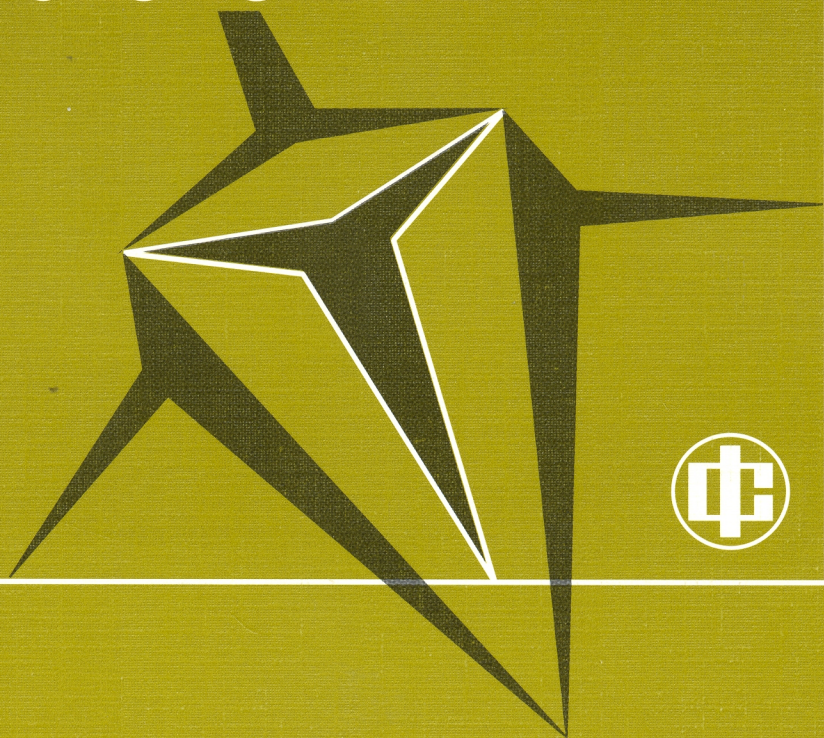


Эк

Е.П. Чураков

**МАТЕМАТИЧЕСКИЕ
МЕТОДЫ
ОБРАБОТКИ
ЭКСПЕРИМЕНТАЛЬНЫХ
ДАННЫХ
В ЭКОНОМИКЕ**



Е.П.Чураков

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ В ЭКОНОМИКЕ

8652

Рекомендовано
Учебно-методическим объединением
по образованию в области
математических методов в экономике
в качестве учебного пособия
для студентов высших учебных заведений,
обучающихся по специальности
061800 "Математические методы в экономике"
и другим экономическим специальностям



МОСКВА
"ФИНАНСЫ И СТАТИСТИКА"
2004

РЕЦЕНЗЕНТЫ:

кафедра экономической теории
Рязанской государственной радиотехнической академии;
Б. А. Лагоша,
доктор экономических наук,
профессор, заведующий кафедрой
экономико-математического моделирования МЭСИ

Чураков Е. П.

Ч-93 Математические методы обработки экспериментальных данных в экономике: Учеб. пособие. — М.: Финансы и статистика, 2004. — 240 с.: ил.

ISBN 5-279-02745-6

Материал пособия базируется на результатах обработки разнообразной информации, определяющей состояние экономических объектов. Большое внимание уделено различным методам оценивания регрессионных параметров. Известные алгоритмы прогнозирования стохастических рядов, основанные на моделях типа AR, MA, ARMA, ARIMA, обобщаются в форме матрично-векторной модели в терминах стохастического вектора состояния, и на ее основе строится рекуррентный алгоритм прогнозирования калмановского вида.

Для студентов экономических специальностей вузов и аспирантов, выполняющих научные исследования в области математических методов.

Ч 1602090000 — 238
010(01) — 2004 302 — 2003

УДК 330.4(075.8)
ББК 65в6я73

ISBN 5-279-02745-6

© Е. П. Чураков, 2004

ПРЕДИСЛОВИЕ

Управленческое решение, как правило, опирается на анализ эмпирических сведений, почерпнутых и представленных тем или иным образом и содержащих в себе информацию, необходимую для его принятия. В простейших ситуациях анализ имеющихся в распоряжении лица, принимающего решение, сведений может быть осуществлен непосредственно автором решения в соответствии с его профессиональной подготовкой, опытом, интуицией, даром предвидения и т.п. Однако сложные ситуации, в том числе характеризующиеся большим объемом данных, часто противоречивых, с не всегда известной степенью взаимодействия, что характерно для большинства реальных экономических явлений и процессов, не позволяют достаточно глубоко проникнуть в их суть только средствами не вооруженного современными математико-вычислительными методами человеческого разума. Новые информационные технологии, базирующиеся на последних достижениях в области вычислительных средств и мощном аппарате прикладной математики, позволяют выработать глубоко обоснованные формализованные рекомендации по принятию конкретного решения, создавая тем самым средства его поддержки.

В общем виде процедура выработки рекомендаций по принятию решения может быть сведена к следующей триаде:

1) разработке средствами математического моделирования формализованной математической модели, с помощью которой должно приниматься решение на основании содержательного существа проблемы и с учетом всех существенных ограничений и имеющихся эмпирических данных, создающих основу принимаемого решения;

2) выбору из арсенала средств прикладной математики метода обработки данных с использованием уже построенной их математической модели и разработкой соответствующего алгоритмического обеспечения;

3) машинной реализации разработанных алгоритмов и передача получаемых рекомендаций заинтересованным лицам.

В настоящем учебном пособии рассматриваются вопросы, частично касающиеся первого элемента этой триады и более фундаментально — второго: математических методов обработки экс-

периментальных данных. Пособие рассчитано прежде всего на студентов специальности 061800 «Математические методы в экономике». Поэтому предполагается, что подлежащие обработке данные имеют «экономическое происхождение». Как следствие, в пособии отражается основная проблематика задач, связанных с обработкой экономических «измерений», и излагаются наиболее характерные математические методы решения этих задач. Хотя сами по себе математические методы носят универсальный характер и применимы для обработки экспериментальных данных практически любой природы, часто их адаптация к экономическим приложениям сопровождается определенными терминологическими построениями и специфической расстановкой акцентов на приоритетах как самих методов, так и способов используемых доказательств. Это обстоятельство учитывается в пособии, что тем не менее не лишает книгу возможности быть полезной студентам других специальностей и направлений, «сопричастных» проблемам обработки экспериментальных данных.

Традиционно включаемые в пособие материалы относят к научной дисциплине, называемой *эконометрикой*. Лаконичное определение этой дисциплины, как отмечается в курсе эконометрики Я.Р. Мангуса, П.К. Катышева и А.А. Перецкого, дать трудно. Тем не менее, следуя курсу прикладной статистики С.А. Айвазяна и В.С. Мхитаряна, условимся под эконометрикой понимать самостоятельную экономико-математическую научную дисциплину, позволяющую на основе экономической теории и исходных статистических данных и в соответствии с математико-статистическими методами придавать конкретное количественное содержание общим (качественным) закономерностям, обусловленным экономической теорией. Традиционно математико-статистические методы эконометрики, как отмечается в указанных курсах, включают в себя:

- классическую и обобщенную линейные модели множественной регрессии;
- классический и обобщенный методы наименьших квадратов;
- модели и методы статистического анализа временных рядов;
- системы одновременных эконометрических уравнений.

Если последнюю позицию этого перечня можно считать специфичным «детисем» эконометрики, то материалы, сосредоточенные в предшествующих трех позициях, активно развиваются

и используются во многих смежных научных направлениях, в частности в теории стохастических систем управления, в статистической радиотехнике, при вторичной обработке траекторной информации о движущихся объектах, при обнаружении сигналов в шумах, в задачах классификации ситуаций и др. Современный специалист в области математических методов не может в своей профессиональной деятельности ограничиваться только «эконометрическими» наработками, но для него весьма полезна информированность о методах и подходах к решению аналогичных задач в смежных областях. Поэтому в данном пособии, не затрагивающем проблематики одновременных уравнений, традиционные эконометрические подходы, широко и всесторонне освещаемые в прекрасных изданиях последнего времени, таких, как уже упоминавшийся курс С.А. Айвазяна и В.С. Мхитаряна, «Введение в эконометрику» К. Доугерти, «Эконометрика» под редакцией И.И. Елисеевой и других, пополняются некоторыми положениями, расширяющими кругозор будущего специалиста в области математических методов и способствующими формированию многоальтернативных подходов к решению задач. К элементам новизны можно отнести такие методы оценивания регрессионных параметров, как рекуррентный метод наименьших квадратов (МНК), максимум апостериорной плотности вероятностей, байесовский и минимаксный методы, проблему обусловленности МНК-оценок, проблему стохастической сходимости, калмановские методы прогнозирования временных рядов и др. Аналогичные соображения определили необходимость включения в пособие фрагментов функционального анализа, кратко затрагивающих понятия банаховых и гильбертовых пространств, ортогональных систем непрерывных и дискретных функций, обобщенных рядов Фурье и способствующих расширению класса моделей трендов и их математической интерпретации. Ограниченное внимание, уделяемое теории случайных процессов в таких дисциплинах, как «Теория вероятностей» и «Математическая статистика», вынудило автора предпослать стохастическим временным рядам элементы этой теории, в основном касающиеся той ее части, которую принято называть корреляционной. В целом представляется, что включенные в пособие материалы могут оказаться полезными при изучении таких «родственных» дисциплин, как «Эконометрика», «Методы социально-экономического прогнозирования», «Эконометрическое моделирование», преду-

смотренных Государственным образовательным стандартом специальности 061800 «Математические методы в экономике».

Материал книги изложен в двух частях, из которых первая посвящена математическим методам восстановления зависимостей по экспериментальным данным, а вторая — математическим методам обработки временных рядов.

В последние годы наблюдается обширная экспансия (в хорошем смысле этого слова) современных информационных технологий в эконометрические методы и их проникновение во все сферы человеческой деятельности. Эта несомненно плодотворная тенденция ярко проявляется и в экономических приложениях. На вооружении экономистов-математиков оказались многочисленные пакеты прикладных программ (например, Statistica, Mathcad, MatLab, Matematica и др.), средствами которых решаются практически все эконометрические задачи. Однако крайне важно, чтобы пользователь этих пакетов не оказался механическим приложением к ним, слепо использующим возможности современных вычислительных систем без глубокого погружения и проникновения в сущность соответствующих математических проблем и их алгоритмического завершения. Творческий подход к решению прикладных задач немыслим без фундаментальной подготовки специалиста в области математико-статистических методов, и самый совершенный вычислительный инструментарий не в состоянии компенсировать пробел в образовании, если таковой был заложен еще в вузовские времена.

Автор надеется, что настоящее пособие будет способствовать уменьшению вероятности появления дискомфортных ситуаций в общении исследователя и компьютера, порожденных недостаточной математической культурой первого. Успех будет сопровождать тех, кто сможет эффективно сочетать глубокую теоретическую подготовку с универсальными возможностями современных информационных технологий.

Автор выражает глубокую благодарность и признательность рецензентам книги — доктору экономических наук, профессору Б.А. Лагоше и кафедре экономической теории Рязанской государственной радиотехнической академии (заведующий кафедрой доктор экономических наук, профессор В.И. Терехин). Все замечания читателей автор примет с благодарностью.

Первая часть

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ВОССТАНОВЛЕНИЯ ЗАВИСИМОСТЕЙ ПО ЭКСПЕРИМЕНТАЛЬНЫМ ДАНЫМ

Глава 1

ЗАДАЧИ РЕГРЕССИОННОГО АНАЛИЗА И ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ ДАННЫХ

1.1. Проблема восстановления зависимостей по экспериментальным данным

Одной из важных проблем, возникающих при исследовании экономических процессов, явлений, систем и т.п. (обобщенно — объектов исследования (ОИ)), выступает поиск и математическое описание зависимостей между различными величинами, тем или иным образом взаимодействующими между собой и определяющими состояние изучаемого объекта. Так, при исследовании сельскохозяйственного процесса может представлять интерес зависимость урожайности некоторой культуры от количества вносимого удобрения; при решении социальных проблем исследованию может быть подвергнута зависимость среднедушевых сбережений от среднедушевых доходов семьи; при анализе технологического процесса небезынтересной может оказаться зависимость производительности технологической установки от некоторых характеристик используемого сырья и т.п. В подобного рода ситуациях все участвующие в формировании изучаемой проблемы величины целесообразно разбить на две группы. Входящие в *первую группу* величины непосредственно характеризуют состояние

объекта исследования, их количественные значения являются отражением наших представлений о том, насколько хорошо или плохо «выглядит» исследуемый объект. Эти величины изменяются под действием различного рода факторов, влияющих на них и образующих *вторую группу* величин. Так, в упомянутых примерах первая группа будет соответственно включать в себя урожайность, среднелюдские сбережения, производительность, вторая — количество удобрений, среднелюдские доходы, характеристики сырья. Механизм взаимодействия между величинами обеих групп, вообще говоря, неизвестен. Поэтому исследуемый объект можно представить как «черный ящик», причем величины первой группы удобно интерпретировать как своеобразные выходы «ящика», а величины второй группы — как входы этого «ящика» (рис. 1.1).

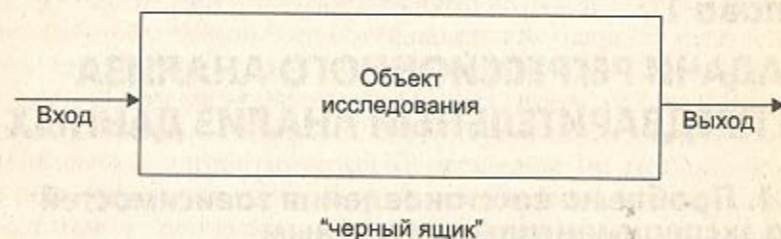


Рис. 1.1. Структура формирования эконометрических данных

Выходы, таким образом, определяются входами и внутренними свойствами ОИ, однако «заглянуть» внутрь «ящика» и выявить механизм взаимодействия выходов и входов в большинстве практических задач невозможно. Вместе с тем в подобных задачах удается установить численное соответствие между входами и выходами, понимаемое в следующем смысле: предполагается, что при различных известных значениях всех или некоторых входов можно измерить (зарегистрировать) соответствующие им численные значения выходов. Полученные экспериментальные данные теперь можно использовать для разработки математической модели, связывающей входы и выходы «черного ящика» без проникновения в существо последнего. В этом и заключается содержательный смысл проблемы восстановления зависимостей по экспериментальным (эмпирическим — полученным из эксперимента) данным.

Прежде чем двигаться далее, полезно преодолеть барьер терминологических особенностей, сопутствующих задаче восстановления зависимостей. В современном научном обиходе выходные величины принято называть зависимыми переменными, откликами, *эндогенными* (внутренними), результирующими, объясняемыми переменными. Входные переменные принято подразделять на две части. К первой из них относят те переменные, которые поддаются количественной оценке, т.е. в процессе проведения исследования их можно «измерить» и указать их численные значения. Эту часть входных переменных называют независимыми, факторами-аргументами, предикторами, *экзогенными* (внешними), объясняющими переменными. Вторая часть входных переменных не поддается экспериментальной регистрации, а о существовании части из них мы можем даже не подозревать. В совокупности их называют *латентными* (скрытыми) переменными и обычно интерпретируют как некие случайные величины с известными или неизвестными вероятностными свойствами.

Обозначим символом $Y = [Y^{(1)} Y^{(2)} \dots Y^{(k)}]^T \in R^k$ вектор эндогенных переменных, где, как обычно, R^k — k -мерное координатное пространство. Аналогично $X = [X^{(1)} X^{(2)} \dots X^{(s)}]^T \in R^s$ — вектор экзогенных переменных и $P = [P^{(1)} P^{(2)} \dots P^{(q)}]^T \in R^q$ — вектор латентных переменных. Здесь и далее T — символ транспонирования. Как уже отмечалось, в результате проведения некоторого (пассивного или активного) эксперимента можно зарегистрировать ряд значений вектора X и соответствующие значения вектора Y . Так, можно указать количество вносимых удобрений и соответствующую урожайность за ряд лет. Можно путем опроса многих семей выявить их среднелюдские доходы и соответствующие сбережения. При различных значениях характеристик сырья можно зарегистрировать производительность технологической установки и т.п.

Предположим, что эксперимент организуется так, что в процессе его проведения экзогенные переменные принимают значения $x_1, x_2, \dots, x_n$, а эндогенные переменные — значения $y_1, y_2, \dots, y_n$. Запись $X = x_i$ будет означать, что вектор X при i -м измерении принял конкретное значение $x_i = [x_i^{(1)} x_i^{(2)} \dots x_i^{(s)}]^T$, при этом $x_i^{(j)}$ — численное значение j -й компоненты $X^{(j)}$ вектора X при i -м измерении ($i = 1, 2, \dots, n$). Аналогична запись для вектора Y . В результате проведения эксперимента получим две числовые матрицы:

$$[x_1 \ x_2 \ \dots \ x_n] = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_n^{(2)} \\ \dots & \dots & \dots & \dots \\ x_1^{(s)} & x_2^{(s)} & \dots & x_n^{(s)} \end{bmatrix} \in R^{s \times n} \quad (1.1)$$

и

$$[y_1 \ y_2 \ \dots \ y_n] = \begin{bmatrix} y_1^{(1)} & y_2^{(1)} & \dots & y_n^{(1)} \\ y_1^{(2)} & y_2^{(2)} & \dots & y_n^{(2)} \\ \dots & \dots & \dots & \dots \\ y_1^{(k)} & y_2^{(k)} & \dots & y_n^{(k)} \end{bmatrix} \in R^{k \times n} \quad (1.2)$$

Здесь и далее символ $R^{a \times b}$ используется для обозначения множества матриц размерностью a на b .

Теперь можем более определенно сформулировать существо проблемы восстановления зависимостей. Располагая экспериментальными данными в объеме (1.1), (1.2), требуется разработать математическую модель, устанавливающую связь между эндогенными Y и экзогенными X переменными и учитывающую присутствие неконтролируемых латентных переменных. Разумеется, это чисто качественная формулировка замысла и она будет уточняться в процессе последующего изучения проблемы. Однако ряд вопросов, порожденных сформулированным намерением, можно выявить уже сейчас. Основные из них следующие.

1. С какой целью строится математическая модель?
2. Имеется ли вообще какая-либо связь между всеми из введенных переменных или только между некоторыми из них? Как оценить степень этой связи?
3. В классе каких математических моделей предполагается искать зависимость эндогенных переменных от экзогенных?
4. Как адаптировать выбранную в некотором классе модель к экспериментальным данным (1.1), (1.2), т. е. как установить конкретный вид модели по экспериментальным данным?
5. Как оценить эффективность построенной модели и в каком смысле эту эффективность следует понимать?

Наиболее просто дать ответ на первый вопрос. Будем полагать, что математическая модель строится для того, чтобы можно было вычислять значения эндогенных переменных при любых

значениях экзогенных переменных, не охваченных экспериментальными данными (1.1), (1.2), и при необходимости управлять эндогенными переменными путем выбора надлежащих значений экзогенных переменных. Подобные цели формируют задачи восстановления (по иной терминологии — интерполяции) и прогноза (предсказания, экстраполяции). Ответы на остальные вопросы дают методы регрессионного (и частично корреляционного) анализа.

1.2. Функция регрессии и регрессионная модель эндогенных переменных

Подчеркнем прежде всего, что эндогенные переменные даже при фиксированных (принявших определенные конкретные значения) экзогенных переменных являются случайными величинами. Это утверждение понимается в следующем смысле. Пусть в некотором эксперименте экзогенные переменные приняли значение x^* ($X = x^*$), а эндогенные переменные — y^* ($Y = y^*$). Если теперь провести второй эксперимент, в котором опять же обеспечить равенство $X = x^*$, то из-за влияния неучитываемых латентных переменных и неизбежных ошибок в результатах измерений окажется $Y \neq y^*$. Аналогичная ситуация возникнет в третьем и последующих экспериментах. Таким образом, последовательность значений эндогенных переменных, полученных в ряде экспериментов при одном и том же значении экзогенных переменных, следует рассматривать как реализации случайной величины Y , имеющей некоторую, как правило, неизвестную плотность вероятностей $\omega(y|X=x)$ или короче $\omega(y|x)$. Если бы условная плотность $\omega(y|x)$ случайной величины Y была известна, можно было бы попытаться найти значение эндогенной переменной Y , соответствующее значению экзогенной переменной $X = x$, поступив следующим образом. Отметим прежде всего, что точное значение эндогенной переменной Y из-за влияния латентных переменных принципиально найти нельзя. Можно отыскать некоторую величину $\bar{Y}$, в каком-то смысле близкую к Y , но не равную Y . Вектор $Y - \bar{Y}$ определяет отклонение того, что можно найти, от того, что хотелось бы найти. В таких случаях вектор $\bar{Y}$ называют оценкой вектора Y , а разность $Y - \bar{Y}$ — ошибкой оценивания. Величину ошибки оценивания как вектора принято характеризовать нормой $\|Y - \bar{Y}\|$, понимаемой для определенности в евклидовом

смысле. Эта норма в силу указанных причин является случайной величиной. Тогда в качестве меры близости величин Y и $\hat{Y}$ можно выбрать какую-либо неслучайную характеристику случайной величины $\|Y - \hat{Y}\|$ или, что математически удобнее, величины $\|Y - \hat{Y}\|^2$. Наиболее простой такой характеристикой является среднее значение (математическое ожидание), найденное при условии $X = x$, т.е. при условии, что экзогенные переменные равны величине x : $M\{\|Y - \hat{Y}\|^2 | X = x\}$, где M — символ усреднения. В развернутом виде

$$M\{\|Y - \hat{Y}\|^2 | X = x\} = \int_{-\infty}^{\infty} \|y - \hat{Y}\|^2 \omega(y | X = x) dy, \quad (1.3)$$

где дифференциал dy понимается как многомерный $dy = dy^{(1)} dy^{(2)} \dots dy^{(k)}$ и соответственно интеграл понимается как k -мерный.

Очевидно, оценка $\hat{Y}$ в среднем наиболее «близка» к Y , если она минимизирует величину (1.3), т.е. находится в результате решения оптимизационной задачи

$$M\{\|Y - \hat{Y}\|^2 | X = x\} \rightarrow \min_{\hat{Y} \in R^k}. \quad (1.4)$$

Если в качестве нормы принять евклидову, т.е. $\|Y - \hat{Y}\|^2 = (Y - \hat{Y})^T (Y - \hat{Y})$, то необходимое условие минимума в задаче (1.4) сведется к уравнению

$$\begin{aligned} \nabla M\{\|Y - \hat{Y}\|^2 | X = x\} &= M\{\nabla (Y - \hat{Y})^T (Y - \hat{Y}) | X = x\} = \\ &= -2M\{(Y - \hat{Y}) | X = x\} = 0_k, \end{aligned} \quad (1.5)$$

где ∇ — градиент по вектору $\hat{Y}$; 0_k — k -мерный нулевой вектор.

Так как в (1.5) вектор $\hat{Y}$ не зависит от Y , то из (1.5) непосредственно получаем

$$\hat{Y} = M\{Y | X = x\} = \int_{-\infty}^{\infty} y \omega(y | X = x) dy. \quad (1.6)$$

Функцию $\hat{Y} = \hat{Y}(x)$, определяемую соотношением (1.6) и по существу представляющую собой условное среднее вектора Y , принято называть *функцией регрессии*, или просто *регрессией* величины Y на X . Традиционно ее обозначают символом $m_Y(x)$. Заметим

еще раз, что здесь и далее выделенные полужирным шрифтом символы указывают на их векторную природу.

Таким образом, при известной условной плотности $\omega(y | X = x)$ величины Y в качестве оценки $\hat{Y}$ этой величины следует принять ее условное среднее — функцию регрессии. Тогда саму случайную величину Y при фиксированном значении экзогенных переменных $X = x$ можно представить как сумму ее среднего значения $m_Y(x)$ и некоторого случайного отклонения $\epsilon(x)$, случайная природа которого порождена как влиянием латентных переменных, так и случайными ошибками, неизбежно сопутствующими процессу измерения эндогенных переменных. Это слагаемое может в общем случае зависеть от значений экзогенных переменных, что и отображается в записи $\epsilon(x) = \epsilon(X = x)$.

Итак, в соответствии с проведенными построениями связь эндогенных и экзогенных переменных, косвенно учитывающая влияние латентных переменных, сводится к соотношению

$$Y = m_Y(X = x) + \epsilon(X = x). \quad (1.7)$$

Из (1.7) непосредственно следует $M\{\epsilon(x) | X = x\} = 0_k$, т.е. вектор ϵ является центрированным. Дополнительно предполагается, что при любых x ковариационная матрица этого вектора является конечной (все элементы матрицы ограничены).

Выражение (1.7) принято называть *регрессионной моделью* эндогенных переменных. Его можно было бы рассматривать как конечный результат наших исследований на пути построения искомой математической модели. К сожалению, это преждевременный оптимизм, так как условная плотность $\omega(y | X = x)$ реально неизвестна и, как следствие, неизвестна функция регрессии. Поэтому выражение (1.7) отражает принципиальную структуру искомой модели, но не содержит той конкретики, которая составила бы инструментарий практической деятельности и алгоритмическое руководство ею.

1.3. Модели, аппроксимирующие функцию регрессии. Гауссовская регрессия

Итак, практически построить модель (1.7) не удастся, так как неизвестна условная плотность $\omega(y | X = x)$ и, как следствие, функция регрессии $m_Y(X = x)$. Однако проблемы, порожденные при-

кладными задачами, заставляют искать выход из возникшей достаточно сложной ситуации. И этот выход таков.

Условимся прежде всего в последующем рассматривать случай одной эндогенной переменной, т.е. положим размерность k вектора Y равной единице. Соответственно функция регрессии в этом случае оказывается скалярной. Основная идея построения искомой модели заключается в следующем: зададимся некоторым классом Ξ функций, которому принадлежит неизвестная функция регрессии. Четко сформулированных и формально обоснованных рекомендаций по выбору этого класса функций нет. Поэтому выбор определяется опытом и интуицией исследователя, предварительным анализом экспериментальных данных (1.1), (1.2), априорными представлениями о сущности искомой зависимости и т.п. Подчеркнем тем не менее, что правильный выбор класса Ξ в значительной степени определяет успех всех последующих исследований, и поэтому удача здесь не только желательна, но и необходима. Наиболее апробированный подход заключается в формировании класса Ξ как множества определенных с точностью до вектора параметров Θ функций (параметрическое семейство функций) $\Xi = \{f(X; \Theta)\}$. Если действительно $m_Y(X=x) \in \Xi$, то параметры Θ можно подобрать так, что $m_Y(X=x) = f(X; \Theta)$. Если все же $m_Y(X=x) \notin \Xi$, то есть надежда на основе эмпирических данных (1.1), (1.2) подобрать такое значение Θ вектора Θ , что с приемлемой точностью, определяемой в том или ином смысле, можно представить $m_Y(X=x) \approx f(X; \Theta)$, и в обоих случаях вместо (1.7) положить

$$Y = f(X; \Theta) + \varepsilon(X), \quad (1.8)$$

считая при этом функцию $f(X; \Theta)$ аппроксимацией неизвестной регрессии.

Чтобы иметь какое-либо начальное представление о формирующемся классе Ξ функций, рассмотрим не претендующий на общность случай двух гауссовских скалярных величин Y и X и найдем регрессию Y на X , используя определение (1.6). В этом случае каждая из величин и совместно обе являются гауссовскими, т.е.

$$Y \sim N_y(m_y, \sigma_y^2), X \sim N_x(m_x, \sigma_x^2), Z = [Y X]^T \sim N_z(m_z, K_z),$$

где $N(\cdot)$ — символ гауссовской плотности с указанными в скобках параметрами.

В развернутом виде, как известно,

$$N_y(m_y, \sigma_y^2) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left\{-\frac{1}{2\sigma_y^2}(y-m_y)^2\right\}, \quad (1.9)$$

$$N_x(m_x, \sigma_x^2) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left\{-\frac{1}{2\sigma_x^2}(x-m_x)^2\right\}, \quad (1.10)$$

$$N_z(m_z, K_z) = \frac{1}{\sqrt{(2\pi)^2 |K_z|}} \exp\left\{-\frac{1}{2}(z-m_z)^T K_z^{-1}(z-m_z)\right\}, \quad (1.11)$$

где $|K_z|$ и K_z^{-1} — определитель ковариационной матрицы и обратная ковариационная матрица соответственно (подобные обозначения используются и далее).

В соответствии с определением представим

$$K_z = \begin{bmatrix} \sigma_y^2 & k_{yx} \\ k_{yx} & \sigma_x^2 \end{bmatrix},$$

и тогда совместная плотность вероятностей $\omega(y, x)$ величин Y и X после проведения соответствующих операций в (1.11) с учетом симметрии $k_{yx} = k_{xy}$ примет вид

$$\omega(y, x) = N_z(m_z, K_z) = \frac{1}{\sqrt{(2\pi)^2 (\sigma_y^2 \sigma_x^2 - k_{yx}^2)}} \times \quad (1.12)$$

$$\times \exp\left\{-\frac{1}{2(\sigma_y^2 \sigma_x^2 - k_{yx}^2)}(\sigma_x^2 \tilde{y}^2 - 2k_{yx} \tilde{y} \tilde{x} + \sigma_y^2 \tilde{x}^2)\right\},$$

где $\tilde{y} = y - m_y$, $\tilde{x} = x - m_x$ — центрированные величины.

Чтобы вычислить регрессию (1.6), необходима условная плотность $\omega(y|x)$, которая находится делением выражения (1.12) на выражение (1.10). Осуществив эту операцию, найдем показатель соответствующей экспоненты

$$\begin{aligned}
& -\frac{1}{2(\sigma_y^2\sigma_x^2 - k_{yx}^2)}(\sigma_x^2 y^2 - 2k_{yx}\tilde{y}\tilde{x} + \sigma_y^2 \tilde{x}^2) + 0,5\sigma_x^{-2}\tilde{x}^2 = \\
& = -\frac{\sigma_x^2}{2(\sigma_y^2\sigma_x^2 - k_{yx}^2)}(y - m_y - k_{yx}\sigma_x^{-2}(x - m_x))^2.
\end{aligned} \quad (1.13)$$

Так как гауссовская величина Y при фиксированном значении величины X по-прежнему остается гауссовской, то условная плотность $\omega(y|x)$ является гауссовской и может быть представлена в виде

$$\omega(y|x) = N_y(m_y(x), D_y(x)) = \frac{1}{\sqrt{2\pi D_y(x)}} \exp\left\{-\frac{1}{2D_y(x)}(y - m_y(x))^2\right\}.$$

Но тогда из сопоставления с (1.13) следует, что слагаемое $m_y + k_{yx}\sigma_x^{-2}(x - m_x)$ в (1.13) есть не что иное как условное математическое ожидание величины Y , т.е. функцию регрессии при гауссовских величинах Y и X удастся найти без проведения операций интегрирования (1.6), а путем анализа структуры условной гауссовской плотности $\omega(y|x)$. Итак, получаем:

$$m_y(x) = M\{Y | X = x\} = m_y + \frac{k_{yx}}{\sigma_x^2}(x - m_x). \quad (1.14)$$

При этом сомножитель $\frac{\sigma_x^2}{\sigma_y^2\sigma_x^2 - k_{yx}^2}$ в (1.13) определяет условную дисперсию $D_y(x)$ величины Y , т.е.

$$D(Y | X = x) = D_y(x) = M\{(Y - m_y)^2 | X = x\} = \frac{\sigma_y^2\sigma_x^2 - k_{yx}^2}{\sigma_x^2}. \quad (1.15)$$

Соотношения (1.14), (1.15) иногда удобнее использовать в другой редакции, если обозначить

$$\frac{k_{yx}}{\sqrt{\sigma_y^2\sigma_x^2}} = r_{yx}, \quad \frac{k_{yx}}{\sigma_x^2} = r_{yx}\sqrt{\frac{\sigma_x^2}{\sigma_y^2}}.$$

Тогда

$$m_y(x) = M\{Y | X = x\} = m_y + r_{yx}\sqrt{\frac{\sigma_x^2}{\sigma_y^2}}(x - m_x), \quad (1.16)$$

$$D(Y | X = x) = \sigma_y^2(1 - r_{yx}^2). \quad (1.17)$$

Проанализируем полученные результаты.

1. Принципиально важным является то, что при гауссовских величинах Y и X регрессия Y на X оказывается линейно зависящей от значения x величины X .

2. Мерой связи величин Y и X оказывается коэффициент корреляции r_{yx} . Если $r_{yx} = 0$, то, как следует из (1.16), условное математическое ожидание величины Y вообще не зависит от значения величины X , т.е. из некоррелированности гауссовских случайных величин следует их независимость.

3. Условная дисперсия (1.17) величины Y не зависит от принимаемых величиной X значений. Если $|r_{yx}| = 1$, то $D(Y | X = x) = 0$, а это означает, что при фиксированном $X = x$ величина Y также оказывается фиксированной, причем, как следует из (1.16), на

уровне $m_y \pm \sqrt{\frac{\sigma_x^2}{\sigma_y^2}}(x - m_x)$. То обстоятельство, что при фиксиро-

ванном X величина Y принимает также фиксированное значение, в свою очередь, означает, что в этом случае, т.е. при $|r_{yx}| = 1$, величины Y и X связаны функциональной зависимостью

$$y = m_y + (\sigma_y^2\sigma_x^{-2})^{-0,5}(x - m_x).$$

Из этих трех выводов обратим внимание на первый: функция регрессии в указанных условиях оказывается линейной. Это обстоятельство позволяет и в иных ситуациях, пусть не гауссовских, исходить из предположения о линейной структуре неизвестной функции регрессии, т.е. класс функций Ξ ограничивать множеством линейных зависимостей вида

$$\Xi = \{(\Theta, X^*)\} = \left\{ \sum_{i=0}^s \Theta_i X^{(i)} \right\}, \quad (1.18)$$

где $X^* = [1 \ X^{(1)} \ X^{(2)} \ \dots \ X^{(s)}]^T$, $X^{(0)} = 1$, (Θ, X^*) — скалярное произведение векторов Θ и X^* .

Здесь вектор экзогенных переменных X расширен до вектора X^* , с тем чтобы в составе функции регрессии можно было управлять «постоянной составляющей», независимой от экзогенных переменных (аналогичная составляющая присутствует в (1.16)).

Класс линейных относительно экзогенных переменных и параметров Θ функций (1.18) является в настоящее время одним из наиболее используемых. Этому способствуют не только строгое обоснование этой модели при гауссовских величинах, но и, как будет показано далее, изящность, и простота получения значений параметра Θ на основе экспериментальных данных (1.1), (1.2). Вместе с тем не следует считать класс функций (1.18) единственно применяемым. Естественным обобщением является множество вида

$$\Xi = \left\{ \sum_{i=0}^m \Theta_i \psi_i(X) \right\}, \quad (1.19)$$

где $\psi_i(X)$, $i = 0, 1, 2, \dots, m$ — выбранные из определенных соображений базисные функции.

Линейный случай (1.18) является частным случаем (1.19). Могут быть применены и более сложные модели элементов множества Ξ с нелинейной зависимостью от вектора параметров Θ :

$$\Xi = \left\{ \sum_{i=0}^q \psi_i(X, \Theta) \right\}, \quad (1.20)$$

где, например, $\psi_0(X, \Theta) = \Theta_0(X^{(1)})^a(X^{(2)})^b \dots (X^{(s)})^c$, $a = \Theta_1$, $b = \Theta_2, \dots$, $c = \Theta_s$.

$$\psi_i(X, \Theta) = 0, i = 1, 2, \dots, q.$$

Подчеркнем еще раз, что выбор и обоснование класса Ξ является наиболее уязвимым и слабо защищенным теоретическими средствами местом в проблеме аппроксимации функции регрессии. В последующем мы еще раз возвратимся к этому вопросу и дадим ряд дополнительных рекомендаций, направленных на упрощение или усложнение аппроксимирующих моделей, если необходимость в этом будет выявлена. В целом же нужно стремиться к тому, чтобы класс Ξ не оказался недопустимо упрощенным (это может привести к потере наиболее характерных свойств ап-

проксимируемой функции регрессии), но и не был бы чрезмерно усложненным, ибо за этим могут последовать вычислительные осложнения обработки экспериментальных данных (1.1), (1.2) без достижения заметного положительного эффекта в качестве решения задачи восстановления зависимостей, понимаемого в том или ином смысле.

1.4. Некоторые специальные случайные величины и их свойства

Приводимые далее сведения известны из курса теории вероятностей и математической статистики. Однако чтобы опрометчиво не полагаться на память читателя, кратко напомним их.

Определение 1.1. Случайная величина $\varepsilon \sim N(0, 1)$, т.е. гауссовская с нулевым математическим ожиданием и единичной дисперсией, называется *стандартной гауссовской случайной величиной*.

Определение 1.2. Пусть $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ — последовательность независимых стандартных гауссовских случайных величин. Тогда случайная величина

$$x(n) = \sum_{i=1}^n \varepsilon_i^2,$$

говорят, имеет χ^2 -распределение с n степенями свободы. В таком случае пишут $x \sim \chi^2(n)$. Доказывается, что $m_x = n$, $\sigma_x^2 = n^2$.

Определение 1.3. Пусть $\varepsilon_0, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ — последовательность независимых стандартных гауссовских случайных величин. Тогда случайная величина

$$y = \frac{\varepsilon_0}{\sqrt{\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2}} = \frac{\varepsilon_0}{\sqrt{\frac{1}{n} x(n)}},$$

говорят, имеет распределение Стюдента, или t -распределение с n степенями свободы. В таком случае пишут $y \sim t(n)$. Доказывается, что при $n > 2$ $m_y = 0$, $\sigma_y^2 = n(n-2)^{-1}$.

Определение 1.4. Пусть $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m, \eta_1, \eta_2, \dots, \eta_n$ — последовательность независимых стандартных гауссовских случайных величин. Тогда говорят, что случайная величина

$$z = \frac{\frac{1}{m} \sum_{i=1}^m \epsilon_i^2}{\frac{1}{n} \sum_{i=1}^n \eta_i^2} = \frac{\frac{1}{m} x(m)}{\frac{1}{n} x(n)}$$

имеет распределение Фишера, или F -распределение с m, n степенями свободы. В таком случае пишут $z \sim F(m, n)$. При $n > 4$ доказывается:

$$m_z = \frac{n}{n-2}, \quad \sigma_z^2 = \frac{2n^2(m+n-2)}{m(n-4)(n-2)^2}.$$

Определение 1.5. Случайный вектор $X = [X_1 \ X_2 \ \dots \ X_n]^T$ называется гауссовским, или нормально распределенным, если совместная плотность вероятностей $\omega_x(x)$ его компонент определяется выражением

$$\omega_x(x) = \frac{1}{\sqrt{(2\pi)^n |K_x|}} \exp\{-0,5(x - m_x)^T K_x^{-1}(x - m_x)\},$$

где $m_x \in R^n$ и $K_x \in R^{n \times n}$ — параметры распределения.

В этом случае сокращенно пишут $X \sim N(m_x, K_x)$. Функцию $\omega_x(x)$ называют n -мерной гауссовской плотностью вероятностей. При $n = 1$ и $n = 2$ мы с нею уже встречались в (1.9) — (1.12). Доказывается, что $m_x = M\{X\}$ — математическое ожидание вектора X , $K_x = M\{(X - m_x)(X - m_x)^T\}$ — его ковариационная матрица.

Определение 1.6. Пусть случайная величина X имеет непрерывную функцию распределения вероятностей $F(x) = P(X < x)$, где $P(\cdot)$ — вероятность соответствующего события; $q \in (0, 1)$ — некоторое число. Тогда квантилью (или квантилем [21]) уровня q , или q -квантилью распределения $F(x)$ называется такое число u_q , что $F(u_q) = P(X < u_q) = q$.

Определение 1.7. Пусть в условиях предыдущего определения $\frac{d}{dx}F(x)$ — симметричная относительно оси ординат плотность вероятностей случайной величины X . Тогда двусторонней q -квантилью распределения $F(x)$ называют такое число t_q , что $P(|X| < t_q) = q$.

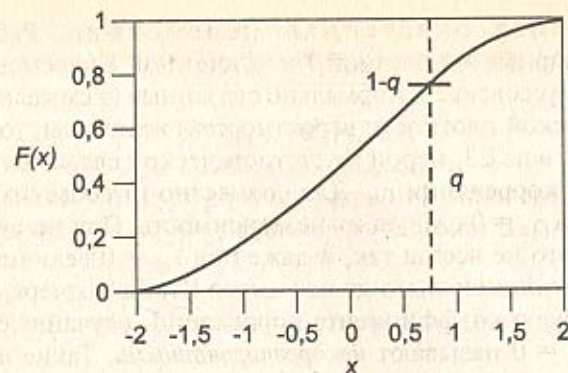


Рис. 1.2. Функция распределения вероятности $u_q = w_{100(1-q)}$

Определение 1.8. Пусть задано число $Q \in (0, 100)$. Тогда Q -процентной точкой непрерывного распределения $F(x)$ называется такое число w_Q , что выполняется условие $1 - F(w_Q) = P(X \geq w_Q) = 10^{-2}Q$. Очевидно, $u_q = w_{100(1-q)}$ (рис. 1.2).

1.5. Предварительный (дорегрессионный) анализ зависимости эндогенной и экзогенных переменных

1.5.1. Общие принципы

Обычно при поиске зависимости между эндогенной и экзогенными переменными предполагается, что еще на этапе предварительного анализа составлен «список» экзогенных переменных, влияющих, по нашему мнению, на эндогенную переменную. Во многих случаях уже из содержательного существа проблемы наличие влияния можно считать непреложной истиной и не подвергать его сомнению. Так, например, покупательные возможности семьи наверняка зависят от ее среднедушевого дохода. Однако в иных ситуациях такая прозрачность в априорной оценке влияния экзогенной переменной на эндогенную отсутствует и необходимо соответствующее обоснование с привлечением определенных формализованных подходов. Трудно заранее, например, утверждать, что производительность технологической установки зависит именно от этой, а не иной характеристики используемого сырья.

Скалярная экзогенная переменная. Рассмотрим случай скалярных эндогенной Y и экзогенной X переменных. Если Y и X — гауссовские и нормально связанные (в смысле совместной гауссовской плотности вероятностей) величины, то, как было показано в п. 1.3, мерой их статистической связи является коэффициент корреляции r_{yx} . Для совместно гауссовских величин из равенства $r_{yx} = 0$ следует их независимость. При негауссовских величинах это не всегда так, и даже при $r_{yx} = 0$ величины могут оказаться функционально зависимыми. Чтобы подчеркнуть факт равенства нулю коэффициента корреляции, случайные величины при $r_{yx} = 0$ называют *некоррелированными*. Такие величины могут оказаться зависимыми, но эту зависимость средствами грубого для исследования подобных ситуаций инструментария в виде коэффициента корреляции зарегистрировать не удастся. Тем не менее коэффициент корреляции используется как своеобразный индикатор связи и при негауссовских величинах. По определению коэффициент корреляции

$$r_{yx} = \frac{1}{\sqrt{\sigma_y^2 \sigma_x^2}} M\{(Y - m_y)(X - m_x)\} =$$

$$= \frac{1}{\sqrt{\sigma_y^2 \sigma_x^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - m_y)(x - m_x) \omega(y, x) dy dx.$$

Однако практически таким аналитическим способом вычислить коэффициент корреляции не удастся, так как обычно неизвестны не только совместная плотность вероятностей $\omega(y, x)$, но и даже числовые характеристики величин Y и X . На помощь приходит предположение о том, что можно провести эксперимент, в котором экзогенной переменной X придаются значения $x_1, x_2, \dots, x_n$ и регистрируются (измеряются) соответствующие значения $y_1, y_2, \dots, y_n$ эндогенной переменной Y . Набор значений экзогенной переменной может быть следствием какого-либо естественного процесса (пассивный эксперимент) или сформирован искусственно из определенных соображений (активный эксперимент). Независимо от природы экспериментальных данных они позволяют найти приближенное значение $\hat{r}_{yx}$ коэффициента корреляции r_{yx} , которое принято называть *эмпирической (выборочной) оценкой*. Хотя принципиально эту оценку можно найти

различными способами, каждый из которых приводит к своему результату, наиболее распространенной оказывается оценка вида

$$\hat{r}_{yx} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (x_i - \bar{x})^2}}, \quad (1.21)$$

$$\text{где } \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

На множестве возможных значений случайной величины Y величина (1.21) является также случайной. Чтобы можно было по ней судить о корреляционной связи величин Y и X , нужны статистические характеристики самой величины (1.21) или какой-либо иной величины, но функционально связанной с $\hat{r}_{yx}$. Пусть такой величиной (часто говорят — *статистикой*) является некая величина $\gamma = \gamma(\hat{r}_{yx})$. Тогда последующий анализ проводится по достаточно типовым для подобных исследований схемам.

В рассмотрение вводятся две гипотезы:

H_0 : корреляционная связь между Y и X отсутствует ($r_{yx} = 0$);

H_1 : величины Y и X коррелированы ($r_{yx} \neq 0$).

Любое последующее решение проблемы сопровождается двумя возможными ошибками:

ошибка первого рода — принять гипотезу H_1 , когда в действительности справедлива гипотеза H_0 ;

ошибка второго рода — принять гипотезу H_0 , когда в действительности справедлива альтернатива H_1 .

Обозначим через $\alpha = P(H_1|H_0)$ вероятность ошибки первого рода, через $\beta = P(H_0|H_1)$ — вероятность ошибки второго рода. Величина $1 - \alpha$ является условной вероятностью правильного решения при выполнении гипотезы H_0 , аналогично $1 - \beta$ есть вероятность правильного решения при условии, что справедлива гипотеза H_1 . Величину α часто называют *уровнем значимости* критерия, величину $1 - \beta$ — *мощностью* критерия.

Решение задачи должно сводиться к обоснованному выбору одной из двух гипотез: H_0 или H_1 на основе значения величины γ , полученного по эмпирическим данным $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$. Величина $\gamma \in R$, где R , как обычно, множество всех вещественных

чисел. Тогда геометрически решение можно интерпретировать так: множество возможных значений величины γ , т.е. R , следует разбить на два подмножества Γ_0 и Γ_1 ($\Gamma_0 \cup \Gamma_1 = R$) так, чтобы наилучшим в некотором смысле образом из условия $\gamma \in \Gamma_0$ следовало принятие гипотезы H_0 , а при условии $\gamma \in \Gamma_1$ предпочтение отдавалось гипотезе H_1 . Чтобы формализовать этот замысел, прежде всего нужно выявить смысл словосочетания «наилучшим образом», т.е., по существу, сформулировать критерий оптимальности, закладываемый в процедуру решения задачи. Возможны следующие варианты (п. 1.5.2 – 1.5.5).

1.5.2. Критерий идеального наблюдателя

Уже отмечалось, что любое решение задачи сопровождается ошибками первого и второго рода с соответствующими вероятностями α и β . Если известны априорные вероятности p_0 и p_1 справедливости гипотез H_0 и H_1 соответственно (иначе можно принять $p_0 = p_1 = 0,5$), то величина $p_0\alpha + p_1\beta$ будет безусловной вероятностью ошибочного решения. Первое слагаемое здесь является безусловной вероятностью ошибки первого рода, т.е. вероятностью выполнения двух событий: справедлива гипотеза H_0 с априорной вероятностью p_0 , но принимается гипотеза H_1 с условной вероятностью α . Аналогична структура второго слагаемого. Очевидно, решение задачи целесообразно искать так, чтобы безусловная вероятность ошибочного решения оказалась наименьшей. Это значит, что подмножества Γ_0 и Γ_1 (или только Γ_1 , так как $\Gamma_0 = R \setminus \Gamma_1$) следует находить в процессе решения оптимизационной задачи

$$p_0\alpha + p_1\beta \rightarrow \min. \quad (1.22)$$

Условие (1.22) называют *критерием идеального наблюдателя* (иногда – *критерием Котельникова*). Рассмотрим его более внимательно.

Пусть известны условные плотности вероятностей $\omega(\gamma|H_0)$ и $\omega(\gamma|H_1)$ величины γ соответственно при выполнении гипотез H_0 и H_1 . Тогда

$$\alpha = \int_{\Gamma_1} \omega(\gamma|H_0) d\gamma, \quad \beta = \int_{\Gamma_0} \omega(\gamma|H_1) d\gamma,$$

и условие (1.22) переписывается так:

$$p_0 \int_{\Gamma_1} \omega(\gamma|H_0) d\gamma + p_1 \int_{\Gamma_0} \omega(\gamma|H_1) d\gamma \rightarrow \min.$$

Поскольку

$$\int_R \omega(\gamma|H_1) d\gamma = \int_{\Gamma_1} \omega(\gamma|H_1) d\gamma + \int_{\Gamma_0} \omega(\gamma|H_1) d\gamma = 1,$$

оптимизационная задача может быть записана в новой редакции:

$$p_1 - \int_{\Gamma_1} (p_1\omega(\gamma|H_1) - p_0\omega(\gamma|H_0)) d\gamma \rightarrow \min.$$

Чтобы эта целевая функция была минимальна, значение интеграла должно быть максимальным. Это достигается, если подмножество Γ_1 выбрано так, что во всех принадлежащих ему точках подинтегральная функция неотрицательна, т.е.

$$p_1\omega(\gamma|H_1) - p_0\omega(\gamma|H_0) \geq 0. \quad (1.23)$$

Таким образом, если при найденном по эмпирическим данным значении величины γ выполняется неравенство (1.23), то принимается гипотеза H_1 . При противоположном неравенстве предпочтение отдается альтернативе H_0 . Лаконично это записывается так:

$$\frac{\omega(\gamma|H_1)}{\omega(\gamma|H_0)} \begin{cases} \geq \frac{p_0}{p_1} \Rightarrow H_1 \\ < \frac{p_0}{p_1} \Rightarrow H_0 \end{cases} \quad (1.24)$$

Выражение (1.24) совместно с правилом вычисления γ представляет собой алгоритм решения задачи по критерию идеального наблюдателя.

1.5.3. Критерий Неймана – Пирсона

Второй возможный подход к решению задачи основывается на так называемом критерии Неймана – Пирсона. Его целесообразно применять в тех случаях, когда последствия от ошибок первого и второго рода не являются равноценными. В этих случаях ре-

шение задачи ищут таким образом, чтобы вероятность одной из ошибок оказалась ограниченной некоторой малой величиной, а вероятность второй при этом приняла наименьшее значение. Например,

$$p_1\beta \rightarrow \min \text{ при } p_0\alpha = \delta = \text{const}, \quad (1.25)$$

где δ — выбранная малая величина.

Эти условия и формируют критерий Неймана — Пирсона. «Рычагом» их реализации по-прежнему является выбор оптимальных подмножеств Γ_0 и Γ_1 . Задача (1.25) относится к классу задач на условный экстремум и решается методом неопределенных множителей Лагранжа. С этой целью составляется функция Лагранжа

$$L = p_1\beta + \lambda(p_0\alpha - \delta),$$

где λ — неопределенный множитель Лагранжа.

Используя предыдущую схему преобразований, записываем функцию Лагранжа в иной форме:

$$\begin{aligned} L &= p_1 \int_{\Gamma_0} \omega(\gamma | H_1) d\gamma + \lambda(p_0 \int_{\Gamma_1} \omega(\gamma | H_1) d\gamma - \delta) = \\ &= p_1 - \int_{\Gamma_1} [p_1 \omega(\gamma | H_1) - \lambda p_0 \omega(\gamma | H_0)] d\gamma - \lambda \delta. \end{aligned}$$

Опять же минимальное значение функции L достигается при максимальном значении интеграла, что, в свою очередь, обеспечивается выбором подмножества Γ_1 таким образом, чтобы во всех принадлежащих ему точках подынтегральная функция была положительной. Отсюда по аналогии с (1.24) вытекает правило

$$\frac{\omega(\gamma | H_1)}{\omega(\gamma | H_0)} \begin{cases} \geq \frac{p_0 \lambda}{p_1} \Rightarrow H_1 \\ < \frac{p_0 \lambda}{p_1} \Rightarrow H_0 \end{cases},$$

отличающееся от (1.24) только выбором порога $\lambda p_0 p_1^{-1}$. Чтобы окончательно найти этот порог, следует вычислить множитель Лагранжа λ . Принципиально это делается на основе ограничения $p_0\alpha = \delta$, но данная задача нетривиальная.

1.5.4. Критерий проверки гипотезы H_0 при скалярной экзогенной переменной

Рассмотренные два подхода предполагают, что известны вероятностные свойства величины γ при обеих гипотезах H_0 и H_1 . Во многих практических задачах такую статистику найти не удастся, но можно установить величину γ с известными вероятностными свойствами при справедливости одной из гипотез. Тогда задача формулируется и решается так.

Пусть проверяется справедливость гипотезы H_0 и известна условная плотность вероятностей $\omega(\gamma | H_0)$. Задав ширину вероятностной α ошибки первого рода (наиболее часто принимают $\alpha = 0,05$), находят такое подмножество $\Gamma_0 \subset R$, что

$$P(\gamma \in \Gamma_0 | H_0) = 1 - \alpha. \quad (1.26)$$

Если теперь по экспериментальным данным найдено конкретное численное значение величины γ и оказалось, что $\gamma \in \Gamma_0$, то с доверительной вероятностью $1 - \alpha$ признается справедливость гипотезы H_0 . Если же окажется $\gamma \notin \Gamma_0$, то гипотеза H_0 отвергается с вероятностью ошибиться α . В задачах эконометрики, в частности применительно к обсуждаемой здесь конкретной проблеме установления связи эндогенной и экзогенной переменных, этот подход используется наиболее широко.

Итак, возвратимся непосредственно к нашей задаче (п. 1.5.1). Уже отмечалось, что коэффициент $\hat{r}_{yx}$ на множестве значений эндогенной переменной Y является случайной величиной, и доказывается (например, [1]), что при совместно гауссовских величинах Y и X , $n > 200$ и $|r_{yx}| < 1$ приближенно $\hat{r}_{yx} \sim N(r_{yx}, (1 - r_{yx}^2)^2/n)$. Однако практически этим свойством воспользоваться не удастся из-за невыполнения условий, при которых оно справедливо.

