модели с четырьмя параметрами ψ_1 , ψ_2 , ψ_4 , ψ_5 . Однако модель остается избыточной, поскольку значения ψ_4 и ψ_5 равны. Получить модель с тремя параметрами можно, перенумеровав уровни фактора β : например, взаимно перенумеровав второй и третий уровни, получим истинную модель с тремя параметрами ψ_1 , ψ_2 , ψ_5 . Таким образом, последний вариант нумерации потенциально приводит к более простому описанию.

Получить процедуру выбора структуры, инвариантную к перенумерации уровней факторов, можно путем перехода к обобщенной за-

даче выбора структуры, сформировав дополнительное множество ограничений, приравнивающих нулю те парные сравнения эффектов уровней каждого фактора, которых нет среди параметров. В нашем примере это следующие ФДО:

$$\varphi_1 = \alpha_1 - \alpha_2 = \psi_2 - \psi_3, \quad \varphi_2 = \beta_1 - \beta_2 = \psi_4 - \psi_5.$$

В результате исходное множество ограничений примет вид

$$\psi_1 = 0,0, \quad \psi_2 = 0,0, \quad \psi_3 = 0,0, \quad \psi_4 = 0,0, \quad \psi_5 = 0,0, \quad \psi_6 = 0,0, \\ \psi_2 - \psi_3 = 0,0, \quad \psi_4 - \psi_5 = 0,0.$$

Таким образом, при исходной нумерации уровней истинная модель будет содержать параметры $\psi_1, \psi_2, \psi_4, \psi_5$ и ограничение, приравнивающее нулю φ_2 .

Далее заметим, что реально на первый отклик влияет только фактор α , а на второй отклик – только фактор β . Следовательно, для получения более компактной модели следует построить модель типа псевдонезависимых регрессий. В этом случае истинная модель будет содержать параметры $\psi_{11}, \psi_{12}, \psi_{21}, \psi_{42}, \psi_{52}$ и ограничение, приравнивающее нулю φ_{22} , где второй индекс ФДО отражает их принадлежность определенному отклику.

5.1.6. ИССЛЕДОВАНИЕ КРИТЕРИЕВ ВЫБОРА СТРУКТУРЫ

С целью определения работоспособности критериев качества многооткликовых регрессионных моделей было проведено экспериментальное исследование.

Рассмотрим результаты исследования критериев качества для традиционной задачи выбора структуры, проведенного совместно с д-ром техн. наук, проф. А.А. Поповым.

Для обобщенной задачи выбора структуры многооткликовой регрессионной модели и моделей с качественными факторами результаты исследования в целом совпадают с представленными ниже [55, 73].

Методика исследования критериев базировалась на методе статистических испытаний (методе Монте-Карло) и состояла в многократном генерировании выборки (наложении различных реализаций псев-

дослучайных величин на «истинные» данные) с последующим нахождением по ней структуры, доставляющей критерию оптимальное значение.

Ошибки представляли собой независимые l -мерные нормальные случайные векторы с нулевым математическим ожиданием и ковариационной матрицей Σ , которая определялась по заданным корреляционной матрице ошибок наблюдений и величине ρ (уровню шума [39]) следующим образом.

Вначале вычислялись диагональные элементы

$$\Sigma_{ii} = \frac{\rho^2}{N + N_c - 1} \sum_{j=1}^{N+N_c} (U_{ji} - \bar{U}_i)^T (U_{ji} - \bar{U}_i), \quad i=1, \dots, l,$$

где U_{ji} – истинное значение i -го отклика в j -й точке; \bar{U}_i – среднее истинных значений i -го отклика в выборке; N_c – количество точек в дополнительной выборке C , наблюдения которой не использовались при построении модели. Затем по дисперсиям Σ_{ii} и корреляционной матрице определялись остальные элементы матрицы Σ .

Критерии исследовались при шуме различного уровня, при разной степени коррелированности элементов вектора ошибок и при различном количестве наблюдений в выборке.

В качестве плана эксперимента для выборки W применяется D -А-оптимальный план (20 точек) для полиномиальной модели второго порядка с тремя факторами, варьируемыми на трех уровнях ($x_i \in [-1, 1]$). Разбиение W на части A и B , по десять наблюдений в каждой, производилось с позиций D -оптимальности плана для выборки A . В качестве плана эксперимента для выборки C использовались точки, дополняющие план выборки W до полного факторного эксперимента (7 точек). В каждой точке плана генерировалось по одному наблюдению. Кроме того, для исследования поведения критериев при изменении количества наблюдений генерировались выборки с повторными наблюдениями в тех же точках плана, а именно: при $N = 40$ в каждой точке генерировалось по два наблюдения; при $N = 160$ – по восемь наблюдений.

Полезность критериев определялась частотой выбора истинной структуры в серии экспериментов, проведенной при одинаковых ха-

рактических характеристиках ошибок наблюдений. Критерий считается работоспособным, если соответствующая частота не меньше 0,5.

Приведем результаты исследования, проведенного для традиционной многооткликной модели. Истинная модель имела вид

$$y_1 = 12 + 7x_1 + 3,5x_2 - 1,7x_3 + 0x_1^2 + 0x_1x_2 + 0,6x_1x_3 + 0x_2x_3 + e_1, \quad (5.39)$$

$$y_2 = 11 + 6x_1 + 4x_2 + 2x_3 + 0x_1^2 + 0x_1x_2 - 0,7x_1x_3 + 0x_2x_3 + e_2. \quad (5.40)$$

С нулевыми коэффициентами указаны регрессоры, входящие в полную модель, но отсутствующие в истинной. Было получено по 20 выборок при уровнях шума 0,05, 0,20, 0,35, 0,50, 0,65, 0,80, 0,95 и коэффициенте корреляции между e_1 и e_2 0,1, 0,5, 0,9. Аддитивные постоянные (регрессоры «1») принудительно включались во все модели.

Обобщения критериев непротиворечивости и вариативности, как и для однооткликных моделей, работоспособны только при малом уровне шума. Обобщения остальных критериев работоспособны в большей степени. В целом большую работоспособность имеют критерии типа скользящего контроля и Меллоуса. На рис. 5.1 представлены максимальные значения величины ρ , при которых различные критерии выбирают истинную структуру не менее чем в половине случаев. Столбцам разного цвета на гистограмме соответствуют различные значения коэффициента корреляции элементов вектора ошибок.

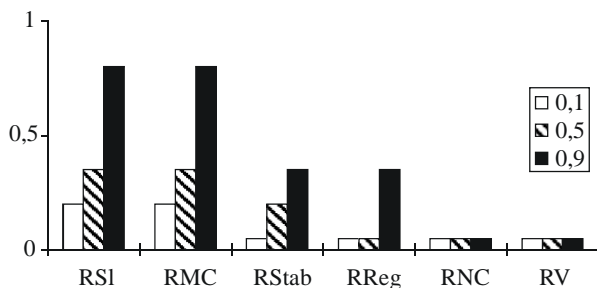


Рис. 5.1. Работоспособность различных критериев

Среди различных сверток для матричных аналогов критериев при высокой степени коррелированности элементов вектора ошибок преимущество показали детерминант и след матрицы с взвешиванием. На

рис. 5.2 представлены максимальные значения величины ρ , при которых различные свертки матричного аналога критерия скользящего контроля выбирают истинную структуру не менее чем в половине случаев. Столбцам разного цвета на гистограмме соответствуют различные значения коэффициента корреляции элементов вектора ошибок. При отсутствии столбца на гистограмме соответствующее максимальное значение ρ меньше 0,05.

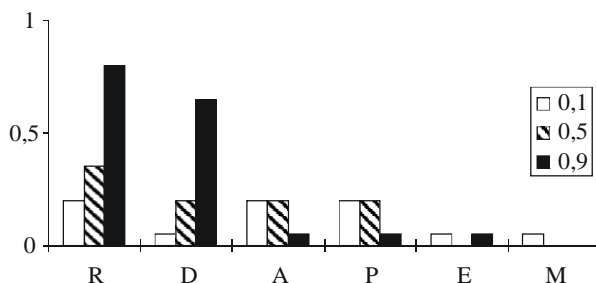


Рис. 5.2. Работоспособность различных сверток критерия скользящего контроля

При увеличении количества наблюдений в выборке указанные тенденции в целом сохраняются. На рис. 5.3 представлены значения частоты выбора истинной структуры различными критериями при значении коэффициента корреляции 0,9 и $\rho = 0,95$. Столбцам разного цвета на гистограмме соответствуют различные значения количества наблюдений.

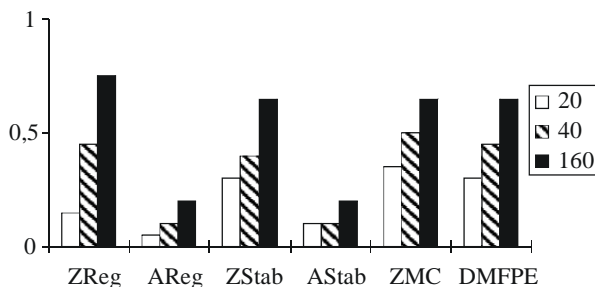


Рис. 5.3. Работоспособность критериев при различном числе наблюдений

Дополнительное исследование обобщений критериев Меллоуса, скользящего контроля и регулярности при одной реализации ошибок показало, что при малом уровне шума минимум критерия соответствует истинной структуре, а при увеличении уровня шума сложность оптимальной структуры уменьшается. Поведение критериев DSI и ZMC при различных значениях уровня шума и при значениях коэффициента корреляции 0,9 представлено на рис. 5.4 и 5.5. По оси абсцисс отложено количество регрессоров в модели, а по оси ординат – значения критерия для лучших структур соответствующей сложности, различным кривым соответствуют различные значения уровня шума.

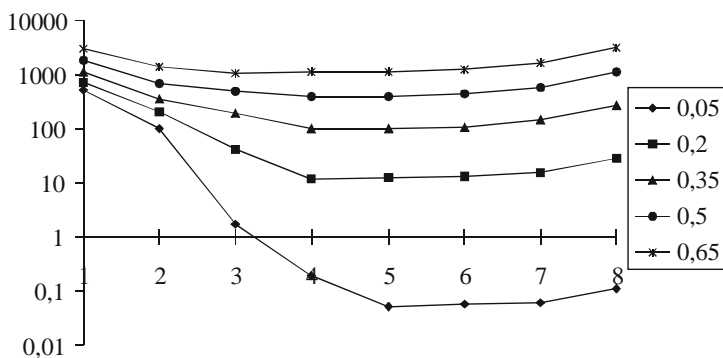


Рис. 5.4. Поведение критерия DSI

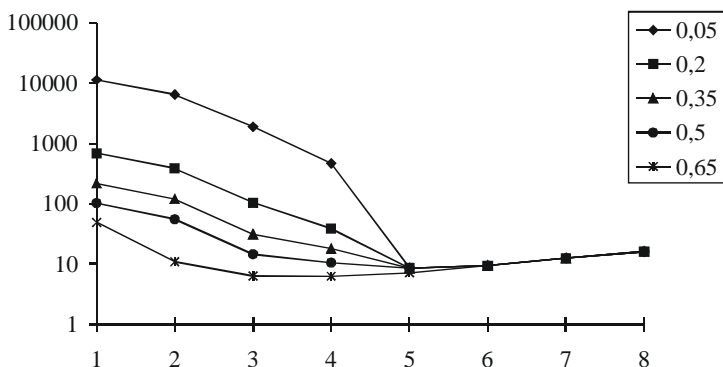


Рис. 5.5. Поведение критерия ZMC

Рассмотрим исследование для модели псевдонезависимых регрессий. Оно имело дополнительную цель: сравнить несколько подходов к выбору структуры. Первый подход предполагает выбор структуры по разработанным критериям качества структур многооткликковых моделей с использованием оценок Зельнера для параметров (используется МНК-оценка матрицы Σ при полной модели). В отличие от него при втором подходе используются МНК-оценки параметров, т. е. при оценивании параметров не учитывается коррелированность и неодинаковость дисперсий элементов вектора ошибок. Третий подход – традиционный, выбор структуры здесь производится по одномерным критериям независимо для каждого отклика.

Истинная модель имела вид

$$y_1 = 12 + 0x_1 + 6x_2 + 1,6x_1^2 + 0x_1x_2 + 3,6x_1x_3 - 2,6x_2^2 + e_1,$$

$$y_2 = 11 + 0x_1 + 7x_3 + 0x_1x_2 + 4,2x_1x_3 + 2,2x_2x_3 - 1,2x_3^2 + e_2.$$

С нулевыми коэффициентами указаны регрессоры, входящие в полную модель, но отсутствующие в истинной. Было получено по 20 выборок при уровнях шума 0,05, 0,20, 0,35, 0,50, 0,75, 1,00 и коэффициенте корреляции между e_1 и e_2 0,1, 0,5, 0,9. Аддитивные постоянные (регрессоры «1») принудительно включались во все модели.

Тенденции, отмеченные при исследовании традиционных многооткликковых моделей, справедливы и в данном случае. Кроме того, выбор

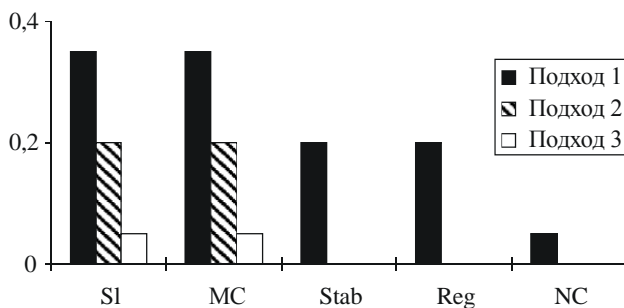


Рис. 5.6. Работоспособность критериев для модели псевдонезависимых регрессий

структуры по критериям качества структур многооткликowych моделей имел определенное преимущество перед традиционным подходом, особенно при высокой степени коррелированности компонентов вектора ошибок, а использование оценок Зельнера имело преимущество перед использованием МНК-оценок.

На рис. 5.6 представлены максимальные значения уровня шума, при которых различные критерии в последовательности экспериментов выбирают истинную структуру не менее чем в половине случаев (значение коэффициента корреляции – 0,9). Для подходов 1 и 2 указаны результаты для критериев в виде Z-свертки матричных аналогов. Столбцам разного цвета на гистограмме соответствуют различные подходы к выбору структуры. При отсутствии столбца на гистограмме соответствующее максимальное значение уровня шума меньше 0,05.

5.2. АЛГОРИТМЫ ВЫБОРА СТРУКТУРЫ

5.2.1. АЛГОРИТМЫ РЕШЕНИЯ ТРАДИЦИОННОЙ ЗАДАЧИ ВЫБОРА СТРУКТУРЫ

Поиск наилучшей по какому-либо критерию качества структуры модели является задачей целочисленного нелинейного программирования. Решить задачу можно перебором различных структур. Алгоритмы, осуществляющие такой поиск, называются переборными.

Гарантированное решение задачи получаем в результате перебора всех структур. Подобный алгоритм называется в работе [33] алгоритмом всех возможных регрессий, а в работе [39] – комбинаторным алгоритмом. Мы будем называть его *алгоритмом полного перебора*.

Количество возможных структур составляет 2^m , где m – количество регрессоров в полном наборе регрессоров, из которых выбирается оптимальный набор. По этой причине реально данным алгоритмом можно воспользоваться при относительно небольших значениях m .

Для сокращения количества перебираемых структур предложены следующие подходы [39]: назначение протекции регрессорам; назначение нижней и верхней границ диапазона сложности структур. В первом случае часть регрессоров обязательно присутствует во всех структурах. Во втором случае назначаются числа m_1 (нижняя граница) и m_2

(верхняя граница) и в переборе участвуют только те структуры, которые содержат количество регрессоров, лежащее в диапазоне $m_1 \dots m_2$.

Генерирование последовательности просматриваемых структур удобно производить с помощью счетчика Гарсайда [106]:

$$S_k = (S_{k-1}, k, S_{k-1}),$$

где $S_1 = 1$, а последовательность S_k содержит номера перебираемых регрессоров в том порядке, в каком они должны включаться-исключаться: если регрессор отсутствует в структуре, то он включается в нее, иначе – исключается. Такая последовательность обладает двумя важными свойствами: переход от структуры к структуре осуществляется добавлением или отбрасыванием одного регрессора, каждая структура формируется ровно один раз.

Другой путь состоит в просмотре только наиболее перспективных вариантов. Этот принцип реализуют алгоритмы направленного перебора [5, 39]. Они осуществляют поиск минимума критерия со значительно меньшими затратами, однако в общем случае не обеспечивают нахождение оптимальной структуры.

Можно выделить две простые стратегии направленного перебора: включения и исключения. При использовании стратегии исключения алгоритм начинает работу с оценки качества модели, содержащей все регрессоры, и состоит из шагов удаления из нее регрессоров до тех пор, пока величина критерия не перестанет уменьшаться. При использовании стратегии включения алгоритм действует в обратном направлении: начиная с модели, содержащей только регрессоры с протекцией, он последовательно добавляет в нее регрессоры.

При описании более сложных стратегий перебора будем пользоваться понятием *опорной структуры*. Опорными в алгоритмах направленного перебора будем называть такие структуры, из которых формируются структуры, просматриваемые на одном шаге. В алгоритмах, реализующих простые стратегии, используется одна опорная структура – наилучшая структура, полученная на предыдущем шаге; структуры, просматриваемые на текущем шаге, получают из нее путем добавления или удаления одного регрессора.

Простые стратегии можно усложнить так, чтобы на каждом шаге алгоритма использовалось несколько опорных структур. Подобное обобщение для стратегии включения реализует алгоритм, называемый

в рамках школы Ивахненко многоэтапным селекционно-комбинаторным [39]; мы будем называть его алгоритмом включения.

Рассмотрим конкретные алгоритмы, предлагаемые для решения традиционной задачи выбора структуры.

Алгоритм включения работает следующим образом. Сначала (на нулевом шаге) вычисляется значение критерия для структуры, содержащей регрессоры с протекцией, которая будет опорной структурой первого шага. На первом и всех последующих шагах алгоритма вычисляются значения критерия для всех структур, получаемых из опорных структур путем добавления одного регрессора, и отбираются наиболее уменьшающие значение критерия структуры, которые будут опорными структурами следующего шага. Алгоритм заканчивает работу, когда все просматриваемые на текущем шаге структуры имеют меньшее качество по сравнению с лучшей структурой, найденной на предыдущем шаге.

Количество опорных структур ν , используемых на каждом шаге, лежит в назначенных пределах $\nu_1 \leq \nu \leq \nu_2$ и может изменяться от шага к шагу в зависимости от поведения критерия. В том случае, если количество структур, уменьшающих значение критерия, меньше ν_1 , то в качестве опорных структур следующего шага отбирается также несколько структур, в наименьшей степени увеличивающих значение критерия. В случае $\nu_1 > 0$ возможно осуществлять принудительное выполнение алгоритма до тех пор, пока в модель не будут включены все регрессоры, даже если ни на одном шаге значение критерия не уменьшается.

Алгоритм исключения действует в обратном порядке по отношению к предыдущему алгоритму. Сначала (на нулевом шаге) вычисляется значение критерия для содержащей все регрессоры структуры, которая будет опорной структурой первого шага. На первом и всех последующих шагах алгоритма вычисляются значения критерия для всех структур, получаемых из опорных структур путем удаления одного регрессора, и отбираются наиболее уменьшающие значение критерия структуры, которые будут опорными структурами следующего шага. Алгоритм заканчивает работу, когда все просматриваемые на текущем шаге структуры имеют меньшее качество по сравнению с лучшей структурой, найденной на предыдущем шаге. В алгоритме также возможно назначать пределы варьирования количества опорных структур

и запрещать останов до тех пор, пока не будут исключены все регрессоры без протекции.

Усложнением простой стратегии является использование, наряду с шагами включения-исключения, шагов обмена регрессора, содержащегося в модели, на регрессор, не содержащийся в ней. Различные алгоритмы, использующие обмены, приведены в [127, 185]. В предлагаемых алгоритмах среди множества пар, которые можно обменять на конкретном шаге, выбирается та, обмен для которой ведет к наибольшему уменьшению значения критерия. Опишем два конкретных варианта.

