

А. А. ГРЕШИЛОВ, В. А. СТАКУН, А. А. СТАКУН

## МАТЕМАТИЧЕСКИЕ МЕТОДЫ ПОСТРОЕНИЯ ПРОГНОЗОВ

МОСКВА «РАДИО И СВЯЗЬ» 1997

УДК 52+55; 35.073.5

Грешилов А. А., Стакун В. А., Стакун А. А. Математические методы построения прогнозов.— М.: Радио и связь, 1997.— 112 с.: ил.— ISBN 5-256-01352-1.

Рассмотрены методы анализа динамических (временных) рядов и построения прогнозов, в том числе методы оценки параметров моделей и диагностической проверки моделей; методы оценки ошибки прогнозов и подправления прогнозов. Рассмотрены интегральные и разностные схемы, методы сглаживания и сезонные ряды.

Книга рассчитана на специалистов, занимающихся задачами построения прогнозов, на студентов вузов и на слушателей системы дополнительного профессионального образования, изучающих методы прогнозирования.

Табл. 3. Ил. 8. Библиогр. 18 назв.

Г  $\frac{1601000000-007}{046(01)-97}$  Заказное издание

ISBN 5-256-01352-1

© Грешилов А. А., Стакун В. А.,  
Стакун А. А., 1997

Научное издание

Грешилов Анатолий Антонович, Стакун Виталия Анатольевна,  
Стакун Андрей Анатольевич

**МАТЕМАТИЧЕСКИЕ МЕТОДЫ ПОСТРОЕНИЯ ПРОГНОЗОВ.**

Редактор Грешилов А. А.  
Технический редактор Романова С. Ф.

**ИБ № 2775**

ЛР № 010164 от 29.01.97

Сдано в набор 10.12.96 Подписано в печать 26.02.97 Формат 60×84/16;  
Бумага офс. № 1 Гарнитура литературная Печать высокая Усл. печ. л. 6,51  
Усл. кр.-отт. 6,74 Уч.-изд. л. 6,35 Тираж 1000 экз. Изд. № 24047 Зак. № 86 С-007

Издательство «Радио и связь» 101000 Москва, Почтамт, а/я 693  
Типография издательства «Радио и связь». 101000 Москва, Почтамт, а/я 693

## ОГЛАВЛЕНИЕ

Введение . . . . .	4
<b>Глава 1. Методы регрессионного анализа как инструмент в построении прогнозов . . . . .</b>	<b>9</b>
1.1. Понятие регрессии; основные определения . . . . .	9
1.2. Линейная регрессия . . . . .	12
1.3. Регрессионный парадокс . . . . .	14
1.4. Ортогональная регрессия . . . . .	15
1.5. Оценка свободных параметров функций, линейных по параметрам, методом наименьших квадратов . . . . .	17
1.6. Оценка параметров моделей с помощью функции правдоподобия . . . . .	22
1.7. Байесовский подход к оцениванию параметров моделей . . . . .	28
1.8. Интервальная оценка линии регрессии и прогнозируемых значений функции . . . . .	29
<b>Глава 2. Конфлюентный анализ как универсальный метод учета неопределенностей в задачах прогнозирования . . . . .</b>	<b>31</b>
2.1. Активный и пассивный эксперименты. Постановка задачи оценивания параметров функции известного вида в пассивном эксперименте . . . . .	31
2.2. О единственности и состоятельности оценок в конфлюентном анализе. Алгоритм получения оценок . . . . .	36
2.3. Оценка параметров прямой линии и ее интервальных границ . . . . .	39
2.4. Оценка параметров линейных функций (и решение систем линейных алгебраических уравнений) . . . . .	42
2.5. Оценка значений параметров в сигноме . . . . .	45
2.6. Оценка параметров полиномиальной зависимости . . . . .	47
2.7. Непосредственное определение оценок параметров экспоненты и функции Гаусса . . . . .	49
2.8. Анализ систем в активном эксперименте . . . . .	51
2.9. Плохая обусловленность и некорректность в задачах оценки параметров функции известного вида . . . . .	55
<b>Глава 3. Методы построения прогнозов . . . . .</b>	<b>62</b>
3.1. Особенности процедуры прогнозирования . . . . .	62
3.2. Модели для получения прогнозов . . . . .	66
3.3. Анализ и сглаживание рядов с помощью скользящей средней . . . . .	70
3.4. Прогнозирование с помощью экспоненциального сглаживания . . . . .	74
3.5. Многофакторное прогнозирование . . . . .	78
3.6. Идентификация моделей типа АРСС . . . . .	82
3.7. Методы получения и уточнения прогнозов . . . . .	85
3.8. Байесовские прогнозы . . . . .	91
3.9. Анализ сезонных рядов . . . . .	93
3.10. Диагностическая проверка моделей и ошибка прогноза . . . . .	97
3.11. Обнаружение выбросов в наблюдениях . . . . .	103
Литература . . . . .	106

## ВВЕДЕНИЕ

Процедуры построения прогнозов используются практически во всех областях знания, в том числе в экономике, социологии, технике, образовании и т. д.

Процесс построения прогноза можно представить в виде двух взаимно связанных задач:

- 1) построение модели исследуемого явления;
- 2) оценка основных характеристик (параметров) модели по базовым данным и получение по этой модели интервальной оценки прогноза.

Как правило, эти задачи дополняют одна другую, процесс построения прогноза часто бывает итерационным — оценка параметров модели по базовым данным и полученные интервальные оценки прогноза могут послужить основанием для изменения исходной модели и последующего пересчета прогноза. В задаче расчета прогноза особое внимание должно быть уделено учету погрешностей исходных данных при оценке характеристик (параметров) модели и процедуре получения интервальных оценок прогнозов.

Первая задача базируется на физических законах, законах развития общества, а часто на интуиции, тогда как вторая задача может быть решена методами математической статистики. Таким образом, если из законов природы и общественного развития может быть получена модель явления, то основная проблема переносится на математические методы построения прогноза.

В процессе построения модели явления опираются, как правило, на наиболее устойчивые события. Например, наилучшее условие для построения модели и последующего прогноза — когда процесс или параметры модели постоянны. Поэтому при построении модели наряду с интегральными методами и элементарными функциями пользуются конечными разностями такого порядка, при которых процесс считается стационарным.

Системы называются стационарными, если их динамические свойства не изменяются с течением времени; если такое изменение имеет место, то системы называют нестационарными. Стационарность означает, что процесс преобразования входных возмущений системой обладает свойством инвариантности относительно сдвига во времени входных возмущений.

Модели бывают концептуальные, физические или математические (другие названия: феноменологические, эмпирические и аналитические) в зависимости от того, какая сторона явления в данном случае наиболее существенна, от методов, которые можно использовать при построении модели, от количества и качества имеющейся информации.

Относительная простота является главной характеристикой модели. Во многих случаях для того, чтобы модель была полезной, ее сложность должна находиться в определенном соотношении со сложностью описываемого объекта. Если физический механизм явления полностью ясен, можно вывести математическое выражение, точно описывающее явление. Во многих случаях для получения таких описаний (моделей) нужны подробные сведения, которых может не быть, и приходится прибегать к эмпирической модели. Оба названных случая представляют собой крайности. Обычно используемые модели занимают промежуточное положение. В частности, можно использовать неполные теоретические представления для указания подходящего класса математических функций, которые могут быть затем эмпирически подогнаны, когда число параметров модели и их числовые значения оцениваются по экспериментальным данным.

При подгонке моделей теоретический анализ не только указывает на подходящий вид модели, но может дать хорошие оценки числовых значений ее параметров. Эти значения затем можно проверить, анализируя реальные данные. Результат такой проверки служит в свою очередь основанием для пересмотра модели.

Итерационный подход к построению моделей включает в себя следующие этапы:

1. На основе теории и практики явления выбирается класс моделей, ориентируясь на те цели, для которых создается модель.

2. Разрабатываются простейшие методы идентификации подклассов этих моделей. Процесс идентификации может быть использован для получения грубых предварительных оценок параметров моделей.

3. Пробная модель подгоняется к экспериментальным данным; оцениваются ее параметры. Грубые оценки, полученные на этапе идентификации, теперь можно использовать как начальные значения в более точных итеративных методах оценивания параметров.

4. Диагностические проверки позволяют выявить возможные дефекты подгонки и диагностировать их причины. Если такие дефекты не выявлены, модель готова к использованию. Если обнаружено какое-либо несоответствие, итеративные циклы идентификации, оценок и диагностической проверки повторяются до тех пор, пока не будет найдено подходящее представление модели.

В практике широко применяются параметрические модели, что повысило интерес к задачам оценивания параметров и к родственным вопросам при построении таких моделей по экспериментальным данным.

При построении модели надо стремиться ответить на следующие вопросы:

- 1) как оценить качество модели;
- 2) как учесть всю имеющуюся информацию;
- 3) в чем состоит оптимальная стратегия получения недостающей информации;
- 4) как поступить с нелинейностями;
- 5) можно ли аппроксимировать сложную систему простой моделью.

Ответы на эти вопросы зависят от конкретного класса систем. На практике отыскание подходящей модели может быть достаточно трудной задачей.

Математическую модель системы называют детерминированной, если входящие в нее описания воздействия и параметры модели являются постоянными или детерминированными функциями переменных состояния и времени. Математическую модель системы называют статистической (стохастической), если функции, описывающие воздействия и параметры модели, являются случайными функциями или случайными величинами. Для стохастических (вероятностных) динамических систем текущее состояние в момент  $t_1$  —  $\mathbf{X}(t_1)$  и входное воздействие  $\omega = \omega(t_1, t_2)$  определяют в момент  $t_2$  не  $\mathbf{X}(t_2)$ , а лишь его вероятностное распределение.

Модели временных рядов и исследуемых процессов, необходимые для получения оптимального прогнозирования, в действительности являются стохастическими, поскольку на изучаемый процесс действует большое число неизвестных факторов и нельзя предложить детерминированную модель, допускающую точное вычисление будущего поведения объекта. Можно вычислить вероятность того, что некоторое будущее значение будет принадлежать определенному интервалу. В дальнейшем будем различать вероятностную модель или стохастический процесс и наблюдаемый временной (вариационный) ряд  $z_1, z_2, \dots, z_N$ , который рассматривается как выборочная реализация.

Важный класс стохастических процессов, рассматриваемых в приложениях, — стационарные процессы. Полезным инструментом для описания поведения стационарных процессов является автокорреляционная функция. Ковариация между значениями  $z_t$  и  $z_{t+k}$  одного вариационного (временного) ряда, отделенными  $k$  интервалами времени, называется автоковариацией с задержкой  $k$  и определяется как

$$\gamma_k = \text{cov}[z_t, z_{t+k}] = M[(z_t - \mu)(z_{t+k} - \mu)],$$

где  $\mu$  — среднее значение процесса.

Автокорреляция с задержкой  $k$  равна

$$\rho_k = \frac{M[(z_t - \mu)(z_{t+k} - \mu)]}{\sqrt{M[(z_t - \mu)^2]M[(z_{t+k} - \mu)^2]}} = \frac{M[(z_t - \mu)(z_{t+k} - \mu)]}{\sigma_z^2} = \frac{\gamma_k}{\gamma_0},$$

где  $\sigma_z^2 \equiv \gamma_0$ ;  $\gamma_k = \rho_k \sigma_z^2$ .

Функция  $\gamma_k$  от задержки  $k$  называется автоковариационной функцией  $\{\gamma_k\}$  стохастического процесса; функция  $\rho_k$  от задержки  $k$  называется автокорреляционной функцией  $\{\rho_k\}$  стохастического процесса.

На практике определяют только выборочные оценки этих функций. Так оценка  $\rho_k$

$$\hat{\rho}_k = \hat{\gamma}_k / \hat{\gamma}_0,$$

где  $\hat{\gamma}_k = \frac{1}{N} \sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z})$ ;  $k=0, \dots, m$ ,  $N$  — число наблюдений,  $\bar{z}$  — среднее значение ряда наблюдений.

Дисперсия выборочного коэффициента автокорреляции

$$D(\hat{\rho}_k) \approx \frac{1}{N} \sum_{v=-\infty}^{\infty} (\rho_v^2 + \rho_{v+k}\rho_{v-k} - 4\rho_k\rho_v\rho_{v-k} + 2\rho_v^2\rho_k^2).$$

Если  $\rho_k = \Phi^{|k|}$  ( $-1 < \Phi < 1$ ) — затухает экспоненциально, то

$$D(\hat{\rho}_k) \approx \frac{1}{N} \left[ \frac{(1+\Phi^2)(1-\Phi^{2k})}{1-\Phi^2} - 2k\Phi^{2k} \right].$$

Но многие встречающиеся на практике временные ряды имеют нестационарные характеристики. В книге рассмотрен класс нестационарных моделей, называемый моделями авторегрессии — интегрированного скользящего среднего (АРПСС).

В данной книге рассматриваются стохастические системы, для которых могут быть получены параметрические модели в виде элементарных функций, систем алгебраических уравнений и конечно-разностных описаний интегральных и дифференциальных уравнений. Предполагается, что исследователь располагает только случайными исходными данными: случайными будут входное воздействие и выход системы, или входное воздействие и параметры системы, случайны соответствующие элементы матриц алгебраических уравнений или ядер интегральных уравнений.

Оцениваемые параметры предполагаются либо неизвестными детерминированными величинами и в процессе решения находится доверительный интервал с определенной вероятностью накрывающий неизвестную величину (подход максимального правдоподобия)

или случайными величинами с априорно известным законом распределения (байесовский подход).

В первых главах книги описывается необходимый математический аппарат, применяемый при построении прогнозов, и рассматриваются особенности его применения. Затем излагаются собственно процедуры прогнозирования.

В первой главе рассматривается регрессионный анализ и метод наименьших квадратов как инструмент для построения прогнозов; подчеркивается, что погрешность аргумента должна быть много меньше, чем погрешность функции. Здесь же рассматривается «регрессионный парадокс» — явление связанное тем, что из-за наличия погрешности в наблюдаемых величинах, по одним и тем же исходным данным, меняя местами аргументы и функции (зависимые и независимые переменные), получим заметно различающиеся прогнозы.

Во второй главе даются методы математической статистики, которые позволяют учитывать в процессе построения прогнозов одновременно погрешности и в значениях функции, и в значениях аргументов — конфлюентный анализ. Эти алгоритмы исключают «регрессионный парадокс». Рассмотрены алгоритмы оценки свободных параметров для практически значимых в задачах построения прогнозов элементарных функций.

В третьей главе рассмотрены разностные методы построения прогнозов, экспоненциальное сглаживание и сглаживание с помощью скользящей средней, прогнозирование сезонных явлений; диагностическая проверка моделей и оценка ошибки прогноза. Приводятся алгоритмы подправления прогнозов.

Книга рассчитана на специалистов, занимающимися задачами построения прогнозов, на студентов вузов и на слушателей системы дополнительного профессионального образования, изучающих методы прогнозирования.

Эта книга вышла в свет благодаря поддержке и реальной помощи декана факультета ИБМ МГТУ им. Н. Э. Баумана д. т. н., профессора Садовской Т. Г. и директора МИПК при МГТУ им. Н. Э. Баумана к. т. н., доцента Ковальчука А. К. Авторы выражают им искреннюю благодарность.

# 1. МЕТОДЫ РЕГРЕССИОННОГО АНАЛИЗА КАК ИНСТРУМЕНТ В ПОСТРОЕНИИ ПРОГНОЗОВ

## 1.1. Понятие регрессии; основные определения

Для построения прогноза необходимо создать математическую модель изучаемого явления. Как правило, эта модель разрабатывается на основе проводящихся наблюдений над явлением, когда наблюдения не свободны от влияния всевозможных помех. Поэтому приходится решать одну из задач математической статистики: установить зависимость между наблюдаемыми случайными величинами.

Начнем с простейшего случая.

Рассмотрим две случайных величины  $\xi$  и  $\eta$ , которые имеют совместное распределение непрерывного типа с плотностью вероятности  $f(x, y)$ . В процессе наблюдений случайная величина  $\xi$  приняла значения  $x_i (i = \overline{1, n})$ , а случайная величина  $\eta - y_i (i = \overline{1, n})$ . Будем рассматривать  $\xi$  как **независимую** переменную. Тогда **фиксированному** значению  $\xi = x$  соответствует распределение вероятностей **зависимой** переменной  $\eta$  с плотностью вероятности

$$f(y | x) = \frac{f(x, y)}{f_1(x)} = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dy},$$

где  $f_1(x)$  — плотность вероятности частного распределения случайной величины  $\xi$ .

Для условного распределения вероятностей переменной  $\eta$  можно найти его числовые характеристики: среднее значение, моду, медиану и т. д., которые будут зависеть от  $x$ . Обозначим выбранную числовую характеристику как  $\bar{y}_x$ . Если  $x$  меняется, то точка  $(x, \bar{y}_x)$  опишет некоторую кривую (рис. 1). По виду этой кривой мы можем судить о расположении условного  $\eta$  — распределения для различных значений  $\xi$ . Кривая называется **кривой регрессии**, и говорят, что она изображает регрессию  $\eta$  на  $\xi$ . В дальнейшем

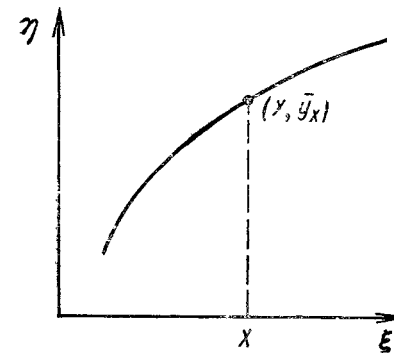


Рис. 1. Регрессия  $\eta$  на  $\xi$

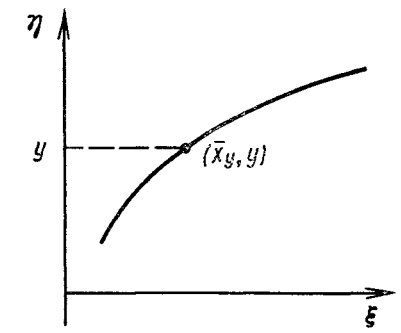


Рис. 2. Регрессия  $\xi$  на  $\eta$

за  $\bar{y}_x$  будем выбирать условное среднее величины  $\eta$ , задаваемое соотношением

$$\bar{y}_x = M(\eta | \xi = x) = \frac{\int_{-\infty}^{\infty} y f(x, y) dy}{\int_{-\infty}^{\infty} f(x, y) dy}.$$

Получим кривую регрессии для **среднего** значения  $\eta$ . Если рассматривать  $\eta$  как **независимую** переменную, то условная плотность вероятности **зависимой** переменной  $\xi$  при фиксированном значении  $\eta = y$  будет

$$f(x/y) = \frac{f(x, y)}{f_2(y)} = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dx} = \frac{f_1(x) f(y/x)}{f_2(y)}.$$

Любая числовая характеристика  $\bar{x}_y$  условного распределения величины  $\xi$  приводит к **кривой регрессии  $\xi$  на  $\eta$**  (рис. 2). Кривая регрессии для условного среднего значения  $\xi$  описывается уравнением

$$\bar{x}_y = M(\xi / \eta = y).$$

Две кривые регрессии  $\bar{y}_x$  и  $\bar{x}_y$  в общем случае не будут совпадать. Кривые регрессии выбираются таким образом, чтобы они отвечали свойству **минимальности**: среди множества всех функций  $g(\xi)$  пытаются найти такую, которая дает **возможно лучшее представление** о случайной величине  $\eta$ . Например, если «возможно лучшее» определить как минимум выражения

$$\begin{aligned}
M\{[\eta - g(\xi)]^2\} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [y - g(x)]^2 f(x, y) dx dy = \\
&= \int_{-\infty}^{\infty} f_1(x) dx \int_{-\infty}^{\infty} [y - g(x)]^2 f(y/x) dy, \quad (1.1)
\end{aligned}$$

то среди всех возможных  $g(\xi)$  надо взять  $g(\xi) = \bar{y}_x$ . Аналогично выражение  $M\{[\xi - h(\eta)]^2\}$  достигает минимума при функции  $h(\eta) = \bar{x}_y$ .

На практике наибольшее применение нашла **другая задача регрессии**: ограничимся функциями  $g(\xi)$  или  $h(\eta)$ , принадлежащими одному классу (например, линейными функциями), и будем путем подбора значений свободных параметров этих функций искать среди них такую, которая дает возможно лучшее представление величины  $\eta$ . В общем случае это будет другая кривая регрессии.

Наиболее простая и распространенная — линейная регрессия  $\eta \equiv g(\xi) = \theta_1 + \theta_2 \xi$ . Если оценки параметров  $\theta_1$  и  $\theta_2$  находятся из условия (1.1), то говорят о **линейной средней квадратической регрессии**.

Обратимся к способам определения **возможно лучшего представления зависимой переменной**. Для этих целей достаточно широко используется концепция метода наименьших квадратов. Здесь можно отметить два подхода. Во-первых, из курса линейной алгебры известно, что представление некоторого элемента (функции)  $y$ , принадлежащего пространству  $L_2$  (функций с интегрируемым квадратом), в виде линейной комбинации заданных элементов (функций)  $\varphi_i(x)$ ,  $i = \overline{1, m}$  сводится к отысканию минимума по  $\alpha_i$  ( $i = \overline{1, m}$ ) следующего выражения:

$$F = \int_x \left[ y - \sum_{i=1}^m \alpha_i \varphi_i(x) \right]^2 dx. \quad (1.2)$$

К этому же функционалу можно придти другим путем. Пусть случайная величина  $\eta$  имеет нормальное распределение, т. е.  $\eta \in N(\sum_i \alpha_i \varphi_i(x); \sigma^2)$ . Для простоты положим, что  $\sigma^2$  известна и постоянна для любого  $x$ . Рассмотрим независимые наблюдения  $x_j$  ( $j = \overline{1, n}$ ), каждому из которых отвечает значение  $y_j$ . Функция правдоподобия для этих наблюдений

$$L(y/x) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ - \frac{1}{2\sigma^2} \sum_{j=1}^n \left[ y_j - \sum_{i=1}^m \alpha_i \varphi_i(x_j) \right]^2 \right\}$$

и будет максимальна, если достигается минимум по  $\alpha_i$  ( $i = \overline{1, m}$ ) выражения

$$F = \frac{1}{2\sigma^2} \sum_{j=1}^n \left[ y_j - \sum_{i=1}^m \alpha_i \varphi_i(x_j) \right]^2, \quad (1.3)$$

эквивалентного выражению (1.2).

Статистический функционал (1.3) дает нам возможность не только найти оценки  $\alpha_i$ , но и определить закон их распределения и интервальные оценки.

Следует заметить, что использование функции правдоподобия приводит к методу наименьших квадратов только в случае нормального закона распределения случайной величины (зависимой переменной). Например, если случайная величина  $\eta$  подчиняется распределению Лапласа

$$f(x) = \frac{1}{2\beta} \exp \left\{ - \frac{|y - \sum_i \alpha_i \varphi_i(x)|}{\beta} \right\},$$

то функция правдоподобия для независимых наблюдений имеет вид

$$L(y/x) = \frac{1}{2^n \beta^n} \prod_{j=1}^n \exp \left\{ - \frac{|y_j - \sum_i \alpha_i \varphi_i(x_j)|}{\beta} \right\},$$

а функционал, аналогичный функционалу (1.3), будет

$$F = \frac{1}{\beta} \sum_{j=1}^n \left| y_j - \sum_i \alpha_i \varphi_i(x_j) \right|. \quad (1.4)$$

Функционалы (1.3) и (1.4) показывают, что то или иное определение «возможно лучшего» зависит от законов распределения исходных данных.

Здесь мы рассмотрели функционалы для регрессии  $\eta$  на  $\xi$ . Аналогично получаются выражения для функционалов, определяющих регрессию  $\xi$  на  $\eta$ . В обоих случаях следует подчеркнуть тот факт, что независимые переменные принимают **фиксированные** (детерминированные, неслучайные) значения. Случайной остается только значение зависимой переменной.

## 1.2. Линейная регрессия $\eta$ на $\xi$ и $\xi$ на $\eta$

Получим уравнения линий регрессии в этом широко распространенном случае. Пусть уравнение регрессии  $\eta$  на  $\xi$  имеет вид

$$\bar{y}_x = \theta_1 + \theta_2 x.$$

Используя определение линейной средней квадратической регрессии, найдем оценки  $\theta_1$  и  $\theta_2$  и выразим уравнение регрессии через данные наблюдений. Для оценки  $\theta_1$  и  $\theta_2$  продифференцируем

$$F = \sum_i (y_i - \theta_1 - \theta_2 x_i)^2 \quad (1.5)$$

по  $\theta_1$  и  $\theta_2$  и приравняем производные к нулю. После преобразований получим оценки

$$\hat{\theta}_2 = \rho \frac{\sigma_y}{\sigma_x}; \quad \hat{\theta}_1 = \bar{y} - \hat{\theta}_2 \bar{x},$$

где  $\hat{\theta}_2$  — коэффициент регрессии  $\eta$  на  $\xi$ ;  $\bar{x}$ ,  $\bar{y}$  — соответственно средние значения  $\{x_i\}$  и  $\{y_i\}$ ;  $\rho$  — коэффициент корреляции случайных величин  $\xi$  и  $\eta$ ;  $\sigma_x$  и  $\sigma_y$  — соответственно средние квадратические отклонения вариационных рядов  $x$  и  $y$  (не путать с величиной  $\sigma^2$  в (1.3) — дисперсией нормального закона распределения случайной величины  $\eta$ ). Отсюда прямая средней квадратической регрессии  $\eta$  на  $\xi$  имеет вид

$$\frac{\bar{y}_x - \bar{y}}{\sigma_y} = \rho \frac{x - \bar{x}}{\sigma_x}. \quad (1.6)$$

Подставив найденные оценки  $\hat{\theta}_1$  и  $\hat{\theta}_2$  в (1.5), найдем минимальное значение функционала  $F$

$$F_{\min} = \sigma_y^2 (1 - \rho^2).$$

Выражение (1.5) определяет сумму квадратов расстояний по вертикали (вдоль оси  $OY$ ) между точками  $(x_i, y_i)$  и прямой  $\bar{y}_x = \theta_1 + \theta_2 x$ .

В случае регрессии  $\xi$  на  $\eta$  надо найти оценки параметров  $\alpha$  и  $\beta$ , при которых

$$M[(\xi - \alpha - \beta\eta)^2] \rightarrow \min_{\alpha, \beta},$$

или

$$\Phi = \sum_i (x_i - \alpha - \beta y_i)^2 \rightarrow \min_{\alpha, \beta},$$

где  $y_i$  — фиксированы. Здесь расстояние между точками  $(y_i, x_i)$  и прямой  $\bar{x}_y = \alpha + \beta y$  измеряется по горизонтали.

Уравнение прямой средней квадратической регрессии  $\xi$  на  $\eta$  имеет вид

$$\frac{\bar{x}_y - \bar{x}}{\sigma_x} = \rho \frac{y - \bar{y}}{\sigma_y}. \quad (1.7)$$

Коэффициент регрессии  $\hat{\beta} = \rho \frac{\sigma_x}{\sigma_y}$  и  $\Phi_{\min} = \sigma_x^2 (1 - \rho^2)$ .

### 1.3. Регрессионный парадокс

По одним и тем же исходным данным согласно уравнениям (1.6) и (1.7) мы получим различные прямые, а следовательно, и различные выводы. Прямые регрессии совпадают только при  $\rho = \pm 1$ ; при  $\rho = 0$  (когда  $\xi$  и  $\eta$  — независимые случайные величины) имеем соответственно уравнения

$$\bar{y}_x = \bar{y}; \quad \bar{x}_y = \bar{x}$$

— две взаимно перпендикулярные прямые.

Если  $\bar{x} = \bar{y} = 0$ ,  $\sigma_x = \sigma_y = 1$ , то линия регрессии  $\eta$  на  $\xi$  описывается уравнением  $\bar{y}_x = \rho x$ ; а линия регрессии  $\xi$  на  $\eta$  уравнением  $y = \frac{\bar{x}_y}{\rho}$ .

В качестве примера рассмотрим линии регрессии, характеризующие зависимость роста сыновей от роста отцов. Пусть переменная  $\xi$  характеризует рост отца, а переменная  $\eta$  — рост сына (рис. 3). Уравнение регрессии  $\eta$  на  $\xi$  имеет вид

$$\frac{\bar{y}_x - \bar{y}}{\sigma_y} = \rho \frac{x - \bar{x}}{\sigma_x}; \quad |\rho| \leq 1; \quad \rho > 0.$$

Положим  $\sigma_y = \sigma_x = 1$  и перенесем начало координат в центр тяжести, в точку  $(\bar{x}, \bar{y})$ , т. е. в уравнениях регрессии надо принять  $\bar{x} = \bar{y} = 0$ . Получим

$$\bar{y}_x = \rho x < x.$$

Это означает, что в среднем сыновья высоких отцов не так высоки, как их отцы. Рост сыновей имеет тенденцию к усреднению.

Согласно регрессии  $\xi$  на  $\eta$

$$\frac{\bar{x}_y - \bar{x}}{\sigma_x} = \rho \frac{y - \bar{y}}{\sigma_y},$$

при сделанных предположениях  $\bar{x}_y = \rho y < y$ .

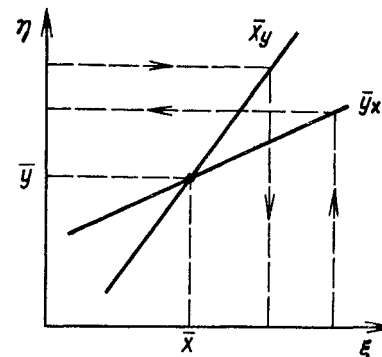


Рис. 3. Регрессионный парадокс

Это означает, что в среднем отцы высоких детей не так высоки, как их дети. Первое поколение имеет тенденцию к усреднению.

Причем, оба этих утверждения имеют место одновременно — налицо регрессионный парадокс. Чтобы устранить этот парадокс, необходимо при определении уравнения регрессии учесть случайный характер как переменной  $\xi$ , так и переменной  $\eta$ .

#### 1.4. Ортогональная регрессия

Учесть одновременно случайный характер переменных  $\xi$  и  $\eta$  позволяет ортогональная регрессия. Пусть нам известны набор случайных точек  $(x_i, y_i)$ ,  $i = \overline{1, n}$  и закон распределения каждого наблюдения. Неопределенность каждой точки теперь, в отличие от регрессий  $\eta$  на  $\xi$  и  $\xi$  на  $\eta$ , имеет вид некоторой области, форма которой зависит от закона распределения результатов наблюдений. И линия регрессии должна возможно лучшим образом пройти по областям неопределенности (рис. 4).

Пусть результаты наблюдений  $x_i$  и  $y_i$  подчиняются нормальным законам распределения соответственно с математическими ожиданиями  $\xi_i$  и  $\eta_i$ , дисперсиями  $\sigma^2(x_i)$  и  $\sigma^2(y_i)$  и коэффициентами корреляции  $\rho_i$ . Тогда функция правдоподобия имеет вид

$$L(x, y/\theta) = \prod_{i=1}^n \frac{1}{2\pi\sigma(x_i)\sigma(y_i)\sqrt{1-\rho_i^2}} \exp\left\{-\frac{1}{2(1-\rho_i^2)} \times \left[\frac{(x_i-\xi_i)^2}{\sigma^2(x_i)} - 2\rho_i \frac{(x_i-\xi_i)(y_i-\eta_i)}{\sigma(x_i)\sigma(y_i)} + \frac{(y_i-\eta_i)^2}{\sigma^2(y_i)}\right]\right\}, \quad (1.8)$$

а соответствующий минимизируемый функционал

$$F = \frac{1}{2} \sum_{i=1}^n \frac{1}{1-\rho_i^2} \left[ \frac{(x_i-\xi_i)^2}{\sigma^2(x_i)} - 2\rho_i \frac{(x_i-\xi_i)(y_i-\eta_i)}{\sigma(x_i)\sigma(y_i)} + \frac{(y_i-\eta_i)^2}{\sigma^2(y_i)} \right]. \quad (1.9)$$

При  $\rho=0$

$$F = \frac{1}{2} \sum_{i=1}^n \left[ \frac{(x_i-\xi_i)^2}{\sigma^2(x_i)} + \frac{(y_i-\eta_i)^2}{\sigma^2(y_i)} \right] \rightarrow \min. \quad (1.10)$$

Геометрически (при  $\sigma(x_i)=\sigma(y_i)=1$ ) это означает, что минимизируется сумма квадратов расстояний между точками  $(x_i, y_i)$  и соответствующими точками кривой регрессии  $(\xi_i, \eta_i)$ .