Известен [1] более полезный для наших целей результат: величина

$$T = \hat{r}_{yx} \frac{\sqrt{n-2}}{\sqrt{1-\hat{r}_{yx}^2}} \quad (1.27)$$

при малых $|r_{yx}|$ и выполнении гипотезы H_0 приблизительно распределена по закону Стьюдента с $n-2$ степенями свободы. Это обстоятельство позволяет величину (1.27) использовать для разра-

ботки критерия проверки гипотезы H_0 в соответствии с принципом (1.26). Учитывая четность и, как следствие, симметричность t -распределения, множество Γ_0 будем искать в виде отрезка $\Gamma_0 = [-g, g]$, причем величину g найдем из условия

$$\int_{-g}^g \omega(\gamma | H_0) d\gamma = 1 - \alpha,$$

где $\omega(\gamma | H_0)$ — плотность вероятности величины γ при гипотезе H_0 , т.е. t -распределение с $n-2$ степенями свободы.

С учетом нормировки плотности вероятности это равенство можно переписать так:

$$\alpha = \int_g^\infty \omega(\gamma | H_0) d\gamma + \int_{-\infty}^{-g} \omega(\gamma | H_0) d\gamma = 2 \int_g^\infty \omega(\gamma | H_0) d\gamma = 2 \int_g^\infty \omega(\gamma | H_0) d\gamma.$$

Отсюда следует, что $-g = u_{\alpha/2}$, $g = w_{100\alpha/2}$ ($u_{\alpha/2} = -w_{100\alpha/2}$), где $u_{\alpha/2} = \alpha/2$ — квантиль распределения Стьюдента с $n-2$ степенями свободы, $w_{100\alpha/2}$ есть $100\alpha/2$ -процентная точка того же распределения. Это позволяет сформулировать следующий критерий проверки гипотезы H_0 .

Пусть проведен активный или пассивный эксперимент и на основе полученных данных по формуле (1.21) найдено конкретное значение эмпирического коэффициента $\hat{r}_{yx}$. Тогда если окажется, что

$$|\hat{r}_{yx}| \frac{\sqrt{n-2}}{\sqrt{1-\hat{r}_{yx}^2}} < u_{\alpha/2} \quad \text{или} \quad \hat{r}_{yx} \frac{\sqrt{n-2}}{\sqrt{1-\hat{r}_{yx}^2}} > w_{100\alpha/2},$$

то гипотеза H_0 об отсутствии корреляционной связи между Y и X отвергается с вероятностью ошибиться α . Эти оба неравенства можно выразить одним:

$$|\hat{r}_{yx}| \frac{\sqrt{n-2}}{\sqrt{1-\hat{r}_{yx}^2}} > w_{100\alpha/2}. \quad (1.28)$$

Таким образом, если по экспериментальным данным найдена величина $\hat{r}_{yx}$, а по соответствующим таблицам (или машинным

образом — см. далее) — величина $w_{100\alpha/2}$ и окажется справедливым неравенство (1.28), то гипотеза H_0 об отсутствии связи между переменными Y и X отвергается с вероятностью α ошибиться. При противоположном неравенстве гипотеза H_0 считается не противоречащей экспериментальным данным с вероятностью $1 - \alpha$ правильности этого решения. Заметим, что таблицы, содержащие характерные точки различных распределений и приведенные во многих литературных источниках (например, [1], [3], [15], [30] и др.), мы не тиражируем, так как эти данные легко получить средствами большинства современных пакетов прикладных программ. Так, при $\alpha = 0,05$ и $n = 15$ величина $w_{100\alpha/2}$ распределения Стьюдента с $n - 2$ степенями свободы легко находится с помощью, например, такой микропрограммы в Mathcad'e

$$\alpha := 0,05 \\ n := 15$$

$$\text{root} \left(pt(x, n-2) - 1 + \frac{\alpha}{2}, x, 0, 10 \right) = 2,161.$$

Если найдена оценка (1.21), можно найти доверительный интервал для истинного значения коэффициента корреляции r_{yx} . С этой целью используется предложенная Р. Фишером статистика

$$z = \frac{1}{2} \ln \frac{1 + \hat{r}_{yx}}{1 - \hat{r}_{yx}}, \quad (1.29)$$

которая уже при небольших n оказывается приближенно гауссовской с параметрами

$$m_z = \frac{1}{2} \ln \frac{1 + r_{yx}}{1 - r_{yx}} + \frac{r_{yx}}{2(n-1)}, \quad \sigma_z^2 = \frac{1}{n-3}.$$

Если задаться доверительной вероятностью $1 - \alpha$, то можно найти соответствующую интервальную оценку величины z , удовлетворяющую традиционному условию

$$P(z_1 \leq z < z_2) = 1 - \alpha, \quad (1.30)$$

где z_1, z_2 — границы интервала.

Так как гауссовская плотность симметрична относительно математического ожидания, границы z_1, z_2 будем искать в виде

$z_1 = m_z - \xi$, $z_2 = m_z + \xi$, где ξ — подлежащая определению величина, обеспечивающая условие (1.30).

Запишем равенство (1.30) в развернутом виде:

$$\frac{1}{\sqrt{2\pi\sigma_z^2}} \int_{m_z-\xi}^{m_z+\xi} \exp\{-(z-m_z)^2/2\sigma_z^2\} dz = 1-\alpha,$$

или же, заменив переменную $(z - m_z)/\sigma_z = s$,

$$\frac{1}{\sqrt{2\pi}} \int_{-b}^b \exp\{-s^2/2\} ds = 1-\alpha, \quad b = \xi/\sigma_z.$$

Это соотношение, если учесть нормировку плотности вероятностей, легко преобразуется к виду

$$\begin{aligned} \alpha &= \frac{1}{\sqrt{2\pi}} \int_b^\infty \exp\{-s^2/2\} ds + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-b} \exp\{-s^2/2\} ds = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-b} \exp\{-s^2/2\} ds, \end{aligned}$$

откуда следует $\xi = -\sigma_z u_{\alpha/2}$, где $u_{\alpha/2}$ есть $\alpha/2$ — квантиль стандартного гауссовского распределения $N(0, 1)$. Следовательно, с вероятностью $1 - \alpha$ имеем

$$\begin{aligned} m_z + \sigma_z u_{\alpha/2} \leq z < m_z - \sigma_z u_{\alpha/2} \Rightarrow \\ z + \sigma_z u_{\alpha/2} < m_z \leq z - \sigma_z u_{\alpha/2}. \end{aligned}$$

С учетом определения m_z получим

$$z + \sigma_z u_{\alpha/2} - \frac{r_{yx}}{2(n-1)} < 0,5 \ln \frac{1+r_{yx}}{1-r_{yx}} \leq z - \sigma_z u_{\alpha/2} - \frac{r_{yx}}{2(n-1)}.$$

Найдем приближенное решение этого неравенства относительно r_{yx} , заменив на границах неравенства величину r_{yx} ее оценкой $\hat{r}_{yx}$.

$$c < 0,5 \ln \frac{1+\hat{r}_{yx}}{1-\hat{r}_{yx}} \leq d, \quad (1.31)$$

$$c, d = 0,5 \ln \frac{1+\hat{r}_{yx}}{1-\hat{r}_{yx}} \pm \frac{u_{\alpha/2}}{\sqrt{n-3}} - \frac{\hat{r}_{yx}}{2(n-1)}. \quad (1.32)$$

Из левого неравенства (1.31) имеем

$$r_{yx} > (e^{2c} - 1) / (e^{2c} + 1) = (e^c - e^{-c}) / (e^c + e^{-c}) = \text{th } c,$$

где $\text{th } c$ — гиперболический тангенс c .

Аналогично из правого неравенства $r_{yx} \leq \text{th } d$. Следовательно, с вероятностью $1 - \alpha$

$$\text{th } c < r_{yx} \leq \text{th } d, \quad (1.33)$$

что и будет доверительным интервалом для истинного коэффициента корреляции r_{yx} . Таким образом, для построения интервала (1.33) следует задаться доверительной вероятностью $1 - \alpha$, найти по эмпирическим данным коэффициент $\hat{r}_{yx}$, воспользовавшись определением (1.21), по соответствующим таблицам или машинным программам выявить значение $u_{\alpha/2}$, т. е. $\alpha/2$ -квантили стандартного гауссовского распределения $N(0, 1)$, по формулам (1.32) рассчитать величины c и d , наконец, по таблицам для гиперболического тангенса или машинным образом найти границы интервала $\text{th } c$, $\text{th } d$. Заметим, что величина (1.29), содержащаяся в (1.32), также может быть найдена по таблицам обратного гиперболического тангенса, так как

$$z = 0,5 \ln \frac{1+\hat{r}_{yx}}{1-\hat{r}_{yx}} = \text{arc th } \hat{r}_{yx}.$$

1.5.5. Критерий проверки гипотезы H_0 при векторной экзогенной переменной

В заключение настоящего раздела остановимся еще на одном достаточно важном обстоятельстве. Ранее предполагалось, что эндогенная переменная определяется единственной экзогенной переменной и что выявляется степень связи между ними. Во многих задачах экзогенных переменных несколько. Если по экспериментальным данным анализируется связь с одной из экзогенных переменных (говорят — *парная* связь), то оставшиеся экзогенные

переменные выступают в роли мешающих параметров и существенно влияют на результаты анализа. Поэтому эксперимент должен быть организован так, чтобы всем значениям исследуемой экзогенной переменной соответствовали одни и те же неизменные (постоянные) значения оставшихся экзогенных (мешающих) переменных. При этом не исключено, что результаты анализа будут зависеть от того, какие именно неизменные значения принимают мешающие экзогенные переменные. Все это существенно усложняет анализ парных связей.

Есть условие, при выполнении которого отмеченные проблемы практически себя не проявляют. Оно заключается в том, что совместно эндогенная переменная Y и экзогенные переменные $X^{(1)}, X^{(2)}, \dots, X^{(s)}$ подчинены $(s+1)$ -мерному гауссовскому распределению. В этом случае частный коэффициент корреляции ρ_{0j} между эндогенной переменной Y и j -й экзогенной переменной $X^{(j)}$ ($j = 1, 2, \dots, s$), вычисленный в предположении, что остальные экзогенные переменные приняли некоторые фиксированные значения, не зависит от уровней, принимаемых остальными (мешающими) экзогенными переменными, и может быть рассчитан по формуле [1]

$$\rho_{0j} = -\frac{R_{0j}}{\sqrt{R_{00}R_{jj}}}, \quad (1.34)$$

где R_{ij} — алгебраическое дополнение ij -го ($i, j = 0, 1, \dots, s$) элемента корреляционной матрицы R случайных величин $Y, X^{(1)}, X^{(2)}, \dots, X^{(s)}$, т.е.

$$R = \begin{bmatrix} 1 & r_{01} & \dots & r_{0s} \\ r_{10} & 1 & \dots & r_{1s} \\ \dots & \dots & \dots & \dots \\ r_{s1} & r_{s2} & \dots & 1 \end{bmatrix}. \quad (1.35)$$

Здесь r_{ij} — коэффициент корреляции величин $X^{(i)}$ и $X^{(j)}$, причем принято $X^{(0)} = Y$. В частности, при $s = 2$ получим:

$$\rho_{01} = \frac{r_{10} - r_{12}r_{20}}{\sqrt{(1-r_{12}^2)(1-r_{02}^2)}}, \quad \rho_{02} = \frac{r_{20} - r_{21}r_{10}}{\sqrt{(1-r_{12}^2)(1-r_{01}^2)}}.$$

Дальнейшая технология практического применения этих соотношений сводится к следующему. Пусть получены экспери-

ментальные данные в объеме $y_i, x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(s)}, i = 1, 2, \dots, n$. По формулам, подобным (1.21), находятся эмпирические коэффициенты корреляции $\hat{r}_{ij}$ величин $X^{(i)}, X^{(j)}, i = 0, 1, 2, \dots, s-1, j = i+1, i+2, \dots, s$. Из этих величин с учетом их симметрии составляется матрица $\hat{R}$ аналогичным (1.35) образом и с помощью (1.34) рассчитываются эмпирические частные коэффициенты корреляции, $\hat{\rho}_{0j}, j = 1, 2, \dots, s$. Для истинного значения каждого из них строится доверительный интервал, подобный (1.33), причем границы интервала находятся подобным (1.32) образом, но с одной существенной поправкой: величина n заменяется на $n-s+1$, где число $s-1$ представляет собой количество мешающих параметров.

В связи со случаем многих экзогенных переменных полезно остановиться на особенностях применения линейных регрессионных множеств вида (1.18), при которых модель (1.8) будет выглядеть так:

$$Y = \Theta_0 + \Theta_1 X^{(1)} + \Theta_2 X^{(2)} + \dots + \Theta_s X^{(s)} + \epsilon. \quad (1.36)$$

Для выявления факта зависимости эндогенной переменной Y от совокупности экзогенных переменных $X^{(1)}, X^{(2)}, \dots, X^{(s)}$ используется множественный коэффициент корреляции $R_{y,x}^2$, определяемый равенством [1]

$$R_{y,x}^2 = 1 - |R|/R_{00}, \quad (1.37)$$

где $|R|$ — определитель матрицы (1.35), R_{00} — как и в (1.34), алгебраическое дополнение элемента $r_{00} = 1$ этой матрицы.

Пусть матрица (1.35) построена по эмпирическим данным. Тогда доказывается, что выборочный коэффициент $\hat{R}_{y,x}$, вычисленный в соответствии с (1.37), но по эмпирической матрице $\hat{R}$, оказывается таков, что величина

$$\gamma = \frac{n-s-1}{s} \frac{\hat{R}_{y,x}^2}{1-\hat{R}_{y,x}^2} \quad (1.38)$$

в случае справедливости гипотезы $H_0: R_{y,x}^2 = 0$ подчинена распределению Фишера с $(s, n-s-1)$ степенями свободы, т.е. $F(s, n-s-1)$ -распределению. Последующий анализ проводится по схеме, по-

добной той, которая ранее привела нас к правилу (1.28). А именно: задаются вероятностью α ошибки первого рода или, что эквивалентно, доверительной вероятностью $1 - \alpha$ справедливости гипотезы H_0 ; по соответствующим справочным или программным материалам находят величину $w_{100\alpha}$, т. е. 100%-ную точку распределения Фишера с числом степеней свободы числителя s и знаменателя $n - s - 1$. Если окажется $\gamma > w_{100\alpha}$, то гипотеза H_0 отвергается с вероятностью ошибиться α (уровень значимости критерия). При противоположном неравенстве предпочтение отдается гипотезе H_0 с вероятностью $1 - \alpha$ правильности этого решения. Полезное свойство модели (1.36) проявляется также в том, что изложенный алгоритм анализа ситуации сохраняет свои свойства и при отклонении совместной плотности вероятностей величин $Y, X^{(1)}, X^{(2)}, \dots, X^{(s)}$ от гауссовской [1].

Глава 2

МЕТОДЫ ОЦЕНИВАНИЯ ПАРАМЕТРОВ РЕГРЕССИОННЫХ МОДЕЛЕЙ

2.1. Проблема оценивания и общие характеристики точечных оценок. Неравенство Рао – Крамера

Пусть, как и ранее, Y — единственная эндогенная переменная, зависящая от s экзогенных переменных $X^{(1)}, X^{(2)}, \dots, X^{(s)}$. Предполагается, что сам факт зависимости установлен на основе предварительного анализа экспериментальных данных в соответствии с вышеизложенными методами или является логическим следствием содержательного существа изучаемого явления. Пусть далее, обоснована модель представления эндогенной переменной в форме (1.8). Аппроксимирующая неизвестную регрессию функция $f(X, \Theta)$ определена с точностью до вектора неизвестных параметров Θ и принадлежит множествам вида (1.18), (1.19). Для определенности будем руководствоваться более общим случаем (1.19); таким образом, связь эндогенной и экзогенных переменных определяется соотношением

$$Y = \sum_{i=0}^m \Theta_i \psi_i(X^{(1)}, X^{(2)}, \dots, X^{(s)}) + \varepsilon(X^{(1)}, X^{(2)}, \dots, X^{(s)}), \quad (2.1)$$

или, более лаконично,

$$Y = \Psi^T(X) \Theta + \varepsilon(X), \quad (2.2)$$

где использованы естественные обозначения $\Theta = [\Theta_0, \Theta_1, \dots, \Theta_m]^T$, $X = [X^{(1)}, X^{(2)}, \dots, X^{(s)}]^T$, $\Psi^T = [\psi_0, \psi_1, \dots, \psi_m]$.

Следствием проведенного эксперимента является совокупность величин (1.1), (1.2) (для скалярной эндогенной переменной в (1.2) следует положить $k = 1$), которая в терминах модели (2.2) опишется соотношениями

$$y_j = \Psi^T(x_j) \Theta + \varepsilon_j(x_j), \quad j = 1, 2, \dots, n, \quad (2.3)$$

где символ $\varepsilon_j(x_j)$ представляет собой ошибку в j -й точке эксперимента (при $X = x_j$).

В матрично-векторных обозначениях n выражений (2.3) сводятся к одному:

$$y = \Psi\Theta + \epsilon, \quad (2.4)$$

где $y = [y_1, y_2, \dots, y_n]^T$ — вектор значений эндогенной переменной и

$$\Psi = \begin{bmatrix} \psi^T(x_1) \\ \psi^T(x_2) \\ \dots \\ \psi^T(x_n) \end{bmatrix} \in R^{n \times (m+1)}, \quad x_i = \begin{bmatrix} x_i^{(1)} \\ x_i^{(2)} \\ \dots \\ x_i^{(s)} \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_n \end{bmatrix}.$$

Последующая задача построения регрессионной модели (2.2) сводится к определению вектора параметров Θ по результатам эксперимента, представленного *апостериорной выборкой* y . Этот вектор связан с параметрами Θ соотношением (2.4), в котором матрица Ψ определена через экспериментально полученные значения экзогенных переменных и известна, а вектор ϵ представляет собой совокупность неизвестных величин, обобщенно трактуемых как *ошибки* эксперимента.

Если бы ошибок эксперимента не было, то для определения величин $\Theta_0, \Theta_1, \dots, \Theta_m$ достаточно было бы провести $m+1$ измерений эндогенной переменной при надлежащем выборе такого же количества значений вектора экзогенных переменных и из $m+1$ уравнений $y_j = \psi^T(x_j)\Theta, j = 1, 2, \dots, m+1$, найти интересующие нас параметры (проблему разрешимости этих уравнений мы здесь не обсуждаем). Однако каждое реальное наблюдение из (2.3), помимо неизвестных величин $\Theta_0, \Theta_1, \dots, \Theta_m$, содержит неизвестную ошибку эксперимента, поэтому сколько бы измерений ни проводилось, точно определить параметры Θ невозможно. Но при достаточно большом числе измерений ($n > m+1$) влияние ошибок можно путем рациональных операций над экспериментальными данными y уменьшить и найти по наблюдениям (2.4) некоторые величины $\hat{\Theta}_0, \hat{\Theta}_1, \dots, \hat{\Theta}_m$, в определенном смысле близкие к истинным, но неизвестным значениям параметров $\Theta_0, \Theta_1, \dots, \Theta_m$. Эти величины называют *точечными оценками* параметров Θ .

В связи с поиском оценок возникают два вопроса: как formalизовать понятие близости вектора оценок $\hat{\Theta} = [\hat{\Theta}_0, \hat{\Theta}_1, \dots, \hat{\Theta}_m]^T$ и оцениваемых параметров Θ (в каком смысле понимать близость)

и как найти оценки, наилучшие с позиций установленного смысла близости. Ответ на первый вопрос приводит к понятию *критерия качества оценивания*. Ответ на второй вопрос позволяет определить вычислительные операции, которые надо провести над экспериментальными данными y , чтобы получить наилучшие в смысле этого критерия оценки как функции экспериментальных данных $\hat{\Theta}_i = \hat{\Theta}_i(y_1, y_2, \dots, y_n) = \hat{\Theta}_i(y), i = 0, 1, \dots, m$, т.е. получить *алгоритм оптимального оценивания*.

В зависимости от объема и характера наших знаний о свойствах оцениваемых параметров и ошибок эксперимента, предшествующих самому эксперименту, применяют тот или иной метод оценивания. Информацию, содержащуюся в вероятностных характеристиках параметров и ошибок, которая может быть как известной, так и неизвестной до проведения эксперимента, называют *априорной*. Так, может быть известна априорная совместная плотность вероятностей $\omega_\epsilon(\epsilon)$ вектора ошибок ϵ . Вектор параметров Θ может классифицироваться как *неизвестный* или как *случайный*. В первом случае он является неслучайным, но априори мы о нем ничего не знаем и полагаем, что его компоненты могут принимать любые значения в диапазоне от $-\infty$ до $+\infty$. Во втором случае считается, что вектор Θ принимает значения в соответствии с априорной плотностью вероятностей $\omega_\Theta(\Theta)$. В общем случае эта плотность исследователю может быть и неизвестна, но объективно существует. Неизвестный вектор Θ часто удобно интерпретировать как случайный с бесконечно большими дисперсиями его компонент и нулевым средним значением. Плотности $\omega_\epsilon(\epsilon)$ и $\omega_\Theta(\Theta)$ устанавливают на основании каких-либо аналитических расчетов или специально организованных экспериментов, предшествующих проведению основного эксперимента с исходными данными (2.4).

Независимо от способа вычисления оценки $\hat{\Theta}(y)$ по результатам y проведенного эксперимента с ней связывают ряд определений.

1. Оценку $\hat{\Theta}$ называют *условной*, если априорная информация, используемая при ее вычислении, ограничена условной плотностью вероятностей $L(y|\Theta)$ экспериментальных данных, найденной в предположении, что вектор параметров Θ принял некоторое фиксированное значение. Условные оценки обычно применяют при решении задач с неслучайными параметрами.

2. Оценку $\hat{\Theta}$ называют *безусловной*, если априорная информация, используемая при ее вычислении, сводится к безусловной

совместной плотности вероятностей $\omega(y, \Theta)$ экспериментальных данных и оцениваемых параметров. Безусловные оценки ищутся в задачах со случайными параметрами, априорные свойства которых в объеме их совместной плотности вероятностей $\omega_{\Theta}(\Theta)$ должны быть известны.

Заметим, что условная оценка может относиться и к случайному параметру, если априорная информация о нем неизвестна или не используется из-за существенного усложнения алгоритма оценивания. Для таких ситуаций безусловная оценка может быть найдена путем усреднения условной оценки по всем возможным значениям вектора параметров Θ .

3. Условную оценку $\hat{\Theta}$ называют *состоятельной*, если при неограниченном объеме выборки ($n \rightarrow \infty$) каждый ее компонент сходится по вероятности к соответствующему компоненту вектора Θ , т.е. если при $\forall \delta > 0$

$$\lim P\{|\hat{\Theta}_i - \Theta_i| \geq \delta\} = 0 \text{ при } n \rightarrow \infty, i = 0, 1, \dots, m.$$

4. Безусловную оценку $\hat{\Theta}$ называют *состоятельной*, если при неограниченном увеличении объема выборки каждый ее компонент по вероятности сходится к среднему значению соответствующего компонента вектора Θ , т.е. если при $\forall \delta > 0$

$$\lim P\{|\hat{\Theta}_i - M\{\Theta_i\}| \geq \delta\} = 0 \text{ при } n \rightarrow \infty, i = 0, 1, \dots, m.$$

5. Условную оценку $\hat{\Theta}$ называют *несмещенной*, если среднее значение этой оценки, полученное ее усреднением по возможным значениям вектора y при фиксированном векторе Θ , равно самому оцениваемому параметру:

$$M_{y|\Theta}\{\hat{\Theta}(y)\} = \int_{-\infty}^{\infty} \hat{\Theta}(y) L(y|\Theta) dy = \Theta. \quad (2.5)$$

Здесь $M_{y|\Theta}\{\dots\}$ — символ условного усреднения.

6. Безусловную оценку $\hat{\Theta}$ называют *несмещенной*, если среднее значение этой оценки, полученное ее усреднением по возможным значениям вектора y при всех возможных значениях вектора Θ , равно среднему значению оцениваемого параметра:

$$M_y\{\hat{\Theta}(y)\} = \int_{-\infty}^{\infty} \hat{\Theta}(y) \omega_y(y) dy = M\{\Theta\}. \quad (2.6)$$

Здесь $M_y\{\dots\}$ — символ безусловного усреднения, $\omega_y(y)$ — безусловная плотность вероятностей вектора данных y и интегралы понимаются как многомерные [как и в (2.5)]:

$$\int_{-\infty}^{\infty} dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} dy_1 dy_2 \dots dy_n.$$

7. Условную оценку $\hat{\Theta}$ называют *эффективной*, если среднее значение квадрата отклонения каждого ее компонента от соответствующего компонента вектора Θ не больше среднего квадрата отклонения для любой другой оценки:

$$M_{y|\Theta}\{(\hat{\Theta}_i - \Theta_i)^2\} = \min_l \hat{M}_{y|\Theta}\{(\hat{\Theta}_i - \Theta_i)^2\}, i = 0, 1, \dots, m.$$

Здесь усреднение проводится по всем значениям вектора y при фиксированном векторе Θ , т.е. понимается в смысле (2.5).

8. Аналогичным образом определяется *эффективная безусловная* оценка, однако усреднение проводится по всем возможным значениям векторов y и Θ , т.е. понимается в смысле (2.6).

9. Оценка называется *достаточной*, если для ее вычисления нет необходимости знать каждый компонент $y_1, y_2, \dots, y_n$ апостериорной выборки y , а достаточно иметь одну или несколько функций от выборки, через которые и выражается оценка. Эти функции называют *достаточными статистиками*.

В настоящее время теория статистических решений и математическая статистика рекомендуют много способов вычисления оценок. Эти способы отличаются объемом используемой априорной информации, критериями оценивания, сложностью вычисления оценок, соответствующих различным критериям, и т.д. Однако как бы ни был совершен метод оценивания, принципиально при конечном числе n экспериментальных данных не удастся добиться полного совпадения оценок $\hat{\Theta}$ и оцениваемых параметров Θ . Чтобы судить о степени приближения к оцениваемым параметрам, в рассмотрение вводят понятие *ошибки оценивания* η , определяемой естественным образом: $\eta(y, \Theta) = \hat{\Theta}(y) - \Theta$. Так как вектор наблюдений y случаен и вектор параметров Θ также может быть случайным, то вектор ошибок всегда случаен и поэтому не может быть надежной мерой качества оценивания. Для описания точности оценивания используют неслучайные показатели, построенные на основе случайной ошибки. Для условной

оценки, которой соответствует неслучайный оцениваемый параметр, такими показателями являются *величина смещения*

$$m_{\eta}(\Theta) = M_{y|\Theta}\{\eta(y, \Theta)\} = M_{y|\Theta}\{\hat{\Theta}(y)\} - \Theta,$$

представляющая собой среднее значение ошибки оценивания, и *ковариационная матрица* ошибок оценивания

$$K_{\eta} = M_{y|\Theta}\{[\eta(y, \Theta) - m_{\eta}(\Theta)] [\eta(y, \Theta) - m_{\eta}(\Theta)]^T\},$$

элементы главной диагонали которой представляют собой дисперсии ошибок оценивания отдельных компонентов вектора Θ , а остальные элементы — ковариации между этими ошибками. Для несмещенных ошибок ($m_{\eta}(\Theta) = 0_{m+1}$) ковариационная матрица ошибок принимает вид

$$K_{\eta} = M_{y|\Theta}\{[\hat{\Theta}(y) - \Theta] [\hat{\Theta}(y) - \Theta]^T\}, \quad (2.7)$$

и элементы ее главной диагонали имеют смысл дисперсий отклонений оценок от оцениваемых параметров, т.е. являются мерой разброса оценки относительно оцениваемого параметра.

Точность безусловной оценки, которая соответствует случайному оцениваемому параметру, принято характеризовать матрицей вторых моментов отклонений оценки от оцениваемых параметров:

$$K_{\eta} = M_{y|\Theta}\{[\hat{\Theta}(y) - \Theta] [\hat{\Theta}(y) - \Theta]^T\}. \quad (2.8)$$

Элементы главной диагонали этой матрицы представляют собой средние квадраты ошибок оценивания отдельных компонентов случайного вектора Θ , а остальные элементы являются вторыми смешанными моментами этих ошибок.

Как уже отмечалось, степень приближения оценок к оцениваемым параметрам является ограниченной снизу. Это значит, что при любом способе оценивания нельзя получить оценки, точность которых будет выше определенных значений, являющихся границами принципиально достижимых результатов. Эти границы задаются с помощью *неравенства Рао — Крамера*. Для несмещенных условных оценок справедливо

$$K_{\eta} = M_{y|\Theta}\{[\hat{\Theta}(y) - \Theta] [\hat{\Theta}(y) - \Theta]^T\} \geq \Phi^{-1}. \quad (2.9)$$

Здесь Φ — так называемая *информационная матрица Фишера*, определяемая двумя эквивалентными способами:

$$\begin{aligned} \Phi &= M_{y|\Theta}\left\{\left[\frac{\partial}{\partial \Theta} \ln L(y|\Theta)\right]^T \left[\frac{\partial}{\partial \Theta} \ln L(y|\Theta)\right]\right\} = \\ &= -M_{y|\Theta}\left[\frac{\partial^2}{\partial \Theta^2} \ln L(y|\Theta)\right], \end{aligned} \quad (2.10)$$

где $\frac{\partial}{\partial \Theta} \ln L(y|\Theta)$ определяется как вектор-строка, а

$\frac{\partial^2}{\partial \Theta^2} \ln L(y|\Theta)$ — матрица вторых производных (матрица Гессе, ее определение дано в п. 2.2). Нижняя граница в неравенстве Рао — Крамера, заметим, достигается при эффективных оценках. Аналогичные соотношения существуют для смещенных и безусловных оценок (например, [29]).

2.2. Операции многомерного дифференцирования

Выше мы уже встречались с операциями дифференцирования, отличающимися от традиционно изучаемых в курсе высшей математики их аналогов. Разумеется, это отличие чисто «организационное», не затрагивающее сущности и свойств дифференцирования как математической операции. Рассмотрим наиболее характерные ситуации, нуждающиеся в соответствующих комментариях.

1. Пусть $y = f(x)$ и $f: R^n \rightarrow R$, т.е. рассматривается скалярная функция y от n переменных $x = [x_1 \ x_2 \ \dots \ x_n]^T$. Тогда по определению

$$\frac{dy}{dx} = \frac{df(x)}{dx} = \left[\frac{\partial f(x)}{\partial x_1} \ \frac{\partial f(x)}{\partial x_2} \ \dots \ \frac{\partial f(x)}{\partial x_n} \right], \quad (2.11)$$

т.е. производная от скалярной функции по векторному аргументу определяется как вектор-строка частных производных. Вектор-столбец

$$\nabla f(x) = \left[\frac{df(x)}{dx} \right]^T \quad (2.12)$$

образует градиент функции $f(x)$ в точке x .

Рассмотрим два характерных случая. Пусть $y = x^T A x$, $A \in R^{n \times n}$. При $n = 2$ ($A = [a_{ij}]$, $i, j = 1, 2$) легко получаем $y = a_{11}x_1^2 + (a_{12} + a_{21})x_1x_2 + a_{22}x_2^2$ и в соответствии с определением

$$\begin{aligned} \frac{dy}{dx} &= [2a_{11}x_1 + (a_{12} + a_{21})x_2, (a_{12} + a_{21})x_1 + 2a_{22}x_2] = \\ &= [x_1 x_2] \begin{bmatrix} 2a_{11} & a_{12} + a_{21} \\ a_{12} + a_{21} & 2a_{22} \end{bmatrix} = \\ &= [x_1 x_2] \left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} \right) = x^T (A + A^T). \end{aligned}$$

По аналогии и для общего случая ($n > 2$) можем получить

$$\frac{d}{dx} x^T A x = x^T (A + A^T), \quad \nabla x^T A x = (A + A^T)x. \quad (2.13)$$

Соответственно при симметрической матрице A , т.е. для квадратичной формы $x^T A x$, имеем $\frac{d}{dx} x^T A x = 2x^T A$, $\nabla x^T A x = 2Ax$. Аналогичным образом для линейной формы $y = c^T x$ находим $\frac{d}{dx} c^T x = c^T$, $\nabla c^T x = c$.

2. Для функции $f: R^n \rightarrow R$ вводится понятие второй производной по вектору x . По определению

$$\nabla^2 f(x) = \frac{d^2}{dx^2} f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \dots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \in R^{n \times n}. \quad (2.14)$$

Эту матрицу называют матрицей вторых производных, или матрицей Гессе, а ее определитель — гессианом. В частности, $\nabla^2 x^T A x = A + A^T$ и при симметрической матрице $\nabla^2 x^T A x = 2A$.

3. Пусть $y = f(x)$ и $f: R^n \rightarrow R^m$, т.е. рассматривается m -мерная вектор-функция $f(x) = [f_1(x) f_2(x) \dots f_m(x)]^T$ от n переменных. Тогда по определению

$$\frac{df(x)}{dx} = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \dots & \frac{\partial f_1(x)}{\partial x_n} \\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_2} & \dots & \frac{\partial f_2(x)}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_m(x)}{\partial x_1} & \frac{\partial f_m(x)}{\partial x_2} & \dots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix} \in R^{m \times n}. \quad (2.15)$$

Эту матрицу принято называть матрицей Якоби, а ее определитель (при $m = n$) — якобианом. В частности, для функции $y = Ax$, где $A \in R^{n \times m}$, имеем

$$\frac{d}{dx} Ax = A. \quad (2.16)$$

4. Если $x \in R^n$ и $u(x) \in R^m$, $v(x) \in R^m$ — две вектор-функции, то производная по вектору x скалярного произведения этих функций определяется как вектор-строка

$$\frac{d}{dx} (u(x), v(x)) = u^T(x) \frac{d}{dx} v(x) + v^T(x) \frac{d}{dx} u(x) \quad (2.17)$$

и, как следствие, производная от квадрата нормы, согласованной со скалярным произведением, и градиент находятся из выражений:

$$\frac{d}{dx} \|u(x)\|^2 = \frac{d}{dx} (u(x), u(x)) = 2u^T(x) \frac{d}{dx} u(x), \quad (2.18)$$

$$\nabla \|u(x)\|^2 = 2 \left[\frac{d}{dx} u(x) \right]^T u(x).$$

2.3. Метод наименьших квадратов

2.3.1. МНК-оценки

Возвратимся теперь к проблеме оценивания параметров Θ регрессионной модели (2.3). Дополнительно уточним некоторые положения, касающиеся этой модели. Прежде всего условимся в последующих процедурах вектор Θ классифицировать как неизвестный, что соответствует отсутствию всякой априорной информации о его свойствах. Далее будем полагать, что ошибки $\epsilon_j(x_j) = \epsilon_j, j = 1, 2, \dots, n$, т.е. не зависят от значений экзогенных переменных, принимаемых при проведении эксперимента. Дополнительно эти ошибки полагаем центрированными, не коррелированными друг с другом и имеющими при всех j одну и ту же, вообще говоря неизвестную, дисперсию σ^2 (в таких случаях измерения (2.3) принято называть *равноточными*). Таким образом,

$$M\{\epsilon_j\} = 0, \quad M\{\epsilon_i \epsilon_j\} = \begin{cases} 0, & i \neq j, \\ \sigma^2, & i = j, \end{cases} \quad i, j = 1, 2, \dots, n. \quad (2.19)$$

Итак, предполагается, что результатом проведения некоторого эксперимента является совокупность наблюдений (апостериорная выборка) (2.3), упорядоченных в форме (2.4). Матрица Ψ определяется выбранной системой функций при аппроксимации неизвестной функции регрессии и значениями экзогенных переменных при проведении эксперимента. Векторы Θ и ϵ в (2.4) соответствуют сделанным выше комментариям. Задача заключается в поиске оценки $\hat{\Theta}(y)$ вектора Θ как функции наблюдений y .

Одним из наиболее распространенных подходов к решению сформулированной задачи в настоящее время является *метод наименьших квадратов* (МНК). Его особенность прежде всего проявляется в отсутствии каких-либо жестких претензий к априорной информации об оцениваемых параметрах и экспериментальных ошибках. И это очень важно для эконометрических задач, которые, как правило, не допускают экспериментального повторения и в этом смысле являются однократными. Поэтому МНК, несмотря на свою большую историю (он применялся еще Гауссом и Лежандром), в экономических исследованиях занимает лидирующую позицию. Существо метода заключается в следующем.

В рассмотрение вводится целевая функция

$$J = \sum_{j=1}^n [y_j - \psi^T(x_j)\Theta]^2, \quad (2.20)$$

представляющая собой сумму квадратов удаления экспериментальных данных от значений, определяемых регрессионной составляющей в модели наблюдений (2.3). Заметим, что именно эти значения наблюдались бы в эксперименте, если бы функция регрессии действительно относилась к классу (1.19), отсутствовали латентные переменные и сам процесс измерений эндогенной переменной был идеален. Структура целевой функции (2.20) является характерным признаком МНК. Наилучшей (оптимальной) оценкой (МНК-оценкой), найденной по этому методу, считается такое значение параметра Θ , при котором функция (2.20) достигает своего минимума. Таким образом, МНК-оценка ищется из условия

$$\hat{\Theta}(y) = \arg \min_{\Theta} J = \arg \min_{\Theta} \sum_{j=1}^n [y_j - \psi^T(x_j)\Theta]^2. \quad (2.21)$$

Если, используя обозначения (2.4), целевую функцию (2.20) переписать в эквивалентной, но более компактной форме

$$J = (y - \Psi\Theta)^T(y - \Psi\Theta) = \|y - \Psi\Theta\|^2,$$

то задача (2.21) будет формулироваться так:

$$\hat{\Theta}(y) = \arg \min_{\Theta} \|y - \Psi\Theta\|^2. \quad (2.22)$$

Запишем необходимое условие минимума для этой задачи, вычислив в соответствии с (2.18), (2.16) градиент функции J и приравняв его нуль-вектору:

$$\nabla \|y - \Psi\Theta\|^2 = -2\Psi^T(y - \Psi\Theta) = 0_{m+1} \Rightarrow \Psi^T\Psi\Theta = \Psi^Ty. \quad (2.23)$$

Уравнение (2.23) представляет собой матрично-векторную форму записи системы линейных неоднородных алгебраических уравнений, состоящей из $m+1$ скалярных уравнений и содержащей такое же число неизвестных $\Theta_0, \Theta_1, \dots, \Theta_m$. Прежде чем приступить к поиску решения данной системы, сформулируем ряд необходимых для этой цели положений.

Определение 2.1. Матрица $A \in R^{m \times n}$ называется *матрицей полного ранга*, если ее ранг $\text{rank } A$ удовлетворяет условию $\text{rank } A = \min(n, m)$.

Утверждение 2.1. Для произвольной матрицы A справедливо: $\text{rank } A = \text{rank } A^T A = \text{rank } A A^T$.

Доказательство утверждения можно найти, например, в [8, 9].

Утверждение 2.2. Пусть в (2.23) матрица $\Psi^T \in R^{(m+1) \times n}$ является матрицей полного ранга и $n > (m+1)$. Тогда матрица $\Psi^T \Psi$ является невырожденной.

Действительно, матрица $\Psi^T \Psi \in R^{(m+1) \times (m+1)}$, и в силу ограничения $n > m+1$ и предыдущего утверждения имеем $\text{rank } \Psi^T \Psi = \text{rank } \Psi = m+1$. Но это значит, что определитель $|\Psi^T \Psi| \neq 0$, т.е. матрица $\Psi^T \Psi$ не вырождена.

Возвратимся теперь к системе уравнений (2.23). Из невырожденности матрицы $\Psi^T \Psi$ вытекает существование обратной матрицы $(\Psi^T \Psi)^{-1}$, что позволяет найти решение системы в форме $\Theta = (\Psi^T \Psi)^{-1} \Psi^T y$. Так как целевая функция J является выпуклой, то необходимое условие ее минимума является и достаточным, поэтому найденное решение системы (2.23) представляет собой решение задачи (2.22) и, следовательно, МНК-оценка вектора Θ определяется выражением

$$\hat{\Theta}(y) = (\Psi^T \Psi)^{-1} \Psi^T y. \quad (2.24)$$

Если учесть определение матрицы Ψ , данное в комментариях к (2.4), то эту оценку можно представить в другой редакции:

$$\hat{\Theta}(y) = \left(\sum_{i=1}^n \psi(x_i) \psi^T(x_i) \right)^{-1} \sum_{i=1}^n \psi(x_i) y_i. \quad (2.25)$$

При линейной относительно оцениваемых параметров модели наблюдений МНК-оценка оказывается линейной относительно вектора наблюдений y . Ограничение $n > m+1$, физически означающее, что число наблюдений при организации эксперимента должно быть больше числа оцениваемых параметров регрессионной модели, является существенным и означает, что лишь в этом случае можно успешно справиться с влиянием неконтролируемых ошибок ϵ . При $n < m+1$ $\text{rank } \Psi = \text{rank } \Psi^T \Psi \leq n$ и матрица $\Psi^T \Psi$ оказывается вырожденной. В этом случае система (2.23), даже если она совместна, не является определенной.

2.3.2. Основные свойства МНК-оценок. Теорема Маркова

Займемся теперь анализом основных свойств найденной оценки. Покажем прежде всего, что МНК-оценка (2.24) является несмещенной. С этой целью, используя (2.4), находим

$$M_{y|\Theta} \{\hat{\Theta}(y)\} = M_{y|\Theta} \{(\Psi^T \Psi)^{-1} \Psi^T y\} = M_{\epsilon|\Theta} \{(\Psi^T \Psi)^{-1} \Psi^T (\Psi \Theta + \epsilon)\} = \Theta,$$

что и является условием несмещенности.

Второй важнейшей характеристикой оценки является ковариационная матрица ошибки. Для несмещенной оценки она определяется как (2.7). Ошибка оценивания находится естественным образом:

$$\begin{aligned} \eta(y, \Theta) &= \hat{\Theta}(y) - \Theta = (\Psi^T \Psi)^{-1} \Psi^T y - \Theta = \\ &= (\Psi^T \Psi)^{-1} \Psi^T (\Psi \Theta + \epsilon) - \Theta = (\Psi^T \Psi)^{-1} \Psi^T \epsilon \end{aligned} \quad (2.26)$$

и, следовательно, $K_\eta = M\{\eta \eta^T\} = (\Psi^T \Psi)^{-1} \Psi^T M\{\epsilon \epsilon^T\} \Psi (\Psi^T \Psi)^{-1} = (\Psi^T \Psi)^{-1} \Psi^T K_\epsilon \Psi (\Psi^T \Psi)^{-1}$, где K_ϵ — ковариационная матрица вектора ϵ . Так как в соответствии с (2.19) $K_\epsilon = \sigma^2 E$, где E — единичная матрица, окончательно находим

$$K_\eta = M\{\eta \eta^T\} = \sigma^2 (\Psi^T \Psi)^{-1}. \quad (2.27)$$

Соотношение (2.27) принципиально позволяет выявить точностные возможности метода наименьших квадратов. К сожалению, практическая значимость этого результата невелика, так как в большинстве эконометрических задач дисперсия σ^2 экспериментальных ошибок неизвестна. Однако МНК-оценкам присуще еще одно важное свойство.

Теорема Маркова. Пусть модель наблюдений имеет структуру (2.4), вектор ϵ удовлетворяет условиям (2.19), матрица Ψ является матрицей полного ранга и $n > m+1$. Тогда МНК-оценка (2.24) является наилучшей (в смысле наименьшей дисперсии ошибок оценивания) среди всех линейных несмещенных оценок.

Для доказательства теоремы предположим, что мы хотим найти наилучшую в указанном смысле линейную несмещенную оценку k -го компонента Θ_k вектора Θ . Этот компонент очевидным образом выразим через сам вектор: $\Theta_k = h_k^T \Theta$, где h_k — век-

тор, у которого на $(k+1)$ -й позиции ($k = 0, 1, \dots, m$) находится единица, а остальные компоненты равны нулю. Оценку $\hat{\Theta}_k$ величины Θ_k будем искать в классе линейных функций наблюдений, т.е. в виде $\hat{\Theta}_k = W_k^T y$, где W_k — вектор подлежащих определению весовых коэффициентов. Пусть η_k — ошибка оценивания, т.е.

$$\eta_k = \hat{\Theta}_k - \Theta_k = W_k^T y - \Theta_k = W_k^T (\Psi \Theta + \varepsilon) - h_k^T \Theta = (W_k^T \Psi - h_k^T) \Theta + W_k^T \varepsilon.$$

Первое слагаемое в этом выражении зависит от неизвестного вектора Θ и поэтому его величину нельзя оценить даже ориентировочно. Чтобы ошибку оценивания сделать независимой от оцениваемого вектора, потребуем

$$\Psi^T W_k - h_k = 0_{m+1}. \quad (2.28)$$

При этом дополнительно оказывается $M\{\eta_k\} = 0$, т.е. соотношения (2.28) оказываются условиями несмещенности. В развернутом виде выражение (2.28) представляет собой систему линейных неоднородных алгебраических уравнений, состоящую из $m+1$ уравнений с n неизвестными в виде компонентов вектора W_k . Так как $n > m+1$ и Ψ — матрица полного ранга, то при указанной структуре вектора h_k эта система совместна, но не определена, т.е. имеет неограниченное число решений. Если равенство (2.28) выполняется, ошибка оценивания будет определяться только ошибками эксперимента и искомой весовой функцией: $\eta_k = W_k^T \varepsilon$, причем дисперсия σ_k^2 этой ошибки оказывается равной

$$\sigma_k^2 = \sigma^2 W_k^T W_k. \quad (2.29)$$

Последующая задача заключается в поиске на множестве решений системы (2.28) такого вектора W_k , который минимизирует величину (2.29). Эта традиционная задача на условный экстремум решается методом неопределенных множителей Лагранжа. Составим функцию Лагранжа

$$L(W_k, \lambda) = \sigma^2 W_k^T W_k + \lambda^T (\Psi^T W_k - h_k),$$

где λ — вектор неопределенных множителей Лагранжа.

Стационарные точки этой функции находятся из условия равенства нулевым векторам ее градиентов по векторам W_k и λ :

$$2\sigma^2 W_k - \Psi \lambda = 0_n, \quad \Psi^T W_k - h_k = 0_{m+1}.$$

Выразив из первого уравнения этой системы вектор W_k и подставив его значение во второе, получим $0,5\sigma^{-2}\Psi^T\Psi\lambda - h_k = 0_{m+1}$. Так как матрица $\Psi^T\Psi$ по доказанному не вырождена, то отсюда однозначно находится вектор λ , подстановка значения которого в первое уравнение предыдущей системы позволяет найти искомый вектор весовых коэффициентов

$$W_k = \Psi(\Psi^T\Psi)^{-1} h_k, \quad (2.30)$$

что приводит к следующему выражению оценки, наилучшей в классе линейных несмещенных оценок:

$$\hat{\Theta}_k = h_k^T (\Psi^T\Psi)^{-1} \Psi^T y, \quad k = 0, 1, \dots, m. \quad (2.31)$$

Упорядочив все эти $m+1$ оценок в форме одной векторной оценки, получим

$$\hat{\Theta} = \begin{bmatrix} h_0^T \\ h_1^T \\ \dots \\ h_m^T \end{bmatrix} (\Psi^T\Psi)^{-1} \Psi^T y.$$

Но первый матричный сомножитель в этом выражении есть не что иное, как единичная матрица, и, следовательно,

$$\hat{\Theta} = (\Psi^T\Psi)^{-1} \Psi^T y.$$

Этот результат полностью совпадает с ранее полученной МНК-оценкой (2.24). Таким образом, МНК-оценка действительно оказывается наилучшей в классе линейных несмещенных оценок.

Значение теоремы Маркова становится еще более важным, если предположить, что вектор экспериментальных ошибок является гауссовским, а именно $\varepsilon \sim N(0, \sigma^2 E)$. Покажем, что в этом случае МНК-оценка оказывается *наилучшей и среди всех нелинейных несмещенных оценок*, т.е. *эффективной*. С этой целью обратимся к неравенству Рао — Крамера (2.9) и найдем информационную

матрицу Фишера, воспользовавшись вторым определением из (2.10). Прежде всего необходимо найти условную плотность экспериментальных данных $L(y|\Theta)$. Если в модели наблюдений (2.4) вектор Θ зафиксирован, то природа «случайности» вектора y порождена аналогичным качеством вектора ϵ , так как иных источников случайности в этой модели нет. Поэтому вектор y при фиксированном Θ будет также гауссовским с ковариационной матрицей $\sigma^2 E$ и математическим ожиданием $\Psi\Theta$, что непосредственно следует из (2.4). Таким образом,

$$L(y|\Theta) = N_y(\Psi\Theta, \sigma^2 E) \Rightarrow \ln L(y|\Theta) = \text{const} - 0,5\sigma^{-2}\|y - \Psi\Theta\|^2,$$

где через const обозначено независимое от Θ слагаемое. Дважды продифференцировав последнюю функцию по Θ (см. (2.14)), найдем

$$\Phi^{-1} = \sigma^2(\Psi^T\Psi)^{-1}.$$

Но этот результат полностью совпадает с ковариационной матрицей ошибки оценивания (2.27). Следовательно, при гауссовских некоррелированных экспериментальных ошибках неравенство Рао – Крамера вырождается в равенство, соответствующее точной нижней грани неравенства, а это и является признаком эффективности МНК-оценки в указанных условиях.

2.3.3. МНК-оценки параметров производственной функции Кобба – Дугласа

В качестве своеобразной иллюстрации техники построения МНК-оценок рассмотрим одну частную, но важную для многих эконометрических приложений задачу.

Производственными функциями принято называть соотношения между используемыми в производстве материальными и трудовыми ресурсами (обобщенно – производственными ресурсами) и выпускаемой продукцией. Пусть некоторое предприятие (отрасль, объединение, фирма и т.п.) в течение определенного промежутка времени производит q наименований продукции соответственно в количествах $u_1, u_2, \dots, u_q$, представленных в натуральных единицах измерения или в денежном эквиваленте. Чтобы производить эту продукцию, необходимы s видов ресурсов со-

ответственно в количествах $v_1, v_2, \dots, v_s$, также представленных в определенных единицах измерения. Тогда неявная функция

$$F(u, v, A) = 0,$$

связывающая вектор выпускаемой продукции $u = [u_1 \ u_2 \ \dots \ u_q]^T$, вектор используемых ресурсов $v = [v_1 \ v_2 \ \dots \ v_s]^T$ и вектор $a = [a_1 \ a_2 \ \dots \ a_p]^T$ параметров, представляет собой производственную функцию. Причина для включения вектора параметров в состав производственной функции та же, что и в случае регрессионной модели: точно указать характер аналитической зависимости между векторами u и v невозможно; однако можно предвидеть параметрический класс функций, которому принадлежит производственная функция; в последующем эти параметры можно оценить по результатам надлежащим образом поставленного эксперимента. Суть последнего может заключаться в следующем: регистрируются объемы выпускаемой продукции при различных объемах используемых ресурсов, что приводит к массиву экспериментальных данных, подобному (1.1), (1.2); на основе этих данных находятся, например, МНК-оценки вектора a .

В настоящее время наиболее часто производственные функции применяют при решении однопродуктовых экономических задач ($q = 1$) с несколькими материальными и трудовыми ресурсами ($s > 1$). В этом случае $u = u$ – скалярная величина, и производственную функцию, разрешив относительно u , представляют в виде

$$u = f(v, a)$$

и часто называют функцией выпуска.

Вид функции $f(\cdot)$ не может быть совершенно произвольным. Она должна удовлетворять определенным условиям. Вот некоторые из них.

1. Производство невозможно при отсутствии хотя бы одного из ресурсов, т.е.

$$\begin{aligned} f(0, v_2, v_3, \dots, v_s, a) &= f(v_1, 0, v_3, \dots, v_s, a) = \\ &= \dots = f(v_1, \dots, v_{s-1}, 0, a) = 0. \end{aligned}$$

Заметим, что если некоторый вид ресурса может компенсироваться другими, записанные условия относительно этого ресурса могут и не выполняться.

2. При увеличении затрат производственных ресурсов выпуск продукции не уменьшается, т.е. для дифференцируемой производственной функции $\nabla_v f(v, a) \geq 0$.

3. Увеличение одного ресурса при неизменных значениях других ресурсов должно приводить ко все меньшим приростам выпускаемой продукции. Это достигается, если

$\frac{\partial^2}{\partial v_i^2} f(v, a) < 0$, $i = 1, 2, \dots, s$, что эквивалентно требованию положительной знакоопределенности матрицы $-\nabla_v^2 f(v, a)$.

Одной из наиболее ранних производственных функций является предложенная в 1928 г. П. Дугласом и Д. Коббом функция для определения влияния величины затрачиваемого капитала (v_1) и объема труда (v_2) на объем выпускаемой продукции (u) в обрабатывающей промышленности США. Эта функция была предложена в форме

$$u = a v_1^b v_2^c, a > 0, b \geq 0, c \geq 0, b + c = 1.$$

Здесь в терминах предыдущих обозначений $a = [a \ b \ c]^T$ — вектор параметров, оцениваемых по экспериментальным данным. В качестве этих данных использовались зафиксированные в промежутке с 1899 по 1922 г. объемы затрачиваемого капитала $v_1^{(j)}$, труда $v_2^{(j)}$ и соответствующий им объем выпускаемой продукции $u^{(j)}$, $j = 1, 2, \dots, 24$.

