

0.13.16

Одесский ордена Трудового Красного Знамени
политехнический институт
Кафедра прикладной математики

На правах рукописи

Машеров Евгений Леонидович

Методы оценивания зависимостей, использующие сингулярное разложение.
Смещенные и несмещенные оценки.

Специальность 05.13.16 "Применение вычислительной техники,
математического моделирования и математических методов в научных
исследованиях"

Диссертация на соискание ученой степени
кандидата технических наук

Научный руководитель
к.т.н. Востров Г.Н.

Машеров

Одесса, 1991

КНИГА ИМЕЕТ

печатных	Выпуск	В перепл. сзип. соедин. №№ вып.	Таблиц	Карт	Иллюстр.	Служебн. №№	№№	списка и порядковые номера	197 г. 200 г. 2
1	1					4	27	60	

Содержание

Введение	3
1. Модель регрессионного анализа и оценки наименьших квадратов.	
1.1 Основные предпосылки регрессионного анализа	7
1.2 Сингулярное разложение матриц	12
1.3 Оценки метода наименьших квадратов. Их построение и статистические свойства.	16
1.4 Оценки метода наименьших квадратов в условиях ошибок вычисления	25
2. Взвешенные оценки регрессионного анализа.	
1.2 Включение в модель весов	32
1.2 Оптимальный выбор весов.	39
1.3 Оценивание значений оптимальных весов	49
1.4 Исключение переменных посредством весов	55
1.5 Учет ограничений общего вида	62
3. Методы проверки предпосылок регрессионного анализа и качества моделей.	
3.1 Скользящий экзамен.	66
3.2 Поиск отклонений от нормальности и построение устойчивой к ним оценки	73
3.3 Случайные регрессоры.	79
3.4 Нелинейное оценивание и классификация с учителем	83
4. Практическое применение W-оценок	
4.1 Имитационный эксперимент для проверки работоспособности W-оценок	89
4.2 Имитационный эксперимент для проверки работоспособности V-оценок	97
4.3 Результаты приложения методов смещенного регрессионного анализа в медицинских задачах.	99
4.4 Результаты приложения методов смещенного регрессионного анализа в экономических задачах.	109
4.5 Результаты приложения методов смещенного регрессионного анализа в задаче прогнозирования дисперсности катализатора	114
Заключение	117
Список литературы	119
Приложения	123

ВВЕДЕНИЕ

Целью данной работы является создание новых методов и алгоритмов регрессионного анализа и разработка пакетов прикладных программ, реализующих эти методы и алгоритмы. Широкое использование регрессионного анализа в разнообразных отраслях науки и практики: экономике[9], технике[30], медицине[33], социальных науках[49] и т.п. определяет практическую важность работы. Теоретическая же ее значимость следует из того, что регрессионный анализ представляет собой один из наиболее разработанных разделов математической статистики, так что построение более совершенных и эффективных методов оценивания параметров регрессионных моделей по сравнению с известными, будет способствовать развитию математической статистики в целом.

В настоящее время предложен ряд оценок, отличных от традиционных, причем в имитационном эксперименте или практическом применении эти оценки показали себя лучше традиционных. К их числу следует отнести прежде всего ридж-регрессию[41], сжатые оценки[43], оценки Марквардта[46], регрессию на главные компоненты[4], метод характеристического корня[38] и т.п. Эти оценки рассматриваются обыкновенно, как конкурирующие между собой и с методом наименьших квадратов. В работе показано, что все эти оценки, включая метод наименьших квадратов, представляют собой частные случаи оценок общего вида, получаемые соответствующим выбором параметров. Доказано, что в этом классе существует оценка, наилучшая одновременно как с точки зрения оценивания коэффициентов модели, так и с точки зрения оценивания математического ожидания вектора отклика. Разработан метод построения аппроксимации такой оценки.

Созданный метод построения регрессионных оценок требует больших затрат машинного времени, чем стандартная вычислительная процедура метода наименьших квадратов, поскольку взамен обращения матриц используется сингулярное или спектральное разложение. Этот недостаток, впрочем, в условиях применения ЭВМ не столь существенный, частично компенсируется возможностью построения с минимальными вычислительными затратами семейств оценок, отличающихся различным выбором множества регрессоров. Более важное преимущество разработанного подхода состоит в высокой устойчивости к погрешностям вычисления, что оказывается особенно важным в условиях мультиколлинеарности.

Для вычисления сингулярного разложения матриц в работе представлен алгоритм, аналогичный алгоритму Голуба-Бизингера-Райнша [37], но несколько более эффективный в вычислительном отношении. На основе этих алгоритмов разработан пакет прикладных программ, ориентированный на ЭВМ типа IBM PC.

Предложенные в работе оценки не обладают, вообще говоря, такими свойствами, как несмещенность, максимальное правдоподобие и т.п. Для обоснования целесообразности их применения был проведен имитационный эксперимент, описание и результаты которого приводятся в работе. В этом эксперименте ошибка оценки вектора коэффициентов модели и ошибка оценки вектора отклика (равная ошибке оценки вектора остатков) для предложенных оценок сравнивалась с соответствующими величинами для наиболее распространенного метода наименьших квадратов. Указанные критерии качества моделей охватывают основные классы задач, решаемых регрессионным анализом. Результаты эксперимента в сочетании с опытом практического применения предложенных оценок позволяют

елать вывод о целесообразности их применения и дальнейшего звития.

Критерии значимости моделей, разработанные для метода именьших квадратов, оказываются для данного класса оценок, обще говоря, непригодными, поэтому существенную часть работы ставляют алгоритмы проверки моделей, основанные на "скользящем замене". Эта процедура, дорогостоящая в вычислительном отношении и использовании сингулярного разложения матриц оказывается сьма эффективной, поскольку затраты на проверку модели малы в авнении с затратами на ее построение. Возможно применение кользящего экзамена" и к модели, полученной методом наименьших адратов или к моделям, полученным варьированием подмножества грессоров, тем самым дополняются традиционные критерии ачимости.

Полученные оценки допускают обобщение на случай оценивания и наличии ограничений-равенств и неравенств. Это позволяет ффективно формализовать априорную инфрмацию, доступную следователю.

Основное внимание в работе уделено случаю детерминированной трицы плана и нормального распределения ошибок. Однако пользование сингулярного разложения позволяет рассмотреть также один из возможных классов оценок со случайной матрицей плана, эдставляющий собой обобщение оценок метода максимального авдоподобия и оценок метода инструментальных переменных.

Отклонение от нормальности распределения вектора ошибок может ественно повлиять на поведение алгоритмов оценивания. едствие этого в работе рассмотрены критерии проверки малльности вектора ошибок, основанные на "скользящем экзамене",

также распространение отклонений от нормальности в промежуточных окончательных результатах.

Использование сингулярного разложения и критерия "скользящего ксзамена" позволило построить эффективный в вычислительном тношении алгоритм выбора наилучшего подмножества переменных.

Вопросы оценивания нелинейной регрессии и классификации с чителем рассмотрены в работе постольку, поскольку эти задачи огут быть сведены к оцениванию (возможно, многократному) линейных оделей.

Результаты работы прошли практическую проверку при решении кономических задач (прогнозирования, построения нормативов, лассификации) в системе управления Черноморским морским ароходством, а также при построении многомерных эмпирических ависимостей при решении научно исследовательских задач Одесского изико-химического института им. А. В. Богатского АН УССР.

Модель регрессионного анализа и оценки наименьших квадратов.

..Основные предпосылки регрессионного анализа.

Регрессионный анализ предполагает наличие одной и более входных переменных (регрессоров) и единственной выходной переменной (отклика), связанных между собой известной с точностью до параметров линейной или нелинейной зависимостью в условиях действия случайных факторов. Ограничиваясь вначале линейной зависимостью отклика от регрессоров и случайного фактора, приходим к модели вида:

$$y = X\beta + \varepsilon \quad (1)$$

где вектор y - выходная переменная (отклик)

матрица X - входные переменные (регрессоры)

β - вектор параметров модели

вектор ε - возмущающее воздействие

Предполагается, что X известны и детерминированы. Если же X получены каким-либо вероятностным механизмом, предполагаем, что нам доступна лишь одна реализация, и в силу этого рассматриваем ее как детерминированную. Известен также и вектор y . Вектор же β неизвестен и подлежит оценке. Он также предполагается детерминированным.

Введем следующие обозначения: наблюдаемые величины или непосредственно вычисляемые из наблюдаемых величин обозначаются латинскими буквами без дополнительных пометок; математические ожидания обозначаются теми же буквами, но с горизонтальной чертой сверху: \bar{y} ; оценки истинных значений параметров выделяются

крышкой" над ними: \hat{y} ; истинные же значения параметров будут, как правило, обозначаться греческими буквами или же латинскими буквами, если это не приводит к неоднозначности. Для удобства будем предполагать, что везде, где это не оговорено особо, число аблюдений (строк матрицы X , элементов векторов y и ε) равно m , а переменных (столбцов X , элементов вектора β) равно n .

Целью регрессионного анализа может быть как оценка вектора параметров, так и оценка вектора отклика, а в некоторых случаях и оценка вектора ошибок (возмущающего воздействия, ε). Иногда же задача сводится к установлению факта зависимости y от X .

Перечислим используемые в работе предпосылки регрессионного анализа [34]:

.. Зависимость y от X является линейной.

Это ограничение не столь существенно, поскольку можно использовать нелинейное преобразование столбцов X и/или вектора y (в последнем случае преобразуется также и спецификация ошибки). Такое преобразование существует для многих, хотя и не для всех нелинейных моделей. Если же подобное преобразование невозможно, оценивание может производиться посредством оценивания последовательности линейных моделей. В действительности существенно лишь более слабое условие:

.. Зависимость y от X линейна по параметрам или может быть приведена к виду, линейному по параметрам.

Вектор ε случаен.

.. Математическое ожидание вектора ε равно нулю.

3. Элементы вектора ε независимы и одинаково распределены.

В некоторых случаях понадобится более сильное предположение:

3+. Элементы вектора ε имеют нормальное распределение с неизвестной, но одинаковой для всех элементов вектора дисперсией σ^2 .

Подобное предположение понадобится, в основном, для обоснования критериев значимости. Для построения методов оценивания будет достаточно сферической симметрии распределения.

Матрица X детерминирована.

Хотя погрешности вычисления и измерения делают матрицу X случайной, однако в большинстве случаев этим можно пренебречь. Если же погрешности столь велики, что матрица не может рассматриваться, как детерминированная, следует отказаться от использования МНК и применять иные методы, один из которых рассмотрен в работе.

Матрица X имеет полный ранг.

Случай неполноты ранга заслуживает особого внимания. Хотя случай истинной неполноты ранга, когда при точной арифметике ранг матрицы X меньше числа столбцов, встречается лишь, когда некоторые столбцы X являются линейными комбинациями других столбцов, а в этом случае исключением части переменных можно без потери информации избавиться от этого, однако приближительная неполнота ранга ("нестрогая мультиколлинеарность") часто встречается на практике. Таковы почти все эконометрические модели или же регрессия на

полиномы высоких порядков.

Вектор $\beta \in \mathbb{R}^n$.

Иначе говоря, на коэффициенты модели не наложено никаких ограничений.

В модели (1) вектор ε неизвестен и, более того, принципиально ненаблюдаем. Поэтому для построения оценки коэффициентов модели ледует принять некоторые дополнительные предположения. Со времен аусса обыкновенно предполагают, что наилучшим выбором вектора $\hat{\beta}$ удет вектор, минимизирующий функционал $\|y - \hat{y}\|_2$. Такой подход ринято называть методом наименьших квадратов (МНК). В альнейшем для оценки качества моделей будем применять также вадратические функционалы, но, возможно, иного вида, в том числе ависящие от непосредственно не наблюдаемых величин (в этом лучае минимизации подлежит оценка этого функционала). Наряду с вадратической функцией применяются также функции, использующие умму абсолютных значений разностей или максимальную по абсолютной эличине разность. Это, соответственно, метод наименьших модулей (МНМ) и чебышевская аппроксимация. Все три перечисленных ункционала суть расстояния в метрике l_1 , l_2 или l_∞ соответственно. литературе предлагались также метрики $l_{1..5}$, $l_{1..8}$ и т.п. [8].

Квадратичный функционал применяется чаще прочих отчасти из-за остоты аналитических преобразований, отчасти же из гатистических соображений. Так, в предположении о нормальности испределения ε МНК есть метод максимального правдоподобия, а кже несмещенно эффективный метод. Однако каждое из этих

нований само по себе, вообще говоря, недостаточно, и в зависимости от задачи приходится отказываться от применения МНК. К, при наличии существенных отклонений от нормальности распределения ε , например, если ε имеет распределение Лапласа первого рода, приходится использовать ММ, а в ряде задач мыслена лишь чебышевская аппроксимация.

3. Сингулярное разложение матриц.

Дальнейшее изложение основывается на возможности представления матриц в виде произведения матриц специального вида. Известно, что квадратные симметричные матрицы представимы в виде

$$X = C^T \Lambda C$$

где C - матрица собственных векторов

Λ - диагональная матрица собственных значений

причем $C^T C = C C^T = I$. Обобщение этого разложения на прямоугольные матрицы размерностью $m \times n$ ($m \geq n$) может быть получено следующим образом. Рассмотрим вначале квадратные несимметричные матрицы. Спектральное их разложение требует, вообще говоря, использования комплексных чисел, поэтому, желая работать лишь с действительными числами, допустим различные левый и правый сомножители в разложении

$$X = S \Lambda C$$

однако по-прежнему будем требовать ортогональности матриц C и S (но не взаимной ортогональности) и диагональности матрицы Λ . Такое разложение возможно. Для его получения рассмотрим матрицу $X^T X = C^T \Lambda S^T S \Lambda C = C^T \Lambda^2 C$, что позволяет, используя спектральное разложение, определить множители C и Λ . Аналогично, посредством матрицы $X X^T$, определяется множитель S . При этом, вычисляя значение λ , выбираем положительное значение корня квадратного.

Обобщение на прямоугольные матрицы очевидно. Следует лишь ослабить требования к ортогональности матрицы S , поскольку она является не квадратной, а наследует размерности матрицы X . Если же потребовать, чтобы матрица S была квадратной, то она будет

определяться неоднозначно, матрица Λ же будет наследовать размерность X . Возможна также неоднозначность, связанная с наличием кратных собственных значений матрицы $X^T X$, при этом собственные вектора будут алгебраически кратными. В условиях конечной точности вычислений вероятность подобного случая весьма мала.

Таким образом, любая действительная матрица может быть представлена в виде:

$$X = S \Lambda C \quad C^T C = C C^T = S^T S = I \quad (2)$$

очевидна связь этого разложения с полярным разложением [29]

$$X = Z A \quad Z^T Z = I \quad A = A^T$$

если положить

$$Z = S C^T \quad A = C^T \Lambda C$$

Используя сингулярное разложение, легко выразить псевдообратную матрицу [3] посредством C, Λ и S (этот способ вычисления псевдообратных матриц является наиболее точным в вычислительном отношении, хотя и не самым экономным):

$$X^+ = C^T \Lambda^+ S^T$$

где $\lambda_i^+ = \begin{cases} \lambda_i^{-1} & \lambda_i > 0 \\ 0 & \lambda_i = 0 \end{cases}$

при этом легко обнаружить потерю ранга матрицы, проявляющуюся в равенстве нулю или малости некоторых λ .

Использование сингулярного разложения упрощает целый ряд выражений, что приводит к повышению устойчивости алгоритма к ошибкам округления. Другой причиной повышения устойчивости алгоритма является то, что сингулярное разложение, вычисляемое

ыкновенно посредством ортогональных преобразований, вообще более
гойчивая по сравнению с обращением матриц процедура. (Иные,
мимо ошибок округления, источники вычислительных ошибок для
льшинства ЭВМ, в частности, для IBM PC и ЕС ЭВМ, в силу
обенностей их архитектуры, несущественны).

Для вычисления сингулярного разложения, в большинстве случаев
целесообразно предварительно вычислять матрицу $X^T X$, поскольку
и этом происходит некоторая потеря точности в силу того, что для
трицы $X^T X$ число обусловленности равно квадрату числа
условленности для матрицы X . Лишь в случае, когда число строк
трицы X многократно превосходит число столбцов, выигрыш в
чности, обусловленный сокращением объема вычислений, превосходит
теру точности, порожденную схемой вычислений.

Для непосредственного вычисления сингулярного разложения
лесообразно использовать алгоритмы, основанные на ортогональных
еобразованиях - отражениях и вращениях. В работе [29] предложено
общение алгоритма Якоби спектрального разложения матрицы.
стоинством его является высокая степень ортогональности матриц S
 S . В справочнике [28] приведен алгоритм, использующий
-разложение двухдиагональных матриц. В вычислительном отношении
ущественно более эффективен, однако ортогональность матриц для
го несколько ниже. Возможно также построение алгоритма, на
рвом шаге которого строится полярное разложение матриц, как
исано в [29], затем вычисляется спектральное разложение
тученной симметричной матрицы.

На основе сравнения вычислительных реализаций был выбран

описанный в [37] алгоритм, однако вместо вращений Гивенса[24] были использованы быстрые вращения Гивенса-Джентлмена. Это позволило приблизительно на 15% ускорить алгоритм при некотором повышении его точности.

Следует отметить, что вычислительные затраты для нахождения сингулярного разложения матриц в 3-5 и более раз превосходят затраты на обращение матриц одинаковой размерности и, в частности, делают невозможными ручные расчеты даже для малых матриц. Следовательно, предлагаемые методы существенно ориентированы на применение вычислительной техники.

3. Оценки метода наименьших квадратов. Их построение и статистические свойства.

Рассмотрим модель вида (1). Для получения оценок вектора положим, что наилучшие оценки есть оценки, минимизирующие в метрике l_2 расстояние между векторами y и $\hat{y} = X\hat{b}$. Выражение $\|y - \hat{y}\|_2^2$, учитывая (2), можно привести к виду:

$$\begin{aligned} \min_b \|y - \hat{y}\|_2^2 &= (y - X\hat{b})^T (y - X\hat{b}) = (y - S\Lambda C\hat{b})^T (y - S\Lambda C\hat{b}) = \\ &= (y - SS^T y + S^T S y - S\Lambda C\hat{b})^T (y - SS^T y + SS^T y - S\Lambda C\hat{b}) = \\ &= ((I - SS^T)y + S(r - \Lambda C\hat{b}))^T ((I - SS^T)y + S(r - \Lambda C\hat{b})) \end{aligned}$$

где $r = S^T y$.

Очевидно, первое слагаемое лежит в пространстве, ортогональном X . Второе же принадлежит пространству, натянутому на столбцы матрицы S . Следовательно, они ортогональны.

$$\begin{aligned} \|y - \hat{y}\|_2^2 &= y^T (I - SS^T)^T (I - SS^T) y + (r - \Lambda C\hat{b})^T S^T S (r - \Lambda C\hat{b}) = \\ &= y^T (I - SS^T) y + (r - \Lambda C\hat{b})^T (r - \Lambda C\hat{b}) = \text{const} + (r - \Lambda C\hat{b})^T (r - \Lambda C\hat{b}) \end{aligned}$$

Первое слагаемое не зависит от \hat{b} , минимум же второго достигается очевидным выбором:

$$b = C^T \Lambda^{-1} r = C^T \Lambda^{-1} S^T y \quad (3)$$

Этот минимум является глобальным в силу свойств неотрицательно определенных квадратичных форм. Единственность и существование оценки обеспечивается полнотой ранга матрицы X , а, следовательно, отличием от нуля всех сингулярных значений этой матрицы. Нарушение этого условия приводит к неоднозначности минимума функционала. При этом, однако, оценка существует, хотя и не может быть вычислена по приведенной формуле.

Легко убедиться, что оценки вида (3) есть, в сущности, лишь

иная форма выражения традиционных оценок метода наименьших квадратов:

$$(X^T X)^{-1} X^T y = (C^T \Lambda^T S A C)^{-1} C^T \Lambda S y = (C^T \Lambda^2 C)^{-1} C^T \Lambda S^T y = \\ C^T \Lambda^{-2} C C^T \Lambda S^T y = C^T \Lambda^{-2} \Lambda S^T y = C^T \Lambda^{-1} S^T y$$

Рассмотрим их статистические свойства.

$$E(\hat{b}) = E(C^T \Lambda^{-1} S^T y) = E(C^T \Lambda^{-1} S^T (X\beta + \varepsilon)) = \beta$$

(Здесь используется тот факт, что $E(\varepsilon) = 0$)

$$D^2(\hat{b}) = E((\hat{b} - \beta)(\hat{b} - \beta)^T) = E((C^T \Lambda^{-1} S^T \varepsilon)(C^T \Lambda^{-1} S^T \varepsilon)^T) = \\ C^T \Lambda^{-2} C E(\varepsilon \varepsilon^T) = \sigma^2 C^T \Lambda^{-1} C = \sigma^2 (X^T X) \quad (4)$$

Следовательно, оценка (3) является несмещенной и ее ковариационная матрица выражается формулой (4).

Оценки вектора y имеют вид:

$$y = Xb = S A C C^T \Lambda^{-1} S^T y = S S^T y \quad (5)$$

Их статистические свойства таковы:

$$E(\hat{y}) = E(S S^T X \beta + S S^T \varepsilon) = S S^T S A C \beta + S S^T E(\varepsilon) = S A C \beta = X \beta$$

Следовательно, оценка \hat{y} также является несмещенной.

$$D^2(\hat{y}) = E((y - E(\hat{y}))(y - E(\hat{y}))^T) = E(S S^T \varepsilon \varepsilon^T S S^T) = S S^T E(\varepsilon \varepsilon^T) = \sigma^2 S S^T \quad (6)$$

Обозначая $e = (y - \hat{y}) = (I - S S^T) y = (I - S S^T) \varepsilon$, получим

$$D^2(e) = E(((I - S S^T) \varepsilon)((I - S S^T) \varepsilon)^T) = \sigma^2 (I - S S^T) \quad (7)$$

Использование сингулярного разложения позволяет характеризовать влияние изменений элементов вектора ε на значения элементов векторов r , \hat{b} , \hat{y} , и e . В качестве меры влияния естественно использовать частные производные элементов этих векторов по элементам вектора ε . Поскольку зависимости эти линейны,

$$r_j = \sum s_{ij} y_i = \sum s_{ij} \bar{y}_i + \sum s_{ij} \varepsilon_i \quad (8)$$

$$\hat{b}_i = \sum_k \sum_j c_{ij} \lambda_j^{-1} s_{kj} y_k = \sum_k \sum_j c_{ij} \lambda_j^{-1} s_{kj} \bar{y}_k + \sum_k \sum_j c_{ij} \lambda_j^{-1} s_{kj} \varepsilon_k \quad (8')$$

$$(8'')$$

$$\hat{y}_i = \sum_k \sum_j s_{ij} s_{kj} y_k = \sum_k \sum_j s_{ij} s_{kj} \bar{y}_k + \sum_k \sum_j s_{ij} s_{kj} \varepsilon_k$$

$$e_i = (1 - \sum_j s_{ij}^2) \varepsilon_k + \sum_{k \neq i} \sum_j s_{kj} \varepsilon \quad (8''')$$

частная производная характеризует влияние не только малых изменений, так что возможно использование ее для изучения влияния больших отклонений (грубых ошибок, выбросов, зашумлений и т.п.) вектора ε , а также для изучения распределения векторов оценок, если распределение ε отлично от нормального. Частные производные имеют вид:

$$\frac{\partial r_i}{\partial \varepsilon_k} = s_{ik} \quad (9)$$

$$\frac{\partial b_i}{\partial \varepsilon_k} = \sum_j c_{ij} \lambda_j^{-1} s_{kj} \quad (9')$$

$$\frac{\partial y_i}{\partial \varepsilon_k} = \sum_j s_{ij} s_{kj} \quad (9'')$$

$$\frac{\partial e_i}{\partial \varepsilon_k} = \begin{cases} (1 - \sum_j s_{ij}^2) & i=k \\ \sum_j s_{ij} s_{kj} & i \neq k \end{cases} \quad (9''')$$

величина частной производной характеризует чувствительность соответствующего элемента вектора оценки к значениям элементов вектора ε и позволяет выявлять точки разбалансировки моделей, то есть такие значения k , для которых ошибки в измерении y_k могут существенно исказить модель. В зависимости от конкретной обстановки исследователь может получить дополнительное наблюдение для значений регрессоров x_{k*} , задать k -тому наблюдению меньший ε , исключить его из выборки или перейти к использованию

робастных оценок. Понятие точки разбалансировки введено в теории робастных оценок [42], однако в качестве меры разбалансировки в указанной работе используется величина, равная норме строки матрицы частных производных \mathbf{b} по ε .