Алгоритм включения с обменами состоит из шагов включения регрессора, после каждого из которых проводится максимально возможное количество шагов обмена, ведущих к уменьшению значения критерия. Алгоритм заканчивает работу, когда более уменьшить значение критерия не удастся. Можно запретить останов алгоритма до тех пор, пока в модель не будут включены все регрессоры.

Алгоритм исключения с обменами аналогичен предыдущему алгоритму, при этом вместо шагов включения используются шаги исключения.

Следующее усложнение простых стратегий заключается в их комбинировании; реализующий его алгоритм – *алгоритм включения-исключения* – состоит из последовательности шагов включения и исключения регрессоров. Правила чередования шагов включения и исключения могут существенно варьироваться (см., например, [5]). В предлагаемых алгоритмах регрессор, включаемый (исключаемый) на каждом шаге, выбирается из условия наибольшего уменьшения значения критерия, выбирается тот вид шага, в результате проведения которого произойдет наибольшее уменьшение значения критерия. Алгоритм начинает работу с вычисления значения критерия для произвольно заданной начальной структуры и заканчивает – когда более уменьшить значение критерия не удастся.

Возможна следующая интерпретация алгоритма включения-исключения: если начальная структура содержит только регрессоры с протекцией, то это алгоритм включения с $\nu=1$ и дополнительными шагами исключения, если же начальная структура содержит все регрессоры, то это алгоритм исключения с $\nu=1$ и дополнительными шагами включения.

Можно усложнить алгоритм включения-исключения, введя дополнительную возможность провести шаги обмена присутствующего в модели регрессора на отсутствующий. Здесь имеются две стратегии:

1) один шаг обмена проводится (если найдется обмен, влекущий уменьшение значения критерия) только в том случае, если ни одно включение-исключение не уменьшает значения критерия, после чего необходимо вернуться к шагам включения-исключения;

2) после каждого шага включения-исключения проводится максимально возможное количество шагов обмена, ведущих к уменьшению значения критерия.

Подытоживая приведенные описания алгоритмов, перечислим используемые ими усложнения простых стратегий:

- останов после включения (исключения) всех регрессоров, участвующих в переборе (кроме алгоритма включения-исключения);
- использование нескольких опорных структур (алгоритмы включения и исключения);
- использование, наряду с основными шагами, шагов противоположного направления (алгоритм включения-исключения) или шагов обмена (алгоритмы включения и исключения с обменами, алгоритм включения-исключения).

Алгоритмы направленного перебора в общем случае могут найти лишь некоторое субоптимальное решение. Повышения работоспособности алгоритмов можно достичь, применяя в них стратегии случайного поиска. Простым и эффективным является *стохастический алгоритм включения-исключения*, который заключается в многократном выполнении алгоритма включения-исключения (в том числе с использованием шагов обмена регрессоров) из случайно формируемых начальных структур; на первом этапе алгоритм начинает работу из начальной структуры, заданной пользователем.

Среди других алгоритмов выбора структуры следует упомянуть методы ветвей и границ (см., например, [119]).

Рассмотрим возможности, которые резонно добавить во все алгоритмы:

1) отбор нескольких структур, наиболее качественных с точки зрения критерия;

2) формирование последовательности структур различной сложности, имеющих лучшее качество с точки зрения критерия;

3) вычисление значений ряда дополнительных критериев качества для структур, отобранных посредством использования двух предыдущих возможностей.

Возможности 1 и 2 реализуют принцип свободы выбора решений Габора [40]. Благодаря им, алгоритм формирует множество структур, из которого впоследствии может быть отобрана одна структура. При окончательном выборе наилучшей структуры может использоваться, например, критерий «сумма мест» [90], когда структуры упорядочиваются в порядке уменьшения предпочтения с точки зрения каждого критерия; структура, имеющая наименьшую сумму мест по всем критериям, считается наилучшей. Это один из способов многокритериального выбора структуры модели, который может быть осуществлен благодаря возможности 3.

Построение последовательностей лучших структур различной сложности позволяет выбрать структуру в кластере структур соответствующей сложности [38]. Анализ графиков зависимости значений критерия для лучших структур различной сложности от сложности структуры может дать дополнительную информацию для выбора наилучшей структуры [106, 164].

Оптимизационный подход и проверка гипотез комбинируются в алгоритме Эфроимсона [139]. Рассмотрим вариант, аналогичный алгоритму включения-исключения, который называется *пошаговой регрессией*. Подробное описание алгоритма для однооткликовой модели содержится в [30]. В [10, 98, 179, 189] дается описание алгоритмов для традиционной многооткликовой модели.

Алгоритм состоит из последовательности шагов включения и исключения регрессора, которые для однооткликовой модели осуществляются по следующим правилам [30]:

1) исключается регрессор, вносящий наименьший вклад в остаточную сумму квадратов среди всех регрессоров, если значение его F -статистики исключения ниже установленного порога;

2) включается регрессор, приводящий к наибольшему уменьшению остаточной суммы квадратов среди всех регрессоров, если значение его F -статистики включения не ниже установленного порога;

3) правило 2 выполняется только в том случае, если нет возможности выполнить правило 1; если ни одно из них не может быть выполнено, наступает остановка.

F -статистика включения (исключения) является статистикой проверки гипотезы о незначимости одного из параметров и вычисляется для включения как

$$F_{in}(k) = \frac{(N - p - 1)h(k)}{g - h(k)},$$

для исключения как

$$F_{out}(k) = -\frac{(N - p)h(k)}{g},$$

где g – остаточная сумма квадратов до включения (исключения) регрессора; $h(k)$ – ее уменьшение (отрицательное в случае исключения регрессора), вызванное включением-исключением k -го регрессора; p – число регрессоров в модели до включения (исключения) k -го регрессора.

Заметим, что из-за наличия операций оптимизации используемая F -статистика не будет иметь F -распределение, поэтому выбор пороговых значений осуществляется эвристически.

Для общей многооткликковой модели (1.38) использование пошаговой регрессии связано с преобразованием к однооткликковой модели (1.40). В случае неизвестной матрицы Σ можно воспользоваться ее оценкой при полной модели.

Для традиционной многооткликковой модели проверяется гипотеза о незначимости всех элементов одной строки матрицы параметров Θ . Вводятся Λ -статистики включения и исключения

$$\Lambda_{in}(k) = \frac{|G - H(k)|}{|G|}, \quad \Lambda_{out}(k) = \frac{|G|}{|G - H(k)|}, \quad (5.41)$$

где G и $H(k)$ – матрицы, аналогичные g и $h(k)$ для однооткликковой модели. Включаемый (исключаемый) регрессор выбирается так, чтобы получающееся в результате значение $|G - H(k)|$ в числителе (знаменателе) Λ -статистики было минимально.

На практике может быть удобен переход от Λ -статистик к F -статистикам [98, 179]. В соответствии с (1.34) получаем:

$$F_{in}(k) = \frac{N-p-l-1-\Lambda_{in}(k)}{l \Lambda_{in}(k)}, \quad F_{out}(k) = \frac{N-p+1-l-1-\Lambda_{out}(k)}{l \Lambda_{out}(k)}.$$

Тогда можно использовать те же правила чередования шагов включения и исключения, что и в однооткликном случае.

Рассмотрим вычислительные аспекты алгоритмов выбора структуры.

Пусть A – квадратная матрица, k -й диагональный элемент которой a_{kk} не равен нулю. Результатом выметания матрицы A по ее k -му диагональному элементу с помощью прямого оператора выметания является новая матрица \tilde{A} того же самого размера с элементами [30]:

$$\tilde{a}_{kk} = -\frac{1}{a_{kk}}, \quad \tilde{a}_{ik} = u \frac{a_{ik}}{a_{kk}}, \quad \tilde{a}_{kj} = u \frac{a_{kj}}{a_{kk}}, \quad \tilde{a}_{ij} = a_{ij} - \frac{a_{ik}a_{kj}}{a_{kk}},$$

где i, j не равны k и $u = 1$.

Прямой оператор выметания можно нейтрализовать с помощью обратного оператора выметания, результат действия которого определяется по тем же формулам при $u = -1$.

Оператор выметания имеет следующие свойства [30]:

1) оператор выметания преобразует симметричные матрицы в симметричные;

2) результат двух последовательных выметаний не зависит от порядка их выполнения;

3) если для разбиения матрицы на блоки, как показано ниже, возможно прямое выметание по каждому диагональному элементу квадратной подматрицы A_{11} в некотором порядке, т. е. если эти элементы не равны нулю, то A_{11} невырождена и результат выметания показан в правой части следующего перехода:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \rightarrow \begin{bmatrix} -A_{11}^{-1} & A_{11}^{-1}A_{12} \\ A_{21}A_{11}^{-1} & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}.$$

Для эффективного вычисления значений критерия при переборе структур традиционных многооткликowych моделей (включая однооткликowe при $l=1$) оператор выметания применяется к матрице [30, 98]

$$A^0 = \begin{bmatrix} X^T X & X^T Y \\ Y^T X & Y^T Y \end{bmatrix} = \begin{bmatrix} A_{11}^0 & A_{12}^0 \\ A_{21}^0 & A_{22}^0 \end{bmatrix}.$$

Выполнив прямые выметания по всем диагональным элементам левого верхнего блока, согласно свойству 3, получим из матрицы A^0 матрицу A вида

$$A = \begin{bmatrix} -(X^T X)^{-1} & (X^T X)^{-1} X^T Y \\ Y^T X (X^T X)^{-1} & Y^T Y - Y^T X (X^T X)^{-1} X^T Y \end{bmatrix}.$$

В результате прямого выметания по диагональному элементу соответствующий регрессор включается в модель, а в результате обратного выметания – исключается из модели. При этом в строках правого верхнего блока, соответствующих включенным регрессорам, находятся МНК-оценки параметров для текущей модели, а правый нижний блок содержит матрицу остаточных сумм квадратов-произведений.

В случае выбора структуры многооткликовой модели (1.38) матрица A^0 формируется для однооткликовой модели, полученной путем преобразования (1.40) с использованием оценки матрицы Σ при полной модели, правый нижний блок при этом содержит величину $\text{tr } \Sigma^{-1} \text{RSS}$, являющуюся остаточной суммой квадратов для однооткликовой модели (1.40).

При выборе структуры алгоритмом пошаговой регрессии для традиционной многооткликовой модели на каждом шаге процедуры необходимо для всех k вычислять значения $|G - H(k)|$ [98, 189]. Матрица $H(k)$, согласно свойству 3 оператора выметания, вычисляется по формуле

$$H(k) = \frac{1}{a_{kk}} h_k h_k^T, \quad (5.42)$$

где $h_k^T = [a_{k(m+1)} \quad a_{k(m+2)} \quad \dots \quad a_{k(m+l)}]$. В результате можно применить формулу для пересчета определителя [6, 34]:

$$|G - H(k)| = |G| \left[1 - \frac{1}{a_{kk}} h_k^T G^{-1} h_k \right]. \quad (5.43)$$

Выражение в скобках в правой части (5.43) можно вычислить следующим образом [98]. Произведем выметания матрицы A по всем диагональным элементам блока $A_{22} = G$, в результате будем иметь матрицу

$$D = \begin{bmatrix} A_{11} - A_{12}G^{-1}A_{21} & A_{12}G^{-1} \\ G^{-1}A_{21} & -G^{-1} \end{bmatrix}.$$

Поскольку k -м столбцом подматрицы A_{12} является вектор-столбец h_k , а k -й строкой подматрицы A_{21} – вектор-строка h_k^T , искомое выражение вычисляется по формуле

$$1 - \frac{h_k^T G^{-1} h_k}{a_{kk}} = \frac{d_{kk}}{a_{kk}}. \quad (5.44)$$

Подставляя его в (5.43), получаем

$$|G - H(k)| = |G| \frac{d_{kk}}{a_{kk}}. \quad (5.45)$$

В результате для F -статистик включения и исключения k -го регрессора справедливы следующие вычислительные формулы:

$$F_{in}(k) = \frac{N - p - l}{l} \left[\frac{a_{kk}}{d_{kk}} - 1 \right], \quad F_{out}(k) = \frac{N - p + 1 - l}{l} \left[\frac{d_{kk}}{a_{kk}} - 1 \right].$$

Для получения матрицы D на каждом шаге алгоритма воспользуемся свойством 2 оператора выметания. В результате вместо вычисле-

ния матрицы D из текущей матрицы A , для которого на каждом шаге требуется l выметаний, применима следующая схема [98]. Путем прямых выметаний из исходной матрицы A^0 получим матрицу D^0 . Затем при переходе от шага к шагу производится одновременное выметание матриц A и D по одному и тому же элементу.

5.2.2. АЛГОРИТМЫ РЕШЕНИЯ ОБОБЩЕННОЙ ЗАДАЧИ ВЫБОРА СТРУКТУРЫ

Для решения обобщенной задачи выбора структуры можно применить алгоритмы, используемые при решении традиционной задачи. Рассмотрим конкретные предлагаемые алгоритмы, их особенности и эффективные вычислительные схемы.

Модель в алгоритмах формируется добавлением или удалением следующих *элементов перебора*: регрессоров; ограничений на параметры из множеств (5.4), (5.5); блоков, состоящих из нескольких регрессоров и ограничений, которые должны одновременно присутствовать или отсутствовать в модели.

Алгоритмы позволяют осуществлять протекцию элементам перебора, отбор нескольких структур, наиболее качественных с точки зрения критерия, формирование последовательности структур различной сложности, имеющих лучшее качество с точки зрения критерия, и вычисление значений ряда дополнительных критериев качества для множества предварительно отобранных структур.

Назначение протекции позволяет учитывать априорную информацию в виде линейных ограничений на параметры, а возможность формировать блоки – информацию об одновременном присутствии и отсутствии в модели каких-либо элементов. Заметим, что формирование блоков можно осуществлять также для уменьшения количества перебираемых структур.

Рассмотрим алгоритмы. Первые два алгоритма аналогичны одноименным алгоритмам, описанным в п. 5.2.1.

Алгоритм полного перебора перебирает структуры в заданном диапазоне сложности, который определяется нижней и верхней границами.

Алгоритм включения-исключения начинает работу с произвольно заданной начальной структуры и имеет дополнительную возможность провести шаги обмена по одной из двух стратегий. Возможны и стохастические варианты данного алгоритма.

При отсутствии блоков шаг обмена – это шаг, позволяющий изменить модель, не изменяя сложности ее структуры. Обменами являются замена присутствующего в модели регрессора на отсутствующий и замена присутствующего в модели ограничения на отсутствующее. Также обменом является одновременное включение или исключение регрессора и ограничения. При наличии блоков обмен может изменить сложность структуры модели, однако производится по тем же правилам. Блок в правилах участвует при этом как регрессор или как ограничение в зависимости от того, регрессоров или ограничений в нем больше.

Алгоритм включения-исключения по F-статистике аналогичен пошаговой регрессии. Он позволяет выбрать начальную структуру, однако не позволяет формировать последовательность лучших структур различной сложности, а вместо множества структур, имеющих наилучшее качество, представляет структуры, выбранные лучшими на нескольких последних шагах.

Рассмотрим вычислительные аспекты построения алгоритмов. Прежде всего обратим внимание на традиционную многооткликтовую модель, так как однооткликтовая модель, используемая в том числе и для представления многооткликтовой модели общего вида (1.38), является ее частным случаем.

Для вычисления значений критериев и F -статистик при переборе структур будем применять оператор выметания к матрице, которая имеет вид

$$A^0 = \begin{bmatrix} X^T X & C & X^T Y \\ C^T & 0 & \Xi \\ Y^T X & \Xi^T & Y^T Y \end{bmatrix} = \begin{bmatrix} A_{11}^0 & A_{12}^0 & A_{13}^0 \\ A_{21}^0 & A_{22}^0 & A_{23}^0 \\ A_{31}^0 & A_{32}^0 & A_{33}^0 \end{bmatrix}.$$

Пусть матрица $\begin{bmatrix} X^T & \vdots & R^T \end{bmatrix}$ имеет полный (строчный) ранг. Разобьем матрицу X на два блока – $X_1 \mid X_2$: в первом содержатся

значения регрессоров, включенных в модель, а во втором – не включенных. Представим множество (5.5) в виде

$$\begin{bmatrix} R_1 & R_2 \\ R_3 & R_4 \end{bmatrix} \begin{bmatrix} \Theta_1 \\ 0 \end{bmatrix} = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix},$$

где верхние блоки соответствуют ограничениям, включенным в модель, а нижние – не включенным; Θ_1 – неизвестные параметры при регрессорах, включенных в модель. Такое разбиение всегда возможно сделать путем изменения нумерации регрессоров и ограничений.

Произведем прямые выметания в матрице A^0 , в данном случае имеющей вид

$$\begin{bmatrix} X_1^T X_1 & X_1^T X_2 & R_1^T & R_3^T & X_1^T Y \\ X_2^T X_1 & X_2^T X_2 & R_2^T & R_4^T & X_2^T Y \\ R_1 & R_2 & 0 & 0 & U_1 \\ R_3 & R_4 & 0 & 0 & U_2 \\ Y^T X_1 & Y^T X_2 & U_1^T & U_2^T & Y^T Y \end{bmatrix},$$

по диагональным элементам первого и третьего диагональных блоков.

Обозначим

$$\begin{bmatrix} X_1^T X_1 & R_1^T \\ R_1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} C_{11} & C_{12} \\ C_{13} & C_{14} \end{bmatrix}.$$

Тогда, воспользовавшись свойством 3 оператора выметания, приходим к следующим выводам.

Блок, содержащий $X_1^T Y$, содержит оценки $\hat{\Theta}_1$ неизвестных параметров при включенных в модель регрессорах, удовлетворяющие ограничениям $R_1 \Theta_1 = U_1$: $\hat{\Theta}_1 = C_{11} X_1^T Y + C_{12} U_1$.

Блок, содержащий U_2 , содержит разность между правой и левой частями не включенных в модель ограничений $R_3 \Theta_1 = U_2$: $U_2 - R_3 \hat{\Theta}_1$, что позволяет без дополнительных вычислений выделить не включенные в модель ограничения, которым автоматически удовлетворяют оценки $\hat{\Theta}_1$.

Блок, содержащий $Y^T Y$, содержит элементы матрицы остаточных сумм квадратов-произведений: $Y^T Y - Y^T X_1 \hat{\Theta}_1 - U_1^T \hat{L}_1$, где \hat{L}_1 – значения множителей Лагранжа при ограничениях, включенных в модель.

Заметим, что в данном случае множество ограничений $R_1 \Theta_1 = U_1$, $R_3 \Theta_1 = U_2$ эквивалентно исходному множеству (5.5).

Обозначим через A матрицу, полученную из A^0 в результате выметаний, соответствующих некоторой структуре модели. Тогда для включения отсутствующего в модели k -го регрессора необходимо осуществить прямое выметание матрицы A по k -му диагональному элементу блока A_{11} , а для исключения присутствующего регрессора – обратное выметание. Для включения отсутствующего k -го ограничения из множества (5.5) необходимо осуществить прямое выметание матрицы A по k -му диагональному элементу блока A_{22} , а для исключения присутствующего ограничения – обратное выметание.

В некоторых случаях осуществить шаг включения-исключения невозможно, при этом соответствующий диагональный элемент матрицы A равен нулю (при вычислениях на компьютере – близок к нулю). Перечислим такие ситуации:

1) невозможно включить ограничение (или исключить регрессор), если соответствующая ему строка матрицы коэффициентов $I_m \mid C^T$ в левой части системы уравнений (5.7) линейно зависима со строками R (ситуация обусловлена видом системы уравнений, решаемой при оценивании параметров);

2) невозможно исключить ограничение (или включить регрессор), если это приведет к неполноте ранга матрицы $\left[X^T \mid \tilde{R}^T \right]$, где \tilde{R} – матрица коэффициентов системы ограничений (1.25) после удаления из нее соответствующего ограничения [106].

Рассмотрим правила проведения выметаний, использующиеся при включении-исключении блоков в описываемых алгоритмах. При включении-исключении блока циклически осуществляются попытки провести выметания по всем еще не выметанным диагональным элементам матрицы A , соответствующим компонентам блока, до тех пор, пока среди них имеется хотя бы один ненулевой диагональный эле-

мент. При обмене блоков осуществляются следующие действия (в порядке их приоритета):

- 1) исключение регрессора;
- 2) исключение ограничения;
- 3) включение ограничения;
- 4) включение регрессора.