Характерная особенность предложенной производственной функции проявляется в нелинейной зависимости от параметров, что существенно усложняет процедуру их последующего оценивания по экспериментальным данным. Поэтому оказалось целесообразным предварительное логарифмирование этой функции, приводящее к очевидному результату:

$$\ln u - \ln v_2 = \ln a + b(\ln v_1 - \ln v_2),$$

уже линейно зависящему от параметров $\ln a$ и b . Тогда математическую модель экспериментальных данных можно отобразить равенством

$$\ln u^{(j)} - \ln v_2^{(j)} = \ln a + b(\ln v_1^{(j)} - \ln v_2^{(j)}) + \varepsilon_j^{(j)}, j = 1, 2, \dots, 24,$$

в котором слагаемое $\varepsilon_j^{(j)}$ моделирует неизбежные отклонения результатов эксперимента от предполагаемых теоретических значе-

ний. Если теперь ввести обозначения: $y_j = \ln u^{(j)} - \ln v_2^{(j)}$, $\Theta_0 = \ln a$, $\Theta_1 = b$, $\Theta = [\Theta_0 \ \Theta_1]^T$, $x_j = \ln v_1^{(j)} - \ln v_2^{(j)}$, $\Psi^T(x_j) = [1 \ x_j]$, $\varepsilon_j = \varepsilon_j^{(j)}$ и положить $n = 24$, экспериментальные данные опишутся соотношением

$$y_j = \Psi^T(x_j)\Theta + \varepsilon_j, j = 1, 2, \dots, n,$$

что полностью совпадает с регрессионной моделью (2.3). В терминах этого совпадения величины y_j и x_j , $j = 1, 2, \dots, n$, можно интерпретировать как экспериментальную выборку соответственно эндогенной и экзогенной переменных данной экономической системы. МНК-оценка вектора параметров Θ находится с помощью уже известного правила (2.24), которое с учетом введенных обозначений и после проведения соответствующих матрично-векторных операций может быть конкретизировано следующим образом:

$$\hat{\Theta}_0 = \bar{y} - \bar{x}\hat{\Theta}_1, \quad \hat{\Theta}_1 = \frac{k_{xy} - \bar{x}\bar{y}}{(x^2) - \bar{x}^2},$$

где, в свою очередь, использованы обозначения

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j, \quad \bar{y} = \frac{1}{n} \sum_{j=1}^n y_j, \quad k_{xy} = \frac{1}{n} \sum_{j=1}^n x_j y_j, \quad (x^2) = \frac{1}{n} \sum_{j=1}^n x_j^2.$$

Зная оценки параметров Θ_0 и Θ_1 , не представляет труда найти оценки исходных параметров a , b , c производственной функции. Следует заметить при этом, что исследование свойств этих оценок может породить самостоятельную задачу. Обратим внимание еще на одну особенность полученных оценок: они выражаются не через отдельные компоненты $y_1, y_2, \dots, y_n$ экспериментальных данных, а через функции $\bar{y}, k_{xy}$ этих данных. Поэтому найденные оценки можно классифицировать как *достаточные*, а функции $\bar{y}, k_{xy}$ рассматривать как *достаточные статистики*.

2.3.4. Идемпотентные матрицы

Для последующего анализа нам потребуются матрицы, обладающие определенными специфическими свойствами. Одной из них является идемпотентная.

Определение 2.2. Матрица $P \in R^{q \times q}$ называется *идемпотентной*, если выполняется условие $P^2 = PP = P$, т.е. квадрат идемпотентной матрицы равен самой матрице.

Как следствие этого определения, очевидно, $P^k = P$ при $\forall k \in N$. Идемпотентные матрицы обладают рядом замечательных свойств. Ограничим рассмотрение важным для нас случаем симметрических матриц.

Утверждение 2.3. Собственные числа идемпотентной матрицы равны нулю или единице.

Действительно, пусть v и λ — собственное число и соответствующий ему собственный вектор матрицы P . Тогда по определению $Pv = \lambda v \Rightarrow P^2v = P(\lambda v) = \lambda Pv = \lambda^2 v \Rightarrow P^2v = \lambda v$, т.е. $\lambda = \lambda^2$, что выполняется при $\lambda = 0$ или $\lambda = 1$.

Утверждение 2.4. Ранг симметрической идемпотентной матрицы P равен ее следу: $\text{rang } P = \text{Sp } P$.

Так как P — симметрическая матрица, то существует ортогональная матрица T , диагонализующая матрицу P , т.е. обладающая свойствами $T^T T = E$, $T^T P T = v$, где $v = \text{diag } [v_i]$, $i = 1, 2, \dots, q$, — диагональная матрица из собственных чисел матрицы P . Отсюда, учитывая, что $T^T = T^{-1}$, получаем $P = T v T^T$. Так как T — невырожденная матрица, то $\text{rang } P = \text{rang } v$. Но ранг диагональной матрицы v равен числу ее ненулевых строк. Число же таких строк совпадает с количеством собственных чисел, равных единице, т.е. со следом матрицы v : $\text{rang } v = \text{Sp } v$. В свою очередь, $\text{Sp } P = \text{Sp } (T v T^T) = \text{Sp } (T T^T v) = \text{Sp } v$ и, следовательно, $\text{rang } P = \text{Sp } P$.

Заметим, что здесь и далее используются свойства следа матрицы: $\text{Sp } (AB) = \text{Sp } (BA)$; $\text{Sp } (A) = \text{Sp } (A^T)$; $\text{Sp } (A + B) = \text{Sp } A + \text{Sp } B$; $\text{Sp } (A A^T) = \text{Sp } (A^T A) = A^T A$, где $A \in R^n$.

Обратим внимание еще на одну особенность идемпотентной матрицы, по существу являющуюся следствием доказанного утверждения: так как ранг ненулевой матрицы является целым положительным числом, сумма диагональных элементов симметрической идемпотентной матрицы представляет собой положительное целое число.

Утверждение 2.5. Пусть $\epsilon \in R^q$ — стандартный гауссовский вектор ($M\{\epsilon\} = 0_q$, $M\{\epsilon \epsilon^T\} = \sigma^2 E$) и P — идемпотентная q -матрица. Тогда случайная величина $\mu = \epsilon^T P \epsilon$ подчинена $\chi^2(r)$ распределению, где $r = \text{rang } P$.

Действительно, $\epsilon^T P \epsilon = \epsilon^T T v T^T \epsilon = (T^T \epsilon)^T v (T^T \epsilon)$. Так как T — ортогональная матрица, то $T^T \epsilon \sim N(0_q, \sigma^2 E)$, т.е. также стандартный гауссовский вектор. Но тогда случайная величина $\epsilon^T P \epsilon = (T^T \epsilon)^T v (T^T \epsilon)$ представляет собой сумму r квадратов стандартных гауссовских величин и распределена, следовательно, по закону $\chi^2(r)$.

2.3.5. Несмещенная оценка дисперсии экспериментальных ошибок и ее свойства

Выше было получено выражение (2.27), определяющее точность МНК-оценок, и одновременно отмечена ограниченность его применения, обусловленная неизвестностью дисперсии σ^2 экспериментальных ошибок в большинстве реальных эконометрических задач. Поэтому оказывается целесообразным оценивание параметров Θ регрессионной модели совместить с оцениванием дисперсии экспериментальных ошибок, чтобы в последующем иметь возможность аргументированно анализировать свойства МНК-оценок регрессионных параметров.

С целью получения оценки дисперсии найдем минимальное значение целевой функции (2.20), соответствующее МНК-оценке регрессионных параметров, т.е.

$$\min_{\Theta} J = \min_{\Theta} \sum_{j=1}^n [y_j - \Psi^T(x_j) \Theta]^2 = \min_{\Theta} \|y - \Psi \Theta\|^2 = \|y - \Psi \hat{\Theta}\|^2. \quad (2.32)$$

Легко устанавливается

$$\|y - \Psi \hat{\Theta}\|^2 = (y - \Psi \hat{\Theta})^T (y - \Psi \hat{\Theta}) = y^T (y - \Psi \hat{\Theta}) - \hat{\Theta}^T \Psi^T (y - \Psi \hat{\Theta}).$$

Так как в силу необходимого условия (2.23) оптимальности МНК-оценки вычитаемое справа в этом выражении представляет собой нуль, то получаем

$$\|y - \Psi \hat{\Theta}\|^2 = y^T (y - \Psi \hat{\Theta}).$$

В свою очередь, если учтем модель экспериментальных данных (2.4) и выражение МНК-оценки (2.24), получим

$$(y - \Psi \hat{\Theta}) = (E_n - \Psi (\Psi^T \Psi)^{-1} \Psi^T) \epsilon, \quad (2.33)$$

где, как обычно, E_n — единичная n -матрица.

Если теперь еще раз воспользоваться моделью (2.4) и учесть то обстоятельство, что матрица $\Psi^T(E_n - \Psi(\Psi^T\Psi)^{-1}\Psi^T)$ является нулевой, то установим окончательно

$$\min J = \varepsilon^T(E_n - \Psi(\Psi^T\Psi)^{-1}\Psi^T)\varepsilon. \quad (2.34)$$

Величина (2.34) является случайной. Найдем ее среднее значение $M\{\min J\}$. Пусть $E_n - \Psi(\Psi^T\Psi)^{-1}\Psi^T = [a_{ij}]$, $i, j = 1, 2, \dots, n$. Тогда

$$\varepsilon^T(E_n - \Psi(\Psi^T\Psi)^{-1}\Psi^T)\varepsilon = \sum_{i=1}^n \sum_{j=1}^n a_{ij} \varepsilon_i \varepsilon_j.$$

Усредняя это выражение, с учетом (2.19) получаем

$$M\{\min J\} = \sigma^2 \sum_{i=1}^n a_{ii} = \sigma^2 \text{Sp}(E_n - \Psi(\Psi^T\Psi)^{-1}\Psi^T).$$

Используя свойства следа матрицы, находим

$$M\{\min J\} = \sigma^2(\text{Sp} E_n - \text{Sp} \Psi(\Psi^T\Psi)^{-1}\Psi^T) = \quad (2.35)$$

$$\sigma^2(n - \text{Sp}(\Psi^T\Psi(\Psi^T\Psi)^{-1})) = \sigma^2(n - \text{Sp} E_{m+1}) = \sigma^2(n - m - 1).$$

Рассмотрим случайную величину

$$\hat{\sigma}^2 = \frac{1}{n-m-1} \min J. \quad (2.36)$$

Из сопоставления с (2.35) следует, что среднее значение этой величины $M\{\hat{\sigma}^2\} = \sigma^2$. Но это означает, что величина $\hat{\sigma}^2$ может быть принята в качестве *несмещенной* оценки дисперсии σ^2 экспериментальных ошибок. Таким образом, окончательно

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-m-1} \sum_{j=1}^n (y_j - \Psi^T(x_j)\hat{\Theta})^2 = \frac{1}{n-m-1} \|y - \Psi\hat{\Theta}\|^2 = \\ &= \frac{1}{n-m-1} y^T(y - \Psi\hat{\Theta}). \end{aligned} \quad (2.37)$$

Оценка (2.37) позволяет конкретизировать точностные свойства МНК-оценки регрессионных параметров, если в подвергав-

шемся критическим комментарием выражению (2.27) дисперсию σ^2 заменить ее оценкой (2.37). Так как эта оценка является случайной величиной, при обширном ее использовании в процедурах эконометрических исследований недостаточно знания только ее математического ожидания, но необходим более обширный спектр вероятностных характеристик. Установим одну из них. Для этого воспользуемся определением (2.36) и откорректируем его в соответствии с (2.34):

$$\hat{\sigma}^2 = \frac{1}{n-m-1} \varepsilon^T(E_n - \Psi(\Psi^T\Psi)^{-1}\Psi^T)\varepsilon. \quad (2.38)$$

Утверждение 2.6. Матрица $E_n - \Psi(\Psi^T\Psi)^{-1}\Psi^T$ является идемпотентной с рангом $\text{rank}(E_n - \Psi(\Psi^T\Psi)^{-1}\Psi^T) = n - m - 1$.

Справедливость утверждения проверяется непосредственно:

$$\begin{aligned} (E_n - \Psi(\Psi^T\Psi)^{-1}\Psi^T)^2 &= (E_n - \Psi(\Psi^T\Psi)^{-1}\Psi^T)(E_n - \Psi(\Psi^T\Psi)^{-1}\Psi^T) = \\ &= E_n - \Psi(\Psi^T\Psi)^{-1}\Psi^T; \end{aligned}$$

$$\text{rank}(E_n - \Psi(\Psi^T\Psi)^{-1}\Psi^T) = \text{Sp}(E_n - \Psi(\Psi^T\Psi)^{-1}\Psi^T) = n - m - 1.$$

Ориентируясь на перспективу применения этого утверждения, следующим образом подкорректируем выражение (2.38):

$$\frac{\hat{\sigma}^2(n-m-1)}{\sigma^2} = \left(\frac{\varepsilon}{\sigma}\right)^T (E_n - \Psi(\Psi^T\Psi)^{-1}\Psi^T) \left(\frac{\varepsilon}{\sigma}\right). \quad (2.39)$$

Так как вектор ε/σ представляет собой стандартный гауссовский вектор, стоящая справа в (2.39) случайная величина в соответствии с утверждением 2.5 подчинена χ^2 -распределению с $n - m - 1$ степенями свободы. Таким образом, получен следующий важный результат.

Утверждение 2.7. Случайная величина $\frac{\hat{\sigma}^2(n-m-1)}{\sigma^2}$ подчинена χ^2 -распределению с $n - m - 1$ степенями свободы, т.е. распределению $\chi^2(n-m-1)$.

Оценки $\hat{\sigma}^2$ и $\hat{\Theta}$ обладают еще одним любопытным свойством. Обе они являются функциями вектора наблюдений y . И тем не менее справедлив следующий результат.

Утверждение 2.8. Оценки $\hat{\sigma}^2(y)$ и $\hat{\Theta}(y)$ некоррелированы, а при гауссовских экспериментальных ошибках ϵ независимы.

Для доказательства МНК-оценку (2.24), используя модель данных (2.4), представим в виде

$$\hat{\Theta} = \Theta + (\Psi^T \Psi)^{-1} \Psi^T \epsilon. \quad (2.40)$$

Воспользовавшись (2.33), построим величину

$$(y - \Psi \hat{\Theta}) \hat{\Theta}^T = (E - \Psi(\Psi^T \Psi)^{-1} \Psi^T) \epsilon (\epsilon^T \Psi(\Psi^T \Psi)^{-1} + \Theta^T)$$

и найдем ее среднее значение

$$M\{(y - \Psi \hat{\Theta}) \hat{\Theta}^T\} = (E - \Psi(\Psi^T \Psi)^{-1} \Psi^T) M\{\epsilon \epsilon^T\} \Psi(\Psi^T \Psi)^{-1} = 0_{n \times (m+1)},$$

где $0_{n \times (m+1)}$ — нулевая $n \times (m+1)$ -матрица. Здесь учитывается, что $M\{\epsilon\} = 0$ и $M\{\epsilon \epsilon^T\} = \sigma^2 E$. Следовательно, векторы $y - \Psi \hat{\Theta}$ и $\hat{\Theta}$ не коррелированы. Но тогда не коррелированы случайная величина $\|y - \Psi \hat{\Theta}\|^2$ и случайный вектор $\hat{\Theta}$, а это влечет за собой некоррелированность самих оценок $\hat{\sigma}^2$ и $\hat{\Theta}$, так как первая из них непосредственно выражается через $\|y - \Psi \hat{\Theta}\|^2$. Если дополнительно вектор ϵ экспериментальных ошибок является гауссовским, то векторы $y - \Psi \hat{\Theta}$ и $\hat{\Theta}$ также гауссовские и из их некоррелированности следует независимость, порождающая независимость и оценок $\hat{\sigma}^2$ и $\hat{\Theta}$.

Зная оценку $\hat{\sigma}^2$ и учитывая то обстоятельство, что при гауссовском векторе ϵ оценка $\hat{\Theta} \sim N(\Theta, \sigma^2(\Psi^T \Psi)^{-1})$, можно построить доверительные интервалы для компонентов вектора $\hat{\Theta}$. Действительно, приближенно примем $\hat{\Theta} \sim N(\Theta, \hat{\sigma}^2(\Psi^T \Psi)^{-1})$, заменив истинное, но неизвестное значение дисперсии ее оценкой, и зададимся доверительной вероятностью $1 - \alpha$. Для i -го компонента $\hat{\Theta}_i$ вектора $\hat{\Theta}$ имеем $\hat{\Theta}_i \sim N(\Theta_i, \hat{\sigma}^2 \psi_{ii})$, где ψ_{ii} — i -й диагональный элемент матрицы $(\Psi^T \Psi)^{-1}$. Найдем такую величину h , при которой $P(\Theta_i - h \leq \hat{\Theta}_i < \Theta_i + h) = 1 - \alpha$, т.е.

$$\frac{1}{\sqrt{2\pi\hat{\sigma}_i^2}} \int_{\Theta_i-h}^{\Theta_i+h} \exp\left\{-\frac{1}{2\hat{\sigma}_i^2}(\hat{\Theta}_i - \Theta_i)^2\right\} d\hat{\Theta}_i = 1 - \alpha,$$

где $\hat{\sigma}_i^2 = \hat{\sigma}^2 \psi_{ii}$. После замены переменной $\hat{\sigma}_i^{-1}(\hat{\Theta}_i - \Theta_i) = z_i$ это равенство преобразуется к виду

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-b} \exp\left\{-\frac{1}{2}z_i^2\right\} dz_i = \frac{\alpha}{2}, \quad b = h\hat{\sigma}_i^{-1} \Rightarrow h = -\hat{\sigma}_i u_{\alpha/2},$$

где $u_{\alpha/2}$ есть $\alpha/2$ -квантиль стандартного гауссовского распределения. Таким образом, с вероятностью $1 - \alpha$ выполняется неравенство $\Theta_i + \hat{\sigma}_i u_{\alpha/2} \leq \hat{\Theta}_i < \Theta_i - \hat{\sigma}_i u_{\alpha/2}$, из которого следует $\hat{\Theta}_i + \hat{\sigma}_i u_{\alpha/2} < \Theta_i \leq \hat{\Theta}_i - \hat{\sigma}_i u_{\alpha/2}$, $i = 0, 1, 2, \dots, m$. Это и будут доверительные интервалы для регрессионных параметров. Далее эти неравенства будут уточнены.

2.3.6. Проблема обусловленности МНК-оценок. Векторные и матричные нормы

Выражение (2.24) для вычисления МНК-оценки внешне подкупает своей простотой и математической изящностью. Однако присутствие в нем операции обращения матрицы представляет собой тот айсберг, встреча с которым может привести к непредвиденным последствиям, если к этой встрече не подготовиться предварительно. Природа опасения заключается в следующем.

Матрица $(\Psi^T \Psi)^{-1}$ в (2.24) преобразует вектор $\Psi^T y$ в оценку $\hat{\Theta}(y)$. В практических задачах вектор y отображает результаты определенного эксперимента, которому сопутствуют неизбежные ошибки и неточности, обусловленные несовершенством измерительных технологий, методологии организации эксперимента и др. Все это приводит к тому, что реально вместо вектора $y^* = \Psi \Theta$, объективно отражающего состояние исследуемого явления, придется оперировать вектором $y \neq y^*$, который после его подстановки в (2.24) приведет к величине $\hat{\Theta}(y) = (\Psi^T \Psi)^{-1} \Psi^T y \neq \hat{\Theta}(y^*) = (\Psi^T \Psi)^{-1} \Psi^T y^*$. Это неравенство порождает крайне существенный вопрос: если $\Psi^T(y^* - y) = \Delta y$, где Δy — некоторое отклонение, то к каким последствиям в смысле величины изменения $\delta \hat{\Theta} = \hat{\Theta}(y^*) - \hat{\Theta}(y)$ МНК-оценки отклонение Δy приведет? Естественным является желание иметь при малых в некотором смысле отклонениях Δy малые изменения $\delta \hat{\Theta}$. Однако это пожелание далеко не всегда достигается. Поэтому нужно выявить механизм, определяющий характер взаимодействия величин $\delta \hat{\Theta}$ и Δy , с тем чтобы, раскрыв этот механизм, суметь достичь желаемого результата. Сформулируем ряд основополагающих в данной проблеме положений.

Пусть $z \in R^m$, $y \in R^n$ и $z = \Phi(y)$ — решение некоторой задачи, найденное как функция вектора y . Тогда:

Определение 2.3. Решение $z = \Phi(y)$ называется *устойчивым*, если при $\forall \varepsilon > 0 \exists \delta(\varepsilon) > 0$ такое, что из неравенства $\|y_1 - y_2\| < \delta(\varepsilon)$ следует $\|z_1 - z_2\| < \varepsilon$, где y_1 и y_2 — произвольные элементы из R^n , $z_1 = \Phi(y_1)$, $z_2 = \Phi(y_2)$.

По существу, это определение означает, что решение устойчиво, если малым по норме изменениям δy аргумента y соответствуют малые по норме изменения δz в решении z , так что $\|\delta z\| \rightarrow 0$ при $\|\delta y\| \rightarrow 0$.

Определение 2.4. Задача поиска решения $z = \Phi(y)$ называется *корректно* поставленной на паре пространств R^n , R^m , если выполняются условия:

- 1) для $\forall y \in R^n$ существует решение $z \in R^m$;
- 2) решение определено однозначно, т.е. единственно;
- 3) решение устойчиво.

Задачи, не удовлетворяющие перечисленным требованиям, называют *некорректно* поставленными. В связи с поиском МНК-оценок наибольшее беспокойство вызывает третье условие, что и порождает, как уже отмечалось, необходимость найти своеобразный критерий устойчивости решения. Осуществим это намерение в связи с поиском решения системы алгебраических уравнений

$$y = Az, A \in R^{n \times n}, \quad (2.41)$$

подобной определяющей МНК-оценку системе (2.23). Для этого нам понадобятся некоторые дополнительные сведения о векторных и матричных нормах.

Определение 2.5. Пусть $y = [y_i] \in R^n$, т.е. $i = 1, 2, \dots, n$. Тогда число

$$\|y\| = (|y_1|^p + |y_2|^p + \dots + |y_n|^p)^{1/p}, p \geq 1, \quad (2.42)$$

называется *гельдеровой* нормой вектора y .

В зависимости от p получают ту или иную частную норму. Так, при $p = 1$ получают октаэдрическую [9], или l_1 -норму [26]; при $p = 2$ имеем уже широко нами использовавшуюся евклидову, или l_2 -норму; при $p = \infty$ получим кубическую, или l_∞ -норму $\|y\| = \max\{|y_i|, i = 1, 2, \dots, n\}$. В зависимости от характера решаемой задачи применяют ту или иную частную норму.

Определение 2.6. Пусть $A = [a_{ij}] \in R^{m \times n}$ и $\|y\| \in R^n$. Тогда число

$$\|A\| = \max_{y \in R^n, y \neq 0} \frac{\|Ay\|}{\|y\|} \quad (2.43)$$

называется *нормой матрицы* A , согласованной с нормой вектора y .

Помимо трех традиционных аксиом, которым удовлетворяет норма любого математического объекта из линейного пространства, матричная норма (2.43) обладает еще одним свойством: если $A \in R^{m \times n}$ и $B \in R^{n \times k}$, то

$$\|AB\| \leq \|A\| \times \|B\|. \quad (2.44)$$

В зависимости от способа определения векторной нормы получают определенную матричную норму. Так, если $\|y\|$ — l_1 -норма, то

$$\|A\| = \max_{1 \leq k \leq n} \sum_{i=1}^m |a_{ik}|$$

и называется максимальной *столбцовой* нормой. Если $\|y\|$ — l_∞ -норма, то

$$\|A\| = \max_{1 \leq i \leq m} \sum_{k=1}^n |a_{ik}|$$

и называется максимальной *строчной* нормой. Если $\|y\|$ — евклидова норма, то

$$\|A\| = \sqrt{\lambda_{\max}},$$

где $\lambda_{\max}$ — максимальное собственное число матрицы $A^T A$ и называется *спектральной* нормой. Используются и другие матричные нормы. Например,

$$\|A\| = \sqrt{\sum_{i=1}^m \sum_{k=1}^n a_{ik}^2} \quad \text{или при } m = n \quad \|A\| = n \max_{1 \leq i, k \leq n} |a_{ik}|.$$

Полезно иметь в виду $\|E\| \geq 1$, $\|A^{-1}\| \geq \|E\|/\|A\|$, где $E, A \in R^{n \times n}$.

Возвратимся теперь к системе (2.41). Пусть вместо точного значения y получено приближенное значение $y = y^* + \delta y$. Тогда вместо точного решения z системы получим приближенное

$z = z^* + \delta z$, причем отклонения связаны уравнением $\delta y = A \delta z$ или $\delta z = A^{-1} \delta y$. Переходя к нормам, получаем с учетом (2.44) $\|\delta z\| \leq \|A^{-1}\| \times \|\delta y\|$. Рассмотрим относительную ошибку решения $\|\delta z\|/\|z\|$. Из (2.44) следует $\|y\| \leq \|A\| \times \|z^*\|$. Перемножив два последних неравенства, получим

$$\begin{aligned} \|\delta z\| \times \|y^*\| &\leq \|\delta y\| \times \|A^{-1}\| \times \|A\| \times \|z^*\| \Rightarrow \\ \Rightarrow \|\delta z\|/\|z^*\| &\leq \|A^{-1}\| \times \|A\| \times \|\delta y\|/\|y^*\|. \end{aligned} \quad (2.45)$$

Отсюда следует, что относительная ошибка в решении, вызванная неточным заданием входных данных y^* , тем больше, чем больше число $\|A^{-1}\| \times \|A\|$, которое может служить своеобразным индикатором устойчивости решения.

Определение 2.7. Число $D_A = \|A^{-1}\| \times \|A\|$ для невырожденной матрицы A и $D_A = \infty$ для вырожденной называется *числом обусловленности* матрицы A .

Некоторые характерные свойства числа обусловленности [19]: $D_A \geq 1$; $D_A \geq |\lambda_{\max}|/|\lambda_{\min}|$, где $|\lambda_{\max}|$ и $|\lambda_{\min}|$ — наибольшее и наименьшее по модулю собственные числа матрицы A ; $D_{AB} \leq D_A D_B$ для произведения AB матриц. Матрицы с большим числом обусловленности называются *плохо обусловленными* в противовес *хорошо обусловленным* матрицам, которым соответствуют малые числа обусловленности.

Таким образом, чем больше число обусловленности матрицы, тем большая относительная ошибка решения порождается одной и той же относительной ошибкой задания входных данных y . При этом следует иметь в виду, что конкретное значение числа обусловленности зависит от способа задания нормы $\|A\|$ матрицы A . Заметим, что современные пакеты прикладных программ предусматривают вычисление чисел обусловленности матриц при различных заданиях их норм.

Выявленные особенности системы (2.41) свойственны, разумеется, и системе (2.23), используемой для поиска МНК-оценки (2.24). Если матрица $\Psi^T \Psi$ оказывается плохо обусловленной, использовать выражение (2.24) как средство практического вычисления МНК-оценки рискованно, так как даже малые ошибки в задании вектора $\Psi^T y$ могут привести к большим отклонениям решения, т.е. к большим ошибкам МНК-оценки. Несложно получить аналогичное (2.45) неравенство для уравнения (2.23). Пусть $y^* = \Psi^T \Theta$ — гипотетический вектор экспериментальных данных,

соответствующих идеальному случаю безошибочных наблюдений, и $y = \Psi^T \Theta + \varepsilon$ — результат реальных наблюдений. Тогда по аналогии с (2.45) получаем

$$\|\hat{\Theta} - \Theta\|/\|\Theta\| \leq \|\Psi^T \Psi\| \times \|(\Psi^T \Psi)^{-1}\| \times \|\Psi^T \varepsilon\|/\|\Psi^T y^*\|.$$

Отсюда следует, что при плохо обусловленной матрице $\Psi^T \Psi$ нужны иные, нежели (2.24), принципы поиска решения системы (2.23). Эти принципы и при плохо обусловленной матрице $\Psi^T \Psi$ должны гарантировать приемлемо малые ошибки оценивания. Успехи вычислительной математики в этом направлении в последнее время связаны с разработкой так называемых *методов регуляризации*, объединенных в весьма разветвленную математическую теорию решения некорректных задач (например, [24]). Далее, не вдаваясь в терминологические и принципиальные особенности методов регуляризации, рассмотрим один из алгоритмов, позволяющих обеспечить устойчивость решения при вычислении МНК-оценок и организовать достаточно рациональную вычислительную схему поиска этих оценок.

2.3.7. Рекуррентный метод наименьших квадратов

Предположим, что поиск МНК-оценки осуществляется не по всему массиву $y_1, y_2, \dots, y_n$ экспериментальных данных, а лишь по части $y_1, y_2, \dots, y_k, k \leq n$. Это означает, что целевая функция (2.20) и соответствующая оптимизационная задача приобретают вид

$$J_k = \sum_{j=1}^k [y_j - \Psi^T(x_j) \Theta]^2 \rightarrow \min \text{ по } \Theta. \quad (2.46)$$

Техника решения задачи (2.46) уже известна, и ее «материализация» может быть представлена в форме (2.25)

$$\hat{\Theta}^{(k)} = \left(\sum_{i=1}^k \Psi(x_i) \Psi^T(x_i) \right)^{-1} \sum_{i=1}^k \Psi(x_i) y_i. \quad (2.47)$$

Здесь индекс y оценки указывает на объем экспериментальных данных, используемых при ее вычислении. Выражение (2.47) введено чисто формально в том отношении, что проблема обра-

щаемости матрицы в (2.47) не обсуждается, так как в последующем она легко снимается.

Утверждение 2.9. Пусть матрицы P, Q, R таковы, что образуют невырожденную матрицу $P^{-1} + Q^T R^{-1} Q$. Тогда справедливо представление

$$(P^{-1} + Q^T R^{-1} Q)^{-1} = P - P Q^T (Q P Q^T + R)^{-1} Q P. \quad (2.48)$$

Доказательство этой так называемой *леммы об обращении матрицы* можно найти, например, в [20], и мы не будем на нем останавливаться.

Введем обозначение

$$\begin{aligned} \mathfrak{R}_k &= \left(\sum_{i=1}^k \psi(x_i) \psi^T(x_i) \right)^{-1} = \left(\sum_{i=1}^{k-1} \psi(x_i) \psi^T(x_i) + \psi(x_k) \psi^T(x_k) \right)^{-1} = \\ &= (\mathfrak{R}_{k-1}^{-1} + \psi(x_k) \psi^T(x_k))^{-1}. \end{aligned}$$

Воспользовавшись леммой об обращении матрицы, представим

$$\mathfrak{R}_k = \mathfrak{R}_{k-1} - \mathfrak{R}_{k-1} \psi(x_k) (\psi^T(x_k) \mathfrak{R}_{k-1} \psi(x_k) + 1)^{-1} \psi^T(x_k) \mathfrak{R}_{k-1}. \quad (2.49)$$

Это позволяет следующим образом преобразовать выражение (2.47):

$$\begin{aligned} \bar{\Theta}^{(k)} &= \mathfrak{R}_k \left(\sum_{i=1}^{k-1} \psi(x_i) y_i + \psi(x_k) y_k \right) = \\ &= (\mathfrak{R}_{k-1} - q_{k-1} \mathfrak{R}_{k-1}) \sum_{i=1}^{k-1} \psi(x_i) y_i + \mathfrak{R}_k \psi(x_k) y_k, \end{aligned}$$

где обозначено $q_{k-1} = \mathfrak{R}_{k-1} \psi(x_k) (\psi^T(x_k) \mathfrak{R}_{k-1} \psi(x_k) + 1)^{-1} \psi^T(x_k)$. По аналогии с (2.47)

$$\mathfrak{R}_{k-1} \sum_{i=1}^{k-1} \psi(x_i) y_i = \bar{\Theta}^{(k-1)}$$

и, следовательно,

$$\bar{\Theta}^{(k)} = \bar{\Theta}^{(k-1)} - q_{k-1} \bar{\Theta}^{(k-1)} + \mathfrak{R}_k \psi(x_k) y_k.$$

В соответствии с введенными обозначениями имеем:

$$\begin{aligned} q_{k-1} \mathfrak{R}_{k-1} &= \mathfrak{R}_{k-1} - \mathfrak{R}_k \Rightarrow q_{k-1} = E - \mathfrak{R}_k \mathfrak{R}_{k-1}^{-1}; \\ \mathfrak{R}_{k-1}^{-1} &= \mathfrak{R}_k^{-1} - \psi(x_k) \psi^T(x_k) \Rightarrow q_{k-1} = \mathfrak{R}_k \psi(x_k) \psi^T(x_k). \end{aligned}$$

Это позволяет окончательно записать

$$\bar{\Theta}^{(k)} = \bar{\Theta}^{(k-1)} + \mathfrak{R}_k \psi(x_k) (y_k - \psi^T(x_k) \bar{\Theta}^{(k-1)}), \quad k = 1, 2, \dots, n. \quad (2.50)$$

Совокупность выражений (2.49), (2.50) обычно и принято называть *рекуррентным методом наименьших квадратов* (РМНК). Вычисления по этому методу организуются последовательно. Вначале полагают $k = 1$; задают начальные условия, наиболее часто в виде $\bar{\Theta}^{(0)} = 0, \mathfrak{R}_0 = \gamma E$, где $\gamma = \text{const} \gg 1$; из (2.49) находят $\mathfrak{R}_1$, а из (2.50) — $\bar{\Theta}^{(1)}$, что формально соответствует поиску оценки $\bar{\Theta}^{(1)}$ регрессионных параметров по единственному экспериментальному результату y_1 . Разумеется, никакого серьезного внимания к оценке $\bar{\Theta}^{(1)}$ проявлять нельзя, она лишена какого-либо практического смысла и должна рассматриваться как формальный «эпизод» на пути получения МНК-оценки. Далее принимают $k = 2$, из (2.49) находят $\mathfrak{R}_2$, а из (2.50) — $\bar{\Theta}^{(2)}$, что соответствует уточненной по второму экспериментальному наблюдению y_2 оценке. Далее аналогичным образом проводятся вычисления при $k = 3, 4, \dots, n$, что приводит к оценке $\bar{\Theta}^{(n)}$, принимаемой за МНК-оценку.

Как видно, при такой схеме вычислений не приходится обращаться плохо обусловленную матрицу, что стимулирует получение устойчивого решения. Чтобы более явно раскрыть механизм этого явления, поступим следующим образом. С использованием матрицы $\mathfrak{R}_k$ МНК-оценку (2.25) можем представить в следующем виде:

$$\bar{\Theta} = \left(\sum_{i=1}^n \psi(x_i) \psi^T(x_i) \right)^{-1} \sum_{i=1}^n \psi(x_i) y_i = \mathfrak{R}_n \sum_{i=1}^n \psi(x_i) y_i. \quad (2.51)$$

Матрица $\mathfrak{R}_n$ вычисляется последовательно. Будем иметь:

$$\begin{aligned} \mathfrak{R}_1 &= (\mathfrak{R}_0^{-1} + \psi(x_1) \psi^T(x_1))^{-1}; \quad \mathfrak{R}_2 = (\mathfrak{R}_1^{-1} + \psi(x_2) \psi^T(x_2))^{-1} = \\ &= (\mathfrak{R}_0^{-1} + \psi(x_1) \psi^T(x_1) + \psi(x_2) \psi^T(x_2))^{-1}; \dots; \\ \mathfrak{R}_n &= \left(\mathfrak{R}_0^{-1} + \sum_{i=1}^n \psi(x_i) \psi^T(x_i) \right)^{-1}. \end{aligned}$$

Если учесть результат последовательного вычисления матрицы $\mathfrak{R}_n$, выражение (2.51) перепишется так:

$$\hat{\Theta} = \left(\mathfrak{R}_0^{-1} + \sum_{i=1}^n \psi(x_i) \psi^T(x_i) \right)^{-1} \sum_{i=1}^n \psi(x_i) y_i =$$

$$= \left(\gamma^{-1} E + \sum_{i=1}^n \psi(x_i) \psi^T(x_i) \right)^{-1} \sum_{i=1}^n \psi(x_i) y_i. \quad (2.52)$$

Как видим, оценки (2.51), (2.52) различаются: если в (2.51) матрица $\Psi^T \Psi = \sum_{i=1}^n \psi(x_i) \psi^T(x_i)$ плохо обусловлена, то ее аналог

$\gamma^{-1} E + \sum_{i=1}^n \psi(x_i) \psi^T(x_i)$ в (2.52) благодаря начальному условию $\gamma^{-1} E$ может оказаться хорошо обусловленной матрицей, что и обеспечивает устойчивость решения.

Чтобы получить представление об эффективности РМНК, приведем результаты одного вычислительного эксперимента. Его существо заключается в следующем. Были заданы конкретные матрица Ψ размерностью $n = 10$ на $m + 1 = 3$ и вектор Θ . Матрица Ψ строилась случайным образом, но так, что один из ее столбцов отличался от соседнего только одним элементом, изменение которого позволяло варьировать число обусловленности матрицы $\Psi^T \Psi$. Далее решалась прямая задача, в результате чего находили вектор $y^* = \Psi \Theta$. Этот вектор искажался случайным «шумом» ϵ с независимыми и равномерно распределенными на $[-c, c]$ компонентами, что приводило к формированию вектора $y = y^* + \epsilon$. Построенный таким образом вектор y использовался при решении обратной задачи, т.е. при поиске оценки $\hat{\Theta}$ вектора Θ в соответствии с регулярным (2.24) и рекуррентным (2.49), (2.50) методами наименьших квадратов. Результаты вычислений содержатся в табл. 2.1.

В таблице использованы следующие обозначения: $D = \|\Psi^T \Psi\|$ $\|(\Psi^T \Psi)^{-1}\|$ — число обусловленности матрицы $\Psi^T \Psi$; $\frac{\|\Psi^T \epsilon\|}{\|\Psi^T y^*\|}$ — относительная ошибка при формировании свободных членов

Таблица 2.1

Точностные характеристики МНК и РМНК

D	$\frac{\ \Psi^T \epsilon\ }{\ \Psi^T y^*\ }$	$\frac{\ \Theta - \hat{\Theta}\ }{\ \Theta\ }$ для МНК	$\frac{\ \Theta - \hat{\Theta}\ }{\ \Theta\ }$ для РМНК	$\text{Diag}(\Psi^T \Psi)^{-1}$
∞	$4 \cdot 10^{-3}$	—	0,258	—
$5,56 \cdot 10^{12}$	$3,8 \cdot 10^{-3}$	$6,93 \cdot 10^3$	0,257	$8 \cdot 10^{-3}$ $1,1 \cdot 10^{10}$ $1,1 \cdot 10^{10}$
$5,56 \cdot 10^{10}$	$3,9 \cdot 10^{-3}$	704,3	0,23	$8 \cdot 10^{-3}$ $1,1 \cdot 10^8$ $1,1 \cdot 10^8$
$5,56 \cdot 10^8$	$3,3 \cdot 10^{-3}$	43,79	0,342	$8 \cdot 10^{-3}$ $1,1 \cdot 10^6$ $1,1 \cdot 10^6$
$5,56 \cdot 10^6$	$5,1 \cdot 10^{-3}$	7,12	3,18	$8 \cdot 10^{-3}$ $1,1 \cdot 10^4$ $1,1 \cdot 10^4$
$2,5 \cdot 10^4$	$4 \cdot 10^{-3}$	0,507	0,506	$8 \cdot 10^{-3}$ 49,9 49,1
$2,3 \cdot 10^3$	$3,37 \cdot 10^{-3}$	0,187	0,187	$8 \cdot 10^{-3}$ 4,69 4,42
595	$3,8 \cdot 10^{-3}$	0,066	0,066	$8 \cdot 10^{-3}$ 1,26 1,1
36	$3 \cdot 10^{-3}$	0,029	0,029	$8 \cdot 10^{-3}$ 0,102 0,044
19,68	$3,2 \cdot 10^{-3}$	$9,8 \cdot 10^{-3}$	$9,8 \cdot 10^{-3}$	$8 \cdot 10^{-3}$ 0,057 0,011

системы уравнений (2.23); $\frac{\|\hat{\theta} - \theta\|}{\|\theta\|}$ для МНК — относительная

ошибка МНК-оценки (2.24); $\frac{\|\hat{\theta} - \theta\|}{\|\theta\|}$ для РМНК — относитель-

ная ошибка оценки, найденной рекуррентным методом наименьших квадратов; $\text{diag}(\Psi^T \Psi)^{-1}$ — столбец диагональных элементов матрицы $(\Psi^T \Psi)^{-1}$, определяющей точность МНК-оценки. При проведении вычислений применялись евклидова векторная и спектральная матричная нормы. Содержащиеся в табл. 2.1 данные получены усреднением по десяти реализациям вектора ϵ , ковариационная матрица которого в эксперименте принималась равной $(2/3) \times 10^{-1} E$. Истинное значение вектора θ формировалось случайным образом и оказалось равным $\theta^T = [4,063 \ 0,053 \ 1,655]$. Первая строка в табл. 2.1 соответствует вырожденной матрице $\Psi^T \Psi$. Из таблицы видно, что при плохо обусловленных матрицах $\Psi^T \Psi$ рекуррентный метод, в отличие от регулярного, ведет себя вполне «пристойно».

По мере уменьшения числа обусловленности матрицы $\Psi^T \Psi$ точностные характеристики методов сближаются. Наблюдаемое ухудшение точности рекуррентного метода в «среднем» диапазоне чисел обусловленности можно объяснить тем, что все исследования проводились при одном и том же значении параметра γ , определяющего начальную матрицу $\mathcal{R}_0 = \gamma E$ в структуре рекуррентного алгоритма (было принято $\gamma = 10^4$). Вместе с тем исследования свидетельствуют об определенной чувствительности алгоритма к значению этого параметра, которое полезно адаптировать к числу обусловленности матрицы $\Psi^T \Psi$. Так, если при $D = 5,56 \times 10^6$ положить $\gamma = 10^3$, то относительная ошибка рекуррентно найденной оценки оказывается равной 0,723 (вместо 3,18 при $\gamma = 10^4$). Выбор оптимального значения параметра γ здесь не обсуждается.

2.3.8. Коэффициент детерминации.

Послерегрессионный анализ регрессионной модели

Коэффициент детерминации. Рассмотренные выше методы регрессионного анализа предполагали, что регрессионная модель (2.2) (число и характер экзогенных переменных, структура и размерность вектор-функции $\Psi(x)$ и т.п.) выбрана и обоснована

средствами предварительного анализа регрессионной модели (см. п.1.5). После проведения регрессионного анализа исследователь располагает большей информацией о свойствах модели, так как для регрессионных параметров модели и дисперсии экспериментальных ошибок получены МНК-оценки и установлены их основные свойства. Это позволяет с позиций новых знаний еще раз возвратиться к модели (2.2), подвергнуть ее дополнительному анализу и при необходимости подкорректировать.

В эконометрических задачах существенная часть послерегрессионного анализа связана с применением так называемого коэффициента детерминации. Чтобы его определить (здесь возможны различные подходы), рассмотрим вариацию (разброс) $\text{var}(y)$ экспериментальных данных (2.4) относительно их среднего значения, которую по определению примем равной

$$\text{var}(y) = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n} \|y - \bar{y}s\|^2 = \frac{1}{n} (y - \bar{y}s)^T (y - \bar{y}s),$$

где $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, $s = [1 \ 1 \ \dots \ 1]^T \in \mathbb{R}^n$.

Введем обозначения: $\hat{f} = \Psi \hat{\theta}$ — оцененные значения регрессионных составляющих в наблюдениях y , $\hat{\epsilon} = y - \hat{f}$ есть оценки экспериментальных ошибок, и представим

$$\begin{aligned} \text{var}(y) &= \frac{1}{n} (y - \hat{f} + \hat{f} - \bar{y}s)^T (y - \hat{f} + \hat{f} - \bar{y}s) = \\ &= \frac{1}{n} (y - \hat{f})^T (y - \hat{f}) + \frac{1}{n} (\hat{f} - \bar{y}s)^T (\hat{f} - \bar{y}s) + \frac{2}{n} (y - \hat{f})^T (\hat{f} - \bar{y}s) = \\ &= \frac{1}{n} \hat{\epsilon}^T \hat{\epsilon} + \text{var}(\hat{f}) + \frac{2}{n} \hat{\epsilon}^T (\hat{f} - \bar{y}s), \end{aligned} \quad (2.53)$$

где величина $\text{var}(\hat{f}) = \frac{1}{n} (\hat{f} - \bar{y}s)^T (\hat{f} - \bar{y}s)$ характеризует разброс регрессионных составляющих наблюдений относительно среднего значения наблюдений. Покажем, что последнее слагаемое в (2.53) равно нулю. Действительно, имеем:

$$\begin{aligned} \hat{\epsilon}^T \hat{f} &= \hat{\epsilon}^T \Psi \hat{\theta} = \hat{\theta}^T \Psi^T \hat{\epsilon}, \\ \hat{\epsilon} &= y - \Psi \hat{\theta} = \Psi \theta + \epsilon - \Psi (\Psi^T \Psi)^{-1} \Psi^T (\Psi \theta + \epsilon) = (E - \Psi (\Psi^T \Psi)^{-1} \Psi^T) \epsilon. \end{aligned}$$

Следовательно,

$$\Psi^T \hat{\epsilon} = \Psi^T (E - \Psi(\Psi^T \Psi)^{-1} \Psi^T) \epsilon = 0_{m+1} \Rightarrow \quad (2.54)$$

$$\Rightarrow \hat{\epsilon}^T \hat{f} = 0. \quad (2.55)$$

Рассмотрим структуру матрицы Ψ^T , вытекающую из ее определения:

$$\begin{aligned} \Psi^T &= [\psi(x_1) \psi(x_2) \psi(x_3) \dots \psi(x_n)] = \\ &= \begin{bmatrix} \psi_0(x_1) & \psi_0(x_2) & \dots & \psi_0(x_n) \\ \psi_1(x_1) & \psi_1(x_2) & \dots & \psi_1(x_n) \\ \dots & \dots & \dots & \dots \\ \psi_m(x_1) & \psi_m(x_2) & \dots & \psi_m(x_n) \end{bmatrix}, \end{aligned} \quad (2.56)$$

где $\psi_0(x)$, $\psi_1(x)$, ..., $\psi_m(x)$ — компоненты вектора $\psi(x)$. Тогда из равенства (2.54) следует

$$\sum_{i=1}^n \Psi_k(x_i) \hat{\epsilon}_i = 0, \quad \forall k = 0, 1, 2, \dots, m, \quad (2.57)$$

где $\hat{\epsilon}_i$ — i -й компонент вектора $\hat{\epsilon}$.

Регрессионные модели обычно конструируют таким образом, что компонент Θ_0 вектора регрессионных параметров Θ имитирует «постоянную составляющую» функции регрессии, не зависящую от экзогенных переменных. Для этого обычно полагают $\psi_0(x) = 1$. Но тогда из (2.57) при $k = 0$ следует

$$\sum_{i=1}^n \hat{\epsilon}_i = \hat{\epsilon}^T \mathbf{1} = 0. \quad (2.58)$$

Совместное выполнение условий (2.55), (2.58) приводит к тому, что $\hat{\epsilon}^T (\hat{f} - \bar{y} \mathbf{1}) = 0$, и выражение (2.53) упрощается:

$$\text{var}(y) = \text{var}(\hat{f}) + \frac{1}{n} \hat{\epsilon}^T \hat{\epsilon}. \quad (2.59)$$

Определение 2.8. Коэффициент

$$K_d^2 = \frac{\text{var}(\hat{f})}{\text{var}(y)} = 1 - \frac{\frac{1}{n} \hat{\epsilon}^T \hat{\epsilon}}{\text{var}(y)} = 1 - \frac{\|\hat{\epsilon}\|^2}{\|y - \bar{y} \mathbf{1}\|^2} \quad (2.60)$$

называется коэффициентом детерминации.

Содержательно этот коэффициент определяет, какая часть вариации экспериментальных данных объясняется разбросом регрессионной (детерминированной) составляющей этих данных. Из определения вытекает $K_d^2 \in [0, 1]$. Если $K_d^2 = 0$, то $\text{var}(\hat{f}) = 0$, т.е. $\hat{f} = \Psi \hat{\Theta} = \bar{y} \mathbf{1} \Rightarrow \Psi^T(x_i) \hat{\Theta} = \bar{y}$, $i = 1, 2, \dots, n$. А это можно интерпретировать как независимость эндогенной переменной от экзогенных переменных. Если же $K_d^2 = 1$, то $\hat{\epsilon} = 0_n$, т.е. регрессионная поверхность проходит точно через все экспериментальные точки y_i ($i = 1, 2, \dots, n$) в смысле выполнения равенств $y_i = \Psi^T(x_i) \hat{\Theta}$ ($i = 1, 2, \dots, n$). Разумеется, если положить $m + 1 = n$ и матрицу $\Psi \in R^{n \times m}$ выбрать невырожденной, это условие будет достигнуто, так как МНК-оценка (2.24) в этом случае приобретает вид

$$\hat{\Theta} = (\Psi^T \Psi)^{-1} \Psi^T y = \Psi^{-1} (\Psi^T)^{-1} \Psi^T y = \Psi^{-1} y \Rightarrow \Psi \hat{\Theta} = y.$$

Однако это условие, вообще говоря, не является признаком хорошо подобранной регрессионной модели, так как такая модель будет «отслеживать» все случайные составляющие ϵ в результатах эксперимента y , а это недопустимо.

Проверка гипотезы $H_0: \Theta_1 = \Theta_2 = \dots = \Theta_m = 0$. Введенный коэффициент детерминации широко используется для подтверждения (или опровержения) предположения о том, что эндогенная переменная действительно зависит от выбранных экзогенных переменных. С этой целью в рассмотрение вводятся две гипотезы:

$$H_0: \Theta_1 = \Theta_2 = \dots = \Theta_m = 0;$$

$$H_1: \Theta_0, \Theta_1, \Theta_2, \dots, \Theta_m \neq 0.$$

В качестве «индикатора» правомочности одной из этих гипотез используется величина

$$\gamma = \frac{K_d^2}{1 - K_d^2} \frac{n - m - 1}{m},$$

которая с учетом определения коэффициента детерминации преобразуется к виду

$$\gamma = \frac{\|\Psi \hat{\Theta} - \bar{y} \mathbf{1}\|^2}{\|y - \Psi \hat{\Theta}\|^2} \frac{n - m - 1}{m}.$$

Для последующих доказательств это выражение целесообразно представить так:

$$\gamma = \frac{\frac{1}{m} \frac{\|\Psi\hat{\Theta} - \bar{y}s\|^2}{\sigma^2}}{\frac{1}{n-m-1} \frac{\|y - \Psi\hat{\Theta}\|^2}{\sigma^2}} = \frac{\frac{1}{m} \frac{\|\Psi\hat{\Theta} - \bar{y}s\|^2}{\hat{\sigma}^2}}{\frac{\sigma^2}{\sigma^2}}. \quad (2.61)$$

Найдем распределение этой величины при справедливости гипотезы H_0 . Прежде всего обратим внимание на знаменатель. Если мысленно его умножить и разделить на число $n - m - 1$, то величина $(n - m - 1) \frac{\hat{\sigma}^2}{\sigma^2}$, как было уже доказано, будет подчинена χ^2 -распределению с $n - m - 1$ степенями свободы. Следовательно, знаменатель в (2.61) можно представить как $\frac{v}{n - m - 1}$, где случайная величина $v \sim \chi^2(n - m - 1)$.

Рассмотрим теперь числитель в (2.61). С учетом модели наблюдений (2.4) и определения МНК-оценки (2.24) находим:

$$\bar{y}s = \frac{1}{n} ss^T y = \frac{1}{n} ss^T \Psi\Theta + \frac{1}{n} ss^T \epsilon,$$

$$\Psi\hat{\Theta} = \Psi(\Psi^T \Psi)^{-1} \Psi^T y = \Psi\Theta + \Psi(\Psi^T \Psi)^{-1} \Psi^T \epsilon$$

и, следовательно,

$$\Psi\hat{\Theta} - \bar{y}s = \left(E_n - \frac{1}{n} ss^T\right) \Psi\Theta + \left(\Psi(\Psi^T \Psi)^{-1} \Psi^T - \frac{1}{n} ss^T\right) \epsilon.$$

Если справедлива гипотеза H_0 , т.е. $\Theta_1 = \Theta_2 = \dots = \Theta_m = 0$, то $\Psi\Theta = s\Theta_0$, так как первый столбец матрицы Ψ состоит из единиц, т.е. представляет собой вектор s . Но тогда

$$\left(E_n - \frac{1}{n} ss^T\right) \Psi\Theta = \left(E_n - \frac{1}{n} ss^T\right) s\Theta_0 = \theta_n$$

и получаем упрощенное выражение

$$\Psi\hat{\Theta} - \bar{y}s = \left(\Psi(\Psi^T \Psi)^{-1} \Psi^T - \frac{1}{n} ss^T\right) \epsilon. \quad (2.62)$$

Докажем теперь важное для последующего вывода утверждение.

Утверждение 2.10. Пусть матрица $\Psi = [\psi_0 \psi_1 \dots \psi_m]$, где $\psi_i \in R^n$ — столбцы матрицы. Тогда справедливы равенства

$$\Psi(\Psi^T \Psi)^{-1} \Psi^T \psi_i = \psi_i, \quad i = 0, 1, \dots, m. \quad (2.63)$$

Векторы $\psi_0, \psi_1, \dots, \psi_m$ являются линейно независимыми, так как $\text{rang } \Psi = m + 1 < n$. Построим подмножество $G^{m+1} \subset R^n$, которое состоит из векторов z , допускающих разложение по векторам

$$\psi_0, \psi_1, \dots, \psi_m \text{ как по базису, т.е. } G^{m+1} = \left\{ z \in R^n : z = \sum_{i=0}^m \alpha_i \psi_i, \alpha_i \in R \right\}$$

и в пространстве R^n образует линейное $(m + 1)$ -мерное подпространство. Рассмотрим вектор $\Psi\Theta = \sum_{i=0}^m \Theta_i \psi_i$. Из этого разложения

следует $\Psi\Theta \in G^{m+1}$. Величина $J = \|y - \Psi\hat{\Theta}\|^2$ представляет собой квадрат расстояния между векторами y и $\Psi\hat{\Theta}$. Тогда величина $\min J$ определяет наименьшее расстояние от вектора y до подпространства G^{m+1} , т.е. расстояние до такого вектора z из G^{m+1} , который является ортогональной проекцией вектора y на множество G^{m+1} : $z^* = \text{Pr}_{G^{m+1}} y$. С другой стороны, $\min J = \|y - \Psi\hat{\Theta}\|^2$ и, следовательно, $\Psi\hat{\Theta} = \Psi(\Psi^T \Psi)^{-1} \Psi^T y = z^*$. Но отсюда вытекает, что матрица $\Psi(\Psi^T \Psi)^{-1} \Psi^T$ является матрицей ортогонального проектирования на подпространство G^{m+1} . Эта матрица любой вектор, принадлежащий множеству G^{m+1} , преобразует в себя, т.е. из условия $z \in G^{m+1}$ следует $\Psi(\Psi^T \Psi)^{-1} \Psi^T z = z$. Так как $\psi_0, \psi_1, \dots, \psi_m \in G^{m+1}$, то равенства (2.63) справедливы. Заметим, что непосредственно это равенство вытекает из тождества $\Psi(\Psi^T \Psi)^{-1} \Psi^T \Psi = \Psi$.