Для исследования распределения оценок в качестве параметров распределения используем семиинварианты этого распределения. Предполагается, что распределение ε имеет конечные семиинварианты всех интересующих нас порядков. Это требование не является слишком ограничительным, поскольку в действительности значения элементов вектора ε ограничены из-за конечности элементов \mathbf{y} , так что ограничены и семиинварианты. Используем соотношения [15]

$$\kappa_i(\eta + \theta) = \kappa_i(\eta) + \kappa_i(\theta)$$

$$\kappa_i(a\eta) = a^i \kappa_i(\eta)$$

где κ_i есть семиинварианты i -того порядка.

Полагая, что все элементы вектора ε распределены одинаково, получим:

$$\kappa_i(r_j) = \kappa_i(\varepsilon) \sum s_{kj}^i \quad (10)$$

$$\kappa_i(\hat{b}_j) = \kappa_i(\varepsilon) \sum_k \left(\sum_l c_{jl} \lambda_l^{-1} s_{kl} \right)^i \quad (10')$$

$$\kappa_i(y_j) = \kappa_i(\varepsilon) \sum_k \left(\sum_l s_{kl} s_{jl} \right)^i \quad (10'')$$

$$\kappa_i(e_j) = \kappa_i(\varepsilon) \left(\left(1 - \sum_l s_{jl}^2 \right)^i + \sum_{k \neq j} \left(\sum_l s_{jl} s_{kl} \right)^i \right) \quad (10''')$$

Рассмотрим прежде всего распределение элементов вектора \mathbf{r} . Дисперсия его элементов одинакова и равна σ^2 , так что

нормированные семиинварианты его выражаются через нормированные семиинварианты ε , как

$$\bar{x}_i(r_j) = \bar{x}_i(\varepsilon) \sum S_{kj}^i$$

где \bar{x}_i — нормированные семиинварианты. Поскольку матрица S нормирована по столбцам, справедливо неравенство:

$$\bar{x}_i(\varepsilon) \geq \bar{x}_i(r_j) \geq \begin{cases} 0 & i \text{ нечетное} \\ m^{-1/2} \bar{x}_i(\varepsilon) & i \text{ четное} \end{cases} \quad (110)$$

При этом верхняя граница соответствует случаю, не встречающемуся, — видимому, на практике, а именно, когда n столбцов матрицы S содержат единичные элементы, следовательно, $m-n$ строк матрицы S , значит, и матрицы X , равны нулю. Нижняя же граница достижима, например, если матрица S представляет собой фрагмент матрицы гамары, умноженный на нормирующий множитель. Матрица такого типа может представлять собой, например, план эксперимента, так что нижняя граница вполне реальна. Возможно появление такой матрицы и в пассивном эксперименте. Из сказанного следует, что элементы r не могут быть распределены "менее нормально", чем элементы ε ; нормализация же элементов r вполне возможна. Рассмотрим это явление на конкретном примере. Для этого возьмем выборку [44], которую используем при тестировании программ регрессионного анализа, и вычислим по формулам (10)–(10''') нормированные семиинварианты. В таблице 1 приведены значения отношений нормированных семиинвариантов вектора r к нормированным семиинвариантам вектора ε . Очевидно, элементы вектора r могут иметь распределение, близкое к нормальному, даже если

распределение ε существенно отличается от нормального. Этот факт используется в главе 2 при построении "оптимально взвешенных оценок". Тем не менее проверка предположения о нормальности распределения ε остается необходимой.

Табл. 1

Номер	$\chi^2_{\text{Асим-}}_{\text{метрия}}$	$\chi^2_{\text{Экс-}}_{\text{цесс}}$	$\chi^2_{\text{Б}}$	$\chi^2_{\text{В}}$
1	-0.128	0.197	-0.057	0.053
2	-0.101	0.150	-0.029	0.032
3	-0.025	0.148	0.003	0.028
4	0.206	0.196	0.069	0.049
5	-0.105	0.138	-0.030	0.023
6	-0.251	0.063	-0.016	0.004

Из рассмотрения семиинвариантов распределения ε следует, что и также, хотя и в меньшей степени, испытывают на себе деморализующее воздействие (табл. 2). Следовательно, тесты проверки нормальности, использующие вектор ε , могут приводить к излишне оптимистичным выводам в отношении нормальности распределения ε . Это объясняется известным фактом, когда тесты χ^2 или Колмогорова-Смирнова, примененные к вектору остатков, не реагируют на редкие выбросы. В связи с этим для проверки нормальности, а также проверки выполнения других предположений регрессии, целесообразно использовать технику "скользящего экзамена". Эти подходы к проверке будут описаны в главе 3.

Табл. 2

Номер	$\alpha_{\text{Асим-}}^3$ Метрия	$\alpha_{\text{Экс-}}^3$ Цесс	α_5	α_6
1	0.509	0.363	0.258	0.193
2	0.304	0.246	0.125	0.090
3	0.587	0.430	0.332	0.263
4	0.577	0.422	0.320	0.250
5	0.716	0.617	0.539	0.476
6	0.649	0.501	0.404	0.333
7	0.534	0.390	0.289	0.222
8	0.445	0.310	0.211	0.153
9	0.438	0.334	0.222	0.166
10	0.768	0.673	0.604	0.545
11	0.715	0.611	0.534	0.470
12	0.433	0.299	0.198	0.141
13	0.578	0.445	0.351	0.282
14	0.725	0.617	0.540	0.476
15	0.618	0.459	0.356	0.282
160.395	0.232	0.126	0.073	

Распределение оценок коэффициентов модели также испытывает нормализующее воздействие (табл. 3). При этом распределение отдельных коэффициентов может отличаться весьма существенно. В таблице 4 приведены сведения о распределении оценки отклика.

Табл. 3

Номер	μ_3 Асим- метрия	μ_4 Экс- цесс	μ_5	μ_6
1	-0.120	0.195	-0.053	0.052
2	0.188	0.170	0.063	0.042
3	0.008	0.151	0.004	0.029
4	0.036	0.214	0.024	0.062
5	-0.106	0.133	-0.030	0.022
6	0.161	0.151	0.049	0.032

Табл. №4

Номер	азим- метрия	ас- цесс	μ_5	μ_6
1	0.391	0.238	0.132	0.083
2	0.434	0.352	0.239	0.183
3	0.396	0.196	0.101	0.055
4	0.402	0.218	0.113	0.063
5	0.282	0.146	0.057	0.028
6	0.420	0.211	0.106	0.055
7	0.376	0.221	0.121	0.073
8	0.389	0.254	0.156	0.105
9	0.353	0.259	0.143	0.100
10	0.363	0.157	0.069	0.032
11	0.296	0.124	0.048	0.020
12	0.403	0.269	0.168	0.115
13	0.328	0.174	0.084	0.047
14	0.342	0.147	0.061	0.027
15	0.437	0.236	0.127	0.071
16	0.578	0.431	0.328	0.256

4. Оценки метода наименьших квадратов в условиях ошибок вычисления.

При рассмотрении свойств оценок регрессии выше предполагалось, что вычисления производятся точно, так что оценки вида (3) эквивалентны традиционным оценкам МНК. Однако реальные вычисления всегда производятся в условиях ошибки, связанной прежде всего с погрешностью округления результатов операций (для ряда ЭВМ, в том числе для ЭВМ Единой Серии, а также для ЭВМ, совместимых с IBM PC, ввиду наличия дополнительного разряда можно считать, что операции производятся без ошибок, и лишь при округлении непосредственно перед записью в память возникает погрешность).

Распространение ошибки зависит от особенностей реализации алгоритма, однако для большинства алгоритмов алгебры, в том числе для алгоритмов обращения и сингулярного разложения матриц [48], включенных в библиотеки программ, справедливо утверждение, что результат вычисления с конечной точностью оказывается равен результату точного вычисления над некоторой возмущенной матрицей, причем норма возмущения не превосходит элементарной ошибки округления, умноженной на зависящий от вида алгоритма и размерности матрицы коэффициент.

Рассмотрим вычислительную погрешность оценки (3) в сравнении с традиционной схемой вычисления такой оценки, использующей обращение матрицы.

Обозначим:

$$\hat{\mathbf{b}} = \mathbf{C}^T \mathbf{\Lambda}^{-1} \mathbf{S}^T \mathbf{y} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

$$\tilde{\mathbf{b}} = \tilde{\mathbf{C}}^T \tilde{\boldsymbol{\lambda}}^{-1} \tilde{\mathbf{S}}^T \mathbf{y}$$

$$\tilde{\mathbf{b}} \approx (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

где символ \sim обозначает величину, вычисленную с ошибкой. При этом справедливо:

$$\begin{aligned} \|\tilde{\mathbf{C}} - \mathbf{C}\| &< K_{m,n}^1 \delta_{\text{маш}} \\ \|\tilde{\mathbf{S}} - \mathbf{S}\| &< K_{m,n}^2 \delta_{\text{маш}} \\ \|\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}\| &< K_{m,n}^3 \delta_{\text{маш}} \end{aligned}$$

$$\tilde{\mathbf{R}}^{-1} = (\mathbf{R} + \Delta \mathbf{R})^{-1}$$

$$\|\Delta \mathbf{R}\| < K_n^4 \delta_{\text{маш}}$$

Где $\delta_{\text{маш}}$ есть величина элементарной ошибки округления, зависящая от типа ЭВМ, а $K_{m,n}^i$ - коэффициент, зависящий от размерности задачи и реализации алгоритма, меняющийся в сравнительно узких пределах.

Пренебрегая величинами порядка δ^2 , получим:

$$\tilde{\mathbf{b}} - \hat{\mathbf{b}} = \tilde{\mathbf{C}}^T \tilde{\boldsymbol{\lambda}}^{-1} \tilde{\mathbf{S}}^T \mathbf{y} - \mathbf{C}^T \boldsymbol{\lambda}^{-1} \mathbf{S}^T \mathbf{y} \approx \Delta_{\mathbf{C}}^T \boldsymbol{\lambda}^{-1} \mathbf{S}^T \mathbf{y} + \mathbf{C}^T \Delta_{\boldsymbol{\lambda}} \boldsymbol{\lambda}^{-2} \mathbf{S}^T \mathbf{y} + \mathbf{C}^T \boldsymbol{\lambda}^{-1} \Delta_{\mathbf{S}} \mathbf{y} \approx \mathbf{C}^T \Delta_{\boldsymbol{\lambda}} \boldsymbol{\lambda}^{-2} \mathbf{S}^T \mathbf{y}$$

Хотя величины $\Delta_{\mathbf{C}}$, $\Delta_{\mathbf{S}}$ и $\Delta_{\boldsymbol{\lambda}}$ примерно одного порядка, однако при наличии малых элементов среди сингулярных чисел второе слагаемое оказывается значительно больше первого и третьего, так что в первую очередь необходимо рассматривать его. Если же все сингулярные числа велики, ошибка вычисления мала в сравнении с погрешностью, вносимой случайностью вектора $\boldsymbol{\varepsilon}$, так что практический интерес представляет ошибка вычисления при наличии малых λ . Подобная ситуация в вычислительной практике именуется

плохой обусловленностью, в статистике же - мультиколлинеарностью. При этом множитель $\tilde{\Lambda}^{-1}$ является мультипликатором, усиливающим малую саму по себе вычислительную ошибку.

Для традиционной схемы МНК вычислительная ошибка может быть представлена, как:

$$\tilde{\hat{b}} - \hat{b} = (X^T X)^{-1} X^T y - (X^T X)^{-1} X y \approx (X^T X)^{-1} \Delta_R (X^T X)^{-1} X^T y = C^T \Lambda^{-2} (C \Delta_R C^T) \Lambda^{-1} S^T y$$

Здесь использовано приведенное в [7] выражение для производной от обратной матрицы и, также как и в предыдущем случае, величины порядка δ^2 не рассматриваются. Через Δ обозначено эквивалентное возмущение для алгоритма обращения матриц. Сомножитель $(C \Delta_R C^T)$, ввиду ортогональности матрицы C , не изменяет нормы эквивалентного возмущения, и сравнение двух оценок показывает, что для традиционной схемы мультипликатор есть Λ^{-2} .

На практике это приводит к тому, что использование сингулярного разложения при одинарной точности вычислений приводит к меньшей погрешности, чем обращение матриц при двойной точности. Так, для данных Лонгли [44], традиционно считающихся образцом мультиколлинеарности и в качестве таковых используемых, как тестовая выборка при проверке алгоритмов регрессионного анализа, использование сингулярного разложения с одинарной точностью ($\delta \approx 10^{-7}$) дало ошибку такого же порядка, как использование стандартной программы из ПНП-ПЛ с двойной точностью ($\delta \approx 10^{-14}$). Использование же стандартной программы при одинарной точности к осмысленным результатам не привело.

Еще более явно проявляются преимущества предлагаемой схемы вычислений при расчете оценки вектора y .

$$\hat{y} = SS^T y = X(X^T X)^{-1} X^T y$$

$$\tilde{y} = SS^T y$$

$$\tilde{y} \approx X(X^T X)^{-1} X^T y$$

$$\|y - \hat{y}\| < K_{m, n}^y \delta_{\text{маш}}$$

$$\tilde{y} - \hat{y} \approx X(X^T X)^{-1} \Delta_R (X^T X)^{-1} X^T y = S \Lambda^{-1} (C \Delta_R C^T) \Lambda^{-1} S^T y$$

Для традиционной схемы вычислений появляется мультипликатор порядка \tilde{K}^c , вовсе отсутствующий при использовании сингулярного разложения.

Разумеется, преимущества сингулярного разложения для вычисления оценок регрессии проявляются лишь при наличии малых сингулярных чисел, чего всегда можно избежать в активном эксперименте, применяя ортогональные или квазиортогональные планы. В пассивном же эксперименте этого избежать зачастую невозможно. Особенно сильно это проявляется при анализе экономических данных, в частности, если регрессоры и регрессанд суть временные ряды, а также, если регрессоры связаны между собой зависимостью, хотя бы и нелинейной, что часто обнаруживается в технических естественно-научных задачах. Традиционный подход к устранению этого недостатка – исключение некоторых переменных из модели. Из теоремы Коши о разделении следует, что сингулярные числа редуцированной матрицы разделяют сингулярные числа исходной, и, учитывая их неотрицательность, можно утверждать, что для редуцированной матрицы они больше, чем для исходной. Однако такое "хирургическое" лечение мультиколлинеарности приводит к потерям содержательной точки зрения и затрудняет интерпретацию результатов.

Иное обоснование необходимости исключения регрессоров – снижение при исключении дисперсии оставшихся коэффициентов. Из (4) и теоремы Коши о разделении следует, что такое снижение будет иметь место (следует отметить, что снижение оценки дисперсии коэффициентов, в общем случае не будет наблюдаться, поскольку при исключении коэффициентов может возрасти оценка $\hat{\sigma}^2$). Предложено несколько способов определения переменных, подлежащих исключению [27]. Однако все они, за исключением весьма трудоемкого с вычислительной точки зрения перебора всех возможных регрессий, носят полуэмпирический характер. Некоторые иные причины для исключения переменных будут рассмотрены в главе 2.

Следует отметить, что оценки, полученные исключением переменных, оказываются смещенными, если только истинное значение коэффициента регрессии на исключенные переменные не равно в точности нулю, несмотря на то, что применяемый к редуцированной матрице метод наименьших квадратов дает несмещенные оценки. Это существенно для сравнения методов исключения переменных с предложенными в настоящей работе методами, изначально ориентированными на смещенное оценивание.

Еще один возможный подход к улучшению модели, характерный для так называемых активно-пассивных экспериментов – дополнение матрицы X новыми строками с тем, чтобы уменьшить степень мультиколлинеарности. Такой подход целесообразен, если информация может быть получена как путем наблюдения, так и в эксперименте, причем стоимость наблюдения существенно ниже стоимости эксперимента в расчете на одну строку X .

В этом случае план эксперимента для пополнения матрицы регрессоров может быть получен, исходя из сингулярного разложения X . В пространстве, натянутом на столбцы S план, в наибольшей степени снижающий мультиколлинеарность, строится, как вектор вида $(0, 0, \dots, 1, 0, \dots)$, причем позиция единицы соответствует номеру наименьшего сингулярного числа матрицы X . Тогда наименьшее сингулярное число увеличивается в наибольшей степени, снижая тем самым мультиколлинеарность.

В пространстве же X план имеет вид:

$$x_+ = \gamma e \Lambda C$$

где x_+ - вектор-план дополнительного эксперимента.

e - вектор из нулей и единицы.

γ - множитель, выбираемый из условий эксперимента.

Если возможно проведение нескольких дополнительных экспериментов, можно применять такой подход многократно, всякий раз пересчитывая сингулярное разложение X . При этом, если план выбирается в точном соответствии с указанной формулой, то пересчет разложения весьма прост, поскольку матрица C остается неизменной, в диагональной матрице Λ меняется лишь один элемент и, возможно, их упорядочение по величине, в матрице же S изменяется лишь нормировка столбца, соответствовавшего минимальному сингулярному числу.

$$\lambda_i = (1 + \gamma) \lambda_i$$

$$s_{ij} = \frac{s_{ij}}{\sqrt{1 + \gamma}}$$

Если же одно или несколько сингулярных чисел равны нулю, то традиционная оценка вовсе не существует. Оценка же вида (4) может быть легко модифицирована с тем, чтобы работать и при неполноте ранга X . Для этого рассмотрим оценку вида

$$b^+ = C^T \Lambda^+ S^T y$$

$$\text{где } \lambda_i^+ = \begin{cases} \lambda_i^{+1} & \lambda_i \neq 0 \\ 0 & \lambda_i = 0 \end{cases}$$

Она дает при неполноте ранга решение минимальной нормы. Однако конечная точность вычислений приводит, как правило, к тому, что ни одно из сингулярных чисел не равно в точности нулю, хотя некоторые из них имеют порядок величины ошибки вычисления. В этом случае предложено [3] использовать оценку вида:

$$b^* = C^T \Lambda^* S^T y$$

$$\text{где } \lambda_i^* = \begin{cases} \lambda_i^{+1} & \lambda_i > \delta \\ 0 & \lambda_i \leq \delta \end{cases}$$

Основная трудность при использовании этой оценки связана с выбором параметра δ или, что эквивалентно, с определением числа априори обнуляемых сингулярных чисел.

Поскольку наибольший вклад в оценку дают наименьшие сингулярные числа, то даже малые колебания параметра δ могут приводить к резким колебаниям оценки вектора коэффициентов и к (не столь сильным) колебаниям оценки вектора отклика. Такое же действие будут оказывать малые колебания величин λ . Приведенная выше оценка оказывается разрывной функцией от δ и λ , что особенно опасно при автоматическом выборе параметра δ .

Взвешенные оценки регрессионного анализа.

1. Включение в модель весов.

Рассмотренные в предыдущей главе оценки \mathbf{b}^* и \mathbf{b}^+ уже не являются несмещенными, как оценки МНК, однако за они приобрели некоторые новые полезные свойства, прежде всего осмысленность при неполноте ранга матрицы регрессоров. Представляется привлекательной возможность использовать отказ от требования несмещенности для улучшения свойств модели и в других отношениях, например, для сокращения дисперсии. Однако оценки \mathbf{b}^+ и \mathbf{b}^* не слишком перспективны, поскольку, во-первых, управляются лишь одним параметром δ , во-вторых, зависимость от этого параметра носит разрывный характер, в-третьих, используется лишь информация о матрице X , заключенная в ее сингулярных значениях, и вовсе не используется информация о векторе y , а также какая-либо априорная информация. Для построения более перспективной системы оценок запишем приведенные выше оценки в несколько ином виде:

$$\mathbf{b}^+ = \mathbf{C}^T \mathbf{W}^+ \Lambda^{-1} \mathbf{S}^T \mathbf{y}$$

$$\mathbf{b}^* = \mathbf{C}^T \mathbf{W}^* \Lambda^{-1} \mathbf{S}^T \mathbf{y}$$

где $w_{ij} = w_{ji} = 0 \quad i \neq j$

$$w_{ii}^+ = \begin{cases} 1 & \lambda_i \neq 0 \\ 0 & \lambda_i = 0 \end{cases}$$

$$w_{ii}^* = \begin{cases} 1 & \lambda_i > \delta \\ 0 & \lambda_i \leq \delta \end{cases}$$

В новой форме представления оценки обладают всеми перечисленными недостатками, однако очевиден путь их совершенствования: использование в качестве значений весов w непрерывных функций от матрицы X (удобнее использовать функции от матриц, образующих сингулярное разложение X) и от некоторых дополнительных параметров, зависящих, вообще говоря, от y .

В качестве матрицы весов W может быть использована любая квадратная матрица, однако в дальнейшем, кроме специально оговоренных случаев, будем предполагать, что она диагональна. Ниже будет показано, что для некоторых критериев качества регрессионной модели, таких, как средний квадрат ошибки оценки, переход к недиагональной матрице к улучшению модели не приводит, так что это ограничение оправдано. Поэтому для упрощения обозначений вместо w_{ii} будем писать w_i .

К настоящему времени предложено большое число [40] различных смещенных оценок регрессионных моделей. Среди них следует отметить сжатые оценки, ридж-регрессию, регрессию на главные компоненты, метод характеристического корня, метод Марквардта (метод дробного ранга) и т. д. [25]. Будет показано, что все эти методы могут быть сведены к одной оценке, зависящей от n весов. Оценка эта имеет вид:

$$\hat{b}_w = C^T W \Lambda^{-1} S^T y \quad (12)$$

Матрицы W и Λ диагональны, а следовательно – перестановочны, так что оценка может быть также записана в виде:

$$\hat{b}_w = C^T \Lambda^{-1} W S^T y \quad (12')$$

Если же в качестве матрицы весов взята недиагональная матрица, эти две формы записи оценки могут различаться, однако для каждой матрицы W существует такая матрица \tilde{W} , что оценки (12) и (12')

совпадают.