Выбор приоритетов осуществлен эвристически и нацелен на возможность работы с идентифицирующими ограничениями.

Для модели (1.38), преобразованной к виду (1.40), выметаемая матрица имеет вид

$$A^0 = \begin{bmatrix} X^T (\hat{\Sigma}^{-1} \otimes I_N) X & C & X^T (\hat{\Sigma}^{-1} \otimes I_N) \tilde{Y} \\ C^T & 0 & \xi \\ \tilde{Y}^T (\hat{\Sigma}^{-1} \otimes I_N) X & \xi^T & \tilde{Y}^T (\hat{\Sigma}^{-1} \otimes I_N) \tilde{Y} \end{bmatrix} =$$

$$= \begin{bmatrix} A_{11}^0 & A_{12}^0 & A_{13}^0 \\ A_{21}^0 & A_{22}^0 & A_{23}^0 \\ A_{31}^0 & A_{32}^0 & A_{33}^0 \end{bmatrix}, \quad (5.46)$$

где $\hat{\Sigma}$ – оценка матрицы Σ при полной модели.

Рассмотрим подробнее алгоритм включения-исключения по F -статистике.

Воспользовавшись тем, что включение (исключение) регрессоров эквивалентно исключению (включению) ограничений из множеств (5.2), (5.3), алгоритм удобно представить состоящим из шагов включения и исключения ограничений из множеств (5.6), (5.7).

Для выбора включаемого-исключаемого ограничения на каждом шаге алгоритма вычисляются значения статистик проверки гипотез, задаваемых одним или несколькими уравнениями.

При включении-исключении блока фактически осуществляется обмен двух подсистем ограничений: исключаяющих ограничений, соответствующих регрессорам блока, и ограничений блока из множества (5.4) или (5.5). Для того чтобы было возможно применять в этом случае методы проверки гипотез, аналогичные используемым в классиче-

ском варианте пошаговой регрессии, необходимо придерживаться специальных правил формирования блоков, участвующих в переборе.

Назовем *контрограничениями* входящие в блок и имеющие нулевую правую часть ограничения на параметры только тех регрессоров, которые входят в данный блок.

Перечислим правила формирования блоков:

- 1) блок содержит только исключающие ограничения (т. е. ограничения из множеств (5.2), (5.3));
- 2) блок содержит только ограничения из множеств (5.4), (5.5);
- 3) блок содержит обмениваемые исключающие ограничения и контрограничения.

Правила 1 и 2 предполагают, что осуществляется либо только включение, либо только исключение нескольких ограничений.

Рассмотрим правило 3. Система контрограничений блока линейно зависима с системой исключающих ограничений блока. Это позволяет интерпретировать контрограничения как содержащиеся во всех перебираемых структурах. Если исключающие ограничения блока включены в модель, то контрограничения автоматически удовлетворяются или, иначе говоря, находятся в модели в неявном виде. Обмен при включении-исключении такого блока не происходит, блок интерпретируется при этом как содержащий только исключающие ограничения.

Рассмотрим традиционную многооткликовую модель.

Пусть на некотором шаге на параметры наложена система линейных ограничений

$$\bar{R}\Theta = \bar{U},$$

состоящая из системы ограничений (1.25) и неявных контрограничений.

На каждом шаге алгоритма проверяется гипотеза вида (1.32), задаваемая одним или несколькими уравнениями из множества (5.7). Вводятся Λ -статистики включения и исключения ограничений (блоков), аналогичные вводимым в рамках традиционной задачи выбора структуры (5.41).

Предположим, что в случае включения ограничения (блока) строки матрицы K линейно независимы со строками матрицы \bar{R} (за исключением строк, соответствующих контрограничениям блока), а в случае исключения блока матрица $\begin{bmatrix} X^T & \vdots & \tilde{R}^T \end{bmatrix}$ имеет полный строчный ранг и

строки матрицы K линейно независимы со строками матрицы \tilde{R} (за исключением строк, соответствующих контрограничениям блока), где матрица \tilde{R} получается из матрицы \bar{R} путем удаления строк матрицы K .

В случае включения-исключения блока ограничений переход к F -статистике производится с помощью аппроксимации (1.33). При этом в случае включения блока

$$\alpha = \operatorname{rg} \begin{bmatrix} \bar{R}^T & \vdots & K^T \end{bmatrix} - \operatorname{rg} \bar{R} = v - t,$$

$$\beta = N - m + \operatorname{rg} \begin{bmatrix} \bar{R}^T & \vdots & K^T \end{bmatrix} - \frac{l + \alpha + 1}{2} = N - m + u + v - t - \frac{l + \alpha + 1}{2},$$

а в случае исключения блока

$$\alpha = \operatorname{rg} \begin{bmatrix} \tilde{R}^T & \vdots & K^T \end{bmatrix} - \operatorname{rg} \tilde{R} = v - t,$$

$$\beta = N - m + \operatorname{rg} \begin{bmatrix} \tilde{R}^T & \vdots & K^T \end{bmatrix} - \frac{l + \alpha + 1}{2} = N - m + u - \frac{l + \alpha + 1}{2},$$

где t – количество контрограничений блока.

Для ограничений, не входящих в блоки, справедливы формулы преобразований (1.34) [89]:

$$F_{in}(k) = \frac{N - m + u + 1 - l}{l} \frac{1 - \Lambda_{in}(k)}{\Lambda_{in}(k)}, \quad F_{out}(k) = \frac{N - m + u - l}{l} \frac{1 - \Lambda_{out}(k)}{\Lambda_{out}(k)}.$$

Так как при выборе ограничений на одном и том же шаге вычисляются значения F -статистик с различным количеством степеней свободы, то для выбора включаемого-исключаемого ограничения используется сравнение не значений F -статистик (как в алгоритмах, описанных в п. 5.2.1), а соответствующих им значений вероятностного интеграла $\int_0^{F_0} g_F(z) dz$, где $g_F(z)$ – плотность F -распределения, F_0 – значение F -статистики (см. [179]).

Используются следующие правила чередования шагов включения и исключения:

1) из модели исключается ограничение (блок ограничений), значение вероятностного интеграла для F -статистики исключения которого максимально и не ниже установленного порогового значения для исключения;

2) в модель включается ограничение (блок ограничений), значение вероятностного интеграла для F -статистики включения которого минимально и ниже установленного порогового значения для включения;

3) приоритет имеет шаг 2.

Распространим вычислительную схему алгоритма пошаговой регрессии из [98] (см. п. 5.2.1) на алгоритм включения-исключения по F -статистике для традиционной многооткликowej модели.

Воспользовавшись свойством 3 оператора выметания, матрицу уменьшения остаточных сумм квадратов-произведений при включении-исключении одного ограничения, являющегося k -м уравнением множества (5.7), можно представить в виде (5.42) с

$$h_k^T = \left[a_{k(m+w+1)} \quad a_{k(m+w+2)} \quad \dots \quad a_{k(m+w+l)} \right].$$

В результате справедливо представление (5.43), для вычисления которого можно воспользоваться матрицей D , получаемой путем выметания матрицы A по всем диагональным элементам блока $A_{33} = G$:

$$D = \left[\begin{array}{ccc} \left(\begin{array}{cc} A_{11} & A_{12} \\ A_{21} & A_{22} \end{array} \right) - \left(\begin{array}{c} A_{13} \\ A_{23} \end{array} \right) G^{-1} & A_{31} \quad \vdots \quad A_{32} & \left[\begin{array}{c} A_{13} \\ A_{23} \end{array} \right] G^{-1} \\ & G^{-1} & A_{31} \quad \vdots \quad A_{32} \\ & & -G^{-1} \end{array} \right].$$

Поскольку k -м столбцом подматрицы $A_{31} \quad \vdots \quad A_{32}$ является вектор-столбец h_k , а k -й строкой подматрицы $\left[\begin{array}{c} A_{13}^T \\ A_{23}^T \end{array} \right]^T$ – вектор-строка h_k^T , справедливы представления (5.44) и (5.45).

Для F -статистик включения и исключения одного ограничения справедливы следующие вычислительные формулы:

$$F_{in}(k) = \frac{N - m + u + 1 - l}{l} \left[\frac{d_{kk}}{a_{kk}} - 1 \right], \quad F_{out}(k) = \frac{N - m + u - l}{l} \left[\frac{a_{kk}}{d_{kk}} - 1 \right].$$

Рассмотрим случай включения-исключения блока ограничений. Обозначим J_k такую последовательность индексов диагональных элементов матрицы A , в порядке которой возможно сделать все выметания, ведущие к включению-исключению k -го блока. Воспользовавшись формулой (5.43), получим

$$|G - H(k)| = |G| \prod_{i \in J_k} \frac{\tilde{d}_{ii}}{\tilde{a}_{ii}},$$

где \tilde{a}_{ii} и \tilde{d}_{ii} для i , являющегося первым элементом последовательности J_k , совпадают с a_{ii} и d_{ii} соответственно, для остальных i — с i -ми диагональными элементами матриц A и D , выметанных по диагональным элементам, соответствующим предыдущим индексам в последовательности J_k . Тогда вычисление значений статистик Уилкса для случаев включения и исключения блока ограничений на каждом шаге алгоритма можно осуществить по формулам

$$\Lambda_{in}(k) = \prod_{i \in J_k} \frac{\tilde{a}_{ii}}{\tilde{d}_{ii}}, \quad \Lambda_{out}(k) = \prod_{i \in J_k} \frac{\tilde{d}_{ii}}{\tilde{a}_{ii}}.$$

Для F -статистик получаем следующие формулы:

$$F_{in}(k) = \frac{f_2}{f_1} \left[\left(\prod_{i \in J_k} \frac{\tilde{d}_{ii}}{\tilde{a}_{ii}} \right)^{1/\gamma} - 1 \right], \quad F_{out}(k) = \frac{f_2}{f_1} \left[\left(\prod_{i \in J_k} \frac{\tilde{a}_{ii}}{\tilde{d}_{ii}} \right)^{1/\gamma} - 1 \right].$$

Перейдем к рассмотрению алгоритма включения-исключения по F -статистике для общей многооткликной модели типа (1.38).

Пусть на некотором шаге алгоритма на параметры наложена система линейных ограничений

$$\bar{R}\theta = \bar{d},$$

состоящая из системы ограничений (1.9) и неявных контрограничений.

При выборе включаемых-исключаемых ограничений и блоков, являющихся подмножествами множества (5.6), используются значения F -статистик проверки соответствующих гипотез.

Пусть проверяемая гипотеза имеет вид (1.17). Предположим, что в случае включения ограничения (блока) строки матрицы K линейно независимы со строками матрицы \bar{R} (за исключением строк, соответствующих контрограничениям блока), а в случае исключения ограничений имеет полный ранг матрица

матрица $\begin{bmatrix} S^{-1} \otimes I_N & X^T & \vdots & \tilde{R}^T \end{bmatrix}$, где матрица X определена в формуле (1.39), матрица \tilde{R} получается из матрицы \bar{R} путем удаления строк матрицы K и строки матрицы K линейно независимы со строками матрицы \tilde{R} (за исключением строк, соответствующих контрограничениям блока).

Используются F -статистики включения и исключения ограничений (блоков), аналогичные применяемым в одномерном случае:

$$F_{in}(k) = -\frac{\text{tr } \hat{\Sigma}^{-1}H(k) / f_1}{\text{tr } \hat{\Sigma}^{-1}G / f_2}, \quad F_{out}(k) = \frac{\text{tr } \hat{\Sigma}^{-1}H(k) / f_1}{\text{tr } \hat{\Sigma}^{-1}G - H(k) / f_2 - f_1},$$

где матрицы G и $H(k)$ имеют тот же смысл, что и для традиционной многооткликковой модели; $\hat{\Sigma}$ – оценка матрицы Σ при полной модели; $f_1 = \text{rg} \begin{bmatrix} \bar{R}^T & \vdots & K^T \end{bmatrix} - \text{rg } \bar{R}^T = v - t$ – для случая включения блока, $f_1 = \text{rg} \begin{bmatrix} \tilde{R}^T & \vdots & K^T \end{bmatrix} - \text{rg } \tilde{R} = v - t$ – для случая исключения блока, t – количество контрограничений блока; для отдельного ограничения $f_1 = 1$, $f_2 = N - m + \text{rg } \bar{R} = N - m + u$ [32].

Значения $\text{tr } \hat{\Sigma}^{-1}G$ на каждом шаге алгоритма находятся в блоке A_{33} матрицы A , получаемой путем выметаний матрицы (5.46), а зна-

чения $\text{tr } \hat{\Sigma}^{-1}H(k)$ можно вычислить исходя из определения оператора

выметания как $\frac{a_{k(m+w+1)}^2}{a_{kk}}$ – для k -го ограничения [30] и

$\sum_{i \in J_k} \frac{\tilde{a}_{i(m+w+1)}^2}{\tilde{a}_{ii}}$ – для блока, где J_k имеет тот же смысл, что и в случае

традиционной многооткликowej модели, а \tilde{a}_{ij} являются i -ми элементами j -го столбца матрицы A , выметенной по диагональным элементам, которые соответствуют предыдущим по отношению к i индексам в последовательности J_k .

5.2.3. ИССЛЕДОВАНИЕ АЛГОРИТМОВ ВЫБОРА СТРУКТУРЫ

В [54, 55] описано исследование алгоритмов, рассмотренных в п. 5.2.1 и 5.2.2, как методом статистических испытаний, так и на одной реализации ошибок. Исследование показало, что все усложнения простых стратегий повышают способность алгоритмов направленного перебора находить минимум и последовательности лучших структур различной сложности.

Рассмотрим результаты сравнения нескольких алгоритмов на одной реализации ошибок, отражающие нюансы их работы.

Алгоритмы исследуются на способность находить оптимальную структуру, соответствующую минимуму критерия, а также последовательности наилучших структур различной сложности.

Алгоритмы решения традиционной задачи выбора структуры исследовались на тех же моделях, что и критерии в п. 5.1.6. Рассмотрим случай традиционной многооткликowej модели.

В табл. 5.2–5.4 отражены процессы минимизации критерия DSI при уровне шума 0,05 и значении коэффициента корреляции элементов вектора ошибок 0,9 с помощью следующих алгоритмов:

1 – алгоритм включения с остановом после включения всех регрессоров, участвующих в переборе, $v = 1$;

2 – алгоритм включения с обмeнами и с остановом после включения всех регрессоров, участвующих в переборе;

3 – алгоритм включения с остановом после включения всех регрессоров, участвующих в переборе, $\nu = 3$.

Заметим, что аддитивная постоянная в процессе работы алгоритмов всегда присутствовала в модели.

Таблица 5.2

Процесс поиска минимума критерия DSI алгоритмом 1

Номер шага	Сложность	Действие	Структура	Значение критерия DSI
0	1**		10000000	503,416
1	2**	Включение	10010000	94,8231
2	3*	Включение	11010000	16,8007
3	4**	Включение	11110000	0,253278
4	5**	Включение	11110010	0,0102066
5	6**	Включение	11110011	0,0108482
6	7**	Включение	11111011	0,0116516
7	8**	Включение	11111111	0,0208706

Таблица 5.3

Процесс поиска минимума критерия DSI алгоритмом 2

Номер шага	Сложность	Действие	Структура	Значение критерия DSI
0	1**		10000000	503,416
1	2**	Включение	10010000	94,8231
2	3	Включение	11010000	16,8007
2а	3**	Обмен	11100000	2,39418
3	4**	Включение	11110000	0,253278
4	5**	Включение	11110010	0,0102066
5	6**	Включение	11110011	0,0108482
6	7**	Включение	11111011	0,0116516
7	8**	Включение	11111111	0,0208706

В таблицах указаны структуры, выбранные в качестве наилучших на текущем шаге, и соответствующие им значения критерия. Присут-

ствии в модели регрессора обозначается цифрой 1, а отсутствие – цифрой 0. Регрессоры следуют в порядке, указанном в уравнениях (5.39) и (5.40). Таким образом, истинной структуре соответствует последовательность 11110010. Эта же структура является оптимальной с точки зрения рассматриваемого критерия. Звездочками во втором столбце отмечены лучшие структуры различной сложности, найденные алгоритмом: одной звездочкой – не совпадающие с наилучшей структурой своей сложности (найденной полным перебором), двумя звездочками – при совпадении.

Кроме того, рассмотрим результаты работы алгоритма 4 (алгоритма включения-исключения с начальной структурой, содержащей только регрессоры с протекцией), который произвел шаги, аналогичные шагам 0–5 алгоритма 1.

Таблица 5.4

Процесс поиска минимума критерия DSI алгоритмом 3

Номер шага	Сложность	Действие	Структура	Значение критерия DSI
0	1**		10000000	503,416
1	2**	Включение	10010000	94,8231
	2	Включение	11000000	143,823
	2	Включение	10100000	477,71
2	3**	Включение	11100000	2,39418
	3	Включение	11010000	16,8007
	3	Включение	10010010	46,3889
3	4**	Включение	11110000	0,253278
	4	Включение	11010010	0,728649
	4	Включение	10110010	2,54252
4	5**	Включение	11110010	0,0102066
	5	Включение	11111000	0,285066
	5	Включение	11110001	0,317844
5	6**	Включение	11110011	0,0108482
	6	Включение	11111010	0,0108947
	6	Включение	11110110	0,0158677
6	7**	Включение	11111011	0,0116516
	7	Включение	11111110	0,0168232
	7	Включение	11110111	0,0197665
7	8**	Включение	11111111	0,0208706

Алгоритмом 4 наилучшая структура сложности 3 была найдена в результате генерирования дополнительных по отношению к алгоритму 1 структур на шаге 4. Алгоритмом 2 эта же структура найдена за счет обмена на шаге 2а, а алгоритмом 3 – за счет использования дополнительных опорных структур (для успеха достаточно было одной дополнительной опорной структуры).

Заметим, что структура 11010000, которая выбрана лучшей алгоритмом 1 на шаге 2, является наилучшей структурой сложности 3 по теоретическим критериям прогнозирования на выборках W и C , но не является таковой с точки зрения критерия DSI . Для алгоритма 3 эта структура является второй по качеству на шаге 2.

Под теоретическим критерием прогнозирования традиционной многооткликной модели понимается критерий

$$TCP(Z) = \mathbf{E} \left\{ \mathbf{E}Y_Z - X_Z \hat{\theta}^T \Sigma^{-1} \otimes I_N \mathbf{E}Y_Z - X_Z \hat{\theta} \right\},$$

где Z – обозначение выборки (W или C); Y_Z, X_Z – вектор значений откликов и матрица значений регрессоров в одномерной модели, полученной преобразованием типа (1.20).

Рассмотрим представленные в табл. 5.5–5.8 результаты поиска минимума критерия $ZReg$ при уровне шума $\rho = 0,65$ и значении коэффициента корреляции элементов вектора ошибок 0,9. Исследовались следующие алгоритмы:

1 – алгоритм исключения с остановом после исключения всех регрессоров, участвующих в переборе, $v = 1$;

2 – алгоритм исключения с обменами и с остановом после исключения всех регрессоров, участвующих в переборе;

3 – алгоритм исключения с остановом после исключения всех регрессоров, участвующих в переборе, $v = 3$;

4 – алгоритм включения-исключения с начальной структурой, содержащей все регрессоры.

Аддитивная постоянная в процессе работы алгоритмов всегда присутствовала в модели.

Заметим, что оптимальной по рассматриваемому критерию является структура 11010011, не совпадающая с истинной.