Следствие. Так как вектор s является столбцом матрицы Ψ , а именно $s = \psi_0$, то справедливы равенства

$$\Psi(\Psi^T \Psi)^{-1} \Psi^T s = s, \quad s^T \Psi(\Psi^T \Psi)^{-1} \Psi^T = s^T. \quad (2.64)$$

Утверждение 2.11. Матрица $\Psi(\Psi^T \Psi)^{-1} \Psi^T - \frac{1}{n} ss^T$ является идемпотентной и ее ранг равен m .

Доказательство проводится непосредственно:

$$\begin{aligned} & \left(\Psi(\Psi^T \Psi)^{-1} \Psi^T - \frac{1}{n} ss^T\right)^2 = \Psi(\Psi^T \Psi)^{-1} \Psi^T \Psi(\Psi^T \Psi)^{-1} \Psi^T + \frac{1}{2} ss^T ss^T - \\ & - \frac{1}{n} ss^T \Psi(\Psi^T \Psi)^{-1} \Psi^T - \frac{1}{n} \Psi(\Psi^T \Psi)^{-1} \Psi^T ss^T = \Psi(\Psi^T \Psi)^{-1} \Psi^T - \frac{1}{n} ss^T, \end{aligned}$$

где учтены условия (2.64) и очевидное $s^T s = n$; аналогично

$$\begin{aligned} \text{rank}(\Psi(\Psi^T \Psi)^{-1} \Psi^T - \frac{1}{n} s s^T) &= \text{Sp}(\Psi(\Psi^T \Psi)^{-1} \Psi^T - \frac{1}{n} s s^T) = \\ &= \text{Sp} \Psi^T \Psi (\Psi^T \Psi)^{-1} - \text{Sp} \frac{1}{n} s s^T = m. \end{aligned}$$

Воспользовавшись выражением (2.62) и утверждением 2.11, теперь можем числитель в (2.61) представить в виде

$$\|\Psi \hat{\Theta} - \bar{y} s\|^2 / \sigma^2 = \frac{\varepsilon^T}{\sigma} \left(\Psi(\Psi^T \Psi)^{-1} \Psi^T - \frac{1}{n} s s^T \right) \frac{\varepsilon}{\sigma},$$

где ε/σ — стандартный гауссовский вектор.

Так как матрица $\Psi(\Psi^T \Psi)^{-1} \Psi^T - \frac{1}{n} s s^T$ является идемпотентной, то в соответствии с утверждением 2.5 эта величина подчинена $\chi^2(m)$ -распределению. Таким образом, выражение (2.61) может быть представлено в эквивалентной форме

$$\gamma = \frac{\frac{1}{m} w}{\frac{1}{n-m-1} v},$$

где $w \sim \chi^2(m)$, $v \sim \chi^2(n-m-1)$ — случайные величины.

Ранее при доказательстве утверждения 2.8 было показано, что при гауссовском векторе ε вектор $y - \Psi \hat{\Theta}$ и оценка $\hat{\sigma}^2$ независимы. Но тогда независимы вектор $\Psi \hat{\Theta} - \bar{y} s$ и та же оценка $\hat{\sigma}^2$, т.е. числитель и знаменатель в (2.61). Последнее порождает независимость случайных величин w и v , а это означает, что случайная величина γ при справедливости гипотезы H_0 подчинена распределению Фишера с m степенями свободы числителя и $n-m-1$ степенями свободы знаменателя, т.е. $\gamma \sim F(m, n-m-1)$, что позволяет сформулировать следующий критерий проверки гипотезы H_0 (или альтернативы H_1).

На основе экспериментальных данных построим величину

$$\gamma = \frac{\|\Psi \hat{\Theta} - \bar{y} s\|^2}{\hat{\sigma}^2} \frac{1}{m}, \quad (2.65)$$

где $\hat{\Theta}$ — МНК-оценка (2.24) регрессионных параметров, $\hat{\sigma}^2$ — оценка (2.37) дисперсии экспериментальных ошибок.

Зададимся доверительной вероятностью q и по таблицам или с помощью вычислительных средств найдем q -квантиль u_q распределения Фишера с $m, n-m-1$ степенями свободы. Тогда если окажется $\gamma < u_q$, то с вероятностью q считается справедливой гипотеза H_0 ; если же $\gamma > u_q$, то с вероятностью $1-q$ ошибиться предпочтение отдается альтернативе H_1 .

Проверка гипотезы $H_0: \Theta_i = \Theta_i^0$ ($i = 0, 1, \dots, m$).

Пусть в соответствии с какими-либо априорными соображениями появились основания предполагать, что параметр Θ_i регрессионной модели принимает гипотетическое значение Θ_i^0 . Требуется на основе имеющихся экспериментальных данных подтвердить или опровергнуть это предположение. С этой целью в рассмотрение вводятся две гипотезы:

$$H_0: \Theta_i = \Theta_i^0; H_1: \Theta_i \neq \Theta_i^0.$$

Анализ ситуации проводится следующим образом. В случае гауссовской регрессионной модели, как было показано выше, $\hat{\Theta} \sim N(\Theta, \sigma^2(\Psi^T \Psi)^{-1})$ и, следовательно, при справедливости гипотезы H_0 $\hat{\Theta}_i - \Theta_i^0 \sim N(0, \sigma_i^2)$, $\sigma_i^2 = \sigma^2 \psi_{ii}$, где ψ_{ii} — ii -й элемент матрицы $(\Psi^T \Psi)^{-1}$. Тогда случайная величина $(\hat{\Theta}_i - \Theta_i^0)/\sigma_i \sim N(0, 1)$, т.е. представляет собой стандартную гауссовскую случайную величину. Построим случайную величину

$$\gamma = \frac{\frac{\hat{\Theta}_i - \Theta_i^0}{\sigma_i}}{\sqrt{\frac{(n-m-1)\hat{\sigma}^2}{\sigma^2} \frac{1}{n-m-1}}},$$

где $\hat{\sigma}^2 = \|y - \Psi \hat{\Theta}\|^2 / (n-m-1)$.

Так как случайные величины, формирующие числитель и знаменатель этого выражения, статистически независимы и

$\frac{(n-m-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-m-1)$, то в соответствии с определением

$\gamma - t(n - m - 1)$. Если учесть, что $\sigma_i^2 = \sigma^2 \psi_{ii}$ и $\hat{\sigma}_i^2 = \hat{\sigma}^2 \psi_{ii}$, то оказывается удобным представить

$$\gamma = \frac{\hat{\theta}_i - \theta_i^0}{\hat{\sigma}_i} \sim t(n - m - 1), \quad (2.66)$$

что справедливо, таким образом, при выполнении гипотезы H_0 . Принятие последующего решения осуществляется по уже известной схеме (см. вывод критерия (1.28)). В соответствии с результатами регрессионного анализа находится величина (2.66). Далее при выбранной доверительной вероятности $1 - \alpha$ по таблицам для распределения Стюдента с $n - m - 1$ степенями свободы или машинным образом находится $100\alpha/2$ -процентная точка $w_{100\alpha/2}$. Если окажется $|\gamma| < w_{100\alpha/2}$, то с вероятностью $1 - \alpha$ считается справедливой гипотеза H_0 ; при противоположном неравенстве, т.е. в случае $|\gamma| > w_{100\alpha/2}$, гипотеза H_0 отвергается как не согласующаяся с экспериментальными данными с вероятностью α ошибки первого рода.

Для линейных регрессионных моделей, построенных с использованием множества (1.18), величина (2.66) позволяет дополнительно решить вопрос о целесообразности включения i -й экзогенной переменной в состав величин, влияющих на эндогенную переменную. Отсутствие такого влияния можно интерпретировать как равенство $\theta_i^0 = 0$. Тогда, если окажется $|\hat{\theta}_i| < \hat{\sigma}_i w_{100\alpha/2}$, то с вероятностью $1 - \alpha$ целесообразно считать, что i -я экзогенная переменная не оказывает влияния на эндогенную переменную; при противоположном неравенстве гипотезу $H_0: \theta_i = 0$ следует отвергнуть с вероятностью α ошибиться.

Выражение (2.66) позволяет построить доверительные интервалы для параметров регрессионной модели. Действительно, с вероятностью $1 - \alpha$ должно выполняться

$$u_{\alpha/2} \leq \frac{\hat{\theta}_i - \theta_i}{\hat{\sigma}_i} < -u_{\alpha/2} \Rightarrow \hat{\theta}_i + \hat{\sigma}_i u_{\alpha/2} < \theta_i \leq \hat{\theta}_i - \hat{\sigma}_i u_{\alpha/2}, \quad i = 0, 1, \dots, m.$$

Здесь $u_{\alpha/2}$ есть $\alpha/2$ -квантиль распределения Стюдента с $n - m - 1$ степенями свободы. Если учесть уже доказанное для данного случая равенство $w_{100\alpha/2} = -u_{\alpha/2}$, то найденный интервал запишется так: $\hat{\theta}_i - \hat{\sigma}_i w_{100\alpha/2} < \theta_i \leq \hat{\theta}_i + \hat{\sigma}_i w_{100\alpha/2}$.

2.3.9. Нелинейные регрессионные модели. Проблема стохастической сходимости

Характерная особенность рассмотренных выше регрессионных моделей проявляется в их линейной зависимости от параметров $\theta_0, \theta_1, \dots, \theta_m$. Вместе с тем не исключается перспектива применения моделей с нелинейной зависимостью от регрессионных параметров, подобно тому, как это было в случае производственной функции Кобба—Дугласа. Так, если множество потенциальных функций регрессии задается соотношением (1.20), то вектор экспериментальных данных y , определенный подобным (2.4) образом, будет представлен выражением

$$y = \Psi(\theta) + \epsilon, \quad y \in R^n, \quad \theta \in R^{m+1}, \quad (2.67)$$

в котором вектор-функция $\Psi(\theta)$ определена естественным для (1.20) образом:

$$\Psi(\theta) = \sum_{k=1}^q \begin{bmatrix} \psi_k(x_1, \theta) \\ \psi_k(x_2, \theta) \\ \dots \\ \psi_k(x_n, \theta) \end{bmatrix}$$

Как и ранее, МНК-оценка $\hat{\theta}$ вектора параметров θ ищется в соответствии с условием

$$J = \|y - \Psi(\theta)\|^2 \rightarrow \min_{\theta}, \quad (2.68)$$

причем стационарные точки $\hat{\theta}$ функции J удовлетворяют традиционному условию $\nabla J(\hat{\theta}) = 0_{m+1}$. Однако из-за нелинейного включения вектора θ в состав минимизируемой функции J найти оценку $\hat{\theta}$ как явную функцию $\hat{\theta}(y)$ экспериментальных данных, в отличие от (2.24), не удастся. Поэтому применяют различные численные алгоритмы нелинейного оценивания [25]. Несмотря на многообразие подобных методов, можно выделить некоторые общие принципы, в соответствии с которыми конструируется большинство из них. Детальный анализ этих принципов не является целью настоящего пособия. Однако некоторые справочные сведения, касающиеся процедуры решения задачи (2.68), привести полезно.

Определение 2.9. Говорят, что вектор $h \in R^{m+1}$ задает направление убывания функции $J(\Theta)$ в точке Θ , если при всех малых $\beta > 0$ выполняется условие

$$J(\Theta + \beta h) < J(\Theta).$$

Утверждение 2.12. Для того чтобы вектор h задавал направление убывания функции $J(\Theta)$ в точке Θ , необходимо и достаточно, чтобы в этой точке выполнялось условие $(h, \nabla J(\Theta)) < 0$, т.е. скалярное произведение вектора h и вычисленного в точке Θ градиента функции $J(\Theta)$ должно быть отрицательным.

Доказательство этого утверждения, несмотря на его простоту, мы опускаем, так как оно несущественно для последующего изложения.

Определение 2.10. Последовательность точек $\bar{\Theta}^{(0)}, \bar{\Theta}^{(1)}, \bar{\Theta}^{(2)}, \dots$ в пространстве R^{m+1} будем называть *минимизирующей* функцию $J(\Theta)$, если выполняются условия $J(\bar{\Theta}^{(0)}) \geq J(\bar{\Theta}^{(1)}) \geq J(\bar{\Theta}^{(2)}) \geq \dots$.

Определение 2.11. Последовательность точек $\bar{\Theta}^{(0)}, \bar{\Theta}^{(1)}, \bar{\Theta}^{(2)}, \dots$ в пространстве R^{m+1} будем называть *сходящейся* к некоторой стационарной точке $\bar{\Theta} \in R^{m+1}$ функции $J(\Theta)$, если при $\forall \delta > 0 \exists N(\delta) > 0$, так что при $k \geq N$ выполняется условие $\|\bar{\Theta}^{(k)} - \bar{\Theta}\| < \delta$ или, что эквивалентно, $\lim_{k \rightarrow \infty} \|\bar{\Theta}^{(k)} - \bar{\Theta}\| = 0$.

Заметим, что в общем случае минимизирующая и сходящаяся последовательности могут не совпадать.

Основываясь на введенных понятиях, можно предложить правило вычисления последовательности точек $\bar{\Theta}^{(0)}, \bar{\Theta}^{(1)}, \bar{\Theta}^{(2)}, \dots$ в пространстве R^{m+1} , минимизирующей функцию $J(\Theta)$. В достаточно общем виде это правило таково:

$$\bar{\Theta}^{(k+1)} = \bar{\Theta}^{(k)} + \alpha_k h_k, \quad k = 0, 1, 2, \dots, \quad (2.69)$$

где вектор h_k задает направление убывания функции $J(\bar{\Theta})$ в точке $\bar{\Theta}^{(k)}$, а скалярная величина $\alpha_k > 0$ определяет величину шага перемещения. В зависимости от того, как конкретно выбираются параметры α_k и h_k , получают конкретный вычислительный алгоритм минимизации функции $J(\Theta)$. В частности, если принять $h_k = -\nabla J(\bar{\Theta}^{(k)})$, что соответствует утверждению 2.12, то получим семейство *градиентных* алгоритмов. Из всего многообразия подобных методов остановимся на одном — методе *последовательной линейаризации*. Суть его заключается в следующем.

Пусть в соответствии с какими-либо предварительными соображениями можно указать предполагаемое значение $\bar{\Theta}^{(0)}$ оцен-

ки $\bar{\Theta}$. Если таковые соображения отсутствуют, примем $\bar{\Theta}^{(0)} = 0_{m+1}$. Назовем эту величину *нулевым приближением* к оценке $\bar{\Theta}$. Полагая функцию $\Psi(\Theta)$ дифференцируемой, разложим ее в окрестности точки $\bar{\Theta}^{(0)}$ в ряд Тейлора, ограничив разложение первой частичной суммой. Приближенно получим:

$$\Psi(\Theta) \approx \Psi(\bar{\Theta}^{(0)}) + D^{(0)} \Delta \Theta^{(0)}, \quad \Delta \Theta^{(0)} = \Theta - \bar{\Theta}^{(0)}, \quad D^{(0)} = d\Psi(\bar{\Theta}^{(0)})/d\Theta.$$

Подставив это разложение в (2.68), получим функцию J , квадратично зависящую от отклонения $\Delta \Theta^{(0)}$. Найдем вектор $\Delta \Theta^{(0)}$, минимизирующий J . Очевидно, эта задача аналогична задаче (2.22), решается теми же средствами и приводит к подобному (2.24) результату: $\Delta \Theta^{(0)} = ((D^{(0)})^T D^{(0)})^{-1} (D^{(0)})^T (y - \Psi(\bar{\Theta}^{(0)}))$. Это позволяет найти *первое приближение* $\bar{\Theta}^{(1)}$ к оценке $\bar{\Theta}$

$$\bar{\Theta}^{(1)} = \bar{\Theta}^{(0)} + ((D^{(0)})^T D^{(0)})^{-1} (D^{(0)})^T (y - \Psi(\bar{\Theta}^{(0)})).$$

Если теперь функцию $\Psi(\Theta)$ линейаризовать в окрестности точки $\bar{\Theta}^{(1)}$ и повторить всю операцию, получим *второе приближение*. Обобщение процесса линейаризации с последующей минимизацией квадратичной целевой функции порождает следующее *релаксационное* правило последовательного приближения к МНК-оценке:

$$\bar{\Theta}^{(k+1)} = \bar{\Theta}^{(k)} + ((D^{(k)})^T D^{(k)})^{-1} (D^{(k)})^T (y - \Psi(\bar{\Theta}^{(k)})), \quad (2.70)$$

$$k = 0, 1, \dots$$

Несложно заметить, что сомножитель $(D^{(k)})^T (y - \Psi(\bar{\Theta}^{(k)}))$ в (2.70) коллинеарен антиградиенту функции J , вычисленному в точке $\bar{\Theta}^{(k)}$:

$$(D^{(k)})^T (y - \Psi(\bar{\Theta}^{(k)})) = -0,5 \nabla J(\bar{\Theta}^{(k)}).$$

Поэтому правило последовательных вычислений (2.70) можно представить в эквивалентной форме:

$$\bar{\Theta}^{(k+1)} = \bar{\Theta}^{(k)} - 0,5 ((D^{(k)})^T D^{(k)})^{-1} \nabla J(\bar{\Theta}^{(k)}), \quad k = 0, 1, \dots \quad (2.71)$$

Следовательно, построенный алгоритм можно классифицировать как *градиентный* со специфичной матрицей весовых коэффициентов. В сопоставлении с (2.69) имеем $h_k = -((D^{(k)})^T \times$

$\times D^{(k)})^{-1} \nabla J(\bar{\Theta}^{(k)})$, $\alpha_k = 0,5 = \text{const}$. Покажем, что в первом приближении этот алгоритм определяет минимизирующую последовательность. С этой целью сопоставим значения функции J в точках $\bar{\Theta}^{(k+1)}$ и $\bar{\Theta}^{(k)}$, снова используя частичную сумму ряда Тейлора:

$$J(\bar{\Theta}^{(k+1)}) = J(\bar{\Theta}^{(k)} - 0,5((D^{(k)})^T D^{(k)})^{-1} \nabla J(\bar{\Theta}^{(k)})) = J(\bar{\Theta}^{(k)}) - 0,5 (\nabla J(\bar{\Theta}^{(k)}))^T ((D^{(k)})^T D^{(k)})^{-1} \nabla J(\bar{\Theta}^{(k)}).$$

Так как матрица $(D^{(k)})^T D^{(k)}$ является положительно определенной [26], получаем $J(\bar{\Theta}^{(k+1)}) - J(\bar{\Theta}^{(k)}) < 0$, т.е. в рамках используемых допущений последовательность величин, найденных в соответствии с (2.71), действительно является минимизирующей. Заметим, что факт выполнения этого неравенства можно непосредственно контролировать в процессе проведения расчетов в соответствии с (2.71) и в случае его нарушения коэффициент 0,5 в (2.71) может быть уменьшен до значения, обеспечивающего получение минимизирующей последовательности. Вычисления прекращаются на некотором s -м шаге, если на этом шаге срабатывает выбранное *правило останова*, имеющее, например, вид $\|\nabla J(\bar{\Theta}^{(s)})\| \leq \nu$, где ν — назначенная малая величина. В этом случае принимается $\bar{\Theta} \approx \bar{\Theta}^{(s)}$. Полезно обратить внимание на следующее: если минимизирующая последовательность оказывается и сходящейся, то в точке $\bar{\Theta}$, как следует из (2.71), должно выполняться условие $\nabla J(\bar{\Theta}) = 0$, т.е. точка $\bar{\Theta}$ является стационарной точкой функции (2.68). Это означает, что в случае сходимости алгоритм (2.71) определяет одну из точек *локального* минимума целевой функции (2.68). Эта функция в силу ее нелинейной структуры может оказаться многоэкстремальной, и алгоритм (2.71) приводит к одному из экстремумов, ближайшему в некотором смысле к начальной точке $\bar{\Theta}^{(0)}$. Поиск глобального минимума в задаче (2.68) требует привлечения дополнительных методов *многоэкстремальной* оптимизации.

Прежде чем обсуждать проблему сходимости алгоритма (2.71), обратим внимание на различие смыслового содержания внешне одинаковых обозначений в алгоритмах (2.50) и (2.71): в рекуррентном методе наименьших квадратов (2.50) символ $\bar{\Theta}^{(k)}$ означает *оптимальную* оценку вектора Θ , найденную по экспериментальным данным $y_1, y_2, \dots, y_k$. В то же время в (2.71) символ $\bar{\Theta}^{(k)}$ представляет собой k -е *приближение* к оценке $\bar{\Theta}$, которая ищется по *всем* наблюдениям $y_1, y_2, \dots, y_n$.

Существующие пакеты прикладных программ, о которых упоминалось во введении, позволяют решать задачи нелинейного оценивания параметров в подобных (2.68) задачах без глубокого проникновения в существо используемого алгоритмического обеспечения и его математических особенностей. Но чтобы общение с этими пакетами было не механистическим, а созидательным процессом, пользователь должен обладать определенной математической культурой, позволяющей ему осознанно воспринимать существо используемых вычислительных процедур и сопутствующих им математических закономерностей. Поэтому коротко остановимся на одной важной характеристике алгоритма (2.71), определяющей его принципиальную «жизнеспособность» и известной как сходимость последовательности точек $\bar{\Theta}^{(0)}, \bar{\Theta}^{(1)}, \bar{\Theta}^{(2)}, \dots$, определяемых средствами этого алгоритма, к некоторой точке $\bar{\Theta}$.

При обсуждении сходимости алгоритма (2.71) возможны два концептуально различных взгляда на эту проблему. При первом из них вектор u экспериментальных данных, участвующий в формировании градиента функции J , рассматривается как конкретный числовой вектор, соответствующий результатам проведенного эксперимента. И в этом случае можно найти условия, обеспечивающие сходимость последовательности полученных с помощью правила (2.71) точек $\bar{\Theta}^{(0)}, \bar{\Theta}^{(1)}, \bar{\Theta}^{(2)}, \dots$ к некоторой точке $\bar{\Theta}$ именно при этом конкретном векторе u . Однако если теперь провести другой эксперимент при тех же значениях экзогенных переменных, то в силу случайной природы эндогенной переменной получим новую реализацию u экспериментальных данных, и ранее полученные условия сходимости при этой реализации, вообще говоря, могут не выполняться. Поэтому нужны такие условия сходимости, которые будут обеспечивать сходимость при *всех* n -мерных реализациях эндогенной переменной, образуемых в соответствии с (2.67). В первом случае принято говорить о *детерминированной* сходимости, во втором — о *стохастической*. Для нас больший интерес представляют условия стохастической сходимости, так как именно они гарантируют успешную работоспособность алгоритма (2.71) при всех возможных исходах проводимого эксперимента.

Проблема стохастической сходимости или, по иной терминологии, стохастической устойчивости решений разностных уравнений интенсивно развивается в последние десятилетия (напри-

мер, [18]). Не ставя перед собой задачи детального проникновения в существо проблемы, кратко остановимся на ряде основополагающих понятий.

Прежде всего отметим, что понятие стохастической сходимости не является однозначным (впрочем, как и детерминированным) — бывают сходимость по вероятности, p -сходимость и частный ее вариант среднеквадратическая сходимость, экспоненциальная сходимость, сходимость с вероятностью единица (почти наверное сходимость) и др. Из всех этих видов сходимости наиболее сильной является сходимость с вероятностью единица, так как при выполнении соответствующих условий все реализации случайных последовательностей, получаемые по правилу (2.71) на множестве значений случайного вектора y , кроме, быть может, реализаций с нулевой вероятностью, сходятся к вектору $\hat{\theta}$ в обычном понимании сходимости, соответствующем определению (2.11). Поэтому нас будут интересовать именно условия сходимости с вероятностью единица.

Для последующего изложения удобно алгоритм (2.71) переписать в отклонениях относительно стационарной точки $\hat{\theta}$. С этой целью введем обозначение $\mu^{(k)} = \hat{\theta}^{(k)} - \hat{\theta}$. Тогда можем записать

$$\mu^{(k+1)} = \mu^{(k)} - 0,5 \int_0^1 Q(\hat{\theta} + \gamma \mu^{(k)}) \mu^{(k)} d\gamma, \quad (2.72)$$

где матрица $Q(\theta)$ является матрицей Якоби вектора $(D^T(\theta)D(\theta))^{-1} \nabla J(\theta)$, вычисленной в точке $(\hat{\theta} + \gamma \mu^{(k)})$, т.е.

$$Q(\hat{\theta} + \gamma \mu^{(k)}) = \frac{d}{d\theta} ((D^T(\theta)D(\theta))^{-1} \nabla J(\theta)),$$

$$\theta = \hat{\theta} + \gamma \mu^{(k)}; \quad \Delta(\theta) = \frac{d}{d\theta} \Psi(\theta).$$

Интеграл в (2.72) формирует отклонение вектора $(D^{(k)})^T D^{(k)})^{-1} \nabla J(\hat{\theta}^{(k)})$ из (2.71) от его значения в стационарной точке $\hat{\theta}$, т.е. относительно θ_{m+1} [14]. В обозначениях (2.72) проблема поиска условий стохастической сходимости последовательности векторов $\hat{\theta}^{(0)}, \hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \dots$ к точке $\hat{\theta}$ эквивалентна задаче поиска аналогичных условий сходимости последовательности $\mu^{(0)}, \mu^{(1)}, \mu^{(2)}, \dots$ к точке θ_{m+1} . Дадим теперь определение понятию стохастической сходимости с вероятностью единица.

Определение 2.12. Будем говорить, что последовательность $\mu^{(0)}, \mu^{(1)}, \mu^{(2)}, \dots$ с вероятностью единица сходится к точке θ_{m+1} , если при $\forall \delta > 0 \exists p(\delta) > 0$ такое, что для всех $\|\mu^{(0)}\| \leq p(\delta)$ выполня-

ется $\lim_{N \rightarrow \infty} P \left\{ \sup_{k \geq N} \|\mu^{(k)}\| > \delta \right\} = 0$, где $P\{\dots\}$, как обычно, — вероятность соответствующего события.

По существу, это определение означает следующее: в случае сходимости с вероятностью единица почти все реализации случайной последовательности векторов $\{\mu^{(k)}\}$ (за исключением множества реализаций меры нуль), образованные по правилу (2.72) при различных значениях вектора y и начальных значениях из некоторой окрестности точки θ_{m+1} , сходятся к этой точке в обычном детерминированном смысле. Величина $\sup_{k \geq N} \|\mu^{(k)}\|$

представляет собой наибольшее отклонение реализаций от нулевого значения, начиная с $(N+1)$ -го элемента последовательности и до бесконечного. Если указанное наибольшее отклонение для всех реализаций при $N \rightarrow \infty$ стремится к нулю, то это означает сходимость к нулю всех реализаций. Неравенство в определении 2.12 означает, что доля несходящихся реализаций стремится к нулю.

Проблема выявления условий сходимости с вероятностью 1 весьма сложна. Наиболее общим математическим подходом к ее решению в настоящее время является применение стохастических функций Ляпунова с использованием свойств супермартингалов.

Определение 2.13. Положительно определенная непрерывная функция $V(x)$ векторного аргумента x , обладающая свойствами $V(0) = 0$, $V(x)$ конечна при всех x с конечной нормой $\|x\|$ и $V(x) \rightarrow \infty$ при $\|x\| \rightarrow \infty$, называется функцией Ляпунова.

Определение 2.14. Пусть случайная последовательность векторов $z_0, z_1, z_2, \dots$ обладает свойствами: $M\{z_k\} < \infty$, $M\{z_k | z_0, z_1, \dots, z_{k-1}\} \leq z_{k-1}$. Тогда эта последовательность называется супермартингалом.

Применение введенных понятий для выявления условий сходимости с вероятностью 1 последовательности векторов $\mu^{(0)}, \mu^{(1)}, \mu^{(2)}, \dots$, образованных в соответствии с (2.72), к точке θ_{m+1} основывается на следующем утверждении.

Утверждение 2.13. Пусть $V(\mu)$ — функция Ляпунова и последовательность $V(\mu^{(0)}), V(\mu^{(1)}), V(\mu^{(2)}), \dots$ значений этой функции, вычисленная на элементах случайной последовательности $\{\mu^{(k)}\}$, является супермартингалом. Тогда последовательность $\mu^{(0)}, \mu^{(1)}, \mu^{(2)}, \dots$ с вероятностью 1 сходится к точке θ_{m+1} .

Мы не будем останавливаться на доказательстве этой теоремы и ограничимся некоторыми комментариями к ее физическому содержанию. С этой точки зрения величину $V(\mu^{(k)})$ можно интерпретировать как «обобщенную энергию» некоторой динамической системы, описываемой разностным уравнением (2.72). То обстоятельство, что функция Ляпунова оказывается супермартингалом, означает уменьшение в среднем энергии в точке $\mu^{(k)}$ по сравнению с ее значением в предыдущей точке $\mu^{(k-1)}$, а это влечет за собой соответствующее уменьшение величины $\|\mu^{(k)}\|$.

Применим утверждение 2.13 для анализа стохастической сходимости последовательности (2.72). С этой целью введем в рассмотрение стохастическую функцию Ляпунова $V(\mu) = \|\mu\|$ и рассмотрим условное математическое ожидание

$$\begin{aligned} M\{\|\mu^{(k+1)}\| \mid \mu^{(0)}, \mu^{(1)}, \mu^{(2)}, \dots, \mu^{(k)}\} &= \\ &= M\{\|\mu^{(k)} - 0,5 \int_0^1 Q(\bar{\theta} + \gamma \mu^{(k)}) \mu^{(k)} d\gamma\| \mid \mu^{(0)}, \dots, \mu^{(k)}\} = \\ &= M\{\|\int_0^1 (E_{m+1} - 0,5Q(\bar{\theta} + \gamma \mu^{(k)})) \mu^{(k)} d\gamma\| \mid \mu^{(0)}, \dots, \mu^{(k)}\}. \end{aligned}$$

Из этого соотношения следует

$$\begin{aligned} M\{\|\mu^{(k+1)}\| \mid \mu^{(0)}, \mu^{(1)}, \mu^{(2)}, \dots, \mu^{(k)}\} &\leq \\ &\leq M\{\|\int_0^1 (E_{m+1} - 0,5Q(\bar{\theta} + \gamma \mu^{(k)})) d\gamma\| \mid \mu^{(0)}, \dots, \mu^{(k)}\}. \end{aligned}$$

Пусть выполняется условие

$$\max_{\theta} M\{\|E_{m+1} - 0,5Q(\theta)\|\} < 1, \quad (2.73)$$

т.е. наибольшее значение математического ожидания нормы случайной матрицы $E_{m+1} - 0,5Q(\theta)$ меньше единицы. Тогда

$$M\{\|\mu^{(k+1)}\| \mid \mu^{(0)}, \mu^{(1)}, \mu^{(2)}, \dots, \mu^{(k)}\} < \|\mu^{(k)}\|$$

или же $M\{V^{(k+1)} \mid \mu^{(0)}, \mu^{(1)}, \mu^{(2)}, \dots, \mu^{(k)}\} < V^{(k)}$, где $V^{(k)} = \|\mu^{(k)}\|$. Из последнего неравенства следует, что последовательность $V^{(0)}, V^{(1)}, V^{(2)}, \dots$ образует супермартингал. Но тогда последовательность $\mu^{(0)}, \mu^{(1)}, \mu^{(2)}, \dots$ в силу утверждения (2.13) оказывается сходящейся с вероятностью 1. Следовательно, условие (2.73) является достаточным для того, чтобы последовательность приближений $\bar{\theta}^{(0)}, \bar{\theta}^{(1)}, \bar{\theta}^{(2)}, \dots$, вычисляемых по правилу (2.71), при любом векторе наблюдений y , удовлетворяющем определению (2.67), с вероятностью 1 сходилась к стационарной точке $\bar{\theta}$. Практическая проверка этих условий требует дополнительных усилий.

Рассмотрим ряд частных случаев. Пусть функция $\Psi(\theta)$ является линейной: $\Psi(\theta) = \Psi\theta$, $\Psi \in R^{n \times (m+1)}$. Тогда $D(\theta) = \frac{d}{d\theta}(\Psi\theta) = \Psi$,

$$\nabla J = -2\Psi^T(y - \Psi\theta), \quad Q = \frac{d}{d\theta}((D^T D)^{-1} \Delta J) = 2E_{m+1} \text{ и условие (2.73)}$$

вырождается в тривиальное $\max_{\theta} M\{\|0_{(m+1) \times (m+1)}\|\} < 1$, где $0_{(m+1) \times (m+1)}$ — нулевая квадратная матрица, т.е. алгоритм (2.71) устойчив при любой матрице Ψ и любой начальной точке $\bar{\theta}^{(0)}$. Природа этого результата совершенно очевидна: из (2.71) в этом случае уже при $k=1$ следует $\bar{\theta}^{(1)} = \bar{\theta}^{(0)} + (\Psi^T \Psi)^{-1} \Psi^T (y - \Psi \bar{\theta}^{(0)}) = (\Psi^T \Psi)^{-1} \Psi^T y = \bar{\theta}$, т.е. алгоритм (2.71) при линейной модели наблюдений за одну итерацию определяет МНК-оценку, что и отражается в тривиальности условий сходимости.

Пусть теперь с целью уменьшения вычислительных затрат в алгоритме (2.71) используется постоянная матрица весовых коэффициентов, т.е. положим $0,5((D^{(k)})^T D^{(k)})^{-1} = G = \text{const}$, где G — некоторая положительно определенная постоянная матрица. Пусть, как и выше, функция $\Psi(\theta) = \Psi\theta$, т.е. является линейной. В этом случае $Q = 4G\Psi^T \Psi$ и условия (2.73) приобретают форму неравенства

$$\max_{\theta} M\{\|E_{m+1} - 2G\Psi^T \Psi\|\} = \|E_{m+1} - 2G\Psi^T \Psi\| < 1,$$

т.е. сводятся к ограничению на норму стационарной матрицы. Если это условие выполняется, алгоритм (2.71) сходится при лю-

бом начальном условии $\hat{\Theta}^{(0)}$ и любом векторе наблюдений y , соответствующем модели (2.4). Известно [26], что при *любой* норме квадратной матрицы C справедливо неравенство $\|C\| \geq \max_j |\lambda_j|$,

где λ_j , $j = 1, 2, \dots$ — собственные числа матрицы C . Но тогда условия сходимости алгоритма (2.71) при постоянной матрице весовых коэффициентов и линейной модели наблюдений можно сформулировать так: для того, чтобы в указанных условиях алгоритм (2.71) сходиллся с вероятностью 1 при любых начальных условиях $\hat{\Theta}^{(0)}$ (в таком случае говорят о сходимости *в целом*), достаточно, чтобы собственные числа матрицы $E_{m+1} - 2G\Psi^T\Psi$ по абсолютному значению были меньше единицы. Можно показать, что эти достаточные условия в данном случае являются и необходимыми. Действительно, положив в (2.71) $0,5((D^{(k)})^T D^{(k)})^{-1} = G$, $\nabla J(\hat{\Theta}^{(k)}) = -2\Psi^T(y - \Psi\hat{\Theta}^{(k)})$, запишем алгоритм (2.71) в форме

$$\hat{\Theta}^{(k+1)} = (E_{m+1} - 2G\Psi^T\Psi)\hat{\Theta}^{(k)} + 2G\Psi^T y.$$

Это соотношение представляет собой линейное неоднородное разностное уравнение, для устойчивости решений которого, как известно (например, [10]), *необходимо и достаточно*, чтобы собственные числа матрицы $E_{m+1} - 2G\Psi^T\Psi$ по абсолютным значениям были меньше единицы. Легко найти установившееся значение $\hat{\Theta}$ решения этого уравнения. В случае сходимости должно выполняться равенство

$$\begin{aligned}\hat{\Theta} &= (E_{m+1} - 2G\Psi^T\Psi)\hat{\Theta} + 2G\Psi^T y \Rightarrow G\Psi^T\Psi\hat{\Theta} = \\ &= G\Psi^T y \Rightarrow \hat{\Theta} = (\Psi^T\Psi)^{-1}\Psi^T y,\end{aligned}$$

т.е. установившимся значением по-прежнему оказывается МНК-оценка.

2.4. Максимально правдоподобные оценки регрессионных параметров

Большим достоинством метода наименьших квадратов, в значительной степени определившим его широкое применение в эконометрических приложениях, являются ограниченные «претензии» к объему априорной информации. По существу, эта информация ограничивается моделью экспериментальных данных и

предположением о центрированности и некоррелированности ошибок эксперимента. Вместе с тем если исследователь располагает большим объемом априорных сведений о переменных, участвующих в постановке проблемы, их нужно пытаться рационально использовать в надежде добиться более высоких по точности результатов, нежели это регламентировано методом наименьших квадратов. Одним из источников потенциального прогресса может явиться более глубокое проникновение в природу влияния латентных переменных и измерительных технологий на эндогенную переменную. Если соответствующий анализ покажет, что вектор ϵ , участвующий в формировании апостериорных данных, не является гауссовским или является гауссовским, но с коррелированными компонентами, имеет смысл методу наименьших квадратов предпочесть нечто иное, способное использовать выявленные особенности экспериментальных ошибок. Методом, рационально учитывающим априорную информацию об ошибках эксперимента в предположении, что регрессионные параметры по-прежнему классифицируются как неизвестные, является *метод максимального правдоподобия*.

Основой метода является совместная плотность вероятностей экспериментальных данных y , полученная при фиксированном значении регрессионных параметров Θ , т.е. условная плотность вероятностей $L(y|\Theta)$. Чтобы ее найти, необходимо знать модель вектора y и статистические свойства вектора ϵ . Будем для определенности ориентироваться на регрессионную модель (2.67) как более полную. Пусть $\omega_\epsilon(\epsilon)$ — плотность вероятностей вектора ϵ и эта плотность известна. Проблема ее поиска здесь не обсуждается. При фиксированных параметрах Θ причиной «случайности» вектора y в (2.67), как уже отмечалось, является вектор ϵ . Поэтому знания плотности $\omega_\epsilon(\epsilon)$ вполне достаточно для вычисления условной плотности $L(y|\Theta)$. Векторы y и ϵ в этом случае различаются только математическими ожиданиями $M\{y|\Theta\} = \Psi(\Theta) + M\{\epsilon\}$, и можем записать $L(y|\Theta) = \omega_\epsilon(\epsilon)|_{\epsilon=y-\Psi(\Theta)}$. Таким образом, условная плотность $L(y|\Theta)$ получается из совместной плотности вероятностей $\omega_\epsilon(\epsilon)$ ошибок ϵ заменой аргумента ϵ на $y - \Psi(\Theta)$.

Определение 2.15. Пусть в процессе проведения эксперимента эндогенная переменная Y приняла значения $y = [y_1, y_2, \dots, y_n]^T$ и построена условная плотность вероятностей $L(y|\Theta)$. Если теперь в выражении условной плотности аргумент y заменить на

конкретные результаты проведенного эксперимента, получим функцию, зависящую только от вектора Θ . Эту функцию называют *функцией правдоподобия*.

В последующем по традиции функцию правдоподобия будем обозначать так же, как и условную плотность, — $L(y|\Theta)$. Однако следует иметь в виду, что условная плотность $L(y|\Theta)$ является $(m+1)$ -параметрической функцией вектора y ; в то же время функция правдоподобия, в составе которой вектор y фиксирован, рассматривается как функция вектора Θ .

Определение 2.16. Значение $\hat{\Theta}$ вектора Θ , при котором функция правдоподобия $L(y|\Theta)$ или, что то же самое, функция $\ln L(y|\Theta)$ достигает наибольшего значения, называется *максимально правдоподобной оценкой* вектора Θ регрессионных параметров.

Таким образом, максимально правдоподобная оценка находится из условия

$$\hat{\Theta} = \arg \max_{\Theta} \ln L(y|\Theta). \quad (2.74)$$

Смысл этого условия таков: если в результате проведения эксперимента эндогенная переменная Y приняла конкретные значения y , то в качестве оценки $\hat{\Theta}$ следует принять то значение вектора Θ , при котором вероятность наблюдать в эксперименте именно этот вектор y оказывается наибольшей.

Вычислительная процедура поиска максимально правдоподобных оценок в случае модели наблюдений (2.67) может быть организована с использованием тех же принципов, что и при поиске МНК-оценок. Для этого достаточно общий подход (2.69) «адаптировать» к задаче максимизации логарифма функции правдоподобия.

Рассмотрим более детально случай линейной регрессионной модели (2.4) и гауссовского вектора ϵ . Пусть $\epsilon \sim N(0, K_{\epsilon})$, т.е.

$$\omega_{\epsilon}(\epsilon) = \frac{1}{\sqrt{(2\pi)^n |K_{\epsilon}|}} \exp \left\{ -\frac{1}{2} \epsilon^T K_{\epsilon}^{-1} \epsilon \right\}.$$

Как следствие, получаем

$$\ln L(y|\Theta) = \text{const} - 0,5(y - \Psi\Theta)^T K_{\epsilon}^{-1}(y - \Psi\Theta),$$

где $\text{const} = \ln \frac{1}{\sqrt{(2\pi)^n |K_{\epsilon}|}}$, т.е. не зависящее от Θ слагаемое.

Воспользовавшись необходимым условием экстремума $\nabla \ln L(y|\Theta) = 0$, которое обычно называют *уравнением правдоподобия*, получим уравнение

$$\Psi^T K_{\epsilon}^{-1}(y - \Psi\Theta) = 0,$$

из которого следует максимально правдоподобная оценка вектора Θ при линейной гауссовской модели

$$\hat{\Theta} = (\Psi^T K_{\epsilon}^{-1} \Psi)^{-1} \Psi^T K_{\epsilon}^{-1} y. \quad (2.75)$$

Если сопоставить эту оценку с МНК-оценкой (2.24), легко обнаружить ее отличительную особенность: оценка (2.75) учитывает коррелированность экспериментальных ошибок в объеме ковариационной матрицы K_{ϵ} . Если эти ошибки не коррелированы, т.е. $K_{\epsilon} = \sigma^2 E_{n \times n}$, то максимально правдоподобная оценка (2.75) вырождается в МНК-оценку (2.24), что является вполне ожидаемым и естественным результатом. Заметим, что иногда оценку (2.75) получают не в терминах правдоподобия, а путем модификации метода наименьших квадратов, сопровождаемой заменой целевой функции (2.20) $J = \|y - \Psi\Theta\|^2$ на функцию $J = (y - \Psi\Theta)^T K_{\epsilon}^{-1}(y - \Psi\Theta)$ с последующей ее минимизацией. При подобном подходе оценку (2.75) принято называть *обобщенной* МНК-оценкой.

Изучим основные свойства оценки (2.75). Прежде всего покажем, что она является несмещенной. Действительно,

$$\begin{aligned} M\{\hat{\Theta}\} &= (\Psi^T K_{\epsilon}^{-1} \Psi)^{-1} \Psi^T K_{\epsilon}^{-1} M\{y\} = \\ &= (\Psi^T K_{\epsilon}^{-1} \Psi)^{-1} \Psi^T K_{\epsilon}^{-1} M\{\Psi\Theta + \epsilon\} = \Theta. \end{aligned}$$

Далее найдем ковариационную матрицу K_{η} ошибки $\eta = \hat{\Theta} - \Theta$. Имеем $\eta = (\Psi^T K_{\epsilon}^{-1} \Psi)^{-1} \Psi^T K_{\epsilon}^{-1} \epsilon$ и, следовательно,

$$K_{\eta} = M\{\eta\eta^T\} = (\Psi^T K_{\epsilon}^{-1} \Psi)^{-1}. \quad (2.76)$$

При некоррелированных ошибках получаем уже известный результат (2.27). Несложно убедиться, что соотношение (2.76) при условиях (2.4) и $\epsilon \sim N(0, K_{\epsilon})$ соответствует эффективной оценке. Действительно, матрица Фишера, определяемая в соответствии с (2.10), в данном случае оказывается равной $\Phi = -\nabla^2 \ln L(y|\Theta) = \Psi^T K_{\epsilon}^{-1} \Psi$. Из сопоставления этого результа-

та с (2.9) и (2.76) следует $K_{\eta} = \Phi^{-1}$, что и является доказательством эффективности максимально правдоподобной оценки при линейной гауссовской модели экспериментальных данных.

Если $K_{\epsilon} = \sigma^2 E_{n \times n}$, но дисперсия σ^2 неизвестна, метод максимального правдоподобия позволяет легко решить проблему оценивания неизвестной дисперсии, совместив соответствующую процедуру с поиском оценки $\hat{\Theta}$. В этом случае функция правдоподобия максимизируется в $(m+2)$ -мерном пространстве по переменным Θ, σ^2 . Уравнения правдоподобия приобретают вид

$\nabla_{\Theta} \ln L(y|\Theta) = 0, \frac{\partial}{\partial \sigma^2} \ln L(y|\Theta) = 0$ и в случае (2.4) конкретизируются следующим образом:

$$\sigma^{-2} \Psi^T (y - \Psi \Theta) = 0, -n\sigma^2 + \|y - \Psi \Theta\|^2 = 0.$$

Решая совместно эту систему, находим $\hat{\Theta} = (\Psi^T \Psi)^{-1} \Psi^T y$, $\hat{\sigma}^2 = \frac{1}{n} \|y - \Psi \hat{\Theta}\|^2$. Оценка $\hat{\Theta}$ вектора Θ , как следует из этих результатов, совпадает с МНК-оценкой (2.24), однако оценка $\hat{\sigma}^2$ дисперсии σ^2 , оптимальная по критерию максимального правдоподобия, в отличие от ее аналога (2.37), оказывается смещенной.

Максимально правдоподобные оценки и в общем случае, особенно при малых объемах экспериментальных данных, имеют некоторое смещение и неминимальную дисперсию. Однако они обладают рядом достоинств, особенно проявляющихся при статистически независимых ошибках $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ эксперимента и заключающихся в следующем.

1. Если для вектора Θ существует эффективная оценка, то уравнение максимального правдоподобия имеет единственное решение.

2. Если для вектора Θ существует достаточная оценка, то каждый корень уравнения правдоподобия является функцией достаточной оценки.

3. Максимально правдоподобная оценка является состоятельной и асимптотически, т.е. при стремлении объема экспериментальной выборки к бесконечности ($n \rightarrow \infty$), эффективной и гауссовской (в смысле ее плотности вероятностей). На практике эта асимптотика проявляется достаточно хорошо уже при десяти независимых наблюдениях на один скалярный параметр, подлежащий оцениванию.

Эти свойства максимально правдоподобных оценок делают их весьма привлекательными при решении многих прикладных задач.

2.5. Метод максимума апостериорной плотности вероятностей

В регрессионных моделях, параметры которых оценивались в соответствии с методами наименьших квадратов и максимального правдоподобия, вектор параметров Θ классифицировался как неизвестный. И это было принципиально в том смысле, что никакая априорная информация об этом векторе при построении процедур оценивания не использовалась. Вместе с тем возможны ситуации, в которых опыт предшествующей работы с регрессионной моделью в аналогичных исследуемой прикладных задачах позволяет считать вектор параметров Θ принадлежащим некоторой генеральной совокупности с известной плотностью вероятностей $\omega_{\Theta}(\Theta)$. В подобных случаях вектор Θ классифицируется как случайный с известной плотностью вероятностей, которая совместно с плотностью вероятностей $\omega_{\epsilon}(\epsilon)$ экспериментальных ошибок и структурой модели отражает наши априорные представления о свойствах регрессионной модели. Этой информацией теперь надо рационально распорядиться, с тем чтобы повысить эффективность оценивания. Одним из способов, позволяющих это сделать, является метод максимума апостериорной плотности вероятностей. Существо метода заключается в следующем.

В рассмотрение вводится условная плотность вероятностей $\mu(\Theta|y)$ вектора параметров Θ , полученная в предположении, что результаты эксперимента приняли некоторое фиксированное значение y . Найти эту плотность формально можно в соответствии с формулой Байеса:

$$\mu(\Theta|y) = \omega_{\Theta}(\Theta) L(y|\Theta) / \omega_y(y),$$

где $\omega_y(y)$ — безусловная плотность вероятностей наблюдений y . Так как

$$\omega_y(y) = \int_{-\infty}^{\infty} \omega_{\Theta}(\Theta) L(y|\Theta) d\Theta,$$

то условная плотность $\mu(\Theta|y)$ выражается непосредственно через функции $\omega_{\Theta}(\Theta)$ и $L(y|\Theta)$:

$$\mu(\Theta|y) = \omega_{\Theta}(\Theta)L(y|\Theta) / \int_{-\infty}^{\infty} \omega_{\Theta}(\Theta)L(y|\Theta)d\Theta. \quad (2.77)$$

Таким образом, если известны априорные плотности вероятностей $\omega_{\Theta}(\Theta)$, $\omega_{\epsilon}(\epsilon)$ и модель экспериментальных данных y , например, в форме (2.4) или (2.67), то условная плотность $\mu(\Theta|y)$ принципиально вычислена.

Определение 2.17. Пусть проведен эксперимент, в котором вектор значений эндогенной переменной принял конкретное численное значение y . Если именно этот конкретный вектор подставить в условную плотность $\mu(\Theta|y)$, то получим функцию $\mu(\Theta|y)$, зависящую только от вектора Θ . Эту функцию называют *апостериорной плотностью вероятностей* вектора регрессионных параметров Θ .

Таким образом, отличие апостериорной плотности вероятностей $\mu(\Theta|y)$ от условной плотности $\mu(\Theta|y)$ проявляется в том, что у первой вектор y принимает не любое фиксированное значение, а именно то, которое соответствует проведенному эксперименту. При этом обе плотности традиционно обозначаются единообразно и из контекста ясно, о какой из них идет речь.

Определение 2.18. Пусть известна апостериорная плотность вероятностей $\mu(\Theta|y)$. Тогда значение $\hat{\Theta}$ вектора Θ , при котором апостериорная плотность $\mu(\Theta|y)$ или, что то же самое, $\ln \mu(\Theta|y)$ достигает максимума, называется *оценкой, оптимальной по критерию максимума апостериорной плотности вероятностей*.

Итак, оценка, оптимальная по критерию максимума апостериорной плотности вероятностей, находится из условия

$$\hat{\Theta} = \arg \max_{\Theta} \ln \mu(\Theta|y).$$

Так как знаменатель в (2.77) не зависит от Θ , то практически этот критерий принимает вид

$$\hat{\Theta} = \arg \max_{\Theta} (\ln \omega_{\Theta}(\Theta) + \ln L(y|\Theta)). \quad (2.78)$$

Снова ограничим рассмотрение случаем линейной гауссовской модели (2.4), положив в ней $\Theta \sim N(m_{\Theta}, K_{\Theta})$, $\epsilon \sim N(0, K_{\epsilon})$.

Векторы Θ и ϵ принимаются некоррелированными. Задача (2.78) при этих ограничениях конкретизируется следующим образом:

$$\hat{\Theta} = \arg \max_{\Theta} (\text{const} - 0,5(\Theta - m_{\Theta})^T K_{\Theta}^{-1}(\Theta - m_{\Theta}) - 0,5(y - \Psi\Theta)^T K_{\epsilon}^{-1}(y - \Psi\Theta)),$$

где, как и выше, под const понимается не зависящее от Θ слагаемое. Необходимое условие экстремума

$$\nabla_{\Theta} (\text{const} - 0,5(\Theta - m_{\Theta})^T K_{\Theta}^{-1}(\Theta - m_{\Theta}) - 0,5(y - \Psi\Theta)^T K_{\epsilon}^{-1}(y - \Psi\Theta)) = 0_{m+1}$$

в этой задаче приобретает форму уравнения

$$K_{\Theta}^{-1}(\Theta - m_{\Theta}) - \Psi^T K_{\epsilon}^{-1}(y - \Psi\Theta) = 0_{m+1},$$

из которого следует искомая оценка

$$\hat{\Theta} = (\Psi^T K_{\epsilon}^{-1} \Psi + K_{\Theta}^{-1})^{-1} (\Psi^T K_{\epsilon}^{-1} y + K_{\Theta}^{-1} m_{\Theta}). \quad (2.79)$$

Таким образом, оценка (2.79), в отличие от максимально правдоподобной оценки (2.75), существенно определяется априорной информацией об оцениваемом векторе Θ в объеме ковариационной матрицы K_{Θ} и математического ожидания m_{Θ} . Однако если эта информация оказывается «расплывчатой» в том смысле, что диагональные элементы матрицы K_{Θ} чрезмерно велики, то оценка (2.79) практически вырождается в максимально правдоподобную. Диагональные элементы матрицы K_{Θ} являются дисперсиями компонентов вектора Θ , и если они велики, влияние априорной информации значительно снижается, и по своим свойствам случайный вектор приближается к неизвестному. Формально неограниченно большие диагональные элементы матрицы K_{Θ} порождают нулевую обратную матрицу K_{Θ}^{-1} , что и превращает оценку (2.79) в оценку (2.75).

С другой стороны, если априорная информация достаточно содержательная, алгоритм (2.79) перестает «доверять» апостериорным данным и в большей степени ориентируется на априорные сведения. Так, если допустить, что $K_{\Theta} \rightarrow 0_{(m+1) \times (m+1)}$, то $K_{\Theta}^{-1} \rightarrow \infty$ и в (2.79) можно слагаемыми $\Psi^T K_{\epsilon}^{-1} \Psi$ и $\Psi^T K_{\epsilon}^{-1} y$ пренебречь, откуда следует $\hat{\Theta} = m_{\Theta}$. В этом предельном случае, таким образом, алгоритм (2.79) вообще не использует экспериментальные данные и полагается только на априорную информацию. Это

вполне объяснимый результат: устремляя матрицу K_{Θ} к нулевой, мы тем самым превращаем гауссовскую плотность $N(m_{\Theta}, K_{\Theta})$ в дельта-функцию $\delta(\Theta - m_{\Theta})$ с «центром» в точке m_{Θ} . Но это, в свою очередь, означает, что априори известно $\Theta = m_{\Theta}$ и нет необходимости не только в доверии к экспериментальным данным, но и в самом эксперименте.