Вообще, любая оценка регрессионного анализа соответствующим выбором W может быть приведена к виду (12). Для этого обозначим оценку, полученную некоторым методом через \hat{b}_R и воспользуемся полнотой ранга X . Ограничение это не столь существенно, поскольку в силу действия ошибок округления даже при неполноте ранга X все сингулярные значения будут отличны от нуля или же могут быть заменены малыми ненулевыми значениями, при чем внесенная ошибка не превысит вычислительной погрешности. Тогда

$$w_i^R = \begin{cases} \lambda_i \sum_j b_j^R c_{ji} / r_j & r_j \neq 0 \\ 0 & r_j = 0 \end{cases}$$

$$\hat{b}_R = C^T W_R \Lambda^{-1} S^T y$$

Следовательно, семейство оценок вида (12) включает в себя все возможные оценки регрессии, так что оценки, наилучшие среди этого семейства, есть наилучшие среди регрессионных оценок вообще. Практическая ценность такого приведения невысока, поскольку требуется предварительное оценивание коэффициентов модели, однако оно позволяет использовать некоторые критерии для оценок произвольного вида, а для ряда оценок возможно получить выражение для весов, не включающее в себя вектор оценок коэффициентов.

Теорема 1.

Оценки ридж-регрессии, регрессии на главные компоненты, Марквардта, сжатые (Джеймса-Стейна) и МНК могут быть представлены в виде W -оценок.

Доказательство.

Для ридж-регрессии (гребневой регрессии) выражение для весов

следует из:

$$\hat{b} = (X^T X + kI)^{-1} X^T y = C^T (\Lambda^2 + kI)^{-1} \Lambda S^T y = C^T ((\Lambda^2 + kI)^{-1} \Lambda^2) \Lambda^{-1} S^T y$$

Следовательно, $W = \Lambda^2 (\Lambda^2 + kI)^{-1}$

Регрессия на главные компоненты [4] сводится к W -оценкам выбором весов из множества $\{0, 1\}$. При этом выбор весов в зависимости от сингулярных значений X или, что эквивалентно, в зависимости от собственных значений $X^T X$, приводит к классической, предложенной Кендаллом, регрессии на главные компоненты; выбор весов в зависимости от абсолютной величины r или, что эквивалентно, от значимости гипотезы об отличии значения ρ от нуля, приводит к модифицированной оценке регрессии на главные компоненты; возможен также выбор на основе некоторого компромиссного критерия или же на основе содержательного анализа в интерактивном режиме.

Оценки Марквардта [1] (оценки дробного ранга) сводятся к W -оценкам выбором весов из множества $\{0, \theta, 1\}$, где $0 < \theta < 1$ и только одно из w_i равно θ . Различные варианты оценок Марквардта могут быть получены изменением способа назначения весов.

Сжатые оценки [6], имеющие вид $\hat{b} = \theta b_{\text{МНК}}$, сводятся к W -оценкам выбором $W = \theta I$.

Наконец, оценки МНК являются W -оценками при $W = I$, а тривиальные оценки $b \equiv 0$ – при $W = 0$. ■

Ниже будет рассмотрено сведение к W -оценкам оценок, полученных отбрасыванием некоторых регрессоров, иначе говоря, приравниванием нулю некоторых коэффициентов, и оценок, полученных при общих линейных ограничениях-равенствах. Для них будут даны прямые выражения. Для оценок же при ограничениях-неравенствах будет построен итеративный алгоритм, основанный на сингулярном

разложений.

Рассмотрим некоторые статистические свойства W -оценок.

$$E(\hat{b}_W) = E(C^T W \Lambda^{-1} S^T y) = E(C^T W \Lambda^{-1} S^T S A C \beta + C^T W \Lambda^{-1} S^T \varepsilon) = C^T W C \beta \neq \beta$$

$$E(\hat{y}_W) = E(X \hat{b}_W) = E(S W S^T y) = E(S W S^T S A C \beta + S W S^T \varepsilon) = S W A C \beta = S W S^T \bar{y}$$

$$D^2(\hat{b}_W) = E(C^T W \Lambda^{-1} S^T \varepsilon)(C^T W \Lambda^{-1} S^T \varepsilon)^T = \sigma^2 (C \Lambda^{-2} W C)$$

$$D^2(\hat{y}_W) = E(S W S^T \varepsilon)(S W S^T \varepsilon)^T = \sigma^2 (S W^2 S^T)$$

$$E(\hat{e}_W) = S(CI - W) A C \beta$$

$$D^2(\hat{e}_W) = E(S(CI - W) S^T \varepsilon)(S(CI - W) S^T \varepsilon)^T = \sigma^2 (S(CI - W))^2 S^T$$

Здесь $\hat{y}_W = X \hat{b}_W$, $e_W = y - \hat{y}_W$.

Оценки становятся, вообще говоря, смещенными, однако дисперсии их при $w_i < 1$ уменьшаются, что будет использовано для улучшения оценок. Семиинварианты распределения оценок коэффициентов и отклика модели выражаются, в предположении об одинаковости распределения элементов векторов ε , как:

$$\varkappa_i(\hat{b}_j^W) = \varkappa_i(\varepsilon) \sum_{k=1}^m \left(\sum_{l=1}^n c_{jl} \lambda_l^{-1} w_l s_{kl} \right)^i \quad (13)$$

$$\varkappa_i(\hat{y}_j^W) = \varkappa_i(\varepsilon) \sum_{k=1}^m \left(\sum_{l=1}^n s_{jl} w_l s_{kl} \right)^i \quad (14)$$

Следует, однако, оговорить, что все приведенные формулы справедливы лишь при условии, что матрица весов является детерминированной. Это возможно, ввиду того, что случайным в рассматриваемой модели является лишь вектор ε , тогда, когда W зависит лишь от детерминированной матрицы X , но не от случайного вектора y , или же не зависит ни от X , ни от y . Последнее, за исключением хорошо изученного МНК, для которого $W \equiv I$, практического интереса не представляет. Если же веса суть функции от вектора y , они случайны, так что приведенные выражения справедливы лишь приближенно. При необходимости характеристики могут быть получены численно, используя метод "складного ножа" и псевдозначения по

Тьюки [22]. Возможно также использование весов для управления распределением оценок (например, для повышения робастности оценок).

Аналогично (9) можно охарактеризовать влияние изменения отдельных переменных на оценки:

$$\frac{\partial b_i^W}{\partial \varepsilon_j} = \sum_{l=1}^n c_{il} w_l \lambda_1^{\varepsilon_l} s_{lj} \quad (15)$$

$$\frac{\partial y_i^W}{\partial \varepsilon_j} = \sum_{l=1}^n s_{il} w_l s_{lj} \quad (15')$$

Эти выражения также справедливы лишь при W , не зависящем от ε .

Было предложено вводить в регрессионную модель веса для учета неравноточности измерений [12], или с целью получения робастного алгоритма. При этом вектор y и матрица X домножаются на матрицу весов Ω слева. В зависимости от использования диагональной матрицы весов или же матрицы общего вида используется термин "взвешенный метод наименьших квадратов (ВМНК)" или "обобщенный метод наименьших квадратов (ОМНК)" соответственно. Следует подчеркнуть различие между W -оценками и ОМНК. Рассматривая оценку, как оператор над y , покажем, что справедлива

Теорема 2.

Оценка ОМНК и W -оценка совпадают в том и только в том случае, если они совпадают с оценкой МНК.

Доказательство.

Записав ОМНК-оценку в виде

$$\hat{b} = (X^T \Omega X)^{-1} X^T \Omega y$$

где Ω — матрица весов, положим, что она совпадает с W -оценкой

$$(X^T \Omega X)^{-1} X^T \Omega = C^T W \Lambda^{-1} S^T$$

Однако $(X^T \Omega X)^{-1} X^T \Omega = (C^T \Lambda S^T \Omega S \Lambda C)^{-1} C^T \Lambda S^T \Omega = C^T \Lambda^{-1} (S^T \Omega S)^{-1} S^T \Omega$. Тогда $C^T \Lambda^{-1} (S^T \Omega S)^{-1} S^T \Omega = C^T \Lambda^{-1} W S^T$ или $(S^T \Omega S)^{-1} S^T \Omega = W S^T$, а после домножения слева на S — $(S^T \Omega S)^{-1} S^T \Omega S = W S^T S$ или $W = I$. ■

Следовательно, эти два класса оценок "ортогональны", пересекаясь лишь в одной точке — обыкновенном методе наименьших квадратов. Этот факт открывает возможность совместного их применения, например, в условиях неравноточности наблюдений. При этом матрица X и вектор y преобразовываются домножением слева на матрицу $\Omega^{+1/2}$, и сингулярное разложению подлжит преобразованная матрица $\Omega^{+1/2} X$.

2. Оптимальный выбор весов.

Введем критерии оптимальности модели, исходя из целей, стоящих перед регрессионным анализом. Чаще всего целью регрессионного анализа является нахождение возможно более точной оценки коэффициентов, почти столь же часто – нахождение улучшенной, очищенной от случайной составляющей, оценки отклика, иногда же – нахождение оценки вектора остатков, иначе говоря, оценки возмущающего фактора ε .

При преюв из названных целей естественно характеризовать модель отклонением найденной оценки от истинного значения вектора коэффициентов. В качестве метрики будем использовать квадратичную метрику. Кроме того, в силу действия случайности мы не можем получить точную оценку отклонения, и поэтому ограничимся оцениванием его матожидания. Таким образом, в качестве критерия оптимальности модели будем использовать средний квадрат ошибки оценки коэффициентов, сокращенно обозначив его СКОК.

Если целью регрессионного анализа является оценивание вектора отклика, складывается впечатление, что наилучшие оценки дает метод наименьших квадратов, для которого вектор оценки отклика наиболее близок в смысле суммы квадратов отклонений наблюдаемому вектору \hat{y} . Однако ситуация, при которой интерес представляет наиболее точная аппроксимация наблюдаемых данных, представляет собой исключение. Гораздо чаще регрессионный анализ рассматривается, как метод, позволяющий "очистить" вектор отклика от влияния возмущающих факторов, например, при построении различного рода нормативов, в задаче восстановления сигналов, а также при поиске резко выделяющихся наблюдений [32]. Поэтому в качестве критерия будем употреблять не отклонение оценки отклика от наблюдаемых его

значений, а отклонение оценки отклика от его математического ожидания. Аналогично предыдущему критерию, используем квадратическую метрику и вместо неизвестного значения отклонения будем использовать оцениваемое математическое ожидание его. Такой критерий будем сокращенно именовать "средний квадрат ошибки отклика" или СКОО.

В случае, когда исследователя интересует оценка возмущающего фактора ε , в качестве критерия используется "средний квадрат ошибки оценки остатков", сокращенно СКООст. Очевидно, его поведение аналогично СКОО.

Предложенные критерии могут быть записаны, как:

$$\begin{aligned} E((\hat{b}-\beta)^T(\hat{b}-\beta)) & \quad \text{СКОК} \\ E((\hat{y}-X\beta)^T(\hat{y}-X\beta)) & \quad \text{СКОО} \\ E((\hat{e}-\varepsilon)^T(\hat{e}-\varepsilon)) & \quad \text{СКООст} \end{aligned}$$

Минимизируя их по w_j , можно получить оценки для $W_{\text{ОПТ}}$.

Теорема 3.

Минимум критериев СКОК, СКОО и СКООст достигается одновременно при

$$w_j = \frac{\rho_j^2}{\rho_j^2 + \sigma^2}$$

Доказательство:

Проведем выкладки для каждого из критериев отдельно.

а. СКОК.

$$\begin{aligned} E((\hat{b}_w - \beta)^T(\hat{b}_w - \beta)) &= E((C^T W \Lambda^{-1} S^T y - \beta)^T (C^T W \Lambda^{-1} S^T y - \beta)) = \\ &= E((C^T W \Lambda^{-1} S^T S A C \beta + C^T W \Lambda^{-1} S^T \varepsilon)^T (C^T W \Lambda^{-1} S^T S A C \beta + C^T W \Lambda^{-1} S^T \varepsilon)) = \\ &= E((C^T (W - I) C \beta + C^T W \Lambda^{-1} S^T \varepsilon)^T (C^T (W - I) C \beta + C^T W \Lambda^{-1} S^T \varepsilon)) = \\ &= \beta^T C^T (W - I)^2 C \beta + E(\varepsilon^T S \Lambda^{-1} W C C^T W \Lambda^{-1} S^T \varepsilon) = \end{aligned}$$

$$\beta^T C^T (W-I)^2 C \beta + \sigma^2 \operatorname{tr}(W^2 \Lambda^{-2})$$

В последнем равенстве использовано тождество:

$$\operatorname{tr}(AB) = \operatorname{tr}(BA)$$

Учитывая, что $\rho = \Lambda C \beta$, получим выражение для критерия:

$$\rho^T \Lambda^{-1} (W-I)^2 \Lambda^{-1} \rho + \sigma^2 \operatorname{tr}(W^2 \Lambda^{-2})$$

Минимизируя его по W , получим:

$$\min! L = \rho^T \Lambda^{-1} (W-I)^2 \Lambda^{-1} \rho + \sigma^2 \operatorname{tr}(W^2 \Lambda^{-2}) =$$

$$\sum_j \rho_j^2 \lambda_j^{-2} (w_j - 1)^2 + \sigma^2 \sum_j w_j^2 \lambda_j^{-2}$$

$$\frac{\partial L}{\partial w_j} = 2\rho_j^2 \lambda_j^{-2} (w_j - 1) + 2\sigma^2 \lambda_j^{-2} w_j = 0$$

$$w_j = \frac{\rho_j^2}{\rho_j^2 + \sigma^2} \quad (16)$$

д. СКОО.

$$E((\hat{y} - X\beta)^T (\hat{y} - X\beta)) = E((SWS^T y - X\beta)^T (SWS^T y - X\beta)) =$$

$$E((S(W-I)\Lambda C\beta + SWS^T \varepsilon)^T (S(W-I)\Lambda C\beta + SWS^T \varepsilon)) =$$

$$\rho^T (W-I)^2 \rho + \sigma^2 \operatorname{tr}(W^2)$$

$$\min! L = \rho^T (W-I)^2 \rho + \sigma^2 \operatorname{tr}(W^2) =$$

$$\sum_j \rho_j^2 (w_j - 1)^2 + \sigma^2 \sum_j w_j^2$$

$$\frac{\partial L}{\partial w_j} = 2\rho_j^2 (w_j - 1) + 2\sigma^2 w_j = 0$$

$$w_j = \frac{\rho_j^2}{\rho_j^2 + \sigma^2} \quad (16')$$

в. СКООст.

$$E((\hat{e} - \varepsilon)^T (\hat{e} - \varepsilon)) = E((SWS^T y - X\beta)^T (SWS^T y - X\beta)) =$$

$$E((S(W-I)\Lambda C\beta + SWS^T \varepsilon)^T (S(W-I)\Lambda C\beta + SWS^T \varepsilon)) =$$

$$\rho^T (W-I)^2 \rho + \sigma^2 \operatorname{tr}(W^2)$$

$$\min! L = \rho^T (W-I)^2 \rho + \sigma^2 \text{tr}(W^2) =$$

$$\sum_j \rho_j^2 (w_j - 1)^2 + \sigma^2 \sum_j w_j^2$$

$$\frac{\partial L}{\partial w_j} = 2\rho_j^2 (w_j - 1) + 2\sigma^2 w_j = 0$$

$$w_j = \frac{\rho_j^2}{\rho_j^2 + \sigma^2} \quad (16'')$$

Таким образом, минимум всех критериев достигается при одинаковых значениях матрицы W ■

Следовательно, существует матрица весов, на которой достигается минимум всех рассмотренных критериев, несмотря на то, что критерии эти часто рассматриваются, как противоречивые, и эта матрица отлична от единичной, так что существует метод, лучший, нежели МНК в смысле всех рассмотренных критериев. Более того, поскольку всякая регрессионная оценка может быть представлена, как W -оценка, оценка (16) является наилучшей из возможных в смысле рассмотренных критериев.

Однако оценка эта получена при условии, что матрица W диагональна. Покажем, что рассмотрение недиагональных матриц не улучшит значения критериев по сравнению с оценками, использующими лишь диагональные матрицы. Ограничимся рассмотрением критерия СКЖК, поскольку для СКЖЖ и СКЖЖст доказательство аналогично.

Теорема 4.

Для каждой недиагональной матрицы весов существует диагональная матрица, при которой достигается меньшее значение критерия СКЖК.

Доказательство:

Запишем выражение для СКЖК:

$$\Phi(W) = \rho^T \Lambda^{-1} (W-I)^2 \Lambda^{-1} \rho + \sigma^2 \text{tr}(W^2 \Lambda^{-2})$$

Если матрица W недиагональна, существует $w_{pq} \neq 0$ при $p \neq q$. Построим матрицу W^* такую, что

$$w_{ij}^* = \begin{cases} w_{ij} & i \neq p \vee j \neq q \\ 0 & i = p \wedge j = q \end{cases}$$

Тогда для матрицы W^* уменьшится как первое, так и второе слагаемое в выражении для Φ . ■

Из теоремы следует, что оптимум достигается лишь при диагональной матрице W . Кроме того, справедлива

Теорема 5

Диагональные элементы матрицы W должны удовлетворять неравенству

$$0 \leq w_i \leq 1.$$

Доказательство.

Если какой-либо из диагональных элементов отрицателен, его можно заменить на равный по абсолютной величине положительный, сокращая тем первое слагаемое в выражении для СКЖК (смещение); если же какой-либо из весов превышает единицу, его можно заменить на $(1-w_i)$, сокращая тем самым второе слагаемое (дисперсию оценки). При каждом из этих преобразований одно из слагаемых сокращается, второе же остается неизменным, так что повторяя эту процедуру, приходим к указанному неравенству. ■

Может показаться, что наличие выражения (16) делает излишним эту процедуру, поскольку полученные веса заведомо удовлетворяют неравенству, однако существуют процедуры оценивания, для которых выполнение неравенства не гарантируется, к примеру, при исключении регрессоров, и может оказаться полезной коррекция весов.

Из диагональности матрицы весов следует, в частности, что в обобщенных ридж-оценках [2] целесообразно рассматривать лишь

матрицы-параметры, перестановочные с матрицей $X^T X$.

$$\hat{b}_{GR} = (X^T X + P)^{-1} X^T y = (C^T \Lambda^2 C + P)^{-1} C^T \Lambda S^T y = C^T \Lambda^{-1} (I + \Lambda^{-1} C P C^T \Lambda^{-1})^{-1} S^T y = C^T \Lambda^{-1} W S^T y$$

где P — матрица-параметр обобщенных ридж-оценок и матрица весов определится выражением

$$W = (I + \Lambda^{-1} C P C^T \Lambda^{-1})^{-1}$$

и матрица W диагональна, лишь если она имеет общую систему собственных векторов с матрицей $X^T X$, то есть перестановочна с ней.

Все приведенные выше утверждения справедливы при условии, что критерием служит минимум ошибки оценки векторов коэффициентов, отклика или же остатков. Однако при выборе иного, не сводимого к названному критерию, может оказаться целесообразным применение недиагональных матриц весов, а также весов, лежащих вне интервала $[0, 1]$. Такая ситуация может возникнуть, например, в задачах прогнозирования, где регрессоры сами подлежат прогнозу и введение весов призвано повысить устойчивость прежде всего к ошибке прогнозирования значения регрессоров [18], или же если веса вводятся для управления распределением оценок, например, для повышения их робастности.

В дальнейшем, однако, будут рассматриваться лишь критерии СКЖК и СКОО. Критерий СКЖК ведет себя аналогично СКОО, а сфера задач, в которых основной интерес представляют регрессионные остатки, значительно уже, так что этот критерий в дальнейшем особо не рассматривается.

Причины, благодаря которым W -оценки оказываются лучше оценок МНК, становятся ясны из геометрических соображений. Прежде всего отметим, что столбцы X образуют аффинный базис некоторого подпространства \mathcal{X} , переход же от X к S есть переход к

ортогональному базису этого пространства. При этом векторы y и ε принадлежат m -мерному пространству, объемлющему подпространство \mathfrak{X} и представляющему прямую сумму пространств \mathfrak{X} и $\mathfrak{E} \perp \mathfrak{X}$, так что эти векторы можно представить в виде суммы двух векторов, принадлежащих указанным подпространствам. Вектор же \bar{y} принадлежит подпространству \mathfrak{X} . Если цель проведения регрессионного анализа состоит в получении наилучшего приближения к вектору \bar{y} , то очевидно, что МНК не способен достичь этой цели в должной степени, поскольку с его помощью возможно выделить лишь составляющую вектора ε , принадлежащую подпространству \mathfrak{E} ; проекцию же вектора ошибки, на подпространство \mathfrak{X} , МНК не способен отличить от вектора \bar{y} , и, следовательно, не может удалить. В то же время W -оценки, за счет введения весов, меньших единицы, для коэффициентов разложения вектора y по базису, образованному столбцами S , уменьшают, хотя и не полностью удаляют составляющую ε , лежащую в пространстве \mathfrak{X} . При этом воздействию весов подвергается и вектор \bar{y} , так что метод с необходимостью оказывается смещенным. Однако, используя для выбора весов отношение "сигнал-шум", выражаемое (16), при малом искажении \bar{y} существенно сокращаем дисперсию его оценки.

Если же цель регрессионного анализа состоит прежде всего в оценивании коэффициентов модели или оценивании вектора ошибки ε , то очевидно, что замена наблюдаемого значения y на улучшенную его оценку приведет к улучшению модели и в отношении критериев СКОК или СКОост.

Необходимо отметить, что смещенность оценки, при оценивании единственной модели не являющаяся существенным недостатком, может оказаться им, если необходимо оценивать большое число моделей, в особенности если оценки векторов коэффициентов или отклика,

полученные на различных выборках, подлежат усреднению.

Представляет интерес степень возможного улучшения W -оценки в сравнении с оценкой МНК. Вычислим СКЖК и СКЖО для оптимальных W -оценок (т.е. при выборе весов в соответствии с (16)) и для оценок МНК.

$$\begin{aligned}
 L_W^b &= E((\hat{b} - \beta)^T (\hat{b} - \beta)) = \sum_j \rho_j^2 \lambda_j^{-2} (w_j - 1)^2 + \sigma^2 \sum_j w_j^2 \lambda_j^{-2} = \\
 &= \sum \rho^2 \lambda^{-2} \frac{\sigma^4}{(\rho_j^2 + \sigma^2)^2} + \sigma \sum \lambda^{-2} \frac{\rho_j^2}{(\rho_j^2 + \sigma^2)^2} = \sigma^2 \sum \lambda^{-2} \frac{\rho_j^2}{(\rho_j^2 + \sigma^2)} \\
 L_{\text{МНК}}^b &= \sigma^2 \sum \lambda^{-2} \\
 L_{\text{МНК}}^b - L_W^b &= \sigma^4 \sum_j \lambda_j^{-2} (\rho_j^2 + \sigma^2)^{-1} \quad (17)
 \end{aligned}$$

Очевидно, что при наличии малых сингулярных значений преимущество W -оценок над оценками МНК возрастает и, если одно из сингулярных значений λ стремится к нулю, W -оценки становятся сколь угодно лучше оценок МНК. Причина этого в том, что возрастание смещения конечно, а дисперсия при $\lambda \rightarrow 0$ возрастает неограниченно.