Таблица 5.5

Процесс поиска минимума критерия ZReg алгоритмом 1

Номер шага	Сложность	Действие	Структура	Значение критерия ZReg
0	8**		11111111	255,409
1	7**	Исключение	11111011	73,6751
2	6**	Исключение	11111010	61,6922
3	5*	Исключение	11110010	59,8291
4	4**	Исключение	11010010	53,2259
5	3**	Исключение	11010000	55,3604
6	2**	Исключение	10010000	65,5258
7	1**	Исключение	10000000	95,7934

Таблица 5.6

Процесс поиска минимума критерия ZReg алгоритмом 2

Номер шага	Сложность	Действие	Структура	Значение критерия ZReg
0	8**		11111111	255,409
1	7**	Исключение	11111011	73,6751
2	6**	Исключение	11111010	61,6922
3	5	Исключение	11110010	59,8291
3а	5**	Обмен	11010011	50,8259
4	4**	Исключение	11010010	53,2259
5	3**	Исключение	11010000	55,3604
6	2**	Исключение	10010000	65,5258
7	1**	Исключение	10000000	95,7934

Алгоритмом 1 в качестве минимума получена наилучшая структура сложности 4, в то время как минимуму соответствует наилучшая структура сложности 5. Алгоритм 2 находит минимум за счет обмена на шаге 3, алгоритм 3 – за счет использования дополнительных опорных структур (для успеха достаточно было одной дополнительной опорной структуры), алгоритм 4 – за счет проведения шага включения (шага 5).

Таблица 5.7

Процесс поиска минимума критерия ZReg алгоритмом 3

Номер шага	Сложность	Действие	Структура	Значение критерия ZReg
0	8**		11111111	255,409
1	7**	Исключение	11111011	73,6751
	7	Исключение	11011111	74,5163
	7	Исключение	11110111	190,67
2	6**	Исключение	11111010	61,6922
	6	Исключение	11010111	63,7192
	6	Исключение	11011011	65,127
3	5**	Исключение	11010011	50,8259
	5	Исключение	11110010	59,8291
	5	Исключение	11011010	59,8548
4	4**	Исключение	11010010	53,2259
	4	Исключение	11110000	57,483
	4	Исключение	11010001	59,9532
5	3**	Исключение	11010000	55,3604
	3	Исключение	10010010	66,1218
	3	Исключение	10110000	67,4563
6	2**	Исключение	10010000	65,5258
	2	Исключение	11000000	92,8066
	2	Исключение	10100000	98,5983
7	1**	Исключение	10000000	95,7934

Таблица 5.8

Процесс поиска минимума критерия ZReg алгоритмом 4

Номер шага	Сложность	Действие	Структура	Значение критерия ZReg
0	8**		11111111	255,409
1	7**	Исключение	11111011	73,6751
2	6**	Исключение	11111010	61,6922
3	5	Исключение	11110010	59,8291
4	4**	Исключение	11010010	53,2259
5	5**	Включение	11010011	50,8259

Заметим также, что «неправильная» структура сложности 5, полученная алгоритмом 1, оказывается оптимальной по теоретическим критериям, более того – это истинная структура.

Перейдем к рассмотрению алгоритмов решения обобщенной задачи выбора структуры. Сравнивались следующие алгоритмы:

1 – алгоритм включения-исключения с начальной структурой, содержащей только регрессоры, вводящие аддитивные постоянные;

2 – алгоритм 1 с дополнительными шагами обмена, проводимыми в соответствии со стратегией 1;

3 – алгоритм 1 с дополнительными шагами обмена, проводимыми в соответствии со стратегией 2;

4 – алгоритм включения-исключения с начальной структурой, содержащей все регрессоры и блоки;

5 – алгоритм 4 с дополнительными шагами обмена, проводимыми в соответствии со стратегией 1;

6 – алгоритм 4 с дополнительными шагами обмена, проводимыми в соответствии со стратегией 2;

7 – алгоритм включения-исключения по F -статистике с той же начальной структурой, что у алгоритма 1, и пороговыми значениями 0,95 для включения и исключения;

8 – алгоритм включения-исключения по F -статистике с той же начальной структурой, что у алгоритма 4, и пороговыми значениями 0,95 для включения и исключения.

Исследуем работу алгоритмов выбора структуры для традиционной многооткликовой модели с качественными факторами.

Предполагается, что на два отклика аддитивно влияют три фактора α , β , γ , имеющих три уровня варьирования, и один фактор η , изменяющийся на двух уровнях. Используются следующие истинные значения:

$$\mu = 16\frac{7}{15}, 15\frac{4}{15}, \quad \alpha_1 = 2,6, 2,2, \quad \alpha_2 = 1,8, 1,6, \quad \alpha_3 = -4,4, -3,8,$$

$$\beta_1 = -2\frac{2}{30}, 2,4, \quad \beta_2 = 1\frac{28}{30}, -2,2, \quad \beta_3 = \frac{4}{30}, -0,2,$$

$$\gamma_1 = 0,2, \frac{16}{30}, \quad \gamma_2 = -0,1, -\frac{8}{30}, \quad \gamma_3 = -0,1, -\frac{8}{30},$$

$$\eta_1 = 0, 0, \quad \eta_2 = 0, 0,$$

где μ – аддитивная постоянная; $\alpha_i, \beta_i, \gamma_i, \eta_i$ – эффекты i -х уровней соответствующих факторов. Эффекты уровней факторов удовлетворяют при этом идентифицирующим ограничениям типа нулевой суммы эффектов уровней для каждого фактора.

В рамках решения обобщенной задачи выбора структуры формируются четыре блока, каждый из которых объединяет все фиктивные переменные, стоящие при эффектах уровней одного качественного фактора, и соответствующее идентифицирующее ограничение.

В качестве ограничений, участвующих в переборе, выступают

$$\alpha_1 - \alpha_3 = 0, \quad \alpha_2 - \alpha_3 = 0, \quad \beta_1 - \beta_3 = 0, \quad \beta_2 - \beta_3 = 0, \quad \gamma_1 - \gamma_3 = 0,$$

$$\gamma_2 - \gamma_3 = 0.$$

Заметим, что аддитивная постоянная в процессе перебора структур всегда присутствовала в модели.

Выборка W состояла из $N = 48$ наблюдений (по 24 наблюдения в выборках A и B). В выборку C включены шесть наблюдений, дополняющих выборку W до реализации полного факторного эксперимента.

В табл. 5.9–5.11 приводятся результаты работы алгоритмов, соответствующие случаю поиска минимума критерия $ZReg$ при уровне шума 0,65 и значении коэффициента корреляции элементов вектора ошибок 0,9.

Таблица 5.9

**Процесс поиска минимума критерия $ZReg$ алгоритмом 2
для обобщенной задачи выбора структуры**

Номер шага	Сложность	Действие	Значение критерия $ZReg$
0	1	Начальная структура содержит μ	671,186
1	3	Включение β	198,545
2	5	Включение α	114,569
3	6	Обмен – включение γ и $\gamma_1 - \gamma_3 = 0$	113,977

Таблица 5.10

**Процесс поиска минимума критерия ZReg алгоритмом 3
для обобщенной задачи выбора структуры**

Номер шага	Сложность	Действие	Значение критерия ZReg
0	1	Начальная структура содержит μ	671,186
1	3	Включение β	198,545
1а	4	Обмен – включение α и $\alpha_2 - \alpha_3 = 0$	167
1б	5	Обмен – включение γ и $\gamma_1 - \gamma_3 = 0$	165,495
1в	5	Обмен – включение $\gamma_2 - \gamma_3 = 0$ и исключение $\alpha_2 - \alpha_3 = 0$	114,569
2	6	Исключение $\gamma_2 - \gamma_3 = 0$	113,977

Таблица 5.11

**Процесс поиска минимума критерия ZReg алгоритмами 4–6
для обобщенной задачи выбора структуры**

Номер шага	Сложность	Действие	Значение критерия ZReg
0	8	Начальная структура содержит $\mu, \alpha, \beta, \gamma, \eta$	132,523
1	7	Включение $\gamma_1 - \gamma_3 = 0$	124,122
2	6	Исключение η	113,977

Оптимальна при этом структура, содержащая аддитивную постоянную μ , факторы α, β, γ и ограничение $\gamma_1 - \gamma_3 = 0$. В табл. 5.9–5.11 указаны действия, ведущие к формированию лучшей структуры текущего шага, а также соответствующее этой структуре значение критерия.

Алгоритм 1 произвел шаги, аналогичные шагам 0–2 алгоритма 2, при этом лучшей была выбрана структура сложности 5, имеющая более высокое качество по теоретическим критериям прогнозирования, чем структура, соответствующая минимуму. Минимум был найден алгоритмом 2 за счет одного дополнительного шага обмена.

Алгоритм 3 также нашел минимум за счет проведения шагов обмена, однако действовал иначе: большая часть выполненных шагов – это шаги обмена. В то же время алгоритм 6, в который заложена та же стратегия проведения обменов, действовал аналогично алгоритму 4, не осуществив ни одного обмена.

Алгоритмы 7 и 8 выбрали ту же структуру, что и алгоритм 1 при минимизации критерия $ZReg$. Заметим, что эта структура является оптимальной по критериям ZMC и ZSI , но не является оптимальной с точки зрения критерия $ZReg$. Последовательность шагов данных алгоритмов была следующей: алгоритм 7 последовательно включал факторы β и α , а алгоритм 8 последовательно исключал факторы γ и η .

5.3. ВЫБОР СТРУКТУРЫ ПРИ НЕГАУССОВСКИХ И ЗАВИСИМЫХ ОШИБКАХ

5.3.1. РОБАСТНЫЕ КРИТЕРИИ ВЫБОРА СТРУКТУРЫ

Перейдем к рассмотрению подходов, которые могут использоваться при выборе структуры регрессионной модели в условиях нарушения классических предположений.

Прежде всего заметим, что в условиях негауссовских ошибок при ММП-оценивании параметров выбор структуры может осуществляться с использованием информационных критериев AIC , BIC , HQ и др.

Для выбора модели однооткликowej регрессии в условиях нестройной нормальности при случайных регрессорах разработан ряд робастных подходов. В частности, в [182] рассматривается обобщенный критерий скользящего контроля; посредством замены МНК-оценок робастными и использования модифицированных F -статистик можно получить робастный алгоритм пошаговой регрессии [99]. Робастные модификации информационного критерия Акаике и критерия Меллоуса получены для более общих моделей [131, 181, 183, 190].

В [181] рассмотрена робастная модификация информационного критерия Акаике для следующей общей модели. Пусть по независимым одинаково распределенным наблюдениям

$$z_i, \quad i=1, \dots, N,$$

строится M -оценка $\hat{\phi}$ вектора ϕ путем минимизации функции

$$Q(\phi) = \sum_{i=1}^N \rho(z_i, \phi),$$

причем используемая структура модели включает в себя истинную, тогда робастный критерий Акаике (AICR) имеет вид

$$\text{AICR} = 2 \sum_{i=1}^N \rho(z_i, \hat{\phi}) + \alpha, \quad (5.47)$$

где $\alpha = 2 \text{tr} \Omega_1^{-1} \Omega_2$, $\Omega_1 = \mathbf{E} \left[\frac{\partial \Psi}{\partial \phi} \right]$, $\Omega_2 = \mathbf{E} \left[\Psi \Psi^T \right]$, $\Psi = \frac{\partial \rho}{\partial \phi}$.

В п. 2.3.1 рассмотрена однородная многооткликковая регрессионная модель с засоренным многомерным нормальным распределением ошибок. Воспользуемся формулой (5.47) для получения робастного критерия выбора структуры данной модели при использовании ЭВ-оценок параметров уравнения регрессии и ковариационной матрицы. Заметим, что ограничения на параметры уравнений регрессии не наложены.

Функция потерь имеет вид (1.55). Обозначим $\text{vec} \Sigma$ – вектор, составленный из $\frac{l(l+1)}{2}$ различающихся элементов матрицы Σ . Тогда вектор всех неизвестных параметров модели ϕ включает элементы векторов θ и $\text{vec} \Sigma$.

Найдем вид матриц Ω_1 , Ω_2 . Имеем для t -го наблюдения

$$\Psi_{\theta}(r_t, \Sigma) = \frac{\partial \rho(q_t, \Sigma)}{\partial \theta} = -|\Sigma|^{-\frac{c}{2(1+c)}} \exp \left\{ -\frac{c}{2} q_t \right\} c X_t^T \Sigma^{-1} r_t,$$

$$\Psi_{\Sigma_{ij}}(r_t, \Sigma) = \frac{\partial \rho(q_t, \Sigma)}{\partial \Sigma_{ij}} = |\Sigma|^{-\frac{c}{2(1+c)}} \exp\left\{-\frac{c}{2} q_t\right\} \frac{c}{2} (2 - \delta_{ij}) \times$$

$$\times \left[\frac{1}{1+c} \Sigma_{ij}^{-1} - \sum_{u=1}^l \Sigma_{iu}^{-1} r_{t,u} \sum_{v=1}^l \Sigma_{jv}^{-1} r_{t,v} \right],$$

где r_t – остаток для t -го наблюдения, $r_{t,u}$ – его u -й элемент, δ_{ij} – символ Кронекера. Можно показать, что матрицы Ω_1 , Ω_2 имеют блочно-диагональный вид:

$$\Omega_i = \begin{bmatrix} \Omega_{i,\theta} & 0 \\ 0 & \Omega_{i,\Sigma} \end{bmatrix},$$

где $\Omega_{i,\theta}$, $\Omega_{i,\Sigma}$ – подматрицы, соответствующие параметрам θ и $\text{vec} \Sigma$, $i=1,2$,

$$\text{tr} \Omega_{1,\theta}^{-1} \Omega_{2,\theta} = c |\Sigma|^{-\frac{c}{2(1+c)}} p \left(\frac{1+c}{1+2c} \right)^{1+l/2},$$

$$[\Omega_{1,\Sigma}]_{i,j,u,v} = \mathbf{E} \frac{\partial \Psi_{\Sigma_{ij}}}{\partial \Sigma_{uv}} = |\Sigma|^{-\frac{c}{2(1+c)}} \frac{c}{4} \frac{(2-\delta_{ij})(2-\delta_{uv})}{(1+c)^{2+l/2}} \left[\Sigma_{ui}^{-1} \Sigma_{vj}^{-1} + \Sigma_{uj}^{-1} \Sigma_{vi}^{-1} \right],$$

$$[\Omega_{2,\Sigma}]_{i,j,u,v} = \mathbf{E} \Psi_{\Sigma_{ij}} \Psi_{\Sigma_{uv}} = |\Sigma|^{-\frac{c}{1+c}} \frac{c^2}{4} (2-\delta_{ij})(2-\delta_{uv}) \times$$

$$\times \frac{1}{(1+2c)^{2+l/2}} \left[\left(\frac{c}{1+c} \right)^2 \Sigma_{ij}^{-1} \Sigma_{uv}^{-1} + \Sigma_{ui}^{-1} \Sigma_{vj}^{-1} + \Sigma_{uj}^{-1} \Sigma_{vi}^{-1} \right],$$

где p – число регрессоров в рассматриваемой структуре модели.

Представим $[\Omega_{2,\Sigma}]_{i,j,u,v}$ в виде суммы

$$[\Omega_{2,\Sigma}]_{i,j,u,v} = [\Omega_{2,\Sigma}^{(1)}]_{i,j,u,v} + [\Omega_{2,\Sigma}^{(2)}]_{i,j,u,v},$$

где

$$[\Omega_{2,\Sigma}^{(1)}]_{i,j,u,v} = |\Sigma|^{\frac{-c}{1+c}} \frac{c^2}{4} (2 - \delta_{ij})(2 - \delta_{uv}) \frac{1}{(1+2c)^{2+l/2}} [\Sigma_{ui}^{-1} \Sigma_{vj}^{-1} + \Sigma_{uj}^{-1} \Sigma_{vi}^{-1}],$$

$$[\Omega_{2,\Sigma}^{(2)}]_{i,j,u,v} = |\Sigma|^{\frac{-c}{1+c}} \frac{c^2}{4} (2 - \delta_{ij})(2 - \delta_{uv}) \frac{1}{(1+2c)^{2+l/2}} \left(\frac{c}{1+c} \right)^2 \Sigma_{ij}^{-1} \Sigma_{uv}^{-1},$$

тогда

$$\text{tr} \Omega_{1,\Sigma}^{-1} \Omega_{2,\Sigma}^{(1)} = c |\Sigma|^{\frac{-c}{2(1+c)}} \frac{l(l+1)}{2} \left(\frac{1+c}{1+2c} \right)^{2+l/2},$$

$$\text{tr} \Omega_{1,\Sigma}^{-1} \Omega_{2,\Sigma}^{(2)} = c |\Sigma|^{\frac{-c}{2(1+c)}} \left(\frac{1+c}{1+2c} \right)^{2+l/2} \left(\frac{c}{1+c} \right)^2 \text{tr} \tilde{\Omega}_1^{-1} \tilde{\Omega}_2,$$

где

$$[\tilde{\Omega}_1]_{i,j,u,v} = \frac{1}{4} (2 - \delta_{ij})(2 - \delta_{uv}) [\Sigma_{ui}^{-1} \Sigma_{vj}^{-1} + \Sigma_{uj}^{-1} \Sigma_{vi}^{-1}],$$

$$[\tilde{\Omega}_2]_{i,j,u,v} = \frac{1}{4} (2 - \delta_{ij})(2 - \delta_{uv}) \Sigma_{ij}^{-1} \Sigma_{uv}^{-1}.$$

Матрица $\tilde{\Omega}_1$ является информационной матрицей, соответствующей ММП-оценке $\hat{\Sigma}$ в нормальной модели с нулевым математическим

ожиданием и ковариационной матрицей Σ , поэтому элементы матрицы $\tilde{\Omega}_1^{-1}$ имеют вид [189]

$$\left[\tilde{\Omega}_1^{-1} \right]_{i,j,u,v} = \Sigma_{ui} \Sigma_{vj} + \Sigma_{uj} \Sigma_{vi},$$

тогда, расписав операцию следа произведения матриц и проведя несложные преобразования, получим

$$\text{tr} \tilde{\Omega}_1^{-1} \tilde{\Omega}_2 = \frac{l}{2}.$$

В результате имеем

$$\alpha = 2c |\Sigma|^{-\frac{c}{2(1+c)}} \left(\frac{1+c}{1+2c} \right)^{1+l/2} \left\{ p + \left(\frac{1+c}{1+2c} \right) \left[\frac{l(l+1)}{2} + \left(\frac{c}{1+c} \right)^2 \frac{l}{2} \right] \right\}.$$

Неизвестное значение матрицы Σ заменяется оценкой по текущей структуре.

Перейдем к рассмотрению критериев робастного выбора структуры однородной однооткликowej регрессионной модели (1.1) без ограничений на параметры уравнений регрессии при использовании ЭВ- L_v -оценок параметров для случая $v \geq 2$ (см. п. 2.4.2, индекс группы здесь и далее опущен).

Получим формулу критерия AICR (5.47) при ЭВ- L_v -оценивании регрессионной модели (1.1). В основу построения критерия положим оптимизационную формулировку (2.44) для оценивания параметров θ и λ , приводящую к функции потерь для t -го наблюдения

$$\rho(r_t, \lambda) = -\frac{1}{\delta} \lambda^{\delta^2/(1+\delta)} g^\delta(r_t, \lambda, v),$$

где $r_t = y_t - f^T(x_t)\theta$ – остаток. Параметр формы v при этом считается известным и одинаковым для всех структур; на практике следует использовать его некоторую оценку, например, оценку при полной модели. В результате вектор неизвестных параметров модели ϕ включает элементы вектора θ и параметр λ .