Однако из функционала  $F$  нельзя найти оценки параметров  $\hat{\theta}_j$  ( $j = \overline{1, m}$ ), так как решить систему уравнений

$$\frac{\partial F}{\partial \theta_j} = \sum_{i=1}^n \frac{y_i - \eta_i}{\sigma^2(y_i)} \frac{\partial \eta_i}{\partial \theta_j} \Big|_{\theta_j = \hat{\theta}_j} = 0, \quad j = \overline{1, m} \quad (1.11)$$

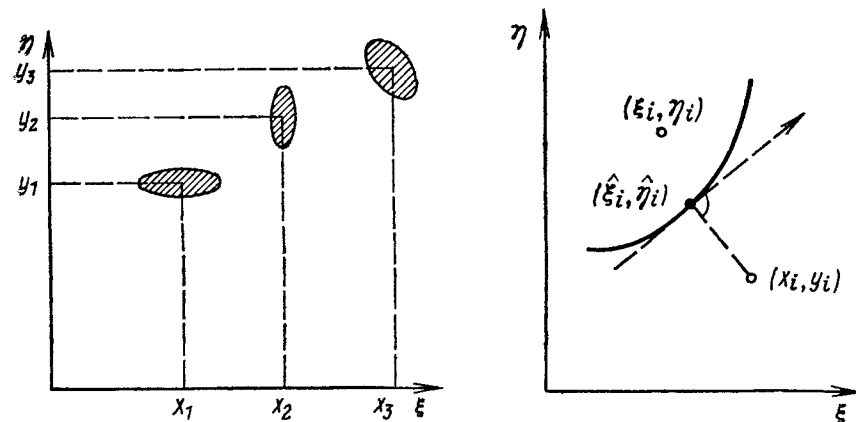


Рис. 4. Области неопределенности

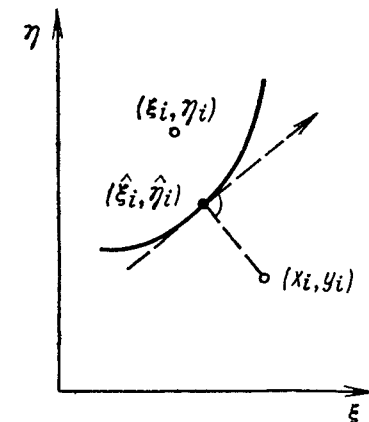


Рис. 5. Наблюдаемые и оцениваемые координаты переменных

нельзя, поскольку неизвестны истинные значения  $\xi_i$ , входящие в выражения для  $\eta_i$ .

Доопределим условие задачи. В качестве оценок  $\hat{\xi}_i$  возьмем те значения  $\xi_i$ , которые обратят в нуль частные производные  $\frac{\partial F}{\partial \xi_i}$ ;  $i = \overline{1, n}$ :

$$\frac{\partial F}{\partial \xi_i} = \frac{x_i - \xi_i}{\sigma^2(x_i)} + \frac{y_i - \eta_i}{\sigma^2(y_i)} \frac{\partial \eta_i}{\partial \xi_i} \Big|_{\xi_i = \hat{\xi}_i} = 0. \quad (1.12)$$

Условие (1.12) при  $\sigma(x_i)=\sigma(y_i)=1$  есть условие ортогональности векторов  $\{(x_i - \hat{\xi}_i); (y_i - \hat{\eta}_i)\}$  и  $\{1; \frac{\partial \eta_i}{\partial \xi_i}\}$ . Причем последний вектор направлен по касательной к кривой  $\eta = f(\xi, \theta)$  или вдоль прямой при линейной регрессии (рис. 5). Условия (1.11), (1.12) определяют третью линию регрессии — ортогональную, не совпадающую ни с  $\eta$  на  $\xi$ , ни с  $\xi$  на  $\eta$ , но одновременно учитывающую случайный характер и  $\xi$ , и  $\eta$ .

Приведенные результаты справедливы для пассивного эксперимента, когда просто наблюдаются текущие значения  $x_i$  и соответствующие им значения  $y_i$ . К результатам активного эксперимента, когда заранее устанавливаются значения  $x_i$ , в общем случае выводы пассивного эксперимента неприменимы.



### 1.5. Оценка свободных параметров функций, линейных по параметрам, методом наименьших квадратов

Примером функций, линейных по параметрам  $\theta$ , могут быть функции

$$\eta = \sum_{i=0}^k \theta_i x^i; \quad \eta = \sum_{i=0}^k \theta_i f_i(x) \text{ и т. д.}$$

Здесь независимая переменная  $x$  (или в общем случае  $\vec{x}$ ) измеряется без ошибок (детерминирована). Результаты наблюдений за случайной величиной  $\eta$  дают значения  $y$ , соответствующие значениям  $x$ . Каждому значению  $x_i$  может отвечать либо единственное значение  $y_i$ , либо множество значений  $y_{i1}, y_{i2}, \dots, y_{im}$ .

Будем считать, что каждое наблюдение  $y_i$  содержит ошибку  $\varepsilon_i$ .

Пусть  $n$  — мерный вектор наблюдений (вектор откликов)  $\vec{y} = \{y_1, y_2, \dots, y_n\}^T$  порождается моделью

$$\vec{y} = \vec{x}\theta + \varepsilon,$$

где  $\vec{x}$  — известная матрица измерений (матрица плана), размером  $n \times p$ , построенная либо по значениям вектора  $\vec{x}$ , либо по значениям  $f_i(x)$ ,  $i = \overline{1, k}$ ;  $\theta = \{\theta_1, \theta_2, \dots, \theta_p\}$  — неизвестный  $p$  — мерный вектор параметров  $\theta \in \mathbb{R}^p$ ;  $\varepsilon$  —  $n$  — мерный случайный вектор ошибок, удовлетворяющий условию  $M(\varepsilon) = 0$ ;  $D(\varepsilon) = M(\varepsilon^T \varepsilon) = \sigma^2 I$ . Здесь  $\sigma^2$  — неизвестный скалярный параметр;  $I$  — единичная матрица.

Линейные модели наблюдений могут быть полного и неполного ранга. Если ранг матрицы измерений  $\text{rang } \vec{x}$  равен  $p$ , то модель наблюдений называется моделью полного ранга. Если  $\text{rang } \vec{x} < p$ , то — неполного ранга. Другими словами, если число точек наблюдений меньше числа искомых параметров  $\theta$ , то имеет место модель наблюдений неполного ранга. В этом случае часть параметров  $\theta$  являются свободными. Модель неполного ранга здесь не рассматривается.

А. Простейший случай:  $\sigma^2 = 1$ .

Воспользуемся методом наименьших квадратов, т. е. будем искать точку минимума квадратичной формы

$$Q(\theta) = \varepsilon^T \varepsilon = (\vec{y} - \vec{x}\theta)^T (\vec{y} - \vec{x}\theta) = \vec{y}^T \vec{y} - 2\theta^T \vec{x}^T \vec{y} + \theta^T \vec{x}^T \vec{x} \theta.$$

Дифференцируем  $Q(\theta)$  по  $\theta$ :

$$-2\vec{x}^T \vec{y} + 2\vec{x}^T \vec{x} \theta = 0 \text{ (система нормальных уравнений).}$$

$$\text{Отсюда } \hat{\theta} = (\vec{x}^T \vec{x})^{-1} \vec{x}^T \vec{y}.$$

Минимальное значение  $S$ , которое имеет  $Q(\theta)$  при  $\theta = \hat{\theta}$ ,

$$\begin{aligned} S &= \vec{y}^T \vec{y} - 2[(\vec{x}^T \vec{x})^{-1} \vec{x}^T \vec{y}]^T \vec{x}^T \vec{y} + [(\vec{x}^T \vec{x})^{-1} \vec{x}^T \vec{y}]^T \vec{x}^T \vec{x} \times [(\vec{x}^T \vec{x})^{-1} \vec{x}^T \vec{y}] = \\ &= \vec{y}^T \vec{y} - 2\vec{y}^T \vec{x} [(\vec{x}^T \vec{x})^{-1} \vec{x}^T \vec{y}] + \vec{y}^T \vec{x} (\vec{x}^T \vec{x})^{-1} \vec{x}^T \vec{x} (\vec{x}^T \vec{x})^{-1} \vec{x}^T \vec{y} = \\ &= \vec{y}^T \vec{y} - \vec{y}^T \vec{x} (\vec{x}^T \vec{x})^{-1} \vec{x}^T \vec{y}; \quad S = \vec{y}^T \vec{y} - \vec{y}^T \vec{x} \hat{\theta}. \end{aligned}$$

Здесь использованы соотношения  $(AB)^T = B^T A^T$ ;  $(AB)^{-1} = B^{-1} A^{-1}$ .

Итак, метод оценивания параметров (метод наименьших квадратов) не требует никаких предположений о нормальном законе распределения исходных данных при получении **точечных оценок**, но они становятся необходимыми при построении доверительных интервалов и для проверки гипотез о параметрах  $\theta$ .

Рассмотрим свойства оценок метода наименьших квадратов.

**Теорема.** Оценка  $\hat{\theta}$  является несмещенной оценкой  $\theta$ .

Доказательство:

$$M[\hat{\theta}] = M[(\vec{x}^T \vec{x})^{-1} \vec{x}^T \vec{y}] = (\vec{x}^T \vec{x})^{-1} \vec{x}^T M[\vec{y}];$$

$$M[\vec{y}] = \vec{x}\theta; \quad M[\hat{\theta}] = (\vec{x}^T \vec{x})^{-1} \vec{x}^T \vec{x} \theta = \theta \text{ (и т. д.).}$$

**Теорема** (Гаусса — Маркова). Среди класса оценок  $\theta^*$  величины  $\theta$ , (которые а) являются несмещенными оценками и б) представляют собой линейные комбинации исходных данных  $\vec{y}$ , с помощью метода наименьших квадратов можно найти такую оценку  $\hat{\theta}$ , что  $D(\hat{\theta}) \leq D(\theta^*)$ , т. е.  $\hat{\theta}$  — наиболее точная оценка  $\theta$  из всех возможных, принадлежащих данному классу (эффективная оценка). Без доказательства.

Получим формулу для оценки  $\vec{D}(\theta)$ . Пусть дисперсия  $D(\varepsilon_j) = \sigma^2$ ;  $j = \overline{1, n}$ , т. е. дисперсия  $D(\vec{y}) = \sigma^2 I$ . В общем случае  $D(\vec{W}\vec{y}) = \vec{W} D(\vec{y}) \vec{W}^T$ . При  $\hat{\theta} = (\vec{x}^T \vec{x})^{-1} \vec{x}^T \vec{y}$

$$D(\hat{\theta}) = (\vec{x}^T \vec{x})^{-1} \vec{x}^T D(\vec{y}) [(\vec{x}^T \vec{x})^{-1} \vec{x}^T]^T = (\vec{x}^T \vec{x})^{-1} \vec{x}^T \sigma^2 I \vec{x} (\vec{x}^T \vec{x})^{-1} = \\ = \sigma^2 (\vec{x}^T \vec{x})^{-1} \vec{x}^T I \vec{x} (\vec{x}^T \vec{x})^{-1} = \sigma^2 (\vec{x}^T \vec{x})^{-1}.$$

Из формулы для  $D(\hat{\theta})$  следует, что дисперсии оценок могут быть получены без знания самих оценок.

Получим формулу для оценки параметра  $\sigma^2$ . В формулу для дисперсии оценок входит неизвестный параметр  $\sigma^2$ . Его оценка  $\hat{\sigma}^2 = \frac{S}{n-p}$ , где  $n$  — число точек наблюдения;  $p$  — число оцениваемых параметров  $\theta$  (без доказательства).

Замечание: если  $D(\varepsilon_j) = \sigma_j^2$ , то можно положить  $D(\varepsilon_j) = \frac{\sigma^2}{\omega_j}$ , где  $\omega_j$  — вес  $j$ -го наблюдения. Вместо  $\varepsilon_j = y_j - \sum_{i=1}^p \hat{\theta}_i x_{ij}$  следует взять  $\varepsilon_j^* = \varepsilon_j \sqrt{\omega_j}$ , тогда

$$D(\varepsilon_j^*) = D(\varepsilon_j \sqrt{\omega_j}) = \omega_j D(\varepsilon_j) = \omega_j \frac{\sigma^2}{\omega_j} = \sigma^2$$

$$\text{и } S = \sum_{j=1}^n (\varepsilon_j^*)^2 = \sum_{j=1}^n \omega_j \varepsilon_j^2.$$

Б. Общий случай: пусть результаты наблюдений подчиняются модели

$$\vec{y} = \vec{x}\theta + \varepsilon.$$

Ошибки  $\varepsilon$  имеют математическое ожидание  $M(\varepsilon) = 0$  и конечные вторые моменты, а также ковариационную матрицу  $D(\vec{y})$ :

$$M[\varepsilon_j^2] = \sigma_j^2 = D(y_j);$$

$$M[\varepsilon_i, \varepsilon_j] = \rho_{ij} \sigma_i \sigma_j = \text{cov}(y_i, y_j),$$

т. е.

$$D(\vec{y}) = \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho_{12} & \dots & \sigma_1 \sigma_n \rho_{1n} \\ \sigma_1 \sigma_2 \rho_{12} & \sigma_2^2 & \dots & \sigma_2 \sigma_n \rho_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_1 \sigma_n \rho_{1n} & \sigma_2 \sigma_n \rho_{2n} & \dots & \sigma_n^2 \end{pmatrix}.$$

Рассмотрим квадратичную форму

$$Q(\vec{\theta}) = (\vec{y} - \vec{x}\vec{\theta})^T D^{-1}(\vec{y}) (\vec{y} - \vec{x}\vec{\theta}) =$$

$$= \vec{y}^T D^{-1}(\vec{y}) \vec{y} - 2\vec{y}^T D^{-1}(\vec{y}) \vec{x}\vec{\theta} + (\vec{x}\vec{\theta})^T D^{-1}(\vec{y}) \vec{x}\vec{\theta} = \\ = \vec{y}^T D^{-1}(\vec{y}) \vec{y} - 2\vec{\theta}^T \vec{x}^T D^{-1}(\vec{y}) \vec{y} + \vec{\theta}^T \vec{x}^T D^{-1}(\vec{y}) \vec{x}\vec{\theta};$$

продифференцируем  $Q(\vec{\theta})$  по  $\vec{\theta}$ :

$2\vec{x}^T D^{-1}(\vec{y}) \vec{y} = 2\vec{x}^T D^{-1}(\vec{y}) \vec{x}\vec{\theta}$  — система нормальных уравнений. Отсюда

$$\hat{\vec{\theta}} = (\vec{x}^T D^{-1}(\vec{y}) \vec{x})^{-1} \vec{x}^T D^{-1}(\vec{y}) \vec{y}.$$

Оценки  $\hat{\vec{\theta}}$  являются также несмещенными и эффективными. Дисперсия оценок

$$D(\hat{\vec{\theta}}) = [\vec{x}^T D^{-1}(\vec{y}) \vec{x}]^{-1}.$$

Если  $D(\vec{y}) = \sigma^2 \vec{N}$ , то  $D^{-1}(\vec{y}) = \frac{1}{\sigma^2} \vec{N}^{-1} = \frac{\vec{W}}{\sigma^2}$  и  $D(\hat{\vec{\theta}}) = \sigma^2 (\vec{x}^T \vec{W} \vec{x})^{-1}$ ;  $\hat{\sigma}^2 = \frac{\varepsilon^T \vec{W} \varepsilon}{n-p}$ .

**Пример.** Найти оценки параметров  $\vec{\theta}$  в модели  $\eta = \theta_1 x_1 + \theta_2 x_2$ , если матрица плана

$$\vec{x} = \begin{pmatrix} 3 & 2 & 3 \\ 1 & 1 & 2 \end{pmatrix}^T;$$

вектор наблюдений  $\vec{y} = \{4, 4, 4\}^T$ ;

$$D(\vec{y}) = \sigma^2 \vec{N}; \quad N = \begin{pmatrix} 4 & 2 & 0 \\ 2 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

**Решение:**

$$D^{-1}(\vec{y}) = \frac{\vec{W}}{\sigma^2}; \quad \vec{W} = \vec{N}^{-1} = \begin{pmatrix} 0,333 & -0,167 & 0,000 \\ -0,167 & 0,333 & 0,000 \\ 0,000 & 0,000 & 0,250 \end{pmatrix}.$$

Так как  $\sigma^2$  постоянна для всех наблюдений, то на точечную оценку она не влияет (сокращается),  $\hat{\vec{\theta}} = \{1,00; 0,75\}^T$ ; вектор невязок  $\vec{\varepsilon} = \vec{y} - \vec{x}\hat{\vec{\theta}} = \{0,25; 1,25; -0,50\}^T$ ;

$$\hat{\sigma}^2 = \frac{\varepsilon^T \vec{W} \varepsilon}{n-p} = \frac{0,500}{3-2} = 0,500;$$

$$D(\hat{\vec{\theta}}) = \hat{\sigma}^2 (\vec{x}^T \vec{W} \vec{x})^{-1} = \begin{pmatrix} 1,000 & -1,750 \\ -1,750 & 3,437 \end{pmatrix}.$$

Отсюда  $\sigma(\hat{\theta}_1) = 1,000$ ;  $\sigma(\hat{\theta}_2) = 1,854$ .

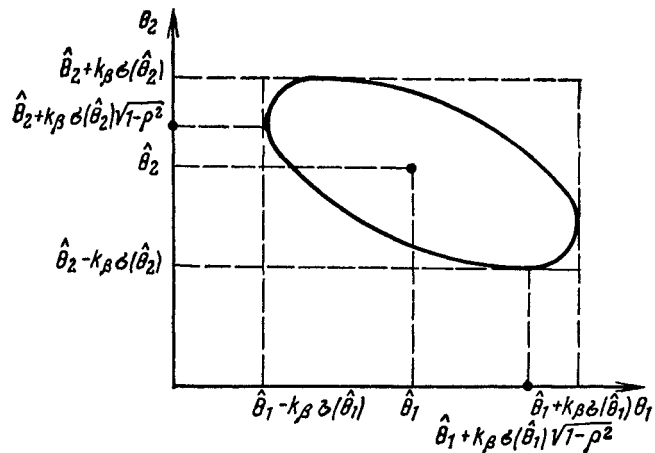


Рис. 6. Эллипс рассеяния оценок  $\vec{\theta}$

Корреляционная матрица оценок параметров  $\vec{\theta}$

$$\rho_{ij} = \begin{pmatrix} 1,00 & -0,94 \\ -0,94 & 1,00 \end{pmatrix},$$

она показывает, что бессмысленно в данном примере приводить только средние квадратические отклонения оценок, так как велики недиагональные элементы (рис. 6). Большая положительная ошибка в  $\hat{\theta}_1$  приведет к большой отрицательной ошибке в  $\hat{\theta}_2$ ,  $\sigma(\hat{\theta}_1)$  и  $\sigma(\hat{\theta}_2)$  — маргинальные средние квадратические отклонения, они определяются проекциями эллипса на соответствующие оси координат.

Более строго, из условия нормального распределения оценок  $\vec{\theta}$  следует, что квадратичная форма

$$Q(\vec{\theta}, \vec{\theta}) = (\vec{\theta} - \vec{\theta})^T D(\vec{\theta}) (\vec{\theta} - \vec{\theta})$$

распределена по закону  $\chi^2(N)$ ,

где  $N$  — размерность вектора  $\vec{\theta}$ .

Поэтому можно записать вероятностное утверждение

$$P[Q(\vec{\theta}, \vec{\theta}) \leq k_\beta^2] = \beta,$$

где  $\beta$  — доверительная вероятность;

$k_\beta^2$  — квантиль для  $\chi^2(N)$  — распределения.

Область в пространстве параметров  $\theta$  задается уравнением

$$Q(\vec{\theta}, \vec{\theta}) = k_\beta^2$$

и имеет вид гиперэллипсоида. Отсюда можно получить доверительную область для  $\vec{\theta}$

В нашем примере для эллиптической доверительной области при  $\beta = 0,68$  имеем  $k_\beta^2 = 3,8$ . Если взять  $k_\beta = 1$ , то получим  $\beta = 0,39$ . При коэффициенте корреляции  $\rho(\hat{\theta}_1, \hat{\theta}_2) = 0,95$  и  $k_\beta = 1$  прямоугольная область, включающая в себя эллипс, имеет доверительную вероятность  $\beta = 0,498$  [5].

### 1.6. Оценка параметров моделей с помощью функции правдоподобия

Пусть случайная величина  $X$  имеет плотность вероятности  $f(X, \vec{\theta})$ , где

$\vec{\theta}$  — неизвестный вектор параметров этого распределения. Проведем независимые выборки и получим реализации  $x_1, x_2, \dots, x_n$ . Дифференциал вероятности совместного распределения имеет вид

$$\begin{aligned} L(x_1, x_2, \dots, x_n; \vec{\theta}) dx_1 dx_2 \dots dx_n = \\ = f(x_1, \vec{\theta}) f(x_2, \vec{\theta}) \dots f(x_n, \vec{\theta}) dx_1 dx_2 \dots dx_n. \end{aligned}$$

Совместная плотность вероятности  $L(X, \vec{\theta}) = \prod_{i=1}^n f(x_i, \vec{\theta})$  называется функцией правдоподобия для выборки  $x_1, x_2, \dots, x_n$ .

Аналогично под  $f(x_i, \vec{\theta})$  можем понимать вероятность того, что дискретная случайная величина  $X$  принимает значение  $x_i$ , т. е.

$$f(x_i, \vec{\theta}) = P(X = x_i).$$

Можно найти такое число (оценку)  $\vec{\theta}$ , которое обращает функцию правдоподобия в максимум;  $\vec{\theta}$  и будет точечной оценкой параметра  $\vec{\theta}$ .

Будем считать, что оценки  $\vec{\theta}$  удовлетворяют следующим условиям (условиям регулярности):

1) Пусть  $\vec{\theta} = \vec{\theta}(x_1, x_2, \dots, x_n)$  — однозначная функция от  $x_1, x_2, \dots, x_n$  непрерывная и имеющая непрерывные частные производные  $\frac{\partial \vec{\theta}}{\partial x_i}$  во всех точках  $X = (x_1, x_2, \dots, x_n) \in R_n$  и операторы  $\frac{\partial}{\partial \theta}$  и  $\int$  коммутируют;

2) Область значений  $X$  не зависит от параметра  $\theta$ .

Предположим, что математическое ожидание оценки  $M[\hat{\theta}] = \theta + b(\theta)$ , где  $b(\theta)$  — смещение оценки  $\hat{\theta}$ .

**Теорема.** В каждом случае регулярной оценки непрерывного типа квадрат среднего квадратического отклонения оценки  $\hat{\theta}$  от истинного значения  $\theta$  удовлетворяет неравенству

$$D(\hat{\theta}) = M[(\hat{\theta} - \theta)^2] \geq \frac{\left(1 + \frac{db}{d\theta}\right)^2}{\int_{-\infty}^{\infty} \left(\frac{\partial \ln L}{\partial \theta}\right)^2 L(x, \theta) dx}.$$

Доказательство. Имеем

$$\int L(\hat{\theta}/\theta) d\hat{\theta} = 1.$$

Дифференцируем это выражение по  $\theta$

$$\int \frac{\partial L}{\partial \theta} d\hat{\theta} = 0 = \int \frac{\partial \ln L}{\partial \theta} \cdot L(\hat{\theta}/\theta) d\hat{\theta},$$

т. к.  $\frac{\partial \ln L}{\partial \theta} = \frac{1}{L} \cdot \frac{\partial L}{\partial \theta}$ .

Имеем

$$M(\hat{\theta}) = \int \hat{\theta} \cdot L(\hat{\theta}/\theta) d\hat{\theta} = \theta + b(\theta);$$

дифференцируем по  $\theta$ , получим

$$\frac{\partial}{\partial \theta} \int \hat{\theta} \cdot L(\hat{\theta}/\theta) d\hat{\theta} = 1 + \frac{d[b(\theta)]}{d\theta}.$$

Поскольку  $\theta + b(\theta)$  не зависит от  $\hat{\theta}$ , то

$$\int [\theta + b(\theta)] \frac{\partial (\ln L)}{\partial \theta} \cdot L(\hat{\theta}/\theta) d\hat{\theta} = 0.$$

Вычитаем последнее выражение из предыдущего

$$\int [\hat{\theta} - \theta - b(\theta)] \frac{\partial (\ln L)}{\partial \theta} \cdot L(\hat{\theta}/\theta) d\hat{\theta} = 1 + \frac{d[b(\theta)]}{d\theta}.$$

Используя неравенство Шварца

$$\left[ \int f(x) g(x) dx \right]^2 \leq \int f^2(x) dx \int g^2(x) dx,$$

преобразуем этот интеграл

$$\int [\hat{\theta} - \theta - b(\theta)]^2 L(\hat{\theta}/\theta) d\hat{\theta} \cdot \int \left[ \frac{\partial (\ln L)}{\partial \theta} \right]^2 L(\hat{\theta}/\theta) d\hat{\theta} \geq \left(1 + \frac{d[b(\theta)]}{d\theta}\right)^2.$$

Первый интеграл — выражение для дисперсии  $D(\hat{\theta})$ , т. е.

$$D(\hat{\theta}) \geq \frac{\left(1 + \frac{d[b(\theta)]}{d\theta}\right)^2}{M\left[\left(\frac{\partial \ln L(\hat{\theta}/\theta)}{\partial \theta}\right)^2\right]}.$$

Это неравенство называется неравенством Крамера — Рао.

Поскольку для нахождения  $D(\hat{\theta})$  требуется  $\ln L(\hat{\theta}/\theta)$  и поскольку максимум функций  $L(\hat{\theta}/\theta)$  и  $\ln L(\hat{\theta}/\theta)$  совпадает, то для определения точечной оценки  $\hat{\theta}$  решается уравнение правдоподобия

$$\frac{\partial [\ln L(\hat{\theta})]}{\partial \theta} = 0.$$

Метод определения оценки  $\hat{\theta}$  с помощью функции правдоподобия называется методом максимума правдоподобия (ММП).

В асимптотике (для состоятельных оценок)

$$M\left[\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right]_{\theta=\hat{\theta}} = -M\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right]_{\theta=\hat{\theta}}.$$

Неравенство Крамера — Рао показывает, что наименьшая величина дисперсии оценки (для эффективной оценки) равна

$$D(\hat{\theta}) = \frac{1}{\int_{-\infty}^{\infty} \left(\frac{\partial \ln L}{\partial \theta}\right)^2 L(x, \theta) dx} = \frac{1}{M\left[\left(\frac{\partial \ln L(\hat{\theta})}{\partial \theta}\right)^2\right]}.$$

Откуда дисперсионная (ковариационная) матрица оценок  $\hat{\theta}$  найдется из условия

$$D(\hat{\theta}) = N^{-1}; \quad N = \left\| -\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \right\|; \quad i, j = 1, \dots, k.$$

**Пример.** Найти точечные оценки математического ожидания  $m$  и дисперсии  $\sigma^2$  нормального закона распределения.

Пусть провели выборку  $\{x_i\}$ ,  $i = \overline{1, n}$  и получили  $n$  независимых значений из нормального распределения. Функция правдоподобия в этом случае имеет вид

$$L(m, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - m)^2}{2\sigma^2}}.$$

Тогда

$$\ln L = -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2 - \frac{n}{2} \ln \sigma^2 + \text{const.}$$

Метод максимума правдоподобия приводит к уравнениям

$$\begin{cases} \frac{\partial \ln L}{\partial m} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - m) = 0 \\ \frac{\partial \ln L}{\partial \sigma^2} = \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - m)^2 - \frac{n}{2\sigma^2} = 0 \end{cases}$$

Отсюда имеем следующие оценки

$$\hat{m} = \frac{1}{n} \sum_{i=1}^n x_i \equiv \bar{x},$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Математическое ожидание  $M[\hat{m}] = M[\bar{x}] = M[x_i] = \mu$  — несмещенная оценка.

Найдем математическое ожидание  $\hat{\sigma}^2$ . Математическое ожидание каждого слагаемого в формуле для  $\hat{\sigma}^2$  равны. Тогда

$$M(\hat{\sigma}^2) = \frac{1}{n} \cdot n \cdot M[(x_i - \bar{x})^2] = M[(x_i - \mu)^2 + (\bar{x} - \mu)^2 - 2(x_i - \mu)(\bar{x} - \mu)] = \sigma^2 + \frac{\sigma^2}{n} - 2 \frac{\sigma^2}{n} = \sigma^2 \cdot \frac{n-1}{n}.$$

Получили, что  $\hat{\sigma}^2$  — смещенная оценка.

Несмещенная оценка (исправленная) дисперсии равна

$$s^2 = \frac{n}{n-1} \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

При достаточно широких условиях оценки максимального правдоподобия состоятельны, и при выполнении условий регулярности (наиболее важное из них — независимость области определения  $f(X, \theta)$  от  $\theta$ ), кроме того, асимптотически нормальны и эффективны.

Если функция правдоподобия имеет вторые производные по  $\theta$  в интервале, включающем истинное значение  $\theta_0$ , и если, кроме того

$$M \left[ \frac{\partial \ln L(X, \theta)}{\partial \theta} \right] = 0;$$

$N(\theta) = -M \left[ \frac{\partial^2 \ln L(X, \theta)}{\partial \theta_i \partial \theta_j} \right]$  — существует и отлично от нуля в указанном интервале, то оценка  $\hat{\theta}$  асимптотически нормально распределена со средним  $\theta_0$  и дисперсией  $[N(\theta)]^{-1}$ .

Итак, оценки ММП в пределе распределены по нормальному закону

$$N[\theta, 1/(nI(\theta))],$$

где  $n$  — число наблюдений;  $I(\theta) = \frac{\partial^2 \ln L}{\partial \theta^2}$  — информация о параметре  $\theta$ , содержащаяся в одном наблюдении (хотя оценка  $\hat{\theta}$  может

быть смещенной). Кроме того асимптотически нормально распределена величина  $\frac{\partial \ln L}{\partial \theta}$  со средним, равным нулю, и дисперсией, равной  $nI(\theta)$ .

В приложениях часто используется приближение, что переменная  $\frac{1}{\sqrt{nI(\theta)}} \cdot \frac{\partial \ln L}{\partial \theta}$  имеет стандартное нормальное распределение.

Важным свойством оценки ММП является ее инвариантность: оценка  $\hat{\tau}$  функции  $\tau(\theta)$  равна  $\hat{\tau} = \tau(\hat{\theta})$ . Например, оценка  $\theta^2$  равна квадрату оценки  $\hat{\theta}$ , но оценка  $\hat{\tau}$  не обязательно несмещенная.

Интервальная оценка параметра  $\theta$  может быть получена по графику функции правдоподобия.

Одномерный случай. Пусть результаты наблюдений подчиняются нормальному распределению  $N(\mu, \sigma^2=1)$ .

Логарифм функции правдоподобия имеет вид параболы

$$\begin{aligned} \ln L(X, \mu) &= \ln C - \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 = \\ &= \ln C - \frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2 - \frac{n}{2} (\mu - \bar{x})^2 \end{aligned}$$

с точкой максимума при  $\mu = \bar{x}$ .  
Смещаем параболу так, чтобы

$$\ln L(\bar{x}) = 0.$$

Пусть  $\ln L = -\frac{1}{2}$ ; при  $n=1$

$$\frac{(\mu - \bar{x})^2}{2} = \frac{1}{2}.$$

Откуда получим доверительный интервал

$$\bar{x} - 1 \leq \mu \leq \bar{x} + 1.$$

Но из свойств нормального распределения следует, что

$$P[(\bar{x} - \mu)^2 \leq 1] = 0,683,$$

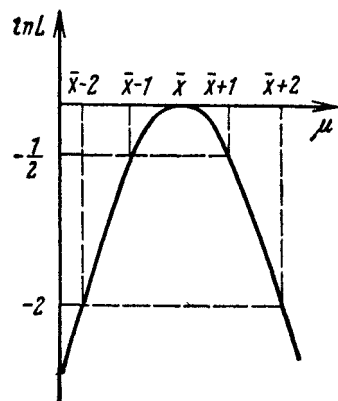
т. е.

$$P(\bar{x} - 1 \leq \mu \leq \bar{x} + 1) = 0,683 \text{ (рис. 7).}$$

Пусть  $\ln L = -2$ ; откуда  $\frac{(\mu - \bar{x})^2}{2} = 2$  и  $|\mu - \bar{x}| = 2$ .

Или  $\bar{x} - 2 \leq \mu \leq \bar{x} + 2$ .

Рис 7. Интервальная оценка по функции правдоподобия



Из условия нормального распределения имеем

$$P[(\bar{x}-\mu)^2 \leq 2] = 0,955$$

или

$$P(\bar{x}-2 \leq \mu \leq \bar{x}+2) = 0,955.$$

В общем случае, когда  $L(\mathbf{X}/\theta)$  — непрерывная функция и имеет один максимум, можно найти преобразование  $g(\theta, \mathbf{X})$ , которое преобразует кривую  $\ln L_\theta(\mathbf{X}/\theta)$  в параболу вблизи функции  $G(\mathbf{X})$  наблюдений

$$\ln L_g(\mathbf{X}/g(\theta)) = \frac{1}{2} [g - G(\mathbf{X})]^2.$$

В асимптотическом случае доказывается, что  $g$  можно выбрать независимо от  $\mathbf{X}$ . В таком случае  $G$  служит оценкой максимального правдоподобия для  $g$ .

Используя свойство инвариантности ММП, можно сделать такие же выводы о параметре  $g$ , какие делались о параметре нормального распределения  $\theta$ .