Рассмотрим аналогичные результаты для оценки отклика:

$$\begin{aligned}
 L_W^b &= E((\hat{y} - \bar{y})^T (\hat{y} - \bar{y})) = \sum_j \rho_j^2 (w_j - 1)^2 + \sigma^2 \sum_j w_j^2 = \\
 &= \sum \rho^2 \frac{\sigma^4}{(\rho_j^2 + \sigma^2)^2} + \sigma \sum \frac{\rho_j^2}{(\rho_j^2 + \sigma^2)^2} = \sigma^2 \sum \frac{\rho_j^2}{(\rho_j^2 + \sigma^2)} \\
 L_{\text{МНК}}^b &= n\sigma^2 \\
 L_{\text{МНК}}^b - L_W^b &= \sigma^4 \sum_j (\rho_j^2 + \sigma^2)^{-1} \quad (18)
 \end{aligned}$$

Здесь также достигается существенное, хотя и не такое

значительное, как для коэффициентов, улучшение качества оценки. Поскольку в выражение для СКОО не входят сингулярные числа, изменение степени мультиколлинеарности не влияет на сравнительное качество оценок.

Состоятельность W -оценок имеет место во всех случаях, когда оквзываются состоятельными оценки МНК, поскольку для W -оценок ошибка приближения к вектору β меньше, чем для МНК. Следует, однако, оговорить, что состоятельность оценок, в которые вместо весов, вычисленных по формуле (16), входят некоторые их оценки, следует доказывать особо.

Веса могут вводиться не с целью сократить влияние возмущающего фактора, а с целью повысить ее устойчивость к погрешностям вычисления. Примером такого применения весов могут служить оценки b^+ и b^* , в форме, приведенной в первом параграфе. Чтобы выбрать в этом случае веса, заменим детерминированные, но неизвестные нам ошибки вычисления (округления) на случайные величины с известными параметрами распределения: нулевым маожиданием (строго говоря, это недостижимо при четном основании системы счисления и детерминированном округлении, однако может быть удобным приближением) и дисперсией, зависящей от способа представления чисел в ЭВМ и особенностей алгоритма. Положим, что с ошибкой вычисляются лишь λ_j , поскольку при наличии мультиколлинеарности наиболее существенна ошибка в λ , в отсутствие же мультиколлинеарности проблема повышения вычислительной устойчивости алгоритма не является главной. Тогда погрешность, обусловленная совместным действием весов, отличных от единицы, и ошибок вычисления, составит:

$$\hat{b}_{i,} - \hat{b} - C^T \Delta_A W \Lambda^{-2} S^T y + C^T \Lambda^{-1} (I - W) S^T y$$

Минимизируя ее по W , получим:

$$L = \|\hat{b}_w - \hat{b}\|^2 = r^T \Lambda^{-4} W^2 E(\Delta_\Lambda^2) r + r^T \Lambda^{-1} (I - W)^2 \Lambda^{-1} r = \\ \delta^2 r^T \Lambda^{-4} W^2 r + r^T \Lambda^{-1} (I - W)^2 \Lambda^{-1} r$$

Здесь $\delta^2 = E(\Delta_\Lambda^T \Delta)$

$$\frac{\partial L}{\partial w_j} = 2\delta^2 r_j^2 \lambda_j^{-4} w_j + 2r_j \lambda_j^{-2} (w_j - 1) = 0$$

$$w = \frac{\lambda_j^2}{\lambda_j^2 + \delta^2}$$

Полученная таким образом оценка зависит лишь от λ_j . Величина δ^2 должна быть оценена, исходя из особенностей ЭВМ и алгоритма.

3. Оценивание значений оптимальных весов.

Выражения (17) и (18), показывающие весьма высокое качество W -оценок, в действительности не применимы, поскольку для получения весов по формуле (16) необходимо знать ρ и σ^2 . Однако знание ρ эквивалентно знанию величины β . Следовательно, необходимо прежде вычисления весов оценить величины ρ и σ^2 и подставить эти оценки в (16), перейдя таким образом от оптимальных весов к субоптимальным. (Представляется возможным также, оценив величины ρ и σ^2 по дополнительной выборке или субъективно, использовать эту оценку для построения регрессии. Однако такой подход приводит к внесению субъективных искажений и вероятностные свойства его неясны. В дальнейшем рассматриваться он не будет.)

Для получения оценки величины σ^2 воспользуемся тем, что вектор $(I-SS^T)y=(I-SS^T)\varepsilon$ представляет собой проекцию не зависящего от ρ вектора ε на подпространство, также не зависящее от ρ , и, следовательно, также не зависит от ρ . Математическое ожидание квадрата его длины равно:

$$E(((I-SS^T)y)^T((I-SS^T)y))=E(\varepsilon^T(I-SS^T)\varepsilon)=\sigma^2 \text{tr}(I-SS^T)=(m-n) \sigma^2$$

(здесь используется идемпотентность матрицы $(I-SS^T)$).

Следовательно, несмещенная оценка σ^2 есть

$$\hat{\sigma}^2 = \frac{1}{m-n} y^T(I-SS^T)y = \frac{1}{m-n} (y^T y - r^T r) \quad (19)$$

Из независимости ее от ρ следует независимость ее от β .

В предположении нормальности распределения ε получим максимально правдоподобную оценку:

$$\hat{\sigma}_{\text{ММП}}^2 = \frac{1}{m} (y^T y - r^T r) \quad (19')$$

Наконец, можно рассмотреть оценку, для которой достигается

минимум среднего квадрата ошибки. Эта оценка имеет вид [12]:

$$\hat{\sigma}^2 = \frac{1}{m-n+2} (\mathbf{y}^T \mathbf{y} - \mathbf{r}^T \mathbf{r}) \quad (19'')$$

Все эти оценки весьма близки между собой, по крайней мере при условии, характерном для реальных задач регрессионного анализа:

$$m \gg 3n \wedge m \gg 10$$

В этом случае относительное различие между оценками составляет величину порядка m^{-2} .

В предположении нормальности вектора ε можно вычислить коэффициент вариации оценки дисперсии:

$$\frac{\sqrt{D^2(\hat{\sigma}^2)}}{E(\hat{\sigma}^2)} = \sqrt{\frac{3}{m-n}}$$

Можно сделать вывод, что замена точного значения σ^2 его оценкой и выбор в качестве оценки любой из приведенных не повлияет существенно на качество оценки регрессии. Вывод этот находит подтверждение в результатах имитационного эксперимента, описанного в главе 4, а также в результатах решения реальных задач.

Иначе обстоит дело с оцениванием вектора ρ . Хотя легко можно получить очевидную оценку $\hat{\rho} = \mathbf{r} = \mathbf{S}^T \mathbf{y}$, являющуюся несмещенной:

$$E(\mathbf{r}) = E(\mathbf{S}^T \mathbf{y}) = \Lambda \mathbf{C} \beta + \mathbf{S}^T E(\varepsilon) = \rho$$

обладающую конечной дисперсией при любых Λ :

$$E((\mathbf{r} - \rho)(\mathbf{r} - \rho)^T) = \sigma^2 \mathbf{I}$$

а при условии нормальности ε максимально правдоподобную, однако замена ρ на \mathbf{r} приводит к существенно ^ббольшим отклонениям от значений оптимальных весов, чем замена истинного значения дисперсии его оценкой. Оценки регрессии, полученные таким образом, оказываются все же состоятельными при тех же условиях, которые

влекут состоятельность оценок МНК [35]. (В последнее время свойство состоятельности не рассматривается уже, как одно из наиболее важных [26], однако оценки, не стремящиеся к истинным значениям при росте объема выборки, вряд ли заслуживают доверия). Для доказательства состоятельности W -оценок воспользуемся тем, что

$$p\lim_{m \rightarrow \infty}(ab) = p\lim_{m \rightarrow \infty}(a) \cdot p\lim_{m \rightarrow \infty}(b)$$

так что, показав, что $p\lim(w) = 1$, где $p\lim$ есть предел по вероятности, покажем состоятельность W -оценок для случаев, когда состоятельны оценки МНК. Условие Эйкера состоятельности оценок МНК требует $\lim_{m \rightarrow \infty} (\min \lambda) \rightarrow \infty$, следовательно, учитывая, что $r = AC\beta + S\epsilon^T$,

оно влечет рост r . Веса есть монотонно возрастающие функции $|w|$ и ограничены сверху единицей. Поэтому предел по вероятности весов при условии состоятельности оценок МНК есть единица, так что W -оценки также состоятельны. В то же время, поскольку средний квадрат ошибки для W -оценок меньше, чем для оценок МНК, возможна состоятельность W -оценок при отсутствии состоятельности МНК.

Рассмотрим несколько иной подход к построению оптимальных весов. Введем в рассмотрение вектор q , полученный из вектора r домножением на ортогональную матрицу V , $V^T V = I$. Тогда можно записать

$$q = Vr = V\rho + V\eta = \tau + \theta$$

Предположим, что выбрана такая матрица V , что $\tau = (\tau_1, 0, 0, \dots, 0)$. Тогда все, кроме первого, элементы вектора q обусловлены вектором ошибки θ , так что, заменяя их на 0, уменьшаем ошибку оценки вектора τ вектором $\hat{q} = (q_1, 0, 0, \dots, 0)$. Однако, в силу ортогональности матрицы V , ошибка оценки вектора ρ вектором $\hat{\tau}$ равна ошибке оценки вектора τ вектором \hat{q} . Таким образом, домножив вектор \hat{q} на V^T ,

получим улучшенную оценку вектора ρ

$$\hat{r} = V^T \hat{q}$$

Матрица V , однако, неизвестна, и ее знание эквивалентно знанию искомого вектора ρ . Кроме того, даже при известном ρ она определена неоднозначно. Поэтому будем искать эту матрицу, используя лишь значения вектора q . Очевидно, что в случае правильного выбора матрицы V элементы вектора q , начиная со второго, будут представлять собой независимые нормально распределенные случайные величины с нулевым матожиданием и дисперсией σ^2 . Вопрос о проверке принадлежности выборки заданному распределению - один из самых разработанных в математической статистике, и здесь можно использовать какой-либо из известных критериев, например, критерии типа Колмогорова-Смирнова или \tilde{F} . Критерии типа $\tilde{\chi}$ менее удобны, поскольку требуют группировки наблюдений, а размерность вектора q для большинства задач (в частности, для задач оценивания регрессии) невелика. Не теряя общности, положим, что элементы вектора $q_2 \dots q_n$ расположены в порядке возрастания. Взяв в качестве критерия сумму квадратов отклонений элементов вектора от матожиданий порядковых статистик нормального распределения, видим, что абсолютный минимум достигается при $q_i = E(\varepsilon_{(i-1)})$, где $\varepsilon_{(k)}$ есть k -тая порядковая статистика нормальной выборки из нормального распределения объемом $(n-1)$, $i=2, \dots, n$. При этом q_1 есть

$$\pm \sqrt{\sum_{i=1}^n r_i^2 - \sigma^2 \sum_{i=2}^n e^2_{(i-1)}}$$

где $e_{(k)}$ есть матожидание k -той порядковой статистики в выборке объема $(n-1)$. Если же выражение под корнем отрицательно, требуемой матрицы V не существует.

Чтобы избавиться от неоднозначности, связанной с произволом в

выборе матрицы V , усредним оценки $\hat{\Gamma}$ по всем матрицам V , переводящим r в q . Избегая явного интегрирования для взятия математического ожидания векторов, заметим, что вектор \hat{q} может быть представлен в виде суммы векторов $q_1 \perp q$ и $q_2 \parallel q$ и при усреднении по всем V вектор q_1 оказывается элиминируемым, поскольку все матрицы V переводят q в r , но при этом для каждой матрицы V^T переводящей q в r_1 найдется матрица \tilde{V}^T , переводящая q в $-r_1$. Следовательно, при усреднении остается лишь вектор q_2 , коллинеарный вектору q и отличающийся множителем

$$\alpha = \frac{q_1^T q}{q^T q} = \left(1 - \frac{\sigma^2 \sum e^2}{\sum r^2} \right)$$

Следовательно, оценка вектора ρ может быть получена умножением вектора r на множитель α . Оценка эта сходна с известной оценкой Джеймса-Стейна, но отличается от нее как способом получения, так и значением множителя.

Очевидно, используя лучшие, чем r , оценки ρ , можно получить лучшие, чем МНК, оценки вектора коэффициентов β . Для задачи регрессии, однако, доступна и дополнительная информация, так что возможно построение оценок лучших, чем оценки, не использующие каких-либо предположений о векторе ρ . Дополнительная информация может состоять либо в ограничениях на элементы искомого вектора (или на функции от них), либо в вероятностном распределении элементов этого вектора.

Последний подход обыкновенно критикуется за то, что предполагается случайная величина, заведомо детерминированная — вектор β . Однако, если коэффициенты модели в действительности и детерминированы, существует по крайней мере два механизма, обуславливающих случайность оцениваемых величин. Прежде всего,

выбор единиц измерения является, по-видимому, случайным и, во всяком случае, не связанным с действительными значениями коэффициентов модели. Затем выбор матрицы плана X также не связан со значениями коэффициентов и в ряде случаев является случайным. Естественно предполагать сферическую симметрию распределения компонентов проекции вектора y на столбцы матрицы S . Предполагая, что длина вектора y (зависящая от выбора единиц измерения и случайная в той степени, в какой этот выбор случаен) может быть аппроксимирована распределением χ^2 , видим, что вектор $\rho = \Lambda V \beta$ может быть аппроксимирован распределением $N(0, \Lambda^2 \delta^2)$.

Задачу оценивания вектора ρ можно разбить на серию подзадач следующим образом: упорядочим элементы вектора g в соответствии с возрастанием значений λ , а, следовательно, с убыванием весов (равных λ^{-2}) при ошибках оценки отдельных элементов в выражении для суммарной ошибки оценки вектора коэффициентов. Рассмотрим векторы, полученные из вектора g отбрасыванием его последних элементов, так что $r_1 = (r_1, 0, 0, \dots, 0)$, $r_2 = (r_1, r_2, 0, 0, \dots, 0) \dots r_n = g$. Для каждого из них может быть вычислен множитель, а для выражения в целом получена взвешенная сумма множителей

$$w_i = \frac{\sigma^2 \lambda_i^2}{\sigma^2 \lambda_i^2 + \sigma^2} \sum_{k \leq i} \lambda_k (\lambda_j^{-2} - \lambda_{j+1}^{-2}) \left[1 - \frac{\sigma^2 \sum_{k \leq i} e^2_{(k)}}{\sum_{k \leq i} r_k^2} \right]$$

($\lambda_{n+1} = \infty$)

4. Исключение переменных посредством весов.

Регрессионные методы во всей их совокупности можно разбить на два класса: редуцирующие и полнопространственные. Первые уменьшают дисперсию оценки коэффициентов и отклика, а также вычислительную погрешность за счет уменьшения числа регрессоров в окончательной модели, вторые же, сохраняя все переменные, изменяют для этой цели лишь значения коэффициентов. К редуцирующим относятся всевозможные варианты отбора переменных: перебор всех регрессий, метод включения-исключения, ступенчатая регрессия и т.п. К классу полнопространственных оценок относятся прежде всего оценки МНК, ридж-регрессия, сжатые оценки. Регрессия на главные компоненты, вместе с оценками Марквардта (дробного ранга), занимает промежуточное положение между названными классами: оценки строятся в пространстве главных компонент, размерность которого меньше, чем у пространства регрессоров, однако затем осуществляется возврат в исходное пространство, так что в окончательную модель входят все регрессоры.

Хотя из (17) и (18) следует, что W -оценки, принадлежащие к классу полнопространственных, оказываются лучшими, как с точки зрения оценивания коэффициентов, так и с точки зрения оценивания значения отклика, необходимость применения редуцирующих оценок сохраняется. Существует по крайней мере три причины (точнее, три группы причин), по которым следует рассматривать, наряду с полнопространственными, оценки, основанные на исключении переменных:

-Статистические. Сокращение числа переменных, включенных в модель, повышает достоверность оценок коэффициентов при других переменных, а в случае использования W -оценок позволяет более

точно оценить веса.

-Экономические. Включение в модель большого числа регрессоров приводит к дополнительным затратам при использовании этой модели: необходимости сбора дополнительных показателей, установки дополнительных датчиков или измерения новых параметров. Следует, однако, отметить, что эти дополнительные затраты относятся к периоду практического использования модели, а не ее построения, поскольку для принятия решения об исключении переменной знание ее значений необходимо.

-Исследовательские. В ряде случаев основная цель построения регрессионных моделей состоит в разделении регрессоров на существенные и несущественные (значимые и незначимые). В этом случае редуцирующие методы, позволяющие утверждать, что модель может быть удовлетворительно представлена с использованием лишь подмножества регрессоров, могут быть лучше полнопространственных, для которых все регрессоры рассматриваются, как оказывающие влияние на отклик.

В связи с этим рассмотрим методы исключения переменных из регрессионных моделей. Эта задача состоит из двух подзадач: определения переменной (переменных), подлежащей исключению и пересчета коэффициентов модели после исключения переменной. В третьей главе будут рассмотрены использующие "скользящий экзамен" методы, которые могут быть применены как к оцениванию значимости моделей, так и к определению переменных, подлежащих исключению. Поэтому в настоящей главе рассматривается лишь вопрос о пересчете модели при исключении одной или нескольких переменных, поскольку эта задача может быть эффективно решена посредством сингулярного разложения и использования весов.

Задачу исключения из регрессионной модели переменной можно рассматривать, как задачу оценивания с ограничением, состоящим в том, что коэффициент при исключаемой переменной полагается равным нулю. Множитель Лагранжа, соответствующий этому ограничению, обозначим через μ . Рассмотрим вначале случай оценок МНК.

$$\begin{cases} \min! (y - X\hat{b})^T (y - X\hat{b}) = \sum_j r_j^2 (1 - w_j)^2 \\ \sum c_{ij} w_j \lambda_j^{i-1} r_j = 0 \end{cases}$$

где i - номер исключаемой переменной

$$L = \sum_j r_j^2 (1 - w_j)^2 + \mu \sum_j c_{ij} w_j \lambda_j^{i-1} r_j$$

$$\begin{cases} \frac{\partial L}{\partial w_j} = 2(w_j - 1) r_j^2 + \mu \sum_j c_{ij} w_j \lambda_j^{i-1} r_j = 0 \\ \frac{\partial L}{\partial \mu} = \sum_j c_{ij} w_j \lambda_j^{i-1} r_j = 0 \end{cases}$$

$$w_j = 1 - \frac{\mu}{2} c_{ij} \lambda_j^{i-1} r_j$$

$$\sum_j c_{ij} \lambda_j^{i-1} r_j = \frac{\mu}{2} \sum_j \hat{c}_{ij} \lambda_j^{2i-2}$$

$$\frac{\mu}{2} = \left(\sum_j c_{ij}^2 \lambda_j^{2i-2} \right)^{-1} \sum_j c_{ij} \lambda_j^{i-1} r_j$$

$$w_j = 1 - c_{ij} \lambda_j^{i-1} r_j \left(\sum_k \hat{c}_{ik} \lambda_k^{2i-2} \right)^{-1} \sum_k c_{ik} \lambda_k^{i-1} r_k \quad (21)$$

Переход от весов к коэффициентам модели и значениям оценки отклика описан выше, однако для проверки полученной модели (расчету критериев "скользящего экзамена", F-критерия, ожидаемого значения среднего квадрата ошибки оценки коэффициентов и отклика и т.п.) достаточно располагать лишь значениями весов. Легко получить аналогичные выражения для W-оценок при некоторой заданной матрице весов U:

$$\begin{cases} \min! (\mathbf{y}-\mathbf{X}\hat{\mathbf{b}})^T(\mathbf{y}-\mathbf{X}\hat{\mathbf{b}})=\sum_j r_j^2(u_j-w_j)^2 \\ \sum c_{ij}w_j\lambda_j^{\pm 1}r_j=0 \end{cases}$$

где i - номер исключаемой переменной

$$L=\sum_j r_j^2(u_j-w_j)^2+\mu \sum c_{ij}w_j\lambda_j^{-1}r_j$$

$$\begin{cases} \frac{\partial L}{\partial w_j} = 2(w_j-u_j) r_j^2+\mu \sum c_{ij}w_j \lambda_j^{-1}r_j=0 \\ \frac{\partial L}{\partial \mu} = \sum c_{ij}w_j\lambda_j^{-1}r_j=0 \end{cases}$$

$$w_j=u_j-\frac{\mu}{2} c_{ij}\lambda_j^{-1}r_j$$

$$\sum_j c_{ij}u_j\lambda_j^{\pm 1}r_j=-\frac{\mu}{2}\sum_j \tilde{c}_{ij}^{\pm 1}\lambda_j^{\pm 1}$$

$$\frac{\mu}{2} = \left(\sum_j c_{ij}^2 u_j \lambda_j^{-2}\right)^{-1} \sum_j c_{ij} \lambda_j^{-1} r_j$$

$$w_j=u_j-c_{ij}\lambda_j^{\pm 1}r_j \left(\sum_j \tilde{c}_{ij}^{\pm 1} u_j \lambda_j^{\pm 1}\right)^{\pm 1} \sum c_{ij}\lambda_j^{\pm 1}r_j \quad (21')$$

Стоимость пересчета модели, которую можно охарактеризовать числом операций ЭВМ, составляет $O(n)$ операций на одно исключение, что существенно меньше, чем $O(n^3)$ при полном перерасчете модели и $O(n^2)$ при использовании усовершенствованных схем типа "выметания"[27]. Особенно удобен такой подход при использовании "скользящего экзамена" или PRESS-статистик[1], поскольку для их расчета достаточно использовать лишь веса.

Аналогичные схемы расчета могут быть предложены для удаления более чем одной переменной. При этом стоимость пересчета модели может возрасти до $O(nk^3)$, где k - число удаляемых переменных, так что схему эту целесообразно применять лишь при малых k . Если же число удаляемых переменных велико, целесообразнее использовать редуцированную матрицу данных, проводя расчеты, как с новыми

данными. Обозначим множество исключаемых переменных M .

$$\begin{cases} \min! (y - X\hat{b})^T (y - X\hat{b}) = \sum_j r_j^2 (u_j - w_j)^2 \\ \sum_{i \in M} c_{ij} w_j \lambda_j^{i+1} r_j = 0 \end{cases}$$

$$L = \sum_j r_j^2 (u_j - w_j)^2 + \sum_{i \in M} \mu_i \sum_j c_{ij} w_j \lambda_j^{-1} r_j$$

$$\begin{cases} \frac{\partial L}{\partial w_j} = 2(w_j - u_j) r_j^2 + \sum_{i \in M} \mu_i \sum_j c_{ij} w_j \lambda_j^{-1} r_j = 0 \\ \frac{\partial L}{\partial \mu} = \sum_j c_{ij} w_j \lambda_j^{-1} r_j = 0 \end{cases}$$

$$w_j = u_j - \sum_{i \in M} \frac{\mu_i}{2} c_{ij} \lambda_j^{-1} r_j$$

$$\sum_j c_{ij} u_j \lambda_j^{i+1} r_j = \sum_{i \in M} \frac{\mu_i}{2} \sum_j \tilde{c}_{ij} \lambda_j^{2i} \quad i \in M \quad (22)$$

Решая эту систему относительно μ и переходя затем переходя к w_j , получим оценки при удалении множества регрессоров M . Если регрессоры удаляются последовательно, общий объем вычислений может быть сокращен до величины $O(nk)$.