Оценку (2.79) часто оказывается удобным представлять в ином виде, поступая следующим образом. Представим

$$\begin{aligned}\hat{\Theta} &= m_{\Theta} + (\Psi^T K_{\epsilon}^{-1} \Psi + K_{\Theta}^{-1})^{-1} (\Psi^T K_{\epsilon}^{-1} y + K_{\Theta}^{-1} m_{\Theta}) - m_{\Theta} = \\ &= m_{\Theta} + (\Psi^T K_{\epsilon}^{-1} \Psi + K_{\Theta}^{-1})^{-1} (\Psi^T K_{\epsilon}^{-1} y + K_{\Theta}^{-1} m_{\Theta} - \\ &\quad - (\Psi^T K_{\epsilon}^{-1} \Psi + K_{\Theta}^{-1}) m_{\Theta}),\end{aligned}$$

откуда следует

$$\hat{\Theta} = m_{\Theta} + (\Psi^T K_{\epsilon}^{-1} \Psi + K_{\Theta}^{-1})^{-1} \Psi^T K_{\epsilon}^{-1} (y - \Psi m_{\Theta}). \quad (2.80)$$

Это выражение имеет определенный «философский» смысл: оценка определяется как сумма двух слагаемых, первое из которых отображает наши априорные представления о среднем значении параметра Θ , а второе формирует апостериорную поправку к этому среднему значению. Сама поправка содержит сомножитель $y - \Psi m_{\Theta}$, который представляет собой отклонение результатов проведенного эксперимента от ориентированного только на априорные представления прогнозируемого исхода эксперимента. «Вес» поправки определяется априорными свойствами регрессионной модели.

Выявим основные свойства оценки (2.80). Покажем прежде всего, что она является несмещенной. Действительно, при $\Theta \sim N(m_{\Theta}, K_{\Theta})$, $\epsilon \sim N(0, K_{\epsilon})$ из (2.4) следует $M\{y\} = \Psi m_{\Theta}$. Но тогда непосредственное усреднение оценки (2.80) приводит к очевидному результату $M\{\hat{\Theta}\} = m_{\Theta}$, т.е. оценка по максимуму апостериорной плотности вероятностей как безусловная оценка оказывается несмещенной.

Далее находим ковариационную матрицу ошибки оценивания $\eta = \hat{\Theta} - \Theta$. Используя соотношения (2.80) и (2.4), получаем

$$\begin{aligned}\eta &= D^{-1} \Psi^T K_{\epsilon}^{-1} \epsilon - (E - D^{-1} \Psi^T K_{\epsilon}^{-1} \Psi) (\Theta - m_{\Theta}) = \\ &= D^{-1} \Psi^T K_{\epsilon}^{-1} \epsilon - D^{-1} (D - \Psi^T K_{\epsilon}^{-1} \Psi) (\Theta - m_{\Theta}) = \\ &= D^{-1} \Psi^T K_{\epsilon}^{-1} \epsilon - D^{-1} K_{\Theta}^{-1} (\Theta - m_{\Theta}),\end{aligned}$$

где для упрощения записей использовано обозначение $D = \Psi^T K_{\epsilon}^{-1} \Psi + K_{\Theta}^{-1}$. Следовательно, ковариационная матрица K_{η} ошибки при условии некоррелированных векторов Θ и ϵ найдется в результате следующей последовательности операций:

$$\begin{aligned}K_{\eta} &= M\{\eta \eta^T\} = M\{(D^{-1} \Psi^T K_{\epsilon}^{-1} \epsilon - D^{-1} K_{\Theta}^{-1} (\Theta - m_{\Theta})) (\epsilon^T K_{\epsilon}^{-1} \Psi D^{-1} - \\ &\quad - (\Theta - m_{\Theta})^T K_{\Theta}^{-1} D^{-1})\} = D^{-1} \Psi^T K_{\epsilon}^{-1} \Psi D^{-1} + D^{-1} K_{\Theta}^{-1} D^{-1} = D^{-1}.\end{aligned}$$

При этом выводе учитывается симметричность матрицы D . Итак, окончательно:

$$K_{\eta} = D^{-1} = (\Psi^T K_{\epsilon}^{-1} \Psi + K_{\Theta}^{-1})^{-1}. \quad (2.81)$$

Интересно сопоставить этот результат с аналогичной формулой (2.76) метода максимального правдоподобия. Принципиальным является присутствие в (2.81) ковариационной матрицы K_{Θ} вектора Θ , которая может существенно влиять на матрицу K_{η} . Достаточно наглядно это проявляется при скалярном параметре $\Theta = \theta \in \mathbb{R}$ ($m + 1 = 1$). В этом случае $\Psi \in \mathbb{R}^{n \times 1}$, $\Psi^T K_{\epsilon}^{-1} \Psi$ — скалярная величина и $\sigma_{\eta}^2 = (\Psi^T K_{\epsilon}^{-1} \Psi + \sigma_{\Theta}^{-2})^{-1}$, где σ_{η}^2 , σ_{Θ}^2 — соответствующие дисперсии. Отсюда следует, что максимально правдоподобная оценка скалярного параметра всегда по точности хуже аналогичной оценки, найденной по методу максимума апостериорной плотности вероятностей. Преимущество последней обусловлено именно априорной информацией — в данном случае дисперсией σ_{Θ}^2 . Чем она меньше, т.е. чем меньше разброс реализаций параметра Θ относительно его математического ожидания, тем это преимущество существеннее. При $\sigma_{\Theta}^2 \rightarrow 0$, что, как уже отмечалось, соответствует, по существу, детерминированной ситуации, дисперсия σ_{η}^2 ошибки оценивания по методу максимума апостериорной плотности вероятностей также стремится к нулю. В методе максимального правдоподобия это свойство, разумеется, не проявляется. Если же $\sigma_{\Theta}^2 \rightarrow \infty$, что практически соответствует полному отсутствию априорной информации об оцениваемом параметре, точности обеих оценок равны. Выявленные закономерности, естественно, проявляются таким же образом и при векторном регрессионном параметре Θ . Однако следует иметь в виду, что за выявленное преимущество приходится платить дополнительными усилиями, направленными на приобретение и обоснование априорной информации.

2.6. Байесовские оценки регрессионных параметров

Среди большого числа методов, рекомендуемых современной теорией статистических решений для поиска оценок параметров, наиболее универсальным является байесовский метод. Условия его применимости те же, что и в методе максимума апостериорной плотности вероятностей: задана модель экспериментальных данных, векторы регрессионных параметров Θ и экспериментальных ошибок ϵ классифицируются как случайные, известна соответствующая априорная информация в объеме плотностей вероятностей $\omega_{\Theta}(\Theta)$, $\omega_{\epsilon}(\epsilon)$. Содержательное существо метода заключается в следующем.

Использование любого метода оценивания сопровождается ошибкой $\eta(y, \Theta) = \hat{\Theta}(y) - \Theta$, которая зависит от конкретной реализации экспериментальных данных y , будучи при одних реализациях малой, при других — большой. Естественным является стремление с большими ошибками встречаться как можно реже. Это намерение можно попытаться реализовать, если процедуру оценивания «наказывать» за большие ошибки. С этой целью в рассмотрение вводят некоторую функцию $C(\Theta, \hat{\Theta})$, которая, как правило, зависит от ошибки $\eta(y, \Theta)$ и называется *функцией стоимости*, или *функцией потерь*. Выбор этой функции в значительной степени субъективен. Однако структура $C(\eta(y, \Theta))$ обычно такова, что функция возрастает (или, по крайней мере, не убывает) с ростом любого компонента аргумента и достигает наименьшего значения при $\eta(y, \Theta) = 0$. Такой характер функции обеспечивает возрастающую стоимость больших ошибок оценивания, за которые система должна подвергаться «наказанию». Среди функций с указанным свойством наиболее часто применяющимися оказываются следующие.

Функция потерь, зависящая от модулей ошибок,

$$C(\eta) = \sum_{i=0}^m |\Theta_i - \hat{\Theta}_i|.$$

Квадратичная функция потерь

$$C(\eta) = \sum_{i=0}^m (\Theta_i - \hat{\Theta}_i)^2 = \|\hat{\Theta} - \Theta\|^2.$$

Прямоугольная функция потерь

$$C(\eta) = \sum_{i=0}^m c_i, \quad c_i = \begin{cases} 1, & |\Theta_i - \hat{\Theta}_i| \geq \Delta, \\ 0, & |\Theta_i - \hat{\Theta}_i| < \Delta, \end{cases}$$

где Δ — некоторая константа.

Находит применение и *простая функция потерь*

$$C(\eta) = c - \sum_{i=0}^m \delta(\Theta_i - \hat{\Theta}_i),$$

где $c > 0$ — константа, $\delta(\Theta_i - \hat{\Theta}_i)$ — дельта-функция Дирака.

Функция стоимости является неслучайной функцией случайного аргумента η и на множестве значений аргумента принимает случайные значения. Среднее значение функции стоимости, полученное усреднением ее по всем возможным значениям векторов Θ и y , называют *средними потерями* или *средним риском* (байесовским риском):

$$J = \int \int C(\eta(y, \Theta)) v(y, \Theta) d\Theta dy, \quad (2.82)$$

где $v(y, \Theta)$ — совместная плотность вероятностей векторов y и Θ .

Индексы y интегралов символически отражают тот факт, что интегрирование ведется по всем пространствам существования векторов y и Θ . Так как ошибка оценивания $\eta(y, \Theta) = \hat{\Theta}(y) - \Theta$ зависит от оценки $\hat{\Theta}$, т.е. от способа оценивания, то и средний байесовский риск зависит от алгоритма оценивания. Очевидно, можно допустить существование такого алгоритма, при котором средние потери окажутся наименьшими по сравнению с потерями, сопутствующими другим алгоритмам.

Определение 2.19. Функция $\hat{\Theta}(y)$ экспериментальных данных, при которой средние потери достигают минимума, называется *оптимальной по Байесу оценкой*, или *байесовской оценкой* регрессионных параметров.

Таким образом, байесовская оценка определяется условием

$$\hat{\Theta}(y) = \arg \min_{\Theta} J.$$

Если воспользоваться представлением $v(y, \Theta) = \omega_y(y) \mu(\Theta | y)$, то сформулированная оптимизационная задача примет вид

$$J = \int_y \omega_y(y) \left(\int_{\Theta} C(\eta(y, \Theta)) \mu(\Theta | y) d\Theta \right) dy \rightarrow \min_{\Theta(y)} \quad (2.83)$$

Так как плотность вероятностей $\omega_y(y)$ неотрицательна, то условие (2.83) выполняется, если минимального значения достигает внутренний интеграл в (2.83) при любом значении вектора y , т.е.

$$\int_{\Theta} C(\eta(y, \Theta)) \mu(\Theta | y) d\Theta \rightarrow \min_{\Theta(y)} \quad (2.84)$$

Последующее решение задачи (2.84) требует конкретизации функции стоимости. В настоящее время из всех вышеприведенных функций наибольшее распространение в прикладных задачах получила квадратичная функция стоимости. Это обусловлено относительной простотой математического аппарата, сопровождающего связанные с этой функцией преобразования, и хорошим соответствием функции требованиям задачи оценивания: с возрастанием ошибки оценивания «наказание» существенно возрастает.

Утверждение 2.14. При квадратичной функции стоимости байесовская оценка регрессионных параметров представляет собой *апостериорное среднее* оцениваемых параметров.

Действительно, при квадратичной функции стоимости задача (2.84) принимает вид

$$\int_{\Theta} \|\hat{\Theta} - \Theta\|^2 \mu(\Theta | y) d\Theta \rightarrow \min_{\Theta(y)}$$

Воспользовавшись необходимым условием минимума

$$\int_{\Theta} \nabla_{\hat{\Theta}} \|\hat{\Theta} - \Theta\|^2 \mu(\Theta | y) d\Theta \rightarrow 0_{m+1},$$

получим уравнение

$$\int_{\Theta} \|\hat{\Theta} - \Theta\|^2 \mu(\Theta | y) d\Theta = 0_{m+1},$$

из которого с учетом независимости величин $\hat{\Theta}$ и Θ и нормировки условной плотности $\mu(\Theta | y)$ получим

$$\hat{\Theta}(y) = \int_{\Theta} \Theta \mu(\Theta | y) d\Theta. \quad (2.85)$$

Если в (2.85) под y понимать конкретный результат проведенного эксперимента, то выражение в правой части (2.85) принято называть *апостериорным средним* вектора Θ .

Итак, байесовская оценка при квадратичной функции стоимости является апостериорным средним оцениваемого параметра. И это очень важный для приложений результат. Для ряда априорных плотностей вероятностей $\omega_{\Theta}(\Theta)$ и $\omega_{\epsilon}(\epsilon)$ и регрессионных моделей (2.67) интегрирование в (2.85) удастся провести аналитически и получить оценку $\hat{\Theta}(y)$ в виде явной зависимости от экспериментальных данных y . Можно вычислить и минимальное значение байесовского риска, соответствующее оптимальной оценке (2.85). Для этого байесовский риск (2.82) при квадратичной функции стоимости представим в виде

$$\begin{aligned} J &= \int_y \int_{\Theta} \hat{\Theta}^T (\hat{\Theta} - \Theta) v(y, \Theta) d\Theta dy - \int_y \int_{\Theta} \Theta^T (\hat{\Theta} - \Theta) v(y, \Theta) d\Theta dy = \\ &= \int_y \hat{\Theta}^T \omega_y(y) \int_{\Theta} (\hat{\Theta} - \Theta) \mu(\Theta | y) d\Theta dy - \int_y \int_{\Theta} \Theta^T \hat{\Theta} v(y, \Theta) d\Theta dy + \\ &\quad + \int_{\Theta} \Theta^T \int_y v(y, \Theta) dy d\Theta. \end{aligned}$$

В силу необходимого условия минимума первое слагаемое в правой части этого выражения обращается в нуль и после несложных преобразований получаем

$$\begin{aligned} \min J &= \int_{\Theta} \Theta^T \Theta \omega_{\Theta}(\Theta) d\Theta - \int_y \int_{\Theta} \Theta^T \mu(\Theta | y) d\Theta \hat{\Theta} \omega_y(y) dy = \\ &= \int_{\Theta} \Theta^T \Theta \omega_{\Theta}(\Theta) d\Theta - \int_y \hat{\Theta}^T \hat{\Theta} \omega_y(y) dy. \end{aligned}$$

Полезно обратить внимание на структуру полученного выражения. Первое слагаемое в нем определяет ту часть средних потерь, которая обусловлена априорными сведениями об оцениваемых параметрах. Второе слагаемое (вычитаемое) показывает, насколько уменьшаются средние потери в связи с проведенной операцией оценивания. Так как $\Theta^T \Theta = \sum_i \Theta_i^2$, после усреднения в первом слагаемом по Θ выражение для минимального риска можно представить так:

$$\min J = Sp K_{\Theta} + m_{\Theta}^T m_{\Theta} - \int_y \bar{\Theta}^T \bar{\Theta} \omega_y(y) dy, \quad (2.86)$$

где K_{Θ} , m_{Θ} — ковариационная матрица и математическое ожидание вектора Θ соответственно.

Аналогичным образом можно представить и второе слагаемое.

Рассмотрим теперь частный случай задачи, позволяющий на основании (2.85) получить алгоритм оценивания как явную функцию от экспериментальных данных y . Этот случай охватывает линейную модель (2.4) при гауссовских некоррелированных векторах Θ и ϵ , т.е., как и ранее, принимается $\Theta \sim N(m_{\Theta}, K_{\Theta})$, $\epsilon \sim N(0, K_{\epsilon})$. Необходимая для оценивания апостериорная плотность $\mu(\Theta | y)$ находится в соответствии с (2.77). Запишем в явном виде формирующие ее функции:

$$\omega_{\Theta}(\Theta) = \frac{1}{\sqrt{(2\pi)^{m+1} |K_{\Theta}|}} \exp\{-0,5(\Theta - m_{\Theta})^T K_{\Theta}^{-1}(\Theta - m_{\Theta})\},$$

$$L(y | \Theta) = \frac{1}{\sqrt{(2\pi)^n |K_{\epsilon}|}} \exp\{-0,5(y - \Psi\Theta)^T K_{\epsilon}^{-1}(y - \Psi\Theta)\}.$$

Общий показатель экспоненты, образующийся после подстановки этих функций в (2.77), представим в удобной для последующего интегрирования форме

$$(\Theta - m_{\Theta})^T K_{\Theta}^{-1}(\Theta - m_{\Theta}) + (y - \Psi\Theta)^T K_{\epsilon}^{-1}(y - \Psi\Theta) = \\ = q + (\Theta - G)^T D(\Theta - G).$$

Величины q , G , D находим из условия тождественного выполнения последнего соотношения

$$q = m_{\Theta}^T K_{\Theta}^{-1} m_{\Theta} + y^T K_{\epsilon}^{-1} y - G^T D G, \quad D = \Psi^T K_{\epsilon}^{-1} \Psi + K_{\Theta}^{-1}, \\ G = D^{-1}(K_{\Theta}^{-1} m_{\Theta} + \Psi^T K_{\epsilon}^{-1} y).$$

Подставив с использованием этих преобразований функцию (2.77) в (2.85), получим

$$\bar{\Theta} = \int_{-\infty}^{\infty} \Theta \exp\{-0,5(\Theta - G)^T D(\Theta - G)\} d\Theta / \int_{-\infty}^{\infty} \exp\{-0,5(\Theta - G)^T D(\Theta - G)\} d\Theta.$$

Дальнейшее многомерное интегрирование осуществляется следующим образом (например, [29]). Так как матрица D симметрическая, существует ортогональная матрица R , диагонализующая матрицу D , т.е. обладающая свойствами $R^T R = E$, $R^T D R = S = \text{diag}[s_i]$, $i = 1, 2, \dots, m+1$, где S — диагональная матрица из собственных чисел s_i матрицы D . Исходя из существования такой матрицы, введем новую переменную Q , которая с вектором Θ связана соотношением $\Theta - G = RQ \Rightarrow d\Theta = |R| dQ$. Так как R — ортогональная матрица, можем положить $|R| = 1$, т.е. $d\Theta = dQ$. В результате выражение для оценки $\bar{\Theta}$ принимает вид

$$\bar{\Theta} = G - R \int_{-\infty}^{\infty} \Theta \exp\{-0,5Q^T S Q\} dQ / \int_{-\infty}^{\infty} \exp\{-0,5Q^T S Q\} dQ.$$

Пусть $Q = [q_1 \ q_2 \ \dots \ q_{m+1}]^T$. Тогда $Q^T S Q = \sum_{i=1}^{m+1} s_i q_i^2$ и многомер-

ный интеграл в числителе оценки $\bar{\Theta}$ в развернутом виде оказывается таков:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \begin{bmatrix} q_1 \\ q_2 \\ \dots \\ q_{m+1} \end{bmatrix} \exp\left\{-0,5 \sum_{i=1}^{m+1} s_i q_i^2\right\} dq_1 dq_2 \dots dq_{m+1}.$$

Так как в силу нечетности подынтегральной функции

$$\int_{-\infty}^{\infty} q_j \exp\{-0,5 s_j q_j^2\} dq_j = 0, \quad j = 1, 2, \dots, m+1,$$

то второе слагаемое в выражении для $\bar{\Theta}$ обращается в нуль и окончательно получаем

$$\bar{\Theta} = G = D^{-1}(\Psi^T K_{\epsilon}^{-1} y + K_{\Theta}^{-1} m_{\Theta}). \quad (2.87)$$

Оптимальная байесовская оценка, таким образом, при линейной гауссовской модели экспериментальных данных и квадратичной функции стоимости оказывается линейной относительно вектора наблюдений y . Линейность проявляется именно при этих условиях. Если же модель экспериментальных данных окажется нелинейной относительно вектора параметров Θ или

хотя бы один из векторов Θ , ϵ будет негауссовским, байесовский алгоритм становится нелинейным.

Легко видеть, что байесовский алгоритм (2.87) полностью совпадает с оценкой (2.79), оптимальной по критерию максимума апостериорной плотности вероятностей. И это не случайно. Гауссовская апостериорная плотность вероятностей вектора Θ достигает наибольшего значения в точке, являющейся апостериорным средним этого вектора. Но и байесовская оценка (2.85) представляет собой апостериорное среднее. Этим и объясняется совпадение обеих оценок. Подобное совпадение обнаруживается при всех унимодальных и симметричных относительно моды апостериорных плотностях вероятностей. Следствием совпадения оценок (2.79), (2.87) оказывается совпадение и их свойств: байесовская оценка (2.87) является несмещенной и имеет ковариационную матрицу ошибки (2.81). Минимальное значение $\min J$ байесовского риска (2.82) в данном случае находится достаточно просто. По определению $\min J = M\{\|\hat{\Theta} - \Theta\|^2\} = M\{(\hat{\Theta} - \Theta)^T(\hat{\Theta} - \Theta)\}$, где $\hat{\Theta}$ — байесовская оценка (2.87). С другой стороны, из (2.81) для этой же оценки имеем $M\{(\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^T\} = D^{-1}$. Но $M\{(\hat{\Theta} - \Theta)^T(\hat{\Theta} - \Theta)\} = Sp M\{(\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^T\}$ и, следовательно, $\min J = Sp D^{-1}$.

Следует отметить, что байесовский подход является наиболее универсальным средством оценивания. Можно показать, например, что максимально правдоподобные оценки соответствуют минимуму так называемого условного байесовского риска (см. далее) при простой функции потерь. Оценки, соответствующие максимуму апостериорной плотности вероятностей, могут быть получены из условия минимума среднего (безусловного) риска (2.82) также при простой функции потерь.

2.7. Минимаксные оценки регрессионных параметров

Познакомимся еще с одним методом оценивания, который в эконометрических задачах не имеет широкого применения, но потенциально может оказаться полезным. Этот метод формирует так называемые минимаксные оценки и применяется в тех случаях, когда априорная плотность $\omega_{\Theta}(\Theta)$ случайного параметра Θ неизвестна и оценивание осуществляется в расчете на наиболее неблагоприятную обстановку.

Снова обратимся к средним байесовским потерям (2.82). Воспользуемся уже неоднократно применявшимся представлением $v(y, \Theta) = \omega_{\Theta}(\Theta)L(y|\Theta)$ и запишем средний риск в новой форме:

$$J = \int_{\Theta} r(\Theta)\omega_{\Theta}(\Theta)d\Theta, \quad (2.88)$$

где функция $r(\Theta)$ называется *условными потерями (условным риском)* и определяется очевидным образом:

$$r(\Theta) = \int_y C(\Theta, \hat{\Theta}(y))L(y|\Theta)dy. \quad (2.89)$$

Пусть каким-либо образом найдена *любая* оценка $\hat{\Theta}_1(y)$. Если эту оценку подставить в (2.89), то, проинтегрировав, найдем условный риск как функцию параметра Θ . Обозначим эту функцию символом $r_1(\Theta)$. При некотором Θ функция $r_1(\Theta)$ достигает наибольшего значения. Пусть это будет в точке $\Theta = \Theta_1$, т.е. $\Theta_1 = \arg \max r_1(\Theta)$, и в этой точке условный риск принимает значение $r_1(\Theta_1) = \max r_1(\Theta)$. Если, далее, взять какую-либо другую оценку $\hat{\Theta}_2(y)$ и проделать те же самые операции, то при сохранении того же принципа обозначений получим $r_2(\Theta_2) = \max r_2(\Theta)$. Для оценки $\hat{\Theta}_3(y)$ аналогичным образом найдем $r_3(\Theta_3) = \max r_3(\Theta)$. Если теперь абстрактно представить, что удалось перебрать все мыслимые оценки, то найдется такая оценка $\hat{\Theta}^*(y)$, которой будет соответствовать значение $r^*(\Theta^*) = \max r^*(\Theta)$, *меньшее*, чем при любой другой оценке. Эта оценка и называется *минимаксной*.

Формально, таким образом, минимаксная оценка $\hat{\Theta}^*$ определяется условием

$$\max_{\Theta} r(\Theta, \hat{\Theta}^*) = \min_{\hat{\Theta}} \max_{\Theta} r(\Theta, \hat{\Theta}), \quad (2.90)$$

т.е. при минимаксной оценке *максимум* условного риска по вектору регрессионных параметров Θ оказывается *наименьшим* среди всех максимумов при любых других оценках, что и определяет название этого метода оценивания.

Критерий (2.90) минимакса на практике часто оказывается не совсем удовлетворительным, так как по сравнению с другими минимаксная оценка может дать выигрыш в очень малой области значений вектора Θ , реальные значения которого этой области могут и не принадлежать. Дополнительно на пути поиска ми-

нимаксных оценок встречаются не всегда преодолимые аналитические трудности.

В теории статистических решений доказывается следующее.

1. Минимаксная оценка существует и представляет собой байесовскую оценку при некоторой априорной плотности вероятностей $\omega_{\Theta}^*(\Theta)$ вектора Θ , которая называется *наименее благоприятной априорной плотностью*. Плотность $\omega_{\Theta}^*(\Theta)$ является такой, что оптимальная байесовская оценка при этой плотности характеризуется большим средним риском, чем оптимальная байесовская оценка при любой другой плотности $\omega_{\Theta}(\Theta)$ и той же функции стоимости. Таким образом, минимаксная оценка дает наилучший результат в наихудших условиях, которые, вообще говоря, могут и не соответствовать решаемой конкретной задаче.

2. Если существует такая байесовская оценка $\hat{\Theta}^+(y)$, при которой условный риск (2.89) оказывается константой, не зависящей от вектора Θ , то эта оценка будет минимаксной, т.е. $\hat{\Theta}^*(y) = \hat{\Theta}^+(y)$. Если же условный риск обращается в константу при *небайесовской* оценке $\hat{\Theta}(y)$, то это не означает, что $\hat{\Theta}(y)$ является минимаксной оценкой.

Вторым свойством иногда удается воспользоваться для нахождения наименее благоприятной плотности $\omega_{\Theta}^*(\Theta)$. Для этого байесовскую оценку как функцию записанной в общем виде априорной плотности $\omega_{\Theta}(\Theta)$ (например, как апостериорное среднее при квадратичной функции стоимости) подставляют в выражение для риска (2.89) и пытаются найти такую плотность $\omega_{\Theta}(\Theta)$, при которой условный риск обращается в не зависящую от Θ константу. В случае удачи соответствующая плотность $\omega_{\Theta}(\Theta)$ оказывается наименее благоприятной плотностью $\omega_{\Theta}^*(\Theta)$, при которой байесовская оценка оказывается минимаксной. В ряде случаев, в частности при линейной модели (2.4) экспериментальных данных и квадратичной функции стоимости, наименее благоприятной плотностью является равномерная.

Вторая часть

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ВРЕМЕННЫХ РЯДОВ

Глава 3

СТРУКТУРНО ДЕТЕРМИНИРОВАННЫЕ ВРЕМЕННЫЕ РЯДЫ

3.1. Математические модели структурно детерминированных временных рядов

Часто анализ какого-либо экономического явления сопровождается изучением свойств некоторой функции $y(t)$, характеризующей его развитие и представляющей собой функцию времени. Такую изменяющуюся во времени функцию принято называть *процессом*. Так, например, могут представлять интерес ежемесячный объем продукции, выпускаемой некоторым предприятием, ежемесячная заработная плата какой-либо семьи, ежегодная урожайность определенной сельскохозяйственной культуры, изменение температуры в течение того или иного промежутка времени и т.п. Область определения подобных процессов может быть как непрерывной, так и дискретной. Температура непрерывно изменяется во времени, что может быть измерено прибором и квалифицировано как функция с непрерывной областью определения. Зарплата обычно фиксируется один раз в месяц и, следовательно, определена на дискретном множестве значений ее аргумента. Но независимо от исходной природы процесса практически регистрация его значений обычно осуществляется в дискретные, как правило, равноотстоящие моменты времени

$t_1, t_2, \dots, t_N, t_{i+1} - t_i = \text{const}$, следствием чего является совокупность величин $y(t_1), y(t_2), \dots, y(t_N)$. Эти величины удобно обозначить более лаконичными символами $y_1, y_2, \dots, y_N$ или $y_i, i = 1, 2, \dots, N$. Упорядоченную во времени последовательность наблюдений $y_i, i = 1, 2, \dots, N$, принято называть *временным рядом*.

Элементы, или *уровни*, $y_1, y_2, \dots, y_N$ временного ряда являются случайными величинами в том смысле, что заранее, до проведения эксперимента, точные значения их предсказать невозможно. Так, едва ли кто-либо совершенно точно может сказать, какая температура будет в шесть часов утра такого-то числа или какая урожайность пшеницы будет в i -м году. Даже заработная плата человека, находящегося на окладе, может претерпевать заранее мало ожидаемые изменения.

Временной ряд, отражающий эволюцию какого-либо экономического процесса, используется для формирования определенных суждений о развитии этого процесса. Чтобы это суждение (решение) выработать, элементы ряда следует подвергнуть математической обработке по определенному правилу (алгоритму). Но для этого необходимо математически описать сам ряд, т.е. составить его математическую модель.

Мы будем придерживаться двух взглядов на математическую природу временного ряда. При первом из них зарегистрированную последовательность уровней $y_1, y_2, \dots, y_N$ интерпретируем как реализацию некоторого случайного процесса (случайной последовательности) и соответствующую математическую модель строим на основе аппарата теории случайных процессов. Полученную таким образом математическую модель временного ряда будем называть *стохастической*. Изучение и применение таких моделей даны в следующей главе.

Вторая точка зрения заключается в том, что мы считаем временной ряд состоящим из двух слагаемых, первое из которых представляет собой полностью определенную с точностью до нескольких параметров функцию (такие функции часто называют квазидетерминированными или структурно детерминированными), а второе слагаемое является последовательностью независимых центрированных случайных величин. Временные ряды с такой трактовкой структуры их модели будем называть *структурно детерминированными* (или квазидетерминированными). Их изучению посвящена настоящая глава.

Итак, в соответствии со второй концепцией полагаем:

$$y_i = f_i + p_i, i = 1, 2, \dots, N, \quad (3.1)$$

где $f_i = f(t_i)$ — квазидетерминированная составляющая, обычно называемая *трендом* временного ряда, p_i — случайная составляющая.

Математически первая из них принимается известной с точностью до некоторой совокупности неизвестных параметров $a_0, a_1, \dots, a_q$, как правило, линейно входящих в выражение тренда $f(t_i)$. Весьма распространенной такой зависимостью является линейная комбинация известных функций $\phi_k(t), k = 0, 1, \dots, q$:

$$f_i = f(t_i) = \sum_{k=0}^q a_k \phi_k(t_i), i = 1, 2, \dots, N. \quad (3.2)$$

Частным случаем модели (3.2) является полиномиальная, в которой $\phi_k(t_i) = t_i^k$ и, следовательно,

$$f_i = \sum_{k=0}^q a_k t_i^k. \quad (3.3)$$

В этом случае детерминированная составляющая временного ряда (тренд) является многочленом q -го порядка с неизвестными коэффициентами $a_k, k = 0, 1, 2, \dots, q$. Этим коэффициентам можно придать определенный физический смысл, если функцию $f(t)$, считая ее q раз дифференцируемой, разложить в ряд Тейлора в окрестности, например, точки t_1 и ограничить разложение $(q+1)$ -й частичной суммой

$$f(t_i) = \sum_{k=0}^q \frac{1}{k!} f^{(k)}(t_1) (t_i - t_1)^k.$$

В этом случае $\phi_k(t_i) = \frac{1}{k!} (t_i - t_1)^k, a_k = f^{(k)}(t_1), k = 0, 1, \dots, q$, т.е. неизвестные параметры физически представляют собой значение функции $f(t)$, скорости ее изменения, ускорения и т.д. в точке $t = t_1$.

Если ввести в рассмотрение $(q+1)$ -мерные векторы

$$a = [a_0 \ a_1 \ \dots \ a_q]^T, \Phi_i = [\phi_0(t_i) \ \phi_1(t_i) \ \dots \ \phi_q(t_i)]^T,$$

то выражение (3.2) можем переписать в более удобной форме:

$$f_i = (\Phi_i, a) = \Phi_i^T a, \quad i = 1, 2, \dots, N, \quad (3.4)$$

где, как и ранее, $(\cdot)$ — символ скалярного произведения, T — символ транспонирования.

Случайная составляющая в (3.1), как уже отмечалось, обладает свойствами:

$$M\{p_i\} = 0, \quad M\{p_i p_j\} = \begin{cases} \sigma^2, & i = j \\ 0, & i \neq j \end{cases} \quad (3.5)$$

Здесь, как и ранее, M — символ усреднения. Дисперсия σ^2 может быть как известной, так и, что наиболее часто наблюдается в реальных задачах, неизвестной. Последовательность случайных величин со свойствами (3.5) обычно принято называть *дискретным белым шумом*. Второе условие в (3.5) является следствием независимости величин p_i и p_j при $i \neq j$, а дисперсия, вообще говоря, может меняться во времени, т.е. $\sigma^2 = \sigma_i^2$, но мы пока ограничим рассмотрение случаем (3.5).

Таким образом, в случае структурно детерминированного временного ряда имеем

$$M\{y_i\} = (\Phi_i, a) = \Phi_i^T a, \quad M\{(y_i^\circ)^2\} = \sigma^2,$$

где y_i° — центрированная случайная величина.

Временной ряд (3.1) имеет детерминированное с точностью до вектора a математическое ожидание, и члены ряда относительно его среднего совершают случайные отклонения со свойствами (3.5). Эти отклонения обусловлены случайными изменениями самого ряда, а также могут быть порождены неточностью измерений, сопровождающих эксперимент.

В связи с моделью (3.1), (3.4) возникает ряд важных вопросов, в частности: как выбрать функции $\Phi_k(t)$, чтобы в последующем получить достаточно «комфортные» алгоритмы обработки ряда; какое число q составляющих модели следует принять, чтобы получить модель, адекватную реальным наблюдениям; как поступать с неизвестной дисперсией σ^2 и, наконец, для какой цели используется временной ряд.

Ответим прежде всего на последний вопрос. С использованием рядов могут решаться различные прикладные задачи. Приоритетной из них мы будем считать следующую: временной ряд используется для того, чтобы по имеющимся наблюдениям y_i , $i = 1, 2, \dots, N$, спрогнозировать будущее значение ряда для какого-либо момента $t_m > t_N$, т.е. найти оценку $\hat{Y}_m = \hat{Y}(t_m) = \hat{Y}_m(y_1, y_2, \dots, y_N)$ будущего значения ряда как функцию имеющихся в нашем распоряжении наблюдений $y_1, y_2, \dots, y_N$. Такую задачу принято называть задачей *прогнозирования* (предсказания, экстраполяции) временного ряда. Ответы на остальные вопросы, так же как и на вопрос о том, как найти «хорошую» в каком-то смысле оценку $\hat{Y}_m$, будем искать в процессе последующего изложения материала.

В заключение обратим внимание на внешнее сходство модели (3.1), (3.4) временного ряда и рассмотренных в первой части пособия регрессионных моделей. По существу они отличаются только физической содержательностью. Если время t в (3.1) интерпретировать как экзогенную переменную, а значения ряда как значения эндогенной переменной, то формально модели окажутся эквивалентными. Это позволяет в теорию временных рядов привнести многое из регрессионного анализа, рассмотренного ранее. По этой же причине структурно детерминированные модели временных рядов иногда называют регрессионными.

3.2. Ортонормированные системы функций

3.2.1. Банаховы и гильбертовы пространства

Как уже было отмечено, при разработке математической модели временного ряда большое внимание должно быть уделено свойствам функций $\Phi_k(t)$, $k = 0, 1, \dots, q$, используемых для описания детерминированной составляющей (3.2) временного ряда. Удачный выбор этих функций существенно определяет вычислительную простоту и точность решения задачи. Хотя универсальных и общепринятых рекомендаций по выбору этих функций нет, можно указать класс функций, во многих задачах оказывающихся предпочтительными. Этот класс составляют ортогональные и ортонормированные системы функций. Дадим соответствующие определения. При этом будем ориентироваться на основные понятия конечномерных линейных пространств, изложенные, на-

пример, в [14], и опустим соответствующие пространственные комментарии.

Определение 3.1. Множество $\mathfrak{F}$, элементами которого являются вещественные функции (для определенности, времени t) или временные последовательности с установленными соотношениями между элементами, будем называть *вещественным функциональным пространством*.

Определение 3.2. Функциональное пространство $\mathfrak{F}$ называют *линейным*, если для любых двух элементов $u(t)$ и $v(t)$ из этого пространства (т.е. $u(t), v(t) \in \mathfrak{F}$) и любых чисел $\lambda_1, \lambda_2 \in \mathbb{R}$ линейная комбинация $z(t) = \lambda_1 u(t) + \lambda_2 v(t)$ также принадлежит множеству $\mathfrak{F}$.

В этом определении операции умножения функций на число и сложения функций понимаются в обычном алгебраическом смысле (в этом же смысле указанные операции понимаются и далее).

Определение 3.3. Линейное функциональное пространство $\mathfrak{F}$ называется *нормированным*, если существует правило, ставящее в соответствие каждому элементу $f(t) \in \mathfrak{F}$ вещественное число, называемое *нормой* (или *длиной*) функции $f(t)$, символически обозначаемое $\|f(t)\|$ и удовлетворяющее трем аксиомам нормы конечномерных линейных пространств [14]:

1. $\|f(t)\| > 0$ при $f(t) \neq 0$ и $\|f(t)\| = 0 \Leftrightarrow f(t) = 0$;
2. $\|\lambda f(t)\| = |\lambda| \|f(t)\|$ при $\forall \lambda \in \mathbb{R}$;
3. $\|f(t) + y(t)\| \leq \|f(t)\| + \|y(t)\|$, при $\forall y(t), f(t) \in \mathfrak{F}$.

Определение 3.4. Функциональное пространство $\mathfrak{F}$ называется *метрическим*, если существует правило, ставящее в соответствие любым двум функциям $u(t)$ и $v(t)$ из $\mathfrak{F}$ неотрицательное вещественное число, называемое *метрикой* (расстоянием) пространства $\mathfrak{F}$, символически обозначаемое $\rho(u(t), v(t))$ и удовлетворяющее трем аксиомам метрики конечномерных линейных пространств [14]:

1. $\rho(u(t), v(t)) = 0 \Leftrightarrow u(t) = v(t)$;
2. $\rho(u(t), v(t)) = \rho(v(t), u(t))$;
3. $\rho(u(t), v(t)) \leq \rho(u(t), z(t)) + \rho(z(t), v(t))$.

Определение 3.5. Линейное функциональное пространство $\mathfrak{F}$ называется *евклидовым* (или пространством со скалярным произ-

ведением), если существует правило, ставящее в соответствие любым двум функциям $u(t)$ и $v(t)$ из $\mathfrak{F}$ вещественное число, называемое *скалярным произведением* функций $u(t)$ и $v(t)$, символически обозначаемое $(u(t), v(t))$ и удовлетворяющее четырем аксиомам скалярных произведений конечномерных линейных пространств [14]:

1. $(u(t), v(t)) = (v(t), u(t))$;
2. $(u(t) + z(t), v(t)) = (u(t), v(t)) + (z(t), v(t))$;
3. $(\lambda u(t), v(t)) = \lambda (u(t), v(t))$ при $\forall \lambda \in \mathbb{R}$;
4. $(u(t), u(t)) > 0$ при $u(t) \neq 0$ и $(u(t), u(t)) = 0 \Leftrightarrow u(t) = 0$.

Если в линейном функциональном пространстве определено скалярное произведение, то обычно принимают:

$$\|f(t)\| = \sqrt{(f(t), f(t))}, \quad (3.6)$$

$$\rho(u(t), v(t)) = \|u(t) - v(t)\| = \sqrt{(u(t) - v(t), u(t) - v(t))} \quad (3.7)$$

и в зависимости от способа определения скалярного произведения получают те или иные нормы и метрику. В таком случае говорят, что норма (метрика) порождена скалярным произведением. Заметим, что если отвлечься от математической аккуратности мышления, то норму можно интерпретировать как своеобразную меру «величины» функции и указать, какая функция по норме больше, а какая меньше. С аналогичной степенью строгости метрику можно рассматривать как «расстояние» между двумя функциями.

Для функциональных пространств одной из фундаментальных является проблема сходимости последовательности элементов этого пространства.

Определение 3.6. Пусть $\{f_n(t)\}_{n=1}^{\infty}$ — последовательность функций из метрического пространства $\mathfrak{F}$. Если для $\forall \epsilon > 0 \exists N(\epsilon) > 0$ такое, что при всех $n \geq N$ выполняется $\rho(f_n(t), f(t)) < \epsilon$, где $f(t) \in \mathfrak{F}$, то говорят, что последовательность $\{f_n(t)\}_{n=1}^{\infty}$ в метрике пространства $\mathfrak{F}$ сходится к функции $f(t) \in \mathfrak{F}$.

Другими словами, последовательность функций $\{f_n(t)\}_{n=1}^{\infty}$ сходится в метрике пространства $\mathfrak{F}$ к функции $f(t)$ из этого пространства, если $\lim_{n \rightarrow \infty} \rho(f_n(t), f(t)) = 0$ при $n \rightarrow \infty$. Это означает, что с ростом n «расстояние» между элементами последовательности

$\{f_n(t)\}$ и функцией $f(t)$ становится сколь угодно малым. Если не возникает недоразумений, то в подобных случаях пишут: $f_n(t) \rightarrow f(t)$ при $n \rightarrow \infty$.

Если пространство нормировано и метрика порождена нормой, то в случае $\lim_{n \rightarrow \infty} \|f_n(t) - f(t)\| = 0$ говорят, что последовательность $\{f_n(t)\}$ по норме сходится к $f(t)$. В последующем, как правило, именно в этом смысле будем понимать сходимость.

Определение 3.7. Последовательность функций $\{f_n(t)\}_{n=1}^{\infty}$ из метрического пространства $\mathcal{F}$ называют *последовательностью Коши* или *фундаментальной* последовательностью, если для $\forall \epsilon > 0 \exists N(\epsilon) > 0$ такое, что для всех $m, n > N(\epsilon)$ выполняется неравенство $\rho(f_m(t), f_n(t)) < \epsilon$.

У фундаментальной последовательности, таким образом, «расстояние» между ее элементами при достаточно больших номерах элементов становится сколь угодно малым.

Утверждение 3.1. Если некоторая последовательность $\{f_n(t)\}_{n=1}^{\infty} \subset \mathcal{F}$, где $\mathcal{F}$ — метрическое пространство, сходится к функции $f(t) \in \mathcal{F}$, то эта последовательность является фундаментальной.

Действительно, пусть последовательность $\{f_n(t)\}_{n=1}^{\infty} \subset \mathcal{F}$ сходится к функции $f(t) \in \mathcal{F}$. Это значит, что при $\forall \epsilon > 0 \exists N(\epsilon) > 0$ такое, что при $n > N$ справедливо неравенство $\rho(f_n(t), f(t)) < \epsilon$. Так как в соответствии с одной из аксиом метрики $\rho(f_n(t), f_m(t)) \leq \rho(f_n(t), f(t)) + \rho(f(t), f_m(t))$, то при $n, m \rightarrow \infty$ выполняется неравенство $\rho(f_m(t), f_n(t)) < 2\epsilon$, т.е. при достаточно больших n, m «расстояние» между любыми двумя элементами сходящейся последовательности становится сколь угодно малым. Но в соответствии с определением 3.7 последовательность $\{f_n(t)\}_{n=1}^{\infty}$ тогда является фундаментальной.

Сходящаяся последовательность, таким образом, всегда фундаментальная. Но фундаментальная последовательность функций, в отличие от числовой последовательности, оказывается не всегда сходящаяся. В некоторых метрических пространствах удастся построить такие последовательности Коши, которые не сходятся ни к какой функции из этого пространства. Вместе с тем существуют и пространства, в которых фундаментальные последовательности оказываются сходящимися, причем сходящимися

именно к какому-либо элементу этого же пространства. Такие пространства принято выделять в самостоятельный класс.

Определение 3.8. Если в метрическом пространстве $\mathcal{F}$ любая фундаментальная последовательность сходится к некоторому элементу этого же пространства, то пространство $\mathcal{F}$ называется *полным*.

Определение 3.9. Нормированное линейное пространство, полное относительно метрики $\rho(u(t), v(t)) = \|u(t) - v(t)\|$, порожденной его нормой, называется *банаховым* пространством.

Определение 3.10. Линейное функциональное пространство $\mathcal{F}$ со скалярным произведением $(u(t), v(t))$ называется *гильбертовым*, если оно полно относительно нормы, порожденной его скалярным произведением.

Таким образом, если в нормированном пространстве «расстояние» между функциями определено через норму пространства и любая фундаментальная последовательность в этом пространстве сходится по норме к элементу этого пространства, то такое нормированное пространство принято называть банаховым. Если дополнительно функциональное пространство является евклидовым, норма в нем выражена через скалярное произведение и любая фундаментальная последовательность по этой норме сходится к некоторому элементу этого пространства, то такое евклидово пространство называют гильбертовым.

3.2.2. Пространство L_2

Абстрактные функциональные пространства в эконометрических приложениях не находят широкого применения. Обычно используют множества с конкретным заданием скалярного произведения, нормы и метрики. Распространенным множеством такого вида является пространство L_2 .

Определение 3.11. Функция $f(t)$ называется функцией с интегрируемым квадратом, или *квадратично интегрируемой* на отрезке $[t_1, t_2]$, если

$$\int_{t_1}^{t_2} f^2(t) dt < \infty, \quad (3.8)$$

т.е. интеграл существует и конечен.

Квадратично интегрируемые функции обладают рядом полезных для последующего свойств. В частности:

- 1) квадратично интегрируемая функция является интегрируемой;
- 2) произведение двух квадратично интегрируемых функций является интегрируемой функцией;
- 3) сумма двух квадратично интегрируемых функций является квадратично интегрируемой функцией;
- 4) если $f(t)$ квадратично интегрируема, то и $\lambda f(t)$ квадратично интегрируема, где $\lambda \in \mathbb{R}$.

Таким образом, линейная комбинация квадратично интегрируемых функций является квадратично интегрируемой, а это значит, что множество квадратично интегрируемых функций образует линейное пространство.

Определим на множестве квадратично интегрируемых функций следующую интегральную операцию над некоторыми функциями $f(t)$ и $u(t)$:

$$(f(t), u(t)) = \int_{t_1}^{t_2} \mu(t) f(t) u(t) dt, \quad (3.9)$$

где $\mu(t)$ — произвольная положительная интегрируемая функция.

Несложно убедиться, что выражение (3.9) удовлетворяет всем четырем аксиомам скалярного произведения и, следовательно, является скалярным произведением в линейном пространстве квадратично интегрируемых функций.

Определение 3.12. Линейное функциональное пространство, состоящее из квадратично интегрируемых на $[t_1, t_2]$ функций, со скалярным произведением (3.9) называется *пространством L_2* .

Норма и метрика в этом пространстве определяются соответственно выражениями:

$$\|f(t)\| = \sqrt{\int_{t_1}^{t_2} \mu(t) f(t)^2 dt}, \quad (3.10)$$

$$\rho(f(t), u(t)) = \sqrt{\int_{t_1}^{t_2} \mu(t) (f(t) - u(t))^2 dt}. \quad (3.11)$$

В евклидовом пространстве, как известно, выполняются *неравенства Коши—Буняковского* и *треугольника* (Минковского), имеющие в общем случае соответственно вид:

$$(f(t), u(t))^2 \leq (f(t), f(t))(u(t), u(t)), \\ \|f(t) + u(t)\| \leq \|f(t)\| + \|u(t)\|.$$

В пространстве L_2 эти неравенства таковы:

$$\left[\int_{t_1}^{t_2} \mu(t) f(t) u(t) dt \right]^2 \leq \int_{t_1}^{t_2} \mu(t) f(t)^2 dt \int_{t_1}^{t_2} \mu(t) u(t)^2 dt, \quad (3.12)$$

$$\sqrt{\int_{t_1}^{t_2} \mu(t) (f(t) + u(t))^2 dt} \leq \sqrt{\int_{t_1}^{t_2} \mu(t) f(t)^2 dt} + \sqrt{\int_{t_1}^{t_2} \mu(t) u(t)^2 dt}. \quad (3.13)$$

Доказывается (например, [14]), что пространство L_2 является полным, т.е. представляет собой гильбертово пространство со скалярным произведением (3.9). Во многих случаях принимается $\mu(t) = 1$, и тогда приведенные соотношения упрощаются.

Сходимость последовательности функций $\{f_n(t)\}_{n=1}^{\infty} \subset L_2$ к функции $f(t) \in L_2$ понимается в смысле равенства

$$\lim_{n \rightarrow \infty} \int_{t_1}^{t_2} \mu(t) (f_n(t) - f(t))^2 dt = 0,$$

и в этом случае говорят, что последовательность функций $\{f_n(t)\}_{n=1}^{\infty}$ сходится к функции $f(t)$ в *среднеквадратическом* смысле.

3.2.3. Ортогональные и ортонормированные системы функций

Рассмотрим некоторую систему функций $\{f_n(t)\}_{n=1}^{\infty} \subset L_2, t \in [t_1, t_2]$, у которой ни одна из функций тождественно не обращается в нуль на $[t_1, t_2]$.

Определение 3.13. Система функций $\{f_n(t)\}_{n=1}^{\infty}$ называется *линейно независимой* на отрезке $[t_1, t_2]$, если любая конечная сово-

купность этих функций $f_{\alpha_1}(t), f_{\alpha_2}(t), \dots, f_{\alpha_n}(t)$ является линейно независимой при всех $t \in [t_1, t_2]$, т.е. равенство $\sum \beta_i f_{\alpha_i}(t) = 0$ возможно лишь тогда, когда все $\beta_i = 0$.

Отметим, что в функциональных пространствах в общем случае можно указать систему из произвольного конечного числа линейно независимых функций. Поэтому такие пространства (в том числе и L_2) называют *бесконечномерными*.

Определение 3.14. Система функций $\{f_n(t)\}_{n=1}^{\infty}$ называется *ортogonalной* на отрезке $[t_1, t_2]$, если любые две функции $f_i(t), f_j(t)$ ($i \neq j$) из этой системы обладают свойством: $(f_i(t), f_j(t)) = 0$. Сами функции в этом случае также называются ортогональными.

Определение 3.15. Система функций $\{f_n(t)\}_{n=1}^{\infty}$ называется *ортонормированной* на отрезке $[t_1, t_2]$, если любые две функции $f_i(t), f_j(t)$ из этой системы ортогональны и каждая из них имеет единичную норму, т.е.

$$(f_i(t), f_j(t)) = \begin{cases} 0, & i \neq j, \\ 1, & i = j. \end{cases}$$

В этом случае говорят, что сама функция *нормирована*.

Ортогональную систему функций легко превратить в ортонормированную путем перехода к функциям $f_i^*(t) = f_i(t)/\|f_i(t)\|$, $i = 1, 2, \dots$. Ортогональная система функций является линейно независимой. Действительно, рассмотрим равенство $\sum \beta_i f_i(t) = 0$.

Умножив обе его части скалярно на функцию $f_k(t)$, с учетом ортогональности получим $\beta_k \|f_k(t)\|^2 = 0 \Rightarrow \beta_k = 0$, т.е. линейная комбинация ортогональных функций обращается в нуль только при равных нулю весовых коэффициентах, что является признаком линейной независимости. Опять же, любую линейно независимую систему функций с помощью специальной операции ортогонализации можно превратить в ортогональную или в ортонормированную.

Определение 3.16. Ортогональная (ортонормированная) система функций называется *полной* в L_2 , если в этом пространстве нет ни одной функции, кроме нулевой, которая была бы ортогональна ко всем функциям данной системы.

Доказывается, что в пространстве L_2 существуют полные ортогональные системы функций. Приведем ряд таких систем, характерных для приложений.

1. Основная тригонометрическая система функций

$$1, \cos t, \sin t, \cos 2t, \sin 2t, \cos 3t, \sin 3t, \dots$$

Система является ортогональной на отрезке $[-\pi, \pi]$. Действительно, пусть $i \neq j$. Тогда при $\mu(t) = 1$

$$(\cos it, \cos jt) = \int_{-\pi}^{\pi} \cos it \cos jt \, dt = \frac{1}{2} \int_{-\pi}^{\pi} (\cos(i-j)t + \cos(i+j)t) \, dt = 0,$$

$$(\sin it, \sin jt) = \int_{-\pi}^{\pi} \sin it \sin jt \, dt = \frac{1}{2} \int_{-\pi}^{\pi} (\cos(i-j)t - \cos(i+j)t) \, dt = 0,$$

$$(\cos it, \sin jt) = \int_{-\pi}^{\pi} \cos it \sin jt \, dt = \frac{1}{2} \int_{-\pi}^{\pi} (\sin(i+j)t - \sin(i-j)t) \, dt = 0,$$

$$(1, \cos it) = \int_{-\pi}^{\pi} \cos it \, dt = 0,$$

$$(1, \sin it) = \int_{-\pi}^{\pi} \sin it \, dt = 0.$$

Система не нормирована, так как

$$\|1\| = \sqrt{\int_{-\pi}^{\pi} dt} = \sqrt{2\pi},$$

$$\|\cos it\| = \sqrt{\int_{-\pi}^{\pi} (\cos it)^2 \, dt} = \sqrt{\pi},$$

$$\|\sin it\| = \sqrt{\int_{-\pi}^{\pi} (\sin it)^2 \, dt} = \sqrt{\pi}.$$

Ортонормированная основная тригонометрическая система имеет вид:

$$\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}} \cos t, \frac{1}{\sqrt{\pi}} \sin t, \frac{1}{\sqrt{\pi}} \cos 2t, \frac{1}{\sqrt{\pi}} \sin 2t, \frac{1}{\sqrt{\pi}} \cos 3t, \frac{1}{\sqrt{\pi}} \sin 3t, \dots$$

2. Тригонометрическая система общего вида

$$1, \cos \omega t, \sin \omega t, \cos 2\omega t, \sin 2\omega t, \cos 3\omega t, \sin 3\omega t, \dots$$

Система ортогональна на отрезке $\left[-\frac{T}{2}, \frac{T}{2}\right]$, где $T = \frac{2\pi}{\omega}$. Доказательство аналогично предыдущему. Ортонормированная тригонометрическая система общего вида такова:

$$\frac{1}{\sqrt{T}}, \sqrt{\frac{2}{T}} \cos \omega t, \sqrt{\frac{2}{T}} \sin \omega t, \sqrt{\frac{2}{T}} \cos 2\omega t, \sqrt{\frac{2}{T}} \sin 2\omega t, \sqrt{\frac{2}{T}} \cos 3\omega t, \sqrt{\frac{2}{T}} \sin 3\omega t, \dots$$

3. Многочлены Чебышева первого рода $T_n(t)$

Эти многочлены задаются рекуррентным правилом

$$T_{n+1}(t) = 2tT_n(t) - T_{n-1}(t), \quad n = 1, 2, 3, \dots$$

при начальных условиях $T_0(t) = 1$, $T_1(t) = t$. Они являются ортогональными на отрезке $[-1, 1]$ с весом $\mu(t) = 1/\sqrt{1-t^2}$, причем

$$(T_i(t), T_j(t)) = \int_{-1}^1 \frac{T_i(t)T_j(t)}{\sqrt{1-t^2}} dt = \begin{cases} 0, & i \neq j, \\ \pi, & i = j \geq 1, \\ \pi, & i = j = 0. \end{cases}$$

Ортогональные многочлены Чебышева, таким образом, имеют вид:

$$\frac{1}{\sqrt{\pi}} T_0(t), \sqrt{\frac{2}{\pi}} T_1(t), \sqrt{\frac{2}{\pi}} T_2(t), \sqrt{\frac{2}{\pi}} T_3(t), \dots$$

4. Многочлены Чебышева второго рода $U_n(t)$

задаются рекуррентным соотношением

$$U_{n+1}(t) = 2tU_n(t) - U_{n-1}(t), \quad n = 1, 2, 3, \dots$$

при начальных условиях $U_0(t) = 1$, $U_1(t) = 2t$. Эти многочлены ортогональны на отрезке $[-1, 1]$ с весом $\mu(t) = \sqrt{1-t^2}$, при этом

$$(U_i(t), U_j(t)) = \int_{-1}^1 \sqrt{1-t^2} U_i(t) U_j(t) dt = \begin{cases} 0, & i \neq j, \\ \frac{\pi}{2}, & i = 1. \end{cases}$$

Ортонормированная система имеет вид

$$\sqrt{\frac{2}{\pi}} U_0(t), \sqrt{\frac{2}{\pi}} U_1(t), \sqrt{\frac{2}{\pi}} U_2(t), \sqrt{\frac{2}{\pi}} U_3(t), \dots$$

5. Многочлены Лежандра $P_n(t)$, вычисляемые по правилу

$$P_{n+1}(t) = \frac{2n+1}{n+1} t P_n(t) - \frac{n}{n+1} P_{n-1}(t), \quad n = 1, 2, 3, \dots$$

при начальных условиях $P_0(t) = 1$, $P_1(t) = t$. Многочлены Лежандра ортогональны на отрезке $[-1, 1]$ с весом $\mu(t) = 1$, причем

$$(P_i(t), P_j(t)) = \int_{-1}^1 P_i(t) P_j(t) dt = \begin{cases} 0, & i \neq j, \\ \frac{2}{2n+1}, & i = j, \end{cases}$$

что позволяет построить ортонормированную систему многочленов Лежандра:

$$\sqrt{\frac{1}{2}} P_0(t), \sqrt{\frac{3}{2}} P_1(t), \sqrt{\frac{5}{2}} P_2(t), \sqrt{\frac{7}{2}} P_3(t), \dots$$

В прикладных задачах находят применение и другие системы функций: многочлены Эрмита, многочлены Лагерра, функции Радемахера, функции Хаара, функции Уолша и др. Заметим, что все перечисленные системы функций, исключая функции Радемахера, являются полными.