Преобразуем доверительный интервал для  $g$  в соответствующий интервал для  $\theta$ . Поскольку

$$\ln L_\theta(\mathbf{X}/\theta) = \ln L_g[\mathbf{X}/g(\theta)],$$

то значения  $\theta$ , соответствующие интервалу по  $g$  при значении  $\ln L_g = -2$  являются точками пересечения  $\ln L = -2$  с графиком функции  $\ln L_\theta(\mathbf{X}/\theta)$ . Таким образом можно сделать выводы о параметре непараболической функции правдоподобия, не находя преобразования к параболической функции правдоподобия. Но эта процедура имеет свои ограничения:

1) способ преобразования правилен с точностью только до порядка  $1/n$ , так как преобразуем экспериментальную функцию правдоподобия, тогда как нужно было бы сделать нормальным (преобразовать) теоретическое распределение;

2) полученный интервал может иметь очень широкие и нецентральные границы для первоначальной переменной;

3) если функция правдоподобия имеет несколько максимумов, то эта процедура дает несколько интервалов; интерпретация такого утверждения в качестве доверительного интервала для  $\theta$  до некоторой степени сомнительна: можно было бы найти один непрерывный интервал с тем же самым вероятностным содержанием;

4) когда предложенный метод дает неопределенные (или бесконечные) доверительные интервалы, то вероятнее всего используемое преобразование недостаточно хорошо по своей природе. Возможно, что для этой задачи требуется более сложная интерпретация, чем интерпретация ее в виде единственной оценки ММП и доверительного интервала.

### 1.7. Байесовский подход к оцениванию параметров моделей

В байесовском подходе неизвестный параметр не детерминированная (постоянная) величина, а случайная величина, для которой априори известно распределение  $\pi(\theta)$  возможных значений  $\theta$ .

После наблюдения  $\mathbf{X}$  получим апостериорное распределение

$$\pi(\theta/\mathbf{X}) = \frac{\left[ \prod_{i=1}^n f(x_i/\theta) \right] \pi(\theta)}{\int \left[ \prod_{i=1}^n f(x_i/\theta) \right] \pi(\theta) d\theta}.$$

Байесовский доверительный интервал  $(\theta_n, \theta_b)$  с вероятностным содержанием  $\beta$  определяется как

$$\int_{\theta_n}^{\theta_b} \pi(\theta/\mathbf{X}) d\theta = \beta.$$

Согласно этому определению доверительный интервал соответствует доле  $\beta$  полной веры наблюдается в параметр  $\theta$ , т. е. на языке пари шанс на выигрыш составляет  $\beta/(1-\beta)$ .

Полное апостериорное распределение  $\pi(\theta/\mathbf{X})$  как бы заменяется дискретным двухточечным распределением:

$$P(\theta_n \leq \theta \leq \theta_b) = \beta;$$

$$P\{\theta \notin [\theta_n, \theta_b]\} = 1 - \beta.$$

Как и в классическом варианте интервал определяется неоднозначно. В обоих случаях не указывается, какое значение внутри интервала является наиболее вероятным. Поэтому желательно дополнительно провести точечную оценку по максимуму  $\pi(\theta/X)$ .

*Замечание:* для нелинейных функций оценку свободных параметров часто проводят после линеаризации заданной функции, например,  $\eta = Ae^{\lambda x} \Rightarrow \ln \eta = A + \lambda x$  или  $\eta_1 = \theta_1 + \theta_2 x$ , где  $\eta_1 = \ln \eta$ ,  $\theta_1 = \ln A$ ;  $\theta_2 = \lambda$ . Но здесь возникают две проблемы: единственности оценок и изменения законов распределения исходных данных.

### 1.8. Интервальная оценка линии регрессии и прогнозируемых значений функции

Пусть получены оценки  $\vec{\hat{\theta}}$  и их дисперсии  $D(\vec{\hat{\theta}})$  в линейной модели  $\vec{\eta} = \vec{x}\vec{\theta}$ . Оценкой для линии регрессии  $\eta$  будет  $\hat{y} = \vec{x}\vec{\hat{\theta}}$ . Надо теперь определить доверительные интервалы для значений  $y$ , которые с заданной доверительной вероятностью  $\gamma$  накроют истинные значения  $\eta$ . Кроме того, надо найти доверительные интервалы для прогнозов, которые накроют новое наблюдение  $y$ , соответствующее заданному значению  $x$ .

Рассмотрим эти оценки для случая прямой линии  $\eta = \theta_0 + \theta_1(x - \bar{x})$ , где  $\bar{x}$  — среднее значение переменной  $x$ . Тогда дисперсия

$$D(\hat{y}) = D(\hat{\theta}_0) + D[\hat{\theta}_1(x - \bar{x})] = D(\hat{\theta}_0) + (x - \bar{x})^2 D(\hat{\theta}_1).$$

В общем случае величина  $\frac{|y - \eta|}{\sqrt{D(\hat{y})}}$  подчиняется распределению Стьюдента с  $n - p$  степенями свободы; при заданном  $\gamma$  квантиль распределения Стьюдента равна  $t_\gamma$  и интервальная оценка имеет вид

$$P\{\hat{y} - t_\gamma \sqrt{D(\hat{y})} < \eta < \hat{y} + t_\gamma \sqrt{D(\hat{y})}\} = \gamma.$$

В общем случае для линии регрессии, описываемой уравнением  $\eta(x, \theta) = f(x, \theta)$ , дисперсия наблюдаемых значений  $y$  определяется выражением

$$D(\hat{y}) = \sum_{j=1}^N \left( \frac{\partial f}{\partial \theta_j} \right)^2 D(\hat{\theta}_j) + 2 \sum_{i>j}^N \frac{\partial f}{\partial \theta_j} \frac{\partial f}{\partial \theta_i} D(\hat{\theta}_j, \hat{\theta}_i),$$

где  $N$  — размерность вектора  $\theta$ ;  $D(\hat{\theta}_j, \hat{\theta}_i)$  — корреляционный момент случайных величин  $\hat{\theta}_j$  и  $\hat{\theta}_i$ . Для линейных функций  $\eta(x, \theta) =$

$$= \sum_{i=1}^N \theta_i, f_i(x) \text{ дисперсия наблюдаемых значений } \hat{y} \text{ в точке } x \text{ равна}$$

$$D(\hat{y}) = f^T(x) D(\hat{\theta}) f(x).$$

Например, для прямой линии  $\eta(x, \theta) = \theta_0 + \theta_1(x - \bar{x})$  вектор  $f(x) = \begin{pmatrix} 1 \\ x - \bar{x} \end{pmatrix}$  и

$$D(\hat{y}) = (1 \quad x - \bar{x}) \begin{pmatrix} D(\hat{\theta}_0) & 0 \\ 0 & D(\hat{\theta}_1) \end{pmatrix} \begin{pmatrix} 1 \\ x - \bar{x} \end{pmatrix} = D(\hat{\theta}_0) + (x - \bar{x})^2 D(\hat{\theta}_1).$$

Новое наблюдение (прогнозируемое)  $y$  отсчитывается от оценки функции регрессии  $\hat{y}$  и распределение Стьюдента имеет величина  $\frac{|y - \hat{y}|}{\sqrt{D(y - \hat{y})}}$ . Новое наблюдение имеет дисперсию  $\sigma^2$  и

$$D(y - \hat{y}) = \sigma^2 + D(\hat{\theta}_0) + (x - \bar{x})^2 D(\hat{\theta}_1).$$

Интервальная оценка для точек прогноза при одном измерении в точке прогноза имеет вид

$$P\{\hat{y} - t_\gamma \sqrt{D(y - \hat{y})} < y < \hat{y} + t_\gamma \sqrt{D(y - \hat{y})}\} = \gamma.$$

Для  $k$  измерений в точке прогноза вместо  $\sigma^2$  в формуле для  $D(y - \hat{y})$  надо взять  $\sigma^2/k$ .

Для прямой линии нетрудно получить

$$\hat{\theta}_0 = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}; \quad \hat{\theta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

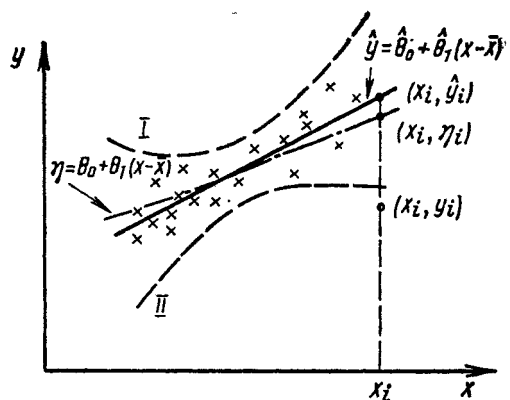
Пусть  $D[y_i] = \sigma^2$ ;  $i = \overline{1, n}$ . Тогда  $D(\hat{\theta}_0) = D(\bar{y}) = \frac{\sigma^2}{n}$ ;

$$D(\hat{\theta}_1) = D \left[ \frac{\sum_i (x_i - \bar{x}) y_i}{\sum_i (x_i - \bar{x})^2} \right] = \frac{\sum_i (x_i - \bar{x})^2 D[y_i]}{[\sum_i (x_i - \bar{x})^2]^2} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

(дисперсия  $D(\hat{y})$  минимальна при  $x = \bar{x}$ ). Задавая различные значения  $x$ , можно найти интервальные оценки линии регрессии (линии I и II) при заданном  $\gamma$  и интервальную оценку нового наблюдения  $y_i$  для прогнозируемой точки  $x_i$  (рис. 8).

Ниже будут рассмотрены алгоритмы, позволяющие находить оценки свободных параметров элементарных функций при наличии

Рис. 8. Интервальная оценка линии регрессии



погрешностей и в значениях аргумента, и в значениях функции. Эти алгоритмы объединены под названием «конфлюэнтный анализ» (confluence — слияние, confluent — сливающийся).

## 2. КОНФЛЮЭНТНЫЙ АНАЛИЗ КАК УНИВЕРСАЛЬНЫЙ МЕТОД УЧЕТА НЕОПРЕДЕЛЕННОСТЕЙ В ЗАДАЧАХ ПРОГНОЗИРОВАНИЯ

### 2.1. Активный и пассивный эксперименты.

#### Постановка задачи оценивания параметров функции известного вида в пассивном эксперименте

Пассивный эксперимент наиболее часто встречается в процессах получения оценок параметров моделей динамических систем во многих отраслях науки и техники.

Требуется найти интервальную оценку параметра  $\vec{\theta}$  функции  $\vec{\eta} = f(\vec{\xi}, \vec{\theta})$ , когда точные значения  $\vec{\eta}$  и  $\vec{\xi}$  мы наблюдать не можем, но можем наблюдать значения случайных величин  $\vec{Y}$  и  $\vec{X}$ , определяемые следующим образом:

$$x_i = \xi_i + \delta_i; y_i = \eta_i + \varepsilon_i, i = \overline{1, n}, \quad (2.1)$$

где  $\delta_i$  и  $\varepsilon_i$  — соответственно ошибки значений переменных и функции (случайные величины).

Пусть имеем статистический ряд экспериментальных значений  $\{x_i\} \in X$  и соответствующий им ряд значений функций  $\{y_i\} \in Y$ ,

$i = 1, 2, \dots, n; n \geq m$ , где  $m$  — число оцениваемых параметров  $\vec{\theta}$ . Будем считать, что переменные  $x_i$  и  $y_i$  не детерминированы, но являются выборками из генеральных совокупностей  $X$  и  $Y$  с известными плотностями функций вероятностей. Переменные  $x_i = \xi_i + \delta_i$  и  $y_i = \eta_i + \varepsilon_i$  могут быть статистически как зависимы, так и независимы; могут быть как коррелированы, так и не коррелированы.

В основном мы будем иметь дело с выборками из  $n$  независимых наблюдений из одного и того же распределения. Пусть  $f_1(x_i/\theta)$  и  $f_2(y_i/\theta)$  — соответственно плотности вероятности случайных величин  $x_i$  и  $y_i$ , если  $x_i$  и  $y_i$  непрерывны, и соответственно вероятности значений  $x_i$  и  $y_i$ , если распределения  $x_i$  и  $y_i$  дискретны;  $x_i, y_i$  и распределения  $f_1(x_i/\theta)$  и  $f_2(y_i/\theta)$  могут быть как одномерными, так и многомерными.

Найдем выражение для совместной плотности вероятности экспериментальных данных при условии, что значения  $\varepsilon_i$  и  $\eta_i$  связаны функциональной зависимостью, но их погрешности  $\delta_i$  и  $\varepsilon_i$  являются независимыми при переходе от одной точки  $(x_i, y_i)$  к другой. Тогда совместная плотность вероятности получить одновременно значения  $x_i$  и  $y_i$

$$P_i = f_1(x_i/\vec{\theta}) \cdot f_2(y_i/\vec{\theta}). \quad (2.2)$$

Совместная плотность вероятности получить  $n$  статистически независимых точек  $(x_i, y_i)$

$$L(x, y/\vec{\theta}) = \prod_{i=1}^n f_1(x_i/\vec{\theta}) \cdot f_2(y_i/\vec{\theta}). \quad (2.3)$$

Аналогично можно вывести формулы совместной плотности вероятности для зависимых или коррелированных экспериментальных точек. Эти формулы могут быть взяты из книг по теории вероятностей.

Для нас важен тот факт, что в выражения для совместной плотности вероятности входят математические ожидания экспериментальных данных, экспериментальные значения и оцениваемые параметры, так как  $f_1(x_i/\vec{\theta})$  — функция математического ожидания  $\xi_i$ , экспериментальных значений  $x_i$  и параметров  $\vec{\theta}$ ;  $f_2(y_i/\vec{\theta})$  — функция математического ожидания  $\eta_i$ , экспериментальных значений  $y_i$ ,  $x_i$  и параметров  $\vec{\theta}$ .

Кроме того, нам известно функциональное соотношение

$$\eta_i = f(\xi_i, \vec{\theta}), \quad (2.4)$$



которое порождает структурное соотношение между наблюдаемыми случайными величинами  $x_i$  и  $y_i$ :

$$y_i = \psi(x_i, \vec{\theta}, \delta_i, \varepsilon_i), \quad (2.5)$$

или

$$y_i = f(x_i - \delta_i, \vec{\theta}) + \varepsilon_i$$

при аддитивных помехах  $\delta_i$ ;  $\varepsilon_i$ .

Таким образом, в поставленной задаче следует отметить две проблемы: 1) каким образом ввести в рассмотрение погрешность в аргументе;

2) как минимизировать функционалы при отыскании оценок параметров  $\theta$  после введения в них оценок аргументов или выражений для них, так как эти функционалы имеют сложную форму и соответствующие системы уравнений для определения этих же оценок нелинейны. Известно, что каждая нелинейная система требует особого рассмотрения своего решения.

Наиболее часто в практике физических измерений встречаются распределения Пуассона и Гаусса. Найдем (в качестве примера) для них вид функционалов, из которых затем могут быть получены оценки искомых параметров (для других функций распределений экспериментальных данных процедура получения минимизируемого функционала будет аналогичной).

Пусть экспериментальные значения  $x_i$  и  $y_i$  — случайные величины, каждая из которых имеет функцию плотности вероятности (ф.п.в.), описываемую функцией Гаусса с математическими ожиданиями  $\xi_i$  и  $\eta_i$ , дисперсиями  $\sigma^2(x_i)$  и  $\sigma^2(y_i)$  и коэффициентом корреляции  $\rho_i = \rho(x_i, y_i)$ . Тогда плотность вероятности получить точку с координатами  $(x_i, y_i)$

$$P_i = \frac{1}{2\pi\sigma(x_i)\sigma(y_i)\sqrt{1-\rho_i^2}} \exp\left[-\frac{u_{1i}^2 - 2\rho_i u_{1i}u_{2i} + u_{2i}^2}{2(1-\rho_i^2)}\right], \quad (2.6)$$

где

$$u_{1i} = \frac{x_i - \xi_i}{\sigma(x_i)}; \quad u_{2i} = \frac{y_i - \eta_i}{\sigma(y_i)}.$$

Совместная плотность вероятности получить  $n$  независимых таких точек  $L(X, Y) = \prod_{i=1}^n P_i$

и

$$\ln L(X, Y) = -\frac{1}{2} \sum_{i=1}^n \left[ \frac{(x_i - \xi_i)^2}{\sigma^2(x_i)} - 2\rho_i \frac{(x_i - \xi_i)(y_i - \eta_i)}{\sigma(x_i)\sigma(y_i)} + \frac{(y_i - \eta_i)^2}{\sigma^2(y_i)} \right] \times \\ \times \frac{1}{1-\rho_i^2} + \text{const.} \quad (2.7)$$

Оценки искомых параметров  $\vec{\theta}$  находятся из условия минимума функционала

$$F = \frac{1}{2} \sum_{i=1}^n \frac{1}{1-\rho_i^2} \left[ \frac{(x_i - \xi_i)^2}{\sigma^2(x_i)} - 2\rho_i \frac{(x_i - \xi_i)(y_i - \eta_i)}{\sigma(x_i)\sigma(y_i)} + \frac{(y_i - \eta_i)^2}{\sigma^2(y_i)} \right]. \quad (2.8)$$

В важном частном случае, когда погрешности  $\delta_i$  и  $\varepsilon_i$  некоррелированы, выражение (2.8) примет вид

$$F = \frac{1}{2} \sum_{i=1}^n \left[ \frac{(x_i - \xi_i)^2}{\sigma^2(x_i)} + \frac{(y_i - \eta_i)^2}{\sigma^2(y_i)} \right]. \quad (2.9)$$

Мы получили явный вид одного из тех функционалов, которые по определению вводятся в методе обобщенной невязки.

Когда все экспериментальные данные статистически независимы и каждое экспериментальное значение имеет функцию вероятности, подчиняющуюся закону Пуассона,

$$L(X, Y) = \prod_{i=1}^n \frac{\xi_i^{x_i}}{x_i!} \cdot \frac{\eta_i^{y_i}}{y_i!} \exp(-\xi_i - \eta_i) \quad (2.10)$$

и

$$F = \sum_{i=1}^n (-\xi_i - \eta_i + x_i \ln \xi_i + y_i \ln \eta_i). \quad (2.11)$$

Аналогично получают вид минимизируемого функционала при других законах распределения исходных данных. Как в выражениях (2.6) ... (2.11) перейти к случаю, когда переменные являются детерминированными, очевидно.

Теперь рассмотрим задачу отыскания минимума функционалов типа (2.8) ... (2.10) по параметрам  $\vec{\theta}$  при условии (2.4).

Нам неизвестны истинные значения абсцисс экспериментальных точек, а известны только их доверительные области, тогда как в регрессионном анализе абсциссы известны. Причем случайная величина  $x$  коррелирована с обобщенной ошибкой функции  $\eta$  (например, для прямой линии  $\eta = \theta_1 + \theta_2 \xi$  обобщенная ошибка функции  $\eta$  равна  $\varepsilon - \theta_2 \delta$ ).

Перед тем, как приступить к определению точки минимума функционалов (2.8), (2.9), (2.11) по  $\vec{\theta}$ , требуется каким-то образом определить  $\xi_i$ , а только затем, подставив выражения для  $\xi_i$  и  $\eta_i$  в функционал, приступить к отысканию минимума получившейся функции нескольких переменных.

Выход системы  $y$  содержит информацию не только о параметрах модели  $\vec{\theta}$ , но и о действительных значениях входного воздей-

ствия  $\xi$ . Искомые значения  $\xi_i$  и оценки  $\theta$  определяются из условия

$$\frac{\partial F}{\partial \xi_i} \Big|_{\xi_i = \hat{\xi}_i} = 0; \quad i = \overline{1, n}; \quad (2.12)$$

$$\frac{\partial F}{\partial \eta} \cdot \frac{\partial \eta}{\partial \theta_j} \Big|_{\theta_j = \hat{\theta}_j} = 0; \quad j = \overline{1, m}. \quad (2.13)$$

Очевидно, что получаемые оценки значений  $\xi_i$  должны принадлежать области неопределенности  $D_i$  измеренных величин  $x_i$ , т. е.  $\hat{\xi}_i \in D_i$ . Когда известен закон распределения погрешности измерения  $x_i$ , это условие может быть выражено в более конкретной форме: при нормальном законе распределения случайной величины  $x_i$  —  $|x_i - \xi_i| \leq k\sigma(x_i)$ , где значение коэффициента  $k$  определяется выбранным уровнем доверия.

Таким образом, например, решение задачи минимизации функционала (2.9) при условии (2.4) эквивалентно решению системы уравнений

$$\sum_{i=1}^n \frac{y_i - \eta_i}{\sigma^2(y_i)} \frac{\partial f}{\partial \theta_j} = 0; \quad j = \overline{1, m}; \quad i = \overline{1, n}; \quad n \geq m \quad (2.14)$$

при

$$\frac{x_i - \xi_i}{\sigma^2(x_i)} + \frac{y_i - \eta_i}{\sigma^2(y_i)} \cdot \frac{\partial f}{\partial \xi_i} = 0. \quad (2.15)$$

Для функций, линейных по параметрам  $\theta$ , система уравнений (2.14) — это система линейных алгебраических уравнений.

Система (2.15) для линейных по  $\xi$  функций представляет  $n$  не связанных между собой систем из  $m$  линейных уравнений.

Замена нормального закона распределения другим, например законом Пуассона или Лапласа и т. д., приводит к системе нелинейных уравнений (2.14).

Учет корреляции погрешностей  $x_i$  и  $y_i$  не имеет принципиальных трудностей, но усложняет формулы; учет корреляции погрешности значений  $x_{i1}$  и  $x_{i2}$  в разных точках не позволяет разбить систему из  $m \cdot n$  — уравнений на  $n$  независимых; в этом случае придется решать систему линейных уравнений в  $n$  раз большей размерности, чем при независимых переменных  $x_{i1}$  и  $x_{i2}$ .

Учет корреляции погрешностей всегда ухудшает обусловленность системы уравнений (2.14).

Условие (2.15) для нелинейных функций  $\eta = f(\xi, \theta)$  дает соответственно  $n$  независимых систем из  $m$  нелинейных уравнений, которые можно решать методом линеаризации. Сходимость итерационного процесса при этом обеспечивается малостью допустимых интервалов для  $\xi$ .

## 2.2. О единственности и состоятельности оценок в конфлюентном анализе. Алгоритм получения оценок

Самый простой функционал (2.9) нелинеен по  $\theta$  даже в случае оценки параметров прямой линии. Поэтому встает вопрос о единственности оценок искомых параметров, который тесно связан с выбором нулевого приближения решения.

Из условия (2.15) следует, что для одной и той же задачи минимальное значение функционала метода наименьших квадратов больше минимального значения функционала (2.9), поскольку условие (2.15) — это условие ортогональности вектора касательной к функции  $f(\xi, \theta)$  в точке, принимаемой за истинную, и вектора, проведенного через две точки: исходную  $(x_{i1}, x_{i2}, \dots, x_{im}, y_i)$ , и принимаемую за истинную. Таким образом, минимизация функционала (2.9) при условии (2.15) приводит к минимизации суммы квадратов наикратчайших расстояний от точки  $(x_{i1}, \dots, x_{im}, y_i)$  до кривой, тогда как в методе наименьших квадратов минимизируется сумма квадратов отклонений при фиксированных значениях  $x_{ij}$ .

Для конкретных видов функции  $f(\xi, \theta)$  можно получить условия существования единственного решения.

В работе [4] устанавливается существование единственного решения для линейных функций  $\eta = \sum_{j=1}^m \theta_j x_j + \theta_0$  в предположении, что

1) погрешности  $\varepsilon$  и  $\delta$  случайных величин  $x$  и  $y$  статистически независимы и подчиняются нормальному закону распределения с известными дисперсиями; 2) за нулевое приближение взято решение регрессионной задачи.

При этих предположениях достаточным условием единственности оценок параметров линейной функции  $\eta = \theta_0 + \sum_{j=1}^m \theta_j x_j$  является

выполнение неравенства  $\bar{x}_i^2 > \left(1 - \frac{m}{n}\right) \hat{\sigma}^2(y)$ , где  $\bar{x}_i^2 = \frac{1}{n} \sum_{i=1}^n x_{ij}^2$ .

Здесь  $\hat{\sigma}^2(y)$  — оценка дисперсии значений  $y_i$  при равноточных измерениях. Кроме того, в функционале (2.9)  $\sigma(x_{ij})$  и  $\sigma(y_i)$  приняты равными единице.

Функционал (2.9) с учетом условия (2.15) для линейных функций имеет вид

$$F = \frac{1}{2} \sum_{i=1}^n \frac{\left(y_i - \theta_0 - \sum_{j=1}^m \theta_j x_{ij}\right)^2}{1 + \sum_{j=1}^m \theta_j^2}. \quad (2.16)$$

При фиксированных  $\theta_j (j = \overline{1, m})$  функционал (2.16) относительно переменной  $\theta_0$  есть парабола и любое начальное значение  $\theta_{00}$  приведет к единственной оценке. С другими переменными  $\theta_j$  дело обстоит сложнее. Функционал (2.16) монотонно растет с увеличением  $(\theta_j - \hat{\theta}_j)$  от точки минимума вправо до горизонтальной асимптоты, равной  $\sum_{i=1}^n x_{ij}^2$ , а влево — монотонно растет до точки максимума, затем монотонно опускается сверху до той же горизонтальной асимптоты.

Можно сформулировать и более строгие условия, когда точка минимума  $\Phi_p$  регрессионного функционала (метода наименьших квадратов) будет лежать под асимптотой и в то же время принадлежать строго выпуклой области функционала (2.16).

Там же в [4] показано, что достаточным условием того, что регрессионное решение при оценке параметров прямой линии  $\eta = \theta_0 + \theta_1 \xi$  принадлежит строго выпуклой области функционала (2.16), является выполнение неравенства

$$\sum_i x_i^2 - \frac{1}{n} \left( \sum_i x_i \right)^2 > (n-2) \hat{\sigma}^2(y).$$

Условия, обеспечивающие попадание регрессионного решения в выпуклую область функционала (2.9), для других функций несложно проверить на ЭВМ, но достаточно трудно выразить в явном виде.

Для функций  $\eta = f(\vec{\xi}, \vec{\theta})$ , линейных по параметрам  $\vec{\theta}$ , единственность оценки и сходимости нулевого приближения к оцениваемым параметрам при нормально распределенных исходных данных определяется только способом нахождения значений  $\vec{\xi}$ , так как при фиксированных значениях  $\vec{\xi}$  (2.14) — система линейных алгебраических уравнений с детерминантом, не равным нулю. Для этих функций из условия (2.15), в принципе, может быть найден не единственный набор значений  $\vec{\xi}$ , каждый из которых принадлежит области возможных значений при заданном наборе значений  $\{x_i\}$ . В такой ситуации у нас нет оснований отдать предпочтение какому-либо набору значений  $\vec{\xi}$  и, строго говоря, решение будет не единственным.

Если оценки параметров прямой  $\eta = \theta_0 + \theta_1 \xi$  находятся как точка минимума функционала (2.16), то они состоятельны.

Когда в качестве  $\xi_i$  берутся значения  $x_i$ , то оценка  $\theta_1$  не будет состоятельной. Для иллюстрации возьмем случай, когда  $\hat{\theta}_0 = 0$ .

Тогда

$$\hat{\theta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \frac{s_{xy}}{s_x^2} \xrightarrow{n \rightarrow \infty} \frac{\theta_1 s_\xi^2}{s_\xi^2 + 1} = \theta_1 \left( 1 - \frac{1}{s_\xi^2 + 1} \right),$$

т. е. оценка несостоятельна.

Если оценки параметров  $\vec{\theta}$  в уравнении  $\eta = \theta_0 + \sum_{j=1}^m \theta_j \xi_j$  находятся как точки минимума функционала (2.9) и оценки  $\xi$  находятся из условий (2.15), то полученные оценки будут состоятельными [4].

По аналогии с МНК при анализе систем, когда закон распределения исходных данных неизвестен, оценки параметров моделей ищутся как точка минимума функционала (2.9) [4, 6].

Учитывая свойства функционала  $F$  и получающихся оценок, задачу минимизации функционала  $F$  при условии (2.12) будем решать по определенной схеме:

1-й шаг: решается регрессионная задача поиска минимума по  $\vec{\theta}$  функционала  $F$  при значениях  $\vec{\xi} = \vec{x}$ . Получают первое приближение для оценок искомых параметров  $\vec{\theta}$ ;

2-й шаг: учитывая условие (2.12) вычисляются значения переменных  $\vec{\xi}$  при полученных оценках параметров  $\vec{\theta}$ . Проверяется принадлежность нового значения  $\hat{\xi}_{ij}$  области возможных значений при заданном значении  $x_{ij}$ ;

3-й шаг: минимизируется функционал  $F$  по  $\vec{\theta}$  при новых точных значениях переменной  $\vec{\xi}$ .

Шаги 2 и 3 выполняются поочередно до тех пор, пока не будет выполнено хотя бы одно из трех условий:

а) на очередном шаге значение функционала (2.9) меньше заданного числа  $\gamma$ ;

б) на соседних итерациях значения функционала  $F$  и значения оценок параметров  $\vec{\theta}$  отличаются несущественно, т. е.

$$\left| \frac{F_v - F_{v+1}}{F_v} \right| \leq \gamma_1; \quad \max \left| \frac{\theta_j^v - \theta_j^{v+1}}{\theta_j^v} \right| \leq \gamma_2; \quad j = \overline{1, m},$$

где  $\gamma_1$  и  $\gamma_2$  — заданные числа;

в) исчерпан заданный лимит итераций.

4-й шаг: после нахождения оценок  $\vec{\theta}$  определяются их дисперсии.

### 2.3. Оценка параметров прямой линии и ее интервальных границ

Работы, посвященные оценке свободных параметров прямой линии с учетом погрешности по осям абсцисс и ординат, известны давно [10, 17].

Рассмотрим задачу определения свободных параметров прямой линии

$$\eta = \theta_1 + \theta_2 \xi \quad (2.17)$$

и покажем основные особенности задач конфлюентного анализа.

При определении значений свободных параметров  $\theta_1$  и  $\theta_2$  прямую линию часто записывают в виде  $\eta = \theta_1^1 + \theta_2 (\xi - \bar{\xi})$ . В этом случае оценки  $\theta_1^1$  и  $\theta_2$  параметров можно получить, не решая совместные системы связанных уравнений, как это приходится делать, если линейная модель записана в виде (2.17). Оценки  $\hat{\theta}_1^1$  и  $\hat{\theta}_2$  статически независимы, что несправедливо для оценок  $\hat{\theta}_1$  и  $\hat{\theta}_2$ .

Для нормально распределенных статистически независимых погрешностей экспериментальных точек  $(x_i, y_i)$  минимизируемый функционал имеет следующий вид:

$$F = \sum_{i=1}^n \left[ \frac{(x_i - \xi_i)^2}{\sigma^2(x_i)} + \frac{(y_i - \eta_i)^2}{\sigma^2(y_i)} \right].$$

Из условия (2.12) и (2.15) получим

$$\xi_i = \frac{\sigma^2(y_i) x_i + \theta_2 \sigma^2(x_i) (y_i - \theta_1)}{\sigma^2(y_i) + \theta_2^2 \sigma^2(x_i)}.$$

Тогда

$$F = \sum_{i=1}^n \frac{(y_i - \theta_1 - \theta_2 x_i)^2}{\sigma^2(y_i) + \sigma^2(x_i) \theta_2^2} = \sum_{i=1}^n \omega_i (y_i - \theta_1 - \theta_2 x_i)^2, \quad (2.18)$$

где  $\omega_i^{-1} = \sigma^2(y_i) + \theta_2^2 \sigma^2(x_i)$ . Введем обозначения [18]:

$$\begin{aligned} \bar{x} &= \sum_i \omega_i x_i / \sum_i \omega_i; \quad x_i' = x_i - \bar{x}; \\ \bar{y} &= \sum_i \omega_i y_i / \sum_i \omega_i; \quad y_i' = y_i - \bar{y}. \end{aligned} \quad (2.19)$$

Тогда  $F = \sum_i \omega_i (\theta_2 x_i' - y_i')^2$ ;  $\bar{y} = \theta_1 + \theta_2 \bar{x}$ .

Условие  $\frac{\partial F}{\partial \theta_2} = 0$  дает

$$\sum_i \omega_i x_i' (\theta_2 \cdot x_i' - y_i') - \omega_i^2 \theta_2 \sigma(x_i)^2 (\theta_2 x_i' - y_i')^2 = 0. \quad (2.20)$$

Уравнение (2.20) может быть решено итерациями. Но нетрудно убедиться, что не всякое нулевое приближение приведет к реше-

нию задачи (как и при непосредственном отыскании точки минимума функции (2.18)).

Так как переменная  $\theta_2$  входит в выражение для  $\omega_i$ , то равенство (2.18) может быть представлено приближенно в виде полинома любой степени относительно  $\theta_2$ , и в работе [18] для нахождения оценки  $\hat{\theta}_2$  получено уравнение

$$\begin{aligned} \theta_2 \sum_i \omega_i [\sigma^2(y_i) x_i' + \theta_2 \sigma^2(x_i) y_i'] x_i' &= \sum_i \omega_i^2 [\sigma^2(y_i) x_i' + \\ &+ \theta_2 \sigma^2(x_i) y_i'] y_i' = 0, \end{aligned}$$

которое обеспечивает сходимость итерационного процесса при любом нулевом приближении  $\theta_2$  и используется в дальнейшем для оценки  $\theta_2$ . При известном  $\theta_2$  из равенства (2.19) находится оценка  $\theta_1$ .