Следует отметить, что для построения W -оценок необходима оценка дисперсии. Для ее получения целесообразно предварительно построить оценку МНК, и затем получить оценку дисперсии, как

$$\hat{\sigma}^2 = \frac{1}{m - n} (y^T y - r^T r)$$

Преимущества последовательного пересчета проявляются, если рассматривается поиск наилучшей модели "в ширину", а не "в глубину", то есть рассматриваются все модели с $(n-1)$ регрессоров прежде, чем перейти к моделям с $(n-2)$ регрессорами. Такой поиск более целесообразен при интерактивной схеме, когда на каждом шаге происходит обращение к исследователю, который, имея в своем

распоряжении информации об изменении критериев качества при удалении того или иного регрессора, принимает решение относительно удаления его.

Возможно построение нового класса регрессионных оценок, отчасти сходных по построению с оценками "характеристического корня"[38]. Для их получения расширим матрицу регрессоров столбком отклика, обозначив такую матрицу, как

$$Z=(X|y)$$

Очевидно, регрессия y на Z даст, за исключением случая полного отсутствия возмущений, т.е. $\varepsilon=0$, вектор коэффициентов вида

$$b=(0,0,\dots,1)^T$$

Однако, введя ограничение на коэффициент при $(n+1)$ -ом регрессоре (равном y) вида $b_{n+1}=0$, получим осмысленные оценки. Они могут быть выражены через (21) или (21').

В отсутствие предварительно заданных весов U оценки эти совпадают с оценками МНК, введение же весов дает оценки, близкие, но не совпадающие с W -оценками при $W=U$. Основанием для рассмотрения такого семейства оценок является тот факт, что оценки весов, основанные на векторе $r=AC\beta+S^T\varepsilon$, могут быть неустойчивы при наличии равных или близких по значению сингулярных чисел λ . При этом малые возмущения X приводят к столь же малым возмущениям Λ , однако возможны значительные изменения матриц C и S . Если веса зависят от r , возможны искажения значений W -оценок. (Недостаток этот, как, по-видимому следует из вычислительного эксперимента, не проявляется слишком сильно; возможная причина этого - малость изменения весов даже при существенных изменениях C и S). Лишены этого недостатка оценки, основанные на весах, не зависящих от r , однако класс их слишком узок. Переход же к расширенной матрице Z

частично закрепляет оси координат в пространстве, сокращая тем самым возможное изменение r вследствие вычислительных погрешностей.

5. Учет ограничений общего вида.

В предыдущем разделе была рассмотрена возможность учета ограничений-равенств простейшего вида $b_i = 0$. Однако иногда доступна информация о неравенствах или равенствах, которым должны удовлетворять коэффициенты или значения отклика в заданных точках. Этот случай также может быть рассмотрен с применением сингулярного разложения, почему и рассматривается в этой главе, хотя здесь и не будет применены веса. Рассмотрим вначале случай ограничений в виде равенств, связывающих коэффициенты. (Ограничения на значения y в заданных точках легко приводятся к указанному ниже виду).

$$\begin{cases} \min! (y - X\hat{b})^T (y - X\hat{b}) \\ F\hat{b} = g \end{cases}$$

или, вводя иную параметризацию регрессионной модели и обозначая через $\hat{r} = \Lambda \hat{b}$

$$\begin{cases} \min! (y - S\hat{r})^T (y - S\hat{r}) \\ F\Lambda^{-1}\hat{r} = g \end{cases}$$

$$(y - S\hat{r})^T (y - S\hat{r}) = ((I - SS^T)y + S(r - \hat{r}))^T ((I - SS^T)y + S(r - \hat{r})) = \\ y^T (I - SS^T) (I - SS^T)y + (r - \hat{r})^T S^T S (r - \hat{r}) = \text{const} + (r - \hat{r})^T (r - \hat{r})$$

Обозначая $p = r - \hat{r}$, $q = F\Lambda^{-1} S^T y - g$ и $H = F\Lambda^{-1} S^T$, можно записать задачу в виде:

$$\begin{cases} \min! p^T p \\ Hp = q \end{cases}$$

Таким образом, построение регрессионной модели при ограничениях в виде равенств сводится к решению недоопределенной системы уравнений относительно p . Наличие в выражении для H

множителя Λ^{-1} делает целесообразным использование для ее решения сингулярного разложения, а малая, по сравнению с размерностью X , размерность N , позволяет пренебречь возникающей в этом случае дополнительной вычислительной работой. Сингулярное разложение в этом случае может быть записано в виде:

$$N=UDV \quad U^T U=UU^T=I \quad V^T V=I$$

Отличия от (2) связаны с тем, что у матрицы N , в отличие от матрицы X , больше столбцов, нежели строк.

Используя оценку типа b^+ , получим

$$p=V^T D^+ U^T q$$

где

$$d_i^+ = \begin{cases} 0 & d_i = 0 \\ d_i^{-1} & d_i \neq 0 \end{cases}$$

и легко перейдем к оценке

$$\hat{b}=C^T \Lambda^{-1} (S^T y - V^T D^+ U^T q) \quad (23)$$

Известная оценка [1]

$$\hat{b}=b_{\text{МНК}} + (X^T X)^{-1} F (F (X^T X)^{-1} F^T)^{-1} (g - F b_{\text{МНК}})$$

эквивалентна оценке (23), однако двукратное обращение матриц, которые, вообще говоря, плохо обусловлены (в частности, ограничения-равенства иногда вводятся для сокращения дисперсии, вызванной плохой обусловленностью $X^T X$), с учетом же вычисления оценок МНК - трехкратное, значительно ухудшает вычислительные свойства этой оценки.

Наличие изометрии между пространствами W и B , выражаемой равенством

$$\sum_j (b_j - \hat{b}_j)^2 = \sum_j r^2 (w_j - \hat{w}_j)^2$$

позволяет легко перенести этот результат на случай W -оценок. Для

этого следует лишь заменить r на $\hat{r} - Wr$.

Следует отметить, однако, что полезность введения ограничений в виде равенств при построении регрессионных моделей сомнительна. При неполноте ранга матрицы X более целесообразно, по-видимому, сокращение числа регрессоров, информация о значениях коэффициентов или значениях отклика при определенных значениях регрессоров редко является точной; приближенную же информацию такого рода удобнее вводить добавлением псевдонаблюдений.

Априорная информация часто может быть представлена в виде неравенств. Задача оценивания в этом случае может быть сведена к общей задаче квадратического программирования. Если ограничения наложены лишь на знак коэффициентов, задача может быть сведена к оцениванию 2^k регрессий без ограничений [27]. Рассмотрим оценку, существенно использующую специфику регрессионного анализа и разработанную для важного частного случая, когда заданы интервалы, содержащие возможные значения регрессоров. (Рассмотрение ее в этой главе оправдано тем, что она также строится с использованием сингулярного разложения).

$$\begin{cases} \min! (y - X\hat{b})^T (y - X\hat{b}) \\ b_- \leq \hat{b} \leq b_+ \end{cases}$$

где b_- и b_+ есть вектора, составленные из действительных чисел или символов $+\infty$ и $-\infty$. В частном случае, когда заданы лишь ограничения на знак, соответствующее b_- равно нулю и $b_+ = +\infty$, или же $b_- = -\infty$, а $b_+ = 0$. Обозначая $d = b - \bar{C} \Lambda^1 r$, $d_- = b_- - \bar{C} \Lambda^1 r$, $d_+ = b_+ - \bar{C} \Lambda^1 r$, запишем

$$\begin{cases} \min! d^T C^T \Lambda^2 C d \\ d_- \leq d \leq d_+ \end{cases}$$

Целевая функция задачи квадратична и допустимое множество

выпукло, следовательно, задача имеет единственный минимум. Для получения его используем метод проекции градиента. Он обладает геометрической сходимостью, однако простота отдельных итераций и относительно малая размерность задачи позволяют решить ее за удовлетворительное время. Проектор на допустимое множество $P(d)$ имеет вид:

$$P_i(d) = \min(d_i^-, \max(d_i^+, d_i))$$

а последовательность шагов к экстремуму:

$$d_{k+1} = P(d_k - \gamma_k C^T \Lambda^2 C d_k) = P(C^T (I - \gamma_k \Lambda^2) C d_k) \quad (24)$$

При этом в качестве d_0 можно принять, например, $P(0)$. Введение ограничений-неравенств в W -оценки также не представляет трудности, если заменить в выражениях для \underline{d} , \underline{d} , d вектор r на вектор Wr , где W - матрица весов, определяющая соответствующий класс W -оценок.

Для проверки значимости полученной модели применение традиционных статистических критериев, вообще говоря, затруднено. Для этой цели может быть эффективно применена техника "скользящего экзамена".

Методы проверки предпосылок регрессионного анализа и качества моделей.

1. Скользящий экзамен.

Несмотря на наличие большого числа критериев для проверки значимости коэффициентов модели и проверки основных предположений регрессионного анализа [10,16], вопрос о соответствии модели реальности, а тем самым о применимости оценок этой модели, может быть окончательно решен лишь при получении новых, не использовавшихся при построении модели, данных. Модель может быть признана работоспособной, если она обнаруживает согласие с новой выборкой данных, в частности, позволяет предсказывать значения y для элементов новой выборки по соответствующим значениям x . Критерии же, использующие для проверки модели те же данные, на основе которых она строилась, могут проверять не столько ее соответствие реальности, сколько ее способность к подгонке, иначе говоря, они ориентированы не столько на генеральную совокупность, сколько на выборку.

Распространение же выводов, полученных на выборке, на генеральную совокупность невозможно без постулирования ряда предположений, многие из которых выполняются не всегда и при этом трудно проверяемы.

Примером этого может служить предположение о нормальности распределения возмущающего фактора ε , от которого зависит, в частности, применимость весьма часто используемого F-критерия Фишера. С одной стороны, нет оснований утверждать, что распределение ε будет всегда нормальным; напротив, отклонение от нормальности, то ли в форме выбросов (грубых ошибок), то ли в виде

отклонения семиинвариантов высшего порядка от нуля для всех элементов ε , скорее правило, чем исключение. С другой стороны, остатки, полученные при подгонке регрессионной модели, как показано во второй главе, распределены "более нормально" в том смысле, что нормированные семиинварианты порядка больше второго оказываются меньше, чем соответствующие величины для распределения возмущающего фактора ε . Следствием этого является сокрытие нарушения нормальности от исследователя и снижение достоверности выводов, делаемых на основе критерия Фишера. Другим примером может служить проверка значимости отличия от нуля коэффициентов регрессии при помощи критерия Стьюдента, для обоснованного применения которого требуется выполнение ряда не всегда выполняющихся и при этом трудно проверяемых условий.

В связи с этим неоднократно предлагался [31] подход, при котором выборка разбивалась на две подвыборки - обучающую и экзаменующую, причем модель, построенная на обучающей выборке, проверялась на экзаменующей. Характеристикой качества модели оказывалась, таким образом, ее прогностическая способность, проявляемая на экзаменующей выборке. При этом ошибки "прогноза" могут быть эффективно использованы для поиска выбросов в элементах вектора ε или для обнаружения ошибок спецификации (в частности, для обнаружения нелинейности связи).

Способ разбиения выборки на две, очевидно, влияет на полученные таким путем результаты. Способов таких может быть $2^m - 2$, где m - объем выборки. Хотя в работе [50] предложен способ использовать это многообразие вариантов для более точного оценивания распределений, однако требуемые затраты ресурсов (прежде всего машинного времени, затем памяти) превышают доступные

уже при умеренных размерах выборки, так что приходится ограничиваться малым числом из всех возможных разбиений. Еще одна проблема состоит в соотношении объемов обучающей и экзаменующей выборок. Малый объем обучающей выборки приводит к низкой точности полученных с ее помощью оценок, а иногда (скажем, если объем обучающей выборки меньше числа параметров регрессионной модели) и к невозможности их получения. В то же время для получения сколько-нибудь достоверных результатов касательно характеристик модели следует столь возможно увеличивать объем экзаменующей выборки.

Возможный выход из этого затруднения состоит в том, чтобы использовать обучающую выборку максимального возможного объема, при котором имеет еще смысл говорить об экзамене, а именно выборку объема $(n-1)$. При этом экзамен проводится лишь на одном наблюдении, но этот недостаток компенсируется тем, что такое разбиение проводится m раз таким образом, что экзаменующую выборку объема 1 будет поочередно образовывать каждое из наблюдений исходной выборки, остальные же $(n-1)$ будут образовывать для этой экзаменующей выборки обучающую.

Такой подход, разумеется, не способен в полной мере заменить проверку модели на новой выборке, однако, располагая лишь единственной выборкой, трудно получить лучший результат. Возможные улучшения подхода могут быть связаны с использованием всевозможных обучающих выборок объема $(n-k)$ при объеме экзаменующих выборок k . Разбиений такого рода существует C_n^k , и вопрос о том, компенсируется ли существенное увеличение вычислительной работы повышением точности оценок, остается открытым. Еще один подход связан с разбиением выборок на обучающую и экзаменующую случайным

образом, так что возможно применение вероятностно-статистических методов.

Непосредственная реализация такого подхода, носящего название "скользящего экзамена", связана с большим объемом вычислений, поскольку на построение одной модели требуется $O(m^2)$ операций, а на проведение всего скользящего экзамена $O(m^2 n^2)$, что превышает доступные ресурсы даже при умеренных m и n . Целесообразным оказывается применение такой схемы вычислений, которая позволит сократить объем вычислений за счет использования уже полученной информации, прежде всего сингулярного разложения X . Такая схема, требующая не более $O(n)$ операций на одно разбиение или не более $O(mn)$ на весь экзамен, приводится ниже.

Введем прежде всего некоторые обозначения. В соответствии с ранее условленным, y_i есть наблюдаемое значение i -того элемента вектора y , \hat{y}_i есть оцененное по полной выборке значение этого элемента, s_{ij} - элементы матрицы S , w_j - веса. В дополнение к этому обозначим $\hat{y}_{(i)}$, оценку i -того элемента вектора y , полученную без использования наблюдаемого значения этого элемента, иными словами, предмет рассмотрения - прогноз на единственный элемент экзаменующей выборки, соответствующей i -тому элементу.

будем рассматривать общий случай W -оценок, поскольку оценки МНК сводятся к ним выбором $W=I$. При этом:

$$\begin{aligned} \hat{y}_i &= \sum_j s_{ij} w_j r_j = \sum_j \sum_k s_{ij} w_j s_{kj} y_k = \\ & \sum_{k \neq i} \sum_j s_{ij} w_j s_{kj} y_k + \sum_j s_{ij}^2 w_j y_i \\ \hat{e}_i &= y_i - \hat{y}_i = \\ & (1 - \sum_j s_{ij}^2 w_j) y_i - \sum_{k \neq i} \sum_j s_{ij} w_j s_{kj} y_k \end{aligned}$$

Чтобы получить оценку $\hat{y}_{(i)}$, заменим в выражении для \hat{y}_i величину

y_i , по условию недоступную, на самую оцениваемую величину:

$$\hat{y}_{(i)} = \sum_{k \neq i} \sum_j s_{ij} w_j s_{kj} y_k + \sum_j s_{ij}^2 w_j \hat{y}_{(i)}$$

или, вынося $\hat{y}_{(i)}$ в левую часть:

$$\begin{aligned} (1 - \sum_j s_{ij}^2 w_j) \hat{y}_{(i)} &= \sum_{k \neq i} \sum_j s_{ij} w_j s_{kj} y_k = \\ &= \sum_i \sum_j s_{ij} w_j s_{kj} y_k - \sum_j s_{ij}^2 w_j y_i \end{aligned}$$

Обозначая "ошибку прогноза i -той экзаменующей выборки" через $e_{(i)}$, получим:

$$\begin{aligned} \hat{e}_{(i)} = y_i - \hat{y}_{(i)} &= \\ y_i - (1 - \sum_j s_{ij} w_j)^{-1} \sum_j \hat{s}_{ij} w_j y_i - (1 - \sum_j \hat{s}_{ij} w_j)^{-1} \sum_k \sum_j s_{ij} w_j s_{kj} y_k &= \\ (1 - \sum_j s_{ij}^2 w_j)^{-1} (y_i - \hat{y}_{(i)}) &= (1 - \sum_j s_{ij}^2 w_j)^{-1} e_i \end{aligned} \quad (25)$$

Полученная оценка не вполне совпадает с оценкой, получаемой полным пересчетом модели, прежде всего потому, что при полном пересчете производится также и исключение x_i , а в схеме (25) исключается лишь y_i . Однако разница составляет величину порядка m^2 , и при обычных значениях m ею можно пренебречь. Вычисление оценки вектора отклика является почти неизменным этапом построения регрессионной модели, и в этом случае объем работы составляет $O(n)$ в расчете на одно разбиение, или $O(mn)$ на весь экзамен. Если же вычисляется также и величина $\sum_j \hat{s}_{ij} w_j$, например, для оценки дисперсии прогноза, то дополнительная работа сокращается до величины порядка $O(m)$ на весь экзамен.

Ошибки прогноза (точнее, псевдопрогноза) $\hat{e}_{(i)}$ могут быть использованы для проверки наличия связи между регрессорами и откликом, для оценки значимости переменных, для исследования распределения остатков, в частности, для поиска выбросов и проверки на наличие гетероскедастичности. Для проверки значимости

модели в целом возможно применение PRESS-статистики Аллена[27]:

$$\sqrt{\frac{1}{m} \sum e_{(i)}^2} \quad (26)$$

Наиболее существенным ее достоинством является способность к обнаружению "сверхсглаживания" (*overfitting*), то есть включения в модель большего, чем необходимо, числа регрессоров. F-критерий Фишера обнаруживает это явление существенно хуже, а, например, такой критерий качества модели, как коэффициент множественной корреляции R, на включение излишних регрессоров реагирует увеличением. Вместе с (26) целесообразно вычислять статистику

$$1.25 \frac{1}{m} \sum |e_{(i)}| \quad (26')$$

которая более устойчива к выбросам. В их отсутствие эти статистики совпадают с точностью до выборочной погрешности. Существенное же различие этих статистик указывает на наличие выбросов.

Чрезвычайно важным является вопрос о линейности связей в модели. Оценки вектора ошибок e для проверки линейности не вполне подходят, поскольку вычислены таким образом, чтобы минимизировать отклонения \hat{y} от y , в том числе отклонения, вызванные неверной спецификацией вида зависимости. Величины же $\hat{e}_{(i)}$ оказываются для этой цели более пригодными. Для установления вида зависимости можно использовать графики остатков[11] или же воспользоваться методом проб и ошибок. В качестве же индикатора наличия нелинейной зависимости можно использовать различие между двумя оценками дисперсии, "внутренней":

$$\frac{1}{m-n} (y^T y - r^T r)$$

и "внешней", вычисленной исходя из ошибок прогноза. Для ее построения воспользуемся тем, что дисперсия прогноза есть:

$$\sigma^2 = \sigma_p^2 (1 + x_p (X^T X)^{-1} x_p)$$

где x_p — есть вектор значений регрессоров в точке прогноза. Учитывая же, что x_p есть x_i , получим:

$$\sigma_i^2 = \sigma^2 (1 + \sum_j s_{ij}^2 w_j)$$

Величины $h_i = \hat{e}_{(i)} / \sigma_i$ распределены с нулевым матожиданием и единичной дисперсией, однако не являются независимыми. Кроме того, они более чувствительны к выбросам, чем e_i , поэтому в качестве "внешней меры дисперсии" целесообразно использовать медиану абсолютных значений (домноженную на соответствующий множитель).

При близости двух оценок дисперсии можно сделать вывод об отсутствии существенных нелинейностей. Если же "внешняя" оценка заметно превышает внутреннюю, можно сделать вывод о неправильной спецификации формы связи. Случай, когда "внутренняя" оценка существенно превышает "внешнюю", может свидетельствовать об отсутствии в модели важных регрессоров.

Оценки Аллена, "внешняя" оценка дисперсии и т.п. могут быть использованы при выборе оптимального подмножества регрессоров совместно с формулами (21) и (22). При этом переменная рассматривается, как незначимая, если исключение ее приводит к уменьшению PRESS-статистики Аллена.

Еще одно возможное применение величин $\hat{e}_{(i)}$ связано с необходимостью обнаружения и оценивания корреляций между элементами вектора ε . Величины $\hat{e}_{(i)}$, как и e_i , можно представить в виде суммы двух слагаемых, причем первое зависит от ε_i , второе же представляет собой взвешенную сумму всех прочих элементов вектора ε . При этом в $\hat{e}_{(i)}$ значение ε_i входит с единичным множителем, а в e_i — с меньшим единицы и притом различным для различных i .

2. Поиск отклонений от нормальности и построение устойчивой к ним оценки.

Хотя при построении методов регрессионного анализа не предполагалась нормальность распределения вектора ε , и для их работоспособности нормальность ε не является необходимым условием, однако же при выполнении условия нормальности свойства оценок значительно улучшаются. С одной стороны, становятся применимы разнообразные статистические критерии, при построении которых использовалось предположение нормальности, с другой же стороны, может быть повышена точность оценивания. Этим и обусловлена целесообразность разработки методов обнаружения отклонений от нормальности. Следует отметить, что непосредственное применение известных критериев нормальности, разработанных для проверки гипотезы о распределении выборки, составленной из независимых наблюдений, к регрессионным остаткам, оказывается бесполезным, поскольку остатки эти зависимы и при этом распределение их ближе к нормальному, чем распределение ε . (Переход к независимым функциям остатков, наподобие BLUE- или BLUS-статистик [9], приводит к еще большей степени нормализации.)

В то же время, поскольку ненормальность распределения ε не всегда обнаруживается, но почти всегда присутствует, целесообразно рассмотреть методы построения устойчивых к отклонениям от нормальности (робастных к распределению) оценок. Выбор между поиском выбросов в сочетании с малоробастными оценками и оценками робастного типа зависит от выбранной спецификации ненормальности, в некоторых же случаях следует использовать как поиск выбросов, так и робастные оценки.

Спецификация ненормальности может быть введена различными

способами, а именно: признанием за всеми элементами ε одинакового, но отличного от нормального распределения, которое можно охарактеризовать его семиинвариантами, конечность которых предполагается; введением понятия выброса, то есть встречающихся с вероятностью $p \ll 1$ наблюдений, дисперсия которых значительно превосходит дисперсию "регулярных" наблюдений, при этом, как правило, распределение и выбросов, и регулярных наблюдений предполагается нормальным; введением различной дисперсии для разных наблюдений. Возможна и комбинация этих способов.

Первый способ спецификации приводит к построению алгоритмов робастного оценивания, второй - к разработке [39] методов выявления выбросов, третий же приводит к оценкам ОМНК и в настоящей работе подробно не рассматривается.