3.2.4. Обобщенные ряды Фурье

Напомним, что в n -мерном конечномерном линейном пространстве (т.е. в пространстве, где максимальное число линейно независимых векторов равняется n , а любые $n+1$, $n+2$ и т.д. векторов линейно зависимы) можно любой элемент этого пространства единственным образом представить в виде линейной комбинации линейно независимых векторов. Если $\{e_i\}_{i=1}^n$ — система линейно независимых векторов и x — произвольный вектор из этого пространства, то, таким образом, справедливо равенство

$$x = \sum_{i=1}^n c_i e_i, \quad (3.14)$$

известное как разложение по базису. Если векторы e_i , $i = 1, 2, \dots, n$, взаимно ортогональны, коэффициенты разложения c_i , обычно называемые координатами вектора x в базисе $\{e_i\}_{i=1}^n$, определяются соотношениями

$$c_i = (x, e_i) / \|e_i\|^2, \quad i = 1, 2, \dots, n. \quad (3.15)$$

Возвратимся теперь к функциональному пространству L_2 . Как уже отмечалось, в этом пространстве существует неограниченное число линейно независимых элементов. Пусть $\{e_i(t)\}_{i=1}^\infty$ — одна из таких систем. Пусть, далее, $f(t)$ — некоторая функция из L_2 . Тогда становится логичной мысль о возможности представить функцию $f(t)$ подобным (3.14) образом, т.е. в виде

$$f(t) \sim \sum_{i=1}^\infty c_i e_i(t), \quad (3.16)$$

где c_i — некоторые весовые коэффициенты.

Здесь, в отличие от (3.14), знак равенства заменен символом соответствия $\sim$ по той причине, что не оговорено, в каком смысле понимается соответствие между функцией $f(t)$ и стоящим справа рядом. Представляется вполне естественным знак соответствия заменить знаком равенства, если правый ряд в каком-то смысле сходится к функции $f(t)$.

Пусть $S_n(t)$ — n -я частичная сумма ряда из (3.16), т.е.

$$S_n(t) = \sum_{i=1}^n c_i e_i(t).$$

Будем говорить, что ряд $\sum_{i=1}^\infty c_i e_i(t)$ сходится по норме к функции $f(t)$, если

$$\lim_{n \rightarrow \infty} \|S_n(t) - f(t)\|^2 = 0. \quad (3.17)$$

В случае выполнения (3.17) будем писать

$$f(t) = \sum_{i=1}^\infty c_i e_i(t) \quad (3.18)$$

и говорить, что функция $f(t)$ задана разложением по системе функций $\{e_i(t)\}_{i=1}^\infty$. Если выполняется (3.18), несложно определить соответствующие этому равенству коэффициенты c_i , $i = 1, 2, \dots$. Для этого обе части (3.18) умножим скалярно на $e_k(t)$, воспользуемся свойствами скалярного произведения, справедливыми и в случае ряда, и учтем ортогональность функций $e_k(t)$ и $e_i(t)$ при $i \neq k$. В результате получим подобное (3.15) выражение

$$c_k = (f(t), e_k(t)) / \|e_k(t)\|^2, \quad k = 1, 2, \dots, \quad (3.19)$$

или

$$c_k = (f(t), e_k(t)), \quad k = 1, 2, \dots, \quad (3.20)$$

если система функций $\{e_i(t)\}_{i=1}^\infty$ ортонормирована. Коэффициенты (3.19), (3.20) называют коэффициентами Фурье.

Итак, предположим, что выполняется (3.17), хотя пока мы и не знаем, при каких условиях это выполняется. В этом случае, по нашей договоренности, справедливо соотношение (3.18).

Определение 3.17. Пусть $\{e_i(t)\}_{i=1}^\infty$ — некоторая система ортогональных (ортонормированных) функций из L_2 и $f(t) \in L_2$. Пусть, далее, выполняется (3.17). Тогда ряд $\sum_{i=1}^\infty c_i e_i(t)$, где c_i , $i = 1, 2, \dots$, — коэффициенты Фурье (3.19), (3.20), называется обобщенным рядом Фурье функции $f(t)$ по системе функций $\{e_i(t)\}_{i=1}^\infty$ и используется запись (3.18).

В этом случае будем говорить, что функция $f(t)$ представлена обобщенным рядом Фурье по системе функций $\{e_i(t)\}_{i=1}^\infty$. Если же для некоторой функции $f(t) \in L_2$ построен ряд $\sum_{i=1}^\infty c_i e_i(t)$ с коэффициентами Фурье (3.19) или (3.20), но не доказана сходимость

ряда к функции $f(t)$ или выявлено, что ряд сходится, но не к функции $f(t)$, то такой ряд принято называть *формальным рядом Фурье* функции $f(t)$ и использовать запись (3.16).

Если некоторая система функций $\{e_i(t)\}_{i=1}^{\infty}$ обеспечивает равенство (3.18) для любой функции $f(t)$ из L_2 , то такую систему функций называют *ортogonalным (ортонормированным) базисом* пространства L_2 .

3.2.5. Минимальное свойство коэффициентов Фурье

Использование обобщенного ряда Фурье как математической модели некоторого процесса $f(t) \in L_2$ при решении практических задач часто приводит к избыточно громоздким результатам. Поэтому стараются ограничиться приближенными моделями, менее точными, но позволяющими получить вполне реализуемые соотношения при обработке временных рядов. Подобная модель может быть получена, если использовать частичную сумму ряда по некоторой системе ортогональных (ортонормированных) функций $\{e_i(t)\}_{i=1}^q$. В этом случае, таким образом, приближенно представляем

$$f(t) = \sum_{i=1}^q b_i e_i(t), \quad (3.21)$$

где $b_i, i = 1, 2, \dots, q$ — некоторые неизвестные пока коэффициенты. Эти коэффициенты, очевидно, следует выбрать так, чтобы обеспечить наивысшую в некотором смысле точность приближения (часто говорят, аппроксимации) (3.21). Функцию

$$\varepsilon(t) = f(t) - \sum_{i=1}^q b_i e_i(t)$$

назовем *ошибкой аппроксимации* (3.21). Величину этой ошибки условимся характеризовать ее нормой или квадратом нормы $\|\varepsilon(t)\|^2$. Тогда коэффициенты $b_i, i = 1, 2, \dots, q$, следует выбрать в соответствии с условием

$$\|\varepsilon(t)\|^2 = \left(f(t) - \sum_{i=1}^q b_i e_i(t), f(t) - \sum_{k=1}^q b_k e_k(t) \right) \rightarrow \min_{b_i, i=1, \dots, q}, \quad (3.22)$$

обеспечивающим наивысшую точность аппроксимации.

Утверждение 3.2. Величина $\|\varepsilon(t)\|^2$ достигает наименьшего значения, если в качестве коэффициентов b_i принять коэффициенты Фурье c_i , т.е. положить $b_i = c_i, i = 1, 2, \dots, q$.

Для доказательства удобно $\|\varepsilon(t)\|^2$ записать в развернутом виде

$$\begin{aligned} \|\varepsilon(t)\|^2 &= (f(t), f(t)) - 2 \sum_{i=1}^q b_i (f(t), e_i(t)) + \sum_{i=1}^q \sum_{k=1}^q b_i b_k (e_i(t), e_k(t)) = \\ &= \|f(t)\|^2 - 2 \sum_{i=1}^q b_i (f(t), e_i(t)) + \sum_{i=1}^q b_i^2 \|e_i(t)\|^2. \end{aligned} \quad (3.23)$$

Запишем необходимое условие минимума величины (3.23)

$$\frac{\partial}{\partial b_k} \|\varepsilon(t)\|^2 = -(f(t), e_k(t)) + b_k \|e_k(t)\|^2 = 0, \quad k = 1, 2, \dots, q, \quad (3.24)$$

откуда непосредственно следует

$$b_k = (f(t), e_k(t)) / \|e_k(t)\|^2 = c_k, \quad k = \overline{1, q}.$$

Так как $\nabla^2 \|\varepsilon(t)\|^2 = 2E > 0$, т.е. матрица вторых производных $\nabla^2 \|\varepsilon(t)\|^2$ является положительно определенной, полученный результат действительно является решением задачи (3.22).

Итак, аппроксимация (3.21) оказывается наилучшей в смысле минимума нормы ошибки, если в качестве весовых коэффициентов b_i использовать коэффициенты Фурье c_i . В этом и проявляется свойство минимальности коэффициентов Фурье. Заметим, что система уравнений (3.24) приобрела простейший вид, при котором каждое из уравнений системы содержит лишь одну неизвестную величину благодаря ортогональности системы функций $\{e_i(t)\}_{i=1}^q$. Если бы эти функции были неортогональными, система уравнений, полученная в соответствии с условием минимума величины (3.23), не распалась бы на q независимых уравнений, каждое из уравнений содержало бы все q неизвестных, и решение системы не свелось бы к коэффициентам Фурье. Найденное решение обладает еще одним замечательным качеством: если по каким-либо причинам количество слагаемых q в аппроксимирующей модели (3.21) оказалось недостаточным и его следует увеличить, то все коэффициенты не надо пересчитывать, а нужно вычислить лишь новые коэффициенты $b_k = c_k, k = q + 1$,

$q + 2$ и т.д., сохранив все предыдущие коэффициенты. Если бы использовалась неортогональная система функций, пришлось бы пересчитывать все коэффициенты.

Найдем минимальное значение величины (3.23) — $\min \|\epsilon(t)\|^2$. С этой целью представим

$$\|\epsilon(t)\|^2 = \|f(t)\|^2 - \sum_{i=1}^q b_i(f(t), e_i(t)) + \sum_{i=1}^q b_i(-f(t), e_i(t)) + b_i \|e_i(t)\|^2.$$

Эта величина в соответствии с доказанным будет минимальна, если выполняется (3.24) и $b_i = c_i$. Но тогда

$$\min \|\epsilon(t)\|^2 = \|f(t)\|^2 - \sum_{i=1}^q c_i(f(t), e_i(t)) = \|f(t)\|^2 - \sum_{i=1}^q c_i^2 \|e_i(t)\|^2. \quad (3.25)$$

Так как $\min \|\epsilon(t)\|^2 \geq 0$, то из (3.25) следует

$$\|f(t)\|^2 \geq \sum_{i=1}^q c_i^2 \|e_i(t)\|^2 \quad (3.26)$$

при ортогональной системе функций $\{e_i(t)\}_{i=1}^q$ и

$$\|f(t)\|^2 \geq \sum_{i=1}^q c_i^2 \quad (3.27)$$

при ортонормированной системе $\{e_i(t)\}_{i=1}^q$. Выражения (3.26), (3.27) называются *неравенствами Бесселя*.

Рассмотрим состоящий из неотрицательных членов числовой ряд $\sum_{i=1}^{\infty} c_i^2 \|e_i(t)\|^2$. Так как в силу (3.26) частичные суммы этого ряда ограничены, то ряд сходится и, следовательно, справедливы общие результаты

$$\|f(t)\|^2 \geq \sum_{i=1}^{\infty} c_i^2 \|e_i(t)\|^2$$

при ортогональной системе функций $\{e_i(t)\}_{i=1}^{\infty}$ и

$$\|f(t)\|^2 \geq \sum_{i=1}^{\infty} c_i^2$$

при ортонормированной системе $\{e_i(t)\}_{i=1}^{\infty}$. Как следствие, из сходимости ряда имеем $c_i^2 \|e_i(t)\|^2 \rightarrow 0$ при $i \rightarrow \infty$ для ортогональной системы функций и $c_i^2 \rightarrow 0$ при $i \rightarrow \infty$ для ортонормированной системы.

3.2.6. Сходимость обобщенных рядов Фурье

Прежде чем обсуждать проблему сходимости рядов Фурье, рассмотрим небольшой пример, не имеющий практической значимости, но иллюстрирующий технику представления (аппроксимации) некоторой функции $f(t) \in L_2$ частичной суммой обобщенного ряда Фурье. Прежде всего отметим, что функция $f(t)$ и используемая ортогональная (ортонормированная) система функций $\{e_i(t)\}_{i=1}^q$ могут быть определены на разных множествах. Чтобы воспользоваться изложенным аппаратом, их нужно «привести» к единому множеству определения. Это можно сделать, например, так.

Пусть функция $f(t)$ определена на отрезке $[a, b]$, т.е. $t \in [a, b]$, а система функций $\{e_i(\tau)\}_{i=1}^q$ ортогональна на отрезке $[c, d]$, т.е. $\tau \in [c, d]$. В качестве аргумента системы функций записана переменная τ именно для того, чтобы подчеркнуть тот факт, что переменные t и τ принадлежат разным множествам. Чтобы привести их к единому множеству, представим

$$t = a + \lambda(b - a), \tau = c + \lambda(d - c), \lambda \in [0, 1].$$

Выразив λ , например, из первого соотношения и подставив во второе, найдем $\tau = c + (t - a)(d - c)/(b - a)$. Если теперь в составе функций $\{e_i(\tau)\}_{i=1}^q$ аргумент τ заменить в соответствии с последним выражением, получим систему функций $\{e_i^*(t)\}_{i=1}^q$, ортогональную на множестве $[a, b]$, на котором определена и функция $f(t)$. Если же подобным образом выразить $t = a + (\tau - c)(b - a)/(d - c)$ и это значение подставить в выражение функции $f(t)$, получим новую функцию $f^*(\tau)$, определенную на $[c, d]$, на котором ортогональна исходная система функций $\{e_i(\tau)\}_{i=1}^q$. Оба приема, разумеется, эквивалентны в смысле получения одного и того же результата и позволяют после такой замены аргументов воспользоваться изложенными результатами.

После сделанного замечания рассмотрим следующий пример. Пусть мы хотим функцию $f(t) = 1 + t$, $t \in [0, 2]$, представить в виде

частичной суммы обобщенного ряда Фурье по многочленам Лежандра, ограничив приближение первыми тремя слагаемыми. В данном случае, таким образом, $e_1(\tau) = P_0(\tau) = 1$, $e_2(\tau) = P_1(\tau) = \tau$, $e_3(\tau) = P_2(\tau) = 3\tau^2/2 - 1/2$, причем $\tau \in [-1, 1]$, т.е. $[a, b] = [0, 2]$, $[c, d] = [-1, 1]$. Переопределим функции Лежандра так, чтобы они оказались ортогональными на отрезке $[0, 2]$. С этой целью выражаем $\tau = t - 1$ и получаем $P_0^*(t) = 1$, $P_1^*(t) = t - 1$, $P_2^*(t) = 3t^2/2 - 3t + 1$. Теперь нам необходимо найти коэффициенты Фурье c_0, c_1, c_2 , которые с наивысшей, как мы выяснили, точностью позволят представить $f(t) \approx c_0 P_0^*(t) + c_1 P_1^*(t) + c_2 P_2^*(t)$. Эти коэффициенты находятся в соответствии с (3.19):

$$c_i = (f(t), P_i^*(t)) / \|P_i^*(t)\|^2, \quad i = 0, 1, 2.$$

Имеем:

$$\|P_0^*(t)\|^2 = \int_0^2 P_0^{*2}(t) dt = 2; \quad \|P_1^*(t)\|^2 = \int_0^2 P_1^{*2}(t) dt = 2/3;$$

$$\|P_2^*(t)\|^2 = \int_0^2 P_2^{*2}(t) dt = 2/5;$$

$$(f(t), P_0^*(t))^2 = \int_0^2 (1+t) dt = 4; \quad (f(t), P_1^*(t))^2 = \int_0^2 (1+t)(t-1) dt = 2/3;$$

$$(f(t), P_2^*(t))^2 = \int_0^2 (1+t)(3t^2/2 - 3t + 1) dt = 0$$

и, следовательно, $c_0 = 2, c_1 = 1, c_2 = 0$, что приводит к окончательному результату: $f(t) = c_0 P_0^*(t) + c_1 P_1^*(t) = 1 + t$. В этом простейшем случае, как видим, с помощью всего лишь двух слагаемых ряда Фурье удалось совершенно точно описать функцию $f(t) = 1 + t$, что, еще раз подчеркнем, объясняется простотой задачи. В общем случае такой «блестящий» результат не достигается, но при удачном выборе системы функций $\{e_i(\tau)\}_{i=1}^q$ и достаточно большом их количестве q можно сколь угодно близко к нему подойти.

Определение 3.18. Ортонормированная (ортогональная) система функций $\{e_i(\tau)\}_{i=1}^\infty$ называется *замкнутой*, если любую функцию $f(t) \in L_2$ можно по норме этого пространства приблизить

с любой точностью линейными комбинациями конечного числа элементов этой системы.

Утверждение 3.3. Если система функций $\{e_i(\tau)\}_{i=1}^\infty$ замкнута, то для любой функции $f(t) \in L_2$ неравенства Бесселя (3.26), (3.27) переходят в точные равенства соответственно

$$\|f(t)\|^2 = \sum_{i=1}^\infty c_i^2 \|e_i(t)\|^2, \quad \|f(t)\|^2 = \sum_{i=1}^\infty c_i^2,$$

называемые *равенствами Парсеваля*.

Ограничим доказательство вторым из этих равенств, относящимся к ортонормированной системе функций. Из определения замкнутой системы следует, что при $\forall \epsilon > 0 \exists q > 0$ такое, что

$$0 \leq \|f(t)\|^2 - \sum_{i=1}^q c_i^2 < \epsilon. \quad (3.28)$$

Но это эквивалентно условию

$$\lim_{q \rightarrow \infty} \left(\|f(t)\|^2 - \sum_{i=1}^q c_i^2 \right) = 0,$$

так как при возрастании q сумма в (3.28) может только возрастать. Последнее эквивалентно второму из записанных равенств Парсеваля.

Утверждение 3.4. Если система функций $\{e_i(\tau)\}_{i=1}^\infty$ замкнута, то формальный ряд Фурье для любой функции $f(t) \in L_2$ сходится по норме пространства L_2 к этой функции, т.е.

$$\lim_{q \rightarrow \infty} \left\| f(t) - \sum_{i=1}^q c_i e_i(t) \right\| = 0.$$

Действительно, в соответствии с (3.25) для ортонормированных функций имеем

$$\|f(t) - \sum_{i=1}^q c_i e_i(t)\|^2 = \|f(t)\|^2 - \sum_{i=1}^q c_i^2.$$

Так как при возрастании q правая часть последнего равенства стремится к нулю, приходим к данному утверждению.

Существует еще одно очень важное свойство функциональных пространств, которое мы приведем без доказательства.

Утверждение 3.5. В пространстве L_2 всякая полная ортогональная (ортонормированная) система функций является замкнутой, и наоборот.

И вообще, в функциональных пространствах справедливы выводы: если ортогональная (ортонормированная) система функций $\{e_i(t)\}_{i=1}^{\infty}$ является полной, то она оказывается замкнутой; с помощью линейной комбинации ее элементов с любой точностью можно представить произвольную функцию $f(t) \in L_2$; формальный ряд Фурье функции $f(t)$ сходится по норме к самой функции; выполняется равенство Парсеваля. Такая система является ортогональным (ортонормированным) базисом в пространстве L_2 . Соответственно, если выполняется равенство Парсеваля, то такая система функций является базисом в пространстве L_2 , т.е. выполнение равенства Парсеваля эквивалентно сходимости формального ряда Фурье функции $f(t)$ к этой функции.

3.2.7. Ортогональные (ортонормированные) системы дискретных функций

При определении временного ряда было указано, что он формируется на основе наблюдений, проводимых в дискретные, как правило равноотстоящие, моменты времени $t_1, t_2, \dots, t_N$. Это в явном виде отражено в модели тренда (3.2). Изложенные же выше основы теории обобщенных рядов Фурье используют концепцию квадратично интегрируемых непрерывных функций с интегральным определением скалярного произведения, нормы, метрики и других основанных на этих понятиях характеристик. Поэтому непосредственно использовать полученные результаты для математического описания временных рядов принципиально невозможно. Однако можно построить системы функций, определенных на дискретных множествах значений их аргумента, которые обладают свойствами и характеристиками, подобными присущим квадратично интегрируемым функциям.

Итак, рассмотрим множество функций, определенных в точках $t_1, t_2, \dots, t_N$. Условимся символом $f(t_i)$, $i = 1, 2, \dots, N$, обозначать некоторую из них. Величину

$$(f(t_i), u(t_i)) = \sum_{i=1}^N f(t_i)u(t_i) \quad (3.29)$$

назовем *скалярным произведением* функций $f(t_i)$ и $u(t_i)$. *Нормой* функции $f(t_i)$ назовем число

$$\|f(t_i)\| = \sqrt{(f(t_i), f(t_i))} = \sqrt{\sum_{i=1}^N f^2(t_i)}. \quad (3.30)$$

Под *расстоянием* между функциями $f(t_i)$ и $u(t_i)$ условимся понимать величину

$$\begin{aligned} \rho(f(t_i), u(t_i)) &= \|f(t_i) - u(t_i)\| = \sqrt{(f(t_i) - u(t_i), f(t_i) - u(t_i))} = \\ &= \sqrt{\sum_{i=1}^N (f(t_i) - u(t_i))^2}. \end{aligned} \quad (3.31)$$

Можно убедиться, что все введенные величины удовлетворяют соответствующим аксиомам.

Функции $f(t_i)$ и $u(t_i)$ будем называть *ортогональными*, если $(f(t_i), u(t_i)) = 0$, и *ортонормированными*, если, помимо условия ортогональности, выполняется нормировка $\|f(t_i)\| = \|u(t_i)\| = 1$. Обратим внимание на то, что все введенные определения подобны их аналогам из пространства L_2 , но интегральные операции заменены конечными суммами. В выражении (3.29) скалярного произведения могут быть предусмотрены весовые коэффициенты, моделирующие функцию $\mu(t)$ в определении (3.9).

Систему функций $\{e_k(t_i)\}_{k=0}^q$, где $i = 1, 2, \dots, N$, будем называть *ортогональной* на множестве $\{t_1, t_2, \dots, t_N\}$ дискретных значений ее аргумента, если любые две функции $e_k(t_i)$ и $e_j(t_i)$ этой системы при $k \neq j$ ортогональны, и *ортонормированной*, если

$$(e_k(t_i), e_j(t_i)) = \delta_{k,j} = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases}$$

где $\delta_{k,j}$ — дельта-символ Кронекера.

Все введенные понятия удобно переформулировать в терминах линейных конечномерных пространств. Действительно, некоторую функцию $f(t_i)$, определенную на множестве дискретных

значений ее аргумента t_i , $i = 1, 2, \dots, N$, можно отождествить с N -мерным арифметическим вектором f , компоненты которого равны соответствующим значениям функции $f(t_i)$, т.е.

$$f^T = [f(t_1) f(t_2) \dots f(t_N)] \in R^N. \quad (3.32)$$

Тогда в соответствии с обычными векторными операциями выразим

$$(f(t_i), u(t_i)) = (f, u) = f^T u = u^T f, \quad (3.33)$$

$$\|f(t_i)\| = \sqrt{(f, f)} = \sqrt{f^T f}, \quad (3.34)$$

$$\rho(f(t_i), u(t_i)) = \|f - u\| = \sqrt{(f - u, f - u)} = \sqrt{(f - u)^T (f - u)}. \quad (3.35)$$

Пусть теперь $\{e_k(t_i)\}_{k=1}^N$, $i = 1, 2, \dots, N$, — это N линейно независимых функций, определенных в N дискретных точках. Тогда N векторов $e_1, e_2, \dots, e_N$, сформированных по принципу (3.32), также линейно независимы в пространстве R^N и, так как пространство R^N — это N -мерное, образуют в нем базис. Любую функцию $f(t_i)$, $i = 1, 2, \dots, N$, можно единственным образом представить в виде линейной комбинации этих функций:

$$f(t_i) = \sum_{k=1}^N \beta_k e_k(t_i), \quad (3.36)$$

причем весовые коэффициенты разложения (координаты) однозначно находятся из системы линейных алгебраических уравнений

$$f = \sum_{k=1}^N \beta_k e_k, \quad (3.37)$$

являющейся совместной и определенной. Если ввести обозначения $E = [e_1 e_2 \dots e_N]$, $\beta = [\beta_1 \beta_2 \dots \beta_N]^T$, система (3.37) приобретает вид $f = E\beta$, откуда следует ее решение

$$\beta = E^{-1}f. \quad (3.38)$$

Поиск этого решения существенно упрощается, если функции $e_k(t_i)$, а следовательно, и векторы e_k ортогональны (ортонормированы), т.е. $e_k^T e_j = 0$ при $k \neq j$. Такие функции, как уже отме-

чалось, линейно независимы и образуют ортогональный (ортонормированный) базис. В этом случае достаточно обе части равенства (3.37) скалярно умножить на e_j , чтобы получить

$$\beta_j = (e_j, f) / \|e_j\|^2 = e_j^T f / e_j^T e_j \quad (3.39)$$

при ортогональных векторах и

$$\beta_j = (e_j, f) = e_j^T f \quad (3.40)$$

при ортонормированных векторах. Последний результат непосредственно следует и из (3.38), так как в этом случае матрица E оказывается ортогональной и для нее $E^{-1} = E^T$. Полученные выражения обладают явным вычислительным преимуществом перед общим результатом (3.38), не будучи связанными с трудоемкими операциями обращения матрицы. Обратим внимание на внешнее совпадение соотношений (3.39), (3.40) и ранее полученных определений коэффициентов Фурье (3.19), (3.20). Поэтому и величины (3.39), (3.40) обычно называют *коэффициентами Фурье*.

Возвратимся теперь к математической модели (3.1), (3.2) временного ряда. Предположим, что в качестве функций $\varphi_k(t_i)$ выбраны ортогональные (ортонормированные) функции $e_k(t_i)$, $k = 0, 1, \dots, q$. Модель тренда должна быть достаточно простой, чтобы работа с ней не порождала дополнительные трудности, но и не настолько элементарной, чтобы оказаться не адекватной изучаемому экономическому явлению. Поэтому соблюдается условие $q + 1 < N$. В этом случае система $\{e_k(t_i)\}_{k=0}^q$ уже не может быть базисом в N -мерном пространстве R^N , в силу чего произвольную функцию $f(t_i)$, $i = 0, 1, \dots, N$, нельзя абсолютно точно задать разложением по этим функциям. Возможно приближение

$$f(t_i) \approx \sum_{k=0}^q c_k e_k(t_i), \quad i = 1, 2, \dots, N, \quad \text{или} \quad f \approx \sum_{k=0}^q c_k e_k. \quad (3.41)$$

Как и в пространстве L_2 , коэффициенты разложения должны обеспечить условие

$$J = \|f - \sum_{k=0}^q c_k e_k\|^2 \rightarrow \min_{c_k}.$$

Эта задача подобна рассмотренной ранее задаче (3.22), решается аналогичным образом и приводит к тому же результату:

$$c_k = (f, e_k) / \|e_k\|^2 = \frac{\sum_{i=1}^N f(t_i) e_k(t_i)}{\sum_{i=1}^N e_k^2(t_i)}, \quad k=0, 1, \dots, q, \quad (3.42)$$

при ортогональных функциях и

$$c_k = (f, e_k) = \sum_{i=1}^N f(t_i) e_k(t_i), \quad k=0, 1, \dots, q, \quad (3.43)$$

при ортонормированных функциях. Точность приближения выражается подобным (3.25) образом:

$$\min \|f - \sum_{k=0}^q c_k e_k\|^2 = \|f\|^2 - \sum_{k=0}^q c_k^2 \|e_k\|^2. \quad (3.44)$$

Обратим внимание на одно важное обстоятельство, касающееся результатов (3.42), (3.43). По существу, используемая целевая функция J та же, что ранее использовалась нами в методе наименьших квадратов при поиске оценки параметров линейной регрессионной модели. Поэтому соотношения (3.42), (3.43) можно рассматривать как обычные МНК-оценки вектора $c = [c_0 \dots c_q]^T$, но полученные с учетом ортогональных свойств приближения (3.41). Действительно, если использовать обычный аппарат метода наименьших квадратов, то минимизация целевой функции

$J = \|f - \sum_{k=0}^q c_k e_k\|^2 = \|f - Ec\|^2$, где $E = [e_0 \ e_1 \ \dots \ e_q]$, приведет к уже

известному результату $c = (E^T E)^{-1} E^T f$. С учетом ортогональности столбцов матрицы E имеем $E^T E = \text{diag} [\|e_k\|^2]$, $k=0, 1, \dots, q$, т.е. $(E^T E)^{-1} = \text{diag} [\|e_k\|^{-2}]$. Так как $E^T f = [e_0^T f \ e_1^T f \ \dots \ e_q^T f]^T$, приходим к вычислительно более предпочтительным соотношениям (3.42). Таким образом, если модель временного ряда (3.1) имеет квазидетерминированную структуру:

$$y_i = \sum_{k=0}^q c_k e_k(t_i) + p_i, \quad i=1, \dots, N,$$

то МНК-оценки параметров модели определяются соотношениями (3.42), (3.43), в которых вектор f и его компоненты следует заменить на вектор $y = [y_1, y_2, \dots, y_N]^T$ и его компоненты соответственно. Это же относится и к (3.44).

3.2.8. Ортогональные системы тригонометрических и полиномиальных дискретных функций

Рассмотрим теперь некоторые конкретные системы функций, ортогональных на дискретных множествах значений их аргументов. Начнем рассмотрение с *тригонометрических функций*. В п. 3.2.3 отмечалось, что тригонометрическая система общего вида $1, \cos \omega t, \sin \omega t, \cos 2\omega t, \sin 2\omega t, \cos 3\omega t, \sin 3\omega t, \dots$ является ортогональной в смысле пространства L_2 на отрезке $[-\pi/\omega, \pi/\omega]$. Строго говоря, эта система является ортогональной на любом отрезке длительностью $2\pi/\omega$.

Предположим теперь, что отрезок $[0, 2\pi/\omega]$ точками $t_0, t_1, \dots, t_N$ разбит на N непересекающихся участков длительностью $2\pi/\omega N$ каждый. В этих точках, таким образом, $t_i = 2\pi i/\omega N$, $i=0, 1, \dots, N$. Рассмотрим свойства тригонометрической системы в N дискретных точках $t_1, t_2, \dots, t_N$. При этом из всей бесконечной совокупности функций, формирующих тригонометрическую систему, ограничимся первыми N функциями, т.е. функциями $1, \cos \omega t, \sin \omega t, \cos 2\omega t, \sin 2\omega t, \dots, \cos N\omega t/2$ при N четном и функциями $1, \cos \omega t, \sin \omega t, \cos 2\omega t, \sin 2\omega t, \dots, \sin (N-1)\omega t/2$ при N нечетном. В i -й точке t_i ($i=1, 2, \dots, N$) эти функции принимают соответственно значения $1, \cos 2\pi i/N, \sin 2\pi i/N, \dots, \cos \pi i$ и $1, \cos 2\pi i/N, \sin 2\pi i/N, \dots, \sin (N-1)\pi i/N$. Доказываются (например, [3]) важные для нас свойства этих функций:

$$\sum_{i=1}^N \cos 2\pi j i/N \cos 2\pi k i/N = \begin{cases} 0, & 0 \leq k \neq j \leq [N/2], \\ N/2, & 0 < k = j < N/2, \\ N, & k = j = 0 \text{ или } N/2, \end{cases} \quad (3.45)$$

$$\sum_{i=1}^N \cos 2\pi j i/N \sin 2\pi k i/N = 0, \quad k, j = 0, 1, \dots, [N/2], \quad (3.46)$$

$$\sum_{i=1}^N \sin 2\pi j i/N \sin 2\pi k i/N = \begin{cases} 0, & 0 \leq k \neq j \leq [N/2], \\ N/2, & 0 < k = j < N/2, \\ 0, & k = j = 0 \text{ или } N/2, \end{cases} \quad (3.47)$$

$$\sum_{i=1}^N \cos 2\pi k i/N = 0, \quad k=1, 2, \dots, [N/2], \quad (3.48)$$

$$\sum_{i=1}^N \sin 2\pi k i/N = 0, \quad k=1, 2, \dots, [N/2], \quad (3.49)$$

где $[N/2] = N/2$, если N четное, и $[N/2] = (N-1)/2$ при нечетном N . Соотношения (3.46), (3.49) показывают: если N четное, то функции $1, \cos 2\pi i/N, \sin 2\pi i/N, \cos 4\pi i/N, \sin 4\pi i/N, \dots, \cos (N-2)\pi i/N, \sin (N-2)\pi i/N, \cos \pi i = (-1)^i$ образуют ортогональную на дискретном множестве точек $t_1, t_2, \dots, t_N$ систему функций; при нечетном N таким свойством обладает система $1, \cos 2\pi i/N, \sin 2\pi i/N, \cos 4\pi i/N, \sin 4\pi i/N, \dots, \cos (N-1)\pi i/N, \sin (N-1)\pi i/N$. Эти функции не нормированы, а именно:

$$\|\cos 2\pi k i / N\|^2 = \sum_{i=1}^N (\cos 2\pi k i / N)^2 = N/2, \quad 0 < k \leq [N/2], \quad (3.50)$$

$$\|\sin 2\pi k i / N\|^2 = \sum_{i=1}^N (\sin 2\pi k i / N)^2 = N/2, \quad 0 < k \leq [N/2], \quad (3.51)$$

$$\|1\|^2 = N, \quad (3.52)$$

$$\|(-1)^i\|^2 = N. \quad (3.53)$$

Имея в виду (3.50)–(3.53), легко построить соответствующие нормированные функции. Но независимо от нормировки эти функции в пространстве R^N образуют базис (ортогональный или ортонормированный) и, следовательно, любая функция $f(t_i)$, $i = 1, 2, \dots, N$, может быть единственным образом представлена разложением по этому базису подобным (3.36) образом:

при четном N

$$f(t_i) = a_0 + \sum_{k=1}^{N/2-1} (b_k \cos 2\pi k i / N + c_k \sin 2\pi k i / N) + b_{N/2} (-1)^i, \quad (3.54)$$

при нечетном N

$$f(t_i) = a_0 + \sum_{k=1}^{(N-1)/2} (b_k \cos 2\pi k i / N + c_k \sin 2\pi k i / N), \quad (3.55)$$

где по аналогии с (3.42)

$$\begin{aligned} a_0 &= \frac{1}{N} \sum_{i=1}^N f(t_i), \quad b_k = \frac{2}{N} \sum_{i=1}^N f(t_i) \cos 2\pi k i / N, \\ c_k &= \frac{2}{N} \sum_{i=1}^N f(t_i) \sin 2\pi k i / N, \quad b_{N/2} = \frac{1}{N} \sum_{i=1}^N f(t_i) (-1)^i. \end{aligned} \quad (3.56)$$

Эти функции применяют для математического представления тренда, содержащего периодические составляющие. Модель тренда формируют подобным (3.41) образом со всеми последующими выводами (3.42) – (3.44).

Вторым важным для эконометрических приложений классом функций являются ортогональные на множестве $\{t_0, t_1, \dots, t_N\}$ многочлены. Обычно задача построения таких многочленов ставится так. На отрезке $[a, b]$ задана некоторая функция $f(t)$. Этот отрезок равноотстоящими точками $t_0 = a, t_1, t_2, \dots, t_N = b$ разбивается на N непересекающихся участков, так что $t_i = a + i(b-a)/N$ и, как следствие,

$$i = \frac{N(t_i - a)}{b - a}, \quad i = 0, 1, \dots, N. \quad (3.57)$$

Требуется построить $N+1$ многочленов $P_0, N(t_i), P_1, N(t_i), \dots, P_N, N(t_i)$ соответственно нулевого, первого и т.д. N -го порядков, определенных и ортогональных на множестве $\{t_0, t_1, \dots, t_N\}$ и позволяющих аналогично (3.36) представить

$$f(t_i) = \sum_{k=0}^N c_k P_{k,N}(t_i), \quad i = 0, 1, \dots, N, \quad (3.58)$$

где c_k — определяемые подобным (3.39) или (3.42) образом коэффициенты Фурье. Чтобы не быть привязанными к конкретным значениям величин $t_0, t_1, \dots, t_N$, обычно ищут многочлены $P_{k,N}[i]$, $k = 0, 1, \dots, N$, ортогональные на целочисленном множестве $\{0, 1, \dots, N\}$ значений их аргумента, а после того как такие многочлены найдены, при необходимости аргумент i в них заменяют на t_i в соответствии с (3.57). Известны [15, 16] явные выражения для многочленов $P_{k,N}[i]$, $k = 0, 1, \dots, N$:

$$P_{k,N}[i] = \sum_{m=0}^k (-1)^m C_m^k C_m^{k+m} (i/N)^m, \quad (3.59)$$

где $C_m^k = k! / m!(k-m)!$ — биномиальные коэффициенты. В частности,

$$P_{0,N}[i] = 1,$$

$$P_{1,N}[i] = 1 - \frac{2i}{N},$$

$$P_{2,N}[i] = 1 - \frac{6i}{N} + \frac{6i(i-1)}{N(N-1)},$$

$$P_{3,N}[i] = 1 - \frac{12i}{N} + \frac{30i(i-1)}{N(N-1)} - \frac{20i(i-1)(i-2)}{N(N-1)(N-2)}.$$

Многочлены (3.59), часто называемые *дискретными полиномами Лежандра—Чебышева*, обладают свойством

$$(P_{k,N}[i], P_{s,N}[i]) = \sum_{i=0}^N P_{k,N}[i] P_{s,N}[i] = \begin{cases} 0, & k \neq s, \\ \frac{(N+k+1)(N+k)^k}{(2k+1)N^k}, & k = s \end{cases} \quad (3.60)$$

т.е. многочлены ортогональны, но не нормированы. Их использование для моделирования тренда аналогично общему случаю (3.41) с естественной заменой функций $e_k(t_i)$ на ортогональные многочлены.

3.3. Предварительное (локальное) сглаживание временных рядов

Одной из часто используемых операций, направленных на уменьшение влияния случайной составляющей ряда на точность оценивания параметров тренда, является локальное сглаживание ряда. Сущность этой операции заключается в следующем.

Пусть наблюдается ряд (3.1). Зададимся некоторой величиной $m \ll N$ и рассмотрим фрагмент ряда, соответствующий отрезку $[t_{i-m}, t_{i+m}]$, т.е. совокупность величин $y_{i-m}^{i+m} = [y_{i-m} y_{i-m+1} \dots y_i \dots y_{i+m}]^T$. Здесь предполагается $i \geq m+1$ и $i \leq N-m-1$. Так как длина отрезка $[t_{i-m}, t_{i+m}]$ существенно меньше общего интервала наблюдения $[t_1, t_N]$, то на отрезке $[t_{i-m}, t_{i+m}]$ можно использовать более простую модель тренда по сравнению с моделью, например (3.2), соответствующей отрезку $[t_1, t_N]$. Обычно в качестве таковой используют многочлен невысокого порядка типа (3.3), т.е. принимают

$$y_s = \sum_{k=0}^n b_k (s-i)^k + p_s, \quad s \in [i-m, i+m], \quad n < 2m, \quad n < q, \quad (3.61)$$

или же в матричной форме

$$y_{i-m}^{i+m} = Mb + p_{i-m}^{i+m},$$

$$M = \begin{bmatrix} 1 & -m & m^2 & \dots & (-m)^n \\ 1 & -(m-1) & (m-1)^2 & \dots & (-(m-1))^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 4 & \dots & 2^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & m & m^2 & \dots & m^n \end{bmatrix}, \quad (3.62)$$

$$b = \begin{bmatrix} b_0 \\ b_1 \\ \dots \\ b_n \end{bmatrix}, \quad p_{i-m}^{i+m} = [p_{i-m} \dots p_{i+m}]^T.$$

Используя наблюдения (3.62), находят МНК-оценку β вектора b :

$$\beta = (M^T M)^{-1} M^T y_{i-m}^{i+m}, \quad (3.63)$$

с помощью которой вычисляют сглаженное значение $\hat{y}_i$ i -го уровня ряда y_i , соответствующего *середине* отрезка локального сглаживания $[t_{i-m}, t_{i+m}]$. В соответствии с моделями наблюдений (3.61), (3.62) при $s = i$ принимают

$$\hat{y}_i = M_{m+1}^T \beta = \beta_0,$$

где M_{m+1}^T — $(m+1)$ -я строка матрицы M , β_0 — первый компонент вектора β . Из выражения (3.63) следует $\beta_0 = Q_1^T y_{i-m}^{i+m}$, где Q_1^T — первая строка матрицы $(M^T M)^{-1} M^T$. Следовательно,

$$\hat{y}_i = Q_1^T y_{i-m}^{i+m} = \sum_{k=i-m}^{i+m} q_{1,i-m+1-k} y_k, \quad (3.64)$$

где $q_{1,j}$ — j -й элемент строки Q_1^T . Несложно рассчитать вектор весов Q_1^T при различных m и n . Так, при $n = 0$ имеем $M = [1 \ 1 \dots 1]^T \in R^{2m+1}$ и

$$\hat{y}_i = \frac{1}{2m+1} \sum_{k=i-m}^{i+m} y_k, \quad (3.65)$$

т.е. происходит простейшее усреднение элементов ряда на отрезке локального сглаживания. Этот алгоритм принято называть *скользящим средним*. При $n = 1$ находим

$$M = \begin{bmatrix} 1 & -m \\ 1 & -(m-1) \\ \dots & \dots \\ 1 & m \end{bmatrix}, \quad M^T M = \begin{bmatrix} 2m+1 & 0 \\ 0 & \sum_{s=-m}^m s^2 \end{bmatrix},$$

$$M^T y_{i-m}^{i+m} = \begin{bmatrix} \sum_{s=-m}^m y_{i+s} \\ \sum_{s=-m}^m s y_{i+s} \end{bmatrix}$$

и после вычисления β_0 приходим к тому же результату (3.65), т.е. скользящее среднее является алгоритмом локального сглаживания как для многочлена нулевого, так и первого порядка. Можно показать, что для многочленов второго и третьего порядков также будет общий алгоритм локального сглаживания и т.д. Приведенная далее машинная распечатка (рис. 3.1) процедуры вычисления весовых коэффициентов локального сглаживания для различных m и n позволяет в этом убедиться.

Нетрудно оценить эффективность локального сглаживания. Предварительно условимся под этим термином понимать отношение ρ дисперсий случайной составляющей i -го уровня ряда после сглаживания и до сглаживания. В результате сглаживания величина y_i преобразуется в $\hat{y}_i$ в соответствии с (3.65), случайная составляющая ε_i которой, что видно из (3.62), оказывается равной $\varepsilon_i = Q_1^T p_{i-m}^{i+m}$. Дисперсия σ_i^2 этой величины легко вычисляется: $\sigma_i^2 = M(Q_1^T p_{i-m}^{i+m} (Q_1^T p_{i-m}^{i+m})^T) = \sigma^2 Q_1^T Q_1$, где σ^2 — дисперсия случайной составляющей исходного наблюдения y_i . Следовательно, $\rho = Q_1^T Q_1$, и эффективность локального сглаживания, таким образом, просто оценивается. Так, например, при $n = 2$ и 3 и $m = 5$ имеем $Q_1^T = [-0,086 \ 0,343 \ 0,486 \ 0,343 \ -0,086]$ и, значит, $\rho = 0,486$. Аналогично для тех же n и $m = 7$ имеем $Q_1^T = [-0,095 \ 0,143 \ 0,286 \ 0,333 \ 0,286 \ 0,143 \ -0,095]$ и $\rho = 0,333$. Вообще следует иметь в ви-

исходные обозначения: $2m+1$ — количество точек, участвующих в локальном сглаживании; n — порядок сглаживающего многочлена; $z(n,m)$ — вектор весовых коэффициентов локального сглаживания

ORIGIN:=1

$z(n,m) :=$ for $j \in 1..n+1$

for $i \in 1..2m+1$

$M_{i,j} \leftarrow (i-m-j)^{j-1}$

$Q \leftarrow M \cdot (M^T \cdot M)^{-1}$

$M1 \leftarrow Q^{(1)}$

$M1^T$

$z(2,2) = (-0.086 \ 0.343 \ 0.486 \ 0.343 \ -0.086)$

$z(2,3) = (-0.095 \ 0.143 \ 0.286 \ 0.333 \ 0.286 \ 0.143 \ -0.095)$

$z(2,4) = (-0.091 \ 0.061 \ 0.169 \ 0.234 \ 0.255 \ 0.234 \ 0.169 \ 0.061 \ -0.091)$

$z(3,2) = (-0.086 \ 0.343 \ 0.486 \ 0.343 \ -0.086)$

$z(3,3) = (-0.095 \ 0.143 \ 0.286 \ 0.333 \ 0.286 \ 0.143 \ -0.095)$

$z(3,4) = (-0.091 \ 0.061 \ 0.169 \ 0.234 \ 0.255 \ 0.234 \ 0.169 \ 0.061 \ -0.091)$

$z(4,2) = (0 \ 0 \ 1 \ 0 \ 0)$

$z(4,3) = (0.022 \ -0.13 \ 0.325 \ 0.567 \ 0.325 \ -0.13 \ 0.022)$

$z(4,4) = (0.035 \ -0.128 \ 0.07 \ 0.315 \ 0.417 \ 0.315 \ 0.07 \ -0.128 \ 0.035)$

$z(5,2)$ не существует, так как число $n+1$ оцениваемых параметров превышает количество $2m+1$ используемых наблюдений и, как следствие, матрица $M^T \cdot M$ является вырожденной

$z(5,3) = (0.022 \ -0.13 \ 0.325 \ 0.567 \ 0.325 \ -0.13 \ 0.022)$

$z(5,4) = (0.035 \ -0.128 \ 0.07 \ 0.315 \ 0.417 \ 0.315 \ 0.07 \ -0.128 \ 0.035)$

и так далее.

Рис. 3.1. Результат машинного вычисления весовых коэффициентов локального сглаживания

ду, что эффективность локального сглаживания возрастает с увеличением m и падает с ростом n . При практических расчетах целесообразно ограничиваться $n \leq 2$ и $m \leq 4$. Дополнительно обратим внимание на то, что при изложенной организации локального сглаживания несглаженными оказываются первые и последние m элементов исходного временного ряда. Этот недостаток можно устранить, применив на этих участках алгоритмы типа рекуррентного метода наименьших квадратов.

3.4. Линейное прогнозирование структурно детерминированных рядов

Приступим теперь к рассмотрению одного из наиболее важных вопросов анализа временных рядов — прогнозированию их последующих значений по результатам наблюдения за ними на некотором фиксированном отрезке времени. Итак, полагаем, что в нашем распоряжении имеются N наблюдений, математически отображаемых моделями (3.1), (3.2). Базисные функции $\varphi_k(t_i)$, $k = 0, 1, \dots, q$, $i = 1, 2, \dots, N$, выбраны или в классе полиномиальных функций, или в более широком классе ортогональных функций. Ряд может быть подвергнут локальному сглаживанию, но мы сохраняем исходные обозначения (3.1). Задача, как уже отмечалось, заключается в поиске наилучшей в некотором смысле оценки $\hat{y}_m$ ненаблюдаемой величины y_m , $m > N$, по результатам наблюдений $y = [y_1 \ y_2 \ \dots \ y_N]^T$.

Прогнозируемое значение ряда y_m будем искать в классе линейных операторов, позволяющем представить

$$\hat{y}_m = W^T y, \quad (3.66)$$

где W — пока неизвестный вектор весовых коэффициентов. Этот вектор попытаемся найти таким образом, чтобы точность прогнозирования, понимаемая в далее определяемом смысле, оказалась наивысшей. Чтобы конкретизировать содержание последней фразы, обозначим символом

$$\eta = y_m - \hat{y}_m \quad (3.67)$$

ошибку прогнозирования и более детально изучим ее структуру. Для этого удобно вектор наблюдений y в соответствии с (3.1), (3.2) представить в матрично-векторной форме

$$y = \Phi a + p, \quad (3.68)$$

где

$$\Phi = \begin{bmatrix} \varphi_0(t_1) & \varphi_1(t_1) & \dots & \varphi_q(t_1) \\ \varphi_0(t_2) & \varphi_1(t_2) & \dots & \varphi_q(t_2) \\ \dots & \dots & \dots & \dots \\ \varphi_0(t_N) & \varphi_1(t_N) & \dots & \varphi_q(t_N) \end{bmatrix}, \quad a = \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_q \end{bmatrix}, \quad p = \begin{bmatrix} p_0 \\ p_1 \\ \dots \\ p_N \end{bmatrix}.$$

Тогда

$$\eta = y_m - W^T(\Phi a + p).$$

Теперь сделаем важное допущение: будем полагать, что модель тренда (3.2), введенная выше на отрезке $[t_1, t_N]$, справедлива и на множестве $[t_1, t_m]$. Это позволяет представить

$$y_m = \varphi_m^T a + p_m,$$

где

$$\varphi_m^T = [\varphi_0(t_m) \ \varphi_1(t_m) \ \dots \ \varphi_q(t_m)].$$

С использованием этого допущения записываем

$$\eta = (\varphi_m^T - W^T \Phi) a + p_m - W^T p. \quad (3.69)$$

Из (3.69) следует, что ошибка прогнозирования содержит детерминированную и стохастическую составляющие. Первая из них определяет среднее значение ошибки прогнозирования и, так как вектор a неизвестен, также является величиной неизвестной. Так как вектор W пока не найден, потребуем

$$\varphi_m - \Phi^T W = 0, \quad (3.70)$$

где 0 — нулевой $(q+1)$ -вектор. Соотношение (3.70) обеспечивает равенство $M\{\eta\} = 0$ и, следовательно, является условием несмещенности. Относительно компонентов вектора W оно представляет собой систему линейных алгебраических уравнений, содержащую $q+1$ уравнений с N неизвестными. Так как $q+1 < N$, то при выполнении условий совместности Кронекера–Капелли система оказывается неопределенной, т.е. имеет бесконечное число решений. Для поиска конкретного решения следует задать до-

полнительное условие, которому это решение должно удовлетворять. За таковое примем следующее.

Если выполняется (3.70), то ошибка прогнозирования (3.69) содержит только случайную составляющую. В этом случае точность прогнозирования можно характеризовать дисперсией ошибки прогнозирования $D\{\eta\} = M\{\eta^2\}$, и точность прогнозирования окажется наивысшей, если вектор W , помимо выполнения условий (3.70), обеспечит минимальное значение дисперсии $D\{\eta\}$. В результате приходим к следующей оптимизационной задаче поиска вектора W :

$$M\{\eta^2\} = \sigma^2(1 + W^T W) \rightarrow \min_{W \in X \subset R^N}, \quad (3.71)$$

$$X = \{W \in R^N: \varphi_m^T - \Phi^T W = 0\}. \quad (3.72)$$

Смысл задачи (3.71), (3.72), таким образом, следующий: на множестве решений системы (3.70) найти то, которое минимизирует дисперсию ошибки прогнозирования. Подобная задача встречалась ранее при доказательстве теоремы Маркова и, как отмечалось, решается методом неопределенных множителей Лагранжа. Поэтому, не прибегая к подробным комментариям, ограничимся основными этапами решения.