В предположении справедливости квадратичной аппроксимации функции (2.18) дисперсии оценок  $\hat{\theta}_1$  и  $\hat{\theta}_2$  находятся по следующим формулам [18]:

$$D(\hat{\theta}_2) = Q^2 \sum_i \omega_i [x_i'^2 \sigma^2(y_i) + y_i'^2 \sigma^2(x_i)];$$

$$D(\hat{\theta}_1) = (\sum_i \omega_i)^{-1} + 2(\bar{x} + 2\bar{z}) \bar{z} Q + (\bar{x} + 2\bar{z})^2 D(\hat{\theta}_2),$$

где

$$Q^{-1} = \sum_i \omega_i [x_i' y_i / \hat{\theta}_2 + 4 z_i' (z_i - x_i')];$$

$$z_i' = z_i - \bar{z}; \quad \bar{z} = \sum_i \omega_i z_i / \sum_i \omega_i;$$

$$z_i = \omega_i [\sigma^2(y_i) \cdot x_i' + \hat{\theta}_2 \sigma^2(x_i) \cdot y_i'].$$

Ковариационную матрицу оценок  $D(\hat{\theta})$  можно найти как матрицу  $M^{-1}$ , обратную матрице  $M$ , элементы которой  $-\frac{\partial^2 F}{\partial \theta_1^2}$ ;  $-\frac{\partial^2 F}{\partial \theta_1 \partial \theta_2}$ ;  $-\frac{\partial^2 F}{\partial \theta_2^2}$ , вычисленные при найденных значениях оценок  $\hat{\theta}_1$  и  $\hat{\theta}_2$ .

После определения оценок  $\hat{\theta}_1$  и  $\hat{\theta}_2$  и их дисперсий  $D(\hat{\theta}_1)$  и  $D(\hat{\theta}_2)$  приступим к нахождению доверительных интервалов. Если число экспериментальных точек достаточно велико и можно считать, что оценки параметров распределены относительно их математических ожиданий по нормальному закону, то для получения доверительных интервалов применяют безразмерную  $t$ -статистику Стьюдента, которая подчиняется  $t$ -распределению с  $\nu = n - 2$  степенями свободы.

Пусть уровень значимости выбран  $\alpha$ . Тогда

$$t = \frac{\hat{\theta}_1 - \theta_1}{\sqrt{D(\hat{\theta}_1)}}$$

и  $100(1-\alpha)$  — процентный доверительный интервал для  $\theta_1$

$$\hat{\theta}_1 - t_{1-\frac{\alpha}{2}} \sqrt{D(\hat{\theta}_1)} \leq \theta_1 \leq \hat{\theta}_1 + t_{1-\frac{\alpha}{2}} \sqrt{D(\hat{\theta}_1)}; \quad (2.21)$$

доверительный интервал для  $\theta_2$

$$\hat{\theta}_2 - t_{1-\frac{\alpha}{2}} \sqrt{D(\hat{\theta}_2)} \leq \theta_2 \leq \hat{\theta}_2 + t_{1-\frac{\alpha}{2}} \sqrt{D(\hat{\theta}_2)}.$$

Как следует из формулы (2.21), интервальные границы в случае прямой линии для  $M(\eta/\xi_0)$  будут

$$\hat{\eta} - t_{1-\frac{\alpha}{2}} \sigma(\hat{\eta}) \leq M(\eta/\xi_0) \leq \hat{\eta} + t_{1-\frac{\alpha}{2}} \sigma(\hat{\eta}), \quad (2.22)$$

где

$$\hat{\eta} = \hat{\theta}_1 + \hat{\theta}_2 \xi_0; \quad \sigma^2(\hat{\eta}) = (1; \xi_0) {}^T D(\hat{\theta}) (1; \xi_0).$$

Огибающая семейства всех возможных прямых будет кривой второго порядка, уравнение которой получим из формулы (2.22) [10]:

$$\sum_{i=1}^n (y_i - \theta_1 - \theta_2 x_i)^2 \leq c_\gamma (1 + \theta_2^2). \quad (2.23)$$

Принимая за начало отсчета наблюдаемые средние значения  $\bar{x}$  и  $\bar{y}$ , перепишем (2.23) в виде

$$\theta_2^2 (s_x^2 - c_\gamma) - 2\theta_2 s_{xy} + \theta_1^2 \leq c_\gamma - s_y^2. \quad (2.24)$$

Здесь  $s_x^2$ ,  $s_y^2$  и  $s_{xy}^2$  — соответственно выборочные дисперсии  $x$ ,  $y$  и выборочная ковариация величин  $x$  и  $y$ .

Условие (2.24) можно считать ограничением, которому удовлетворяет истинная прямая. Отсюда найдем огибающую семейства всех возможных прямых. Уравнение огибающей

$$\frac{(y - \hat{\theta}_2 x)^2}{c_\gamma - b_1} - \frac{(\hat{\theta}_2 y + x)^2}{b_2 - c_\gamma} = 1 + \hat{\theta}_2^2,$$

где

$$b_1 = s_x^2 - \frac{s_{xy}}{\hat{\theta}_2}; \quad b_2 = s_x^2 + \theta_2 s_{xy}; \quad \hat{\theta}_2 > 0; \quad b_2 > b_1.$$

Программы определения свободных параметров прямой линии часто применяются для оценки параметров функции, которую при помощи некоторых преобразований привели к прямой линии.

Но при такой обработке нелинейных функций следует иметь в виду, что для того, чтобы оценки  $\hat{\theta}$  соответствующих параметров  $\theta$ , полученные из преобразованного уравнения, обладали оптимальными свойствами (несмещенностью, минимальной дисперсией

и т. д.) необходимо, чтобы предположение об аддитивной ненаблюдаемой случайной ошибке было справедливым для преобразованной, а не для первоначальной модели.

Влияние аддитивной ошибки в преобразованной и исходных моделях можно установить, только исследуя каждую конкретную модель.

Иногда на практике можно пренебречь неоптимальностью оценок, например при многократных однотипных исследованиях. Учитывая, что заложенные в расчеты предположения могут быть не абсолютно точны, при относительных измерениях важно «одинаково ошибаться» каждый раз. Но этот прием неприменим, если необходимы абсолютные значения параметров для непосредственного сравнения результатов расчетов и экспериментов.

Важно и то, что разные алгоритмы приводят к различным интервальным оценкам. Наиболее оптимальные оценки должны получаться в тех алгоритмах, в которых операции ведутся непосредственно с исходными экспериментальными данными, когда не нарушается принцип аддитивности помехи и не изменяется существенно закон ее распределения.

Большое практическое значение имеет положение экспериментальных точек на исследуемом интервале возможных значений. Оптимальный закон их распределения (при минимуме дисперсии оценок) зависит от конкретных значений экспериментальных величин и их погрешностей, и без упрощающих предположений о последних получить его в явном виде не удастся.

Исследовалась также возможность при оценке свободных параметров экспоненты использовать минимизируемый функционал, в котором в качестве математических ожиданий временных координат точек брались значения, получающиеся из обратных соотношений. Проведенный анализ показал, что в целом ряде случаев смещение оценок незначительно (находится в пределах погрешности оценок).

#### 2.4. Оценка параметров линейных функций (и решение систем линейных алгебраических уравнений)

Рассмотрим применение описанного подхода к задаче определения оценок параметров линейной модели

$$\vec{\eta} = f(\vec{\xi}, \vec{\theta}) = \sum_{i=1}^n \theta_i \xi_i$$

и к решению систем линейных алгебраических уравнений с погрешностями в матрице системы и в правой части уравнений,

если ошибки измерений суть независимые нормально распределенные случайные величины с нулевыми средними значениями и известными дисперсиями  $\sigma^2(x_{ij})$  и  $\sigma^2(y_j)$ .

В этом случае функционал (2.9) будет иметь вид

$$F = \frac{1}{2} \sum_{j=1}^m \left\{ \sum_{i=1}^n \frac{(x_{ij} - \xi_{ij})^2}{\sigma^2(x_{ij})} + \frac{(y_j - \sum_i \theta_i \xi_{ij})^2}{\sigma^2(y_j)} \right\}, \quad (2.25)$$

а ограничение  $\hat{\xi}_i \in D_i$  можно записать следующим образом:

$$|x_{ij} - \hat{\xi}_{ij}| \leq 3\sigma(x_{ij}). \quad (2.26)$$

Здесь для упрощения вида функционала (2.25) ошибки измерений считаются статистически независимыми.

Задачу минимизации по  $\vec{\theta}$  функционала (2.25) при фиксированных значениях  $\vec{\xi}$  можно рассматривать как обычную задачу регрессионного анализа. Другим способом является непосредственная минимизация функционала (2.25). Исследования на ЭВМ показали, что для обеспечения минимального времени счета и большей точности лучше всего на первом шаге решать систему линейных уравнений при  $\xi_{ij} = x_{ij}$ , методом Гаусса с выбором максимального элемента. А в дальнейшем непосредственно минимизировать по  $\vec{\theta}$  функционал (2.25) при помощи метода сопряжения градиентов [4].

Учитывая вид функционала (2.25), задачу минимизации по  $\vec{\theta}$  легко свести к задаче поиска минимума квадратичной формы

$$G(\vec{\theta}) = \frac{1}{2} \vec{\theta}^T \vec{A} \vec{\theta} + \vec{a}^T \vec{\theta},$$

которая и решалась методом сопряженных градиентов. Здесь  $\vec{\theta}$  — вектор искомых оценок параметров;  $\vec{A}$  — матрица квадратичной формы, элементы которой вычисляются по формуле

$$A_{rp} = \sum_{j=1}^m \frac{1}{\sigma^2(y_j)} x_{rj} x_{pj}; \quad r = \overline{1, n}; \quad p = \overline{1, n}$$

( $x_{rj}$  — значение  $r$ -й переменной, полученное в  $j$ -м измерении);

$\vec{a}^T$  — вектор, компоненты которого равны

$$a_r^T = - \sum_{j=1}^m \frac{y_j}{\sigma^2(y_j)} \cdot x_{rj}; \quad r = \overline{1, n}.$$

Пересчет точных значений  $\vec{\xi}$  на основании условий (2.15) сводится к решению  $m$  несвязанных между собой систем из  $n$  линейных уравнений вида

$$\sum_{r=1}^n \frac{\theta_r \theta_p}{\sigma^2(y_j)} \xi_{rj} + \frac{\xi_{pj}}{\sigma^2(x_{pj})} = \frac{x_{pj}}{\sigma^2(x_{pj})} + \frac{\theta_p \cdot y_j}{\sigma^2(y_j)}; \quad p = \overline{1, n}; \quad j = \overline{1, m}.$$

Полученные новые значения  $\vec{\xi}$  должны удовлетворять условию (2.26). В противном случае те  $\xi_{ij}$ , которые выходят за указанные границы, заменяются на значение ближайшей граничной точки. В связи с этим иногда можно ожидать увеличения функционала (2.25) на новых точных значениях переменных по сравнению с предыдущим шагом итерационного процесса. Это приводит к снижению скорости сходимости процесса и даже к возникновению колебаний. Для устранения таких нежелательных последствий после пересчета  $\vec{\xi}$  те наборы  $\xi_j$ , на которых произошло увеличение соответствующих слагаемых функционала

$$F_j = \sum_{i=1}^n \frac{(x_{ij} - \xi_{ij})^2}{\sigma^2(x_{ij})} + \frac{[y_j - f(\xi_j; \vec{\theta})]^2}{\sigma^2(y_j)}$$

по сравнению с предыдущей итерацией, заменяются значениями  $\vec{\xi}$  с предыдущего шага.

Применение описанного приема позволило в большинстве случаев резко увеличить скорость сходимости процесса (примеры, для решения которых не хватало 30 итераций, сошлись за 7 ... 10 итераций). Когда же первоначальный процесс не сходился (возникали колебания), решение достигалось за 12 ... 17 итераций. Элементы дисперсионной матрицы ошибок (матрицы рассеивания) для оценок искомых параметров подсчитываются как элементы матрицы, обратной матрице  $M$ :

$$M_{ij} = - \left. \frac{\partial^2 F}{\partial \theta_i \partial \theta_j} \right|_{\vec{\theta} = \vec{\hat{\theta}}},$$

где  $\vec{\hat{\theta}}$  — полученные оценки параметров  $\vec{\theta}$ .

Описанную выше задачу можно интерпретировать как задачу решения переопределенной системы из  $m$  линейных уравнений с  $n$  переменными ( $m > n$ ):

$$\vec{Y} = \vec{X} \vec{\theta},$$

где  $\vec{Y} = (y_1, \dots, y_m)^T$  — столбец значений функции;  $\vec{X}$  — матрица размерностью  $(m \times n)$ , элементы которой  $x_{ij}$  суть значения  $i$ -й переменной в  $j$ -м измерении;  $\vec{\theta} = (\theta_1, \dots, \theta_n)^T$  — столбец неизвестных (искомых) параметров.

## 2.5. Оценка значений параметров в сигноме

Рассмотрим задачу оценки параметров  $\vec{\theta}$  сигнома

$$\eta = \sum_{l=1}^k \theta_l \prod_{j=1}^m x_{ij}^{\alpha_{jl}}$$

при условии, что все ошибки измерений — это независимые нормально распределенные случайные величины с нулевыми средними и известными дисперсиями  $\sigma^2(x_{ij})$  и  $\sigma^2(y_i)$ .

Функционал (2.9) будет иметь вид

$$F = \frac{1}{2} \sum_{i=1}^n \left\{ \sum_{j=1}^m \frac{(x_{ij} - \xi_{ij})^2}{\sigma^2(x_{ij})} + \frac{[y_i - f(\vec{\xi}_i; \vec{\theta})]^2}{\sigma^2(y_i)} \right\}. \quad (2.27)$$

Ограничение  $\vec{\xi}_i \in D_i$  можно записать следующим образом:

$$|x_{ij} - \xi_{ij}| \leq 3\sigma(x_{ij}). \quad (2.28)$$

Чтобы упростить вид функционала (2.27), ошибки измерений считаем статистически независимыми.

Задачу минимизации по  $\vec{\theta}$  сводим к задаче поиска минимума квадратной формы:

$$G(\vec{\theta}) = \frac{1}{2} \vec{\theta} \vec{A} \vec{\theta} + \vec{a}^T \vec{\theta}. \quad (2.29)$$

Точка минимума  $G(\vec{\theta})$  определялась методом сопряженных градиентов.

В выражении (2.29)  $\vec{A}$  — матрица квадратичной формы, элементы которой имеют вид

$$A_{rp} = \sum_{i=1}^n \frac{1}{\sigma^2(y_i)} \prod_{j=1}^m x_{ij}^{\alpha_{jr} + \alpha_{jp}}; \quad r = \overline{1, k}; \quad p = \overline{1, k}$$

$x_{ij}$  — значение  $i$ -й переменной, полученное в  $j$ -м измерении;

$\vec{a}^T$  — вектор, компоненты которого равны

$$a_r^T = - \sum_{i=1}^n \frac{y_i}{\sigma^2(y_i)} \prod_{j=1}^m x_{ij}^{\alpha_{jr}}; \quad r = \overline{1, k}.$$

Пересчет точных значений  $\vec{\xi}$  с учетом условий (2.15) сводится к решению  $n$  несвязанных между собой систем из  $m$  нелинейных уравнений вида

$$\frac{x_{ij} - \xi_{ij}}{\sigma^2(x_{ij})} + \frac{y_i - f(\vec{\xi}_i; \vec{\theta})}{\sigma^2(y_i)} \frac{\partial f}{\partial \xi_{ij}} = 0, \quad i = \overline{1, n}; \quad j = \overline{1, m}. \quad (2.30)$$

где

$$\frac{\partial f}{\partial \xi_{ij}} = \sum_{l=1}^k \theta_l \alpha_{jl} \cdot x_{i1}^{\alpha_{1l}} \dots x_{ij}^{\alpha_{jl} - 1} \dots x_{im}^{\alpha_{ml}};$$

$\vec{\xi}_i = (\xi_{i1}, \dots, \xi_{im})$  — искомые точные значения переменных в  $i$ -м измерении.

Система уравнений (2.30) решалась методом итераций с использованием линеаризации исходных уравнений на каждом шаге [4]. Такое упрощение вычислений возможно потому, что на основании условия (2.25) все расчеты ведутся в небольшой области пространства переменных.

Таким образом, на каждой итерации для решения системы (2.30) решались системы линейных уравнений

$$\sum_{r=1}^m \frac{\partial \varphi_j(\xi_i)}{\partial \xi_{ir}} \Big|_{\xi_i = \xi_i^0} (\xi_{ir} - \xi_{ir}^0) = - \varphi_j(\xi_i^0); \quad i = \overline{1, n}; \quad j = \overline{1, m};$$

где  $\varphi_j(\xi_i^0)$  — значение левой части (2.30) в точке  $\xi_i^0$  (точке разложения)

$$\frac{\partial \varphi_i}{\partial \xi_{ir}} = \frac{y_i - f(\vec{\xi}_i; \vec{\theta})}{\sigma^2(y_i)} f'_{\xi_{ij} \xi_{ir}} - \frac{f'_{\xi_{ji}} f'_{\xi_{ir}}}{\sigma^2(y_i)}, \quad \text{если } r \neq j;$$

$$\frac{\partial \varphi_i}{\partial \xi_{ir}} = \frac{y_i - f(\vec{\xi}_i; \vec{\theta})}{\sigma^2(y_i)} f'_{\xi_{ij}^2} - \frac{f'_{\xi_{ij}^2}}{\sigma^2(y_i)} - \frac{1}{\sigma^2(x_{ij})}, \quad \text{если } r = j.$$

$$\text{Здесь } f'_{\xi_{ij} \xi_{ir}} = \sum_{l=1}^k \theta_l \alpha_{jl} \alpha_{rl} x_{i1}^{\alpha_{1l}} \dots x_{ij}^{\alpha_{jl} - 1} \dots x_{ir}^{\alpha_{rl} - 1} \dots x_{im}^{\alpha_{ml}};$$

$$f'_{\xi_{ij}^2} = \sum_{l=1}^k \theta_l \alpha_{jl} (\alpha_{jl} - 1) x_{i1}^{\alpha_{1l}} \dots x_{ij}^{\alpha_{jl} - 2} \dots x_{im}^{\alpha_{ml}}.$$

В связи с тем, что уравнения в системе (2.30) замещались приближенными выражениями, в некоторых случаях можно также ожидать увеличения функционала (2.27) при новых точных значениях переменных по сравнению с предыдущим шагом итерационного процесса, что приведет к снижению скорости сходимости процесса и даже к возникновению колебаний. Для устранения этих нежелательных последствий после пересчета  $\vec{\xi}$  те значения  $\vec{\xi}_i$ , на которых произошло увеличение соответствующих слагаемых функционала

$$F_i = \sum_{j=1}^m \frac{(x_{ij} - \xi_{ij})^2}{\sigma^2(x_{ij})} + \frac{[y_i - f(\vec{\xi}_i, \vec{\theta})]^2}{\sigma^2(y_i)}$$

по сравнению с предыдущей итерацией, заменялись значениями с предыдущего шага.

Элементы дисперсионной матрицы ошибок (матрицы рассеивания) для оценок искомых параметров подсчитываются как элементы матрицы, обратной матрице М:

$$M_{rp} = - \left. \frac{\partial^2 F}{\partial \theta_r \partial \theta_p} \right|_{\vec{\theta} = \vec{\hat{\theta}}}; \quad r = \overline{1, k}; \quad p = \overline{1, k},$$

где  $\vec{\hat{\theta}}$  — полученные оценки параметров  $\vec{\theta}$ .

Многочисленные вычислительные эксперименты позволяют сделать вывод о состоятельности рассматриваемых оценок.

## 2.6. Оценка параметров полиномиальной зависимости

Рассмотрим задачу оценки параметров полинома. Требуется определить оценки коэффициентов полинома степени  $m$

$$\eta = \sum_{l=0}^m \theta_l \xi^l$$

по совокупности значений аргумента и функции, содержащих погрешности измерений как в значениях функции, так и в значениях аргумента. Пусть исходные данные имеют вид

$$y_i = \eta_i + \varepsilon_i; \quad x_i = \xi_i + \delta_i; \quad i = \overline{1, n}.$$

Предполагается, что ошибки измерений  $\varepsilon$  и  $\delta$  — нормально распределенные случайные величины с нулевыми средними значениями и с известными дисперсиями (соответственно  $\sigma^2(x_i)$  и  $\sigma^2(y_i)$ ), и известными коэффициентами корреляции  $\rho_i$  между измеренными значениями  $x_i$  и  $y_i$ .

Тогда искомые оценки коэффициентов  $\vec{\theta}$  получают путем минимизации функционала

$$F = \frac{1}{2} \sum_{i=1}^n \frac{1}{1 - \rho_i^2} \left[ \frac{(x_i - \xi_i)^2}{\sigma^2(x_i)} + \frac{\left( y_i - \sum_{l=0}^m \theta_l \xi_i^l \right)^2}{\sigma^2(y_i)} - 2\rho_i \frac{(x_i - \xi_i)(y_i - \eta_i)}{\sigma(x_i)\sigma(y_i)} \right] \quad (2.31)$$

при условиях

$$\left. \frac{\partial F}{\partial \xi_i} \right|_{\xi_i = \hat{\xi}_i} = 0; \quad |x_i - \hat{\xi}_i| \leq 3\sigma(x_i). \quad (2.32)$$

Пересчет значений  $\xi_i$  согласно условию (2.32) сводится к решению системы из  $m$  нелинейных уравнений вида

$$\begin{aligned} \frac{x_i - \xi_i}{\sigma^2(x_i)} - \rho_i \frac{y_i - \sum_{l=0}^m \theta_l \xi_i^l + (x_i - \xi_i) \sum_{l=0}^m l \theta_l \xi_i^{l-1}}{\sigma(x_i)\sigma(y_i)} + \\ + \frac{y_i - \sum_{l=0}^m \theta_l \xi_i^l}{\sigma^2(y_i)} \sum_{l=1}^m l \theta_l \xi_i^{l-1} = 0; \quad i = \overline{1, n} \end{aligned} \quad (2.33)$$

на отрезке  $|x_i - \xi_i| \leq 3\sigma(x_i)$ .

В процессе решений системы уравнений (2.33) может оказаться что:

1) корень уравнения (2.33) принадлежит отрезку  $|x_i \xi_i| \leq 3\sigma(x_i)$ . Значение корня находится методом хорд;

2) корень уравнения (2.33) не принадлежит отрезку  $|x_i \xi_i| \leq 3\sigma(x_i)$ . Тогда с помощью алгоритма поиска глобального минимума функции находится наименьшее значение функции

$$\begin{aligned} \varphi(\xi_i) = \frac{x_i - \xi_i}{\sigma^2(x_i)} + \rho_i \frac{y_i - \sum_{l=0}^m \theta_l \xi_i^l + (x_i - \xi_i) \sum_{l=1}^m l \theta_l \xi_i^{l-1}}{\sigma(x_i)\sigma(y_i)} + \\ + \frac{y_i - \sum_{l=0}^m \theta_l \xi_i^l}{\sigma^2(y_i)} \sum_{l=1}^m l \theta_l \xi_i^{l-1}, \end{aligned}$$

принадлежащее отрезку  $|x_i - \xi_i| \leq 3\sigma(x_i)$ .

Поставленная задача минимизации функционала (2.31) решается по итерационной схеме, изложенной в п. 2.2.

Дисперсии полученных оценок находились с помощью матрицы вторых производных функционала (2.31). Процедура нахождения оценок и их дисперсий аналогична ранее описанной в п. 2.4, 2.5.

**Пример.** Найти оценки параметров  $\vec{\theta}$  в модели

$$\eta = \theta_1 \xi_1 \xi_2^3 + \theta_2 \xi_1^2 \xi_2^2 + \theta_3 \xi_1^3.$$

Предварительно были взяты значения  $\theta_1 = 1$ ;  $\theta_2 = -1$ ;  $\theta_3 = 2$ ;  $\vec{\xi}_i = \{1, 1, 2, 2, 3, 3\}$ ;  $\vec{\xi}_2 = \{1, 2, 1, 2, 1, 2\}$  и рассчитаны значения  $\vec{\eta} = \{2, 6, 14, 16, 48, 42\}$ .



С помощью датчика случайных чисел на точные данные накладывались помехи с математическим ожиданием, равным нулю, и дисперсиями для  $\xi_1 = 0,020$ ; для  $\xi_2 = 0,015$ ; для  $\eta = 0,1$ . Получены следующие реализации:

$$\vec{y} = \{2,2057; 6,1090; 14,0602; 15,7987; 47,8996; 42,1214\};$$

$$\vec{x}_1 = \{1,0070; 0,9604; 2,0123; 2,0427; 3,0027; 2,9781\};$$

$$\vec{x}_2 = \{1,0011; 2,0159; 1,0142; 1,9916; 1,0233; 1,9861\}.$$

Результат решения регрессионной задачи

$$\hat{\theta}_{1p} = 0,895; \hat{\theta}_{2p} = -0,893; \hat{\theta}_{3p} = 1,975.$$

Сумма квадратов невязок  $\sum_{i=1}^6 (y_i - f(\vec{x}_i, \hat{\theta}_p))^2$  равна 0,0232.

Результат решения конъюгентной задачи:  $\hat{\theta}_1 = 0,984$ ,  $\hat{\theta}_2 = -0,979$ ,  $\hat{\theta}_3 = 1,999$ . Сумма квадратов невязок равна 0,0007. Матрица рассеяния имеет вид

$$\begin{pmatrix} 0,00025 & -0,00022 & -0,00002 \\ -0,00022 & 0,00023 & -0,00004 \\ -0,00002 & -0,00004 & 0,00004 \end{pmatrix}.$$

## 2.7. Непосредственное определение оценок параметров экспоненты и функции Гаусса

Оценки параметров экспоненты и функции Гаусса могут быть определены после преобразований, приводящих исходную информацию к прямой линии.

Подобный способ не всегда является желательным по следующим причинам:

1) искажаются функции плотностей вероятности исходных данных, что требует изменения алгоритмов оценок;

2) не всегда преобразования обеспечивают однозначность физической интерпретации результатов. Так, при обработке пиков полного поглощения в гамма-спектрометрии оценка дисперсии функции Гаусса в ряде работ определяется угловым коэффициентом прямой линии, проведенной по точкам  $(x_i; \sqrt{\ln y_{\max}/y_i})$ . Здесь  $(x_i, y_i)$  — координаты  $i$ -й экспериментальной точки;  $y_{\max}$  — амплитуда пика полного поглощения. Строго говоря, оценка дисперсии таким путем получается смещенной: высокоэнергетический спад пика полного поглощения менее искажен, чем низкоэнергетический и вершина пика полного поглощения может не описываться функ-

цией Гаусса с теми же параметрами, что и остальная его часть. Расчеты показывают, что ошибка в  $y_{\max}$  на  $n$  процентов для идеального гауссиона приводит к погрешности в определении полуширины пика полного поглощения на  $2n$ ;

3) при обработке экспериментальных данных, поступающих с дискретных цифровых регистраторов, значение  $y_{\max}$  для функции Гаусса может быть просто неизвестно и выполнить указанное преобразование невозможно.

Параметры экспоненты  $\eta = A \exp(\lambda \xi)$  оцениваются путем нахождения координат точки минимума функционала

$$F = \frac{1}{2} \sum_{i=1}^n \left[ \frac{(x_i - \xi_i)^2}{\sigma^2(x_i)} + \frac{(y_i - \eta_i)^2}{\sigma^2(y_i)} \right] \quad (2.34)$$

при дополнительном условии

$$\frac{x_i - \xi_i}{\sigma^2(x_i)} + \frac{y_i - \eta_i}{\sigma^2(y_i)} A \lambda \exp(\lambda \xi_i) = 0.$$

За оценки параметров  $A, \sigma, \xi_0$  функции Гаусса

$$\eta = \frac{A}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\xi - \xi_0)^2}{2\sigma^2}\right]$$

принимались координаты точки минимума функционала (2.34), а оценки истинных значений определялись следующим образом:

$$\hat{\eta}_i = \frac{A}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i - x_0)^2}{2\sigma^2}\right]; \quad (2.35)$$

$$\hat{\xi}_i = x_0 + \sigma \sqrt{2 \ln \frac{A}{y_i \sqrt{2\pi}\sigma}}. \quad (2.36)$$

Условия (2.35) и (2.36) позволили значительно упростить вычислительный процесс отыскания оценок.

В процессе вычисления оценок параметров функции Гаусса и экспоненты возникают определенные трудности: функционал (2.34) имеет сложный вид, содержит локальные экстремумы и континуумы, велико взаимодействие искомых параметров.

Отыскание точки минимума функционала (2.34) при приведенных условиях проводилось методом сопряженных градиентов.

Вычислительные эксперименты показали, что в практически реальных случаях итерационный процесс отыскания точки минимума функционала (2.34) сходится, когда нулевые приближения отличаются от точных значений до 40%. Подобной точности можно добиться, если предварительно провести оценку искомых параметров экспоненты и функции Гаусса по преобразованным (линеаризованным) данным. Полученные при этом оценки затем ис-

пользуются как нулевые приближения для более строгого алгоритма.

Дисперсия получаемых оценок определялась с помощью матрицы вторых производных функционала (2.34), вычисленной при найденных оценках параметров.

Приведенные соображения остаются в силе и в том случае, когда экспериментальные данные — интегральные значения функции Гаусса или экспоненты, полученные на интервале  $(x_i, x_{i+1})$ ;  $i = \overline{1, n-1}$ .

## 2.8. Анализ систем в активном эксперименте

Пусть проводится эксперимент, ставящий своей целью определение свободных параметров модели (функции)  $\vec{\eta} = f(\vec{\xi}, \vec{\theta})$ , описывающей поведение системы. Причем мы можем заставить  $\vec{x}$  принимать любые заданные значения и затем измерить соответствующие значения  $\vec{y}$ , которые можно рассматривать как результат влияния случайной ошибки  $\vec{\varepsilon}$  на истинное значение  $\vec{\eta}$ :

$$\vec{y} = \vec{\eta} + \vec{\varepsilon}. \quad (2.37)$$

Однако и на значения  $\vec{x}$  тоже может накладываться ошибка  $\vec{\delta}$ :

$$\vec{\xi} = \vec{x} + \vec{\delta}, \quad (2.38)$$

где  $\vec{x}$  называется контролируемой переменной, т. е. ее значение фиксированно заранее, тогда как неизвестное истинное значение  $\vec{\xi}$  случайно колеблется.

Пусть ошибки  $\vec{\delta}$  имеют нулевое среднее значение. Тогда среднее значение  $\vec{\xi}_i$  будет равно  $x_i$ , коэффициент корреляции между  $\vec{\xi}_i$  и  $\delta_i$  равен (+1) и

$$\vec{y} = f(\vec{x} + \vec{\delta}, \vec{\theta}) + \vec{\varepsilon}. \quad (2.39)$$

Поскольку  $\vec{x}$  теперь уже не является случайной величиной, то ни  $\vec{\varepsilon}$  ни  $\vec{\delta}$  не коррелированы с  $\vec{x}$ . Таким образом, выражение (2.39) представляет собой обычное уравнение регрессии, к которому без всяких изменений можно применять методы регрессионного анализа, например метод наименьших квадратов.

Закон распределения случайной величины  $f(\vec{x} + \vec{\delta}, \vec{\theta})$  в выражении (2.39) может быть рассчитан по формулам переноса ошибок. Затем, в соответствии с полученным законом распределения случайной величины для выражения (2.39), методом максимума правдоподобия может быть составлен функционал, координаты точки минимума которого будут определять оценки искомых параметров  $\vec{\theta}$ .

Если случайная величина в выражении (2.39) подчиняется нормальному закону распределения, то для оценки параметров  $\vec{\theta}$  получим функционал метода наименьших квадратов, но дисперсия случайной величины обязательно будет зависеть от значений  $\vec{\theta}$ , т. е. соответствующие уравнения для определения оценок  $\vec{\theta}$  будут нелинейными.

Если значения, в которых контролируются  $\vec{x}$ , сами являются случайными величинами (т. е. определяются с помощью некоторого процесса случайного выбора), полученные выше выводы остаются в силе, когда  $\vec{\delta}$  и  $\vec{\xi}$  некоррелированы с  $\vec{x}$  [10]. Предположение относительно  $\vec{\varepsilon}$  обычно выполнено, но относительно  $\vec{\delta}$  дело обстоит сложнее. Например, требование некоррелированности  $\vec{\delta}$  и  $\vec{x}$  в данном случае означает, что с большими значениями  $\vec{x}$  не должна быть связана тенденция к большим или меньшим ошибкам в определении истинного  $\vec{x}$  по сравнению с малыми  $\vec{x}$ . Выполнение этого условия может быть проверено только эмпирическим путем.

Заметим, что при неизвестной дисперсии  $\sigma^2(\delta)$  или ее оценке в активной схеме эксперимента не идентифицируемы даже свободные параметры кубической параболы [10].