Для обоснования применения "скользящего экзамена" для поиска выбросов рассмотрим распределение остатков e_i и ошибок псевдопрогноза $e_{(i)}$, причем характеризовать эти распределения будем их семиинвариантами \varkappa_i и нормированными семиинвариантами $\bar{\varkappa}_i$, где i -порядок. Используем соотношения

$$\varkappa_i(\eta + \theta) = \varkappa_i(\eta) + \varkappa_i(\theta)$$

$$\varkappa_i(a\eta) = a^i \varkappa_i(\eta)$$

$$\varkappa_i(e_j) = \varkappa_i(\varepsilon) \left((1 - \sum_{j=1}^i s_{j1}^2 w_1)^i + \sum_{k \neq j} (\sum_{j=1}^i s_{j1} s_{k1} w_1)^i \right) \quad (27)$$

$$\bar{\varkappa}_i(e_j) = \varkappa_i(\varepsilon) \frac{\left((1 - \sum_{j=1}^i s_{j1}^2 w_1)^i + \sum_{k \neq j} (\sum_{j=1}^i s_{j1} s_{k1} w_1)^i \right)}{\left((1 - \sum_{j=1}^i s_{j1}^2 w_1)^2 + \sum_{k \neq j} (\sum_{j=1}^i s_{j1} s_{k1} w_1)^2 \right)^{1/2}} \quad (27')$$

$$\hat{e}_{(j)} = \varepsilon_i (\varepsilon) \left(1 + \frac{\sum_{k \neq j} (\sum_1 s_{j1} s_{k1} w_1)^i}{(1 - \sum_1 s_{j1}^2 w_1)^i} \right) \quad (28)$$

$$\bar{e}_{(j)} = \varepsilon_i (\varepsilon) \frac{(1 + \sum_{k \neq j} (\sum_1 s_{j1} s_{k1} w_1)^i / (1 - \sum_1 s_{j1}^2 w_1)^i)}{(1 + \sum_{k \neq j} (\sum_1 s_{j1} s_{k1} w_1)^2 / (1 - \sum_1 s_{j1}^2 w_1)^2)^{1/2}} \quad (28')$$

Из сравнения (27') и (28') видно, что нормированные семиинварианты ошибок скользящего экзамена ближе к соответствующим величинам для ε , чем нормированные семиинварианты остатков, и, что более важно, дисперсии их более стабильны. Следовательно, применение критериев нормальности к ошибкам скользящего экзамена принесет более достоверную информацию о распределении ε , чем применение тех же критериев к остаткам e . При этом, однако, и ошибки скользящего экзамена, так и остатки коррелированы.

В выражение для $\hat{e}_{(i)}$, величина ε_1 входит с множителем 1, а в выражение для $e_{(i)}$ - с множителем $(1 - \sum_1 s_{j1}^2 w_1)$. Следовательно, ошибки скользящего экзамена более информативны как для поиска выбросов, так и для обнаружения отклонения от нормальности распределения каждого ε_1 . В качестве индикатора выброса можно использовать отношение

$$t_i = e_{(i)} / \sigma(e_{(i)})$$

при нормальности ε имеющее распределение Стьюдента с $m-n$ степенями свободы.

При этом, однако, не следует автоматически отбрасывать наблюдения, удовлетворяющие критерию выброса, поскольку наличие "выброса" может в действительности свидетельствовать о неправильной спецификации формы связи (что особенно вероятно при расположении "выброса" вблизи границы изменения какого-либо из

регрессоров), ошибке измерения, регистрации или подготовки данных (что требует исправления, но не удаления), изменении формы связи с течением времени и т.п.

Кроме того, особо следует рассмотреть наблюдения, для которых дисперсия прогноза существенно превышает дисперсию прогноза для прочих наблюдений. Оказывая существенно большее влияние на регрессию, они должны составить "зону особого внимания" для исследователя. Значения x и особенно y в этих точках должны проверяться дополнительно. В качестве критерия для выделения таких точек могут быть использованы также значения частных производных по y . Разработка же для этой цели статистических критериев затруднена, поскольку матрица X предполагается детерминированной.

Проблема построения робастных оценок считается [42] "ортогональной" к проблеме мультиколлинеарности, однако из рассмотрения видно, что наличие мультиколлинеарности может существенно понизить устойчивость модели к возмущениям ε . Если проекция вектора выбросов на столбец S , соответствующий малому сингулярному числу, окажется достаточно большой, модель будет сильно искажена, что следует из (9).

При разработке алгоритма робастного оценивания для работы в условиях мультиколлинеарности целесообразно использовать сингулярное разложение X . Однако, в силу его высокой стоимости, желательно использовать его единожды. Поэтому оценки, основанные на итеративном взвешивании [23], оказываются неудобны, и за основу взят алгоритм [42].

Используем вместо квадратической функции потерь линейно-квадратическую, что соответствует использованию для малых остатков среднего, а для больших медианы, и запишем ее в виде:

$$\rho(x) = \begin{cases} 1/2 x^2 & |x| \leq c \\ c|x| - 1/2 c^2 & |x| > c \end{cases}$$

где c - параметр метода. Кроме того, введем функцию

$$\psi(x) = \frac{d\rho(x)}{dx} = \begin{cases} x & |x| \leq c \\ c \operatorname{sign}(x) & |x| > c \end{cases}$$

и вспомогательную, зависящую от c величину:

$$a = (m-n) \int_{-\infty}^{+\infty} \psi(x) \varphi(x) dx$$

где $\varphi(x)$ есть плотность нормального распределения. Выбрав $1 \leq c \leq 2$ и вычислив a , вычисляем в соответствии со следующим алгоритмом:

1. $X \rightarrow SAC$

2. $r = S^T y$ - вычисление

3. $b = C^T \Lambda^{-1} S^T y$ начальных

4. $\sigma^2 = (y^T y - r^T r) / (m-n)$ приближений

5. $e = y - Sr$ - вычисление остатков

6. $\sigma^2 = (\sum \psi^2(e_i / \sigma)) / (a\sigma^2)$ - вычисление псевдодисперсии

7. $z = \psi(e/\sigma)\sigma$ и псевдоостатков

8. $\Delta r = S^T z$ - уточнение вектора r

9. $r = r + \Delta r$

10. Если поправка меньше $\sigma \sum \lambda w$ для всех элементов r , $b = C^T \Lambda^{-1} W S^T y$ иначе повторить с шага 5.

Затраты на итерацию не превышают $O(mn)$, что по-видимому, компенсирует большее потребное число итераций по сравнению со схемой, использующей итеративное перевзвешивание, в которой затраты на итерацию возрастают до величины $O(m^2 n)$. Робастные оценки могут использоваться совместно с методами поиска выбросов, являясь начальным приближением. Возможно также использование всех рассмотренных выше подходов к выбору матрицы весов W . Использование W -оценок приводит к повышению устойчивости оценок

как к искажениям, вызванным отклонениями распределения ε от нормальности, так и к искажениям, вызванными робастными методами оценивания, поскольку они используют некоторую корректировку данных. Это относится как к M-оценкам, так и к L-оценкам. Что касается R-оценок, то распространение их на задачи многомерной регрессии, в частности, на построение W-оценок, весьма затруднительно и приводит к внесению иных искажений.

3. Случайные регрессоры.

Детерминированность матрицы регрессоров есть одно из основных предположений регрессионного анализа. Однако зачастую оно нарушается. В отдельных случаях это не приводит к существенным отклонениям. Например, если дисперсия ошибки мала, случайность может рассматриваться, как вычислительная погрешность, так что методы борьбы с вычислительной погрешностью, рассмотренные в главах 1 и 2, могут быть эффективно применены. Другим случаем, когда можно пренебречь случайностью матрицы X , является активный эксперимент, в котором контролируется математическое ожидание значений регрессоров, реализация же подвержена случайным отклонениям. Как показано в [27], в этом случае можно применять МНК, получая правильные результаты.

Рассмотрим, однако, случай, когда случайность существенна. Это имеет место, например, если математические ожидания регрессоров и отклика связаны линейно, доступны же реализации, отличающиеся от матожиданий на случайную величину [13]. Такая спецификация модели соответствует, в частности, случаю измерений с достаточно большой ошибкой. При этом случайное отклонение будем считать распределенным нормально с нулевым матожиданием и одинаковой, но неизвестной дисперсией. Случай, когда для различных регрессоров и отклика дисперсии различны, но известно, что они пропорциональны множителю α_1 , то есть

$$\sigma^2(x_i) = \alpha_1 \sigma^2$$

сводится к рассмотренному домножением столбцов X и вектора y на множитель $\alpha^{1/2}$.

Оценка максимального правдоподобия может быть получена при минимизации суммы квадратов отклонений от X и y . Предположим, что

в нашем распоряжении $k \geq 1$ откликов, линейно, хотя и различным образом связанных с X , причем дисперсии приведены к одной величине. Запишем задачу в виде:

$$y_{i1} + \eta_{i1} = a_{1k} (x_{ik} + \varepsilon_{ik})$$

причем $\sigma^2(\eta) = \sigma^2(\varepsilon) = \sigma^2$. Введем объединенную матрицу $Z = (X|y)$. Тогда оценка максимального правдоподобия может быть получена, как:

$$\min! \|E\|_2$$

$$Z = (X|y) = Z_R + E$$

$$\text{rang}(Z_R) = n$$

Условие на ранг матрицы Z_R получено, исходя из характера зависимости $E(y)$ от $E(X)$. Целевая же функция пропорциональна логарифму функции правдоподобия в предположении о нормальности ошибок. При этом полагаем, что размерность X есть $(m \times n)$, y - $(m \times k)$, Z - $(m \times (n+k))$. Матрицу E можно представить в виде суммы двух матриц E_S и E_N , причем столбцы E_S лежат в пространстве, натянутом на столбцы Z , а столбцы E_N ортогональны этому пространству. Следовательно,

$$\|E\|_2^2 = \|E_S\|_2^2 + \|E_N\|_2^2$$

При этом возмущение в плоскости, ортогональной Z , не может уменьшить ранга Z_R . Таким образом, в минимальной по норме матрице E можно положить $E_N \equiv 0$ и Z_R лежит в пространстве Z .

Лемма. Сумма квадратов элементов матрицы равна сумме квадратов ее сингулярных значений.

Доказательство.

$$\sum_{i,j} z_{ij}^2 = \text{tr}(Z^T Z) = \text{tr}(C^T \Lambda^2 C) = \text{tr}(\Lambda C C^T \Lambda) = \text{tr}(\Lambda^2) = \sum_j \lambda_j^2$$

Таким образом, задача сводится к минимизации суммы квадратов сингулярных значений E . Поскольку $E \perp Z_R$, в противном случае можно

было бы уменьшить норму E, задача может быть, с учетом приведенной леммы, записана в виде:

$$\begin{aligned} \min! \operatorname{tr}(\Lambda_{\mathbf{E}}^2) \\ S\Lambda C = S\Lambda_{\mathbf{R}} C + S\Lambda_{\mathbf{E}} C \\ \operatorname{rang}(\Lambda_{\mathbf{R}}^2) = n \end{aligned}$$

Полагая $\lambda_i \geq \lambda_{i+1} \geq 0$ при $i = \overline{1, n-1}$, получим λ , минимизирующее целевую функцию:

$$\lambda_i^{\mathbf{R}} = \begin{cases} \lambda_i & i = \overline{1, n} \\ 0 & i = \overline{n+1, n+k} \end{cases}$$

$$\lambda_i^{\mathbf{E}} = \begin{cases} 0 & i = \overline{1, n} \\ \lambda_i & i = \overline{n+1, n+k} \end{cases}$$

Полученная оценка в частном случае $k=1$ совпадает с оценкой ортогональной регрессии, являющейся оценкой максимального правдоподобия и состоятельность которой доказана. В случае же $k>1$ оценки проявляют сходство с оценками инструментальных переменных, однако пополняется здесь не множество регрессоров, а множество откликов.

Для оценки коэффициентов регрессии, составляющих уже не вектор, а матрицу B, разобьем C на квадранты:

$$C = \left(\begin{array}{c|c} C_{11} & C_{12} \\ \hline C_{21} & C_{22} \end{array} \right) \begin{matrix} n \\ k \end{matrix}$$

При этом справедливо

$$\begin{aligned} S\Lambda_{\mathbf{R}} &= X C_{11}^{-1} = Y C_{21}^{-1} \\ S\Lambda_{\mathbf{E}} &= X C_{12}^{-1} = Y C_{22}^{-1} \\ Y &= X B = X C_{12} C_{22}^{-1} = X C_{11}^{-1} C_{21} C_{22}^{-1} \\ \hat{B} &= C_{12} C_{22}^{-1} = C_{11}^{-1} C_{21} C_{22}^{-1} \end{aligned}$$

Выбор формул для вычисления \hat{Y} , \hat{X} и \hat{B} в точной арифметике

несущественен, однако следует иметь в виду, что, хотя матрица C и ортогональна, квадранты ее могут этим свойством не обладать, так что обращение их сопряжено с ошибками и выбором одной из двух формул можно ошибки уменьшить. Как правило, целесообразно выбирать формулу, требующую обращения матриц меньшей размерности. В оценку для B не вошла матрица Λ , подобно тому, как в оценку для u в классической модели регрессии она также не входит.

Применение подобной, многооткликковой модели возможно в различных ситуациях:

1. При наличии многих откликов, измеряемых одновременно, в частности, при многократных отягченных ошибкой измерениях отклика. Альтернативой в этом случае является усреднение различных измерений отклика.
2. При возможности рассмотрения, помимо отклика, переменных, не представляющих сами по себе интереса, но легко измеряемых и несомненно связанных с регрессорами (откликом). В этом случае модель аналогична модели инструментальных переменных.
3. Наличие двух равнозначных групп переменных, каждая из которых может рассматриваться в качестве многомерного отклика, когда другая – регрессор. В этом случае модель аналогична каноническим корреляциям.

4. Нелинейное оценивание и классификация с учителем.

Рассмотренные в предыдущих главах методы предназначены для оценивания зависимостей, линейных по параметрам. Это ограничение не столь обременительно, поскольку нелинейным преобразованием регрессоров и, возможно, отклика (последнее может, вообще говоря, приводить к изменениям в спецификации ошибки, например, к появлению гетероскедастичности) многие нелинейные модели [5] приводятся к линейному по параметрам виду. Другой подход к оцениванию нелинейных зависимостей состоит в использовании системы подходящих функций, таких, как полиномы, тригонометрические функции или сплайны, и построении линейной регрессии на них.

Первый из этих подходов не охватывает все возникающие в приложениях функции. Например, если часто используемая в приложениях функция Кобба-Дугласа вида

$$V = K^{\alpha} L^{\beta}$$

приводится к линейному виду логарифмированием, то функция вида

$$V = (\alpha K^p + (1-\alpha)L^p)^{1/p}$$

где α и p суть параметры, подлежащие оцениванию, по-видимому, не может быть приведена к виду, линейному по параметрам. Кроме того, при таком преобразовании могут возникнуть затруднения с интерпретацией.

Второй подход, охватывая все возможные виды зависимостей, если только выбрана достаточно полная система функций, в то же время слабо пригоден для экстраполяции полученных результатов, а интерпретация результатов оказывается в ряде случаев почти невозможна.

Поэтому задача оценивания существенно нелинейных зависимостей представляется важной с практической точки зрения. При этом, хотя

возможно решение ее посредством общих методов минимизации нелинейных функций, затраты ресурсов при этом весьма велики. Методы, использующие специфику регрессионных задач, существенно эффективнее в вычислительном отношении и позволяют использовать разработанные в математической статистике методы проверки значимости. Рассмотрим зависимость вида:

$$y=f(X, b)+\varepsilon$$

где y - отклик

f - нелинейная функция

X - регрессоры

b - вектор параметров

ε - аддитивная помеха

Для приведения модели к такому виду могут понадобиться предварительные преобразования, например, для перехода от мультипликативной помехи к аддитивной. Кроме того, преобразования могут способствовать упрощению зависимости, что может оказаться полезным как с точки зрения оценивания, так и интерпретации. Функция f предполагается известной с точностью до параметров b .

Полагая известным приближенное значение вектора параметров b_0 (оно может быть задано, исходя из содержательных соображений, получено каким-либо методом приближенной минимизации, например, поиском по сетке или же выбрано случайным образом, в последнем случае оценивание должно проводиться несколько раз), разложим функцию f в ряд Тейлора в окрестности точки b_0 . Ограничившись членами первого порядка, запишем

$$y=f(X, b_0)+\sum \frac{\partial f(X, b_0)}{\partial b_j} (b_j - b_j^0) + \varepsilon^*$$

где вектор ε^* включает в себя как случайную составляющую, так и

погрешность, вызванную заменой нелинейной зависимости линейным приближением, или, обозначая

$$v = y - f(X, b_0)$$

$$d = b - b_0$$

$$Z_{ij} = \frac{\partial f(x_i, b_0)}{\partial b_j}$$

$$v = Zd + \varepsilon^*$$

Полученная зависимость линейна по параметрам и ее возможно оценивать посредством линейных методов. Оцененный вектор d дает поправку к значению вектора оценок параметров, причем на следующей итерации строятся новая матрица Z и вектор v , по которым оценивается новое значение вектора поправок d . В качестве критерия сходимости может быть использована стабилизация нормы вектора ε^* при приближении к экстремуму. Используя для вычисления оценки d выражение вида $(Z^T Z)^{-1} Z^T v$, приходим к методу Гаусса-Ньютона. При этом число обусловленности матрицы Z , как правило, весьма велико, и ошибки округления могут препятствовать достижению экстремума. По этой причине предлагались [45] оценки вида

$$d = (Z^T Z + \alpha I)^{-1} Z^T v$$

или оценки, использующие для повышения устойчивости сингулярное разложение [17]. Очевидно, что усовершенствование метода оценивания в линейной подзадаче приводит к ускорению сходимости задачи в целом, так что применение W -оценок оказывается оправдано. Следует отметить, что, несмотря на применение к оцениванию линейных подзадач W -оценок, для нелинейной задачи по-прежнему строится оценка МНК. Опыт применения оценок такого типа [19] показывает, что они эффективны для функций различного класса, причем наиболее существенный недостаток метода Гаусса-Ньютона - потеря сходимости

при возрастании дисперсии вектора ε - существенно ослабевает. При этом следует иметь в виду, что вектор ε^* , фигурирующий в линейной подзадаче, включает в себя как вектор помех ε , так и погрешность аппроксимации функции f линейной зависимостью в окрестности b_0 . По этой причине целесообразно на первых итерациях, когда b_0 далеко от b_{opt} , при оценивании весов заменять оценку дисперсии на большую величину (возрастание оценки σ^2 приближает оценку метода Гаусса-Ньютона к оценке градиентного поиска).

Для проверки значимости полученных моделей могут быть применены разнообразные критерии математической статистики, например, критерий максимального правдоподобия. Применение же техники "скользящего экзамена" возможно либо к линейной подзадаче, полученной на последней итерации, либо, при наличии большого ресурса машинного времени, с уточнением модели, полученной исключением наблюдения, пересчетом матрицы Z и вектора v и проведением дополнительной итерации (при этом в качестве начального приближения используется оценка, полученная для линейной подзадачи).

Разработанные для оценивания регрессионных моделей методы могут быть эффективно применены и для оценивания моделей других классов. Очевидно применение рассмотренных в настоящей работе методов к дисперсионному и ковариационному анализу. Использование их в задачах классификации возможно, по крайней мере, в задаче классификации с учителем. При классификации на два класса связь с регрессионным анализом очевидна, если ввести характеристическую функцию, принимающую на объектах одного класса значение 1, на объектах другого же класса - значение 0 (разумеется, выбор множества значений $\{0,1\}$ произволен, существенно лишь их различие

между классами и равенство внутри классов). Тогда задача обучения классификации сведется к оцениванию регрессионной зависимости вида:

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

где y есть вектор значений характеристической функции. Применение W -оценок может оказаться целесообразным, особенно в условиях мультиколлинеарности. Возможно также применение техники "скользящего экзамена", развитой в главе 3. При этом, однако, следует иметь в виду, что распределение вектора ε вряд ли будет хотя бы приблизительно нормальным.

При разбиении на $k \geq 3$ классов "оцифровка" посредством множества значений $\{0, 1, \dots, k-1\}$ оказывается невозможной, и при этом следует использовать либо структуру дерева, строя иерархическую классификацию, либо строить k моделей, каждая из которых рассматривает принадлежность к i -тому классу, $i \in \{0, \dots, k-1\}$ против принадлежности ко всем остальным классам.

Связь между методами классификации и регрессией становится яснее, если рассмотреть соотношение между матрицами X и S . Евклидово расстояние между строками матрицы S совпадает с расстоянием Махаланобиса между строками матрицы X , предполагая, что строки матрицы X есть векторные случайные величины, имеющие одинаковую дисперсионную матрицу.

$$X = SA C$$

$$S = X C^T \Lambda^{-1}$$

$$(s_1 - s_2)(s_1 - s_2)^T = (x_1 C^T \Lambda^{-1} - x_2 C^T \Lambda^{-1})(x_1 C^T \Lambda^{-1} - x_2 C^T \Lambda^{-1})^T =$$

$$(x_1 - x_2) C^T \Lambda^{-2} C (x_1 - x_2)^T = (x_1 - x_2)(X^T X)^{-1}(x_1 - x_2)^T$$

Это дает возможность, используя сингулярное разложение, строить также и алгоритмы классификации без учителя. Кроме того, возможно

исследовать структуру матрицы X , выявляя близкие или резко различающиеся совокупности наблюдений. Такая возможность оказывается полезной, если наблюдения образуют заведомо неоднородную совокупность, причем в различных частях совокупности данных функция регрессии, возможно, различна. В подобном случае целесообразно построить классификацию наблюдений и проверить одинаковость регрессионных зависимостей для различных частей совокупности, используя для этого обычные статистические критерии или "скользящий экзамен".

1. Имитационный эксперимент для проверки работоспособности W-оценок.

Вопрос о пригодности того или иного статистического метода для решения практических задач может быть решен лишь на основе комплексного подхода, сочетания аналитических (как правило, носящих асимптотический характер и при этом ограниченных узкими рамками допущений) выводов, имитационного эксперимента, распространяющего аналитические выводы на случай малой выборки, иных законов распределения и т.п., с практическим опытом применения аттестуемого метода. Ни один из этих этапов, вообще говоря, не может быть опущен. Аналитические выводы позволяют справедливо судить о методе за пределами хотя бы и многочисленных частных случаев, имитационный эксперимент расширяет рамки предположений, сделанных при аналитическом исследовании метода, опыт же практического использования позволяет проверить, существуют ли в действительности условия, при которых применение испытуемого метода оправдано, при этом проверяются также и те условия, которые трудно не только формализовать, но и четко сформулировать. Ситуации, при которых можно судить о статистическом методе на основе аналитических рассуждений и практического опыта, достаточно редки, и для аттестации W-оценок необходимы все три этапа.

Необходимо оговорить здесь, что под имитационным

экспериментом понимается не событийное моделирование, а более широкий класс моделей, в которых воспроизводится внутренняя структура исследуемого объекта и взаимодействие его составных частей [14], причем учитывается случайный характер этого взаимодействия. В описываемом ниже эксперименте моделируется внутренняя структура алгоритма, предназначенного для построения W -оценок. При этом, хотя моделирующая программа располагает истинными значениями оцениваемых параметров, исследуемый алгоритм оперирует лишь с данными, доступными в реальных условиях исследователю, в данном случае вектором y и матрицей X ; истинные же значения параметров используются лишь для сравнения с оцениваемыми.