Функция Лагранжа

$$L(W, \lambda) = \sigma^2 W^T W + 2\lambda^T (\varphi_m - \Phi^T W).$$

Необходимые условия рассматриваемого условного минимума

$$\sigma^2 W - \Phi \lambda = 0_N, \quad \varphi_m - \Phi^T W = 0_{q+1}.$$

Решение:

$$W = \Phi(\Phi^T \Phi)^{-1} \varphi_m. \quad (3.73)$$

Алгоритм прогнозирования:

$$\hat{Y}_m = W^T y = \varphi_m^T (\Phi^T \Phi)^{-1} \Phi^T y. \quad (3.74)$$

Таким образом, наилучшее линейное прогнозирование временного ряда в сформулированных условиях заключается в линейном преобразовании вектора наблюдений y линейным опера-

тором, отождествляемым с вектором (3.73). Легко обнаружить, что множитель $(\Phi^T \Phi)^{-1} \Phi^T y$ в (3.74) представляет собой не что иное, как МНК-оценку $\hat{a}$ вектора параметров a , найденную по наблюдениям y . Поэтому кратко можно записать: $\hat{Y}_m = \varphi_m^T \hat{a}$. Точность оптимального прогнозирования определяется минимальным значением дисперсии (3.71), которая в случае (3.73) оказывается равной

$$\sigma_\eta^2 = \min M\{\eta^2\} = \sigma^2(1 + \varphi_m^T (\Phi^T \Phi)^{-1} \varphi_m). \quad (3.75)$$

Можно, вычислив (3.74), построить доверительные интервалы для истинного значения y_m прогнозируемой величины. Пусть вначале дисперсия σ^2 известна. Так как вектор y является гауссовским, то ошибка прогнозирования η является центрированной гауссовской величиной: $\eta \sim N(0, \sigma_\eta^2)$ и, следовательно, $\eta/\sigma_\eta = (y_m - \hat{Y}_m)/\sigma_\eta \sim N(0, 1)$ — стандартная случайная гауссовская величина. При доверительной вероятности α должно выполняться $P\{|\eta/\sigma_\eta| < \gamma_\alpha\} = \alpha$, где γ_α — двухсторонняя α -квантиль стандартного гауссовского распределения, P — символ вероятности. Но тогда $-\gamma_\alpha \leq (y_m - \hat{Y}_m)/\sigma_\eta < \gamma_\alpha$ и с вероятностью α выполняется

$$\hat{Y}_m - \gamma_\alpha \sigma_\eta \leq y_m < \hat{Y}_m + \gamma_\alpha \sigma_\eta. \quad (3.76)$$

Если дисперсия σ^2 неизвестна, то, как это делалось в регрессионном анализе, вычисляется ее МНК-оценка

$$\hat{\sigma}^2 = \frac{1}{N - q - 1} \sum_{k=1}^N (y_k - \Phi_k^T \hat{a})^2 = \frac{1}{N - q - 1} \|y - \Phi \hat{a}\|^2,$$

где Φ_k^T — k -я строка матрицы Φ ,

и находится оценка дисперсии (3.75)

$$\hat{\sigma}_\eta^2 = \hat{\sigma}^2(1 + \varphi_m^T (\Phi^T \Phi)^{-1} \varphi_m).$$

Далее находится величина $\eta/\hat{\sigma}_\eta$ и средствами, подобными применявшимся в аналогичной ситуации в п. 2.3.8, доказываем, что эта величина распределена по закону Стьюдента с $N - q - 1$ степенями свободы: $\eta/\hat{\sigma}_\eta \sim t(N - q - 1)$. Но тогда приходим к подобному (3.76) результату:

$$\hat{Y}_m - \gamma_\alpha \hat{\sigma}_\eta \leq y_m < \hat{Y}_m + \gamma_\alpha \hat{\sigma}_\eta. \quad (3.77)$$

где γ_α — двухсторонняя α -квантиль распределения Стюдента с $N - q - 1$ степенями свободы.

Алгоритм прогнозирования существенно упрощается, если столбцы матрицы Φ ортогональны. Используя общий результат (3.74), несложно найти:

$$\hat{Y}_m = \sum_{k=0}^q \varphi_k(t_m) \left(\frac{\sum_{i=1}^N \varphi_k(t_i) y_i}{\sum_{i=1}^N \varphi_k^2(t_i)} \right), \quad (3.78)$$

$$\sigma_\eta^2 = \sigma^2 \left(1 + \sum_{k=0}^q \left(\frac{\varphi_k^2(t_m)}{\sum_{i=1}^N \varphi_k^2(t_i)} \right) \right). \quad (3.79)$$

Решение (3.74) легко распространяется на случай $p \sim N(0, K_p)$, $M\{p_m p\} = 0$, т.е. на случай коррелированного гауссовского вектора p , но не коррелированного с величиной p_m . Соответствующие величины оказываются равными:

$$W = K_p^{-1} \Phi (\Phi^T K_p^{-1} \Phi)^{-1} \Phi_m,$$

$$\hat{Y}_m = \Phi_m^T (\Phi^T K_p^{-1} \Phi)^{-1} \Phi^T K_p^{-1} y = \Phi_m^T \hat{a},$$

$$\sigma_\eta^2 = \sigma^2 + \Phi_m^T (\Phi^T K_p^{-1} \Phi)^{-1} \Phi_m.$$

3.5. Рекуррентное прогнозирование структурно детерминированных рядов

Распространенный метод прогнозирования временных рядов основан на применении рекуррентных алгоритмов типа экспоненциального сглаживания. В этом методе используется полиномиальная модель тренда

$$f_i = \sum_{k=0}^q a_k t_i^k / k!$$

и в зависимости от порядка полинома q , который обычно не превышает 2, применяют тот или иной алгоритм экспоненциально-

го сглаживания. Начнем рассмотрение с простейшего случая $q = 0$, которому соответствует модель ряда

$$y_i = a_0 + p_i, \quad i = 1, 2, \dots, N. \quad (3.80)$$

Существо метода заключается в последовательном использовании вычислительного алгоритма

$$S_i = \alpha y_i + \beta S_{i-1}, \quad i = 1, 2, \dots, N, \quad \alpha + \beta = 1, \quad \alpha \in [0, 1], \quad (3.81)$$

или, что то же самое,

$$S_i = S_{i-1} + \alpha(y_i - S_{i-1}) \quad (3.82)$$

при некотором начальном условии S_0 .

Величину S_i принято называть экспоненциальной средней, параметр α — коэффициентом (параметром) сглаживания. Из (3.81) последовательно имеем:

$$\begin{aligned} S_1 &= \alpha y_1 + \beta S_0, \quad S_2 = \alpha y_2 + \beta S_1 = \alpha(y_2 + \beta y_1) + \beta^2 S_0, \quad S_3 = \\ &= \alpha(y_3 + \beta y_2 + \beta^2 y_1) + \beta^3 S_0, \dots, \end{aligned} \quad (3.83)$$

$$S_i = \alpha(y_i + \beta y_{i-1} + \beta^2 y_{i-2} + \dots + \beta^{i-1} y_1) + \beta^i S_0 = \alpha \sum_{k=1}^i \beta^{i-k} y_k + \beta^i S_0.$$

Таким образом, величина S_i представляет собой линейную комбинацию всех предшествующих i -му моменту элементов ряда и i -го элемента, причем вес каждого элемента ряда в формировании величины S_i в соответствии с показательной закономерностью, определившей название метода, становится тем меньше, чем дальше во времени этот элемент отстоит от i -го момента. Последнее можно интерпретировать как стремление алгоритма в большей степени доверять новым наблюдениям и в меньшей — более ранним, устаревающим. С ростом i количество участвующих в образовании S_i наблюдений возрастает, причем вес ранних из них становится все меньшим. В случае (3.80)

$$\begin{aligned} S_i &= a_0 \alpha \sum_{k=1}^i \beta^{i-k} + \alpha \sum_{k=1}^i \beta^{i-k} p_k + \beta^i S_0 = \\ &= (1 - \beta^i) a_0 + \alpha \sum_{k=1}^i \beta^{i-k} p_k + \beta^i S_0. \end{aligned}$$

Математическое ожидание и дисперсия этой величины оказываются соответственно равными:

$$M\{S_i\} = (1 - \beta^i) a_0 + \beta^i S_0, \quad (3.84)$$

$$D\{S_i\} = \alpha^2 M \left\{ \sum_{k=1}^i \sum_{j=1}^i \beta^{i-k} \beta^{i-j} p_k p_j \right\} = \alpha^2 \sigma^2 \sum_{j=0}^{i-1} \beta^{2j} = \frac{\sigma^2 \alpha (1 - \beta^{2i})}{2 - \alpha}. \quad (3.85)$$

Так как $\beta < 1$, то при $i \rightarrow \infty$ $M\{S_i\} \rightarrow a_0$, $D\{S_i\} \rightarrow \frac{\sigma^2 \alpha}{2 - \alpha}$. Таким образом, при достаточно больших i , что, в свою очередь, возможно при большом объеме N выборки временного ряда, величина S_i может рассматриваться как несмещенная оценка не изменяющегося во времени тренда a_0 . При этом параметр ρ , характеризующий эффективность сглаживания, оказывается равным $\rho = \frac{\alpha}{2 - \alpha} < 1$. Формально наивысшая эффективность достигается при $\alpha = 0$. Однако этот результат лишен практического содержания, так как в этом случае $S_i = S_0 = \text{const}$ и говорить о каком-либо сглаживании ряда бессмысленно.

При использовании алгоритма (3.82) возникают по крайней мере, два важных в прикладном отношении вопроса: как выбрать начальное условие S_0 и параметр сглаживания α . Однозначных ответов на эти вопросы нет. Из приведенных соотношений следует, что малые значения α обеспечивают высокую эффективность сглаживания, но при этом может недопустимо «затянуться» процесс достижения величинами $M\{S_i\}$ и $D\{S_i\}$ их предельных значений, соответствующих наилучшему сглаживанию и условию несмещенности. Большие значения α ускоряют переходный процесс, но ухудшают сглаживающие свойства алгоритма. Требуется компромиссное решение. Часто рекомендуют выбирать α в пределах $[0, 1; 0, 3]$. Но известны [16] и аргументированные возражения против этих рекомендаций. Аналогичные проблемы возникают и при выборе S_0 . В идеале хотелось бы иметь $S_0 = a_0$ и тогда из (3.84) следует $M\{S_i\} = a_0$ независимо от i . Но это практически недостижимое фантазирование, так как тренд a_0 неизвестен (иначе не было бы задачи). Иногда вполне приемлемым оказывается решение $S_0 = \frac{1}{m} \sum_{k=1}^m y_k$, где m — некоторое количество началь-

ных членов ряда. В качестве прогнозируемого значения $\hat{Y}_m$ ряда в случае (3.80), (3.82) принимается

$$\hat{Y}_m = S_N.$$

При выполнении (3.80) в качестве альтернативного метода экспоненциального сглаживания можно рассмотреть рекуррентный метод наименьших квадратов, изложенный выше. В этом случае МНК-оценка $\hat{a}_{0i}$ величины a_0 , найденная по наблюдениям $y_1, y_2, \dots, y_i$, рекуррентно выражается через оценку $\hat{a}_{0(i-1)}$, соответствующую наблюдениям $y_1, y_2, \dots, y_{i-1}$, и i -е наблюдение:

$$\hat{a}_{0i} = \hat{a}_{0(i-1)} + \frac{1}{i} (y_i - \hat{a}_{0(i-1)}), \quad i = 1, 2, \dots \quad (3.86)$$

Внешне алгоритмы (3.82) и (3.86) очень похожи. Различие проявляется в том, что вес α второго слагаемого справа в (3.82) постоянен, а в (3.86) аналогичный весовой коэффициент $\frac{1}{i}$ является переменным во времени. Но это казалось бы незначительное различие порождает принципиально разные последствия. Распространяв технологию преобразований (3.83) на (3.86), получим

$$\begin{aligned} \hat{a}_{01} &= y_1, \quad \hat{a}_{02} = \frac{1}{2} y_2 + \frac{1}{2} \hat{a}_{01} = \frac{1}{2} (y_1 + y_2), \\ \hat{a}_{03} &= \frac{1}{3} y_3 + \frac{2}{3} \hat{a}_{02} = \frac{1}{3} (y_1 + y_2 + y_3), \dots, \hat{a}_{0i} = \frac{1}{i} \sum_{k=1}^i y_k. \end{aligned} \quad (3.87)$$

МНК-оценка (3.87), таким образом, как и экспоненциальное среднее (3.83), является линейной комбинацией наблюдений $y_1, y_2, \dots, y_i$, но, в отличие от (3.83), в (3.87) эти наблюдения входят с одним и тем же весом $\frac{1}{i}$, т.е. алгоритм не «забывает» более ранние наблюдения и для него все наблюдения одинаково важны. Если тренд действительно не изменяется на отрезке $[1, M]$, т.е. выполняется (3.80), то это свойство алгоритма (3.86) не порождает каких-либо критических замечаний. Но если вопреки (3.80) величина a_0 изменит своё значение в какой-то момент $t_s \in (t_1, t_N)$, алгоритм (3.86) будет медленно реагировать на это изменение, так как проявляется сильное влияние ранних наблюдений, соответству-

ющих прежнему значению тренда a_0 . Алгоритм (3.82) более оперативно среагирует на это изменение, ибо в его структуре ранние наблюдения играют менее значимую роль, нежели поздние, соответствующие произошедшему изменению в тренде. В силу аналогичных причин алгоритм экспоненциального сглаживания может оказаться предпочтительнее и в случаях медленно «дрейфующего» тренда a_0 . Однако если (3.80) выполняется, то

$$M\{\hat{a}_{0i}\} = a_0 \text{ при } \forall i \in [1, N],$$

т.е. рекуррентный МНК формирует несмещенную оценку тренда, начиная с первого наблюдения y_1 , а не асимптотически, как в случае (3.84). Дисперсия

$$D\{\hat{a}_{0i}\} = (1/i^2) M\left\{\sum_{k=1}^i \sum_{j=1}^i p_k p_j\right\} = \frac{\sigma^2}{i} \rightarrow 0 \text{ при } i \rightarrow \infty,$$

что также предпочтительнее (3.84). Дополнительно обратим внимание на то, что для алгоритма (3.86) не существует проблемы выбора начального условия $\hat{a}_{00}$, так как алгоритм нечувствителен к этой величине. Но еще раз подчеркнем, несмотря на эти достоинства, рекуррентный МНК менее чувствителен к возможным эволюциям тренда, нежели алгоритм экспоненциального сглаживания. И объясняется это именно тем обстоятельством, что при формировании МНК-оценки все наблюдения принимают участие с одним и тем же весом. В то же время при экспоненциальном сглаживании более ранние по отношению к текущему моменту времени наблюдения сопровождаются существенно меньшими весами, нежели наблюдения, приближенные к этому моменту. Это в значительной степени объясняет широкую популярность экспоненциального сглаживания при обработке временных рядов, включая и задачи прогнозирования.

Алгоритм (3.82) ориентирован на случай (3.80). Несложно выявить его поведение, если в действительности тренд окажется не постоянным, а, например, линейно меняющимся:

$$f_i = a_0 + a_1 i, i = 1, 2, \dots, N. \quad (3.88)$$

Чтобы это сделать, удобно (3.82) переписать в виде

$$S_{i+1} - (1 - \alpha)S_i = \alpha f_{i+1}, i = 0, 1, \dots, N-1, \quad (3.89)$$

где принято $p_{i+1} = 0$. Выражение (3.89) представляет собой простейшее неоднородное разностное уравнение первого порядка с постоянными параметрами. Найдем его решение, соответствующее начальному условию S_0 . С этой целью предварительно получим общее решение этого уравнения. Как известно, оно складывается из общего решения S_i^* однородного уравнения

$$S_{i+1} - (1 - \alpha)S_i = 0 \quad (3.90)$$

и какого-либо частного решения S_i^{**} неоднородного уравнения (3.89). Уравнению (3.90) соответствует характеристическое уравнение $z - (1 - \alpha) = 0$ с единственным корнем $z_1 = 1 - \alpha$. Следовательно, $S_i^* = c(1 - \alpha)^i$, где c — постоянная «интегрирования». Частное решение уравнения (3.89) в случае (3.88) ищем в виде $S_i^{**} = b_0 + b_1 i$, где b_0, b_1 — некоторые пока неизвестные константы. Для их определения функции S_i^* и (3.88) подставляем в (3.89) и, рассматривая получающееся выражение как тождество, составляем систему уравнений $\alpha b_0 + b_1 = \alpha(a_0 + a_1)$, $\alpha b_1 = \alpha a_1$, из которой следует $b_0 = a_0 - a_1 \beta / \alpha$, $b_1 = a_1$. Таким образом, $S_i^* = a_0 - a_1 \beta / \alpha + a_1 i$ и общее решение уравнения (3.89) оказывается равным $S_i = c(1 - \alpha)^i + a_0 - a_1 \beta / \alpha + a_1 i$. Для решения задачи Коши используем начальное условие $S_0 = c + a_0 - a_1 \beta / \alpha$, из которого следует $c = S_0 - a_0 + a_1 \beta / \alpha$, что позволяет окончательно записать

$$S_i = (S_0 - a_0 + a_1 \beta / \alpha)(1 - \alpha)^i + a_0 - a_1 \beta / \alpha + a_1 i, \quad (3.91)$$

$$i = 1, 2, \dots, N.$$

Таким образом, если тренд изменяется по линейному закону, алгоритм экспоненциального сглаживания (3.82) формирует последовательность величин (3.91), которые отличаются от (3.88). Величину

$$\delta_i = S_i - f_i = (S_0 - a_0 + a_1 \beta / \alpha)(1 - \alpha)^i - \beta a_1 / \alpha \quad (3.92)$$

можно назвать ошибкой преобразования линейно изменяющегося тренда оператором экспоненциального сглаживания. При достаточно больших i первое слагаемое в (3.92) становится сколь угодно малым и, как говорят, в установившемся режиме, т.е. после завершения переходного процесса, будет наблюдаться установившаяся ошибка $\delta_{уст} = -\beta a_1 / \alpha$. Этот результат следует пони-

мать так: при достаточно больших i рассчитанная по правилу (3.82) величина S_i будет меньше соответствующей величины f_i на $\beta a_1/\alpha$, т.е. расчетные значения «запаздывают» относительно истинных. Аналогичный анализ можно провести и при трендах с более сложной моделью, выявив соответствующие ошибки. Наличие ошибок говорит о том, что при отличии модели тренда от (3.80) необходимо алгоритм экспоненциального сглаживания (3.82) видоизменить таким образом, чтобы он оказался более приспособленным к новой модели тренда, нежели (3.82). Такие алгоритмы разработаны и, по существу, представляют собой многократное применение изложенного алгоритма экспоненциального сглаживания. Приведем некоторые из них, не прибегая к детальному анализу.

Пусть $q = 1$, т.е. модель тренда представлена выражением (3.88). В этом случае алгоритм экспоненциального сглаживания представлен двумя рекуррентными соотношениями:

$$S_i = S_{i-1} + \alpha(y_i - S_{i-1}), \quad (3.93)$$

$$Q_i = Q_{i-1} + \alpha(S_i - Q_{i-1}), \quad i = 1, 2, \dots, N, \quad (3.94)$$

из которых второе представляет собой также алгоритм экспоненциального сглаживания, но в предположении, что «входом» для него служат «выходы» первого алгоритма. Для организации вычислений теперь нужно иметь два начальных условия. Они задаются в виде:

$$S_0 = \hat{a}_{00} - \frac{\beta}{\alpha} \hat{a}_{10}, \quad Q_0 = \hat{a}_{00} - \frac{2\beta}{\alpha} \hat{a}_{10}, \quad (3.95)$$

где $\hat{a}_{00}$, $\hat{a}_{10}$ — некоторые начальные оценки параметров модели тренда. Их можно найти, например, методом наименьших квадратов на основании небольшого числа начальных элементов обрабатываемого временного ряда. После обработки всего ряда оценки параметров тренда находят по формулам:

$$\hat{a}_{0N} = 2S_N - Q_N, \quad \hat{a}_{1N} = \alpha(S_N - Q_N)/\beta,$$

что позволяет прогноз выразить соотношением

$$\begin{aligned} \hat{Y}_m &= \hat{a}_{0N} + (m - N)\hat{a}_{1N} = \\ &= (2 + \alpha(m - N)/\beta)S_N - (1 + \alpha(m - N)/\beta)Q_N. \end{aligned} \quad (3.96)$$

Заметим, что вычисления могут быть организованы не обязательно в терминах экспоненциальных средних (3.93), (3.94), но и в терминах текущих оценок $\hat{a}_{0i}$ и $\hat{a}_{1i}$ параметров модели ряда a_0 , a_1 . Действительно, по аналогии с (3.95) имеем:

$$S_i = \hat{a}_{0i} - \beta \hat{a}_{1i}/\alpha, \quad Q_i = \hat{a}_{0i} - 2\beta \hat{a}_{1i}/\alpha.$$

Подставив эти выражения в (3.93), (3.94), получим:

$$\hat{a}_{0i} - \beta \hat{a}_{1i}/\alpha = \alpha y_i + \beta(\hat{a}_{0(i-1)} - \beta \hat{a}_{1(i-1)}),$$

$$\hat{a}_{0i} - 2\beta \hat{a}_{1i}/\alpha = \alpha(\hat{a}_{0i} - \beta \hat{a}_{1i}/\alpha) + \beta(\hat{a}_{0(i-1)} - 2\beta \hat{a}_{1(i-1)}/\alpha).$$

Из этой системы находим:

$$\hat{a}_{0i} = \hat{a}_{0(i-1)} + \hat{a}_{1(i-1)} + (1 - \beta^2)e_i, \quad e_i = y_i - \hat{a}_{0(i-1)} - \hat{a}_{1(i-1)},$$

$$\hat{a}_{1i} = \hat{a}_{1(i-1)} + \alpha e_i, \quad i = 1, 2, \dots, N. \quad (3.97)$$

Аналогичные соотношения для тренда, описываемого полиномом второго порядка $f_i = a_0 + a_1 t_i + a_2 t_i^2/2$, приобретают вид:

$$S_i = S_{i-1} + \alpha(y_i - S_{i-1}),$$

$$Q_i = Q_{i-1} + \alpha(S_i - Q_{i-1}),$$

$$R_i = R_{i-1} + \alpha(Q_i - R_{i-1}), \quad i = 1, 2, \dots, N$$

при

$$S_0 = \hat{a}_{00} - \frac{\beta}{\alpha} \hat{a}_{10} + \frac{\beta(2-\alpha)}{2\alpha^2} \hat{a}_{20},$$

$$Q_0 = \hat{a}_{00} - \frac{2\beta}{\alpha} \hat{a}_{10} + \frac{\beta(3-2\alpha)}{\alpha^2} \hat{a}_{20},$$

$$R_0 = \hat{a}_{00} - \frac{3\beta}{\alpha} \hat{a}_{10} + \frac{3\beta(4-3\alpha)}{2\alpha^2} \hat{a}_{20};$$

текущие оценки:

$$\hat{a}_{0n} = 3S_n - 3Q_n + R_n,$$

$$\hat{a}_{1n} = \alpha((6 - 5\alpha)S_n - 2(5 - 4\alpha)Q_n + (4 - 3\alpha)R_n)/2\beta^2,$$

$$\hat{a}_{2n} = \alpha^2(S_n - 2Q_n + R_n)/\beta^2;$$

прогноз

$$\hat{y}_m = \hat{a}_{0N} + (m - N)\hat{a}_{1N} + (m - N)^2\hat{a}_{2N}/2.$$

Как и в предыдущем случае, этот алгоритм может быть выражен через текущие оценки параметров модели тренда.

3.6. Анализ адекватности модели тренда временного ряда

Правильно обоснованная модель тренда в значительной степени определяет успешность решения задачи прогнозирования временного ряда. К сожалению, универсальных рекомендаций по выбору модели, гарантирующей последующий успех, нет. Интуитивные догадки, следующие из результатов начального визуального обзора ряда, умозрительные заключения, основанные на анализе природы ряда и обуславливающих его причинно-следственных явлений, опыт решения прогностических задач, квалификация исследователя — все это способствует удаче, приближает, но не обеспечивает ее. Поэтому часто оказывается целесообразным задаться несколькими моделями, а в последующем, подвергнув их надлежащему дополнительному анализу, отдать предпочтение наиболее соответствующей (адекватной) результатам наблюдений. Но для этого необходимо иметь набор критериев, выявляющих данную адекватность. В подборе таких критериев также нет общепризнанного мнения. Остановимся на одном частотном подходе. В его основе лежит исходная гипотеза о том, что случайные составляющие p_i , $i = 1, 2, \dots, N$, в составе ряда (3.1) образуют последовательность центрированных независимых нормально распределенных случайных величин. Поэтому если модель тренда выбрана удачно и правильно оценены ее параметры, остаток ряда

$$e_i = y_i - \sum_{k=0}^q \hat{a}_{kN} \phi_k(t_i), \quad i = 1, 2, \dots, N, \quad (3.98)$$

должен также образовывать последовательность типа дискретного белого шума. Поэтому адекватной признают модель, которая порождает остаточный ряд (3.98) со случайными центрированными некоррелированными нормально распределенными

элементами. Тогда проверка адекватности сводится к выявлению перечисленных свойств остаточного ряда. Это осуществляется так.

Проверка случайности элементов остаточного ряда проводится по критерию серий или критерию поворотных точек. При первом из них по результатам сравнения двух соседних элементов остаточного ряда составляется последовательность нулей и единиц. Если первая разность $\Delta e_i = e_{i+1} - e_i > 0$, то в последовательности ставится ноль, иначе — единица. Далее подсчитывается число серий $\nu(N)$, представляющих собой фрагменты последовательности, состоящие только из нулей или единиц, и продолжительность $k_{\max}$ самой длинной серии. Остаточный ряд с вероятностью 0,95 считается случайным, если

$$k_{\max} < k_0(N); \quad \nu(N) > [(2N - 1)/3 - 2\sqrt{(16N - 29)/90}].$$

Здесь $k_0(N) = 5$ при $N < 26$ и $k_0(N) = 6$ при $N > 26$; [...] — символ целой части.

При использовании менее строгого критерия поворотных точек поступают так: сравнивают элемент ряда остатков с двумя соседними; если он окажется меньше или больше их, то соответствующая точка признается поворотной; далее подсчитывается число s всех поворотных точек; если окажется

$$s > [2(N - 2)/3 - 2\sqrt{(16N - 29)/90}],$$

остаточный ряд считается состоящим из случайных элементов.

Проверка центрированности проводится с использованием t -критерия Стьюдента. С этой целью формируется статистика

$$\gamma = |m_e| \sqrt{N} / \sigma_e, \quad (3.99)$$

где

$$m_e = \frac{1}{N} \sum_{i=1}^N e_i, \quad \sigma_e = \sqrt{\frac{1}{N-1} \sum_{i=1}^N e_i^2 - m_e^2},$$

т.е. среднее значение и среднее квадратичное отклонение остаточного ряда. Далее задаются уровнем значимости α или доверительной вероятностью $1 - \alpha$ и находят $100\alpha/2$ -процентную точку $w_{100\alpha/2}$ t -распределения с $N - 1$ степенями свободы. Если окажется $\gamma > w_{100\alpha/2}$, то гипотеза о центрированности остаточного ряда

отвергается как не соответствующая экспериментальным данным с вероятностью ошибиться α . При противоположном неравенстве ряд признается центрированным с вероятностью $1 - \alpha$ правильности этого решения.

Проверка независимости уровней остаточного ряда преследует цель подтвердить отсутствие систематической составляющей в составе ряда и проводится с применением критерия Дарбина–Уотсона. В соответствии с этим критерием вычисляется величина

$$d = \frac{\sum_{k=2}^N (e_k - e_{k-1})^2}{\sum_{k=1}^N e_k^2} \approx 2 \left(1 - \frac{\sum_{k=2}^N e_k e_{k-1}}{\sum_{k=1}^N e_k^2} \right) = 2(1 - R),$$

где R — так называемый коэффициент автокорреляции первого порядка.

Таблица 3.1

Нижний d_1 и верхний d_2 пороги

Объем выборки N	Сложность модели тренда q									
	1		2		3		4		5	
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
20	1,20	1,41	1,10	1,54	1,00	1,68	1,90	1,83	0,79	1,99
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,47
100	1,65	1,69	1,63	1,72	1,61	1,71	1,59	1,76	1,57	1,78

Величина d следующим образом подвергается анализу. Прежде всего если окажется $d > 2$, то заменяют d на $d^* = 4 - d$ и последующая работа ведется с d^* по тому же алгоритму, что и с d . В рассмотрение вводятся два порога: нижний d_1 и верхний d_2 . Значения этих порогов определяются объемом выборки N , сложностью

модели тренда q , уровнем значимости α и при $\alpha = 0,05$ систематизированы в табл. 3.1 [27].

Если d (или d^*) $\in (0, d_1)$, то это является признаком сильной автокоррелированности элементов остаточного ряда, и предполагаемая модель тренда признается неадекватной. Если $d(d^*) \in (d_2, 2)$, то элементы остаточного ряда классифицируются как независимые, а модель тренда — адекватной. При $d(d^*) \in (d_1, d_2)$ однозначный вывод не делается и применяют дополнительные методы исследования.

Дополнительный анализ проводят с использованием корреляционной функции $r(m)$ остаточного ряда, которую определяют следующим образом:

$$r(m) = \frac{\sum_{k=1}^{N-m} e_k e_{k+m}}{\sum_{k=1}^N e_k^2}.$$

Значение этой функции $R = r(1)$, называемое коэффициентом автокорреляции, уже встречалось при формировании критерия Дарбина–Уотсона, и оно же используется для анализа независимости. Величина R сравнивается с порогом γ , зависящим от объема выборки и доверительной вероятности. При уровне значимости 0,05 (доверительной вероятности 0,95) значения порога содержатся в табл. 3.2 [11].

Таблица 3.2

Зависимость порога от объема выборки

Объем N	10	15	20	25	30
Порог γ	0,360	0,328	0,300	0,276	0,257

Если окажется $R > \gamma$, то принимается решение о существенной корреляции элементов остаточного ряда и, следовательно, о неадекватности модели тренда.

Иногда оказывается целесообразным применение более комплексного критерия, основанного на корреляционной функции $r(m)$. Рассчитывается коэффициент

$$G = N \sum_{j=1}^J (r(j))^2,$$

где $J \leq N/3$. Доказывается, что статистика G подчинена χ^2 -распределению с $(N - J - 1)$ степенями свободы. Тогда если окажется $G < w_{100\alpha}$, где $w_{100\alpha}$ есть 100 α -процентная точка χ^2 -распределения с $(N - J - 1)$ степенями свободы, то с вероятностью $1 - \alpha$ элементы остаточного ряда признаются некоррелированными, а если они гауссовские, то и независимыми.

Проверка на нормальное распределение остаточного ряда может проводиться многочисленными методами, разработанными в математической статистике. Остановимся на двух из них, полагая, что предыдущие тесты по анализу случайности, центрированности и независимости дали положительные результаты.

Распространенным методом проверки гипотезы о нормальном распределении является χ^2 -критерий Пирсона. В кратком изложении его сущность заключается в следующем.

Пусть $e_{\min}$ и $e_{\max}$ соответственно наименьший и наибольший элементы ряда остатков. Отрезок $[e_{\min}, e_{\max}]$ разбивают на s равных непересекающихся интервалов так, чтобы в каждый из них попало не менее пяти элементов ряда. Иногда рекомендуют $s = \ln N / \ln 2$. Символами h и N_j ($j = 1, 2, \dots, s$) обозначены соответственно длина интервала и число элементов остаточного ряда, попавших в j -й интервал. Границы j -го интервала обозначены символами $x_j, x_{j+1}, x_j^* = (x_j + x_{j+1})/2$ — центр интервала. Далее подсчитывают теоретические вероятности P_j попадания случайной величины на j -й интервал:

$$P_j = h \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp \left\{ -\frac{(x_j^*)^2}{2\sigma_e^2} \right\}$$

и строят статистику Пирсона

$$\gamma = \sum_{j=1}^s \frac{(N_j - NP_j)^2}{NP_j}.$$

Затем задаются уровнем значимости α и находят 100 α -процентную точку $w_{100\alpha}$ χ^2 -квадрат распределения с $s - 3$ степенями свободы. Если окажется $\gamma < w_{100\alpha}$, то с вероятностью $1 - \alpha$ нет оснований отвергать гипотезу о нормальном распределении эле-

ментов остаточного ряда; при противоположном неравенстве гипотеза отвергается с вероятностью α совершить ошибку.

Более простым, хотя и менее обоснованным, является метод асимметрии и эксцесса. Его сущность такова [21].

По экспериментальным данным (остаточному ряду) строятся эмпирические коэффициенты асимметрии K_a и эксцесса K_3 :

$$K_a = \frac{\frac{1}{N} \sum_{k=1}^N e_k^3}{\sqrt{\left(\frac{1}{N} \sum_{k=1}^N e_k^2 \right)^3}}, \quad K_3 = \frac{\frac{1}{N} \sum_{k=1}^N e_k^4}{\left(\frac{1}{N} \sum_{k=1}^N e_k^2 \right)^2} - 3.$$

Если эти коэффициенты близки к нулю, то появляются основания считать остаточный ряд гауссовским. Для усиления этих оснований вычисляются среднеквадратические отклонения коэффициентов [11, 21]:

$$\sigma_a = \sqrt{\frac{6(N-1)}{(N+1)(N+3)}},$$

$$\sigma_3 = \sqrt{\frac{24N(N-2)(N-3)}{(N-1)^2(N+3)(N+5)}}.$$

Если $|K_a| \leq 1,5\sigma_a, |K_3| \leq 1,5\sigma_3$, то считают, что распределение остаточного ряда не противоречит гипотезе о нормальном распределении. Если хотя бы один из коэффициентов оказывается больше двух среднеквадратических отклонений, гипотеза о нормальности отвергается.

Важной характеристикой модели является ее точность. Существуют различные определения этого понятия. Достаточно распространенной и простой мерой точности является относительная ошибка

$$\epsilon_{\text{отн}} = \frac{1}{N} \sum_{i=1}^N \frac{|e_i|}{|y_i|} 100\%.$$

Если окажется $\epsilon_{\text{отн}} < 15\%$, то точность признается достаточной.

Глава 4

СТОХАСТИЧЕСКИЕ ВРЕМЕННЫЕ РЯДЫ

4.1. Случайные процессы (начальные определения и классификация)

Теория стохастических временных рядов базируется на соответствующей теории случайных (стохастических) процессов. Приведем некоторые исходные положения, полагая, что читатель знаком с терминологией теории вероятностей и элементами вероятностных операций.

Пусть проводится случайный эксперимент и (Ω, F, P) — соответствующее ему вероятностное пространство, $\omega \in \Omega$ — элементарное событие.

Определение 4.1. Вещественная функция $X = \varphi(\omega)$ элементарного события ω называется *случайной величиной*, если для $\forall x \in (-\infty, \infty)$ множество тех ω , для которых $\varphi(\omega) < x$, принадлежит множеству F , т.е. $\{\omega: \varphi(\omega) < x\} \subset F$, и $P(\varphi(\omega) \in (-\infty, \infty)) = 1$.

Так как для каждого события из F определена вероятность P , то условие $\{\omega: \varphi(\omega) < x\} \subset F$ означает, что при любом x может быть определена вероятность события $\varphi(\omega) < x$. Случайная величина, таким образом, представляет собой величину, значение которой до проведения эксперимента точно предсказать нельзя, однако можно указать вероятность того, что в эксперименте случайная величина X окажется меньше любого числа x .

Определение 4.2. Вещественная функция $X(t) = \varphi(\omega, t)$, где $t \in T$ и имеет смысл времени, называется *случайным*, или *стохастическим*, *процессом*, если при каждом фиксированном $t = t^*$ величина $X(t^*) = \varphi(\omega, t^*)$ является случайной величиной.

Множество T задает область определения процесса. Сам случайный процесс, что следует из его определения, можно рассматривать как параметрически заданную на множестве T случайную величину. Значение случайного процесса при фиксированном аргументе t принято называть *сечением* процесса. Если в эксперименте элементарное событие ω примет определенное значение $\omega = \omega \in F$, то функция $\varphi(\omega, t)$ оказывается детерминированной

функцией времени и называется *реализацией* (траекторией) процесса. Таким образом, случайный процесс как функцию двух переменных t и ω можно рассматривать или как семейство зависящих от параметра t случайных величин, или как семейство реализаций (содержащее, вообще говоря, бесконечное их число).

Если множество значений случайного процесса и область его определения непрерывны, то процесс принято называть *непрерывнозначным*. Если, кроме того, в отдельных реализациях отсутствуют разрывы, случайный процесс называют *непрерывным*.

Случайный процесс называют *дискретным* (разрывным, дискретным процессом с непрерывным временем), если множество

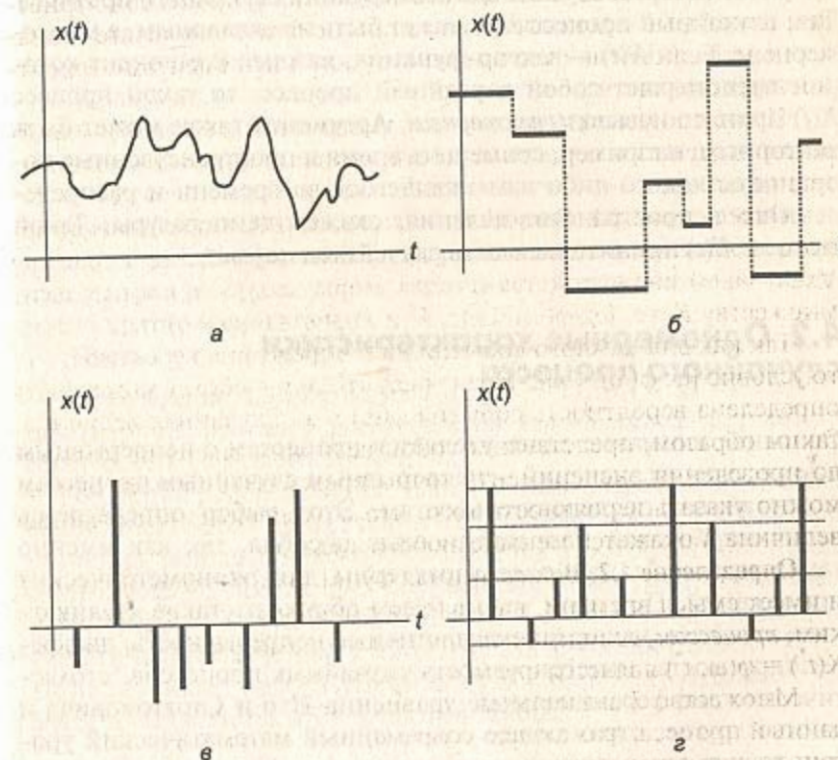


Рис. 4.1. Реализации случайного процесса:

- а — непрерывный процесс, б — дискретный процесс,
в — случайная последовательность,
г — дискретная случайная последовательность

значений представляет собой дискретное множество, а область определения T непрерывна.

Если у случайного процесса множество значений непрерывно, а область определения представляет собой конечное или счетное дискретное множество $T = \{t_1, t_2, \dots\}$, то его принято называть *случайной последовательностью*.

Наконец, случайный процесс принято называть *дискретной случайной последовательностью*, если и область определения, и множество значений являются дискретными множествами. На рис. 4.1 представлены примеры реализаций $x(t)$ различных случайных процессов $X(t)$ в соответствии с данной классификацией.

Возможны более сложные конструкции случайного процесса. Так, случайный процесс $X(t)$ может быть не скалярным, а многомерным. Если $X(t)$ — вектор-функция, каждый компонент которой представляет собой случайный процесс, то такой процесс $X(t)$ принято называть *векторным*. Аргумент t также может быть векторным: например, совмещать время и пространственные координаты какого-либо изменяющегося во времени и распределенного в пространстве явления, скажем температуры. Такой процесс $X(t)$ принято называть *случайным полем*.

4.2. Одномерные характеристики случайного процесса

Основное внимание далее уделяется процессам с непрерывным множеством их значений — непрерывным случайным процессам и случайным последовательностям. Этот выбор определяется эконометрической направленностью пособия, так как именно такие процессы наиболее характерны для эконометрических приложений. При этом мы не будем обсуждать такие «деликатные» понятия, как среднеквадратичная непрерывность, дифференцируемость, интегрируемость случайных процессов, стохастические дифференциальные уравнения Ито и Стратоновича и многое другое, отражающее современный математический уровень теории стохастических процессов, а ограничимся тем минимумом сведений, который необходим для решения основных прикладных задач. Математическое описание случайных последовательностей будем совмещать с аналогичным представлением непрерывных процессов, делая соответствующие комментарии в

случае необходимости в них. Непрерывные случайные процессы условимся обозначать прописными символами — $X(t)$, $Y(t)$ и т.п., а их реализации строчными — соответственно $x(t)$, $y(t)$ и т.п. Аналогичным образом за случайными последовательностями закрепим обозначения X_i , Y_i и т.п. ($i = 1, 2, \dots$), а за их реализациями — x_i , y_i и т.п. При этом если множество значений индекса i конечно, т.е. $i = 1, 2, \dots, N$, то случайную последовательность будем называть *стохастическим временным рядом*. Будем полагать, что как последовательность, так и ряд сформированы на основе равномерно поступающих во времени данных, т.е. $t_i = t_{i-1} + \xi$, где $\xi = \text{const}$ — так называемый *период дискретизации*. При рассмотрении характеристик, общих для непрерывных процессов и последовательностей, будем использовать единый символ $X(t)$, если это не порождает каких-либо недоразумений.

Характеристики случайных процессов принято разделять на одномерные, относящиеся к одному конкретному сечению, и многомерные, отражающие свойства процесса совместно в нескольких сечениях. Так как в сечении случайный процесс является случайной величиной, то его одномерные характеристики совпадают с аналогичными характеристиками случайных величин, известными из курса теории вероятностей, что позволяет достаточно кратко изложить их.

Пусть $t = t_i$ — фиксированное сечение процесса $X(t)$.

Определение 4.3. Функция $F_X(x; t_i) = P(X(t_i) < x)$ называется *одномерной функцией распределения вероятностей* случайного процесса $X(t)$ в момент времени t_i .

Функция распределения, таким образом, представляет собой вероятность того, что в сечении t_i случайный процесс $X(t)$ примет значение, меньшее некоторой величины x . Хотя формально она записана как функция двух переменных, ее аргументом является переменная x , а присутствие величины t_i объясняется желанием указать «адрес» сечения, к которому эта функция относится. Если не возникает недоразумений, эту величину не указывают. Характерные свойства этой функции таковы:

1. $F_X(-\infty; t_i) = 0$;
2. $F_X(\infty; t_i) = 1$;
3. $F_X(b; t_i) - F_X(a; t_i) = P(a \leq X(t_i) < b)$, где $a, b \in \mathbb{R}$;
4. $F_X(b; t_i) \geq F_X(a; t_i)$ при $b \geq a$;

5. Функция распределения непрерывна слева, т.е. $F_X(x_0; t_i) = \lim_{x \rightarrow x_0 - 0} F_X(x; t_i)$ и $P(X(t_i) < x_0) = F_X(x_0; t_i)$;

6. $P(X(t_i) = a) = 0$, если функция $F_X(x; t_i)$ в точке a непрерывна, и $P(X(t_i) = a) = \lim_{b \rightarrow a+0} F_X(b; t_i) - F_X(a; t_i)$, если в этой точке имеется разрыв первого рода.

Одномерная функция распределения, таким образом, является неубывающей, непрерывной слева с множеством значений $[0, 1]$; она имеет участки монотонного возрастания, участки постоянства и в некоторых точках может иметь разрывы первого рода.

Определение 4.4. Пусть существует функция $f_X(x; t_i)$ такая, что можно представить

$$\int_{-\infty}^x f_X(x; t_i) dx = F_X(x; t_i), \quad (4.1)$$

или при дифференцируемой функции $F_X(x; t_i)$

$$f_X(x; t_i) = \frac{d}{dx} F_X(x; t_i). \quad (4.2)$$

Тогда функция $f_X(x; t_i)$ называется одномерной *плотностью вероятностей* случайного процесса $X(t)$ в сечении t_i .

Смысл плотности вероятностей заключается в том, что с точностью до бесконечно малых высших порядков выполняется равенство

$$f_X(x; t_i) dx = P(x \leq X(t_i) < x + dx),$$

и функция $f_X(x; t_i)$ показывает, как эта вероятность распределяется вдоль оси x . В справедливости равенства несложно убедиться, если представить

$$\begin{aligned} \int_x^{x+dx} f_X(x; t_i) dx &= \int_{-\infty}^{x+dx} f_X(x; t_i) dx - \int_{-\infty}^x f_X(x; t_i) dx = \\ &= F_X(x+dx; t_i) - F_X(x; t_i) \end{aligned} \quad (4.3)$$

и функцию $F_X(x+dx; t_i)$ «линеаризовать» в окрестности точки x .

Из (4.1) следует равенство

$$\int_{-\infty}^{\infty} f_X(x; t_i) dx = 1,$$

обычно называемое *условием нормировки* плотности вероятностей. Поступая аналогичным (4.3) образом, получаем еще один важный результат:

$$\int_a^b f_X(x; t_i) dx = F_X(b; t_i) - F_X(a; t_i) = P(a \leq X(t_i) < b). \quad (4.4)$$

Наконец, полезно заметить, что плотность вероятностей $f_X(x; t_i)$ как производная от неубывающей функции $F_X(x; t_i)$ является неотрицательной функцией. Если функция $F_X(x; t_i)$ имеет разрывы, то плотность вероятностей и в этих точках формально может быть определена введением δ -функций с интенсивностями, равными величинам скачков (разрывов).

Введенные характеристики процесса $X(t)$ полностью отражают в сечении t_i его вероятностные свойства. Однако во многих случаях можно ограничиться более скромной информацией о процессе, сосредоточенной в его *числовых характеристиках*. В качестве таковых, как и для случайных величин, используют математическое ожидание (среднее значение) $m_X(t_i)$, средний квадрат $C_X(t_i)$, дисперсию $D_X(t_i)$ и среднеквадратическое отклонение $\beta_X(t_i)$. По определению имеем:

$$m_X(t_i) = M\{X(t_i)\} = \int_{-\infty}^{\infty} x f_X(x; t_i) dx;$$

$$C_X(t_i) = M\{X^2(t_i)\} = \int_{-\infty}^{\infty} x^2 f_X(x; t_i) dx;$$

$$D_X(t_i) = M\{(X^\circ(t_i))^2\} = \int_{-\infty}^{\infty} (x - m_X(t_i))^2 f_X(x; t_i) dx;$$

$$\beta_X(t_i) = \sqrt{D_X(t_i)}.$$

Здесь, как и ранее, $M\{\dots\}$ — символ усреднения, $X^\circ(t_i)$ — центрированная случайная величина. В соответствии с определениями все эти характеристики являются неслучайными числовыми величинами, причем математическое ожидание $m_X(t_i)$ представ-

ляет собой ту величину, относительно которой разбросаны отдельные реализации процесса $X(t)$ в сечении t_i , а дисперсия является мерой этого разброса. Полезно отметить ряд очевидных свойств:

$$M\left\{\sum_{j=1}^k c_j X_j(t_i)\right\} = \sum_{j=1}^k c_j M\{X_j(t_i)\},$$

где $c_j \in \mathbb{R}$, $X_j(t)$ — случайные процессы;

$$C_X(t_i) = D_X(t_i) + (m_X(t_i))^2;$$

если $g(t)$ — неслучайная функция, то

$$M\{g(t_i)\} = g(t_i), M\{Y(t_i) = g(t_i)X(t_i)\} = g(t_i)m_X(t_i), D_Y(t_i) = g^2(t_i)D_X(t_i).$$

Если процесс $X(t)$ является непрерывным и аргумент t_i пробегает все возможные значения в соответствии с областью определения T процесса, то его математическое ожидание и дисперсия могут рассматриваться как детерминированные функции времени соответственно $m_X(t)$ и $D_X(t)$. Для случайной последовательности X_i аналогичным образом получим детерминированные решетчатые функции $m_X[i] = m_X(t_i)$ и $D_X[i] = D_X(t_i)$, $i = 1, 2, \dots$.

4.3. Многомерные характеристики случайного процесса. Марковские процессы

Как уже отмечалось, одномерные характеристики случайного процесса $X(t)$ отражают его свойства в одном сечении и, следовательно, не содержат никаких данных о поведении процесса во времени, т.е. в динамике. Этой цели служат многомерные распределения — двумерные, трехмерные и т.д. Рассмотрим вначале двумерный случай.

Итак, пусть $X(t)$ — случайный процесс. Выделим какие-либо два сечения $X(t_i)$ и $X(t_j)$ этого процесса, соответствующие моментам времени t_i и t_j , и зададимся произвольными числами x_1 и x_2 , относящимися соответственно к i -му и j -му сечениям.

Определение 2.5. Функция $F_X(x_1, x_2; t_i, t_j) = P((X(t_i) < x_1) \cap (X(t_j) < x_2))$ называется *двумерной функцией распределения вероятностей* случайного процесса $X(t)$ в моменты времени t_i и t_j .

Таким образом, двумерная функция $F_X(x_1, x_2; t_i, t_j)$ представляет собой вероятность совместного выполнения двух событий: процесс $X(t)$ в момент времени t_i принимает значение, меньшее некоторой величины x_1 , а в момент t_j — меньшее x_2 . Аргументами этой функции являются переменные x_1 и x_2 , а величины t_i, t_j называют сечения, к которым эта функция относится. Основные ее свойства.

1. Функция $F_X(x_1, x_2; t_i, t_j)$ неотрицательная, неубывающая и непрерывная слева по каждому из аргументов x_1 и x_2 .

2. Если хотя бы один из аргументов стремится к $-\infty$, функция распределения вероятностей стремится к 0.

$$3. F(\infty, \infty; t_i, t_j) = 1.$$

$$4. F_X(x_1, \infty; t_i, t_j) = F_X(x_1; t_i); F_X(\infty, x_2; t_i, t_j) = F_X(x_2; t_j).$$

$$5. F(x_1, x_2; t_i, t_j) = F(x_2, x_1; t_j, t_i).$$

$$6. 0 \leq F(x_1, x_2; t_i, t_j) \leq 1 \text{ при } \forall x_1, x_2.$$

Определение 4.6. Пусть существует функция $f_X(x_1, x_2; t_i, t_j)$, такая, что можно представить

$$\int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_X(x_1, x_2; t_i, t_j) dx_1 dx_2 = F_X(x_1, x_2; t_i, t_j), \quad (4.5)$$

или при дифференцируемой функции $F(x_1, x_2; t_i, t_j)$

$$f_X(x_1, x_2; t_i, t_j) = \frac{\partial^2}{\partial x_1 \partial x_2} F(x_1, x_2; t_i, t_j). \quad (4.6)$$

Тогда функция $f_X(x_1, x_2; t_i, t_j)$ называется *двумерной плотностью вероятностей* случайного процесса $X(t)$ в моменты времени t_i и t_j .

Характерные свойства этой функции.

$$1. f_X(x_1, x_2; t_i, t_j) dx_1 dx_2 = P(x_1 \leq X(t_i) < x_1 + dx_1 \cap x_2 \leq X(t_j) < x_2 + dx_2).$$

$$2. \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_X(x_1, x_2; t_i, t_j) dx_1 dx_2 = 1 \text{ (условие нормировки).}$$

$$3. f_X(x_1, x_2; t_i, t_j) = f_X(x_2, x_1; t_j, t_i).$$

$$4. f_X(x_1, x_2; t_i, t_j) \geq 0 \text{ при } \forall x_1, x_2.$$

$$5. \int_{-\infty}^{\infty} f_X(x_1, x_2; t_i, t_j) dx_1 = f_X(x_2; t_j);$$

$$\int_{-\infty}^{\infty} f_X(x_1, x_2; t_i, t_j) dx_2 = f_X(x_1; t_i).$$

Определение 4.7. Функция

$$f_X(x_1, t_1 | x_2, t_2) = \frac{f_X(x_1, x_2; t_1, t_2)}{f_X(x_2; t_2)}, \quad (4.7)$$

где принимается $f_X(x_2; t_2) \neq 0$, называется *условной плотностью вероятностей* процесса $X(t)$ в момент времени t_1 при условии, что в момент t_2 процесс принял значение x_2 . По аналогии вводится условная плотность

$$f_X(x_2, t_2 | x_1, t_1) = \frac{f_X(x_1, x_2; t_1, t_2)}{f_X(x_1; t_1)}, \quad f_X(x_1; t_1) \neq 0. \quad (4.8)$$

Условные плотности $f_X(x_1; t_1 | x_2; t_2)$, $f_X(x_2; t_2 | x_1; t_1)$ не следует путать с одномерными плотностями $f_X(x_1; t_1)$, $f_X(x_2; t_2)$, которые часто называют в целях противопоставления *безусловными*: безусловная плотность вероятностей в одном из сечений строится независимо от поведения процесса в другом сечении; в то же время условная плотность в одном из сечений строится в предположении, что процесс во втором сечении принял некоторое фиксированное значение. Как плотность вероятностей, условная плотность обладает теми же свойствами, что и безусловная плотность вероятностей.

В соответствии с (4.7), (4.8) двумерная плотность выражается через одномерные

$$\begin{aligned} f_X(x_1, x_2; t_1, t_2) &= f_X(x_1; t_1) \times f_X(x_2; t_2 | x_1; t_1) = \\ &= f_X(x_2; t_2) \times f_X(x_1; t_1 | x_2; t_2). \end{aligned} \quad (4.9)$$

Если же поведение процесса в одном из сечений не зависит от его поведения во втором, то

$$\begin{aligned} f_X(x_1; t_1 | x_2; t_2) &= f_X(x_1; t_1), \quad f_X(x_2; t_2 | x_1; t_1) = \\ &= f_X(x_2; t_2), \quad f_X(x_1, x_2; t_1, t_2) = f_X(x_1; t_1) f_X(x_2; t_2), \end{aligned} \quad (4.10)$$

и сечения $X(t_1)$, $X(t_2)$ называют *независимыми*. При независимых сечениях двумерная плотность оказывается равной *произведению* одномерных.