Рассмотрим задачу оценки свободных параметров в уравнении прямой линии

$$\eta = \theta_1 + \theta_2 \xi.$$

При выполнении условий (2.37) и (2.38) получим

$$y_i = \theta_1 + \theta_2 x_i + \theta_2 \delta_i + \varepsilon_i.$$

Пусть независимые случайные величины  $\vec{\delta}$  и  $\vec{\varepsilon}$  подчиняются нормальному закону распределения с нулевыми математическими ожиданиями и известными дисперсиями  $\sigma^2(x_i)$  и  $\sigma^2(y_i)$ . Функции

плотности вероятности случайных величин будут иметь следующие числовые характеристики:

$$M[\theta_2 \delta_i + \varepsilon_i] = 0; \quad i = \overline{1, n};$$

$$\sigma^2[\theta_2 \delta_i + \varepsilon_i] = \theta_2^2 \sigma^2(x_i) + \sigma^2(y_i); \quad i = \overline{1, n}.$$

Функционал метода наименьших квадратов, точка минимума которого определяет оценки параметров  $\vec{\theta}$ , имеет вид

$$F = \sum_{i=1}^n \frac{(y_i - \theta_1 - \theta_2 x_i)^2}{\sigma^2(y_i) + \theta_2^2 \sigma^2(x_i)}. \quad (2.40)$$

Рассуждая аналогично, получим функционал метода наименьших квадратов для оценки параметров  $\vec{\theta}$  линейных функций:

$$\eta = \theta_0 + \sum_{j=1}^m \theta_j \xi_j;$$

$$y_i = \theta_0 + \sum_{j=1}^m \theta_j x_{ji} + \sum_{j=1}^m \theta_j \delta_{ji} + \varepsilon_i;$$

$$M \left[ \sum_{j=1}^m \theta_j \delta_{ji} + \varepsilon_i \right] = 0; \quad i = \overline{1, n};$$

$$\sigma^2 \left[ \sum_{j=1}^m \theta_j \delta_{ji} + \varepsilon_i \right] = \sigma^2(y_i) + \sum_{j=1}^m \theta_j^2 \sigma^2(x_{ji});$$

$$F = \sum_{i=1}^n \frac{\left( y_i - \sum_{j=1}^m \theta_j x_{ji} \right)^2}{\sigma^2(y_i) + \sum_{j=1}^m \theta_j^2 \sigma^2(x_{ji})}. \quad (2.41)$$

Функционалы (2.40) и (2.41) — те же, что и в пассивной схеме эксперимента.

Рассмотрим функцию

$$\eta = \theta_0 + \sum_{j=1}^m \theta_j \varphi_j(\xi).$$

Тогда

$$y_i = \theta_0 + \sum_{j=1}^m \theta_j \varphi_j(x_i + \delta_i) + \varepsilon_i. \quad (2.42)$$

Здесь не всегда можно выделить случайную компоненту, присутствующую в  $\varphi_j(x_i + \delta_i)$ .

Метод максимума правдоподобия (ММП) позволяет получить функционал, точка минимума которого дает оценки искомых параметров  $\vec{\theta}$ . Вид функционала ММП и способ получения оценок определяется конкретным видом функции  $\varphi(x)$ .

В работе [7], чтобы упростить задачу получения оценок функции (2.42),  $\varphi_j(\xi)$  раскладываются в ряд в окрестности точки  $x_i$ :

$$M[y_i] = \vec{\theta}^T \psi(x) + O[\sigma^2(x_i)^3];$$

$$\varphi_j(x_i) = \varphi_j(x_i) + \frac{1}{2} \sigma^2(x_i) \text{tr} \frac{\partial^2 \varphi_j(x, \vec{\theta})}{\partial x \partial x^T},$$

где  $\text{tr} A$  — след матрицы  $A$ ;  $M[y_i]$  — знак операции вычисления математического ожидания.

Дисперсия значений  $\varphi_j(\xi)$  принималась равной

$$M[(y_i - M[y_i])^2] = \lambda^{-1}(x_i, \vec{\theta}) + O[\sigma^3(x_i)];$$

где  $\lambda^{-1}(x_i, \vec{\theta}) = \sigma^2(y_i) + \sigma^2(x_i) \vec{\theta}^T \frac{\partial \varphi}{\partial x^T} \frac{\partial \varphi}{\partial x} \cdot \vec{\theta}$ .

Задача свелась к классической регрессионной задаче

$$\varphi_j(x) = \varphi_j(x) + \frac{1}{2} \sigma^2(x) \sum_{i=1}^n \frac{\partial^2 \varphi_j(x)}{\partial x_i^2}.$$

Оценки  $\hat{\theta}$  находятся итерациями по методу наименьших квадратов, учитывая, что

$$y_i = [\vec{\theta}^T \psi(x)]_i + \mu_i;$$

$$M[\mu_i] = 0; \quad M[\mu_i^2] = \lambda^{-1}(x_i, \vec{\theta}).$$

При нелинейной параметризации в работе [7] применяется также метод наименьших квадратов с заменой

$$\eta(\vec{x} + \vec{\delta}, \vec{\theta}) \text{ на } \eta(\vec{x}, \vec{\theta}) + \frac{1}{2} \sigma^2(\vec{x}) \text{tr} \frac{\partial^2 \eta(\vec{x}, \vec{\theta})}{\partial \vec{x} \partial \vec{x}^T};$$

$$\sigma^2(\eta(x_i, \vec{\theta})) \text{ на } \sigma^2(y_i) + \sigma^2(x_i) \frac{\partial \eta(\vec{x}, \vec{\theta})}{\partial \vec{x}} \frac{\partial \eta(\vec{x}, \vec{\theta})}{\partial \vec{x}^T}.$$

Оценка  $\hat{\theta}$  находится методом итераций.

С позиции, изложенной в работе [7], для линейных функций получим функционал метода наименьших квадратов (2.41) — тот же, что и в пассивной схеме эксперимента. Для нелинейных функций отличие оценок, полученных при обработке одних и тех же данных различными методами, будет заметным (очевидно, что экспериментальные данные должны обрабатываться тем методом, который следует из условий проведения эксперимента).

### 2.9. Плохая обусловленность и некорректность в задачах оценки параметров функции известного вида

Итерационный алгоритм решения конъюгентной задачи состоит из двух основных элементов: решение системы алгебраических уравнений для получения оценки  $\hat{\theta}$  и решение методом линеаризации систем нелинейных алгебраических уравнений для определения оценок истинных значений аргументов  $\hat{\xi}$ . На обоих этапах возможна ситуация, когда бесконечно малым приращениям в исходных данных могут отвечать сколь угодно большие изменения в решении [15]. Такие системы называются плохообусловленными, а задачи — некорректными. Приращения в исходных данных могут быть вызваны как ошибками в измерениях, так и округлением величин в процессе расчетов на ЭВМ. В плохообусловленных системах строгое математическое решение может не соответствовать физической постановке задачи. Критической величиной, которая определяет физическую надежность строгого математического решения, является отношение наибольшего собственного значения симметричной матрицы  $\vec{A}^T \vec{A}$ , где  $\vec{A}$  — матрица системы алгебраических уравнений, к наименьшему. Квадратный корень этого отношения показывает увеличение помех в направлении, соответствующем наименьшему собственному значению [14, 15].

В качестве другой характеристики обусловленности системы с квадратной матрицей  $\vec{A}$  размерностью  $n$  вводят  $N$ -обуславливающие числа,

$$N = \frac{1}{n} \|\vec{A}\| \cdot \|\vec{A}^{-1}\|,$$

где  $\|\vec{A}\|$  и  $\|\vec{A}^{-1}\|$  — соответственно нормы матрицы  $\vec{A}$  и обратной матрицы  $\vec{A}^{-1}$ .

Наилучшими обусловленными матрицами являются ортогональные, для которых  $N$ -обуславливающие числа равны единице.

В процессе обработки реальных экспериментов и аппроксимации исходных данных функциями различных видов с большим числом оцениваемых параметров в решении появляются осцилляции. Они возникают не только из-за погрешностей эксперимента, но и в результате неадекватного представления исследуемого явления выбранными функциями. С одной стороны, желательно описать как можно точнее изучаемое явление большим числом параметров, с другой стороны, увеличение размерности задачи ухудшает обусловленность систем и задача становится некорректной. Обусловленность систем линейных алгебраических уравнений еще больше ухудшается, когда мы имеем дело с коррелированными исходными случайными величинами.

Задача считается поставленной корректно, если ее решение удовлетворяет требованиям (условиям) Адамара, которые для операторного уравнения  $\vec{Y} = \vec{A}\vec{\theta}$  формулируются следующим образом [15]:

решение  $\vec{\theta}$  существует для любого  $\vec{Y} \in Q_A \subseteq F$ ;

решение  $\vec{\theta}$  единственно в пространстве  $U$ ;

решение  $\vec{\theta}$  непрерывно зависит от  $\vec{Y}$ , т. е. если приращения  $\Delta \vec{Y}$  стремятся к нулю, то приращение  $\Delta \vec{\theta}$  также стремится к нулю.

Другими словами, задача некорректна, если определитель линейного оператора  $\vec{A}$  равен нулю. Здесь  $\vec{\theta}$  является элементом метрического пространства  $U$ , а  $\vec{Y}$  — элементом метрического пространства  $F$ . Область определения оператора  $\vec{A}$ , действующего из  $U$  в  $F$ , есть  $D_A \subseteq U$ , область его значений  $Q_A = \vec{A}(D_A) \subseteq F$ .

Иногда плохо обусловленные и некорректные задачи называют просто некорректными. Этим термином мы будем пользоваться в дальнейшем.

Таким образом, чтобы корректно учитывать реальную экспериментальную информацию, необходимо иметь метод решения некорректных задач. Для этой цели применяется метод регуляризации А. Н. Тихонова и большая серия методов, развитых на его основе [14, 15].

Строго говоря, в задачах, перечисленных в главе 2, мы должны были бы применять на каждой итерации метод регуляризации А. Н. Тихонова для решения плохообусловленных систем линей-

ных алгебраических уравнений. В действительности имеет место более простая ситуация.

Во-первых, если области допустимых значений аргументов не пересекаются, что имеет место в большинстве практических задач, и учитываются ограничения на допустимые значения оценок  $\hat{\xi}$  при вычислениях, то процесс нахождения оценок  $\hat{\xi}$  является корректным.

Во-вторых, если в практических задачах число определяемых параметров невелико, то одна из возможных причин некорректности задачи — большая размерность системы — не проявляется и соответствующая система линейных алгебраических уравнений может быть достаточно хорошо обусловлена. Тем не менее с некорректными задачами в процессе построения прогнозов приходится иметь дело.

Пусть требуется найти решение  $\vec{\theta}$  системы линейных алгебраических уравнений

$$\vec{Y} = \vec{A}\vec{\theta},$$

минимизируя невязку  $\rho^2(\vec{A}\vec{\theta}, \vec{Y}) = J(\vec{\theta})$ . При определенных условиях задача минимизации  $\rho^2(\vec{A}\vec{\theta}, \vec{Y})$  может быть некорректной. В методе регуляризации А. Н. Тихонова вводится функция  $\Omega(\vec{\theta})$ , определенная на непустом множестве  $U_\alpha \subseteq U$ , называемая стабилизатором, если [15]

$$1) \Omega(\vec{\theta}) \geq 0 \text{ для всех } \vec{\theta} \in U_\alpha;$$

2) множество  $\Omega_C = \{\vec{\theta} | \vec{\theta} \in U_\alpha; \Omega(\vec{\theta}) \leq C\}$  является  $\rho$ -компактным при любом  $C = \text{const} \geq 0$ , т. е. из любой последовательности  $\{\vec{\theta}_k\} \in \Omega_C$  можно выбрать подпоследовательность  $\{\vec{\theta}_{k_i}\}$ ,  $\rho$ -сходящуюся к некоторой точке  $\vec{\theta} \in \Omega_C$ ;

3) множество  $U_\alpha^* = U_\alpha \cap U_*$  непустое, ( $U_*$  — множество точек минимума функции  $J(\vec{\theta}) = \rho^2(\vec{A}\vec{\theta}, \vec{Y})$ ).

Далее берется какая-либо положительная последовательность  $\{\alpha_k\}$ , сходящаяся к нулю, и при каждом  $k=1, 2, \dots$  на множестве  $U_\alpha$  определяется функция А. Н. Тихонова

$$T_k(\vec{\theta}) = J(\vec{\theta}) + \alpha_k \Omega(\vec{\theta}); \quad \vec{\theta} \in U_\alpha.$$

Минимум функции А. Н. Тихонова для различных значений  $k$  определяет минимизирующую последовательность  $\{\vec{\theta}_k\}$ , сходящуюся к регуляризованному решению  $\hat{\theta}_\rho$ .

Существуют алгоритмы, в которых по величине погрешности исходных данных определяется единственное (оптимальное) значение параметра регуляризации  $\alpha$  в функции А. Н. Тихонова и сразу находится регуляризованное решение  $\hat{\theta}_\rho$ .

Нас будет интересовать не только регуляризованное решение  $\hat{\theta}_\rho$ , но и интервальная оценка этого решения, поскольку мы имеем дело с исходными случайными величинами. Рассмотрим функцию А. Н. Тихонова как функцию Лагранжа следующей задачи: минимизировать функцию  $\Omega(\vec{\theta})$  на множестве  $U_\alpha \subseteq U$  при условии

$$\rho^2(\vec{A}\vec{\theta}, \vec{Y}) = \delta^2.$$

Здесь  $\delta$  — величина погрешности исходных данных,  $\alpha$  в функции А. Н. Тихонова играет роль множителя Лагранжа.

Выбор стабилизатора  $\Omega(\vec{\theta})$  в методе регуляризации неоднозначен, часто для систем линейных алгебраических уравнений его выбирают в виде  $\Omega(\vec{\theta}) = \|\vec{\theta}\|^2$ .

Если функция  $\rho^2(\vec{A}\vec{\theta}, \vec{Y})$  для  $\vec{Y} = \vec{A}\vec{\theta} + \vec{\varepsilon}$  может быть записана в виде  $(\vec{Y} - \vec{A}\vec{\theta})^T D^{-1}(\vec{Y} - \vec{A}\vec{\theta})$ , где  $D(\vec{Y})$  — ковариационная матрица погрешностей исходных данных, и стабилизатор

$$\Omega(\vec{\theta}) = \vec{L}\vec{\theta} - \vec{R}, \text{ то}$$

1) вектор оценок  $\vec{\theta}$  точки минимума функции А. Н. Тихонова

$$T = (\vec{Y} - \vec{A}\vec{\theta})^T D^{-1}(\vec{Y} - \vec{A}\vec{\theta}) + \vec{\alpha}^T (\vec{L}\vec{\theta} - \vec{R})$$

имеет вид  $\hat{\theta} = \vec{F}\vec{A}^T D^{-1}(\vec{Y})\vec{Y} + \vec{G}^T \vec{R}$ ,

2) матрица вторых моментов оценок

$$D(\hat{\theta}) = \vec{C}^{-1} - \vec{C}^{-1} \vec{L}^T (\vec{L} \vec{C}^{-1} \vec{L}^T)^{-1} \vec{L} \vec{C}^{-1}; \quad (2.43)$$

3) дисперсия параметра регуляризации  $\hat{\alpha}$  (множителя Лагранжа)

$$D(\hat{\alpha}) = (\vec{L} \vec{C}^{-1} \vec{L}^T)^{-1}. \quad (2.44)$$

Здесь  $\vec{F} = \vec{C}^{-1} - \vec{C}^{-1} \vec{L}^T (\vec{L} \vec{C}^{-1} \vec{L}^T)^{-1} \vec{L} \vec{C}^{-1}$ ;  $\vec{G} = (\vec{L} \vec{C}^{-1} \vec{L}^T)^{-1} \vec{L} \vec{C}^{-1}$ ;  $\vec{C} = \vec{A}^T D^{-1}(\vec{Y}) \vec{A}$ .

Диагональные элементы второго члена выражения (2.43) служат мерой уменьшения дисперсий оценок в методе регуляризации.

Что касается смешанных вторых моментов оценок параметров  $\vec{\theta}$ , то они могут увеличиваться или уменьшаться в зависимости от конкретной задачи.

Смешанный второй момент для  $\vec{\theta}$  и  $\vec{\alpha}$  равен нулю: оценки  $\vec{\theta}$  и  $\vec{\alpha}$  не коррелированы.

Из выражений (2.43) и (2.44) следует, что матрица вторых моментов параметров  $\vec{\theta}$  является подматрицей, обратной матрице исходных уравнений, получаемой после дифференцирования функции А. Н. Тихонова  $T(\vec{\theta})$  по  $\vec{\theta}$  и  $\vec{\alpha}$ , а матрица вторых моментов  $\vec{\alpha}$  — подматрицей со знаком минус. Таким образом дисперсии оценок  $\vec{\theta}$  и  $\vec{\alpha}$  можно определить с помощью матрицы  $M(\vec{\theta})$ , элементы которой — вторые производные функции А. Н. Тихонова по  $\vec{\theta}$  и  $\vec{\alpha}$ , взятые со знаком минус и вычисленные при найденных значениях оценок  $\vec{\theta}$  и заданных значениях  $\vec{\alpha}$ . Дисперсия  $\vec{\alpha}$  показывает возможный интервал выбора параметра регуляризации  $\vec{\alpha}$ . Обратив матрицу  $M(\vec{\theta})$ , получим подматрицы, определяющие ковариационные матрицы оценок  $\vec{\theta}$  (2.43) и  $\vec{\alpha}$  (2.44). Отметим, что значение константы  $\vec{R}$  не влияет на значения получаемых дисперсий (2.43) и (2.44).

Однако в методе регуляризации не применяется линейный стабилизатор. Для нахождения интервальных оценок можно, предварительно определив точечные оценки  $\vec{\theta}$  и  $\vec{\alpha}$  с помощью других известных методов [15], воспользоваться только что полученными результатами, линеаризовав стабилизатор  $\Omega(\vec{\theta})$  в окрестности  $\vec{\theta}$ . Определив интервальные оценки, нетрудно непосредственно проверить, не вносит ли линеаризация стабилизатора искажений, превышающих погрешность исходных данных.

**Пример.** Рассмотрим функцию А. Н. Тихонова

$$T(\theta_1, \theta_2) = (\theta_1 - 1)^2 + (\theta_2 - 1)^2 + \alpha(\theta_1^2 + \theta_2^2) \rightarrow \min.$$

Тогда  $\Omega(\theta_1, \theta_2) = \theta_1^2 + \theta_2^2$ . Координаты точки минимума  $T(\theta_1, \theta_2)$  при малых  $\alpha$  близки к (1; 1). Разложим  $\Omega(\theta_1, \theta_2)$  в окрестности

точки (1; 1):  $\Omega(\theta_1, \theta_2) \approx 2 + 2(\theta_1 - 1) + 2(\theta_2 - 1)$ . После замены переменных  $\theta_1 = \theta_1 - 1$ ,  $\theta_2 = \theta_2 - 1$  функция А. Н. Тихонова будет иметь вид

$$T(\theta_1, \theta_2) = \theta_1^2 + \theta_2^2 + \alpha(2\theta_1 + 2\theta_2 + 2) \rightarrow \min.$$

Вычислив матрицу, элементы которой есть вторые производные функции  $T(\theta_1, \theta_2)$  по  $\theta_1, \theta_2$  и  $\alpha$ , обратив ее, получим матрицу вторых моментов оценок  $\theta$  и  $\alpha$

$$\begin{pmatrix} 0,25 & -0,25 \\ -0,25 & 0,25 \\ & & 0,25 \end{pmatrix}.$$

Элементы матрицы (3,1), (3,2), (1,3) и (2,3) здесь не указаны, так как они не несут информации (учитываются только соответствующие подматрицы). Значения элементов матрицы не зависят от величины константы в разложении  $\Omega(\theta_1, \theta_2)$ . Поэтому можно записать эквивалентную задачу так:

$$f(\theta_1, \theta_2) = \theta_1^2 + \theta_2^2 \rightarrow \min$$

при  $2(\theta_1 + \theta_2) = \text{const} = C$ .

Отсюда  $f(\theta_1) = \theta_1^2 + \left(\frac{C}{2} - \theta_1\right)^2 \rightarrow \min;$

$$\frac{d^2 f(\theta_1)}{d\theta_1^2} = 4;$$

дисперсии оценок  $\hat{\theta}_1$  и  $\hat{\theta}_2$  равны 1/4, что совпадает с ранее полученными результатами. Без учета условия-ограничения дисперсии оценок равны 1/2.

Вопрос о некорректности задачи получения оценок в конфликтном анализе можно было бы рассмотреть в самом начале, при определении функционала, дающего конфликтное решение задачи. К этому функционалу, например (2.8); (2.9); (2.11), следует добавить стабилизатор  $\Omega$ . Если стабилизатор  $\Omega = \Omega(\vec{\theta})$ , то мы приходим к изложенным здесь результатам.

Действительно, в общем случае функция А. Н. Тихонова имеет вид

$$T = \ln L(\vec{X}, \vec{Y}/\vec{\theta}) + \alpha\Omega(\vec{\theta}),$$

где  $\Omega(\vec{\theta})$  — дифференцируемая функция. Тогда оценка параметров  $\vec{\theta}$  при фиксированном значении  $\hat{\alpha}$  находится из системы уравнений

$$\frac{\partial}{\partial \theta_j} [\ln L(\vec{X}, \vec{Y}/\vec{\theta}) + \hat{\alpha}\Omega(\vec{\theta})] = 0; \quad j = \overline{1, m}.$$

Состоятельность и асимптотическая нормальность таких оценок  $\hat{\theta}$  в регрессионной модели доказана.

Дисперсия оценок

$$D(\hat{\theta}) = \vec{A}^{-1} - \vec{A}^{-1} \vec{B} (\vec{B}^T \vec{A}^{-1} \vec{B})^{-1} \vec{B}^T \vec{A}^{-1};$$

$$D(\hat{\alpha}) = (\vec{B}^T \vec{A}^{-1} \vec{B})^{-1},$$

где матрицы

$$\vec{A} = \left\| \frac{\partial^2 T}{\partial \theta_i \partial \theta_j} \right\|; \quad \vec{B} = \left\| \frac{\partial^2 T}{\partial \theta_i \partial \alpha} \right\|; \quad i = \overline{1, m}; \quad j = \overline{1, m}.$$

Смешанный второй момент  $\hat{\theta}$  и  $\hat{\alpha}$  равен нулю: оценки параметров  $\hat{\theta}$  не коррелированы с  $\hat{\alpha}$ .

С помощью матрицы вторых производных получим ковариационную матрицу оценок

$$D(\hat{\theta}, \hat{\alpha}) = \begin{pmatrix} \vec{A}^{-1} - \vec{A}^{-1} \vec{B} (\vec{B}^T \vec{A}^{-1} \vec{B})^{-1} \vec{B}^T \vec{A}^{-1} & \vec{A}^{-1} \vec{B} (\vec{B}^T \vec{A}^{-1} \vec{B})^{-1} \\ (\vec{B}^T \vec{A}^{-1} \vec{B})^{-1} \vec{B}^T \vec{A}^{-1} & -(\vec{B}^T \vec{A}^{-1} \vec{B})^{-1} \end{pmatrix}, \quad (2.45)$$

но учитывать в (2.45) надо только диагональные элементы. Например, для  $T = (\theta_1 - 1)^2 + (\theta_2 - 1)^2 + \alpha(\theta_1^2 + \theta_2^2) \rightarrow \min$  ковариационная матрица

$$D(\hat{\theta}, \hat{\alpha}) = \begin{pmatrix} \frac{\hat{\theta}_2^2}{2(1+\hat{\alpha})(\hat{\theta}_1^2 + \hat{\theta}_2^2)} & \frac{-\hat{\theta}_1 \hat{\theta}_2}{2(1+\hat{\alpha})(\hat{\theta}_1^2 + \hat{\theta}_2^2)} \\ \frac{-\hat{\theta}_1 \hat{\theta}_2}{2(1+\hat{\alpha})(\hat{\theta}_1^2 + \hat{\theta}_2^2)} & \frac{\hat{\theta}_1^2}{2(1+\hat{\alpha})(\hat{\theta}_1^2 + \hat{\theta}_2^2)} \\ & & -\frac{1+\hat{\alpha}}{2(\hat{\theta}_1^2 + \hat{\theta}_2^2)} \end{pmatrix}.$$

При  $\hat{\theta}_1 = \hat{\theta}_2 = 1$  и  $\hat{\alpha} \ll 1$  получим ранее найденную матрицу. Из выражения  $D(\hat{\theta}, \hat{\alpha})$  следует, что с увеличением значения  $\hat{\alpha}$  (с увеличением вклада ограничений) дисперсии оценок параметров уменьшаются.

Показано, что с помощью ряда методов безусловной оптимизации функции  $J(\vec{\theta})$ , сопряженных градиентов, наискорейшего (градиентного) спуска и т. д., можно получить регуляризованное решение, если ограничить число итераций  $n \leq n_0$  поиска экстремума таким образом, чтобы при выполнении условия  $J(\vec{\theta}) = \delta$ , где  $\delta$  определяется погрешностью исходных данных, процесс миними-

зации прекращался. Полученное решение и будет регуляризованным. При увеличении числа итераций  $n > n_0$  решение может стать неустойчивым. Параметром регуляризации здесь является число итераций  $n_0$ .

Подобный подход привлекает своей простотой. Для определения дисперсий и оценок в этом случае необходимо учесть ограничение, накладываемое траекторией движения от начального приближения к решению, при условии, что в точке регуляризованного решения значения первых производных функционала малы по сравнению со значениями вторых производных.

### 3. МЕТОДЫ ПОСТРОЕНИЯ ПРОГНОЗОВ

#### 3.1. Особенности процедуры прогнозирования

Рассмотрим приложение стохастических (статистических) моделей, адекватно описывающих поведение систем и их элементов, для целей прогнозирования поведения систем. Другими словами, математические модели будут использованы для получения будущих значений динамических (временных) рядов, характеризующих исследуемые характеристики систем.

Этапу прогнозирования предшествует итеративная процедура построения математической модели, основанная на идентификации, оценивании и диагностической проверке.

Под идентификацией подразумевается использование любой информации о том, как были получены рассматриваемые числовые значения (как был генерирован ряд), с целью отыскания набора экономичных моделей, заслуживающих опробования. Экономичные модели должны обладать максимальной простотой и минимальным числом параметров, но при этом адекватно описывать наблюдения. Методы идентификации определяют класс пробных моделей, для которых применяются более формальные и эффективные методы оценивания.

Под оцениванием подразумевается процедура получения оценок параметров моделей, определяющих адекватность моделей, ибо неадекватность выбранной модели может быть вызвана неэффективностью процедуры оценки параметров модели, а не тем, что неадекватен вид модели. Процедуры получения оценок рассмотрены в гл. 1 и 2.

Под диагностической проверкой подразумевается проверка согласования подготовленной модели с исходными данными, чтобы вскрыть недостатки модели и улучшить ее. В частности, метод диагностической проверки модели может базироваться на введении избыточного числа параметров, т. е. в оценивании параметров

для несколько более общей модели, чем ожидаемая. Этот подход исходит из того, что мы можем угадать неадекватные свойства модели, и требует исследования остаточных ошибок после подгонки модели. Он позволяет определить, какие необходимы изменения модели.

При проведении прогнозирования необходимо иметь в виду, что все факторы, влияющие на поведение системы в базовом (исследуемом) и прогнозируемом периодах, должны быть неизменны или изменяться по известному закону. Первый случай реализуется в однофакторном прогнозировании, второй — при многофакторном.

Можно выделить два вида прогнозируемых характеристик системы, зависящих от времени: переменные состояния и переменные интенсивности. Переменная состояния определяется периодически, и ее значение в течение небольшого интервала времени не зависит от времени, прошедшего с момента начала наблюдения. Переменная интенсивности также определяется периодически, но ее значение пропорционально времени, прошедшему с момента предыдущего наблюдения. Если переменная состояния характеризуется количеством, то переменная интенсивности — скоростью его изменения.

Величина промежутков времени между измерениями входных переменных системы с целью проверки и уточнения ранее сделанных прогнозов о выводных переменных зависит главным образом от длительности времени упреждения и наибольшей частоты циклических изменений в системе, которые должна отражать модель. Поэтому временные интервалы пересмотров прогнозов могут изменяться в широких пределах. Однако эти интервалы должны быть достаточно велики, чтобы обеспечивалась вероятность осуществления ожидаемых изменений системы. Когда имеют место какие-то периодические процессы, то частота наблюдений должна быть по крайней мере вдвое больше частоты изучаемого процесса.

Если случайная ошибка при определении входных переменных велика по сравнению с измеряемой величиной, интервал уточнения прогноза для переменной интенсивности целесообразно увеличить, усредняя таким образом случайную ошибку. Однако для переменной состояния в аналогичном случае интервал уточнения прогноза лучше уменьшить, что позволяет для выделения полезного сигнала использовать соответствующие методы фильтрации. Использование доступных к моменту  $t$  наблюдений динамического ряда для прогнозирования его значений в некоторый момент  $(t+l)$  в будущем является основой для управления и оптимизации промышленных процессов, экономических систем и т. д. Интервал  $l$  называют временем упреждения. Функция  $\hat{y}_t(l)$ , дающая в момент  $t$  прогнозы будущих времен упреждения  $l$ , называется

прогнозирующей функцией в момент  $t$ . Прогнозирующая функция определяется заданной априори целью. Например, находится такая прогнозирующая функция  $\hat{y}_t(l)$ , у которой среднее значение квадрата отклонения  $(y_{t+l} - \hat{y}_t(l))^2$  истинного  $y_{t+l}$  от прогнозируемого значений является наименьшим для каждого упреждения  $l$ .

Вычисление наилучшего прогноза должно сопровождаться указанием его точности, например, чтобы можно было оценить риск, связанный с принятием решения по данному прогнозу. Точность прогноза может быть выражена вероятностными пределами (интервальными оценками).

Динамический ряд рассматривается как сумма детерминированной и случайной компонент. Детерминированная компонента выражается некоторой аппроксимирующей функцией, отражающей закономерность развития исследуемого явления. Появление случайной компоненты определяется сложным переплетением параметров системы, влиянием на их величину большого числа неизвестных факторов, действующих в разных направлениях, что находит свое выражение в отклонении значений показателей системы от основной тенденции развития. Дополнительный вклад вносит и аппроксимирующая модель, которая не в состоянии описать все особенности системы.

Наиболее часто отклонения от основной тенденции развития рассматриваются как стационарный случайный процесс, к которому применимы методы прогнозирования стационарных случайных процессов. Если случайная компонента не является стационарной, то производят определенные преобразования, чтобы сделать случайную компоненту хотя бы стационарной в определенных условиях.

В прогнозировании важное значение имеет предварительный анализ характера изучаемого явления для определения вида его описания: процесс хорошо описывается основной тенденцией (трендом) или процесс зависит от изменения некоторого набора показателей, отражающих структуру процесса. Выбор вида описания предопределяет точность прогноза на будущее.

Методы прогнозирования могут быть разделены на три большие группы: статистические (описательные), причинно-следственные и их комбинация. Для изучения исследуемого процесса следует задать закон изменения входных переменных во времени. Выходные переменные системы могут быть описаны с помощью некоторой модели, значения коэффициентов которой определяются подбором. При этом различные наблюдения учитываются с различными весовыми множителями. По таким моделям, включающим описание предистории системы, прогноз можно составить путем расчета состояния системы для некоторого будущего момента времени.



Если удастся построить модель окружающей среды, позволяющую выявить причины изменений в системе (вторая группа методов), то прогноз, полученный с помощью такой модели, объясняет будущее системы. Подобные методы охватывают широкий круг моделей.

Между двумя любыми автокоррелированными временными рядами всегда существует статистическая корреляция. Следовательно, существует опасность бессмысленного использования множественной регрессии в поисках «хорошего» коэффициента корреляции между прогнозируемыми (выходными) переменными и различными потенциально информативными входными переменными. Известная предыстория представляет собой только элемент полного временного ряда и если искать достаточно долго, связь между переменными в любом случае будет найдена. Но если с увеличением объема информации коэффициенты модели становятся равными нулю, то модель не пригодна для принятия правильного решения, так как именно в критический момент основные показатели перестали быть таковыми. Наилучшие результаты получаются при использовании комбинации статистических и причинно-следственных методов прогнозирования. Исходные данные обычно представляют собой результаты выборочных наблюдений либо переменной интенсивности, либо переменной состояния. Результаты наблюдений регистрируются с ошибками, которые возникают как при наблюдениях, так и при регистрации данных. Кроме того, изучаемый процесс может иметь стохастическую природу. Результаты наблюдений могут содержать и аномальные эффекты. Поэтому не каждую совокупность зарегистрированных по мере поступления реальных данных следует считать подходящим рядом, на основании которого можно составлять прогноз. Перед тем как подобрать коэффициенты модели по исходным данным, из последних должны быть исключены выбросы, т. е. результаты наблюдений, которые не характеризуют прогнозируемый процесс.

Для описания стохастических элементов рядов и их прогноза используются три различных понятия — помехи, остатки и ошибки. Понятие помехи связано с собственной изменчивостью процесса и неопределенностью, вносимой при наблюдении за ним. Следовательно, помеха является составной частью используемых данных. Под остатками понимается разность между результатами наблюдений и соответствующими значениями, вычисленными с помощью прогнозирующей их модели. Таким образом, остатки связаны с прошлыми данными и моделью, которая использовалась для их оценок. Ошибки прогноза представляют собой разницу между прогнозом, сделанным в настоящее время, и тем, что будет наблюдаться позднее в момент времени, для которого составлен прогноз.

Любой процесс, представленный рядом результатов наблюдений, можно описать набором разностных уравнений (когда процесс дискретен) или дифференциальных уравнений (когда он непрерывен). Если в этих уравнениях коэффициенты не зависят от времени, процесс называется стационарным; если же зависят, то нестационарным. Эта зависимость может носить вероятностный или регулярный характер. Известное регулярное изменение коэффициентов во времени может быть описано дополнительными уравнениями.