Как и в случае многооткликовой модели, матрицы Ω_1 , Ω_2 имеют блочно-диагональный вид с блоками, соответствующими параметрам θ и λ . Имеем для t -го наблюдения

$$\Psi_{\theta}(r_t, \lambda) = \frac{\partial \rho(r_t, \lambda)}{\partial \theta} = -\frac{\lambda^{\delta^2/(1+\delta)} \exp(-\delta|r_t|^v/\lambda^v) v|r_t|^{v-1} \text{sign}(r_t)}{2\lambda\Gamma(1+1/v) \delta} f(x_t),$$

$$\Psi'_{\theta}(r_t, \lambda) = \frac{\partial \Psi_{\theta}(r_t, \lambda)}{\partial \theta} = \frac{\lambda^{\delta^2/(1+\delta)-v} v \exp(-\delta|r_t|^v/\lambda^v)}{2\lambda\Gamma(1+1/v) \delta} \times$$

$$\times \left[(v-1)|r_t|^{v-2} - \frac{\delta v}{\lambda^v} |r_t|^{2v-2} \right] f(x_t) f^T(x_t),$$

$$\Psi_{\lambda}(r_t, \lambda) = \frac{\partial \rho(r_t, \lambda)}{\partial \lambda} = \frac{\lambda^{\delta^2/(1+\delta)-1} \exp(-\delta|r_t|^v/\lambda^v)}{2\lambda\Gamma(1+1/v) \delta} \left[\frac{1}{1+\delta} - \frac{v|r_t|^v}{\lambda^v} \right],$$

$$\Psi'_{\lambda}(r_t, \lambda) = \frac{\partial \Psi_{\lambda}(r_t, \lambda)}{\partial \lambda} = \frac{\lambda^{\delta^2/(1+\delta)-2} \exp(-\delta|r_t|^v/\lambda^v)}{2\lambda\Gamma(1+1/v) \delta} \times$$

$$\times \left\{ \frac{v|r_t|^v}{\lambda^v} \left[\frac{1+3\delta}{1+\delta} + v \right] - \frac{1+2\delta}{(1+\delta)^2} - \frac{v^2\delta|r_t|^{2v}}{\lambda^{2v}} \right\},$$

$$\text{tr} \left[\mathbf{E} \Psi'_{\theta}{}^{-1} \mathbf{E} \Psi_{\theta} \Psi_{\theta}^T \right] = \frac{p\lambda^{\delta^2/(1+\delta)}}{2\lambda\Gamma(1+1/v) \delta} \left[\frac{\delta+1}{2\delta+1} \right]^{2-1/v},$$

$$\mathbf{E} \Psi_{\lambda}^2 = \frac{\lambda^{2\delta^2/(1+\delta)-2}}{2\lambda\Gamma(1+1/v) 2\delta} \left[\frac{\delta^2}{(1+\delta)^2} + v \right] \frac{1}{(2\delta+1)^{2+1/v}},$$

$$\mathbf{E}\Psi'_\lambda = \frac{\lambda^{\delta^2/(1+\delta)-2\nu}}{2\lambda\Gamma(1+1/\nu)} \frac{1}{\delta^{1+\delta} (1+\delta)^{2+1/\nu}}.$$

В результате с точностью до несущественного множителя критерий AICR имеет вид

$$\begin{aligned} \text{AICR} = & \left\{ -\frac{1}{\delta} \sum_{i=1}^N \exp \left[-\delta \frac{|r_i(\hat{\theta})|^\nu}{\lambda^\nu} \right] + p \left[\frac{\delta+1}{2\delta+1} \right]^{2-1/\nu} + \right. \\ & \left. + \left[\frac{\delta+1}{2\delta+1} \right]^{2+1/\nu} \left[\frac{\delta^2}{(\delta+1)^2} \frac{1}{\nu} + 1 \right] \right\} \lambda^{-\delta/(1+\delta)}. \end{aligned}$$

Неизвестное значение параметра λ заменяется соответствующей ЭВ- L_ν -оценкой. В случае неопениваемого параметра масштаба данный критерий можно использовать и при стойких оценках параметров уравнения регрессии.

В [125, 126] предложено осуществлять выбор структуры однооткликной регрессии путем минимизации критерия типа MSEF – ожидаемой среднеквадратической погрешности регрессионного предсказания для нового наблюдения (x_0, y_0) :

$$s^2 = \mathbf{E} (y_0 - \hat{y}_0)^2.$$

Для оценок параметров модели (1.1), получаемых путем решения системы оценочных уравнений (1.47), на его основе может быть введен критерий

$$\text{RFPE} = \sigma^2 + \frac{\Delta}{N},$$

где $\Delta = p \frac{\mathbf{E}\psi^2}{\mathbf{E}\psi'^2}$, $\psi' = \frac{\partial\psi(z/\lambda)}{\partial z}$.

В условиях нормального распределения данный критерий эквивалентен критерию финальной ошибки прогнозирования FRE при использовании МНК-оценок неизвестных параметров модели. В связи с этим он может быть назван робастным критерием финальной ошибки прогнозирования, что отражено в его обозначении RFRE.

Согласно (4.19) и (4.20) критерий RFPE имеет вид

$$\text{RFPE} = \lambda^2 \left\{ \frac{\Gamma(3/\nu)}{\Gamma(1/\nu)} + \frac{p}{N} \frac{\Gamma(1/\nu)}{\nu^2 \Gamma(2-1/\nu)} \left[\frac{(\delta+1)^2}{2\delta+1} \right]^{2-1/\nu} \right\}.$$

Неизвестные значения параметров λ и ν заменяются соответствующими ЭВ- L_ν -оценками. Данный критерий может быть применен и при использовании стойкой оценки, являющейся специальным случаем ЭВ- L_ν -оценок.

Перейдем к рассмотрению критериев робастного выбора структуры модели, основанных на использовании экзаменационной выборки.

При построении критериев в качестве мер ошибок предсказания могут применяться робастные функции потерь, как в [182]. В том случае, если для M -оценок нет оптимизационных формулировок вида (1.46), критерии выбора структуры однооткликowej модели можно строить на основе значений оценок дисперсии ошибок наблюдений (преимущественно для нормальных моделей). Для многооткликowej модели (1.19), (1.38) критерии могут строиться с использованием сверток робастных оценок ковариационной матрицы, таких как детерминант, след и т. д.

Для получения конкретных критериев, основанных на применении экзаменационной выборки, воспользуемся интерпретацией критериев, приведенной в п. 5.1.2. В этом случае вводится некоторая вспомогательная модель или несколько моделей, а также специальные линейные ограничения-равенства на параметры вспомогательной модели. В результате оптимальной будет считаться та структура, для которой вспомогательная модель со специальными ограничениями имеет наибольшее качество.

Данный способ конструирования критериев, обобщающий принцип использования экзаменационной выборки, назовем *принципом варьирования модели*.

Для многооткликковой модели, представленной в виде (1.19), вспомогательные модели, специальные ограничения и определения критериев приведены в табл. 5.1.

В соответствии с табл. 5.1 можно определить симметричные и несимметричные критерии регулярности и стабильности. При этом $\hat{\Theta}_W$, $\hat{\Theta}_A$, $\hat{\Theta}_B$ – робастные оценки параметров Θ по выборкам W , A , B соответственно; G – значение функции Q качества оценок параметров (или значение свертки оценки ковариационной матрицы Σ), соответствующее вспомогательной модели без специальных ограничений; H – изменение G , вызванное введением в модель специальных ограничений.

Критерии могут быть построены и на основе разности значений критерия для вспомогательной модели со специальными ограничениями и без них. В этом случае критерии несмещенности могут быть построены в соответствии с определениями из первой и четвертой строк табл. 5.1, критерии вариативности – в соответствии с определениями из третьей и шестой строк.

Для симметричного критерия регулярности можно ввести еще одну интерпретацию, отличную от приведенной в табл. 5.1: используются две вспомогательные модели

$$Y_A = X_A \Theta_1 + E_A, \quad Y_B = X_B \Theta_2 + E_B \quad (5.48)$$

со специальными ограничениями

$$\Theta_1 = \hat{\Theta}_B, \quad \Theta_2 = \hat{\Theta}_A \quad (5.49)$$

и критерий строится как сумма значений функции качества оценок параметров на каждой из подвыборок $Q_A + Q_B$ (или как сумма свертки оценок ковариационных матриц для подвыборок), что может не совпадать с исходным вариантом, в котором критерий строится на основе значения функции качества оценок параметров на всей выборке (свертки оценки ковариационной матрицы для всей выборки).

В результате критерий качества структуры вычисляется по следующей схеме:

1) вычисляются оценки, которые необходимы для формирования специальных ограничений;

2) вычисляются, если необходимо, оценки параметров в рамках вспомогательных моделей со специальными ограничениями и без них;

3) вычисляются значения функции качества или значения свертки оценки ковариационной матрицы ошибок наблюдений, на основании их оценивается качество структуры.

Заметим, что можно считать варьируемыми и такие элементы модели, как мешающие параметры распределения ошибок наблюдений (например, параметры масштаба и формы), и использовать их при формировании специальных ограничений.

Например, при ЭВ- L_ν -оценивании можно предположить, что для вспомогательной модели ошибки в подвыборках A и B имеют законы распределения с параметрами λ_A, ν_A и λ_B, ν_B соответственно. Тогда, например, для критерия $\text{Reg}(B)$, построенного с использованием функции потерь, можно ввести ограничения

$$\lambda_B = \hat{\lambda}_A, \quad \nu_B = \hat{\nu}_A.$$

Одну из модификаций обобщенного критерия скользящего контроля (cross-validation), в которой используется метод Монте-Карло, можно рассматривать как стохастический вариант несимметричного критерия регулярности: наблюдения разбиваются случайным образом на две части и вычисляется значение критерия $\text{Reg}(B)$; данные действия повторяются заданное число раз с усреднением вычисленных значений критерия.

Заметим, что критерии, основанные на использовании принципа варьирования модели, могут быть использованы при решении обобщенной задачи выбора структуры, а также в условиях неоднородных данных.

5.3.2. ИССЛЕДОВАНИЕ РОБАСТНЫХ КРИТЕРИЕВ ВЫБОРА СТРУКТУРЫ

С целью проверки работоспособности предложенных подходов к выбору структуры было проведено экспериментальное исследование методом статистических испытаний. Решалась традиционная задача выбора структуры.

Рассмотрим результаты исследования критериев выбора структуры однооткликowej модели с использованием ЭВ- L_V -оценивания. Набор данных состоял из 200 наблюдений. Уравнение регрессии имело вид

$$y = 10 + x + 0,2x^2 + 0x^3 + 0x^4 + e,$$

где $x \in [-1, 1]$, с нулевыми коэффициентами указаны регрессоры, не влияющие на отклик, но присутствующие в наборе регрессоров, из которых выбирается оптимальный набор.

Распределение ошибок наблюдений представляло собой смесь двух распределений. Первое (основное) распределение в смеси было двусторонним экспоненциальным с параметром формы 3 и дисперсией, варьируемой в эксперименте. Дисперсия определялась по заданному уровню шума, который вычисляется как квадратный корень из отношения дисперсии к среднеквадратическому значению разности незашумленного отклика и его среднего значения в выборке. Вторым (засоряющим) было распределение модуля случайной величины, имеющей двустороннее экспоненциальное распределение с параметром формы 3 и дисперсией, в 25 раз большей дисперсии основного распределения. Вероятности компонентов смеси равнялись 0,9 и 0,1 для основного и засоряющего распределений соответственно.

Использовались ЭВ- L_V -оценки с параметром оценочной функции $\delta = 0,5$. Параметр формы всегда оценивался в рамках исходной модели. Выбор структуры осуществлялся методом полного перебора, при этом аддитивная постоянная всегда присутствовала в модели. В критериях на основе оценок дисперсии использовались специальные ограничения на параметры уравнения регрессии, а в критериях на основе функции качества оценок – специальные ограничения на параметры уравнения регрессии и параметры масштаба.

При вычислении критериев, использующих принцип варьирования модели, в данном и во всех приведенных ниже исследованиях производилось разбиение выборки на две подвыборки одинакового объема. При вычислении критерия GSI случайное разбиение выборки осуществлялось пять раз, для критерия Reg использовался альтернативный вариант его определения по модели (5.48), (5.49).

Значения частоты выбора истинной структуры в 20 испытаниях приведены в табл. 5.12. Верхнее число в ячейке соответствует исполь-

зованию при выборе структуры оценки параметра формы при полной модели, ниже – при оцениваемой модели.

Результаты показывают, что наибольшую работоспособность в целом показывают критерии AICR и GSI. В то же время заметим, что вычисление критерия GSI очень трудоемко, а работоспособность во многом зависит от выбора значений его параметров.

Рассмотрим результаты исследования критериев выбора структуры многооткликковой регрессии на двух моделях типа псевдонезависимых регрессий.

Модель 1. Рассматривалась двухоткликковая регрессионная модель вида

$$y_1 = 20 + x_1 + 0,5x_2 + 0x_3 + e_1, \quad y_2 = 20 + 2x_1 + 0,1x_2 + 0x_3 + e_2,$$

где y_1, y_2 – элементы вектора откликов; x_1, x_2, x_3 – значения факторов, лежащие в области $[-1, 1]$; e_1, e_2 – элементы вектора ошибок.

Т а б л и ц а 5.12

Работоспособность критериев

Уровень шума	AICR	RFPE	Критерий – функция потерь					Критерий – оценка дисперсии				
			Reg(B)	Reg	Stab(A)	Stab	GSI	Reg(B)	Reg	Stab(A)	Stab	GSI
0,05	0,6	0,45	0,35	0,45	0,25	0,15	0,7	0,4	0,4	0,3	0,2	0,65
	–	0,35	0,6	0,55	0,25	0,25	0,7	0,05	0,05	0,3	0,3	0,65
0,15	0,55	0,45	0,3	0,45	0,25	0,15	0,5	0,4	0,35	0,3	0,2	0,5
	–	0,25	0,5	0,45	0,2	0,2	0,6	0,1	0,05	0,15	0,15	0,55
0,25	0,55	0,35	0,3	0,35	0,2	0,15	0,45	0,3	0,3	0,25	0,2	0,45
	–	0,25	0,4	0,4	0,15	0,15	0,55	0,35	0,3	0,2	0,2	0,45
0,35	0,45	0,35	0,3	0,3	0,2	0,15	0,35	0,3	0,3	0,25	0,2	0,35
	–	0,2	0,25	0,25	0,1	0,1	0,35	0,25	0,2	0,1	0,1	0,35
0,45	0,4	0,35	0,25	0,2	0,15	0,15	0,2	0,3	0,2	0,25	0,2	0,35
	–	0,2	0,2	0,25	0,05	0,05	0,4	0,15	0,15	0,05	0,1	0,3
0,55	0,35	0,35	0,2	0,15	0,15	0,15	0,2	0,25	0,2	0,2	0,2	0,3
	–	0,15	0,2	0,3	0,1	0,1	0,35	0,15	0,15	0,1	0,15	0,3

Набор данных состоял из 40 наблюдений. Распределение вектора ошибок – смесь двух распределений. Основное распределение в смеси – нормальное с нулевым математическим ожиданием и дисперсиями элементов, определявшимися по уровню шума, коэффициент корреляции между элементами вектора ошибок был равен 0,5. Засоряющее распределение – нормальное с той же ковариационной матрицей, что и у основного распределения, и вектором математических ожиданий, все элементы которого были равны 7. Вероятности компонентов смеси равнялись 0,85 и 0,15 для основного и засоряющего распределений соответственно.

Сравнивались два подхода к выбору структуры – с использованием робастных ЭВ-оценок при $c=0,25$ и с использованием ММП для нормального распределения.

Выбор структуры осуществлялся методом полного перебора, при этом аддитивная постоянная и регрессор x_1 всегда присутствовали в модели каждого отклика. При вычислении критерия GSI выборка случайно разбивалась на части 5, 10 или 20 раз, при вычислении критерия Reg использовался альтернативный вариант его определения по модели (5.48), (5.49). В критериях на основе оценок ковариационной матрицы использовались специальные ограничения на параметры уравнения регрессии, а в критериях на основе функции качества оценок – специальные ограничения на параметры уравнения регрессии и значения ковариационной матрицы.

Подходы сравнивались по частоте выбора истинной и корректной структур в 20 испытаниях (к корректным относятся истинная и все избыточные структуры) и значению критерия прогнозирования

$$TSP = \mathbf{E} \left[\theta - \hat{\theta} \quad \tilde{X}^T \left[\Sigma^{-1} \otimes I_N \right] \tilde{X} \quad \theta - \hat{\theta} \right],$$

записанному для представления (1.35). При вычислении критерия TSP использовалась робастная оценка $\hat{\theta}$, а математическое ожидание заменялось усреднением по 20 испытаниям.

Значения названных характеристик приведены в табл. 5.13–5.15. Первое число в ячейке – частота выбора истинной структуры, второе – среднее значение критерия TSP для структур, признанных оптимальными, третье – частота выбора корректной структуры. Для обобщен-

ного критерия скользящего контроля с k разбиениями выборки используется обозначение $GSI(k)$. Варианты вычисления (типы) критериев имеют следующие обозначения: Q – критерий на основе функции качества оценок, D, Z, A – критерии в виде соответствующих сверток оценки дисперсионной матрицы (см. п. 5.1.2).

Таблица 5.13

Работоспособность информационных критериев (модель 1)

Уровень шума	AICR	AIC	BIC	HQ	KIC
0,05	0,75	0,05	0,1	0,05	0,1
	0,2280	57,64	57,64	57,64	57,64
	1	0,1	0,1	0,1	0,1
0,15	0,4	0,15	0,05	0,1	0,15
	15,16	14,45	13,92	14,24	14,25
	0,65	0,2	0,05	0,15	0,15

Таблица 5.14

Работоспособность критериев, основанных на принципе варьирования модели, при уровне шума 0,05 (модель 1)

Тип	ЭВ-оценки				ММП-оценки			
	Reg	GSI(5)	GSI(10)	GSI(20)	Reg	GSI(5)	GSI(10)	GSI(20)
Q	0,6	0,55	0,75	0,7	0	0,1	0	0
	0,3904	0,28999	0,2796	0,3326	110,1	144,5	172,7	150,9
	0,85	0,95	0,95	0,9	0,1	0,1	0	0,05
D	0,4	0,45	0,75	0,65	0,4	0,5	0,5	0,8
	4,735	0,2979	0,2813	0,3340	18,15	13,77	13,79	9,155
	0,9	0,95	0,95	0,9	0,55	0,6	0,55	0,8
Z	0,35	0,45	0,75	0,65	0,05	0,25	0,15	0,35
	4,738	0,2979	0,2813	0,3340	33,79	14,09	38,13	22,72
	0,9	0,95	0,95	0,9	0,2	0,3	0,2	0,35
A	0,4	0,35	0,7	0,65	0,05	0,1	0,05	0
	4,740	0,2992	0,2810	0,3339	29,45	14,14	49,08	27,49
	0,9	0,95	0,95	0,9	0,2	0,25	0,1	0

Таблица 5.15

Работоспособность критериев, основанных на принципе варьирования модели, при уровне шума 0,15 (модель 1)

Тип	ЭВ-оценки				ММП-оценки			
	Reg	GSI(5)	GSI(10)	GSI(20)	Reg	GSI(5)	GSI(10)	GSI(20)
Q	0,4	0,2	0,55	0,3	0,2	0,25	0,1	0
	15,05	14,57	15,28	14,37	18,32	15,56	17,83	16,94
	0,55	0,3	0,7	0,3	0,3	0,25	0,15	0
D	0,2	0,15	0,35	0,15	0,15	0,25	0,15	0,05
	16,31	15,11	14,81	14,04	18,32	15,56	17,39	17,07
	0,5	0,25	0,4	0,15	0,3	0,25	0,2	0,05
Z	0,2	0,1	0,35	0,15	0,15	0,25	0,1	0
	16,54	14,98	14,81	15,23	17,55	16,07	17,28	16,95
	0,6	0,2	0,45	0,2	0,2	0,25	0,15	0
A	0,2	0,1	0,25	0,15	0	0	0	0
	19,38	17,06	15,87	16,32	18,11	16,19	18,34	17,58
	0,5	0,15	0,4	0,2	0,25	0,2	0,1	0

Модель 2. Рассматривалась двухоткликовая регрессионная модель вида

$$y_1 = 20 + x + 0x^2 + e_1, \quad y_2 = 20 + 2x + 0x^2 + e_2,$$

где x – входная переменная.

Набор данных состоял из 40 наблюдений, при этом два наблюдения являлись выбросами. Распределение вектора ошибок основной части набора данных – двумерное нормальное, с теми же значениями параметров, что и в модели 1, значения входной переменной лежали в области $[-1,1]$. Наблюдения-выбросы были смоделированы в точках $x = -2$ и $x = -2,5$, элементы вектора ошибок имели то же распределение, что и в основной части выборки, но со сдвигом (3, 8) для первой точки и (7, 12) – для второй.

Таблицы 5.16 – 5.18, аналогичные таблицам 5.13 – 5.15, содержат значения показателей работоспособности критериев.