В процессе эксперимента генерируется вектор коэффициентов β и матрица регрессоров X . При этом матрица X генерируется таким образом, что ее число обусловленности равно заданной величине. Для этого генерируется матрица X_0 и вычисляется ее сингулярное разложение. Затем, исходя из диагональной матрицы сингулярных чисел Λ_0 , строится матрица Λ такая, что $\Lambda = \Lambda_0^\alpha$, где показатель степени выбран так, чтобы обусловленность матрицы Λ была равна заданной.

$X=SAC$

Затем генерируется вектор ε , при этом дисперсия его выбирается с тем расчетом, чтобы F -отношение равнялось бы заданному числу. Эксперимент проводится многократно на одинаковых матрицах X при различных ε . Цель такого повторения - прежде всего сократить время моделирования за счет исключения в части реализаций наиболее

трудоемкого этапа, каким является вычисление сингулярного разложения, и, кроме того, повысить сравнимость оценок в различных реализациях. Исходя из векторов β и ε и матрицы X , строится вектор отклика y .

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

Входом для испытуемого метода является матрица X и вектор y , по которым он строит оценку \hat{b} и истинного значения отклика $\bar{y} = X\beta$. Для сравнения ту же процедуру выполняем методом наименьших квадратов, ошибка оценки для которого принимается за единицу. Вычисляется средний квадрат ошибки оценки коэффициентов (СКОК) и средний квадрат ошибки оценки отклика (СКОО).

Поскольку в эксперименте вводится элемент случайности, выводы неизбежно носят статистический характер. Поэтому для каждого значения β и X эксперимент повторяется многократно. Число повторений на каждой точке равно 800, что позволяет снизить среднеквадратическую ошибку приблизительно в 28 раз. Для более точного сравнения оценок были применены методы снижения дисперсии. Как показали предварительные исследования, эффективными с учетом специфики задачи явились такие методы понижения дисперсии, как общие случайные величины и контрольные переменные. Первый из них, кроме того, чрезвычайно просто реализуется. Такие же методы, как дополняющие случайные величины и псевдовыборки, оказались менее эффективными и в окончательном варианте не использовались. Метод общих случайных величин состоит в том, что один и тот же вектор ε , используется для моделирования как контрольного, так и испытуемого методов – МНК и W -оценок. Он не только не требует дополнительных

затрат, но и приводит к некоторой экономии машинного времени за счет сокращения затрат на генерацию случайных чисел. Метод контрольных величин состоит в том, что для используемых случайных чисел рассчитываются некоторые характеристики, значения которых для генеральной совокупности известны [14]. В рассматриваемом эксперименте в этом качестве использовались следующие характеристики распределения: выборочное среднее, отклонение дисперсии от единичной, показатели асимметрии и эксцесса. Затем строилась регрессионная модель, связывающая оцениваемую величину, то есть относительную, в сравнении с МНК, ошибку W-оценок, с контрольными величинами:

$$L = a_0 + a_1 \theta_1 + a_2 \theta_2 + a_3 \theta_3 + a_4 \theta_4$$

При этом, в силу того, что в генеральной совокупности контрольные величины равны нулю, уточненное значение показателя качества дается свободным членом. Если совокупность контрольных величин полностью характеризует распределение ε , метод контрольных величин дает эффект, сходный с увеличением размера выборки до генеральной совокупности. Эффект, достигаемый при конечной совокупности контрольных величин, зависит от того, насколько хорошо они способны характеризовать распределение случайной величины. Метод же общих случайных величин, не сокращая влияние выборочных отклонений на испытываемые методы, делает это влияние одинаковым, так что эти два метода удачно дополняют друг друга.

Ограниченность ресурса машинного времени при необходимости многократного моделирования повлекла за собой упрощение моделирования за счет сокращения наиболее машиноемких этапов:

Прежде всего, как это было описано, сингулярного разложения, затем вычисления оценок коэффициентов и отклика в явном виде. При этом используется изометрия между пространствами \mathcal{X} и \mathcal{Z} и пространствами \mathcal{Y} и \mathcal{U} соответственно.

$$\|b - \beta\|_2^2 = (C^T \Lambda^{-1} \rho - C^T W \Lambda^{-1} r)^T (C^T \Lambda^{-1} \rho - C^T W \Lambda^{-1} r) = (\rho - W r)^T \Lambda^{-1} (\rho - W r)$$

$$\|y - \hat{y}\|_2^2 = (S \rho - S W r)^T (S \rho - S W r) = (\rho - W r)^T (\rho - W r)$$

Оценка для МНК получается при $W = I$.

При необходимости вычислять расстояние между истинными и оцененными векторами в метрике, отличной от l_2 , явное вычисление оценок необходимо, однако, если ограничиться лишь этой метрикой, такой подход сокращает требуемое число операций с $O(n^2)$ до $O(n)$.

Генерация случайных чисел производилась методом, основанным на центральной предельной теореме и использующем 12 различных линейных конгруэнтных генераторов, имеющих различные взаимно простые периоды. За счет этого коррелированность отдельных реализаций случайных чисел была снижена, а период увеличен до приблизительно 10^{57} . Генератор реализован на языке Ассемблера.

Выбор метода, основанного на центральной предельной теореме, для генерации нормально распределенных случайных чисел, был произведен из следующих соображений. Генераторы, использующие нелинейное преобразование, например, на основе полярного метода, сильно зависят от недостатков базового генератора равномерно распределенных случайных чисел. Хотя распределение должно в точности совпасть с нормальным, однако, с одной стороны, большие значения производных используемых в генераторах функций становятся усилителем недостатков базового генератора, с другой стороны,

использование приближенных формул для вычисления логарифма, корня квадратного и т. п. порождает вычислительную погрешность, что также ухудшает характеристики генератора. Кроме того, преобразования эти требуют больших затрат машинного времени.

В противоположность этому, для метода, основанного на ЦПТ, не требуется сложных вычислений. Если же взять для вычисления равномерно распределенных случайных чисел различные генераторы, то погрешности их реализации будут взаимно компенсировать друг друга. Периоды их целесообразно брать взаимно простыми с тем, чтобы, с одной стороны, увеличить период, равный в этом случае произведению периодов генераторов равномерно распределенных случайных чисел, с другой стороны, уменьшит корреляцию отдельных чисел. Для ускорения работы эти генераторы не вызываются, как подпрограммы, а включены в тело одной процедуры. Распределение суммы от 12 генераторов отлично от нормального, так что оно подвергается преобразованию вида

$$\eta = (((((0.029899776\eta^2 + 0.008355968)\eta^2 + 0.076543912)\eta^2 + 0.252408784)\eta^2 + 3.949846138)\eta/4$$

предложенному в [51]. Оно непрерывно, так что на него не распространяются возражения против нелинейных генераторов.

Проведенный эксперимент показал эффективность и пригодность к практическому применению W -оценок даже при замене оптимальных, выражаемых через ρ , весов на субоптимальные веса, выражаемые через вектор g . При этом, как и следовало ожидать, относительная эффективность W -оценок по сравнению с МНК была выше как при оценивании коэффициентов, так и при оценивании отклика, однако

преимущество W -оценок было существеннее при оценивании коэффициентов и возрастало при возрастании мультиколлинеарности. Несколько неожиданным оказалось возрастание относительной эффективности оценивания отклика с возрастанием числа обусловленности, что не следует из (18). В качестве объяснения этого эффекта можно принять факт косвенной зависимости ρ от λ

$$\rho = \lambda C \beta$$

Однако степень улучшения невысока.

С ростом дисперсии возмущающего фактора ε (иначе говоря, с уменьшением F -отношения) преимущество W -оценок над оценками МНК возрастало, причем это проявлялось как для СКЖК, так и для СКОО

Хотя в описываемом эксперименте не ставилась цель исследовать влияние на относительную эффективность степени отклонения распределения ε от нормального, использование характеристик распределения в качестве контрольных величин позволяет это влияние измерить. При этом следует иметь в виду, что распределение возмущающего фактора выбиралось нормальным, так что отклонения от нормальности связаны прежде всего с конечностью выборки и незначительны. В случае внесенной искусственно ненормальности результаты могли бы быть иными. В данном же случае можно сделать вывод, что асимметрия распределения приводит к снижению относительной эффективности W -оценки, наличие же положительного эксцесса несколько повышает ее относительную эффективность. Это согласуется с представлением об усилении воздействия ненормальности мультиколлинеарностью, так что оценки, устойчивые к мультиколлинеарности, оказываются и несколько более робастными.

Снижение же относительной эффективности при наличии асимметрии связано, по-видимому, с тем, что весовая функция является нечетной функцией по g .

Результаты эксперимента, приведенные в таблицах 5 и 6, позволяют утверждать, что рассмотренные методы оценивания регрессионных зависимостей заслуживают внимания. Применение их целесообразно прежде всего при наличии в данных сильной мультиколлинеарности, что более характерно для пассивного, нежели для активного эксперимента, особенно же часто - для экономических данных.

Таблица 5.
Ошибка оценки коэффициентов.

g, F	10	15	25	50
10	0.676	0.732	0.755	0.843
15	0.561	0.683	0.695	0.763
20	0.497	0.575	0.638	0.751
25	0.484	0.530	0.559	0.683
30	0.454	0.496	0.550	0.636
50	0.368	0.414	0.477	0.510

Таблица 6.
Ошибка оценки отклика.

g, F	10	15	25	50
10	0.868	0.900	0.934	0.967
15	0.822	0.870	0.912	0.947
20	0.809	0.851	0.883	0.937
25	0.804	0.837	0.867	0.926
30	0.803	0.826	0.864	0.918
50	0.764	0.799	0.831	0.868

2. Имитационный эксперимент для проверки работоспособности

V-оценок.

План вычислительного эксперимента в основных чертах сходен с таковым для проверки W-оценок и предполагает получение большого числа реализаций случайных векторов ρ и η , причем для каждого из них сравнивается оценка $\hat{\rho}$, полученная на основе вектора γ с применением различных методов оценивания, как то байесовых, сжатых (Джеймса-Стейна), оценок, Бхаттачариа-Стейна, традиционных, а также предлагаемых.

В приводимой ниже таблице собраны результаты эксперимента, проведенного на ЭВМ РС XT с вектором длиной 16, соотношением b и σ , приблизительно соответствующим $R=0.85$, и величинами λ , изменявшимися приблизительно как геометрическая прогрессия при $\lambda_{\max}/\lambda_{\min}=100$ (программа на языке Pascal прилагается). При этом ошибка испытуемого метода относилась к ошибке метода наименьших квадратов, принятого, таким образом, за единицу. В таблице приведены (относительные) среднеквадратические ошибки для методов: Джеймса-Стейна ; E-оценок, использующих множитель (5); байесовых оценок; W-оценок; оценок Джеймса-Стейна для регрессии; V-оценок . Кроме того, приведена бесполезная с практической точки зрения, но важная теоретически оптимальная оценка, служащая нижней границей ошибки. Она принята за 100%.

Метод	Коэффициенты	Отклик
"Оптимум"	0.12709 0%	0.51144 0%
Джеймс-Стейн	0.87193 85%	0.89287 78%
Б-С(5)	0.65627 60%	0.78288 56%
Байес	0.14258 2%	0.61481 21%
W-оценки	0.80733 78%	0.89022 78%
Дж-Ст регр	0.86333 84%	0.85378 70%
V-оценки	0.12877 0.2%	0.58335 15%

Таким образом, можно сделать вывод, что предложенная оценка является наилучшей из представленных и может быть использована для построения регрессионных моделей. При этом следует отметить, что с ростом длины оцениваемого вектора качество оценок Джеймса-Стейна, Бхаттачариа и V-оценок улучшается, в отличие от оценок ММП, байесовых и W-оценок.

3. Результаты приложения методов смещенного регрессионного анализа в медицинских задачах.

В настоящее время в медицине начинают все более широко применяться методы математической статистики, в том числе регрессионный анализ. Они находят применение как при решении научно-исследовательских задач, так и при создании прикладных систем, например, консультативно-диагностического характера. Применение методов регрессионного анализа в медицине осложняется зачастую наличием мультиколлинеарности, малостью выборки, отклонениями распределений ошибок от нормальности и т. п. Нами показано, что предложенные в настоящей работе методы могут быть эффективно применены для решения ряда научных и практических задач в гастроэнтерологии.

Поставленная в совместной работе с кафедрой морской медицины Одесского медицинского института задача прогнозирования течения язвенной болезни у моряков имеет как практическое, так и теоретическое значение. Практическое значение ее связано прежде всего с тем, что условия труда и быта моряков торгового флота в значительной степени способствуют развитию у них язвенной болезни желудка и двенадцатиперстной кишки, своевременное выявление которой затруднено из-за отсутствия на судах врачей-специалистов и диагностического оборудования, что приводит, с одной стороны, к

необходимости обращаться за помощью в инопортах, расходуя при этом валюту, с другой стороны, к обострению болезненных состояний в то время, когда оказание помощи затруднено. В настоящее время грузовые и пассажирские суда Черноморского пароходства оснащены судовыми компьютерами, используемыми преимущественно для решения экономических задач учетного типа, а также некоторых задач механики корабля. При этом загрузка компьютеров невелика, так что становится возможным создание систем диагностики и контроля состояния здоровья экипажа на их основе.

Такие системы имеются на рынке программного продукта, но стоимость их высока, притом, что возможности их ограничены, как правило, доврачебной помощью. Так, предлагаемая фирмой "Olivetti" система такого рода предназначена лишь для консультаций при оказании доврачебной помощи при травмах и неотложных состояниях.

Задачи, решаемые системой подобного рода, можно подразделить на несколько групп: прогнозирование факторов риска и развития болезни; выдача рекомендаций по методам лечения; прогнозирования хода заживления язвенного дефекта и некоторые другие. Прогнозирование хода заживления язвенного дефекта целесообразно применять, наряду с судовыми системами, также и на берегу, в лечебных учреждениях морского флота, которые в настоящее время также широко оснащаются персональными ЭВМ. Это позволяет, в

частности, сократить количество тяжелых для больного и связанных с риском травмы обследований, таких, как гастроскопия (фиброскопия), в то же время уменьшив срок пребывания больного в стационаре.

Система факторов, используемых при построении оценки скорости заживления язвы, включает в себя данные анамнеза и лабораторных обследований; сведения о роде работы больного, в частности, относится ли больной к береговому или же к плавсоставу, от чего зависит нагрузка на нервную и другие системы организма, пищевой режим и другие факторы; применявшихся методах лечения и лекарствах (холинолитики, H-2 блокаторы, большие и малые репаранты, седативные средства, антибиотики, аутогемотерапия, лазерная терапия, антациды и т.п.); геофизические параметры, такие, как магнитное поле Солнца и число Вольфа. Большое число включенных в модель факторов, многие из которых связаны между собой статистической зависимостью, при относительной малости обучающей выборки, весьма затруднили применение обычных алгоритмов регрессионного анализа ввиду возникшей мультиколлинеарности. Это делает целесообразным применение описанных выше алгоритмов регрессионного анализа, основанных на сингулярном разложении матриц и приводящих, вообще говоря, к смещенным оценкам.

Использование модели, включающей все предложенные для прогнозирования факторы, возможно лишь при использовании ЭВМ. В

связи с этим были построены две существенно отличающиеся по своему объему модели, первая из которых включает более двухсот факторов и не может быть использована в отсутствие ЭВМ. В связи с этим, а также с большим объемом ввода исходной информации ее использование целесообразно в научных исследованиях или же в сложных случаях, в качестве "советчика".

Большой объем вычислений для построения модели подобного рода затрудняет быстрое обновление коэффициентов в случае их "старения", которое может быть связано с изменением условий труда и быта моряков, сменой лекарственных препаратов и т.п. Возможно, однако, производить перерасчет коэффициентов модели с относительно малыми затратами вычислительных ресурсов. Поскольку в предложенных выше алгоритмах основные затраты связаны с вычислением сингулярного разложения, был разработан специальный алгоритм, на вход которого поступают компоненты сингулярного разложения исходной матрицы и добавляемая к матрице строка. Алгоритм этот основан на связи между сингулярным разложением матрицы X и спектральным разложением матрицы $X^T X$. При этом для вычисления спектрального разложения используется модификация метода вращений Якоби, использующая выбор максимального элемента, что, в случае если норма добавляемой строки, как это обычно случается, не превышает существенно нормы строк исходной матрицы, существенно

сокращает время вычисления по сравнению как с обычной реализацией метода Якоби, так и по сравнению с такими методами, как QR-разложение. При этом сохраняется высокая степень ортогональности собственных векторов при достаточной, хотя и несколько меньшей, чем при перевычислении сингулярного разложения непосредственно, точности вычисления сингулярных чисел.

Наряду с машинными методами решения представленных выше задач целесообразно использовать также представление моделей в виде расчетных таблиц или номограмм, что позволяет употреблять их в отсутствие ЭВМ, в частности, в большинстве береговых медицинских учреждений, а также на небольших судах торгового флота. При этом число включаемых в модель параметров необходимо ограничено, так что при построении модели оказывается необходимым использовать методы отбора параметров, описанные выше. Следует отметить, что отбор параметров полезен также и с точки зрения определения степени влияния отдельных факторов на скорость заживления язвенного дефекта.

Ниже представлены две расчетные таблицы, предназначенные для использования врачом-гастроэнтерологом при назначении больного на повторную эндоскопию. При их помощи удастся, с одной стороны, сократить срок нахождения больного на обследовании, не назначая при этом излишних процедур, неприятных и тяжелых для больного, с

Ожидаемый срок заживления язвы желудка

: Корпусный отдел : Прочие отделы желудка

Диаметр язвы	Служащие			Плав.состав			Служащие			Плав.состав		
	Группы крови			Группы крови			Группы крови			Группы крови		
	0; A	B	AB									
35.31	34	39	46	27	31	37	43	50	59	34	40	47
30	32	38	45	26	30	35	41	48	57	33	38	45
29	32	38	44	25	30	35	41	48	57	33	38	45
28	32	37	44	25	30	35	41	48	56	32	38	45
27	32	37	44	25	29	35	40	47	56	32	38	44
26	31	37	43	25	29	34	40	47	55	32	37	44
25	31	36	43	25	29	34	40	46	55	31	37	43
24	31	36	42	24	29	34	39	46	54	31	37	43
23	30	36	42	24	28	33	39	46	54	31	36	43
22	30	35	42	24	28	33	39	45	53	31	36	42
21	30	35	41	24	28	33	38	45	52	30	35	42
20	29	35	41	23	27	32	38	44	52	30	35	41
19	29	34	40	23	27	32	37	44	51	30	35	41
18	29	34	40	23	27	31	37	43	51	29	34	40
17	28	33	39	23	26	31	36	42	50	29	34	40
16	28	33	39	22	26	31	36	42	49	28	33	39
15	28	32	38	22	26	30	35	41	49	28	33	39
14	27	32	37	22	25	30	35	41	48	28	32	38
13	27	31	37	21	25	29	34	40	47	27	32	37
12	26	31	36	21	24	29	33	39	46	27	31	37
11	26	30	35	20	24	28	33	38	45	26	31	36
10	25	29	35	20	23	27	32	38	44	25	30	35

Ожидаемый срок заживления язвы двенадцатиперстной кишки

Возраст 40:49 лет : Иной возраст

Задняя ст. Иная лок. : Задняя ст. Иная лок.

Диаметр язвы	ТВС в		ТВС в :		ТВС в		ТВС в	
	анамнезе							
	Да	Нет	Да	Нет	Да	Нет	Да	Нет
35.31	42	32	39	30	45	34	42	32
30	41	31	38	29	43	33	40	30
29	40	31	37	28	43	33	40	30
28	40	30	37	28	43	32	39	30
27	40	30	37	28	42	32	39	30
26	39	30	36	28	42	32	39	29
25	39	30	36	27	41	31	38	29
24	39	29	36	27	41	31	38	29
23	38	29	35	27	41	31	38	29
22	38	29	35	27	40	31	37	28
21	37	28	35	26	40	30	37	28
20	37	28	34	26	39	30	36	28
19	37	28	34	26	39	29	36	27
18	36	27	33	25	38	29	35	27
17	36	27	33	25	38	29	35	27
16	35	27	32	25	37	28	34	26
15	35	26	32	24	37	28	34	26
14	34	26	31	24	36	27	33	25
13	33	25	31	23	35	27	33	25
12	33	25	30	23	35	26	32	24
11	32	24	30	22	34	26	31	24
10	31	24	29	22	33	25	31	23

другой стороны, выявить группу больных с аномальным течением болезни, что позволяет индивидуально уточнить тактику лечения, а также выявить факторы, не включенные в модель, но оказывающие влияние на течение язвенной болезни.

Для построения приведенных таблиц была использована модель вида

$$T = \exp(2.6762 + 0.2265 \cdot \ln D - 0.0605 \cdot x_1 + 0.2757 \cdot x_2 + 0.0782 \cdot x_3)$$

где T - ожидаемый срок рубцевания язвенного дефекта двенадцатиперстной кишки, D - начальный диаметр язвы, переменные x_1, x_2, x_3 принимают значения 1 в случаях, когда больной принадлежит к определенному классу, а именно: x_1 для больных в возрасте 40-49 лет, x_2 для больных, перенесших ранее туберкулез, x_3 - для локализации язвы на задней стенке. Для язвы желудка модель имеет вид:

$$T = \exp(2.7075 + 0.2305 \cdot \ln D + 0.2307 \cdot x_1 + 0.1577 \cdot x_2 + 0.3207 \cdot x_3 - 0.2453 \cdot x_4)$$

где T, D имеют тот же смысл, что и в предыдущем выражении, а переменные x_i принимают значение 1, когда больной принадлежит к определенному классу, а именно: x_1 - для больных в возрастном интервале 40-49 лет, x_2 - для больных с группой крови B(II), x_3 для больных с группой крови AB(IV), x_4 при локализации язвы в corpus.

Разработана также программа, предназначенная для решения этой

задачи при использовании персонального компьютера, однако в связи с тем, что в настоящее время компьютеры в диагностическом центре Черноморского морского пароходства и Черноморской бассейновой клинической больницы только устанавливаются, ее эксплуатация производится лишь на кафедре морской медицины ОМИ и носит пробный характер.

Среди других задач, для решения которых используются предложенные в настоящей работе методы, следует отметить задачу выявления групп риска по язвенной болезни желудка и двенадцатиперстной кишки среди моряков и работников Черноморского морского пароходства, а также задачу прогнозирования периодических обострений язвенной болезни.

В первой из них требуется, исходя из анамнестических данных, а также из данных инструментальных исследований, выделить группу лиц, склонных к развитию язвенной болезни с целью проведения профилактических мероприятий. Дополнительная трудность, возникающая в этой задаче, связана с применением для ее решения некоторых новых инструментальных методик, например, доплерометрии для измерения скорости кровотока в желудке, при отсутствии нормативных значений этого параметра. Применение для ее решения статистических методов кажется весьма целесообразным, однако применение регрессионного анализа затруднено тем, что зависимая

переменная принимает значения из множества $\{0,1\}$, что может рассматриваться, как ненормальность вектора помехи ε .