Для достаточно больших интервалов времени большинство прогнозируемых рядов являются нестационарными. Однако их все же можно считать квазистационарными, если прогноз составлять для какого-то одного момента времени.

Ни в одном из статистических методов прогнозирования не может быть заранее предусмотрено изменение модели прогнозируемого процесса. Имеется ряд методов быстрого обнаружения изменений в последующих процессах с ней соответствующей реакцией на них. Модель прогноза может все более усложняться, когда это экономически оправдано и позволяет глубже проникнуть в механизм наблюдаемых явлений.

Во многих случаях изменения в изучаемом процессе можно предвидеть заранее, но в модель прогноза они тем не менее не включаются, так как последствия таких изменений не могут быть точно рассчитаны. Тем не менее на основе тщательного анализа различных вариантов можно предсказать характер изменений. В любой отдельный период времени существует, очевидно, несколько серий прогнозов, отличных от простого описательного прогноза. Это позволяет минимизировать время, затрачиваемое на внесение изменений.

### 3.2. Модели для получения прогнозов

Векторная форма. Пусть в момент времени  $T$  задана последовательность результатов наблюдений  $y_t$  для некоторого множества моментов времени  $t \leq T$ . Прогнозирующая модель задает множество выходных переменных  $Y_{T+l}$ , которые могут быть выражены в векторной форме ( $l > 0$ ). В общем виде выражение для модели записывается как  $Y_{T+l} = a_l F(l)$ , где вектор  $a_l$  представляет собой коэффициенты модели, получаемые по результатам наблюдений до момента  $T$  включительно, а матрица  $F$  — набор аппроксимирующих функций. Строки матрицы соответствуют элементам модели, а столбцы — моментам времени. В большинстве практических приложений  $F$  определяет время относительно момента самого последнего наблюдения, а значения коэффициентов  $a_l$  зависят от выбора начала отсчета времени.

Широкий класс моделей процесса составляют модели, которые могут быть представлены с помощью полиномов  $F_i$ :  $F_1(t)=1$ ,  $F_2(t)=t$ ,  $F_3(t)=1/2t(t-1)$  и т. д. Применение различных методов прогнозирования позволяет получить различные модели, например за счет использования разностей или суммы функций  $F_i$ . Поскольку решение линейного дифференциального уравнения может быть представлено в виде полинома, то различные математические описания моделей означают лишь возможность выбора удобной формы анализа, а не фактическое различие в математических свойствах таких моделей. Тем не менее точность прогноза будет зависеть от математического описания явления.

Математические методы позволяют представить прогнозирующую модель в виде полинома любого порядка. Однако, если нет необходимости в использовании полиномов более высокого порядка по другим причинам, применение полиномов второго порядка для описания прогнозируемых рядов часто оказывается достаточным.

В процессе прогнозирования, как правило, нельзя предложить детерминированную модель, допускающую точное вычисление будущего поведения объекта, так как в ней может участвовать ряд неизвестных факторов. С большей вероятностью можно предложить модель, позволяющую вычислить вероятность того, что некоторое будущее значение будет лежать в определенном интервале — стохастическая вероятностная модель (или стохастический процесс).

Динамический ряд  $y_1, y^2, \dots, y_N$  из  $N$  последовательных наблюдений рассматривается как выборочная реализация из бесконечной популяции рядов, генерируемых стохастическим процессом.

Стохастические модели подразделяются на стационарные, когда процесс остается в равновесии относительно постоянного среднего уровня, и на нестационарные, в частности, не имеющие естественного среднего значения. Например, стохастическая модель, для которой прогнозирование экспоненциально взвешенным скользящим средним является оптимальным, относится к классу нестационарных процессов, называемых процессами авторегрессии — проинтегрированного скользящего среднего (АРПСС). Этот класс процессов обеспечивает множество как стационарных, так и нестационарных моделей, которые адекватно описывают многие встречающиеся на практике ряды.

Для построения моделей будем пользоваться некоторыми операторами: оператор сдвига назад  $B$  определяется как  $By_t=y_{t-1}$ ;  $B^m y_t=y_{t-m}$ ; оператор сдвига вперед  $F=B^{-1}$ ;  $Fy_t=y_{t+1}$ ;  $F^m y_t=y_{t+m}$ ; разностный оператор со сдвигом назад  $\Delta$ ;  $\Delta y_t=y_t-y_{t-1}=(1-B)y_t$ ; оператор суммирования  $S=\Delta^{-1}$ ;  $Sy_t=\Delta^{-1}y_t=y_t+y_{t-1}+y_{t-2}+\dots=(1+B+B^2+\dots)y_t=(1-B)^{-1}y_t$ .

Модель авторегрессии. Рассмотрим основные разностные модели для построения прогнозов. Здесь текущее значение процесса выражается как конечная линейная совокупность предыдущих значений процесса и независимого импульса  $a_t$ , который есть реализация случайной величины с фиксированным распределением, часто полагаемым нормальным с нулевым средним и дисперсией  $\sigma^2(a)$  (белый шум), т. е.

$$y_t = \mu + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + a_t,$$

где  $\mu$  — параметр, определяющий «уровень» процесса (тренд);  $\Phi_i$  ( $i=\overline{1, p}$ ) — неизвестные параметры. Пусть  $\tilde{y}_t, \tilde{y}_{t-1}, \tilde{y}_{t-2}, \dots$  будут отклонениями от  $\mu$ , например  $\tilde{y}_t = y_t - \mu$ . Тогда

$$\tilde{y}_t = \Phi_1 \tilde{y}_{t-1} + \Phi_2 \tilde{y}_{t-2} + \dots + \Phi_p \tilde{y}_{t-p} + a_t$$

называется процессом авторегрессии (АР) порядка  $p$ . Введем оператор авторегрессии порядка  $p$

$$\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p,$$

то модель авторегрессии имеет вид

$$\Phi(B) \tilde{y}_t = a_t.$$

Она содержит  $(p+2)$  неизвестных параметра:  $\mu, \Phi_1, \Phi_2, \dots, \Phi_p, \sigma^2(a)$ , которые надо впоследствии оценить перед вычислением прогноза по наблюдениям. Применяя линейный оператор

$$\Psi(B) = 1 + \Psi_1 B + \Psi_2 B^2 + \dots = \Phi^{-1}(B),$$

модель авторегрессии можно записать в виде

$$\tilde{y}_t = \Psi(B) a_t;$$

здесь  $\tilde{y}_t$  выражается как бесконечная взвешенная сумма  $a_t, a_{t-1}, \dots$ . Процессы авторегрессии могут быть стационарными или нестационарными. Чтобы процесс был стационарным, надо чтобы  $\Psi_1, \Psi_2, \dots$  в операторе  $\Psi(B) = \Phi^{-1}(B)$  образовывали сходящийся ряд.

Модель скользящего среднего. Пусть  $\tilde{y}_t$  линейно зависит от конечного числа  $q$  предыдущих значений  $a$ . Процесс  $\tilde{y}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$  называется процессом скользящего среднего (СС) порядка  $q$ .

Введем оператор скользящего среднего порядка  $q$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q,$$

тогда  $\tilde{y}_t = \theta(B) a_t$ .

Модель содержит  $(q+2)$  неизвестных параметра:  $\mu, \theta_1, \theta_2, \dots, \theta_q, \sigma^2(a)$ , подлежащих оценке по наблюдениям. Последовательность независимых случайных величин  $a_t, a_{t-1}, a_{t-2}, \dots$  (белый шум) вводится для генерации динамических рядов, в которых последовательные значения сильно зависимы. Например, белый шум  $a_t$  можно трансформировать в процесс  $y_t$  при помощи линейного фильтра  $\Psi(B)$ :

$$y_t = \mu + a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \dots = \mu + \Psi(B) a_t.$$

Смешанные модели авторегрессии — скользящего среднего. Для достижения большей гибкости в подгонке модели к наблюдениям вводят комбинированную модель авторегрессии — скользящего среднего

$$\tilde{y}_t = \Phi_1 \tilde{y}_{t-1} + \dots + \Phi_p \tilde{y}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

или

$$\Phi(B) \tilde{y}_t = \theta(B) a_t,$$

в которой  $p+q+2$  параметра:  $\mu, \Phi_1, \dots, \Phi_p, \theta_1, \dots, \theta_q, \sigma^2(a)$ , — подлежат оценке по наблюдениям.

Нестационарные модели. Для многих рядов уровни, относительно которых происходят флуктуации, являются различными для разных элементов ряда, но с учетом этого различия поведение их оказывается схожим. Так, для однородного нестационарного процесса можно ввести ряд, составленный из  $d$ -ых разностей, который может быть стационарным. Для описания однородного нестационарного процесса вводят обобщенный оператор авторегрессии

$$\varphi(B) = \Phi(B) (1-B)^d,$$

где  $\Phi(B)$  — стационарный оператор.

Обобщенная модель для описания однородного нестационарного процесса имеет вид

$$\varphi(B) y_t = \Phi(B) (1-B)^d y_t = \theta(B) a_t,$$

или

$$\Phi(B) \omega_t = \theta(B) a_t; \quad \omega_t = \Delta^d y_t,$$

т. е.

$$\omega_t = \Phi_1 \omega_{t-1} + \dots + \Phi_p \omega_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}.$$

Этот процесс называется процессом **авторегрессии — проинтегрированного** (просуммированного) **скользящего среднего** (АРПСС) порядка  $(p, d, q)$ . Очевидно, что АРПСС является общей для всех рассмотренных здесь моделей. На практике адекватное описание стационарных и не стационарных процессов происходит при значениях  $p$  и  $q$  не больше, чем два и при  $d$ , равных нулю или единице (максимум два).

### 3.3. Анализ и сглаживание рядов с помощью скользящей средней

При представлении совокупности результатов наблюдений в виде рядов фактически используется предположение о том, что наблюдаемые величины принадлежат некоторому распределению, параметры которого и их изменение можно оценить. По этим параметрам (как правило, по среднему значению и дисперсии, хотя иногда используется и более полное описание) можно построить одну из моделей вероятностного представления процесса.

Другим вероятностным представлением является модель в виде частотного распределения с параметрами  $p_j$  для относительной частоты наблюдений, попадающих в  $j$ -й интервал. При этом, если в течение принятого времени упреждения не ожидается изменения распределения, то решение принимается на основании имеющегося эмпирического частотного распределения.

Важнейшей задачей анализа является выявление основной тенденции поведения системы, как результата влияния комплекса причин, действующих на изучаемый процесс. Основная тенденция характеризуется **трендом**. Для выявления основной тенденции применяют сглаживание рядов динамики. Процесс сглаживания требует тщательного анализа, чтобы исключить возможность «сгладить» под видом случайных выбросов и отклонений существенные кратковременные изменения показателей, отражающих важные моменты в поведении системы.

Для сглаживания рядов динамики часто применяется метод наименьших квадратов. При сглаживании рядов динамики следует подчеркнуть одну особенность. Процесс в действительности может быть описан гладкой кривой, а наблюдаемые выбросы есть результат наложения случайных помех. В этом случае процедура проведения сглаживания методом наименьших квадратов имеет четкое статистическое объяснение. Но в других случаях отклонение сглаженной от экспериментальной кривой нельзя объяснить только влиянием случайных помех. И в подобных случаях применяют метод наименьших квадратов, но здесь он определяет некоторую гладкую кривую наилучшим образом (в выбранной метрике) проходящую по исходным данным. И ничего общего со статистическим толкованием метода наименьших квадратов в данном случае нет.

В результате предварительного анализа (из физических условий задачи) выбирается класс функций, которыми может быть описано изучаемое явление. Параметры этих функций подлежат определению. В ряде случаев параметры функций могут иметь определенный физический или экономический смысл. Для процедуры прогнозирования при выборе сглаживающей функции надо иметь в виду следующее: функция должна отражать основную

закономерность развития явления, в отношении которой можно выдвинуть гипотезу, что эта закономерность сохранится и в будущем. Тогда как, например, для процесса интерполяции достаточно, чтобы выбранная функция наиболее близко описывала значения ряда.

Для сглаживания рядов динамики нецелесообразно брать функции с большим числом параметров, в то же время выбранная функция должна быть адекватна исследуемому процессу. Следует иметь в виду, что короткие ряды в редких случаях дают возможность получить объективную информацию.

После определения параметров выбранных функций из них надо выбрать наиболее приемлемую. Знание оценок свободных параметров функций и их дисперсий дает возможность применить критерии согласия для выбора основной функции. В крайнем случае этот выбор можно провести по величине среднеквадратического значения отклонения числовых значений элементов ряда и тех же значений, вычисленных по аппроксимирующей функции.

Как было уже отмечено, метод наименьших квадратов применим при детерминированных значениях аргументов (факторов) выбранных функций. В ряде случаев условие детерминированности факторов не выполняется и следует использовать описанные в гл. 2 методы конъюгентного анализа. Все необходимые для сглаживания рядов динамики формулы могут быть взяты в гл. 2. Подчеркнем еще раз необходимость учета всех погрешностей и неопределенностей исходных данных. В п. 2.6 показывается например, что для функции  $\eta = \xi_1 \xi_2^3 - \xi_1^2 \xi_2^2 + 2\xi_1^3$  регрессионные оценки параметров отличались до 10% от истинных значений, а конъюгентные — до 2% при отклонении в значениях функции от идеальных в базисном периоде на 2—3%. Нетрудно теперь будет указать те значения  $\xi$ , когда ошибка прогноза по регрессионной модели достигнет 10%, а по конъюгентной модели — только 2%.

Практически во всех применяемых в настоящее время методах прогнозирования коэффициенты моделей сначала определяются путем подгонки модели к некоторым данным предыстории, а затем проверяются и уточняются по мере поступления новых данных. Выбор коэффициентов, как правило, осуществляется из условия минимизации суммы квадратов остаточных разностей между данными  $y_{T-j}$  и результатами расчета по модели  $\hat{y}_{T-j}$  с учетом различных весовых множителей  $\omega_j$ , приписываемых остаткам в различные моменты времени. Так, в рамках модели МНК, записанной в общем виде, коэффициенты  $a$  выбираются на основе минимизации величины  $S = \sum_j \omega_j (y_{T-j} - \hat{y}_{T-j})^2$ , где суммирование проводится по всем  $j$  вплоть до самого последнего наблюдения в

момент времени  $T$ . При оценке тех или иных достоинств различных систем весовых множителей  $\omega_j$  необходимо учитывать не только достигаемую при их использовании точность прогноза, но и степень сложности соответствующих вычислений.

Скольльзящее среднее представляет собой оценку по методу наименьших квадратов единственной константы для представления исходных данных с одинаковыми весовыми множителями. Этому случаю соответствует простая функция  $F(t) = 1$  для всех  $t$ . При этом весовые множители  $\omega_j = 1$  для  $N$  последних наблюдений в интервале  $0 \leq j \leq N-1$  и  $\omega_k = 0$  для  $j \geq N$ . Некоторые коэффициенты модели могут определяться с помощью полиномов более высокого порядка путем подгонки модели к результатам  $N$  самых последних наблюдений в каждый момент поступления новых данных. Таким образом, в самом общем случае значения коэффициентов в любой заданный момент времени зависят от предыдущих значений коэффициентов, ошибки в прогнозе при использовании самого последнего наблюдения, а также от вида используемого полинома и, конечно, числа  $N$  результатов наблюдений.

В методе скольльзящей средней первоначальные значения элементов ряда заменяются средней арифметической величиной внутри выбранного интервала. Полученное значение относится к середине выбранного интервала. Затем интервал сдвигается на одно наблюдение и расчет средней повторяется. Интервалы определения средней берутся все время одинаковыми. Чем шире интервал, тем более плавным получается новый ряд. Сглаженный ряд короче первоначального на  $(k-1)$  наблюдений ( $k$  — величина интервала сглаживания).

Выбор интервала сглаживания определяется конкретной задачей. Если число членов интервала сглаживания нечетное, то полученные значения скольльзящей средней

$$\hat{y}_{i+m} = \frac{\sum_{k=0}^{2m} y_{i+k}}{2m+1}$$

приходятся на средний член интервала сглаживания. При четном числе членов интервала сглаживания значения скольльзящих средних будут располагаться в промежутках между элементами ряда

$$\hat{y}_{i+m} = \frac{\frac{y_i}{2} + \sum_{k=0}^{2m-1} y_{i+k} + \frac{1}{2} y_{i+2n}}{2m}$$

Здесь  $y_i$  — значение  $i$ -го члена ряда;  $m$  — целое положительное число, определяющее величину интервала сглаживания;

$$1 \leq m \leq n/2 = \begin{cases} n/2, & \text{если } n \text{ четное число;} \\ \frac{n-1}{2}, & \text{если } n \text{ нечетное число;} \end{cases}$$

$n$  — число членов ряда.

Если известно, что внутри интервала сглаживания имеет место нелинейная тенденция, применяют взвешенные скользящие средние: внутри каждого интервала сглаживания значения элементов ряда описывают полиномом  $p$ -й степени

$$y = a_0 + \sum_{i=1}^p a_i t^i.$$

Параметры полинома  $\{a_0, a_1, \dots, a_p\}$  находят, например, методом наименьших квадратов. Взвешенную скользящую среднюю для взятого интервала определяют как средний член сглаженных на основании полинома значений исходного ряда.

Простота и наглядность — достоинства метода скользящей средней. Но при малом числе наблюдений метод приводит к искажению тенденции, величина интервала сглаживания влияет на форму тренда, теряются начальные и конечные элементы ряда.

Сглаживание рядов можно проводить и с помощью других функций и полиномов, а также с помощью конечных разностей. Для элементов ряда можно вычислить первые разности  $\Delta_i^{(1)} = y_{i+1} - y_i$ ; вторые разности  $\Delta_i^{(2)} = \Delta_{i+1}^{(1)} - \Delta_i^{(1)}$  и т. д.

Если принять, что элементы ряда можно описать полиномом  $p$ -й степени, то  $(p+1)$ -я производная этого полинома всюду равна нулю, т. е.  $(p+1)$ -я разность элементов такого ряда будет равна нулю. Таким образом, взяв  $p$ -ю разность от членов ряда, рассматриваем ряд, в котором исключен тренд, выраженный полиномом  $p$ -й степени.

#### 3.4. Прогнозирование с помощью экспоненциального сглаживания

Для стационарных рядов существует система весовых множителей, позволяющая обеспечить минимальную ошибку прогноза. Эти множители определяются видом автоковариационной функции. Известен метод вычисления таких оптимальных весовых функций для постоянного уровня, тренда и сезонных коэффициентов. В каждом из этих случаев весовые множители экспоненциально уменьшаются по закону  $\alpha^t$ , а различные значения  $\alpha$ , получаемые для уровня, тренда и сезонных коэффициентов, определяются систематическими исследованиями точности прогнозов, получаемых при различных комбинациях весовых функций.

Во многих случаях целесообразно использовать последовательность  $\omega_j = \alpha^j$ ,  $\alpha < 1$ , придающую более высокий вес более поздней информации и позволяющую относительно просто оценивать значения коэффициентов даже достаточно сложных моделей, таких, в которых для описания сезонных циклов используются полиномы в сочетании с преобразованиями Фурье (подобное представление можно рассматривать как сложные полиномы). Для модели экспоненциально взвешенного скользящего среднего предложены способы, с помощью которых в те периоды времени, когда средняя ошибка прогноза близка к нулю (благодаря правильности модели и ее коэффициентов), скорость затухания  $\alpha$  может быть увеличена, а в те периоды времени, когда средняя ошибка прогноза существенно отличается от нуля и существует опасность того, что модель может «забыть» старую информацию (в этом случае требуется уточнение прогноза), скорость затухания  $\alpha$  может быть уменьшена.

Уточнение прогноза производится по принципу обратной связи — новые прогнозы корректируются на основе учета ошибок в предшествующих прогнозах. Если при выборе весовых множителей в процессе составления прогноза также используется обратная связь, то не только строгий анализ областей устойчивости данной системы, но и любой анализ вообще становится фактически невозможным. Для анализа эффективности какого-либо метода достаточно привести примеры, подтверждающие его полезность. Необходимо также выявить области (если они существуют), в которых применение рассматриваемого метода невозможно или неэффективно.

Многие методы позволяют отыскать наилучшее значение скорости затухания весовых множителей путем многократного анализа имеющегося ряда данных. При этом в качестве критерия используется достигаемая точность прогноза (минимальная дисперсия ошибок). Однако такой подход содержит и недостатки. Во-первых, если в средней ошибке есть значимые разности (они должны быть равны нулю), то более вероятно, что эти разности больше зависят от способа выбора начальных значений коэффициентов модели, чем от различий в скорости затухания весовых множителей или постоянной сглаживания. Во-вторых, еще более важный источник возможной ошибки можно проиллюстрировать с помощью следующего примера. Будем рассматривать очень длинный ряд чисел как некоторый коррелированный процесс. Этот ряд, стационарность и однородность которого гарантирована самим способом его получения, разделим на короткие отрезки, содержащие достаточно данных для отыскания наилучшего значения скорости затухания весовых множителей. Проанализируем результаты, получаемые для каждого из этих отрезков. Хотя все отрезки

относятся к одному и тому же процессу, тем не менее существует широкое распределение соответствующих значений скорости затухания. Определить скорость затухания для данного отрезка можно только после того, как он стал историей. Величина скорости на следующем отрезке, которая может быть другой, будет известна после его прохождения, и, следовательно, для прогноза такие значения вообще бесполезны.

Разность между точностью, получаемой при постоянном использовании некоторой стандартной скорости затухания весовых множителей, например  $\alpha = 0,9$  и точностью, которая может быть достигнута при заранее известном оптимальном значении скорости затухания, относительно мала по сравнению с выборочной ошибкой скорости затухания в том случае, когда наилучшее значение весовых множителей для длинного ряда выбирается по данным для коротких отрезков этого ряда (содержащих не менее 50 результатов наблюдений).

Рассмотрим метод экспоненциального сглаживания для полинома  $y_t = \sum_{i=0}^p \frac{a_i}{i!} t^i$ , представленного  $n$  членами ( $t=1, 2, \dots, n$ ).

Экспоненциальной средней первого порядка для ряда  $y_t$  назовем

$$S_t^{[1]}(y) = \alpha + \sum_{i=0}^n (1-\alpha)^i y_{t-i},$$

где  $0 < \alpha < 1$  — параметр сглаживания, выбирается априори.

Экспоненциальная средняя  $k$ -го порядка для  $y_t$

$$S_t^{[k]}(y) = \alpha \sum_{i=0}^n (1-\alpha)^i S_{t-i}^{[k-1]}(y).$$

Справедлива следующая рекуррентная формула [11]:

$$S_t^{[k]}(y) = \alpha S_t^{[k-1]}(y) + (1-\alpha) S_{t-1}^{[k]}(y).$$

Откуда  $S_t^{[1]}(y) = \alpha y_t + (1-\alpha) S_{t-1}^{[1]}(y) = \alpha y_t + (1-\alpha)[\alpha y_{t-1} + (1-\alpha) S_{t-2}^{[1]}(y)] = \alpha y_t + \alpha(1-\alpha) y_{t-1} + (1-\alpha)^2 [\alpha y_{t-2} + (1-\alpha) S_{t-3}^{[1]}(y)] = \alpha \sum_{i=0}^{t-1} (1-\alpha)^i y_{t-i} + (1-\alpha)^t y_0$ .

Веса, присвоенные наблюдениям, убывают в геометрической прогрессии

$$S_t^{[2]}(y) = \alpha S_t^{[1]}(y) + (1-\alpha) S_{t-1}^{[2]}(y);$$

$$S_t^{[n]}(y) = \alpha S_t^{[n-1]}(y) + (1-\alpha) S_{t-1}^{[n]}(y).$$

Прогноз значений динамического ряда  $y_t$  в момент  $t+l$  после  $t=n$  строится с помощью ряда Тейлора:

$$\hat{y}_{t+l} = y_t^{(0)} + l y_t^{(1)} + \dots + \frac{l^p}{p!} y_t^{(p)},$$

где  $y_t^{(p)}$  —  $p$ -я производная, взятая в момент  $t$ .

Для линейной модели  $y_t = a_0 + a_1 t$  прогноз вычисляется как  $\hat{y}_{t+l} = \hat{a}_0 + l \hat{a}_1$ ; для квадратичной модели  $y_t = a_0 + a_1 t + \frac{1}{2} a_2 t^2$  получим  $\hat{y}_{t+l} = \hat{a}_0 + l \hat{a}_1 + \frac{1}{2} l^2 \hat{a}_2$ .

Оценки параметров  $\hat{a}_0, \hat{a}_1$  и  $\hat{a}_0, \hat{a}_1, \hat{a}_2$  находятся соответственно из систем уравнений, связывающих искомые параметры с экспоненциальными средними. В первом случае

$$\begin{cases} S^{[1]}(y) = \hat{a}_0 + \frac{1-\alpha}{\alpha} \hat{a}_1; \\ S_t^{[2]}(y) = \hat{a}_0 + \frac{2(1-\alpha)}{\alpha} \hat{a}_1. \end{cases}$$

Откуда

$$\hat{a}_0 = 2S_t^{[1]}(y) - S_t^{[2]}(y),$$

$$\hat{a}_1 = \frac{\alpha}{1-\alpha} [S_t^{[1]}(y) - S_t^{[2]}(y)].$$

Ошибка прогноза при этом определяется следующим образом:

$$\sigma(\hat{y}_{t+l}) = \sigma_0 \sqrt{\frac{\alpha}{(1-\alpha)^3} [1 + 4(1-\alpha) + 5(1-\alpha^2) + 2\alpha(4-3\alpha)l + 2\alpha^2 l^2]},$$

где

$$\sigma_0^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-p-1}.$$

Во втором случае получаем систему уравнений

$$\begin{cases} S_t^{[1]}(y) = \hat{a}_0 - \frac{1-\alpha}{\alpha} \hat{a}_1 + \frac{(1-\alpha)(2-\alpha)}{2\alpha^2} \hat{a}_2; \\ S_t^{[2]}(y) = \hat{a}_0 - \frac{2(1-\alpha)}{\alpha} \hat{a}_1 + \frac{(1-\alpha)(3-2\alpha)}{\alpha^2} \hat{a}_2; \\ S_t^{[3]}(y) = \hat{a}_0 - \frac{3(1-\alpha)}{\alpha} \hat{a}_1 + \frac{3(1-\alpha)(4-3\alpha)}{2\alpha^2} \hat{a}_2. \end{cases}$$

Откуда

$$\hat{a}_0 = 3[S_t^{[1]}(y) - S_t^{[2]}(y)] + S_t^{[3]}(y);$$

$$\hat{a}_1 = \frac{\alpha}{2(1-\alpha)^2} [(6-5\alpha) S_t^{[1]}(y) - 2(5-4\alpha) S_t^{[2]}(y) + (4-3\alpha) S_t^{[3]}(y)];$$

$$\hat{a}_2 = \frac{\alpha^2}{(1-\alpha)^2} [S_t^{[1]}(y) - 2S_t^{[2]}(y) + S_t^{[3]}(y)].$$

Ошибка прогноза рассчитывается по формуле

$$\sigma(\hat{y}_{t+i}) \approx \sigma_0 \sqrt{2\alpha + 3\alpha^2 + 3\alpha^3 t^2}.$$

Для применения этих формул необходимо знать начальную величину в рекуррентной формуле  $S_{t-1}^{[k]}(y)$ . Обычно она определяется из анализа только части динамического ряда. При отсутствии предварительных соображений рекомендуется для нахождения начальных условий пользоваться следующими формулами:

для линейной модели

$$S_{t-1}^{[1]}(y) = a_0 - \frac{1-\alpha}{\alpha} a_1;$$

$$S_{t-1}^{[2]}(y) = a_0 - \frac{2(1-\alpha)}{\alpha} a_1;$$

для квадратичной

$$S_{t-1}^{[1]}(y) = a_0 - \frac{1-\alpha}{\alpha} a_1 + \frac{(1-\alpha)(2-\alpha)}{2\alpha^2} a_2;$$

$$S_{t-1}^{[2]}(y) = a_0 - \frac{2(1-\alpha)}{\alpha} a_1 + \frac{(1-\alpha)(3-2\alpha)}{\alpha^2} a_2;$$

$$S_{t-1}^{[3]}(y) = a_0 - \frac{3(1-\alpha)}{\alpha} a_1 + \frac{3(1-\alpha)(4-3\alpha)}{2\alpha^2} a_2.$$

Значения параметров для этих формул рекомендуется получать методом наименьших квадратов.

Выбор начальных условий сказывается на оценке погрешностей метода. Другой важной проблемой является выбор параметра сглаживания  $\alpha$ . При  $\alpha$ , близких к единице, — учитываются только последние члены ряда; при  $\alpha$  близких к нулю, практически все члены ряда. Точных рекомендаций по выбору значений  $\alpha$  нет.

Метод экспоненциального сглаживания разработан для рядов, состоящих из большого числа наблюдений, при увеличении числа наблюдений точность прогноза должна возрастать. При анализе коротких рядов метод не «срабатывает», так как часто не «успевает» отразить изменения при быстрых темпах роста. Если явление протекает в одних и тех же условиях, то точность прогноза определяется величиной периода предистории явления (базисного периода) и длительности прогнозируемого периода. Количественно это влияние можно оценить по имеющемуся ряду динамики,

если одну часть ряда рассматривать как предисторию, а вторую — как прогнозируемую. Получив прогнозирующую функцию по базисному периоду, по второй части ряда можно оценить реальные ошибки прогноза. Изменяя число элементов рядов предистории и прогноза, получим зависимость точности прогноза от периода предистории и величины прогнозируемого периода.

### 3.5. Многофакторное прогнозирование

Рассмотрим динамическое многофакторное прогнозирование. Многофакторные динамические модели должны учитывать пространственные и временные изменения факторов (аргументов), а также (при необходимости) запаздывание влияния этих факторов на зависимую переменную (функцию). Многофакторное прогнозирование позволяет учитывать развитие взаимосвязанных процессов и явлений. Основой его является системный подход к изучению исследуемого явления, а так же процесс осмысливания явления как в прошлом, так и в будущем. В многофакторном прогнозировании одной из основных проблем является проблема выбора факторов, обуславливающих поведение системы, которая не может быть решена чисто статистическим путем, а только при помощи глубокого изучения существа явления. Здесь следует подчеркнуть примат анализа (осмысливания) перед чисто статистическими (математическими) методами изучения явления.

В традиционных методах (например, в методе наименьших квадратов) считается, что наблюдения независимы друг от друга (по одному и тому же аргументу). В действительности существует автокорреляция и ее учет приводит к неоптимальности статистических оценок, затрудняет построение достоверных интервалов для коэффициентов регрессии, а также проверку их значимости.

Автокорреляция определяется по отклонениям от трендов. Она может иметь место, если не учтено влияние существенного фактора или нескольких менее существенных факторов, но направленных «в одну сторону», либо неверно выбрана модель, устанавливающая связь между факторами и функцией. Для выявления наличия автокорреляции применяется критерий Дурбина—Уотсона

$$d = \frac{\sum_{i=1}^n (\epsilon_{i+1} - \epsilon_i)^2}{\sum_{i=1}^n \epsilon_i^2},$$

где  $\epsilon_i$  — случайные отклонения от регрессионной модели.

Для исключения или уменьшения автокорреляции применяется переход к случайной компоненте (исключение тренда) или введение времени в уравнение множественной регрессии в качестве аргумента.

В многофакторных моделях возникает проблема и мультиколлинеарности — наличие сильной корреляции между факторами, которая может существовать вне всякой зависимости между функцией и факторами.

При мультиколлинеарности между аргументами существует (линейная) связь, и наряду с уравнением регрессии имеются и другие (линейные) отношения; искажается физический смысл коэффициентов регрессии, появляется некорректность — слабая обусловленность систем алгебраических уравнений. Соответствующие диагональные элементы матрицы, обратной матрице вторых производных от функции правдоподобия, определяющие дисперсии оценок коэффициентов регрессии стремятся в бесконечность и порождают ошибочную (в данном случае) тенденцию исключения существенных переменных из уравнения регрессии.

Мультиколлинеарность определяется как степень отклонения от ортогональности множества независимых переменных (факторов). Для проверки наличия мультиколлинеарности во множестве независимых переменных используется показатель

$$\chi^2 = - \left[ n - 1 - \frac{1}{6} (2p + 5) \right] \ln |X^T X|,$$

где  $X$  — матрица факторов (аргументов);  $p$  — число факторов;  $n$  — число наблюдений.

Показатель  $\chi^2$  приближенно имеет распределение  $\chi^2$ .

Главное же состоит в последующем выявлении тех факторов, которые наиболее сильно взаимозависимы. Это можно сделать по диагональным элементам названной ранее матрицы, обратной матрице вторых производных от функции правдоподобия. Вводится величина

$$\Phi = (c_{ii} - 1) \left( \frac{n-p}{p-1} \right),$$

где  $c_{ii}$  —  $i$ -й диагональный элемент матрицы. Величина  $\Phi$  имеет  $F$ -распределение с  $(n-p)$  и  $(p-1)$  степенями свободы.

Выявив, какие факторы являются мультиколлинеарными, можно определить характер взаимозависимости между мультиколлинеарными элементами множества независимых переменных. Для этой цели используются недиагональные элементы обратной матрицы и полученные в них коэффициенты частной корреляции меж-

ду мультиколлинеарными переменными и всеми остальными переменными. Анализ проводится по  $t$ -критерию для коэффициента частной корреляции.