Таблица 5.16

Работоспособность информационных критериев (модель 2)

Уровень шума	AICR	AIC	BIC	HQ	KIC
0,05	0,4 0,4642 1	0 0,8636 1	0 0,8636 1	0 0,8636 1	0 0,8636 1
0,7	0,25 1,820 1	0 3,246 1	0 3,246 1	0 3,246 1	0 3,246 1

Результаты, приведенные в табл. 5.16 – 5.18, показывают работоспособность предложенных устойчивых подходов и их определенное преимущество перед традиционными. Нельзя однозначно сделать выбор в пользу тех или иных предложенных подходов, каждый из них показал определенное преимущество в одной из ситуаций.

В то же время трудоемкость вычисления критерия $GSI(k)$ быстро растет с увеличением k ; по этой причине может быть целесообразным предварительно отбирать множество структур-кандидатов с использованием менее трудоемких методов (как рекомендуется в [182]), а критерий $GSI(k)$ привлекать при окончательном выборе.

Подводя итоги исследования, можно сделать следующие выводы. Робастные методы целесообразно использовать в условиях сильного искажения наблюдений. Среди наиболее работоспособных в первую очередь можно назвать обобщенный критерий скользящего контроля; робастные критерии Акаике также часто имели высокую работоспособность. Вычисление обобщенного критерия скользящего контроля, однако, может быть слишком трудоемким, в этом случае целесообразно его применение при выборе оптимальной структуры из множества структур, предварительно отобранных с использованием менее трудоемких подходов.

Таблица 5.17

Работоспособность критериев, основанных на принципе варьирования модели, при уровне шума 0,05 (модель 2)

Тип	ЭВ-оценки				ММП-оценки			
	Reg	GSI(5)	GSI(10)	GSI(20)	Reg	GSI(5)	GSI(10)	GSI(20)
Q	0,7	0,7	0,8	0,8	0	0	0	0
	0,2944 1	0,3015 1	0,2566 1	0,2637 1	542,1 0	55,60 0,9	55,60 0,9	0,8636 1
D	0,25	0,7	0,75	0,75	0	0,1	0	0
	0,5971 1	0,3203 1	0,2907 1	0,2979 1	547,9 0	355,7 0,35	438,8 0,2	438,8 0,2
Z	0,25	0,65	0,75	0,75	0	0	0	0
	0,5971 1	0,3357 1	0,2907 1	0,2979 1	547,6 0	163,3 0,7	136,8 0,75	0,8636 1
A	0,25	0,45	0,7	0,7	0	0	0	0
	0,6016 1	0,4184 1	0,3061 1	0,3015 1	423,3 0	115,8 0,75	41,93 0,9	0,8636 1

Таблица 5.18

Работоспособность критериев, основанных на принципе варьирования модели, при уровне шума 0,7 (модель 2)

Тип	ЭВ-оценки				ММП-оценки			
	Reg	GSI(5)	GSI(10)	GSI(20)	Reg	GSI(5)	GSI(10)	GSI(20)
Q	0,6	0,75	0,55	0,75	0	0	0	0
	0,5584 0,9	0,5407 1	0,7564 1	0,5497 1	3,755 0,05	3,200 0,85	3,016 0,7	3,200 0,85
D	0,5	0,85	0,65	0,85	0,05	0	0	0
	0,3860 1	0,2388 1	0,3281 1	0,2478 1	3,246 0,2	3,383 0,6	3,049 0,35	2,846 0,55
Z	0,5	0,75	0,65	0,8	0,05	0	0,05	0
	0,3905 1	0,2834 1	0,3281 1	0,2701 1	2,931 0,3	3,380 0,55	3,160 0,4	2,840 0,55
A	0,45	0,45	0,45	0,6	0	0	0	0
	0,4083 1	0,4038 1	0,4083 1	0,3504 1	3,266 0,15	3,343 0,6	3,136 0,4	2,986 0,6

5.3.3. ВЫБОР СТРУКТУРЫ ПРИ ЗАВИСИМЫХ ОШИБКАХ

До сих пор рассматривался выбор структуры модели, в которой имелась лишь зависимость откликов, относящихся к одному наблюдению (в многооткликовом случае). Рассмотрим особенности выбора структуры при произвольной форме зависимости наблюдений.

В данном случае критерии, связанные с разбиением выборки, необходимо конструировать с учетом сохранения информации о зависимости. В простейшем случае выборка должна разбиваться так, чтобы зависимые элементы выборки находились в одной подвыборке. Если же такое разбиение невозможно или нецелесообразно, следует воспользоваться одним из двух подходов.

Во-первых, можно осуществить преобразование типа (1.7), приводящее к некоррелированным наблюдениям. Для гауссовской выборки наблюдения в модели (1.7) будут независимыми, а потому можно использовать обычные подходы к выбору структуры. Для негауссовской выборки наблюдения останутся зависимыми.

Во-вторых, можно воспользоваться принципом варьирования модели, при этом вспомогательная модель должна описывать *всю* выборку.

Например, в модели (1.1), (1.6) для определений критериев из табл. 5.1 (при $l=1$) $\hat{\Theta}_W$ обозначает оценку параметров Θ исходной модели, $\hat{\Theta}_A$ и $\hat{\Theta}_B$ – оценки параметров Θ_1 и Θ_2 модели (5.25). Качество вспомогательной модели со специальными ограничениями определяется с использованием значений функции качества оценок параметров Q или сверток оценок ковариационных матриц ошибок $\hat{V} = V(\hat{\phi})$. Заметим, что несимметричные критерии регулярности, непротиворечивости и вариативности применяются только при возможности разделения исходной выборки на две независимые группы наблюдений.

Обобщенный критерий скользящего контроля можно построить как стохастический вариант симметричного критерия регулярности.

Рассмотрим результаты исследования подходов к выбору структуры при зависимых ошибках.

Решалась традиционная задача выбора структуры традиционной многооткликовой модели в условиях наличия повторных наблюдений

и коррелированных гетероскедастичных нормальных ошибок. Ковариационная матрица ошибок имела мультипликативную структуру вида (3.2).

При исследовании выявлялась способность критериев находить истинную структуру (цель 1), корректные (цель 2) и удовлетворительно прогнозирующие (цель 3) структуры. Способность к прогнозированию определялась по теоретическому критерию

$$TSP = \mathbf{E} \left\{ \mathbf{E}Y - X\hat{\theta} \quad V^{-1} \quad \mathbf{E}Y - X\hat{\theta} \right\},$$

где $\hat{\theta}$ – ОМНК-оценка, получаемая при истинном значении ковариационной матрицы V .

Наборы данных содержали по шесть повторных измерений шести откликов в 16 точках факторного пространства. На отклики предположительно влияют три фактора, варьируемые на трех уровнях. В качестве истинной была выбрана модель

$$y_1 = 10 + 0x_1 + 1x_2 + 0,05x_3 + e_1, \quad y_2 = 12 + 0x_1 + 2x_2 + 0,05x_3 + e_2,$$

$$y_3 = 14 + 0x_1 + 1x_2 + 0,05x_3 + e_3, \quad y_4 = 12 + 0x_1 + 2x_2 + 0,005x_3 + e_4,$$

$$y_5 = 14 + 0x_1 + 1x_2 + 0,05x_3 + e_5, \quad y_6 = 10 + 0x_1 + 2x_2 + 0,02x_3 + e_6,$$

где x_1, x_2, x_3 – значения входных переменных, лежащие в области $[-1, 1]$.

Оценивание параметров осуществлялось по ММП (см. п. 3.1); критерии выбора структуры, использующие принцип варьирования модели, построены на основе функции правдоподобия. Аддитивная постоянная в процессе полного перебора структур всегда присутствовала в модели.

Кросс-ковариационные матрицы Σ , Ω и Λ определялись по заданным кросс-корреляционным матрицам и диагональным элементам. Диагональные элементы матриц были выбраны следующими: для матрицы Σ – все единичные, для матрицы Ω – первые два равны 0,01, следующие два равны 0,1, остальные единичные; для матрицы Λ – первые два равны 100, следующие четыре равны 10, остальные еди-

ничные. Во всех кросс-корреляционных матрицах задавалось значение внедиагональных элементов, равное 0,9. Скалярный параметр ковариационной матрицы имел значение $\delta \approx 0,00469076$.

Было получено 10 выборок при различных реализациях ошибок с описанными свойствами.

Решение задачи выбора структуры осуществлялось с использованием следующих подходов:

1) использование принципа варьирования модели (матричные ковариационные параметры во время выбора структуры не оценивались – использовались их оценки при полной модели, для каждой структуры оценивался скалярный ковариационный параметр);

2) использование принципа варьирования модели (во время выбора структуры матричные ковариационные параметры оценивались);

3) использование преобразования (1.7) исходной модели (для преобразования (1.7) и во время выбора структуры применялись оценки матричных ковариационных параметров при полной модели, для каждой структуры оценивался скалярный ковариационный параметр).

В табл. 5.19 представлены частоты достижения критериями целей 1 и 2 в последовательности экспериментов, а также средние значения критерия ТСР для структур, признанных оптимальными.

Таблица 5.19

Работоспособность критериев при различных подходах

Критерий	Подход 1			Подход 2			Подход 3		
	Цель 1	Цель 2	Цель 3	Цель 1	Цель 2	Цель 3	Цель 1	Цель 2	Цель 3
Reg	0,9	1	18,6	1	1	18	0,2	1	22,8
Stab	1	1	18	1	1	18	0,1	1	23,4
Stab(A)	1	1	18	1	1	18	0	1	24

Результаты показывают преимущество использования принципа варьирования модели перед подходом, связанным с преобразованием (1.7), а также нечувствительность подхода, связанного с варьированием модели, к использованию оценки ковариационной матрицы при полной модели. Заметим здесь же, что оценка ковариационной матрицы при полной модели, необходимая для преобразования (1.7), не все-

гда может быть вычислена (потенциальных регрессоров больше, чем наблюдений; мультиколлинеарность регрессоров).

Рассмотрим результаты сравнения обобщенного критерия скользящего контроля с другими критериями. Использовалась та же модель, но со значением внедиагональных элементов кросс-корреляционных матриц, равным 0,5. Из трех рассмотренных выше подходов к выбору структуры для данной части исследования был выбран подход 1. Показатели работоспособности критериев вычислялись по 20 испытаниям. При вычислении обобщенного критерия скользящего контроля случайное разбиение выборки осуществлялось пять раз. В табл. 5.20, аналогичной табл. 5.19, сравнивается работоспособность ряда критериев при различных значениях параметра δ . Результаты свидетельствуют о высокой работоспособности исследуемого критерия и о его преимуществе в ряде случаев перед остальными.

Т а б л и ц а 5.20

Работоспособность критериев при различном уровне шума

Критерий	Значение параметра δ								
	0,0046907578125			0,46907578125			1,876303125		
	Цель 1	Цель 2	Цель 3	Цель 1	Цель 2	Цель 3	Цель 1	Цель 2	Цель 3
Reg	0,85	18,9	1	0,5	29,9077	0,65	0,2	21,4215	0,25
Stab	0,85	18,9	1	0,75	23,3176	0,85	0,35	20,6173	0,4
Stab(A)	0,9	18,6	1	0,6	30,5803	0,6	0,2	20,9899	0,2
GSI	0,7	19,8	1	0,65	21,3725	0,95	0,5	20,8083	0,7

Рассмотрим работоспособность критериев при различных способах упрощения решения задачи выбора структуры.

Рассматривалась та же задача, что в начале данного пункта, однако модель отклика y_4 имела вид

$$y_4 = 12 + 0x_1 + 2x_2 + 0,05x_3 + e_4.$$

Задача выбора структуры решалась с использованием следующих подходов, основанных на принципе варьирования модели:

1) ковариационные параметры во время выбора структуры не оценивались – использовались их истинные значения;

2) ковариационные параметры во время выбора структуры не оценивались – использовались их оценки при полной модели;

3) во время выбора структуры оценивался только параметр Σ , для остальных использовались оценки при полной модели;

4) во время выбора структуры оценивались все ковариационные параметры;

5) во время выбора структуры оценивался только параметр Σ , остальные были единичными;

6) во время выбора структуры оценивался только параметр δ , остальные были единичными.

Заметим, что во всех случаях, за исключением подхода 6, параметр δ не оценивался ни предварительно, ни при выборе структуры.

В табл. 5.21 представлены частоты достижения критериями целей 1 и 2 в 10 испытаниях.

Таблица 5.21

Работоспособность критериев при различных способах упрощения решения

Критерий	Подходы 1, 3, 4 Цели 1, 2	Подход 2		Подход 5		Подход 6	
		Цель 1	Цель 2	Цель 1	Цель 2	Цель 1	Цель 2
Reg	1	0,9	1	0,4	0,4	0,1	0,1
Stab, Stab(A)	1	1	1	0,3	1	0,2	0,3

Анализ результатов, приведенных в табл. 5.21, позволяет сделать вывод о том, что отказ от учета значения ковариационной матрицы при выборе структуры модели снижает работоспособность процедур. В случае использования оценок ковариационных параметров при полной модели рассматриваемые критерии сохраняют работоспособность, что позволяет применять при выборе структуры с их использованием быстрые, хотя и грубые, методы оценивания.

5.3.4. АЛГОРИТМ ПОСЛЕДОВАТЕЛЬНОГО УТОЧНЕНИЯ ОПТИМАЛЬНОЙ СТРУКТУРЫ

Задачу выбора структуры можно решать путем перебора структур – полного или направленного. При использовании оценок параметров по методу, отличному от МНК и ОМНК, проведение даже направленного перебора может быть слишком трудоемким.

Для упрощения решения задачи выбора структуры рассмотрим *алгоритм последовательного уточнения оптимальной структуры*, основанный на возможности представления M -оценки в виде ОМНК-оценки с весами наблюдений специального вида. Подобный подход используется в [80] для построения устойчивой процедуры пошаговой регрессии.

Сначала осуществляется выбор структуры при использовании ОМНК с равными весами наблюдений и вычисляется M -оценка для оптимальной структуры, затем осуществляется выбор структуры при использовании ОМНК с фиксированными весами наблюдений, вычисленными после M -оценивания, и т. д. Процесс заканчивается, когда на очередном этапе в качестве оптимальной выбрана структура, уже рассматривавшаяся ранее.

Таким образом, метод сводится к последовательности этапов, на которых чередуются шаг выбора структуры с использованием ОМНК и шаг вычисления M -оценок. В результате M -оценки вычисляются только для предварительно отобранных структур-кандидатов.

На каждом этапе выбора структуры можно отбирать несколько оптимальных структур, для каждой из которых продолжать процесс уточнения структуры.

Во время перебора структур используются критерии качества, вычисляемые с использованием ОМНК, при этом выбор структуры может быть осуществлен с использованием эффективных вычислительных схем (см. п. 5.2). Для отобранных на каждом этапе структур вычисляются значения критерия качества, использующего M -оценки. На их основе определяется оптимальная модель.

Рассмотрим результаты исследования алгоритма последовательного уточнения оптимальной структуры.

Целью исследования была проверка работоспособности этого алгоритма при решении обобщенной задачи выбора структуры с исполь-

зованием устойчивых оценок параметров. Исследование проведено методом статистических испытаний.

Решалась обобщенная задача выбора структуры многооткликowej модели типа псевдонезависимых регрессий в условиях засоренного многомерного нормального распределения с ковариационной матрицей Σ . Использовались ЭВ-оценки Мешалкина с параметром $c = 0,25$.

Данные генерировались в соответствии с моделью

$$y_1 = 20 + 6x_1 + 4,6x_2 + 0x_3 + 0x_4 + e_1,$$

$$y_2 = 21 + 7x_1 + 0,5x_2 + 0x_3 + 0x_4 + e_2,$$

$$x_i \in [-1, 1].$$

Множество перебираемых ограничений-кандидатов имело вид

$$\theta_{kl} = 0, \quad k = 1, 2, \quad l = 1, \dots, 4, \quad \theta_{11} = \theta_{21}, \quad \theta_{12} = \theta_{22},$$

где θ_{kl} – параметр при регрессоре x_l в уравнении k -го отклика.

Ошибки наблюдений имели те же характеристики, что и ошибки при исследовании методов оценивания в п. 2.3.2.

Выбор структуры в рамках ОМНК осуществлялся путем полного перебора, использовался критерий регулярности, построенный по модели (5.48), (5.49), в виде Z-свертки оценки ковариационной матрицы Σ . После работы алгоритма последовательного уточнения оптимальной структуры из предварительно сформированного им множества моделей-кандидатов отбирались оптимальные модели при помощи различных сверток критерия этого же типа с использованием M -оценивания.

Алгоритм исследовался на способность выбирать структуры, удовлетворяющие определенным требованиям. Ставились цели нахождения истинной и корректной структур. Кроме того, ставились цели нахождения структур, способных давать удовлетворительный прогноз значений откликов в точках, составляющих набор данных (основная

выборка – W), или в новых точках (выборка C). Способность к прогнозированию определялась по критерию

$$CP(D) = \theta - \hat{\theta}^T \tilde{X}_D^T \left[\Sigma^{-1} \otimes I_N \right] \tilde{X}_D \theta - \hat{\theta} ,$$

где D – индекс выборки (W или C); \tilde{X}_D – матрица регрессоров (см. модель (1.35)) соответствующей выборки.

В табл. 5.22 приведены усредненные по 100 экспериментам значения критерия CP , вычисленного по точкам выборок W и C , частоты нахождения истинной и корректной структур, а также характеристики структуры, найденной на первом этапе алгоритма (путем однократного выбора структуры в рамках ОМНК с одинаковыми весами наблюдений), как до M -оценивания, так и после него.

Таблица 5.22

Качество отбираемых структур

Метод выбора	Критерий	Истинная	Корректная	$CP(W)$	$CP(C)$
Одноэтапный	С ОМНК-оценкой	0,23	0,69	25,6565	6,05431
	С M -оценкой			3,04247	0,724976
Многоэтапный	D-свертка	0,31	0,98	1,16403	0,271675
	Z-свертка	0,32	0,98	1,15633	0,270175
	A-свертка	0,3	0,98	1,15437	0,269673

Данные табл. 5.22, показывают что применение алгоритма последовательного уточнения оптимальной структуры имеет преимущество перед подходом, использующим один этап уточнения структуры.

Таблица 5.23

Работа алгоритма

Шаги	Структура	D-свертка	Z-свертка	A-свертка	$CP(W)$	$CP(C)$
1	$\theta_{23} \neq 0$ $\theta_{24} \neq 0$	0,31810	7,0408	2,2159	0,79547	0,18910
2	$\theta_{24} \neq 0$ $\theta_{12} = \theta_{22}$	32,398	225,09	47,050	1251,9	210,17
3	$\theta_{24} \neq 0$	0,22978	6,0417	1,9467	0,65895	0,15775
4	Истинная	0,20854	5,7783	1,8653	0,48979	0,10361
	Полная	0,37929	7,6771	2,3891	1,0473	0,29327

В табл. 5.23 отражена работа исследуемого алгоритма на одной реализации вектора ошибок, кроме того, там даны характеристики структуры максимальной сложности. Было проведено четыре этапа, на последнем из которых найдена истинная структура. Во втором столбце указаны отличия наилучшей структуры этапа от истинной, в следующих трех столбцах – значения сверток критерия регулярности, в последних двух – значения критериев CP .

ЗАКЛЮЧЕНИЕ

Постоянное повышение сложности моделируемых многофакторных объектов и увеличение производительности вычислительной техники приводят к тому, что разработка методов моделирования при нарушении классических предположений о свойствах наблюдений является одним из интенсивно развивающихся направлений прикладной математической статистики. В рамках этого направления выполнена и работа, результаты которой изложены в данной монографии.

Разработанные методы планирования эксперимента, оценивания параметров и выбора структуры позволяют ослабить требования, предъявляемые к данным (например, не требуется точное знание закона распределения наблюдений), и более полно учесть априорную информацию (например, неоднородность, зависимость моделируемых переменных, шкалу, в которой они измерены). Все это позволяет увеличивать точность и обоснованность моделей, снижать затраты на проведение экспериментов.