Задача прогнозирования периодических обострений также весьма важна для проведения профилактических мероприятий, однако ее решение в настоящее время затруднено из-за неточности регистрации момента обострения, что может рассматриваться со статистической точки зрения, как наличие в выборке грубых ошибок.

В связи с этим для обеих названных задач оказывается полезным применение описанных выше методов поиска выбросов посредством "скользящего экзамена", возможно, совместно с методами построения робастных моделей.

4. Результаты применения методов смещенного регрессионного анализа в экономических задачах.

К задачам экономического характера, решавшимся при помощи предложенных методов, относится задача прогнозирования прибыли инвалютной и прибыли в сов. рублях, полученной Черноморским морским пароходством. В этой задаче необходимость применения указанных методов была связана прежде всего с малостью выборки, в свою очередь обусловленной тем, что условия работы ЧМП с течением времени подвержены изменениям и данные за отдаленные периоды времени не могут быть включены в модель из-за их непредставительности, так что для прогнозирования реально были пригодны данные за 7 последних лет, то есть за 28 кварталов. Наряду с этим имела место также мультиколлинеарность, обусловленная тем, что используемые для прогноза регрессоры в действительности не представляли собой независимые переменные, а находились между собой в нелинейной статистической зависимости. При этих условиях применение более традиционных методов регрессионного анализа привело к неудаче. Так, программа регрессионного анализа, входящая в известный пакет SSP PL, завершалась аварийно, поскольку вычислительные ошибки превышали по порядку величины результат. Предложенные же в настоящей работе методы показали себя достаточно надежными и эффективными.

По результатам содержательного анализа были отобраны 16 показателей, оказывающие, по мнению экспертов, наибольшее влияние на прогнозируемый показатель. В их число входят коэффициент стояночного времени, коэффициент балластных пробегов, средняя грузоподъемность судов флота, количество груза, коэффициент использования грузоподъемности, количество судов, средняя эксплуатационная скорость, количество рейсов, средняя дальность перевозки одной тонны груза, средняя доходная ставка, средняя

продолжительность рейсооборота, коэффициент загрузки, среднесуточная чистая норма грузовых работ, среднесуточная валовая норма грузовых работ, средняя стоимость основных производственных фондов, производительность одной тонны грузоподъемности.

Полученная модель может быть применена различными способами, в том числе для анализа прибыли путем сравнения фактических и оцененных по модели значений этого показателя, в этом случае все названные показатели берутся из отчетов; для получения предварительных оценок полученной прибыли в условиях, когда приведенные выше показатели, относящиеся большей частью к оперативной отчетности, имеются в наличии, величина же прибыли, в связи с методикой ее расчета, еще неизвестна; для получения прогноза прибыли на длительный период времени, что требует получения прогноза регрессоров, в качестве которого могут быть использованы плановые их значения, экспертные оценки, прогноз методом экспоненциального сглаживания и т. п.

Наряду с моделью, включающей все приведенные регрессоры, рассматривались также модели, включающие меньшее их число, что, в связи с наличием статистической зависимости между ними, позволяет получать прогноз со средней погрешностью около 3 процентов, что существенно ниже погрешности экспертного прогноза, составляющего 7-9 процентов.

Примером экономической задачи иного характера, также связанной с деятельностью Черноморского морского пароходства, является задача о распределении круизных судов по регионам плавания. Необходимость применения математических методов для ее решения обусловлено прежде всего тем, что тарифы на круизные рейсы весьма существенно, и притом различно в различных регионах (Средиземное море, Восточное побережье США, Западное побережье США и т.п.), зависит от комфортности судов. В то же время данные исследований, проводимых западными судовладельцами для назначения тарифов, являются коммерческой тайной, а проведение такого рода исследований советскими судовладельцами затруднено по причинам как финансового, так и политического характера.

Возможный путь решения этой проблемы состоит в том, чтобы, располагая установленными западными судовладельцами тарифами на круизные рейсы в различных регионах и на различных судах, а также техническими характеристиками судов, построить модель, связывающую характеристики комфортности судов (площадь кают, площадь ресторанов в расчете на одного пассажира, вместимость кают, наличие бассейна и т.п.) и тарифы в зависимости от региона. Эта модель позволяет, зная соответствующие характеристики советских круизных судов, оценить величину тарифов по регионам, а, следовательно, доход от рейса (а зная расходы на рейс - также и

прибыль).

Полученная оценка дохода (прибыли), обычно скорректированная, исходя из содержительных соображений (например, в связи с занижением тарифов в целях завоевания рынка), может быть использована в качестве входных данных в задаче распределения судов по регионам.

К задачам экономического характера относится также задача прогнозирования себестоимости изделий электронной промышленности и задача оценивания эффективности алгоритмов календарного планирования, внедренные в учебный процесс кафедры АУП в рамках курсов "Экономика промышленности" и "Математические методы календарного планирования".

В первой из них предлагается, исходя из технических параметров изделий (устройств внешней памяти на магнитных носителях для ЭВМ серии СМ), оценить их себестоимость и затем сравнить ее с фактической. Целесообразность применения предложенных в настоящей работе методов вытекает из наличия мультиколлинеарности, вызванной согласованностью технических параметров изделий (например, емкость диска, Мбайт, и скорость обмена, Кбайт/сек) или одновременностью разработки изделий, так что тенденции технического прогресса приводят к одновременному изменению не связанных содержательно между собой характеристик

изделия (например, масса, кг, и наработка на отказ, часов; в данном случае имеет место статистическая зависимость с коэффициентов корреляции $R=-0.80$).

В задаче оценивания алгоритмов календарного планирования проводится имитационный эксперимент, в котором моделируется производственный процесс на машиностроительном предприятии, причем для разрешения возникающих в процессе оперативного планирования конфликтов (то есть для выбора работ, выполняемых на данной единице оборудования в случае, когда имеется две или более готовых к выполнению работ) используется какой-либо эвристический алгоритм (FIFO, LIFO и т.п.). Цель регрессионного анализа в данном случае - построить модель зависимости критериального параметра (например, коэффициента загрузки оборудования) от использованных алгоритмов и характеристик технологического процесса, играющих в данном случае роль мешающих параметров. Степень мультиколлинеарности в данном случае относительно невысока, и применение отличных от стандартных методов регрессионного анализа оправдывается прежде всего малостью выборки, увеличивающей дисперсию оценок параметров модели.

5. Применение смещенных регрессионных оценок
в задаче прогнозирования дисперсности катализатора.

Примером научно-технической задачи, решенной при помощи предложенных методов, может служить задача прогнозирования степени дисперсности катализатора. Прогнозируемая величина, как то следует из теоретических соображений, зависит от температуры и концентрации вещества известным с точностью до параметров образом. Однако величина эта может быть измерена лишь с систематической ошибкой, зависящей от метода измерения (с использованием H_2 , O_2 или CO), а также от неконтролируемых условий эксперимента (последняя величина может быть принята одинаковой для серии опытов, проводимых в одной лаборатории).

Использованная при построении модели выборка состоит из 236 данных измерения степени дисперсности, объединенных в 57 серий опытов. При этом к одной серии отнесены были опыты, проведенные в одной лаборатории в одно время одним и тем же методом. В качестве прогнозируемой переменной выступал натуральный логарифм дисперсности катализатора, а в качестве регрессоров были включены температура, произведение температуры на корень квадратный из концентрации, а также "качественные переменные", принимающие значение 1 при принадлежности катализатора или метода исследования к соответствующей группе или значение (температура) * (корень

квадратный из концентрации) при принадлежности к соответствующей группе серии опытов и значение 0 в противном случае.

Таким образом, размерность пространства параметров равна 68 (шесть металлов-катализаторов - рутений, родий, палладий, никель, иридий, платина, три метода измерения - H_2 , O_2 , CO, температура, $T \cdot \sqrt{C}$ и 57 переменных, указывающих на серии опытов). Высокая размерность пространства параметров, несмотря на относительно малую коррелированность их меж собой, также приводит к явлению мультиколлинеарности, что, в свою очередь, делает целесообразным применение приведенных выше методов построения регрессионных моделей.

При этом необходимо также произвести частичный отбор влияющих переменных, оставляя безусловно характеризующие метод, металл-катализатор, температуру и концентрацию, и производя селекцию среди качественных переменных, указывающих на серию опытов. В случае, если какая-либо из переменных "серия опыта" окажется значимой, это укажет на то, что условия в этой серии сильно отличались от типичных для всей выборке, и, следовательно, результаты этих опытов не могут быть использованы, по крайней мере, не могут быть использованы без введения поправочных коэффициентов. Оказалось, что подобных серий опытов выявилось 11; коэффициенты полученной модели могут быть использованы в качестве

поправочных коэффициентов для приведения условий опыта к типичным, однако до выявления причин, по которым эти опыты дали значимо отличные результаты, использование данных этих опытов даже и с поправочными коэффициентами представляется неосторожным.

В целом, однако, полученная модель оказалась достаточно хорошо точно представляющей данные опытов. Коэффициент корреляции в модели был равен 92,8%, что указывает на высокую объясняющую способность модели, но также и на необходимость включения в модель дополнительных параметров или уточнения результатов некоторых опытов.

В настоящее время разрабатывается пакет программ математической статистики StatArt, подсистема регрессионного анализа в котором основана на описанных выше методах. Это должно обеспечить высокую работоспособность пакета в условиях мультиколлинеарности, малости выборки, выбросов и т.п., что, наряду с широким использованием графического представления данных, удобством импорта файлов из таких распространенных баз данных, как dBase и Clipper, использованием для управления процессом работы системы меню и т.п. должно обеспечить высокую степень дружелюбности пакета к пользователю. Эти методы построения регрессионных моделей также используются в подсистеме анализа временных рядов указанного пакета.

Заключение

В настоящей работе рассмотрены некоторые новые подходы к построению регрессионных моделей. Они объединены с теоритической точки зрения отказом от требования несмещенности и переходом к более общему критерию минимума среднего квадрата ошибки, с точки зрения же реализации - использованием сингулярного разложения в качестве инструмента, упрощающего построение методов и повышающего их численную устойчивость. Следствием этого явились, с одной стороны, большая эффективность применения методов в условиях мультиколлинеарности [36], с другой же - существенное возрастание затрат машинного времени.

Результаты имитационного эксперимента, приведенные в главе 3, вместе с приведенными в главе 4 примерами практического применения предложенных оценок, показывают возможность и целесообразность их использования для решения реальных задач. При этом особенно эффективным оказывается использование диалогового режима.

Направление дальнейшего развития описанных методов связаны, во-первых, с построением лучших, нежели приведенные, оценок оптимальных весов, а также получением весов, оптимальных для критериев, отличных от рассмотренных в диссертационной работе, например, весов, оптимальных для задач прогнозирования или оптимальных при наличии корреляции между элементами вектора ε ; во-вторых, с созданием способов учета и включения в модель априорной информации, лучших, чем введение ограничений-равенств и ограничений-неравенств, в частности, используя теорию нечетких множеств; в третьих, с совершенствованием вычислительных процедур,

прежде всего процедуры сингулярного разложения, в том числе для случая разреженных матриц; в четвертых, с построением программных средств, ориентированных на пользователя с минимальным уровнем математической и тем более программистской подготовки, что особенно важно в условиях применения персональных ЭВМ. Кроме того, целесообразно более подробно рассмотреть связь между регрессионным анализом и классификацией, имея в виду как установление теоретического соответствия между методами, в частности, построение для задач классификации аналога W -оценок, так и построение процедур анализа данных, использующих эти методы.

Предложенные методы скорее дополняют, чем заменяют более традиционные методы, такие, как метод наименьших квадратов. В ряде случаев целесообразно строить как W -оценки, как и оценки МНК, тем более что их одновременное построение возможно с минимальными дополнительными затратами. Полученное таким образом семейство оценок позволяет более глубоко рассмотреть объект исследования. При этом для получения доверительных интервалов следует использовать методы, основанные на "скользящем экзамене".

Список литературы.

1. Айвазян С. А. , Енюков И. С. , Мешалкин Л. Д. Прикладная статистика. Исследование зависимостей. М. : Финансы и статистика, 1985.
2. Айвазян С. А. , Енюков И. С. , Мешалкин Л. Д. Прикладная статистика. Основы моделирования и первичная обработка данных М. : Финансы и статистика, 1983.
3. Алберт А. Регрессия, псевдоинверсия и рекуррентное оценивание. Пер. с англ. М. : , Наука, 1977.
4. Афифи А. , Эйзен С. Статистический анализ. Подход с применением ЭВМ. Пер. с англ. М. : Мир, 1982.
5. Бард И. Нелинейное оценивание параметров. Пер. с англ. М. : Статистика, 1979.
6. Вапник В. Н. Восстановление зависимостей по эмпирическим данным. М. : Наука, 1979.
7. Гантмахер Ф. Р. Теория матриц. М. : Наука, 1988.
8. Демиденко Е. Э. Линейная и нелинейная регрессии. М. : Статистика, 1981.
9. Джонстон. Эконометрические методы. Пер. с англ, М. : Статистика, 1980.
10. Дрейпер Н. , Смит Г. Прикладной регрессионный анализ. В 2 кн. Пер. с англ. М. : Финансы и статистика, кн.1 - 1986, кн.2 - 1987.
11. Езекиел М. , Фокс К. А. Метода анализа корреляций и регрессий. Пер. с англ. М. : Статистика, 1966.
12. Кендалл М. , Стьюарт А. Теория распределений. Пер. с англ. М. : Наука, 1966.

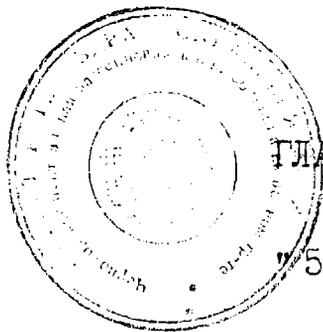
13. Кендалл М., Стьюарт А. Статистические выводы и связи. Пер. с англ. М.: Наука, 1973.
14. Клейнен Дж. Статистические методы в имитационном моделировании. Пер. с англ. М.: Статистика, 1978.
15. Крамер Г. Математические методы статистики. 2-е изд. Пер. с англ. М.: Мир, 1978.
16. Линник Ю.В. Метод наименьших квадратов и основы математико-статистической обработки наблюдений. М.: Физматгиз, 1958.
17. Лоусон И., Хенсон Р. Численное решение задач метода наименьших квадратов. Пер. с англ. М.: Наука, 1986.

22. Мостеллер Ф., Тьюки Дж. Анализ данных и регрессия. вып. 1, 2. М.: Финансы и статистика, 1982.
23. Мудров В.И., Кушко В.Л. Методы обработки измерений. М.: Советское радио, 1976.
24. Парлетт П. Симметрическая проблема собственных значений. Численные методы. М.: Мир, 1983.
25. Петрович М.Л. Регрессионный анализ и его математическое обеспечение на ЕС ЭВМ. М.: Финансы и статистика, 1982.
26. Рао С.Р. Линейные статистические методы и их применения. М.: Наука, 1968.
27. Себер Дж. Линейный регрессионный анализ. Пер. с англ. М.: Мир, 1980.
28. Уилкинсон Дж., Райнш С. Справочник алгоритмов на языке АЛГОЛ. Линейная алгебра. Пер. с англ. М.: Машиностроение, 1976.

29. Фаддеев Д.К., Фаддеева Е.Н. Вычислительные методы линейной алгебры. М.: Физматгиз, 1963.
30. Химмельблау Д. Анализ процессов статистическими методами. Пер. с англ. М.: Мир, 1973.
31. Atkinson A.C. Plots, Transformations and Regression. Oxford. 1985.
32. Belsley D.A., Kuh E., Welsch R.E. Regression diagnostics: Identifying influential data and source of collinearity. N.-Y. John Wiley and sons, 1980.
33. BMDP Biomedical Computer Programs. UCLA Press, 1979.
34. Daniel C., Wood F.S. Fitting equation to data. N.-Y. John Wiley and sons, 1971.
35. Eicker F. Asymptotic normality and consistency of the least squares estimations for families of linear regressions. Ann. Math. Stat., 1963, v. 34.
36. Farrar D.E., Glauber R.R. Multicollinearity in regression analysis. The Review of Economics and Statistics, 1967, v. 49, N°1.
37. Golub G. Reinsch C. Singular value decompositions and least squares solutions. Numer. Math., 1970, v. 14 N°4.
38. Gunst R.F., Webster I.I., Mason R.L. A comparison of least squares and latent root regression estimators. Technometrics, 1976, v. 18, N°1.
39. Hawkins D.M. Identification of outliers. L. 1980.
40. Hocking R.R., Speed F.M., Lynn M.J. A class of biased estimators in linear regression. Technometrics, 1976, v. 18, N°4.

41. Hoerl R., Shuenemeyer J., Hoerl A. A simulation of biased estimation and subset selection regression techniques. *Technometrics*, 1986, v. 28, N°4.
42. Huber P. *Robustness and designs*. Amsterdam: North Holland, 1975.
43. James W., Stein C. Estimation with quadratic loss. in: *Proceedings of the 4-th Berkeley Symp. Math. Stat. and Prob.*, 1961, v. 1.
44. Longley J.W. An appraisal of least-squares programs for the electronic computer from the point of view the user. *JASA*, 1967, v. 2, N°3.
45. Marquardt D.W. An algorithm for least squares estimation of nonlinear parameters. *Journal Soc. of Appl. Math.*, 1963, v. 2, N°4
46. Marquardt D.W., Snee R. Ridge regression in practice. *American statistician*, 1975, v. 29, N°2.
47. Siegel J.B. *Statistical Software for Microcomputers*. N.-Y., 1985.
48. Stewart G.W. *Collinearity, Scaling and Rounding Error in: Computer Science and Statistics*, Amsterdam, 1986.
49. Wetherill G.B. *Regression analysis with applications*. L. 1986.
50. Эфрон Б. Нетрадиционные методы многомерного статистического анализа. М.: Финансы и статистика, 1986.
51. Кнут Д.Е. *Искусство программирования для ЭВМ. Т.2. Получисленные алгоритмы*. М.: Мир, 1977.

П Р И Л О Ж Е Н И Е



"УТВЕРЖДАЮ"

ГЛАВНЫЙ ВРАЧ ЦЕНКБВТ

А. Шаши

"5" ноября 1990 г.

А К Т В Н Е Д Р Е Н И Я

1. Математическая модель построения нормативов и программа для подготовки расчетных таблиц сроков заживления язвенного дефекта желудка и двенадцатиперстной кишки.
/ наименование предложения для внедрения/
2. Одесский медицинский институт им. Н.И. Пирогова, кафедра морской медицины ФУВ, 270049, ул. Судостроительная, I, ассистент Сивый М.И.

Одесский политехнический институт, кафедра прикладной математики, 270044, г. Одесса, пр. Шевченко, I, асп. Машеров Е.Л.
/ учреждение-разработчик, его почтовый адрес, автор/
3. Нормативные материалы "Нормативы и расчетные таблицы сроков заживления язвенного дефекта желудка и двенадцатиперстной кишки".
4. Внедрено в гастроэнтерологическом отделении Черноморской центральной бассейновой клинической больницы на водном транспорте
5. Сроки внедрения: с "7" мая 1990 г. по "26" октября 1990 г.
6. Общее количество наблюдений: 412
7. Эффективность внедрения в соответствии с критериями, изложенными в источнике информации:

Показатели	По данным разработчиков	По данным Внедряющей организации
Сокращение сроков контрольной фиброскопии	7,2 дня	5,8 дней
8. Замечания, предложения

"5" ноября 1990 г.

Ответственный за внедрение
Зав. гастроэнтерологическим
отделением

Гринченко П.А.

Утверждаю
Проректор ОПИ
по учебно-методической
работе



" 12 " октября 1990 г.

С П Р А В К А

о внедрении диссертационной работы аспиранта кафедры прикладной математики Одесского политехнического института Машерова Е. Л. в учебный процесс института.

Результаты диссертационной работы аспиранта Машерова Е. Л. использованы при разработке программного обеспечения для проведения практических занятий и дипломного проектирования по дисциплинам "Экономика, организация и планирование производства" для студентов IV курса специальности 1738 "Организация механизированной обработки экономической информации" и в курсе "Математические методы оперативно-календарного планирования" для студентов V курса той же специальности.

Зав. кафедрой автоматизации управления
и планирования производства,
доктор экономических наук,
профессор

 В. М. Португал

Министерство морского флота
(наименование министерства, ведомства)
Черноморское морское пароходство
(наименование учреждения, организации, предприятия)

А К Т

15.05.86 № 14

г. Одесса

О внедрении научно-исследовательской работы **Разработка и приме-**

нение методов математической статистики для долгосрочного прогнозирования основных показателей работы пароходства

Основание: **хоз/** договор с Одесским политехническим институтом от **3 января 1983 г. № 812-28**
Составлен комиссией в составе:

Председатель

Члены комиссии:

Присутствовали:

Нач. отдела АСУ ЧМП Гайдученко И.А.

Нач. ЦЭО Кляуз О.Г.

Гл. бухгалтер Олин В.И.

Доц. кафедры АУИИ Востров Г.Н.

Ст. н.с. кафедры АУИИ Яромич С.А.

(должность, инициалы, фамилия)

В период с **1.12.85** по **15.12.85** комиссия провела работу по установлению фактического внедрения **комплекса программ: прогнозирования показателей работы пароходства**

(что конкретно внедрено, объем внедрения, место внедрения)

Работа по внедрению была предусмотрена планом в срок **декабрь**

Фактически внедрена **III кв. 84 - III кв. 85**

Годовой экономический эффект составил: фактический **130** тыс. руб.
ожидаемый _____ тыс. руб.

Акт составлен в трех экземплярах, прилагается к утвержденному расчету экономической эффективности, полученной от внедрения работы.

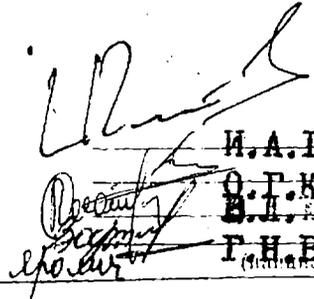
1-й экземпляр — **АСУ ЧМП**

2-й экземпляр — **ОПИ**

3-й экземпляр — **ОПИ**

Председатель комиссии

Члены комиссии:


И.А. Гайдученко
О.Г. Кляуз
В.И. Олин
Г.Н. Востров
С.А. Яромич

Расчет экономической эффективности находится в отделе АСУ ЧМП

ВЫПИСКА

из протокола заседания кафедры прикладной математики N
от 24 октября 1990 года.

СЛУШАЛИ: доклад аспиранта третьего года обучения
Машерова Е. Л. по завершенной диссертационной
работе и информацию научного руководителя
аспиранта доцента Вострова Г. Н.

ВЫСТУПИЛИ: Юхименко Б. И. , Сурков И. В. , Решта А. В. ,
Нечаев В. А. , Камалова И. Г.

ПОСТАНОВИЛИ: 1. Считать план третьего года аспирантуры
полностью выполненным.
2. Представить диссертацию в специализированный
Совет к защите.
3. Принять к сведению использование научных
результатов аспиранта Машерова Е. Л. в хозяйственной
научно-исследовательской работе N 812-28 и
считать долевое участие аспиранта в экономическом
эффекте в размере 40% или 52 тыс. рублей.

Зав. кафедрой прикладной
математики



Востров Г. Н.

Секретарь кафедры



Корнилова С. В.