В многофакторном прогнозировании без методов конъюнктного анализа нельзя обойтись в принципе. Так, в многофакторном анализе необходимо наряду с оценкой параметров сглаживающей (исследуемой) функции построить прогноз каждого фактора (по неким другим функциям или моделям). Естественно, что значения факторов, полученные в эксперименте в базисном периоде не совпадают с аналогичными значениями, найденными по прогнозирующим моделям для факторов. Это различие должно быть объяснено либо случайными отклонениями, величина которых выявлена указанными различиями и должна быть учтена сразу же при оценке параметров сглаживающей функции, либо это различие не случайно и никакого прогноза делать нельзя, т. е. в задаче многофакторного прогнозирования исходные значения факторов как и значения сглаживающей функции должны быть взяты с соответствующими ошибками, закон распределения которых должен быть определен при соответствующем анализе, предшествующем процедуре прогнозирования.

Пример построения многофакторного прогноза. Рассмотрим задачу определения вида зависимости, связывающей в 1972—1980 гг. показатель использования производственной мощности междугородной телефонной связи  $\alpha$  с факторами: долей исходящих каналов автоматической системы коммутации  $d_a$ , телефонной плотности местной телефонной сети  $u$  и долей абонентов городской телефонной сети, имеющих возможность выхода на автоматическую междугородную телефонную связь  $d_{ab}$ . По полученной зависимости необходимо затем получить прогноз величины  $\alpha$  на 1981—1985 гг., зная оценки значений  $d_a$ ,  $u$  и  $d_{ab}$  и их дисперсий в эти годы. Эта задача многофакторного анализа. Строго говоря, построить прогноз в данной задаче без применения методов конъюнктного анализа нельзя. Дело здесь в том, что прогнозы значений факторов получают по своим моделям, и значения факторов, полученных в базовом периоде не будут совпадать со значениями тех же факторов, рассчитанными по моделям прогноза. Различие этих значений может быть объяснено только влиянием случайных процессов и необходим учитывать различий значений факторов в процессе анализа. Исходные данные для определения вида зависимости  $\alpha$  от  $d_a$ ,  $u$ ,  $d_{ab}$  приведены в табл. 3.1

Погрешность исходных данных принималась равной 5 единицам первого не указанного разряда после запятой. Рассматривалось множество функций, для каждой из которых определялись оценки свободных параметров. Затем по критерию  $\chi^2$  выбиралась наиболее приемлемая.



Таблица 3.1

Исходные данные для определения зависимости  $\alpha$  от  $d_a$ ,  $u$ ,  $d_{ab}$ 

$\alpha$ , тыс разг канал	$d_a$	$u$	$d_{ab}$
6,72	0,112	5,4	0,431
6,97	0,148	4,8	0,484
7,17	0,190	6,3	0,552
7,37	0,234	6,8	0,593
7,73	0,273	7,2	0,630
7,99	0,312	7,6	0,649
8,16	0,347	8,1	0,663
8,50	0,361	8,5	0,687
8,23	0,423	8,9	0,724

Таблица 3.2

«Истинные» значения факторов

$\hat{d}_a$	0,1121; 0,1479; 0,1903; 0,2346; 0,2730; 0,3103; 0,3464; 0,3610; 0,4248
$\hat{u}$	5,3811; 5,8120; 6,2622; 6,7592; 7,200; 7,6456; 8,1396; 8,5017; 8,8466
$\hat{d}_{ab}$	0,4537; 0,5028; 0,5460; 0,5840; 0,6175; 0,6470; 0,6730; 0,6959; 0,7161

Таблица 3.3

Прогнозируемые значения  $\hat{\alpha}$  и  $\sigma(\hat{\alpha})$ 

Годы	1986	1987	1988	1989	1990
$\hat{\alpha}$ , тыс. разг. канал	10,64	11,01	11,22	11,66	12,31
$\sigma(\hat{\alpha})$	0,29	0,31	0,32	0,33	0,34

В процессе анализа была выбрана функция

$$\alpha = a_1 + a_2 d_{ab} + a_3 u + a_4 d_a + a_5 d_a / u;$$

Оценки параметров вектора  $\mathbf{a}$  и их дисперсий получены следующие:  $\hat{a}_1 = 0,067$ ;  $\hat{a}_2 = 0,515$ ;  $\hat{a}_3 = 1,411$ ;  $\hat{a}_4 = -11,719$ ;  $\hat{a}_5 = 7,513$ ;  $D(\hat{a}_1) = 0,010$ ;  $D(\hat{a}_2) = 0,018$ ;  $D(\hat{a}_3) = 0,0003$ ;  $D(\hat{a}_4) = 0,085$ ;  $D(\hat{a}_5) = 0,008$ .

Значения  $\hat{\alpha}$  при найденных оценках вектора  $\mathbf{a}$  не отличались от заданных значений  $\alpha$  больше, чем на 0,9%, величина  $\chi^2 = 0,0034$ ; вероятность того, что модель не может быть отвергнута равнялась 0,999. «Истинные» значения факторов  $\hat{d}_a, \hat{u}, \hat{d}_{ab}$  даны в табл. 3.2.

Прогнозируемые значения  $\hat{\alpha}$  и их средние квадратические отклонения на 1981—1990 гг. приведены в табл. 3.3

### 3.6. Идентификация моделей типа АРПСС

Приступим к процедуре идентификации модели. Рассмотрим общую модель— модель АРПСС. Конкретная цель состоит в том, чтобы получить некоторое указание на то, какие значения  $p, d$  и  $q$  нужны в общей линейной модели АРПСС, и выбрать некоторые начальные значения параметров. Полученная таким образом пробная модель является отправной точкой для применения более формальных и эффективных методов оценивания описанных в гл. 1 и 2. Идентификация и оценивание перекрываются. Идентификация неизбежно неточна. Она неточна потому, что вопрос о том, какие типы моделей встречаются на практике и в каких обстоятельствах— это черта поведения физического мира и она не может быть решена чисто математическими рассуждениями. Так как на этапе идентификации нельзя точно сформулировать задачу, приходится прибегать к статистически «неэффективным» методам. На этом этапе особенно полезны графические методы. Предварительная идентификация не требует ничего, кроме опробования класса моделей, который будет позднее эффективно подгоняться и проверяться.

Например, идентифицировать подходящий класс моделей из общего семейства моделей АРПСС

$$\Phi(B) \Delta^d y_t = \theta_0 + \theta(B) a_t,$$

который может быть использован для описания данного ряда можно следующим образом:

а) будем брать конечную разность от  $y_t$  столько раз, сколько необходимо, чтобы обеспечить стационарность, в надежде свести

изучаемый процесс к смешанному процессу авторегрессии — скользящего среднего

$$\Phi(B)\omega_t = \theta_0 + \theta(B)d_t,$$

где

$$\omega_t = (1-B)^d y_t = \Delta^d y_t;$$

б) идентифицировать конечный ряд с помощью автокорреляционной и частной автокорреляционной функций. Эти функции используются также для получения приближенных оценок параметров чтобы получить начальные значения для используемых на этапе оценивания итеративных процедур. В процессе идентификации прежде всего надо оценить порядок разности  $d$ .

Автокорреляционная функция стационарного смешанного процесса авторегрессии — скользящего среднего удовлетворяет разностному уравнению

$$\Phi(B)\rho_k = 0, \quad k > q.$$

Кроме того, если  $\Phi(B) = \prod_{i=1}^p (1-G_i B)$ , то решение этого разностного

уравнения для  $k$ -й автокорреляции  $\rho_k$  в предположении об отсутствии кратности корней имеет вид

$$\rho_k = A_1 G_1^k + A_2 G_2^k + \dots + A_p G_p^k, \quad k > q - p.$$

Условие стационарности требует, чтобы нули  $\Phi(B)$  лежали вне единичного круга, и приводит к тому, что корни  $G_1, G_2, \dots, G_p$  лежат внутри единичного круга. Для стационарной модели, у которой ни один из корней не лежит близко к границе единичного круга, автокорреляционная функция быстро затухает при средних и больших  $k$ . Если хотя бы один действительный корень, например  $G_1$  приближается к 1, так что  $G_1 = 1 - \delta$ , где  $\delta$  малое положительное число, то поскольку для больших  $k$   $\rho_k \approx A_1(1 - k\delta)$ , автокорреляционная функция не будет быстро затухать, а будет спадать медленно и почти линейно. Следовательно, отсутствие у автокорреляционной функции тенденции к затуханию может рассматриваться как свидетельство того, что существует корень, близкий к 1. Выборочная автокорреляционная функция похожа на теоретическую. Отсюда — отсутствие затухания выборочной автокорреляционной функции логично истолковать в том смысле, что процесс  $y_t$  ведет себя нестационарно, хотя возможно, его разность  $\Delta y_t$  или какая-либо более высокая разность стационарна. Предполагается, что необходимая для получения стационарности степень разности  $d$  достигнута, если автокорреляционная функция ряда  $\omega_t = \Delta^d y_t$  быстро затухает. На практике  $d$  обычно равно 0, 1 или 2, и достаточно просмотреть примерно 20 первых значений автокорреляции исходного ряда, его первых и вторых разностей. При-

няв предварительное решение о величине  $d$ , по общему виду выборочных автокорреляционной и частной автокорреляционной функций соответствующего разностного ряда находим указания к выбору порядков  $p$  и  $q$  операторов авторегрессии и скользящего среднего, т. е. проведем идентификацию результирующего стационарного процесса АРСС.

В то время как автокорреляционная функция процесса авторегрессии порядка  $p$  спадает плавно, ее частная автокорреляционная функция имеет обрыв после  $p$ -й задержки. Но, автокорреляционная функция процесса скользящего среднего порядка  $q$  обрывается после задержки  $q$ , в то время как ее частная автокорреляция плавно спадает с ростом задержки.

Автокорреляционная функция смешанного процесса, содержащая компоненту авторегрессии порядка  $p$  и компоненту скользящего среднего порядка  $q$ , после первых  $q-p$  задержек представляется в виде суммы экспонент и затухающих синусоид. В то же время частная автокорреляционная функция смешанного процесса приближенно представляется суммой экспонент и затухающих синусоид после  $p-q$  задержек.

Выборочные автокорреляции могут иметь довольно большие дисперсии и быть сильно коррелированы друг с другом и нельзя ожидать детального сходства выборочной автокорреляционной функции с теоретической. В выборочной функции могут наблюдаться всплески и тренды, не имеющиеся в теоретической функции. При использовании выборочной автокорреляционной функции для идентификации обычно можно быть уверенным только в ее главных характеристиках. Поэтому может потребоваться исследование на этапах оценивания и диагностической проверки двух или более возможных моделей процесса. Важно знать какие отличия выборочной автокорреляционной функции допустимы. В частности, необходимо уметь оценить, становятся ли автокорреляции и частные автокорреляции практически нулевыми при задержках больше некоторого  $k$ . Для больших задержек можно вычислить стандартные ошибки оцениваемых автокорреляций по формуле

$$\delta[\rho_k] \approx \frac{1}{n^{1/2}} \{1 + 2(\rho_1^2 + \rho_2^2 + \dots + \rho_q^2)\}^{1/2}, \quad k > q,$$

в которой  $r_k$  — выборочная оценка  $k$ -й автокорреляции. Стандартная ошибка частной автокорреляции порядка  $p+1$  и выше равна

$$\delta[\hat{\Phi}_{kk}] \approx \frac{1}{n^{1/2}}, \quad k > p.$$

Даже для не слишком больших  $n$  выборочный коэффициент автокорреляции распределен примерно нормально со средним значением нуль. При гипотезе, что теоретическая автокорреляция рав-

на нулю, оценка  $\rho_k$ , деленная на ее стандартную ошибку, также будет распределена примерно по нормальному закону с единичной дисперсией. То же справедливо и для частных автокорреляций. Эти факты можно использовать для получения нестрогих правил проверки, являются ли теоретические автокорреляции и частные автокорреляции при задержках, больших некоторой, практически нулевыми.

Необходимо иметь в виду следующее:

1) по рядам ограниченной длины никогда не удастся доказать, что корень оператора авторегрессии точно равен единице;

2) не существует, конечно, резкого перехода от стационарного поведения ряда к нестационарному.

В неясных случаях некоторыми преимуществами может обладать использование нестационарных моделей. Это, в частности, верно в задачах и прогнозирования, и регулирования. Когда  $\Phi_1$  близко к единице, мы фактически не знаем, имеет ли среднее значение ряда какой-либо смысл или нет. Следовательно, более удобно использовать нестационарную модель, которая не содержит среднего значения  $\mu$ . Если будем использовать такую модель, прогнозы будущего поведения не будут зависеть от выборочного среднего, найденного по предыдущему периоду, которое может не иметь никакого отношения к будущему уровню ряда.

### 3.7. Методы получения и уточнения прогнозов

Методы прогнозирования должны приводить практически к одному результату — выбору одного и того же варианта из ряда возможных и одинаковой оценке последствий выбранного варианта. Правильное ведение прогнозирования предполагает, в частности, применение соответствующих критериев при пересмотре прогнозов.

Для уточнения прогноза применяется метод адаптивного сглаживания, основы которого были заложены при разработке метода экспоненциального сглаживания и затем развиты путем включения в рассмотрение сезонных моделей. Этот метод целесообразно использовать в тех случаях, когда явления достаточно хорошо описываются с помощью постоянного уровня, тренда и сезонных моделей, которые не претерпевают резких изменений для большинства рядов. Во многих случаях адаптивное сглаживание позволяет получать хорошие результаты в течение короткого промежутка времени, что иногда имеет более важное значение, чем получение более точного прогноза в течение длительного времени, но с применением значительных усилий.

Построение прогнозов по векторной модели. Как отмечалось, модель, позволяющая проводить прогнозирование

для любых рядов, имеет вид  $\mathbf{Y} = \mathbf{a}^T \mathbf{F}$ , где  $\mathbf{a}^T$  — вектор  $m$  коэффициентов, вычисленных отдельно для каждого временного ряда,  $\mathbf{F}$  — матрица соответствующих функций размерности  $m \times T$ , строки которой соответствуют каждому из  $m$  членов модели, и столбцы — каждому из  $T$  интервалов прогнозирования. Компоненты вектора  $\mathbf{Y}$  представляют собой прогнозируемые значения для каждого из интервалов.

Коэффициенты  $\mathbf{a}^T$  вычисляются сначала по методу наименьших квадратов в соответствии с  $N$  исходными данными ( $m \times N$ ) матрицы  $\mathbf{F}$ . Гармоники включаются в модель вплоть до той частоты, при которой квадрат амплитуды составляет по крайней мере 0,25 от величины дисперсии остатков, получаемых при подгонке этой модели с учетом имеющихся степеней свободы для ошибок.

С каждым новым наблюдением вектор  $\mathbf{a}_t$  заменяется на вектор  $\mathbf{a}_t = \mathbf{L}^T \mathbf{a}_{t-1} + \mathbf{h}\mathbf{e}$ , где  $\mathbf{L}$  — матрица перехода ( $n \times m$ ), которая определяется с помощью выражения  $f(t) = \mathbf{L}f(t-1)$  и переводит значения соответствующих функций из одного периода времени в следующий. Столбец матрицы  $\mathbf{F}$ , имеющий номер  $t$ , обозначается как  $f(t)$ , и  $\mathbf{e}$  представляет собой ошибку (скаляр) самого последнего прогноза, сделанного на один интервал вперед:  $\mathbf{e} = \mathbf{y}_t - \mathbf{y}_{t-1}$ . Вектор  $\mathbf{h}$ , характеризующий  $m$  постоянных сглаживания, определяется следующим образом. Используя  $z$ -преобразования и некоторые элементарные тригонометрические соотношения, можно вывести выражение для матрицы  $\mathbf{H}$ :  $\mathbf{H} = \sum_j \alpha^j f(-j) f^T(-j)$ . Так как  $\alpha < 1$ , то значимость более старых данных снижается и увеличивается вес, приписываемый более поздним данным.

Следовательно, вектор сглаживания можно записать как  $\mathbf{h} = \mathbf{H}^{-1} \mathbf{f}(0)$ . Величины  $\mathbf{L}$  и  $\mathbf{h}$  необходимо вычислить только один раз, так как при дальнейшем пересмотре прогнозов они не изменяются.

Если прогнозы составляются для переменных интенсивности и результаты наблюдений получены в нерегулярные интервалы времени, то исходные данные преобразуются к величинам, характеризующим скорость их изменения на равномерных интервалах времени. Получаемые в результате прогнозы могут быть приведены в соответствие с величиной будущих интервалов времени. Заметим, что подобное использование сечения какой-либо входной переменной одинаково пригодно для любых рядов и не приводит к увеличению дисперсии, какое имело бы место в случае получения сечения из данных предыстории.

Получение прогноза для модели АРПСС. Для того чтобы в полной мере использовать достоинства метода требуются достаточно длинные ряды (содержащие не менее 100 результатов наблюдений); это в значительной степени помогает понять изучаемый процесс с точки зрения получения предваритель-

ных данных относительно возможных изменений процесса в будущем. С помощью данного метода может быть проанализирован гораздо более широкий круг моделей, чем с помощью метода адаптивного сглаживания. Кроме того, он особенно полезен в тех случаях, когда требуется получить как можно более точный прогноз.

Анализ моделей начинается с формирования разностей из исходных данных  $y_t$  и вычисления частных автокорреляционных функций  $\Phi$  для разностных рядов  $\omega_t = \Delta^d y_t$ . Затем разностные ряды вводятся в модель с соответствующим числом коэффициентов авторегрессии и скользящего среднего:

$$\omega_t = \varphi_1 \omega_{t-1} + \dots + \varphi_p \omega_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}.$$

Для оценки коэффициентов модели такого типа и для вычисления дисперсии шума в результатах наблюдений используется максимизация функции правдоподобия.

Операторы разности скользящего среднего могут быть разложены таким образом, чтобы выразить модель прогнозов через исходные данные наблюдений:

$$y_t = \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + \theta_0 + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}.$$

Сначала прогнозируется изменение преобразованных входных переменных, а затем осуществляется обратное преобразование.

Если необходимо проанализировать сезонные циклы, то вклад от каждого из них вычисляется путем составления разностей для определенных отрезков цикла. Для годового цикла, например, в качестве таких отрезков можно принять месяцы, и тогда прогноз для каких-то месяцев в будущем будет основываться на известных данных для этого месяца в прошлые годы.

Приведем основные формулы для получения прогнозов, вероятностных пределов для прогнозов и для подправления прогнозов. Пусть  $y_t$  будет отклонением наблюденного ряда от любой известной детерминированной функции  $f(t)$ . В частности, для стационарного ряда функция  $f(t)$  может быть равна среднему значению ряда  $\mu$  или нулю.

В последнем случае  $y_t$  образуют наблюденный ряд. Рассмотрим общую модель АРПСС

$$\Phi(B) \Delta^d y_t = \theta(B) a_t$$

или

$$\varphi(B) y_t = \theta(B) a_t.$$

Пусть известны значения ряда до момента  $t$ . Тогда прогноз  $\hat{y}_t(l)$  ( $l > 0$ ) с минимальной среднеквадратической ошибкой — это условное математическое ожидание величины  $y_{t+l}$  при заданных значениях  $y_t; y_{t-1}; \dots$

$$\hat{y}_t(l) = [y_{t+l}] \equiv M[y_{t+l} | y_t, y_{t-1}, \dots].$$

Отсюда следует, что ошибки прогноза с упреждением, равным единице (на шаг вперед), — это не коррелированные между собой импульсы, генерируемые моделью.

На практике простейший способ вычисления прогнозов — непосредственное использование разностного уравнения, а именно

$$\hat{y}_t(l) = \varphi_1 [y_{t+l-1}] + \dots + \varphi_{p+d} [y_{t+l-p-d}] + [a_{t+l}] - \theta_1 [a_{t+l-1}] - \dots - \theta_q [a_{t+l-q}]. \quad (3.1)$$

Для вычисления условных математических ожиданий вместо  $y$ , известных к этому моменту, подставляются их значения; вместо будущих значений  $y$  — их прогноз, вместо известных  $a$  — их значения и вместо будущих  $a$  — нули. Процесс прогнозирования может быть начат аппроксимацией неизвестных  $a$  нулями.

Пусть требуется найти прогнозы с упреждением  $1, 2, \dots, l$  и их доверительные пределы как бесконечную взвешенную сумму текущего и предшествующих импульсов  $a_j$  [2]:

$$z_{t+l} = \sum_{j=-\infty}^{t+l} \psi_{t+l-j} a_j = \sum_{j=0}^{\infty} \psi_j a_{t+l-j},$$

где  $\psi_0 = 1$  и остальные веса  $\psi$  находят из уравнения

$$\varphi(B) \psi(B) = \theta(B),$$

приравнивая коэффициенты при одинаковых степенях  $B$  в выражении

$$(1 - \varphi_1 B - \dots - \varphi_{p+d} B^{p+d}) (1 + \psi_1 B + \psi_2 B^2 + \dots) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q).$$

Зная значения  $\varphi$  и  $\theta$ , после оценки их по результатам наблюдений, используем выражения для  $\psi$ :

$$\begin{aligned} \psi_1 &= \varphi_1 - \theta_1, \\ \psi_2 &= \varphi_1 \psi_1 + \varphi_2 - \theta_2, \\ &\vdots \\ \psi_j &= \varphi_1 \psi_{j-1} + \dots + \varphi_{p+d} \psi_{j-p-d} - \theta_j, \end{aligned}$$

где  $\psi_0 = 1$ ;  $\psi_j = 0$  при  $j < 0$  и  $\theta_j = 0$  при  $j > q$ . Если  $k$  — наибольшее из целых  $p+d-1$  и  $q$ , то при  $j > k$  веса  $\psi$  удовлетворяют разностному уравнению

$$\psi_j = \varphi_1 \psi_{j-1} + \varphi_2 \psi_{j-2} + \dots + \varphi_{p+d} \psi_{j-p-d}.$$

Следовательно,  $\psi$  вычисляются рекуррентным способом. Подставляем вес  $\psi$  в  $y_{t+l}$ , получим

$$y_{t+l} = \hat{y}_t(l) \pm u_{t/2} \left( 1 + \sum_{j=1}^{l-1} \psi_j^2 \right)^{1/2} \sigma(a)$$

и получаем значение прогноза для каждого нужного уровня вероятности  $\epsilon$  и для каждого упреждения  $l$ . Здесь  $\sigma(a)$  можно заменить практически его оценкой  $s(a)$  — выборочным стандартным отклонением белого шума  $a_t$ , а  $u_{\epsilon/2}$  — квантиль уровня  $1-\epsilon/2$  стандартного нормального распределения.

Дисперсия ошибки прогноза на  $l$  шагов вперед для любого момента  $t$ , определяемая как математическое ожидание величины

$$e_t^2(l) = \{y_{t+l} - \hat{y}_t(l)\}^2,$$

равна

$$V(l) = \left\{ 1 + \sum_{j=1}^{l-1} \psi_j^2 \right\} \sigma^2(a).$$

Из предположения, что  $a$  подчиняется нормальному закону, следует, что при известных значениях процесса до момента  $t$  условное распределение вероятности  $P(y_{t+l} | y_t, y_{t-1}, \dots)$  будущего значения процесса  $y_{t+l}$  будет также нормальным со средним значением  $\hat{y}_t(l)$  и стандартным отклонением

$$\left\{ 1 + \sum_{j=1}^{l-1} \psi_j^2 \right\}^{1/2} \sigma(a).$$

Подправка прогнозов. Можно выразить прогнозы  $\hat{y}_{t+1}(l)$  и  $\hat{y}_t(l+1)$  будущего наблюдения  $y_{t+l+1}$ , сделанные в моменты  $t+1$  и  $t$ :

$$\begin{aligned} \hat{y}_{t+1}(l) &= \psi_l a_{t+1} + \psi_{l+1} a_t + \psi_{l+2} a_{t-1} + \dots, \\ \hat{y}_t(l+1) &= \psi_{l+1} a_t + \psi_{l+2} a_{t-1} + \dots \end{aligned}$$

Вычитая второе уравнение из первого, получаем

$$\hat{y}_{t+1}(l) - \hat{y}_t(l+1) = \psi_l a_{t+1},$$

Отсюда видно, что если прогноз величины  $y_{t+l+1}$ , сделанный в момент  $t$ , «подправить» добавлением ошибки прогноза на шаг вперед  $a_{t+1}$  с коэффициентом  $\psi_l$ , то получится прогноз той же величины  $z_{t+l+1}$ , но в момент  $t+1$ .

Положим, что делаются прогнозы в момент  $t$  с упреждениями  $1, 2, \dots, l$ . Затем, как только становится известным  $y_{t+1}$ , можно найти  $a_{t+1} = y_{t+1} - \hat{y}_t(1)$  и пропорционально подправить прогнозы  $\hat{y}_{t+1}(l) = \hat{y}_t(l+1) + \psi_l a_{t+1}$  в момент  $t+1$  для упреждений  $1, 2, \dots, l-1$ . Новый прогноз  $y_{t+1}(l)$  для упреждения  $l$  этим способом найти нельзя, но он легко вычисляется по прогнозам с меньшими упреждениями из разностного уравнения.

Когда становится известным новое отклонение  $y_{t+1}$ , прогноз для момента  $t+1$  может быть скорректирован. Для этого нужно вычислить новую ошибку прогноза  $a_{t+1} = y_{t+1} - \hat{y}_t(1)$  и использовать разностное уравнение (3.1), где  $t$  заменено на  $t+1$ . Другой возможный метод — использовать прогнозы  $\hat{y}_t(1), \hat{y}_t(2), \dots, \hat{y}_t(l)$  на момент  $t$ , получить первые  $l-1$  прогнозов  $\hat{y}_{t+1}(1), \hat{y}_{t+1}(2), \dots, \hat{y}_{t+1}(l-1)$  на момент  $t+1$  из формулы

$$\hat{y}_{t+1}(l) = \hat{y}_t(l+1) + \psi_l a_{t+1}$$

и затем найти последний прогноз  $\hat{y}_{t+1}(l)$ , используя разностное уравнение (3.1).

Другие способы представления прогнозов:

1. Прогнозы в проинтегрированной форме. Для  $l > q - p - d$  прогнозы задаются одной функцией

$$\hat{y}_t(l) = b_0^{(t)} y_0(l) + b_1^{(t)} y_1(l) + \dots + b_{p+d-1}^{(t)} y_{p+d-1}(l), \quad (3.2)$$

определяемой «опорными» значениями  $\hat{y}_t(q), \hat{y}_t(q-1), \dots, \hat{y}_t(q-p-d+1)$ , где  $\hat{y}_t(-j) = y_{t-j}$  ( $j=0, 1, 2, \dots$ ). Если  $q > p+d$ , первые  $q-p-d$  прогнозов не лежат на этой функции. В общем случае стационарный оператор авторегрессии приводит к появлению затухающих экспоненциальных и синусоидальных членов, а нестационарный оператор  $\Delta^d$  — полиномиальных членов.

Подстраивающиеся коэффициенты  $b_j^{(t)}$  в (3.2) при переходе от момента  $t$  к моменту  $t+1$  могут быть скорректированы на величину, зависящую от последней ошибки прогноза на шаг вперед  $a_{t+1}$  согласно общей формуле

$$b^{(t+1)} = L^T b^{(t)} + g a_{t+1},$$

где  $L = F_l^1 F_{l+1}$ ;  $g = F_l^{-1} \psi_l$ ;  $b^{(t)} = (b_0^{(t)}, b_1^{(t)}, \dots, b_{p+d-1}^{(t)})$ ;  $\psi_l = (\psi_l, \psi_{l+1}, \dots, \psi_{l+p+d})^T \dots$

2. Прогноз как взвешанная сумма прошлых наблюдений. С теоретической точки зрения полезно представить прогноз как взвешенную сумму прошлых наблюдений. Так, если модель представлена в обращенной форме

$$a_t = \pi(B) y_t = (1 - \pi_1 B - \pi_2 B^2 - \dots) y_t,$$

где веса  $\pi$  можно получить из выражения  $\varphi(B) = (1 - \pi_1 B - \pi_2 B^2 - \dots) \theta(B)$ , приравняв коэффициенты при одинаковых степенях  $B$ , прогноз с упреждением  $l$  равен

$$\hat{y}_t(l) = \pi_1 y_t + \pi_2 y_{t-1} + \dots,$$

а прогнозы для больших упреждений можно получить из выражения

$$\hat{y}_t(l) = \pi_1[y_{t+l-1}] + \pi_2[y_{t+l-2}] + \dots,$$

где условные математические ожидания известных  $y$  надо заменить их значениями, а для будущих  $y$  — их прогнозом.

Прогноз для любого упреждения можно записать как линейную функцию имеющихся наблюдений. Тогда

$$\hat{y}_t(l) = \sum_{j=1}^{\infty} \pi_j^{(l)} y_{t+j-1},$$

где  $\pi_j^{(l)}$  — функции весов  $\pi$ .

### 3.8. Байесовские прогнозы

При байесовском подходе каждому наблюдению до его проведения ставится в соответствие ряд первичных вероятностных значений коэффициентов модели. Построение модели можно начинать из состояния полной неопределенности, так как даже в этом случае метод позволяет довольно точно устанавливать вероятностные значения коэффициентов.

После того как получены результаты наблюдений  $Y$ , по правилу Байеса определяют апостериорные вероятности, основываясь на которых вычисляют распределение вероятностей прогнозируемой величины  $Y_l$ .

Апостериорная вероятность получить прогнозируемое значение  $Y_l$  при заданной априорной вероятности  $P(Y_l)$  и условной вероятности  $P(Y/Y_l)$  равна

$$P(Y_l/Y) = \frac{P(Y_l) \cdot P(Y/Y_l)}{\int P(Y_l) P(Y/Y_l) dY}.$$

Такой подход позволяет игнорировать кратковременные изменения прогнозируемой переменной и в то же время четко регистрировать те изменения в основном процессе, которые на графике временной зависимости изображаются в виде ступеней и изменения наклона соответствующих кривых. Многократное повторение приведенной схемы вычислений дает возможность уменьшить число вероятностных распределений до числа состояний изучаемого процесса. В число состояний обычно входят нейтральное состояние, которое характеризуется неизменными значениями всех элементов процесса, и состояния, которые соответствуют различным возможным значениям каждого коэффициента и помех. В этих состояниях дисперсия распределения намного выше, чем в нейтральном состоянии, что указывает на отклонение того или иного ко-

эффициента модели от нормы. Возможность уменьшения числа состояний существенна, так как в противном случае число распределений возрастало бы пропорционально квадрату числа наблюдений.

Объем вычислений при использовании рассматриваемого метода гораздо больше, чем при адаптивном сглаживании, но успешное прогнозирование гарантируется и в тех случаях, когда могут возникнуть непредвиденные и существенные изменения в прогнозируемом процессе. Если есть предварительные сведения об ожидаемых событиях, то можно начать прогнозирование при весьма скромных данных или вообще при их отсутствии. Вместе с тем, если имеется большой объем результатов наблюдений, прогнозирование можно начать при отсутствии исходных предпосылок о поведении прогнозируемой величины. Алгоритм позволяет построить модель процесса по данным его предыстории.

Типичная модель для результатов текущих наблюдений может быть представлена выражением  $d_t = \mu_t + \varepsilon_t$ , где  $\mu$  — уровень полезной составляющей и  $\varepsilon$  — шум текущего наблюдения. Текущий уровень связан с трендом  $\beta_t$  выражением  $\mu_t = \mu_{t-1} + \beta_t + \gamma_t$ , где тренд изменяется по закону  $\beta_t = \beta_{t-1} + \delta_t$ . Будем учитывать возможность скачкообразного изменения  $\gamma$  основного уровня, изменение  $\delta$  тангенса угла наклона кривой и шума  $\varepsilon$ . Предполагается, что скачкообразное изменение  $\gamma$  основного уровня, изменение  $\delta$  тангенса угла наклона кривой, а также шум  $\varepsilon$  независимы и распределены по нормальным законам с нулевым средним и известными (но не обязательно постоянными) дисперсиями  $V_\varepsilon$ ,  $V_\gamma$  и  $V_\delta$ .

В модели, которая описывает много состояний  $j$ , обычно имеются два состояния для каждого возможного возмущения, причем каждое из них характеризуется своим значением дисперсии  $V^j$ . В устойчивом состоянии дисперсия мала, а в состоянии изменения велика. Имеются также априорные вероятности  $\pi_j$  того, что состояние  $j$  не изменяется. Ниже приведен пример четырех состояний и соответствующих им условных вероятностей и дисперсий, использованных авторами метода в качестве типичных начальных

Характеристики состояния	$j$	$\pi_j$	$R_\varepsilon$	$R_\gamma$	$R_\delta$
Стабильное	1	0,900	1	0	0
Скачкообразное изменение уровня	2	0,003	1	100	0
Постепенное изменение тангенса угла наклона	3	0,003	1	0	1
Переходное	4	0,094	101	0	0

условий. Дисперсия шума  $V_0$  зависит от соотношения между фактическими изменениями процесса и текущим прогнозом и описывается соотношением  $R^{(i)} = V^{(i)}/V_0$ .

Пусть вектор  $\varphi$  содержит следующие элементы: ожидаемое значение уровня, ожидаемое значение тангенса угла наклона, дисперсию ошибок при определении уровня, ковариацию наклона и уровня, а также дисперсию ошибок при определении тангенса угла наклона. Тогда, если в момент времени  $t-1$  совместное распределение уровня и тангенса угла наклона является двумерным нормальным и характеризуется параметрами  $\varphi_{t-1}$ , то соответствующее распределение в момент времени  $t$  также будет двумерным нормальным, а его параметры будут определяться вектором  $\varphi_t$ .