Рассмотренные методы могут использоваться в практике моделирования сложных объектов, процессов и явлений в различных областях науки и техники.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Айвазян С. А. Прикладная статистика: Основы моделирования и первичная обработка данных / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – М. : Финансы и статистика, 1983. – 471 с.
2. Айвазян С. А. Прикладная статистика: Исследование зависимостей / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – М. : Финансы и статистика, 1985. – 487 с.
3. Айвазян С. А. Прикладная статистика и основы эконометрики / С. А. Айвазян, В. С. Мхитарян. – М. : ЮНИТИ, 1998. – 1022 с.
4. Акаике Х. Развитие статистических методов / Х. Акаике // Современные методы идентификации / под ред. П. Эйкхоффа. – М. : Мир, 1983. – С. 148–176.
5. Алгоритмы и программы восстановления зависимостей / под ред. В. Н. Вапника. – М. : Наука, 1984. – 816 с.
6. Аренс Х. Многомерный дисперсионный анализ / Х. Аренс, Ю. Лейтер. – М. : Финансы и статистика, 1985. – 232 с.
7. Богданович В. А. Теория устойчивого обнаружения, различения и оценивания сигналов / В. А. Богданович, А. Г. Вострецов. – М. : Физматлит, 2004. – 320 с.
8. Болдин М. В. Знаковый статистический анализ линейных моделей / М. В. Болдин, Г. И. Симонова, Ю. Н. Тюрин. – М. : Наука. Физматлит, 1997. – 288 с.
9. Боровков А. А. Математическая статистика / А. А. Боровков. – Новосибирск : Наука ; Изд-во Ин-та математики, 1997. – 772 с.
10. Беспалова В. А. Пошаговая регрессия при многих откликах и ее применение к отбору динамических факторов / В. А. Беспалова, А. П. Коростылев, В. В. Федоров // Применение многомерного статистического анализа в экономике и оценке качества продукции : тез. докл. II Всесоюз. науч.-техн. конф., Тарту, 1981. – М. : ЦЭМИ АН СССР, 1981. – С. 196–198.
11. Важенин И. А. Исследование стратегии последовательного планирования эксперимента при оценивании регрессионной модели с разнотипными откликами / И. А. Важенин, Д. В. Лисицин // Информатика и проблемы телекоммуникаций : материалы рос. науч.-техн. конф., Новосибирск, 27–28 апр., 2006. – Новосибирск, 2006. – Т. 1. – С. 138–141.
12. Важенин И. А. Байесовское планирование эксперимента при оценивании регрессионной модели с разнотипными откликами / И. А. Важенин, Д. В. Лисицин // Информатика и проблемы телекоммуникаций : материалы рос. науч.-техн. конф., Новосибирск, 26–28 апр., 2007. – Новосибирск, 2007. – Т. 1. – С. 109–111.
13. Вапник В. Н. Восстановление зависимостей по эмпирическим данным / В. Н. Вапник. – М. : Наука, 1979. – 447 с.
14. Гантмахер Ф. Р. Теория матриц / Ф. Р. Гантмахер. – М. : Наука, 1988. – 552 с.

15. Горский В. Г. Планирование кинетических экспериментов / В. Г. Горский. – М. : Наука, 1984. – 241 с.
16. Губарев В. В. Вероятностные модели : справочник. В 2-х ч. / В. В. Губарев ; Новосиб. электротехн. ин-т. – Новосибирск, 1992.
17. Гусев В. А. Использование подвыборок и понятия устойчивости в задаче определения общего вида искомой зависимости / В. А. Гусев // Завод. лаб. – 1987. – Т. 53, № 1. – С. 48–53.
18. Демиденко Е. З. Линейная и нелинейная регрессии / Е. З. Демиденко. – М. : Финансы и статистика, 1981. – 302 с.
19. Денисов В. И. Математическое обеспечение системы ЭВМ–экспериментатор / В. И. Денисов. – М. : Наука, 1977. – 252 с.
20. Денисов В. И. О численном построении оптимальных планов / В. И. Денисов, В. Н. Лаптев ; Новосиб. электротехн. ин-т. – Новосибирск, 1973. – Деп. в ВИНТИ, № 5476–73.
21. Денисов В. И. Оптимальное группирование, оценка параметров и планирование регрессионных экспериментов / В. И. Денисов, Б. Ю. Лемешко, Е. Б. Цой ; Новосиб. гос. техн. ун-т. – Новосибирск, 1993. – 346 с.
22. Денисов В. И. Оценивание параметров регрессионной модели с эллиптическим распределением и мультипликативной ковариационной структурой ошибок / В. И. Денисов, Д. В. Лисицин // Сиб. журн. индустр. матем. – 2002. – Т. V, № 3(11). – С. 92–102.
23. Денисов В. И. О свойствах оценок параметров регрессионной модели с эллиптическим распределением и мультипликативной ковариационной структурой ошибок / В. И. Денисов, Д. В. Лисицин // Сиб. журн. индустр. матем. – 2003. – Т. VI, № 2(14). – С. 37–45.
24. Денисов В. И. Планирование эксперимента с учетом появления пропусков в данных / В. И. Денисов, Д. В. Лисицин // Науч. вестн. НГТУ. – Новосибирск, 2004. – № 1(16). – С. 53–61.
25. Денисов В. И. Планирование эксперимента при оценивании параметров многофакторной модели по неоднородным наблюдениям / В. И. Денисов, Д. В. Лисицин, К. В. Гаврилов // Сиб. журн. индустр. матем. – 2002. – Т. V, № 4(12). – С. 14–28.
26. Денисов В. И., Условия оптимальности планов экспериментов для функционалов типа дискретного минимакса / В. И. Денисов, А. А. Попов ; Новосиб. электротехн. ин-т. – Новосибирск, 1978. – Деп. в ВИНТИ, № 3731–78. – 21 с.
27. Денисов В. И. Алгоритмы построения точных оптимальных планов регрессионных экспериментов / В. И. Денисов, А. А. Попов ; Новосиб. электротехн. ин-т. – Новосибирск, 1979. – Деп. в ВИНТИ, № 560–79.
28. Денисов В. И. Пакет программ оптимального планирования эксперимента / В. И. Денисов, А. А. Попов. – М. : Финансы и статистика, 1986. – 159 с.

29. Денисов В. И. Знаковый метод: преимущества, проблемы, алгоритмы / В. И. Денисов, В. С. Тимофеев // Науч. вестн. НГТУ. – Новосибирск, 2001. – № 1(10). – С. 21–35.
30. Дженнрич Р. И. Пошаговая регрессия / Р. И. Дженнрич // Статистические методы для ЭВМ / под ред. : К. Энслейна, Э. Рэлстона, Г. С. Уилфа. – М. : Наука, 1986. – С. 77–94.
31. Джонстон Дж. Эконометрические методы / Дж. Джонстон. – М. : Статистика, 1980. – 450 с.
32. Дисперсионный анализ и синтез планов на ЭВМ / Е. В. Маркова, В. И. Денисов, И. А. Полетаева, В. В. Пономарев. – М. : Наука, 1982. – 196 с.
33. Дрейпер Н. Прикладной регрессионный анализ. В 2 кн. / Н. Дрейпер, Г. Смит. – М.: Финансы и статистика, 1987. – Кн. 2. – 351 с.
34. Ермаков С. М. Математическая теория оптимального эксперимента / С. М. Ермаков, А. А. Жиглявский. – М. : Наука, 1987. – 320 с.
35. Ершов А. А. Стабильные методы оценки параметров / А. А. Ершов // Автоматика и телемеханика. – 1978. – № 8. – С. 66–100.
36. Загоруйко Н. Г. Прикладные методы анализа данных и знаний / Н. Г. Загоруйко. – Новосибирск : Изд-во Ин-та математики, 1999. – 270 с.
37. Зельнер А. Байесовские методы в эконометрии / А. Зельнер. – М. : Статистика, 1980. – 439 с.
38. Ивахненко А. Г. Задача регуляризации и унимодальности критерия непротиворечивости и ее решение в алгоритмах объективного системного анализа и объективной компьютерной кластеризации / А. Г. Ивахненко // Автоматика. – 1988. – № 3. – С. 12–17.
39. Ивахненко А. Г. Помехоустойчивость моделирования / А. Г. Ивахненко, В. С. Степашко. – Киев : Наукова думка, 1985. – 216 с.
40. Ивахненко А. Г. Моделирование сложных систем по экспериментальным данным / А. Г. Ивахненко, Ю. П. Юрачковский. – М. : Радио и связь, 1987. – 120 с.
41. Каминскас В. А. Идентификация динамических систем по дискретным наблюдениям / В. А. Каминскас. – Вильнюс : Моклас, 1982. – Ч. 1. – 245 с.
42. Кендалл М. Дж. Статистические выводы и связи / М. Дж. Кендалл, А. Стьюарт. – М. : Наука, 1973. – 900 с.
43. Кильдишев Г. С. Многомерные группировки / Г. С. Кильдишев, Ю. И. Аболенцев. – М. : Статистика, 1978. – 160 с.
44. Клейнер Г. Б. Эконометрические зависимости / Г. Б. Клейнер, С. А. Смоляк. – М. : Наука, 2003. – 104 с.
45. Кокс Д. Теоретическая статистика / Д. Кокс, Д. Хинкли. – М. : Мир, 1978. – 560 с.
46. Котоков В. И. Многофакторные кусочно-линейные модели / В. И. Котоков. – М. : Финансы и статистика, 1984. – 216 с.
47. Лбов Г. С. Методы обработки разнотипных экспериментальных данных / Г. С. Лбов. – М. : Наука, 1981. – 160 с.

48. Лбов Г. С. Логические решающие функции и вопросы статистической устойчивости решений / Г. С. Лбов, Н. Г. Старцева. – Новосибирск : Изд-во Ин-та математики, 1999. – 212 с.

49. Леман Э. Теория точечного оценивания / Э. Леман. – М. : Наука, 1991. – 448 с.

50. Лемешко Б. Ю. К вопросу о робастности оценок по группированным данным / Б. Ю. Лемешко, С. Н. Постовалов // Сб. науч. тр. НГТУ. – Новосибирск : Изд-во НГТУ. – 1996. – № 2(4). – С. 9–18.

51. Лемешко Б. Ю. Группирование наблюдений как способ получения робастных оценок / Б. Ю. Лемешко // Надежность и контроль качества. – 1997. – № 5. – С. 26–35.

52. Лемешко Б. Ю. Робастные методы оценивания и отбраковка аномальных измерений / Б. Ю. Лемешко // Завод. лаб. – 1997. – Т. 63, № 5. – С. 43–49.

53. Лисицин Д. В. Шаговая процедура выбора многооткликowej модели, оперирующая линейными гипотезами и ограничениями произвольного вида / Д. В. Лисицин // Сб. науч. тр. НГТУ. – Новосибирск : Изд-во НГТУ, 1995. – Вып. 2. – С. 31–34.

54. Лисицин Д. В. Алгоритмы выбора структуры для многомерных моделей / Д. В. Лисицин // Сб. науч. тр. НГТУ. – Новосибирск : Изд-во НГТУ, 1997. – Вып. 2(7). – С. 33–38.

55. Лисицин Д. В. Выбор структуры многомерной линейной модели при построении зависимостей по статистическим данным : дис. ... канд. техн. наук / Д. В. Лисицин. – Новосибирск, 1998. – 272 с.

56. Лисицин Д. В. Обобщенная задача выбора структуры многомерной модели / Д. В. Лисицин // АПЭП-98. Актуальные проблемы электронного приборостроения : тр. 4 междунар. конф. – Новосибирск, 1998. – Т. 3. – С. 69–72.

57. Лисицин Д. В. Оценивание параметров многооткликowej регрессии при мультипликативной параметризации ковариационной матрицы ошибок / Д. В. Лисицин // Науч. вестн. НГТУ. – Новосибирск, 2000. – № 1(8). – С. 12–22.

58. Лисицин Д. В. Адаптивные методы построения многооткликowej регрессионных моделей / Д. В. Лисицин // АПЭП-2000. Актуальные проблемы электронного приборостроения : тр. 5 междунар. конф. – Новосибирск, 2000. – Т. 3. – С. 14–18.

59. Лисицин Д. В. Устойчивые методы оценивания параметров и выбора структуры многооткликowej линейно параметризованной модели / Д. В. Лисицин // Науч. вестн. НГТУ. – Новосибирск, 2001. – № 2(11). – С. 53–66.

60. Лисицин Д. В. Регрессионная модель с эллиптическим распределением и мультипликативной ковариационной структурой ошибок / Д. В. Лисицин // Науч. вестн. НГТУ. – Новосибирск, 2001. – № 2(11). – С. 67–76.

61. Лисицин Д. В. EM-алгоритмы оценивания регрессионной модели с мультипликативной ковариационной структурой ошибок / Д. В. Лисицин // Автoметрия. – 2004. – № 1. – С. 15–26.

62. Лисицин Д. В. О критериях выбора структуры многооткликowej регрессионной модели / Д. В. Лисицин // Сиб. журн. индустр. матем. – 2004. – Т. VII (17). – С. 61–72.
63. Лисицин Д. В. Планирование эксперимента при робастном оценивании параметров регрессионной модели по неоднородным наблюдениям / Д. В. Лисицин // Сиб. журн. индустр. матем. – 2004. – Т. VII, № 4(20). – С. 92–106.
64. Лисицин Д. В. Конструирование робастных оценок параметров регрессии при неоднородных наблюдениях / Д. В. Лисицин // Науч. вестн. НГТУ. – Новосибирск, 2004. – № 3(18). – С. 43–55.
65. Лисицин Д. В. Оценивание параметров многофакторной модели при наличии разнотипных откликов / Д. В. Лисицин // Науч. вестн. НГТУ. – Новосибирск, 2005. – № 1(19). – С. 11–20.
66. Лисицин Д. В. Планирование эксперимента при оценивании параметров регрессионной модели с разнотипными откликами / Д. В. Лисицин // Науч. вестн. НГТУ. – Новосибирск, 2005. – № 3(21). – С. 53–66.
67. Лисицин Д. В. Выбор структуры многооткликowej регрессионной модели / Д. В. Лисицин // Науч. вестн. НГТУ. – Новосибирск, 2006. – № 1(22). – С. 17–32.
68. Лисицин Д. В. Об устойчивом оценивании параметров модели по неоднородным наблюдениям / Д. В. Лисицин // Науч. вестн. НГТУ. – Новосибирск, 2006. – № 2(23). – С. 35–42.
69. Лисицин Д. В. Критерии выбора структуры регрессионной модели при негауссовских и зависимых ошибках / Д. В. Лисицин // Сиб. журн. индустр. матем. – 2006. – Т. IX, № 2(26). – С. 90–106.
70. Лисицин Д. В. О локально устойчивом оценивании параметров распределений / Д. В. Лисицин, К. В. Гаврилов // Сб. науч. тр. НГТУ. – Новосибирск : Изд-во НГТУ, 2004. – Вып. 2(36). – С. 37–46.
71. Лисицин Д. В. Конструирование критериев селекции многомерных регрессионных моделей / Д. В. Лисицин, А. А. Попов // Сб. науч. тр. НГТУ. – Новосибирск : Изд-во НГТУ, 1996. – Вып. 1. – С. 13–20.
72. Лисицин Д. В. Исследование критериев селекции многооткликowych регрессионных моделей / Д. В. Лисицин, А. А. Попов // Сб. науч. тр. НГТУ. – Новосибирск : Изд-во НГТУ, 1996. – Вып. 2(4). – С. 19–28.
73. Лисицин Д. В. Исследование критериев селекции многомерных моделей при наличии разнотипных факторов / Д. В. Лисицин, А. А. Попов // АПЭП-96. Актуальные проблемы электронного приборостроения : тр. 3 междунар. науч.-техн. конф. – Новосибирск, 1996. – Т. 6, Ч. 1. – С. 54–58.
74. Лисицин Д. В. Исследование работоспособности критериев выбора многомерных моделей / Д. В. Лисицин, А. А. Попов // Информатика и проблемы телекоммуникаций : материалы междунар. науч.-техн. конф., Новосибирск, 24–25 апр., 1997. – Новосибирск, 1997. – С. 103–105.

75. Лисицин Д. В. Исследование адаптивных робастных оценок параметров регрессии при неоднородных данных / Д. В. Лисицин, Д. С. Форманчук // Информатика и проблемы телекоммуникаций : материалы рос. науч.-техн. конф., Новосибирск, 21–22 апр., 2005. – Новосибирск, 2005. – Т. 1. – С. 134–137.
76. Лисицин Д. В. Исследование робастных и адаптивных робастных оценок параметров регрессии при неоднородных данных / Д. В. Лисицин, Д. С. Форманчук // Науч. вестн. НГТУ. – Новосибирск, 2007. – № 1(26). – С. 23–30.
77. Лисицин Д. В. Робастное оценивание параметров гетероскедастичной регрессии / Д. В. Лисицин, Д. С. Форманчук // Докл. АН ВШ РФ. – 2007. – № 2(9). – С. 17–23.
78. Литтл Р. Дж. А. Статистический анализ данных с пропусками / Р. Дж. А. Литтл, Д. Б. Рубин. – М. : Финансы и статистика, 1990. – 336 с.
79. Маркова Е. В. Планирование эксперимента в условиях неоднородностей / Е. В. Маркова, А. Н. Лисенков. – М.: Наука, 1973. – 219 с.
80. Мостеллер Ф. Анализ данных и регрессия / Ф. Мостеллер, Дж. Тьюки. – М. : Финансы и статистика, 1982. – Вып. 2. – 239 с.
81. Мудров В. И. Методы обработки измерений: Квазиправдоподобные оценки / В. И. Мудров, В. Л. Кушко. – М. : Радио и связь, 1983. – 304 с.
82. Налимов В. В. Логические основания планирования эксперимента / В. В. Налимов, Т. И. Голикова. – М. : Металлургия, 1981. – 152 с.
83. Никифоров А. М. Статистический анализ наблюдений со случайными пропусками / А. М. Никифоров // Пятая международная Вильнюсская конференция по теории вероятностей и математической статистике : тез. докл. – Вильнюс, 1989. – С. 98–99.
84. Новицкий П. В. Оценка погрешностей результатов измерений / П. В. Новицкий, И. А. Зограф. – Л. : Энергоатомиздат, 1991. – 304 с.
85. Перельман И. И. Методология выбора структуры модели при идентификации объектов управления / И. И. Перельман // Автоматика и телемеханика. – 1983. – № 11. – С. 5–29.
86. Планирование эксперимента в задачах нелинейного оценивания и распознавания образов / Г. К. Круг, В. А. Кабанов, Г. А. Фомин, Е. С. Фомина. – М. : Наука, 1981. – 172 с.
87. Полотнов М. М. Разработка и исследование методов построения регрессионных моделей объектов управления при воздействии неконтролируемого дискретного фактора неоднородности : автореф. дис. ... канд. техн. наук / М. М. Полотнов. – М., 1986. – 20 с.
88. Полотнов М. М. Оценивание параметров модели объекта по неоднородным экспериментальным данным / М. М. Полотнов, Г. А. Фомин // Завод. лаб. – 1986. – № 7. – С. 37–43.