Существуют также выражения для коэффициентов сезонных сечений при анализе сезонных рядов методом Байеса [9].

### 3.9. Анализ сезонных рядов

К сезонным относят такие явления, которые обнаруживают в своем развитии закономерности, более или менее регулярно повторяющиеся в определенные промежутки времени. В анализе сезонных явлений ставятся следующие задачи: численно выразить проявление сезонных колебаний, вскрыть факторы, вызывающие сезонные колебания, дать прогноз сезонных колебаний. Сезонные ряды можно представить в виде компонент: тенденций (тренд), сезонные волны (кратковременные колебания) и помехи (случайные колебания). Тенденция отражает общее изменение ряда за длительный промежуток времени (постоянный подъем или постоянное снижение).

Сезонные волны — это более или менее регулярные изменения временного ряда и повторяющиеся с небольшими изменениями через определенные промежутки времени. Сезонные колебания обычно имеют постоянный период. Случайные колебания вызываются внешними случайными причинами. Они искажают тенденцию, а также сезонные и циклические колебания. Выявив закономерные изменения динамики составляющих динамического (временного) ряда, можно их использовать для экстраполяции значений ряда на перспективу.

Методы нахождения основной тенденции развития явления (тренда) идентичны рассмотренным: методу наименьших квадратов, методу скользящих средних и т. д. В последнем случае интервал сглаживания определяется периодом сезонных колебаний.

Для описания сезонных моделей существуют два совершенно разных подхода, каждый из которых имеет множество вариантов.

Если дано  $N$  результатов наблюдений за какой-то период времени, то можно записать  $N$  поправочных членов (положительных и отрицательных) или  $N$  коэффициентов (каждый из которых больше или меньше единицы), которые затем либо суммируются с результатами прогноза, либо умножаются на него. Значения этих поправочных коэффициентов можно определить по результатам наблюдений в соответствующие моменты времени в прошлых циклах.

Весь ряд поправочных членов или коэффициентов может быть назван сезонным «сечением». Во многих случаях оказывается, что существует небольшое семейство сечений, и тогда прогнозы для отдельных временных рядов могут быть получены путем выбора наиболее подходящего представления этого семейства.

Сечения играют важную роль в методике прогнозирования, так как их сущность легко объяснить, а эффективность их использования легко оценить на практике. Однако существенным недостатком применения сечений является связанная с ними нестабильность прогнозов. Если случайная величина  $x$  имеет дисперсию  $\sigma^2(x)$ , то дисперсия обратной величины ( $y=1/x$ ) порядка  $\sigma^4(x)$ . В большинстве методов, использующих сезонные или циклические сечения, в процессе проведения вычислений необходимо деление на случайную переменную (или ее оценку, имеющую отличную от нуля дисперсию). Если при этом знаменатель имеет тот же порядок величины, что и помеха, то относительно малые колебания исходных данных могут привести к большим изменениям прогноза. Подобная нестабильность обычно проявляется как раз тогда, когда принимаемые решения могут иметь самые серьезные последствия (т. е. перед началом следующего цикла).

Для уменьшения подобной нестабильности используется предположение, согласно которому дисперсия помехи является функцией только среднего по всем наблюдениям за цикл и не зависит от сезонных изменений. При этом чувствительность метода к колебаниям данных в точках наивысших и наименьших значений прогнозируемых переменных соответственно повышается и понижается.

Другой подход заключается в представлении циклических изменений прогнозируемой переменной рядами Фурье, т. е.  $\hat{x}_t = \sum_k A_k \cos \omega kt + \sum_k B_k \sin \omega kt$ . Здесь  $A_k$  и  $B_k$  — коэффициенты, полученные по исходным данным с помощью соответствующей регрессии;  $\omega = 2\pi/p$  — основная частота;  $p$  — число наблюдений в одном сезонном цикле,  $t$  — время, за которое должен быть вычислен прогноз,  $k=1, 2, \dots, N$  и суммирование ведется по всем частотам вплоть до частоты Найквиста (наивысшей частоты гармонического разложения дискретного ряда, которая определяется половиной интервала между наблюдениями).

Достоинство такой модели состоит в том, что она обеспечивает стабильность прогноза даже в точках цикла с наименьшими значениями прогнозируемой переменной, так как коэффициенты вычисляются путем усреднения всего набора имеющихся данных, а не только результатов наблюдений в пределах одного цикла.

Если для составления прогноза используются только первые члены разложения Фурье, то сезонное сечение представляет собой простую синусоиду с некоторой амплитудой и фазой, зависящими от коэффициентов  $A_1$  и  $B_1$ . При увеличении числа членов разложения форма сечения может измениться. В соответствии с теоремой Фурье, если в дискретных рядах величина переменной повторяется через каждые  $p$  наблюдений, то такие ряды могут быть представлены точно  $p$  членами гармонического разложения.

Амплитуда  $j$ -й гармоники, включенной в модель, равна  $C_j = \sqrt{A_j^2 + B_j^2}$  (квадрат амплитуды иногда называют мощностью данной гармоники). Пусть  $\sigma^2$  — дисперсия остаточных разностей между исходными данными и результатами расчетов по модели, включающей гармоники вплоть до  $j$ -й. Тогда модель будет соответствовать предыстории, охватывающей от трех до шести предыдущих циклов, если удовлетворяется соотношение  $CA_j \geq 1/2\sigma$ . Проверка на соответствие предыстории обычно проводится по критерию наименьших квадратов. Отметим, что при вычислении  $\sigma^2$  необходимо учесть потерю степеней свободы некоторым числом переменных в модели. Это означает, что можно достичь минимальной дисперсии, используя модель с числом переменных, которое меньше  $p$ . Однако такая модель будет неверной, если мощность включенной в нее высшей гармоники не будет достаточно большой по сравнению с остаточной дисперсией.

Пусть сезонные волны описываются бесконечным рядом синусоидальных и косинусоидальных функций (ряды Фурье). Если в качестве переменной взять  $\left\{\frac{2\pi}{p} it\right\}$ ,  $i = \overline{1, p}$ , где  $p$  — полный исследуемый период, то значения ряда в момент  $t$  равны

$$y = a_0 + \sum_{i=1}^{i=p/2} \left[ A_i \cdot \sin\left(\frac{2\pi}{p} it\right) + B_i \cdot \cos\left(\frac{2\pi}{p} it\right) \right],$$

где коэффициенты  $a_0$ ,  $\{A_i\}$  и  $\{B_i\}$  определяются формулами

$$a_0 = \frac{1}{p} \sum_{i=1}^p y_i; \quad A_i = \frac{2}{p} \sum_{i=1}^p y_t \sin\left(\frac{2\pi}{p} it\right);$$

$$B_i = \frac{2}{p} \sum_{i=1}^p y_t \cos\left(\frac{2\pi}{p} it\right).$$

Отсюда  $a_0$  — среднее значение ряда за период. Для  $p$  дискретных наблюдений число гармоник не превышает  $p/2$ . Дисперсия, учитываемая одной гармоникой, равна  $\sigma_i^2 = \frac{1}{2} C_i^2$ ,  $C_i = \sqrt{A_i^2 + B_i^2}$ ;  $i = \overline{1, (p/2-1)}$ . Для последней гармоники  $\alpha_{p/2}^2 = C_{p/2}^2$ . Так как никакие две гармоники не коррелируют между собой, то они не будут учитывать одну и ту же часть общей дисперсии, т. е. дисперсии, учитываемые различными гармониками, складываются.

В процессе прогнозирования следует объединить и функцию тренда и функцию сезонных колебаний. При этом трудно оценить погрешность. Самое простое — перенести рассчитанную ошибку аппроксимации в будущее и получить доверительные интервалы.

Естественно, чем больший период берется для составления функции прогноза, тем больше надо определить гармоник, но тем сильнее будет осреднена амплитуда колебаний и полученные амплитуды могут не соответствовать амплитудам прогнозируемого периода. Здесь предпочтительнее взять небольшой промежуток, предшествующий предсказываемому.

Для прогнозирования сезонных рядов применима также модель АРПСС. Прогнозы удобнее всего вычислять прямо из самого разностного уравнения. Для частного случая сезонных колебаний с периодом в один год мультипликативная модель имеет вид [2]

$$y_{t+l} = y_{t+l-1} + y_{t+l-12} - y_{t+l-13} + a_{t+l} - \theta a_{t+l-1} - \Theta a_{t+l-12} + \theta \Theta a_{t+l-13}.$$

Прогноз  $[y_{t+l}] = M[y_{t+l} | \theta, \Theta, y_t, y_{t-1}, \dots]$  — условное математическое ожидание  $y_{t+l}$  в момент  $t+l$ . В этом выражении параметры считаются известными точно, и ряд  $y_t, y_{t-1}, \dots$  предполагается известным достаточно далеко в прошлое.  $\theta$  и  $\Theta$  — разные величины.

Практическое применение модели зависит от ряда фактов:

а) обратимые модели после подгонок к реальным рядам дают прогнозы, существенно зависящие только от сравнительно недавних значений ряда;

б) прогнозы не чувствительны к малым изменениям значений параметров, вносимым, например, ошибками оценивания.

Имеем

$$[y_{t+l}] = \begin{cases} y_{t+l} & j \leq 0 \\ \hat{y}_t(j) & j > 0, \end{cases}$$

$$[a_{t+l}] = \begin{cases} a_{t+l} & j \leq 0 \\ 0 & j > 0. \end{cases}$$

Отсюда, чтобы получать прогнозы, заменяем неизвестные  $y$  прогнозами, а неизвестные  $a$  — нулями. Известные  $a$  — это уже вы-



численные ошибки прогноза на шаг вперед, т. е.  $a_t = y_t - \hat{y}_{t-1}(1)$ ;  $\hat{y}_{t-1}(1)$  — прогноз  $y_{t-1+1}$ .

Веса  $\Psi$ . Чтобы определить формулы коррекции и получить дисперсию ошибки прогноза  $e_t(l)$ , нам нужно знать значения весов  $\Psi$  в выражении для модели вида  $y_t = \sum_{j=0}^{\infty} \Psi_j a_{t-j}$  [2].

Коррекция прогноза. Общая формула для коррекции прогноза имеет вид

$$\hat{y}_{t+1}(l) = \hat{y}_t(l+1) + \Psi_l a_{t+1}.$$

Дисперсия ошибок прогноза. Зная веса  $\Psi$ , можем вычислить дисперсию ошибок прогноза для любого упреждения  $l$ .

$$V(l) = [1 + \Psi_1^2 + \dots + \Psi_{l-1}^2] \sigma^2(a).$$

### 3.10. Диагностическая проверка моделей и ошибка прогноза

Диагностическая проверка моделей может быть начата с оценки параметров моделей. Оценку параметров по результатам наблюдений будем вести по алгоритмам, описанным в гл. 1 и 2. В общем случае перед применением формальной процедуры получения оценок, желательно исследовать поведение функции правдоподобия. В каждом случае функция правдоподобия содержит нужную нам информацию. Так, существование двух максимумов приблизительно равной высоты указывает, что здесь имеются две группы значений параметров, которые могут объяснить данные.

Необходимо соблюдать осторожность при интерпретации функции правдоподобия. В общем случае предположение, что логарифмическая функция правдоподобия вблизи ее максимума близка к квадратичной функции, не всегда применима. Однако эти примеры не типичны.

Особая осторожность необходима, когда максимум функции правдоподобия может лежать на границе или около нее. Первая производная функции правдоподобия в этом случае не равна нулю в максимуме и квадратичная аппроксимация определенно не является адекватным представлением функции правдоподобия.

Рассмотрим еще раз оценивание параметров при помощи теоремы Байеса. Теорема Байеса гласит, что если  $P(\theta)$  — априорное распределение вероятности для  $\theta$ , то апостериорное распределение вероятности  $P(\theta|Y)$  для  $\theta$  после получения данных  $Y$  можно следующим образом выразить через априорную вероятность и функцию правдоподобия:

$$P(\theta|Y) = \frac{P(\theta) L(Y|\theta)}{\int P(\theta) L(Y,\theta) d\theta}.$$

Интеграл от  $P(\theta|Y)$  равен единице. Апостериорная вероятность пропорциональна априорной вероятности, умноженной на функцию правдоподобия. В частности если согласиться, что в интересующей нас области аргумента любые значения  $\theta$  равновероятны,  $P(\theta)$  можно считать в этой области равномерным, и, следовательно,  $P(\theta|Y)$  будет пропорционально функции правдоподобия  $L$ . Говоря о равномерности, подразумевается, что распределение приближенно постоянно в области, где функция правдоподобия существенна, и не принимает очень больших значений вне этой области. Для случая нескольких параметров совместное априорное распределение параметров считается пропорциональным

$$|J_\theta|^{1/2} = \left| -M \left\{ \frac{\partial^2 L}{\partial \theta_i \partial \theta_j} \right\} \right|^{1/2}.$$

Точное выражение для функции правдоподобия временного ряда  $Y$  длиной  $N = n + d$  для процесса АРПСС ( $p, d, q$ ) имеет вид

$$L(\Phi, \theta|Y) = \sigma_a^{-n} f(\Phi, \theta) \exp \left\{ -\frac{S(\Phi, \theta)}{2\sigma_a^2} \right\},$$

где

$$S(\Phi, \theta) = \sum_{t=-\infty}^n [a_t | Y, \Phi, \theta]^2 = \sum_{t=-\infty}^n [a_t | W, \Phi, \theta]^2.$$

Если априорная информация о  $\sigma_a$ ,  $\Phi$  и  $\theta$  отсутствует, то, поскольку информация о  $\sigma_a$  не дает нам информации о  $\Phi$  и  $\theta$ , рационально принять априорное распределение  $\Phi$ ,  $\theta$  и  $\sigma_a$  в виде

$$P(\Phi, \theta, \sigma_a) \sim |J(\Phi, \theta)|^{1/2} \sigma_a^{-1}.$$

Отсюда следует, что апостериорное распределение имеет вид

$$P(\Phi, \theta, \sigma_a | Y) \sim \sigma_a^{-(n+1)} |J(\Phi, \theta)|^{1/2} f(\Phi, \theta) \exp \left\{ -\frac{S(\Phi, \theta)}{2\sigma_a^2} \right\}.$$

Если теперь проинтегрировать  $P(\Phi, \theta, \sigma_a|Y)$  от нуля до бесконечности по  $\sigma_a$ , то мы получим точное совместное апостериорное распределение параметров  $\Phi$  и  $\theta$ :

$$P(\Phi, \theta|Y) \sim |J(\Phi, \theta)|^{1/2} f(\Phi, \theta) \{S(\Phi, \theta)\}^{-n/2}.$$

Если  $y_t$  есть процесс  $(p, d, 0)$ , то  $\omega_t = \Delta^d y_t$  будет чистым процессом авторегрессии порядка  $p$ . Для такого процесса множителя  $|J(\Phi)|^{1/2}$  и  $f(\Phi)$ , которые во всех случаях являются существенно менее важными, чем множитель  $S(\Phi)$ , практически взаимно сокращаются. При данных предположениях параметры  $\Phi$  процесса АР( $p$ ) для  $\omega$  имеют апостериорное распределение

$$P(\Phi | Y) \sim \{S(\Phi)\}^{-n/2}.$$

В этом случае поверхность для суммы квадратов, приближенно совпадающая с поверхностью функции правдоподобия при отсутствии априорной информации, совпадает также с поверхностью апостериорной вероятности.

После того как модель идентифицирована и параметры оценены, подгоняемая модель подвергается диагностической проверке. Один из полезных методов проверки модели состоит в использовании избыточного числа параметров, т. е. в оценивании параметров для несколько более общей модели, чем ожидаемая. Этот метод исходит из того, что мы можем угадать неадекватные свойства модели. Следовательно, полезно дополнить этот подход более общими способами проверки, основанными на остаточных ошибках после подгонки модели. Они позволяют найти в данных указание, какие изменения модели необходимы. Проверку можно провести, используя 1) автокорреляционную функцию остаточных ошибок; 2) кумулятивную периодограмму остаточных ошибок.

Если анализ не выявил, какие добавления нужны, это, конечно, не означает, что наша модель верна. Модель может лишь выдержать испытание.

Если будут обнаружены свидетельства неадекватности, возникает необходимость узнать, как нужно изменить модель на следующем итеративном цикле. Нужно обнаружить, что именно неадекватно в модели, чтобы узнать, как ее изменить.

При недостаточном объеме данных статистическое испытание может отвергнуть модель, которая тем не менее вполне адекватна решаемой задаче. Напротив, испытания могут не выявить серьезных отклонений от сделанных предположений, потому что эти испытания нечувствительны к некоторым видам отклонений.

Методы введения избыточных параметров путем обобщения модели в некотором направлении хороши в том случае, если мы знаем, какого рода отклонений следует бояться. Методики, менее зависящие от таких представлений, основываются на анализе остаточных ошибок. Хотя некоторые представления о том, чего искать, полезны и в этом случае, эти методы создают больше возможностей находить способ изменения моделей непосредственно из данных. Рассмотрим диагностическую проверку при помощи автокорреляций. Пусть подгонялась модель

$$\Phi(B)\tilde{\omega}_t = \theta(B)a_t$$

с  $\tilde{\omega}_t = \Delta^d y_t$ , и были получены оценки ММП параметров  $(\hat{\Phi}, \hat{\theta})$ . Тогда величины

$$\hat{a}_t = \hat{\theta}^{-1}(B)\hat{\Phi}(B)\tilde{\omega}_t$$

мы будем называть остаточными ошибками. Можно доказать, что для адекватной модели

$$\hat{a}_t = a_t + O\left(\frac{1}{\sqrt{n}}\right).$$

По мере увеличения длины ряда  $\hat{a}_t$  становится все ближе к белому шуму  $a_t$ . Следовательно, можно ожидать, что изучение  $\hat{a}_t$  может выявить и указать природу неадекватности модели. В частности, некоторые характерные черты выборочной автокорреляционной функции  $\hat{a}_t$  могут указать на необходимость определенных изменений модели.

Предположим, что модель верна, и мы знали точные значения параметров  $\Phi$  и  $\theta$ . Тогда выборочные автокорреляции  $\rho_k(\mathbf{a})$  ряда  $\mathbf{a}$  должны быть некоррелированы и распределены приближенно нормально относительно нулевого среднего значения с дисперсией  $n^{-1}$  и, следовательно, со стандартной ошибкой  $n^{-1/2}$ . Эти факты можно использовать для оценки статистической значимости кажущихся отклонений этих автокорреляций от нуля.

На практике мы не знаем истинных значений параметров. Мы располагаем только выборочными оценками  $(\hat{\Phi}, \hat{\theta})$ , по которым можем вычислить не  $\mathbf{a}$ , а  $\hat{\mathbf{a}}$ . Автокорреляции  $\rho_k(\hat{\mathbf{a}})$  ряда  $\hat{\mathbf{a}}$  могут дать ценную информацию о недостаточно хорошей подгонке и возможной природе неадекватности модели. Однако может оказаться рискованным придавать статистическую значимость кажущимся отклонениям этих автокорреляций  $\hat{\rho}_k(\hat{\mathbf{a}})$  от их теоретических нулевых значений исходя из стандартной ошибки  $n^{-1/2}$ , соответствующей  $\hat{\rho}_k(\hat{\mathbf{a}})$ . Показано, например, что для процесса  $AP(1)$  с параметром  $\Phi$  дисперсия  $\hat{\rho}_k(\hat{\mathbf{a}})$  равна  $\Phi^2 n^{-1}$ , что может оказаться существенно меньше, чем  $n^{-1}$ . Хотя во всех случаях для малых задержек возможно уменьшение дисперсии и сильная корреляция  $\hat{\rho}_k(\hat{\mathbf{a}})$ , при больших задержках эти эффекты быстро исчезают. Поэтому использование  $n^{-1/2}$  в качестве стандартной ошибки для  $\hat{\rho}_k(\hat{\mathbf{a}})$  будет приводить к недооценке статистической значимости кажущегося отклонения от нуля для автокорреляций при малых задержках, но обычно вполне оправдано для средних и больших задержек.

При не слишком больших задержках можно рассматривать  $n^{-1/2}$  как верхнюю границу стандартных ошибок  $\hat{\rho}_k(\hat{\mathbf{a}})$ , а не как сами стандартные ошибки. Если мы пользуемся  $n^{-1/2}$  как стандартной ошибкой  $\hat{\rho}_k(\hat{\mathbf{a}})$  при малых задержках, то можем серьезно недооценить значимость кажущихся расхождений. Во многих случаях предпочтительнее не рассматривать отдельные  $\hat{\rho}_k(\hat{\mathbf{a}})$ , а

оценить неадекватность модели по нескольким автокорреляциям, рассматриваемым как единое целое.

В других случаях, особенно при подгонке сезонных рядов кумулятивная периодограмма является эффективным средством обнаружения периодического отклонения от случайности [2, 11].

Нормированная кумулятивная периодограмма  $C(\rho_j)$  ряда  $a_t$  определяется как

$$C(\rho_j) = \frac{\sum_{t=1}^j I(\rho_t)}{ns^2},$$

где

$$I(\rho_t) = \frac{2}{n} \left[ \left( \sum_{i=1}^n a_t \cos 2n\rho_t t \right)^2 + \left( \sum_{i=1}^n a_t \sin 2n\rho_t t \right)^2 \right];$$

$\rho_t = i/n$  — частота;  $s^2$  — оценка  $\sigma^2(a)$ .

В любом из методов прогнозирования рядов собственно прогноз представляет собой, по существу, оценку ожидаемого распределения результатов наблюдений в будущем. Для того чтобы на основе полученного прогноза можно было принимать решение, в большинстве случаев необходимо знать исходное распределение. Если это распределение описывается стандартной функцией, то, определив один или два его параметра, можно оценить вероятности возможных результатов будущих наблюдений.

Для целей формального анализа часто удобно принять предположения о том, что помеха (ошибка) в исходных данных подчиняется нормальному закону распределения. Однако это предположение нельзя считать обоснованным для всех наблюдений.

Прогнозируемая выходная переменная характеризуется распределением, дисперсия которого зависит от дисперсии шума в исходных данных. Однако, поскольку прогнозы вычисляются на основе результатов многих наблюдений, часто принимается, что входная переменная описывается нормальным законом распределения независимо от того, каков вид распределения, описывающего шум в исходных данных.

Распределение ошибок при прогнозировании зависит как от распределения самой выходной переменной, так и от распределения соответствующего ей шума, который обычно считается не зависящим от шума в исходных данных. При прогнозировании суммы значений переменной интенсивности в течение времени упреждения, которое намного больше интервала между наблюдениями распределение ошибок представляет собой распределение суммы выборок из нескольких таких распределений и с возраста-

нием времени упреждения все более и более стремится к нормальному.

В большинстве случаев, представляющих практический интерес, исходное распределение может быть восстановлено из полученного по прогнозу среднего значения и какого-либо другого параметра, характеризующего распределение выходной переменной.

В качестве такого параметра, как правило, выбирается стандартное отклонение.

Если имеется какой-то набор прогнозируемых рядов, то обычно оказывается, что для каждого ряда существует определенная зависимость между разбросом ошибок и уровнем прогноза.

Среднее абсолютное отклонение представляет собой оценку отклонений, которая имела бы место, если бы прогнозы разрабатывались на основе минимизации суммы абсолютных величин остаточных разностей между результатами наблюдений и медианой распределения вместо обычно проводимой минимизации суммы квадратов отклонений от среднего. Среднее абсолютное отклонение использовалось вместо стандартного отклонения только потому, что его вычисление требовало меньше времени и меньшего объема памяти. Особенно широко этот показатель использовался в программах ЭВМ для прогнозирования спроса. При современных ЭВМ применение среднего абсолютного отклонения нецелесообразно.

Если распределение описывается нормальным законом, то среднее абсолютное отклонение  $m = 0,8\sigma$  (точное значение коэффициента при  $\sigma$  равно  $\sqrt{2/\pi} = 0,7979$ ). Так как средняя ошибка прогноза должна быть равна нулю, то дисперсия определяется среднеквадратичной ошибкой. Выбрав подходящую модель процесса, можно пересматривать значения коэффициентов этой модели по среднеквадратичной ошибке каждый раз, когда поступает новая информация и заново измеряется ошибка прогноза.

Если шум результатов наблюдений автокоррелирован незначительно, то дисперсия ошибок прогноза переменной интенсивности в течение времени упреждения будет равна сумме дисперсий ошибок прогнозов для отдельных интервалов, на которые разбивается время упреждения  $l$  с целью периодического уточнения прогнозов. Следовательно, стандартное отклонение, используемое для принятия решений, приблизительно равно  $\sigma\sqrt{l}$ . В тех случаях, когда имеет место автокорреляция шума, выражение для стандартного отклонения может быть записано в виде  $\sigma_t = l^p \sigma_1$ , где  $\sigma_1$  — стандартное отклонение ошибок одного знака,  $l$  — время упреждения, измеренное в интервалах между проверками прогноза, и  $\delta$  — постоянная, которая является характеристикой всего семейства прогнозируемых рядов.

Как уже отмечалось, точность прогноза зависит от шума входных переменных. При этом разумно считать, что прогнозы не содержат систематической ошибки, так как средняя ошибка прогноза должна быть равна нулю. Пока модель соответствует изучаемому процессу ошибки прогноза должны колебаться около нуля. Если же модель прогноза ошибочна или станет ошибочной из-за того, что резко изменится сам процесс, то появится последовательность «положительных» или «отрицательных» ошибок и, следовательно, средняя ошибка прогноза больше не будет равна нулю. Имеет смысл систематически осуществлять проверку близости средней ошибки прогноза к нулю. Если средняя ошибка превысит некоторый заранее установленный предел, пользователю необходимо принять меры для своевременной корректировки прогноза. Для осуществления такой корректировки прежде всего требуется установить конкретную причину смещения прогноза; после этого внести в модель достаточно обоснованные изменения; принять новые значения для одного или большего количества коэффициентов, изменить некоторые или все весовые множители, или, наконец, изменить сам вид модели.

### 3.11. Обнаружение выбросов в наблюдениях

Для обнаружения выбросов следящих сигналов и соответственно для выявления недопустимых отклонений средней ошибки прогноза могут быть использованы следующие виды контрольных величин.

**Кумулятивные суммы.** Величина кумулятивной суммы ошибок прогноза, задаваемой выражением  $Y_t = Y_{t-1} + e_t$  при  $Y_0 = 0$  колеблется около нуля, если прогнозы не смещены, и быстро возрастает при появлении последовательности ошибок прогноза одного знака. Для интенсивности затухания весовых множителей  $\beta$  в модели, которая может быть описана полиномом  $(n-1)$ -й степени, дисперсия кумулятивной суммы  $\sigma_Y^2 = [(1+\beta)/2(1-\beta^{2n})]\sigma_e^2$ , где  $\sigma_e^2$  — дисперсия ошибок прогноза. Выброс может быть зарегистрирован, например, тогда, когда выполняется соотношение  $|Y| \geq 3\sigma_Y$ . Однако очевидно, что в этом соотношении для уменьшения числа ложных регистраций вместо множителя 3 может быть использован множитель  $k > 3$ , а для уменьшения вероятности пропуска фактических изменений ошибок прогноза — множитель  $k < 3$ .

**Сглаживание ошибки.** Простое экспоненциальное сглаживание алгебраической суммы ошибок прогноза в виде  $Y_t = \beta Y_{t-1} + (1-\beta)e_t$  создает основу для регистрации выброса, например в случае выполнения соотношения  $|Y| \geq 0,4m$  (вместо множителя 0,4 может быть использован и какой-либо другой мно-

житель). При этом среднее абсолютное отклонение  $m$  заменяется величиной  $m_t = \beta m_{t-1} + (1-\beta)|e_t|$ .

Такой подход имеет два существенных преимущества по сравнению с предыдущим. Предположим, во-первых, что произошла единственная большая ошибка, но все же не настолько большая, чтобы вызвать выброс следящего сигнала. Начиная с этого момента при использовании первого метода величина (невзвешенной) кумулятивной суммы колебалась бы не около нуля, а около уровня, соответствующего этой единственной большой ошибке. Так что даже при вполне допустимых ошибках последующих прогнозов должны были бы регистрироваться выбросы следящих сигналов. Допустим, во-вторых, что следящий сигнал близок к выбросу (кумулятивная сумма близка к одному из пределов) и мы получили идеальный прогноз. При применении первого метода величина невзвешенной кумулятивной суммы не изменилась бы, однако оценка стандартного отклонения  $\sigma_t$  должна была бы уменьшиться, что привело бы к сжатию области допустимых отклонений и, следовательно, увеличению вероятности регистрации выбросов следящего сигнала. Использование сглаженных ошибок позволяет успешно преодолеть обе указанные трудности.

Дисперсия сглаженных ошибок прогноза может быть выражена соотношением  $\sigma_Y^2 = [(1-\beta)/(1+\beta)^2]\sigma_e^2$ .

**V-образные маски.** Маски подобного типа были разработаны на основе метода последовательного анализа Вальда. V-образная маска помещается вершиной вблизи самого последнего значения (невзвешенной) кумулятивной суммы ошибок прогноза, при этом ее раскрытая часть обращена в сторону предыдущих членов ряда. Если хотя бы одно из предыдущих значений кумулятивной суммы лежит вне маски, то это является признаком значительного смещения прогнозов. Угол наклона и точка пересечения двух прямых, составляющих V-образную форму маски, вычисляются на основе анализа ошибок и вероятностей пропустить значительное отклонение и ошибочно обнаружить какое-то отклонение, когда его на самом деле не было.

Внешние воздействия и ошибки, имевшие место при регистрации исходных данных, могут привести к тому, что результаты последующих наблюдений нельзя будет использовать для уточнения прогноза. Если отклонение полученного результата от прогнозируемого значения превышает  $4\sigma$ , то рекомендуется по возможности скорректировать входные переменные и только после этого продолжать прогнозирование на основе выбранной модели.

## ЛИТЕРАТУРА

1. **Бородюк В. И., Вошинин А. П.** Ошибки регистрации независимых переменных в задачах множественной регрессии // Заводская лаборатория. 1973. Т. 39, № 7.— С. 217—222.
2. **Бокс Дж., Дженкинс Г.** Анализ временных рядов. Прогноз и управление. / Пер. с англ. / Под ред. В. Ф. Писаренко.— М.: Мир, 1974.— 406 с.
3. **Браунли К. А.** Статистическая теория и методология в науке и технике / Пер. с англ.; Под ред. Л. Н. Большева.— М.: Наука, 1977.— 407 с.
4. **Грешилов А. А.** Некорректные задачи цифровой обработки информации и сигналов.— М.: Радио и связь, 1984.— 161 с.
5. **Грешилов А. А.** Анализ и синтез стохастических систем. Параметрические модели и конфлюентный анализ.— М.: Радио и связь, 1991.— 320 с.
6. **Грешилов А. А.** Метод наименьших квадратов и элементы конфлюентного анализа: Учебное пособие.— М.: Изд-во МГТУ, 1990.— 67 с.
7. **Жилинская Е. И., Товмаченко Н. Н., Федоров В. В.** Методы регрессионного анализа при наличии ошибки в предикторных переменных.— М.: Научный совет по комплексной проблеме «Кибернетика», 1978.— 25 с.
8. **Исследование операций.** Методологические основы и математические методы. Пер. с англ. / Под ред. И. М. Макарова, И. М. Бескровного — т. 1.— М.: Мир, 1981.— 712 с.
9. **Исследование операций.** Модели и применения. / Пер. с англ. / Под ред. И. М. Макарова, И. М. Бескровного — т. 2.— М.: Мир, 1981.— 677 с.
10. **Кендалл М., Стьюарт А.** Статистические выводы и связи / Пер. с англ. / Под ред. А. Н. Колмогорова.— М.: Наука, 1973.— 590 с.
11. **Кильдишев Г. С., Френкель А. А.** Анализ временных рядов и прогнозирование.— М.: Статистика, 1973.— 103 с.
12. **Крамер Г.** Математические методы статистики / Пер. с англ. / Под ред. А. Н. Колмогорова.— М.: Мир, 1975.— 648 с.
13. **Линник Ю. В.** Метод наименьших квадратов и основы математико-статистической теории обработки информации.— М.: Физматгиз,— 1958.— 333 с.
14. **Морозов В. А.** Линейные и нелинейные некорректные задачи // Итоги науки и техники: Математический анализ.— Л.— М.; ВИНТИ, 1973.— 400 с.
15. **Тихонов А. Н., Арсенин В. Я.** Методы решения некорректных задач.— М.: Наука, 1979.— 288 с.
16. **Худсон Д.** Статистика для физиков / Пер. с англ. / Под ред. Е. М. Лейкина.— М.: Мир, 1970.— 296 с.

17. **Frisch R.** Statistical confluence analysis by means of complete regression systems — Oslo: 1934 — 192 p
18. **Williamson J. H.** Least — squares fitting of straight line // Canad J Phys. 1968 T 46, № 16,— 1845 p.
19. **Yohai V. J., Zamar R. H.** Bounded influence estimation in the error — in — variables model // statist Anal Meas Error Models and Appl: Proc AMS — IMS — STAM It Summer Res Conf, Arcata, Calif., June 10—16, 1989 — Providence, 1990 — P 243—248