89. Пономарев В. В. Диалоговая система обработки данных многооткликовых экспериментов с качественными и количественными факторами : дис. ... канд. техн. наук / В. В. Пономарев. – Новосибирск, 1985. – 193 с.
90. Попов А. А. Дисперсионный анализ моделей с качественными факторами как задача структурного моделирования / А. А. Попов // Машинные методы оптимизации, моделирования и планирования эксперимента. – Новосибирск : Новосиб. электротехн. ин-т, 1988. – С. 130–133.
91. Попов А. А. Конструирование дискретных и непрерывно-дискретных моделей регрессионного типа / А. А. Попов // Сб. науч. тр. НГТУ. – Новосибирск : Изд-во НГТУ, 1996. – Вып. 1. – С. 21–30.
92. Попов А. А. Композиционный подход построения адекватных регрессионных моделей в схемах активного эксперимента / А. А. Попов // Сб. науч. тр. НГТУ. – Новосибирск : Изд-во НГТУ, 1996. – Вып. 2(4). – С. 29–38.
93. Попов А. А. Оптимальное планирование эксперимента в задачах структурной и параметрической идентификации моделей многофакторных систем : дис. ... д-ра техн. наук / А. А. Попов. – Новосибирск, 1997. – 424 с.
94. Прикладная статистика: Классификация и снижение размерности / С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин. – М. : Финансы и статистика, 1989. – 607 с.
95. Пуарье Д. Эконометрия структурных изменений / Д. Пуарье. – М. : Финансы и статистика, 1981. – 183 с.
96. Рао С. Р. Линейные статистические методы и их применения / С. Р. Рао. – М. : Наука, 1968. – 548 с.
97. Редько М. Ю. Квазиправдоподобные оценки для линейной регрессии / М. Ю. Редько. – Новосибирск : Новосиб. электротехн. ин-т, 1988. – 31 с. – Деп. в ВИНТИ, № 4821–В88.
98. Редько М. Ю. Реализация алгоритма пошаговой регрессии для многооткликовой модели / М. Ю. Редько // Электронная техника. Сер. 7: Технология, организация производства и оборудование. – 1993. – Вып. 2(177) – 3(178). – С. 41–44.
99. Робастность в статистике. Подход на основе функций влияния / Ф. Хампель, Э. Рончетти, П. Рассеу, В. Штаэль. – М. : Мир, 1989. – 512 с.
100. Розин Б. Б. Теория распознавания образов в экономических исследованиях / Б. Б. Розин. – М. : Статистика, 1973. – 224 с.
101. Розин Б. Б. Экономико-статистические модели с переменной структурой / Б. Б. Розин, В. И. Котюков, М. А. Ягольницер. – Новосибирск : Наука, 1984. – 242 с.
102. Романов В. Л. Выбор наилучшей линейной регрессии: сравнение формальных критериев (Обобщающая статья) / В. Л. Романов // Завод. лаб. – 1990. – Т. 56, № 1. – С. 90–95.
103. Савараги Е. «Классические» методы и оценивание временных рядов / Е. Савараги, Т. Созда, Т. Накамизо // Современные методы идентификации / под ред. П. Эйкхоффа. – М. : Мир, 1983. – С. 74–147.

104. Салуквадзе М. Е. Задачи векторной оптимизации в теории управления / М. Е. Салуквадзе. – Тбилиси : Мецниереба, 1975. – 201 с.
105. Сарычев А. П. Усредненный критерий регулярности метода группового учета аргументов в задаче поиска наилучшей регрессии / А. П. Сарычев // Автоматика. – 1990. – № 5. – С. 28–33.
106. Себер Дж. Линейный регрессионный анализ / Дж. Себер. – М. : Мир, 1980. – 456 с.
107. Смоляк С. А. Устойчивые методы оценивания: (Статистическая обработка неоднородных совокупностей) / С. А. Смоляк, Б. П. Титаренко. – М. : Статистика, 1980. – 208 с.
108. Спектор А. А. Гауссовские дискретные поля с разделимыми спектрально-корреляционными характеристиками / А. А. Спектор // Прикладная теория случайных процессов и полей / под ред. К. К. Васильева, В. А. Омельченко. – Ульяновск : УлГТУ, 1995. – С. 143–164.
109. Степашко В. С. Методы и критерии решения задач структурной идентификации / В. С. Степашко, Ю. Л. Кочерга // Автоматика. – 1985. – № 5. – С. 29–37.
110. Стогов Г. В. Устойчивые методы обработки результатов измерений / Г. В. Стогов, А. В. Макшанов, А. А. Мусаев // Зарубежная радиоэлектроника. – 1982. – № 9. – С. 3–46.
111. Ступина Т. А. Построение логико-вероятностной модели прогнозирования системы разнотипных переменных : автореф. дис. ... канд. физ.-мат. наук / Т. А. Ступина. – Новосибирск, 2006. – 17 с.
112. Федоров В. В. Свойства и методы построения планов, минимизирующих выпуклые функционалы от информационной матрицы / В. В. Федоров. – М. : Изд-во Москов. ун-та, 1970. – 11 с.
113. Федоров В. В. Теория оптимального планирования эксперимента (планирование регрессионных экспериментов) / В. В. Федоров. – М. : Наука, 1971. – 312 с.
114. Федоров В. В. Оценивание параметров регрессии в случае векторного наблюдения / В. В. Федоров // Регрессионные эксперименты (Планирование и анализ) / под ред. В. В. Налимова. – М. : Изд-во Москов. ун-та, 1977. – С. 112–122.
115. Федоров В. В. Активные регрессионные эксперименты / В. В. Федоров // Математические методы планирования эксперимента. – Новосибирск : Наука, 1981. – С. 19–73.
116. Финн Дж. Д. Многомерный дисперсионный и ковариационный анализ / Дж. Д. Финн // Статистические методы для ЭВМ / под ред. К. Энслейна, Э. Рэлстона, Г. С. Уилфа. – М. : Наука, 1986. – С. 219–268.
117. Хеттманспергер Т. П. Статистические выводы, основанные на рангах / Т. П. Хеттманспергер. – М. : Финансы и статистика, 1987. – 333 с.

118. Хогг Р. В. Введение в помехоустойчивое оценивание / Р. В. Хогг // Устойчивые статистические методы оценки данных / под ред. Р. Л. Лонера, Г. Н. Уилкинсона. – М. : Машиностроение, 1984. – С. 12–26.
119. Хокинг Р. Р. Выбор наилучшего подмножества регрессионных переменных / Р. Р. Хокинг // Статистические методы для ЭВМ / под ред. К. Энслейна, Э. Рэлстона, Г. С. Уилфа. – М. : Наука, 1986. – С. 53–77.
120. Хьюбер П. Робастность в статистике / П. Хьюбер. – М. : Мир, 1984. – 303 с.
121. Цыпкин Я. З. Основы информационной теории идентификации / Я. З. Цыпкин. – М. : Наука, 1984. – 320 с.
122. Шеффе Г. Дисперсионный анализ / Г. Шеффе. – М. : Наука, 1980. – 512 с.
123. Шуленин В. П. Введение в робастную статистику / В. П. Шуленин. – Томск : Изд-во Том. ун-та, 1993. – 227 с.
124. Шурыгин А. М. Оценки параметров нормального распределения с экспоненциальным взвешиванием наблюдений: асимптотическая теория / А. М. Шурыгин // Алгоритмическое и программное обеспечение прикладного статистического анализа. – М. : Наука, 1980. – С. 241–259.
125. Шурыгин А. М. Прикладная стохастика: робастность, оценивание, прогноз / А. М. Шурыгин. – М. : Финансы и статистика, 2000. – 224 с.
126. Шурыгин А. М. Асимптотическая теория устойчивого оценивания : автореф. дис. ... д-ра техн. наук / А. М. Шурыгин. – М., 2002. – 28 с.
127. Яковлев А. А. Алгоритм выбора субоптимальной совокупности предикторов для множественной многомерной регрессии / А. А. Яковлев, Н. А. Ставицкая // Вопросы кибернетики. – М. : ВИНТИ, 1981. – Вып. 73. – С. 110–118.
128. Allen D. M. The relationship between variable selection and data augmentation and a method for prediction / D. M. Allen // *Technometrics*. – 1974. – Vol. 16. – P. 125–127.
129. Atkinson A. C. *D*-optimum designs for heteroscedastic linear models / A. C. Atkinson, R. D. Cook // *J. Amer. Statist. Assoc.* – 1995. – Vol. 90. – P. 204–212.
130. Biernacki C. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models / C. Biernacki, G. Celeux, G. Govaert // *Computational Statistics & Data Analysis*. – 2003. – Vol. 41. – P. 561–575.
131. Cantoni E. Variable selection for marginal longitudinal generalized linear models / E. Cantoni, J. Mills Flemming, E. Ronchetti // *Biometrics*. – 2005. – Vol 61. – P. 507–514.
132. Cavanaugh J. E. A large-sample model selection criterion based on Kullback's symmetric divergence / J. E. Cavanaugh // *Statistics and Probability Letters*. – 1999. – Vol. 44. – P. 333–344.

133. Dempster A. P. Maximum likelihood estimation from incomplete data via the EM algorithm / A. P. Dempster, N. M. Laird, D. B. Rubin // *J. of the Royal Statistical Society, ser. B.* – 1977. – Vol. 39. – P. 1–38.
134. Dempster A. P. Iteratively reweighted least squares for linear regression when errors are normal/independent distributed / A. P. Dempster, N. M. Laird, D. B. Rubin // *Multivariate analysis V* / P. R. Krishnaiah eds. – North Holland, Amsterdam, 1980. – P. 35–57.
135. Dette H. Bayesian and maximin optimal designs for heteroscedastic regression models / H. Dette, L. M. Haines, L. A. Imhof // *The Canad. J. of Statistics.* – 2005. – Vol. 33. – P. 221–241.
136. Dette H. Optimal Bayesian designs for models with partially specified heteroscedastic structure / H. Dette, W. K. Wong // *The Annals of Statistics.* – 1996. – Vol. 24. – P. 2108–2127.
137. Dette H. Bayesian D-optimal designs on a fixed number of design points for heteroscedastic polynomial models / H. Dette, W. K. Wong // *Biometrika.* – 1998. – Vol. 85. – P. 869–882.
138. Dragalin V. Adaptive designs for dose-finding based on efficacy–toxicity response / V. Dragalin, V. Fedorov // *J. of Statistical Planning and Inference.* – 2006. – Vol. 136. – P. 1800–1823.
139. Efron B. Multiple regression analysis / M. A. Efron // *Mathematical methods for digital computers* / A. Ralston, H. S. Wilf eds. – New York : John Wiley, 1960. – P. 191–203.
140. Efron B. Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information / B. Efron, D. V. Hinkley // *Biometrika.* – 1978. – Vol. 65. – P. 457–487.
141. Fedorov V. V. Design of experiments with unknown parameters in variance / V. V. Fedorov, R. C. Gagnon, S. L. Leonov // *Appl. Stochastic Models Bus. Ind.* – 2002. – Vol. 18. – P. 207–218.
142. Gupta V. K. Robustness of designed experiments against missing data / V. K. Gupta, L. Bhar, K. Lal // *J. of Applied Statistics.* – 2001. – Vol. 28. – P. 63–79.
143. Hannan E. J. The determination of the order of an autoregression / E. J. Hannan, B. G. Quinn // *J. of the Royal Statistical Society, ser. B.* – 1979. – Vol. 41. – P. 190–195.
144. Heise M. A. Optimal designs for bivariate logistic regression / M. A. Heise, R. H. Myers // *Biometrics.* – 1996. – Vol. 52. – P. 613–624.
145. Hogg R. V. Adaptive robust procedures: A partial review and some suggestions for future applications and theory / R. V. Hogg // *J. Amer. Statist. Assoc.* – 1974. – Vol. 69. – P. 909–923.
146. Hosmer D. W. *Applied Logistic Regression* / D. W. Hosmer, S. Lemeshow. – New York : John Wiley, 2000. – 375 c.
147. Huber P. J. Minimax aspects of bounded-influence regression / P. J. Huber // *J. Amer. Statist. Assoc.* – 1983. – Vol. 78. – P. 66–80.

148. Adaptive mixtures of local experts / R. A. Jacobs, M. I. Jordan, S. J. Nowlan, G. E. Hinton // *Neural Computation*. – 1991. – Vol. 3. – P. 79–87.
149. Jamshidian M. A note on parameter and standard error estimation in adaptive robust regression / M. Jamshidian // *J. Statist. Comput. Simul.* – 2001. – Vol. 71. – P. 11–27.
150. Jennrich R. I. Unbalanced repeated-measures models with structured covariance matrices / R. I. Jennrich, M. D. Schluchter // *Biometrics*. – 1986. – Vol. 42. – P. 805–820.
151. Karlis D. Choosing initial values for the EM algorithm for finite mixtures / D. Karlis, E. Xekalaki // *Computational Statistics & Data Analysis*. – 2003. – Vol. 41. – P. 577–590.
152. Kassam S. A. Asymptotically robust detection of a known signal in contaminated non-gaussian noise / S. A. Kassam, J. B. Thomas // *IEEE Trans. Information Theory*. – 1976. – Vol. 22. – P. 22–26.
153. Kent J. T. Redescending M -estimates of multivariate location and scatter / J. T. Kent, D. E. Tyler // *Annals of Statistics*. – 1991. – Vol. 19. – P. 2102–2119.
154. Kent J. T. Constrained M -estimation of multivariate location and scatter / J. T. Kent, D. E. Tyler // *Annals of Statistics*. – 1996. – Vol. 24. – P. 1346–1370.
155. Kiefer N. M. Discrete parameter variation: efficient estimation of a switching regression model / N. M. Kiefer // *Econometrica*. – 1978. – Vol. 46. – P. 427–434.
156. Lange K. L. Robust statistical modeling using the t distribution / K. L. Lange, R. J. A. Little, J. M. G. Taylor // *J. Amer. Statist. Assoc.* – 1989. – Vol. 84. – P. 881–896.
157. Little R. J. A. Maximum likelihood estimation for mixed continuous and categorical data with missing values / R. J. A. Little, M. D. Schluchter // *Biometrika*. – 1985. – Vol. 72. – P. 497–512.
158. Liu C. ML estimation of the multivariate t -distribution and the EM algorithms / C. Liu // *J. of Multivariate Analysis*. – 1997. – Vol. 63. – P. 296–312.
159. Liu C. The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence / C. Liu, D. B. Rubin // *Biometrika*. – 1994. – Vol. 81. – P. 633–648.
160. Liu C. ML estimation of the t distribution using EM and its extensions, ECM and ECME / C. Liu, D. B. Rubin // *Statistica Sinica*. – 1995. – Vol. 5. – P. 19–39.
161. Liu C. Ellipsoidally symmetric extensions of the general location model for mixed categorical and continuous data / C. Liu, D. B. Rubin // *Biometrika*. – 1998. – Vol. 85. – P. 673–688.
162. Liu C. Parameter expansion to accelerate EM: the PX-EM algorithm / C. Liu, D. B. Rubin, Y. Wu // *Biometrika*. – 1998. – Vol. 85. – P. 755–770.
163. Lucas A. Robustness of the Student t based M -estimator / A. Lucas // *Commun. Statist. Theory Meth.* – 1997. – Vol. 26. – P. 1165–1182.

164. Mallows C. L. Some comments on C_p / C. L. Mallows // *Technometrics*. – 1973. – Vol. 15. – P. 661–675.
165. Mansson R. A. Missing values in replicated Latin squares / R. A. Mansson, P. Prescott // *J. of Applied Statistics*. – 2001. – Vol. 28. – P. 743–757.
166. Mardia K. V. Spatial-temporal analysis of multivariate environmental monitoring data / K. V. Mardia, C. Goodall // *Multivariate Environmental Statistics* / G. P. Patil, C. R. Rao eds. – N. York : Elsevier ; Amsterdam : North-Holland, 1993. – Vol. 6. – P. 347–385.
167. Maronna R. A. Robust M -estimators of multivariate location and scatter / R. A. Maronna // *Annals of statistics*. – 1976. – Vol. 4. – P. 51–67.
168. Meng X.-L. Maximum likelihood estimation via the ECM algorithm: A general framework / X.-L. Meng, D. B. Rubin // *Biometrika*. – 1993. – Vol. 80. – P. 267–278.
169. McLachlan G. J. An algorithm for unsupervised learning via normal mixture models / G. J. McLachlan, D. Peel // *ISIS: Information, Statistics and Induction in Science* / D. L. Dowe, K. B. Korb, J. J. Oliver eds. – Singapore : World Scientific Publishing, 1996. – P. 354–363.
170. Mineo A. M. A software tool for the exponential power distribution: The normal package / A. M. Mineo, M. Ruggieri // *J. of Statistical Software*. – 2005. – Vol. 12, iss. 4. – 24 p.
171. Mitchell T. J. An algorithm for construction of D -optimal experimental designs / T. J. Mitchell // *Technometrics*. – 1974. – Vol. 16. – P. 203–210.
172. Moberg T. F. An adaptive multiple regression procedure based on M -estimators / T. F. Moberg, J. S. Ramberg, R. H. Randles // *Technometrics*. – 1980. – Vol. 22. – P. 213–224.
173. Müller Ch. H. High breakdown point designs / Ch. H. Müller // *Robust statistics, data analysis, and computer intensive methods* / H. Rieder eds. – *Lecture Notes in Statistics*. – Vol. 109. – N. York : Springer, 1996. – P. 353–360.
174. Müller Ch. H. Optimal breakdown point maximizing designs / Ch. H. Müller // *Tatra Mountains Math. Publ.* – 1996. – Vol. 7. – P. 79–85.
175. Naik D. N. Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix / D. N. Naik, S. S. Rao // *J. of Applied Statistics*. – 2001. – Vol. 28. – P. 91–105.
176. Olkin I. Multivariate correlation models with discrete and continuous variables / I. Olkin, R. F. Tate. – *Ann. Math. Statist.* – 1961. – Vol. 32. – P. 448–465.
177. Pinheiro J. C. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t -distribution / J. C. Pinheiro, C. Liu, Y. N. Wu // *J. of Computational and Graphical Statistics*. – 2001. – Vol. 10. – P. 249–276.
178. Prescott P. Robustness of balanced incomplete block designs to randomly missing observations / P. Prescott, R. A. Mansson // *J. of Statistical Planning and Inference*. – 2001. – Vol. 92. – P. 283–296.

179. Rencher A. C. *Methods of multivariate analysis* / A. C. Rencher. – N. York : John Wiley, 2002. – 708 p.
180. Rissanen J. Minmax entropy estimation of models for vector processes / J. Rissanen // *System identification: Advances and case studies* / R. K. Mehra, D. S. Lainiotis eds. – N. York : Academic Press, 1976. – P. 97–119.
181. Ronchetti E. Robustness aspects of model choice / E. Ronchetti // *Statistica Sinica*. – 1997. – Vol. 7. – P. 327–338.
182. Ronchetti E. Robust linear model selection by cross-validation / E. Ronchetti, C. A. Field, W. Blanchard // *J. Amer. Statist. Assoc.* – 1997. – Vol. 92. – P. 1017–1023.
183. Ronchetti E. A robust version of Mallow's C_p / E. Ronchetti, R. G. Staudte // *J. Amer. Statist. Assoc.* – 1994. – Vol. 89. – P. 550–559.
184. Rothman D. Letter to editor / D. Rothman // *Technometrics*. – 1968. – Vol. 10. – P. 432.
185. SAS/STAT. User's guide release 6.03 edition. – Cary : SAS Inst. Inc., 1988. – 1028 p.
186. Schwarz G. Estimating the dimension of a model / G. Schwarz // *Annals of Statistics*. – 1978. – Vol. 6. – P. 461–464.
187. Searle S. R. *Linear models* / S. R. Searle. – N. York : John Wiley, 1971. – 532 p.
188. Shao J. Linear model selection by cross-validation / J. Shao // *J. Amer. Statist. Assoc.* – 1993. – Vol. 88. – P. 486–494.
189. Siotani M. *Modern multivariate statistical analysis: A graduate course and handbook* / M. Siotani, T. Hayakawa, Y. Fujikoshi. – Columbus, Ohio : Amer. Sci. Press, 1985. – 759 p.
190. Sommer S. Robust variable selection in regression in the presence of outliers and leverage points / S. Sommer, R. G. Staudte // *Austral. J. Statist.* – 1995. – Vol. 37. – P. 323–336.
191. Srivastava R. Robustness of some designs against missing observations / R. Srivastava, V. K. Gupta, A. Dey // *Communications in Statistics – Theory and Methods*. – 1990. – Vol. 19. – P. 121–126.
192. Taylor J. Joint modelling of location and scale parameters of the t distribution / J. Taylor, A. Verbyla // *Statistical Modelling*. – 2004. – Vol. 4. – P. 91–112.
193. Wiens D. P. Robust designs for approximately linear regression: M -estimated parameters / D. P. Wiens // *J. of Statistical Planning and Inference*. – 1994. – Vol. 40. – P. 135–160.
194. Zellner A. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias / A. Zellner // *J. Amer. Statist. Assoc.* – 1962. – Vol. 57. – P. 348–368.
195. Zocchi S. S. Optimum experimental designs for multinomial logistic models / S. S. Zocchi, A. C. Atkinson // *Biometrics*. – 1999. – Vol. 55. – P. 437–444.