Принцип определения двумерных законов распределения вероятностей и плотностей вероятностей распространяется на произвольное число n сечений случайного процесса $X(t)$, что приво-

дит к понятию соответствующих n -мерных характеристик. Рассмотрим n сечений случайного процесса, относящихся к каким-либо n моментам времени $t_1, t_2, \dots, t_n$, и зададимся n числами $x_1, x_2, \dots, x_n$. Тогда по определению принимается

$$F_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = P\left(\bigcap_{i=1}^n (X(t_i) < x_i)\right),$$

т.е. n -мерный закон распределения вероятностей представляет собой вероятность совместного выполнения n указанных событий и является функцией n аргументов $x_1, x_2, \dots, x_n$. Аналогичным образом определяют n -мерные плотности вероятностей

$$\begin{aligned} \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_X(x_1, x_2, \dots, x_n; t_1, \dots, t_n) dx_1 dx_2 \dots dx_n = \\ F_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n), \end{aligned}$$

или же при дифференцируемой функции $F_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n)$

$$\begin{aligned} \frac{\partial^n}{\partial x_1 \partial x_2 \dots \partial x_n} F_X(x_1, x_2, \dots, x_n; t_1, \dots, t_n) = \\ = f_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n). \end{aligned}$$

Эти функции обладают свойствами, подобными перечисленным выше в связи с двумерными характеристиками, и мы их опускаем. Особо отметим, что

$$\begin{aligned} f_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) &= f_X(x_1; t_1) f_X(x_2; t_2 | x_1; t_1) f_X(x_3; t_3 | x_1, x_2; t_1, t_2) \dots \\ &\dots f_X(x_n; t_n | x_1, x_2, \dots, x_{n-1}; t_1, t_2, \dots, t_{n-1}), \end{aligned} \quad (4.11)$$

где смысловое содержание условных плотностей сохраняется прежним. Если же выбранные сечения независимы, соотношение (4.11) существенно упрощается и n -мерная плотность выражается через одномерные

$$f_X(x_1, x_2, \dots, x_n; t_1, \dots, t_n) = \prod_{i=1}^n f_X(x_i; t_i). \quad (4.12)$$

Выражение (4.12) подкупает своей простотой по сравнению с общим выражением (4.11). К сожалению, его применимость ограничена случаем независимых сечений, что в практических задачах встречается не часто. Вместе с тем есть определенный класс случайных процессов, для которых представление многомерной плотности занимает промежуточное положение между (4.11) и (4.12). Такими процессами являются марковские.

Определение 4.8. Случайный процесс $X(t)$ называется *марковским*, если при любом выборе последовательности аргументов $t_1, t_2, \dots, t_n$ выполняется равенство

$$f_X(x_n; t_n | x_1, x_2, \dots, x_{n-1}; t_1, t_2, \dots, t_{n-1}) = f_X(x_n; t_n | x_{n-1}; t_{n-1}), \quad (4.13)$$

т.е. условная плотность процесса в n -м сечении при фиксированных значениях процесса в предыдущих $(n-1)$ сечениях зависит от значения процесса только в предыдущем $(n-1)$ -м сечении и не зависит от «предыстории». Это означает, что если марковский процесс в какой-то момент времени t принял значение x , то распределение вероятностей в любой последующий момент времени $t > t$ определяется его значением x именно в этот момент t и совершенно не зависит от его реализаций до момента t .

Для марковского процесса соотношение (4.11) принимает вид:

$$\begin{aligned} f_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = \\ = f_X(x_1; t_1) f_X(x_2; t_2 | x_1; t_1) f_X(x_3; t_3 | x_2; t_2) \dots f_X(x_n; t_n | x_{n-1}; t_{n-1}). \end{aligned}$$

Иными словами, n -мерная плотность марковского процесса выражается через одномерную плотность первого сечения и условные одномерные плотности последующих сечений при фиксированных значениях только единственного предыдущего сечения. Если воспользоваться определениями (4.7), (4.8), получим

$$\begin{aligned} f_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = \\ = \frac{\prod_{i=1}^{n-1} f(x_i, x_{i+1}; t_i, t_{i+1})}{\prod_{i=2}^{n-1} f(x_i; t_i)} = \frac{\prod_{i=1}^{n-1} f(x_i, x_{i+1}; t_i, t_{i+1})}{\prod_{i=2}^{n-1} \int_{-\infty}^{\infty} f(x_i, x_{i+1}; t_i, t_{i+1}) dx_{i+1}}. \end{aligned} \quad (4.14)$$

Таким образом, n -мерная плотность вероятностей марковского процесса выражается через двумерные плотности. Это бо-

лее сложное представление, чем простейшее (4.12), но значительно проще общего случая (4.11). И если случай независимых сечений является прикладной «экзотикой», то марковская модель широко распространена при описании весьма разнообразных реальных процессов и явлений.

4.4. Ковариационные и взаимные ковариационные функции случайных процессов. Белый шум

Двумерные плотности вероятностей характеризуют свойства случайного процесса совместно в двух сечениях. Однако поиск этих плотностей не является тривиальной процедурой. Поэтому при решении прикладных задач стремятся использовать более доступные в практическом отношении характеристики случайного процесса, пусть менее точно, но тем не менее отражающие взаимные свойства двух сечений процесса. В качестве своеобразной меры статистической связи двух сечений случайного процесса, относящихся к моментам времени t_i и t_j , применяют так называемую ковариационную функцию $K_X(t_i, t_j)$. Прежде чем дать соответствующее определение, условимся *центрированным* называть процесс $X^o(t) = X(t) - m_X(t)$, имеющий нулевое математическое ожидание.

Определение 4.9. Неслучайная функция переменных t_i и t_j

$$K_X(t_i, t_j) = M\{X^o(t_i) X^o(t_j)\} = \quad (4.15)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - m_X(t_i))(x_2 - m_X(t_j)) f_X(x_1, x_2; t_i, t_j) dx_1 dx_2$$

называется *ковариационной функцией* случайного процесса $X(t)$.

Таким образом, ковариационная функция представляет собой среднее значение произведения двух центрированных сечений случайного процесса. Хотя в ее формальном определении присутствует двумерная плотность вероятностей, предполагается, что при практическом вычислении ковариационной функции удастся обойтись без этой плотности. Функцию

$$R_X(t_i, t_j) = \frac{K_X(t_i, t_j)}{\sqrt{D_X(t_i) D_X(t_j)}} \quad (4.16)$$

называют *корреляционной функцией* случайного процесса $X(t)$, или *нормированной ковариационной функцией*. Если $R_X(t_i, t_j) = 0$, то говорят, что сечения $X(t_i)$ и $X(t_j)$ процесса $X(t)$ не коррелированы. В частности, если эти сечения независимы, то они и не коррелированы. Действительно, при независимых сечениях выполняется (4.10), двойной интеграл в (4.15) распадается на два одномерных, каждый из которых равен нулю. Из равенства $R_X(t_i, t_j) = 0$ в общем случае не следует независимость сечений, но эту зависимость с помощью ковариационной (корреляционной) функции не удается зарегистрировать. Из неравенства $R_X(t_i, t_j) \neq 0$ вытекает зависимость сечений.

Приведем наиболее характерные свойства ковариационной функции.

1. $K_X(t_i, t_j) = K_X(t_j, t_i)$, что непосредственно следует из определения (4.15).

2. $K_X(t_i, t_i) = D_X(t_i) \geq 0$, что опять же следует из (4.15).

3. Для любых m вещественных чисел $q_1, q_2, \dots, q_m$ и моментов времени $t_1, t_2, \dots, t_m$ выполняется неравенство

$$\sum_{i=1}^m \sum_{j=1}^m q_i q_j K_X(t_i, t_j) \geq 0. \quad (4.17)$$

Чтобы убедиться в справедливости неравенства, достаточно построить случайную величину $\eta = \sum_{i=1}^m q_i X^\circ(t_i)$ и вычислить $M\{\eta^2\}$.

Функции со свойством (4.17) принято называть неотрицательно определенными.

4. $(K_X(t_i, t_j))^2 \leq K_X(t_i, t_i) K_X(t_j, t_j) = D_X(t_i) D_X(t_j)$, что является аналогом известного неравенства Коши-Буняковского. Это неравенство следует из (4.17), если положить $m = 2$, $q_1 = q$, $q_2 = 1$.

5. Если $Z(t) = g(t) + X(t)$, где $g(t)$ — детерминированный процесс, $X(t)$ — случайный процесс, то $K_Z(t_i, t_j) = K_X(t_i, t_j)$, т.е. неслучайное слагаемое не изменяет ковариационную функцию случайного процесса. Это объясняется легко получаемым равенством $Z^\circ(t) = X^\circ(t)$.

6. Если $Z(t) = g(t)X(t)$, где $g(t)$ — детерминированный процесс, $X(t)$ — случайный процесс, то $K_Z(t_i, t_j) = g(t_i)g(t_j)K_X(t_i, t_j)$. Это неравенство несложно получить, воспользовавшись определением ковариационной функции.

Полезно подчеркнуть, что все приведенные определения и свойства справедливы как для непрерывных случайных процессов, так и для случайных последовательностей. Однако для непрерывного процесса моменты t_i, t_j могут быть любыми в непрерывной области определения процесса, а для случайной последовательности они должны совпадать с дискретными моментами существования элементов последовательности.

При построении математических моделей непрерывных процессов большое значение имеет случайный процесс, называемый *белым шумом*.

Определение 4.10. Случайный процесс $X(t)$ называется *белым шумом*, если $M\{X(t)\} = 0$ и $K_X(t_i, t_j) = c\delta(t_j - t_i)$, где $c = \text{const}$ — интенсивность белого шума, $\delta(t)$ — дельта-функция Дирака, обладающая свойствами

$$\delta(t) = \begin{cases} \infty, & t = 0 \\ 0, & t \neq 0 \end{cases}, \quad \int_{-\epsilon}^{\epsilon} \delta(t) dt = 1 \quad \text{при } \forall \epsilon > 0.$$

Из этого определения следует, что любые два сечения белого шума при $t_i \neq t_j$ некоррелированы, а дисперсия $D_X(t) = \infty$. В силу этих обстоятельств белый шум является чисто гипотетическим процессом, реально не существующим, но представляющим собой весьма полезную математическую модель, широко применяемую при решении многих практических задач. Используется и более широкое толкование белого шума, при котором $c = c(t)$ — функция времени.

При решении задач, основанных на концепции случайных последовательностей, используют другое понятие белого шума. Хотя применительно к иной ситуации это понятие уже использовалось, дадим соответствующее определение.

Определение 4.11. Случайная последовательность $X_1, X_2, \dots$ называется *дискретным белым шумом*, если элементы этой последовательности независимы, $M\{X_i\} = 0$ и $K_X(t_i, t_j) = M\{X_i X_j\} = c\delta_{i,j}$, где, как и выше, $c = \text{const}$ — интенсивность, $\delta_{i,j}$ — дельта-символ Кронекера, уже встречавшийся ранее и определяемый так:

$$\delta_{i,j} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}.$$

Таким образом, дискретный белый шум представляет собой последовательность независимых центрированных случайных величин с постоянной дисперсией c . При более общем определении дискретного белого шума допускается $c = c(t_i)$.

Ковариационная функция $K_X(t_i, t_j)$ случайного процесса $X(t)$, как уже отмечалось, является мерой статистической связи двух сечений этого процесса. Для количественной оценки аналогичной связи, но двух различных процессов используют понятие взаимной ковариационной функции. Дадим соответствующие определения. Пусть $X(t)$ и $Y(t)$ — два случайных процесса.

Определение 4.12. Функция $F_{X,Y}(x, y; t_i, t_j) = P((X(t_i) < x) \cap (Y(t_j) < y))$ называется совместным распределением вероятностей процессов $X(t)$ и $Y(t)$ в моменты времени t_i, t_j .

Определение 4.13. Если $F_{X,Y}(x, y; t_i, t_j)$ — дифференцируемая функция, то функция $f_{X,Y}(x, y; t_i, t_j) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y; t_i, t_j)$ называется совместной плотностью вероятностей процессов $X(t)$ и $Y(t)$ соответственно в моменты времени t_i и t_j .

Определение 4.14. Взаимной ковариационной функцией $K_{XY}(t_i, t_j)$ двух случайных процессов $X(t)$ и $Y(t)$ называется неслучайная функция переменных t_i, t_j :

$$K_{XY}(t_i, t_j) = M\{X^\circ(t_i)Y^\circ(t_j)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - m_X(t_i))(y - m_Y(t_j)) f_{X,Y}(x, y; t_i, t_j) dx dy.$$

Эта функция, представляющая собой среднее значение произведения двух центрированных сечений процессов $X(t)$ и $Y(t)$, используется в качестве меры статистической связи этих процессов. Если при некоторых t_i, t_j $K_{XY}(t_i, t_j) = 0$, то говорят, что сечения $X(t_i)$ и $Y(t_j)$ некоррелированы. Если же это равенство выполняется при любых t_i и t_j , то процессы называют некоррелированными на всем множестве их определения. В соответствии с определением $K_{XY}(t_i, t_j) = K_{YX}(t_j, t_i)$. Если $Z(t) = X(t) + Y(t)$, где $X(t)$ и $Y(t)$ — два случайных процесса, то

$$K_Z(t_i, t_j) = M\{Z^\circ(t_i)Z^\circ(t_j)\} = K_X(t_i, t_j) + K_{XY}(t_i, t_j) + K_{YX}(t_i, t_j) + K_Y(t_i, t_j).$$

Если же процессы не коррелированы, то $K_Z(t_i, t_j) = K_X(t_i, t_j) + K_Y(t_i, t_j)$.

Ковариационные и взаимные ковариационные функции естественным образом распространяются и на векторные случайные процессы, но в этих случаях соответствующие ковариационные функции оказываются матричными. Так, если $X(t) \in \mathbb{R}^n$, $Y(t) \in \mathbb{R}^m$ — два векторных случайных процесса, то по определению:

$$\begin{aligned} K_X(t_i, t_j) &= M\{X^\circ(t_i)(X^\circ(t_j))^T\} \in \mathbb{R}^{n \times n}, \\ K_Y(t_i, t_j) &= M\{Y^\circ(t_i)(Y^\circ(t_j))^T\} \in \mathbb{R}^{m \times m}, \\ K_{X,Y}(t_i, t_j) &= M\{X^\circ(t_i)(Y^\circ(t_j))^T\} \in \mathbb{R}^{n \times m}, \\ K_{Y,X}(t_i, t_j) &= M\{Y^\circ(t_i)(X^\circ(t_j))^T\} \in \mathbb{R}^{m \times n}. \end{aligned}$$

4.5. Стационарные и эргодические случайные процессы

Все случайные процессы принято делить на два широких класса — стационарные и нестационарные процессы, что связано с рядом принципиальных различий в их характеристиках.

Определение 4.15. Случайный процесс $X(t)$ называется стационарным в узком смысле, если все конечномерные функции распределения вероятностей любой размерности инвариантны относительно сдвига во времени, т.е. при $\forall n, t^*$:

$$\begin{aligned} F_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) &= \\ &= F_X(x_1, x_2, \dots, x_n; t_1 + t^*, t_2 + t^*, \dots, t_n + t^*). \end{aligned} \quad (4.18)$$

Из определения следует, что процессы $X(t)$ и $X(t + t^*)$ имеют одинаковые распределения. Процессы, не удовлетворяющие определению (4.18), принято называть нестационарными в узком смысле. Из (4.18) вытекает ряд свойств.

$$\begin{aligned} 1. f_X(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) &= \\ &= f_X(x_1, x_2, \dots, x_n; t_1 + t^*, t_2 + t^*, \dots, t_n + t^*), \end{aligned}$$

так как плотность вероятностей как производная от (4.18) также инвариантна относительно сдвига t^* .

2. Пусть $n = 1$ и выберем $t^* = -t_i$. Тогда $f_X(x; t_i) = f_X(x; t_i + t^*) = f_X(x)$, т.е. одномерная плотность вероятностей стационарного в узком смысле процесса одна и та же во всех сечениях процесса.

3. Пусть $n = 2$ и выберем $t^* = -t_i$. Тогда $f_X(x_1, x_2; t_i, t_j) = f_X(x_1, x_2; t_j - t_i) = f_X(x_1, x_2; \tau)$, $\tau = t_j - t_i$, т.е. двумерная плотность вероятностей стационарного в узком смысле случайного процесса не зависит от того, как выбраны сечения t_i, t_j , а зависит лишь от расстояния τ между сечениями.

4. Математическое ожидание стационарного в узком смысле процесса является постоянной величиной:

$$m_X = M\{X(t)\} = \int_{-\infty}^{\infty} x f_X(x) dx = \text{const.}$$

5. Дисперсия стационарного в узком смысле процесса является постоянной величиной:

$$D_X = M\{(X(t))^2\} = \int_{-\infty}^{\infty} (x - m_X)^2 f_X(x) dx = \text{const.}$$

6. Ковариационная функция стационарного в узком смысле процесса не зависит от моментов времени t_i, t_j , а зависит от расстояния $\tau = t_j - t_i$ между ними:

$$K_X(t_i, t_j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - m_X)(x_2 - m_X) f_X(x_1, x_2; \tau) dx_1 dx_2 = K_X(\tau).$$

7. $D_X = K_X(0)$.

Помимо процессов, стационарных в узком смысле, выделяют класс процессов, стационарных в широком смысле.

Определение 4.16. Случайный процесс $X(t)$ называется *стационарным в широком смысле*, если его математическое ожидание и дисперсия постоянны, а ковариационная функция зависит только от расстояния между сечениями, т.е. $m_X(t) = m_X = \text{const}$, $D_X(t) = D_X = \text{const}$, $K_X(t_i, t_j) = K_X(t_j - t_i) = K_X(\tau)$.

Из сопоставления двух последних определений следует, что стационарный в узком смысле процесс одновременно является стационарным и в широком, но не наоборот. Понятия стационарности в обоих смыслах совпадают лишь для так называемых гауссовских процессов, у которых одномерные и многомерные распределения являются гауссовскими. Аналогичным образом вводится понятие совместно стационарно связанных случайных

процессов. В этом случае $M\{X(t_i)Y(t_j)\} = K_{XY}(t_j - t_i) = K_{XY}(\tau)$. Полезно заметить, что сумма двух нестационарных случайных процессов может оказаться процессом стационарным.

Свойства числовых характеристик стационарных процессов не противоречат их общим свойствам, изложенным выше. Тем не менее удобно заново их указать с учетом стационарности процессов.

1. $K_X(\tau) = K_X(-\tau)$ — четная функция.

2. $|K_X(\tau)| \leq D_X$, $|R_X(\tau)| \leq 1$.

3. Если функция $K_X(\tau)$ непрерывна в точке $\tau = 0$, то она непрерывна при $\forall \tau$.

4. Для многих стационарных процессов $K_X(\tau) \rightarrow 0$ при $\tau \rightarrow \infty$. Величина τ_k такая, что $|K_X(\tau)| \approx 0$ при $\tau \geq \tau_k$, называется *временем корреляции* процесса. Часто принимают $\tau_k = \frac{1}{D_X} \int_0^{\infty} |K_X(\tau)| d\tau$.

5. $K_{XX}(\tau) = K_{XX}(-\tau)$.

6. $(K_{XX}(\tau))^2 \leq K_X(0)K_X(0) = D_X D_Y$.

7. Если $Z(t) = X(t)Y(t)$, где $X(t)$ и $Y(t)$ — стационарные центрированные независимые случайные процессы, то $K_Z(\tau) = M\{Z^0(t)Z^0(t+\tau)\} = K_X(\tau)K_Y(\tau)$.

8. Если $Y(t) = g(t)X(t)$, где $g(t)$ — детерминированная функция, $X(t)$ — стационарный случайный процесс, то $K_Y(t_i, t_j) = g(t_i)g(t_j)K_X(\tau)$.

Понятие стационарности распространяется как на непрерывные процессы, так и на случайные последовательности. Однако в последнем случае необходимо иметь в виду, что расстояние между сечениями представляет собой целое число периодов дискретизации ($\tau = m\xi$, $m = 0, \pm 1, \pm 2, \dots$), а ковариационная функция является решетчатой и ее обозначают символом $K_X[m]$.

Среди стационарных случайных процессов выделяют класс так называемых эргодических процессов, наиболее привлекательный с практической точки зрения. Пусть $X(t)$ — стационарный случайный процесс и $x(t)$ — некоторая его реализация, причем теоретически $t \in (-\infty, \infty)$. Построим по этой реализации следующие величины:

$$\bar{m} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) dt, \quad \bar{D} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T (x(t) - \bar{m})^2 dt,$$

$$\hat{K}_\tau = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T (x(t) - \bar{m})(x(t+\tau) - \bar{m}) dt.$$

Далее положим, что процесс $X(t)$ имеет математическое ожидание, дисперсию и ковариационную функцию, определяемые обычным образом:

$$m_X = \int_{-\infty}^{\infty} x f_X(x) dx, \quad D_X = \int_{-\infty}^{\infty} (x - m_X)^2 f(x) dx,$$

$$K_X(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - m_X)(x_2 - m_X) f_X(x_1, x_2; \tau) dx_1 dx_2.$$

Определение 4.17. Стационарный (в узком смысле) случайный процесс $X(t)$ называется *эргодическим* по математическому ожиданию и ковариационной функции, если с вероятностью единица выполняются равенства: $\bar{m} = m_X$, $\bar{K}(\tau) = K_X(\tau)$ и, как следствие, $\bar{D} = D_X$.

Таким образом, принципиальная особенность эргодического процесса проявляется в том, что основные его числовые характеристики могут быть определены по *одной* реализации процесса достаточно большой длительности без использования плотностей вероятностей. Усреднение проводится по времени, т.е., как говорят, «вдоль процесса». Для неэргодического процесса эти же характеристики приходится искать в соответствии с их классическим определением, использующим плотности вероятностей, поиск которых, в свою очередь, основывается на обработке множества реализаций процесса. Поэтому говорят, что характеристики неэргодического процесса находятся усреднением по множеству реализаций или «поперек процесса». Не любой случайный процесс является эргодическим. Существуют специальные условия, выполнение которых обеспечивает эргодичность. В частности, гауссовский процесс является эргодическим, если $\lim_{\tau \rightarrow \infty} K_X(\tau) = 0$.

Понятие эргодичности распространяется и на случайные последовательности. Однако интегральные операции при проведении временного усреднения здесь уже принципиально неприменимы и усреднение, если $x_i, i = 0, \pm 1, \pm 2, \dots$ — реализация случайной последовательности, принимает вид:

$$\bar{m} = \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{i=-N}^N x_i, \quad \bar{D} = \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{i=-N}^N (x_i - \bar{m})^2,$$

$$\bar{K}[m] = \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{i=-N}^N (x_i - \bar{m})(x_{i+m} - \bar{m}).$$

4.6. Спектральная плотность случайного процесса

Определение 4.18. Пусть $X(t)$ — непрерывный стационарный случайный процесс с абсолютно интегрируемой ковариационной функцией $K_X(\tau)$. Тогда *спектральной плотностью* $S_X(\omega)$ процесса $X(t)$ называют двустороннее преобразование Фурье от его ковариационной функции

$$S_X(\omega) = \int_{-\infty}^{\infty} K_X(\tau) e^{-j\omega\tau} d\tau, \quad (4.19)$$

где ω — параметр, обычно называемый частотой; j — мнимая единица ($j^2 = -1$).

Так как $K_X(\tau)$ — четная функция и $e^{-j\omega\tau} = \cos \omega\tau - j \sin \omega\tau$, то можно записать

$$S_X(\omega) = 2 \int_0^{\infty} K_X(\tau) \cos \omega\tau d\tau,$$

откуда следует, что $S_X(\omega)$ является четной функцией ω .

Справедливо обратное представление

$$K_X(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega) e^{j\omega\tau} d\omega = \frac{1}{\pi} \int_0^{\infty} S_X(\omega) \cos \omega\tau d\omega.$$

Из последнего равенства вытекает важное следствие

$$D_X = K_X(0) = \frac{1}{\pi} \int_0^{\infty} S_X(\omega) d\omega. \quad (4.20)$$

Выявим физический смысл спектральной плотности. С этой целью дополнительно предположим, что $X(t)$ — эргодический центрированный процесс. Тогда

$$D_X = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T (x(t))^2 dt.$$

Последнее соотношение показывает, что дисперсия представляет собой среднюю мощность процесса. Это обстоятельство позволяет следующим образом интерпретировать равенство (4.20): случайный процесс $X(t)$ условно можно представить как

линейную комбинацию бесконечного количества составляющих гармонического типа, частоты которых непрерывно заполняют весь диапазон $[0, \infty)$; тогда величину $\frac{1}{\pi} S_X(\omega) d\omega$ можно трактовать как среднюю мощность составляющих процесса, частоты которых принадлежат отрезку $[\omega, \omega + d\omega]$; суммирование (интегрирование) этих мощностей по всему диапазону частот приводит к средней мощности всего процесса. Следовательно, сама спектральная плотность с точностью до множителя $1/\pi$ представляет собой плотность распределения мощностей гармонических составляющих процесса по частотам.

Спектральные плотности для большинства случайных процессов, моделирующих реальные явления, построены и систематизированы в различных литературных источниках. Приведем некоторые.

$$1. K_X(\tau) = c\delta(\tau), c = \text{const} \Rightarrow S_X(\omega) = c.$$

$$2. K_X(\tau) = \sigma^2 \exp\{-\alpha|\tau|\} \Rightarrow S_X(\omega) = \frac{2\sigma^2\alpha}{\omega^2 + \alpha^2}.$$

$$3. K_X(\tau) = \sigma^2 \exp\{-\alpha|\tau|\} \cos \beta\tau \Rightarrow S_X(\omega) = \frac{2\sigma^2\alpha(\omega^2 + \alpha^2 + \beta^2)}{(\omega^2 - \alpha^2 - \beta^2)^2 + 4\alpha^2\omega^2}.$$

$$4. K_X(\tau) = \sigma^2 \exp\{-\alpha|\tau|\} \left(\cos \beta\tau + \frac{\alpha}{\beta} \sin \beta|\tau| \right) \Rightarrow \\ \Rightarrow S_X(\omega) = \frac{4\sigma^2\alpha(\omega^2 + \alpha^2)}{(\omega^2 - \alpha^2 - \beta^2)^2 + 4\alpha^2\omega^2}.$$

Полезно отметить еще одно свойство пары «ковариационная функция — спектральная плотность»: чем шире график ковариационной функции, тем уже график спектральной плотности, и наоборот. Объясняется это следующим обстоятельством. Широкий график ковариационной функции говорит о том, что процесс характеризуется большим временем корреляции, т.е. его реализации медленно меняются во времени. Но это означает, что мощность процесса определяют низкочастотные составляющие, и график спектральной плотности концентрируется в области низких частот. Узкому графику ковариационной функции соответствует малое время корреляции, реализации процесса меняются

быстро и в них превалирует роль высокочастотных составляющих, что «расширяет» график спектральной плотности. Имеются и более строгие доказательства этого свойства.

Понятие спектральной плотности распространяется и на случайные последовательности. Однако поскольку ковариационная функция $K_X[m]$ случайной последовательности является решетчатой, применить к ней интегральное преобразование вида (4.19) невозможно. Поэтому по определению под спектральной плотностью $S_X(z)$ стационарной случайной последовательности понимают двустороннее z -преобразование ковариационной функции $K_X[m]$:

$$S_X(z) = \sum_{m=-\infty}^{\infty} K_X[m] z^{-m}, \quad (4.21)$$

где z — комплексная переменная, и предполагается, что функция $K_X[m]$ удовлетворяет условиям сходимости ряда (4.21). Часто функцию (4.21) удобнее представить в виде:

$$S_X(z) = S_X^+(z) + S_X^+(z^{-1}) - K_X[0], \quad S_X(z) = \sum_{m=0}^{\infty} K_X[m] z^{-m}, \quad (4.22)$$

где $S_X^+(z)$ — одностороннее z -преобразование ковариационной функции, для которого существуют обширные справочные материалы. Заменой $z = e^{j\omega\xi}$, $\omega\xi \in [-\pi, \pi]$ спектральную плотность $S_X(z)$ часто переводят в частотную область, но в данном случае в этом нет необходимости.

4.7. Преобразование случайного процесса линейным оператором

Пусть $X(t)$ — случайный процесс с известными математическим ожиданием $m_X(t)$ и ковариационной функцией $K_X(t_i, t_j)$ — в результате воздействия заданным линейным оператором A преобразуется в случайный процесс $Y(t)$, т.е.

$$Y(t) = \tilde{A}X(t). \quad (4.23)$$

Задача заключается в поиске числовых характеристик случайного процесса $Y(t)$ и взаимных ковариационных функций процессов $X(t)$ и $Y(t)$.

Решение задачи начнем с вычисления математического ожидания $m_Y(t)$ процесса $Y(t)$. По определению имеем $m_Y(t) = M\{Y(t)\} = M\{\tilde{A}X(t)\}$. Оператор усреднения M , как и оператор $\tilde{A}$, является линейным, причем оба оператора действуют в одном и том же функциональном пространстве случайных процессов. Поэтому, хотя в общем случае линейные операторы не коммутативны, в данном случае можно изменить последовательность их действия и записать $m_Y(t) = \tilde{A}M\{X(t)\}$, что приводит к первому важному результату:

$$m_Y(t) = \tilde{A}m_X(t). \quad (4.24)$$

Ковариационную функцию $K_Y(t_i, t_j)$ процесса $Y(t)$ ищем, используя тот же принцип рассуждений. По определению $K_Y(t_i, t_j) = M\{Y^\circ(t_i)Y^\circ(t_j)\}$. Так как $Y^\circ(t) = Y(t) - m_Y(t) = \tilde{A}X(t) - \tilde{A}m_X(t) = \tilde{A}X^\circ(t)$, получаем $K_Y(t_i, t_j) = M\{\tilde{A}_i X^\circ(t_i) \tilde{A}_j X^\circ(t_j)\}$, где индексы у операторов подчеркивают, по каким переменным (t_i или t_j) они действуют. Снова изменяя последовательность действия операторов M , $\tilde{A}_i$ и $\tilde{A}_j$, получаем второй важный результат:

$$K_Y(t_i, t_j) = \tilde{A}_i \tilde{A}_j K_X(t_i, t_j) \quad (4.25)$$

и, как следствие, $D_Y(t) = K_Y(t, t)$.

Поиск взаимных ковариационных функций не будем сопровождать развернутыми комментариями:

$$K_{YX}(t_i, t_j) = M\{Y^\circ(t_i)X^\circ(t_j)\} = M\{\tilde{A}_i X^\circ(t_i)X^\circ(t_j)\} = \tilde{A}_i K_X(t_i, t_j), \quad (4.26)$$

$$K_{XY}(t_i, t_j) = M\{X^\circ(t_i)Y^\circ(t_j)\} = M\{X^\circ(t_i)\tilde{A}_j X^\circ(t_j)\} = \tilde{A}_j K_X(t_i, t_j). \quad (4.27)$$

Аналогичным образом решается задача поиска вероятностных характеристик процесса $Y(t)$, полученного преобразованием двух случайных процессов $X(t)$ и $Z(t)$ соответственно линейными операторами $\tilde{A}$ и $\tilde{B}$ по правилу

$$Y(t) = \tilde{A}X(t) + \tilde{B}Z(t).$$

Прежняя методика преобразований позволяет установить, что:

$$m_Y(t) = \tilde{A}m_X(t) + \tilde{B}m_Z(t),$$

$$K_Y(t_i, t_j) = \tilde{A}_i \tilde{A}_j K_X(t_i, t_j) + \tilde{A}_i \tilde{B}_j K_{XZ}(t_i, t_j) + \tilde{B}_i \tilde{A}_j K_{ZX}(t_i, t_j) + \tilde{B}_i \tilde{B}_j K_Z(t_i, t_j),$$

$$K_{YX}(t_i, t_j) = \tilde{A}_i K_X(t_i, t_j) + \tilde{B}_i K_{ZX}(t_i, t_j),$$

$$K_{YZ}(t_i, t_j) = \tilde{A}_i K_{XZ}(t_i, t_j) + \tilde{B}_i K_Z(t_i, t_j),$$

$$K_{XY}(t_i, t_j) = \tilde{A}_j K_X(t_i, t_j) + \tilde{B}_j K_{XZ}(t_i, t_j),$$

$$K_{ZY}(t_i, t_j) = \tilde{A}_j K_{ZX}(t_i, t_j) + \tilde{B}_j K_Z(t_i, t_j),$$

где символика обозначений не требует дополнительных пояснений.

4.8. Преобразование случайного процесса оператором свертки

Оператор свертки является частным случаем линейного оператора, широко применяемым при исследовании самых разнообразных динамических процессов и явлений, математическая модель которых представлена линейным дифференциальным или разностным уравнением с постоянными (для простоты) параметрами. Природа оператора такова.

Пусть два случайных процесса $Y(t)$ и $X(t)$ связаны друг с другом дифференциальным уравнением вида

$$a_0 y(t) + a_1 y^{(1)}(t) + a_2 y^{(2)}(t) + \dots + a_n y^{(n)}(t) = b_0 x(t) + b_1 x^{(1)}(t) + \dots + b_m x^{(m)}(t), \quad (4.28)$$

где $y^{(i)}(t)$ — производная i -го порядка и $n \leq m$. Такие уравнения принято называть *стохастическими*. Изучение современной теории стохастических уравнений требует высокой математической культуры, и мы не будем в нее погружаться, ограничив рассмотрение стандартными приемами интегрирования дифференциальных уравнений в рамках операционного подхода (см. приложение 1). Преобразовав это уравнение по Лапласу при отсутствии искусственных ограничений на начальные условия, сведем его к следующему алгебраическому соотношению:

$$\tilde{y}(s) = W(s)\tilde{x}(s), \quad W(s) = \frac{B(s)}{A(s)}, \quad (4.29)$$

где s — комплексная переменная Лапласа, $\tilde{y}(s)$ — изображение функции $y(t)$, $\tilde{x}(s)$ — изображение функции $x(t)$, $A(s) = a_0 + a_1 s +$

$+ a_2 s^2 + \dots + a_n s^n$, $B(s) = b_0 + b_1 s + b_2 s^2 + \dots + b_m s^m$. Чтобы из (4.29) выразить функцию $y(t)$, необходимо это выражение подвергнуть обратному преобразованию Лапласа, которое на основании одного из свойств преобразования Лапласа можно представить в виде

$$y(t) = \int_0^t k(\mu) x(t - \mu) d\mu, \quad (4.30)$$

где $k(t)$ — обратное преобразование Лапласа от функции $W(s)$. Выражение (4.30) и принято называть *оператором свертки*. Таким образом, оператор свертки является одной из форм представления решения дифференциального уравнения (4.28) при отсутствии предварительных ограничений на начальные условия.

Итак, теперь предположим, что $X(t)$ — случайный процесс с известными числовыми характеристиками и связан с процессом $Y(t)$ оператором (4.30). Необходимо найти аналогичные характеристики процесса $Y(t)$ и установить статистическую связь между обоими процессами. Поиск решения задачи проводим в соответствии с общими результатами (4.24)–(4.27):

$$1. m_Y(t) = \int_0^t k(\mu) m_X(t - \mu) d\mu;$$

$$2. K_Y(t_i, t_j) = \tilde{A}_i \int_0^{t_j} k(\mu) K_X(t_i, t_j - \mu) d\mu =$$

$$= \int_0^{t_i} \int_0^{t_j} k(v) k(\mu) K_X(t_i - v, t_j - \mu) dv d\mu;$$

$$3. D_Y(t) = \int_0^t \int_0^t k(v) k(\mu) K_X(t - v, t - \mu) dv d\mu;$$

$$4. K_{XY}(t_i, t_j) = \int_0^{t_j} k(\mu) K_X(t_i, t - \mu) d\mu;$$

$$5. K_{YX}(t_i, t_j) = \int_0^{t_i} k(v) K_X(t_i - v, t_j) dv.$$

Если процесс $X(t)$ является стационарным, эти выражения упрощаются:

$$1. m_Y(t) = m_X \int_0^t k(\mu) d\mu;$$

$$2. K_Y(t_i, t_j) = \int_0^{t_i} \int_0^{t_j} k(v) k(\mu) K_X(\tau - \mu + v) dv d\mu, \quad \tau = t_j - t_i;$$

$$3. D_Y(t) = \int_0^t \int_0^t k(v) k(\mu) K_X(v - \mu) dv d\mu;$$

$$4. K_{XY}(t_i, t_j) = \int_0^{t_j} k(\mu) K_X(\tau - \mu) d\mu;$$

$$5. K_{YX}(t_i, t_j) = \int_0^{t_i} k(v) K_X(\tau + v) dv.$$

Эти равенства показывают, что в общем случае процесс $Y(t)$ даже при стационарном процессе $X(t)$ является нестационарным. Однако существуют условия, при выполнении которых процесс $Y(t)$ при достаточно больших t_i, t_j становится стационарным.

Пусть процесс $X(t)$ — стационарный и $k(t) \rightarrow 0$ при $t \rightarrow \infty$ таким образом, что все интегралы в предыдущих выражениях сходятся при $t, t_i, t_j \rightarrow \infty$. Можно показать, что сходимость интегралов обеспечивается, если все корни уравнения $A(s) = 0$, называемого *характеристическим*, имеют отрицательные вещественные части. Тогда:

$$1. m_Y(t) = m_X \int_0^\infty k(\mu) d\mu = m_Y = \text{const};$$

$$2. K_Y(t_i, t_j) = \int_0^\infty \int_0^\infty k(v) k(\mu) K_X(\tau - \mu + v) dv d\mu = K_Y(\tau);$$

$$3. D_Y(t) = \int_0^\infty \int_0^\infty k(v) k(\mu) K_X(v - \mu) dv d\mu = D_Y = \text{const};$$

$$4. K_{XY}(t_i, t_j) = \int_0^\infty k(\mu) K_X(\tau - \mu) d\mu = K_{XY}(\tau);$$

$$5. K_{YX}(t_i, t_j) = \int_0^\infty k(v) K_X(\tau + v) dv = K_{YX}(\tau).$$

Таким образом, при выполнении указанных условий процесс $Y(t)$, будучи при малых t нестационарным, со временем «превращается» в стационарный с перечисленными характеристиками. В этом, как говорят, установившемся режиме можем дополнительно определить его спектральную плотность

$$S_Y(\omega) = \int_{-\infty}^{\infty} \left[\int_0^{\infty} \int_0^{\infty} k(v)k(\mu)K_X(\tau - \mu + v) d\mu dv \right] e^{-j\omega\tau} d\tau.$$

Изменив последовательность операций интегрирования и обозначив $\tau - \mu + v = \gamma$, $d\tau = d\gamma$, получим

$$\begin{aligned} S_Y(\omega) &= \int_0^{\infty} \int_0^{\infty} k(v)k(\mu) \left[\int_{-\infty}^{\infty} K_X(\tau - \mu + v) e^{-j\omega\tau} d\tau \right] d\mu dv = \\ &= S_X(\omega) \int_0^{\infty} k(v) e^{-j\omega v} dv \int_0^{\infty} k(\mu) e^{j\omega\mu} d\mu. \end{aligned}$$

Так как по определению $\int_0^{\infty} k(v) e^{-j\omega v} dv = W(j\omega)$, $\int_0^{\infty} k(\mu) e^{j\omega\mu} d\mu = W(-j\omega)$, окончательно находим

$$\begin{aligned} S_Y(\omega) &= W(j\omega)W(-j\omega)S_X(\omega) = \\ &= |W(j\omega)|^2 S_X(\omega) = \left| \frac{B(j\omega)}{A(j\omega)} \right|^2 S_X(\omega). \end{aligned} \quad (4.31)$$

Таким образом, если эндогенная и экзогенная переменные связаны друг с другом дифференциальным уравнением (4.28), экзогенная переменная представляет собой стационарный случайный процесс со спектральной плотностью $S_X(\omega)$ и корни характеристического уравнения имеют отрицательные вещественные части, то спектральная плотность эндогенной переменной в установившемся режиме достаточно просто выражается через спектральную плотность экзогенной переменной и параметры уравнения (4.28).

Аналогичные соотношения можно установить и для случайных последовательностей $Y_n, X_n, n = 0, 1, 2, \dots$, связанных друг с

другом линейным разностным уравнением k -го порядка с постоянными параметрами:

$$\begin{aligned} a_0 y_n + a_1 y_{n+1} + a_2 y_{n+2} + \dots + a_k y_{n+k} = \\ = b_0 x_n + b_1 x_{n+1} + b_2 x_{n+2} + \dots + b_m x_{n+m}, \quad k \geq m. \end{aligned} \quad (4.32)$$

Подвергнув это уравнение при отсутствии умышленных ограничений на начальные условия z -преобразованию (см. приложение 1), получим, подобно (4.29),

$$\tilde{y}(z) = W(z)\tilde{x}(z), \quad W(z) = \frac{B(z)}{A(z)}, \quad (4.33)$$

где $\tilde{y}(z), \tilde{x}(z)$ — изображения соответственно функций y_n, x_n и $A(z) = a_0 + a_1 z + a_2 z^2 + \dots + a_k z^k, B(z) = b_0 + b_1 z + b_2 z^2 + \dots + b_m z^m$. Переходя от (4.33) в область оригиналов, получаем

$$y_n = \sum_{i=0}^n k_i x_{n-i}, \quad (4.34)$$

что является оператором свертки, отождествляемым с решением разностного уравнения (4.33). Решетчатая функция k_i в (4.34) является оригиналом изображения $W(z)$. Последующая работа с оператором (4.34) проводится точно так же, как и с оператором (4.30), но в конечных соотношениях интегральные операции следует заменить суммированием. Так как соответствующее редактирование сложностей не порождает, приведем окончательные результаты для случая стационарной случайной последовательности $X_i, i = 0, 1, 2, \dots$ при дополнительном ограничении: корни характеристического уравнения $A(z) = 0$ по модулю меньше единицы, что обеспечивает сходимость соответствующих рядов и позволяет представить:

$$m_Y = m_X \sum_{i=0}^{\infty} k_i;$$

$$K_Y[\tau] = \sum_{v=0}^{\infty} \sum_{\mu=0}^{\infty} k_v k_{\mu} K_X[\tau - \mu + v];$$

$$D_Y = \sum_{v=0}^{\infty} \sum_{\mu=0}^{\infty} k_v k_{\mu} K_X[v - \mu];$$

$$K_{XY}[\tau] = \sum_{\mu=0}^{\infty} k_{\mu} K_X[\tau - \mu];$$

$$K_{YX}[\tau] = \sum_{\nu=0}^{\infty} k_{\nu} K_X[\tau + \nu].$$

Эти соотношения, как и их непрерывные аналоги, напомним, относятся к установившемуся режиму, в котором процесс Y_i можно рассматривать как стационарный. В предшествующем переходном режиме последовательность Y_i является нестационарной. Подвергнув ковариационную функцию $K_Y[\tau]$ двустороннему z -преобразованию, найдем спектральную плотность процесса Y_i , относящуюся к установившемуся режиму:

$$S_Y(z) = \sum_{\tau=-\infty}^{\infty} \left[\sum_{\nu=0}^{\infty} \sum_{\mu=0}^{\infty} k_{\nu} k_{\mu} K_X[\tau - \mu + \nu] \right] z^{-\tau}.$$

Снова, изменив последовательность операций суммирования и обозначив $\tau - \mu + \nu = \gamma$, получим

$$\begin{aligned} S_Y(z) &= S_X(z) \sum_{\mu=0}^{\infty} k_{\mu} z^{-\mu} \sum_{\nu=0}^{\infty} k_{\nu} z^{\nu} = \\ &= W(z)W(z^{-1})S_X(z) = \frac{B(z)B(z^{-1})}{A(z)A(z^{-1})}S_X(z). \end{aligned} \quad (4.35)$$

Соотношение (4.35), таким образом, позволяет найти спектральную плотность случайной последовательности Y_i , если задано разностное уравнение (4.32) и известна спектральная плотность $S_X(z)$ последовательности X_i .

4.9. Формирующие фильтры

При разработке математических моделей стохастических процессов важную роль играет понятие формирующего фильтра. С математической точки зрения под формирующим фильтром понимают некоторое стохастическое уравнение, решение которого, будучи обусловленным порождающим белым шумом, представ-

ляет собой случайный процесс с заданными вероятностными свойствами. Универсального отработанного способа построения такого уравнения нет, однако для процессов с дробно-рациональными спектральными плотностями соответствующая методология известна. Рассмотрим вначале случай непрерывного случайного процесса.

Пусть $Y(t)$ — стационарный случайный процесс со спектральной плотностью $S_Y(\omega)$. Ранее уже отмечалось, что эта плотность является четной функцией частоты ω и для большинства имеющих практическую значимость процессов аппроксимируется дробно-рациональным выражением

$$S_Y(\omega) = \frac{C(\omega^2)}{D(\omega^2)},$$

где $C(\omega^2) = c_0 + c_1\omega^2 + c_2\omega^4 + \dots + c_m\omega^{2m}$, $D(\omega^2) = d_0 + d_1\omega^2 + d_2\omega^4 + \dots + d_n\omega^{2n}$, $n \geq m$.

Характерная особенность многочленов, зависящих от четных степеней аргумента ω , проявляется в том, что их можно представить в виде произведения двух одинаковых многочленов, но один из которых зависит от $j\omega$, а второй — от $(-j\omega)$. Например, $\alpha^2 + \omega^2 = (\alpha + j\omega)(\alpha - j\omega)$. Это позволяет представить $C(\omega^2) = B(j\omega)B(-j\omega)$, $D(\omega^2) = A(j\omega)A(-j\omega)$, где $B(j\omega)$ и $A(j\omega)$ — многочлены по $j\omega$ соответственно порядков m и n . Подобную операцию принято называть факторизацией. В случае ее проведения получим факторизованную спектральную плотность

$$S_Y(\omega) = \frac{B(j\omega)B(-j\omega)}{A(j\omega)A(-j\omega)} = \left| \frac{B(j\omega)}{A(j\omega)} \right|^2. \quad (4.36)$$

Сопоставим это выражение и ранее полученное выражение (4.31). Если в (4.31) положить $S_X(\omega) = 1$, то оба выражения совпадают, причем соотношение (4.31) получено на основе уравнения (4.28) при предположении, что $X(t)$ является случайным процессом со спектральной плотностью $S_X(\omega)$. Следовательно, и (4.36) можно считать результатом, порожденным уравнением (4.28), но при условии, что процесс $X(t)$ имеет единичную спектральную плотность $S_X(\omega) = 1$. Как уже отмечалось, таким процес-

сом является белый шум с единичной интенсивностью. Но тогда процесс $Y(t)$ можно интерпретировать как решение уравнения (4.28), в котором левая и правая части построены по результатам факторизации (4.36), при порождающем белом шуме $X(t)$ с единичной интенсивностью. Это уравнение при такой его содержательности и принято называть *формирующим фильтром*. Процедура его построения, таким образом, сводится к следующему. Проводится факторизация заданной спектральной плотности $S_Y(\omega)$, следствием чего являются многочлены $A(j\omega)$ и $B(j\omega)$. Технология этой операции отработана, содержится, например, в [7] и может быть усовершенствована с привлечением современных вычислительных возможностей. Полезно обратить внимание на то, что при правильно выполненной факторизации корни многочленов $A(s)$ и $B(s)$ имеют отрицательные вещественные части. По результатам факторизации составляется уравнение $A(p)y(t) = B(p)x(t)$, где $p = \frac{d}{dt}$ — оператор дифференцирования. Если $X(t)$ — белый шум с единичной интенсивностью, то это уравнение и будет представлять собой формирующий фильтр. Заметим, что если природа уравнения (4.28) не связана с результатами факторизации, но $X(t)$ — белый шум, то это уравнение позволяет белый шум преобразовать в процесс $Y(t)$, не обязательно стационарный, со свойствами, определяемыми операторами $A(p)$ и $B(p)$.

Подобная идея построения формирующего фильтра распространяется и на случайные последовательности, но здесь формирующий фильтр отождествляется не с дифференциальным, а с разностным стохастическим уравнением и дискретным белым шумом. В основе соответствующего подхода лежит соотношение (4.35), вытекающее из разностного уравнения (4.32). Итак, пусть рассматривается случайная последовательность Y_i , $i = 1, 2, \dots$ с известной спектральной плотностью $S_Y(z)$, построенной в соответствии с (4.22). Функции $S_Y^+(z)$ и $S_Y^-(z^{-1})$ обычно являются дробно-рациональными функциями аргументов z и z^{-1} соответственно, причем нули и полюсы функции $S_Y^+(z)$ лежат внутри окружности единичного радиуса. Если правую часть в выражении функции $S_Y(z)$, соответствующем определению (4.22), привести к общему знаменателю, то снова получим дробно-рациональное выражение. Его знаменатель представляет собой произведение двух одинаковых многочленов, но один из них зависит от z , а второй — от z^{-1} . В числителе же окажется так называемый *возврат-*

ный многочлен вида $q_0 z^{-k} + q_1 z^{-k+1} + \dots + q_{k-1} z^{-1} + q_k + q_{k+1} z + \dots + q_{k+l} z^{l+1} + q_{k+l+1} z^k$. Особенность таких многочленов проявляется в том, что их можно представить в виде произведения двух одинаковых многочленов, из которых один зависит от z , а второй — от z^{-1} . Для этого достаточно вычислить корни многочлена, разложить его на элементарные сомножители, сгруппировать сомножители, содержащие корни, по модулю меньшие единицы, и сомножители с большими по модулю корнями и затем преобразовать к требуемому виду. Например: $2z^{-1} + 5 + 2z = 2z^{-1}(z^2 + \frac{5}{2}z + 1) = 2z^{-1}(z + \frac{1}{2})(z + 2) = 2(z + \frac{1}{2})(1 + 2z^{-1}) = 2(z + \frac{1}{2})2(z^{-1} + \frac{1}{2})$; искомые сомножители $2(z + \frac{1}{2})$ и $2(z^{-1} + \frac{1}{2})$. С использованием современных вычислительных средств подобную процедуру несложно реализовать для возвратных многочленов произвольного порядка. В результате этой процедуры, которую также называют факторизацией, спектральную плотность $S_Y(z)$ удастся представить в виде

$$S_Y(z) = \frac{B(z)B(z^{-1})}{A(z)A(z^{-1})}, \quad (4.37)$$

где $A(z)$ — знаменатель функции $S_Y^+(z)$, $B(z)$ — результат факторизации ее числителя.

Из сопоставления выражений (4.37) и (4.35) следует, что при $S_X(z) = 1$ оба выражения совпадают. Но соотношение (4.35) получено на основе разностного уравнения (4.32), в котором последовательность X_i принималась случайной стационарной со спектральной плотностью $S_X(z)$. Тогда выражение (4.37) можно также считать следствием уравнения (4.32), но для последовательности X_i , имеющей единичную спектральную плотность. Такой последовательностью является дискретный белый шум с ковариационной функцией $K_X[m] = \delta_{m,0}$ и, как следствие, со спектральной плотностью $S_X(z) = 1$. Стохастическое разностное уравнение (4.32), в котором последовательность X_i принимается дискретным белым шумом с единичной интенсивностью, принято называть *дискретным формирующим фильтром* для стационарной случайной последовательности со спектральной плотностью

(4.37). Построение этого уравнения, таким образом, сводится к следующему.

Пусть Y_i , $i = 0, 1, \dots$, — стационарная случайная последовательность с ковариационной функцией $K_Y[m]$. Подобным (4.22) образом строятся функции

$$S_Y^+(z) = \sum_{m=0}^{\infty} K_Y[m] z^{-m}, \quad S_Y(z) = S_Y^+(z) + S_Y^+(z^{-1}) - K_Y[0].$$

После приведения к общему знаменателю функция $S_Y(z)$ приобретает вид $S_Y(z) = \frac{Q(z)}{A(z)A(z^{-1})}$, где $A(z)$ — знаменатель функции

$S_Y^+(z)$, а $Q(z)$ — некоторый возвратный многочлен. Этот многочлен подвергается факторизации, в результате чего его удастся представить в виде $Q(z) = B(z)B(z^{-1})$ с очевидным следствием

$$S_Y(z) = \frac{B(z)B(z^{-1})}{A(z)A(z^{-1})}. \quad \text{Функция } W(z) = \frac{B(z)}{A(z)}, \text{ называемая часто } \textit{переплаточной},$$

определяет разностное уравнение (4.32) формирующего фильтра. Принцип его построения легко обнаруживается из сопоставления уравнения (4.32) и структуры многочленов $A(z)$ и $B(z)$.

В иллюстративных целях рассмотрим следующий пример.

Пусть $K_Y[m] = \sigma^2 e^{-\alpha|m|}$. Тогда $S_Y^+(z) = \sigma^2 \sum_{m=0}^{\infty} (e^{-\alpha} z^{-1})^m = \frac{\sigma^2 z}{z - e^{-\alpha}}$,

что справедливо при $e^{-\alpha}|z^{-1}| < 1$. Далее находим $S_Y(z) = \frac{\sigma^2(1 - e^{-2\alpha})}{(z - e^{-\alpha})(z^{-1} - e^{-\alpha})}$, т.е. в этом простейшем случае $Q(z) =$

$\sigma^2(1 - e^{-2\alpha})$ и, следовательно, $B(z) = \sigma\sqrt{1 - e^{-2\alpha}}$, $A(z) = z - e^{-\alpha}$.

Это позволяет уравнение формирующего фильтра представить в виде

$$y_{n+1} - e^{-\alpha} y_n = \sigma\sqrt{1 - e^{-2\alpha}} x_n.$$

Обратим внимание на одно очень важное обстоятельство. Ранее уже отмечалось, что выражения (4.31), (4.35) справедливы по истечении большого, теоретически бесконечного, промежутка

времени. При ограниченном времени процессы, порождаемые уравнениями (4.28), (4.32), оказываются нестационарными. Чтобы их сделать стационарными при любых значениях времени, необходимо надлежащим образом подобрать начальные условия. Это, разумеется, касается и формирующих фильтров. В продолжение предыдущего примера покажем, как это можно осуществить. В целях лаконичности перепишем уравнение формирующего фильтра в виде

$$y_{n+1} = ay_n + bx_n, \quad a = e^{-\alpha}, \quad b = \sigma\sqrt{1 - e^{-2\alpha}}. \quad (4.38)$$

Для формирования последовательности y_n , $n = 0, 1, \dots$, образованной в соответствии с (4.38), предполагается, что с помощью некоторого генератора создается последовательность x_n , $n = 0, 1, \dots$ как реализация дискретного белого шума единичной интенсивности и при каждом n по правилу (4.38) вычисляется соответствующее значение y_{n+1} . Но чтобы «запустить» алгоритм, необходимо задать начальное условие y_0 . Попробуем это сделать таким образом, чтобы процесс Y_n , реализации которого образуются на основании (4.38), имел постоянную дисперсию σ^2 при всех n . С этой целью положим, что y_0 является реализацией случайной величины Y_0 , имеющей нулевое математическое ожидание $m_Y[0] = 0$ и пока неизвестную дисперсию $D_Y[0] = D_0$. Найдем эту дисперсию.

Усредняя обе части рекуррентного выражения (4.38), получаем

$$m_Y[n+1] = am_Y[n] + bm_X[n] = 0 \Rightarrow m_Y[n] = 0 \text{ при } \forall n,$$

т.е. последовательность Y_n является центрированной. Далее введем обе части того же выражения в квадрат и усредним. В результате получим:

$$D_Y[n+1] = a^2 D_Y[n] + b^2 D_X[n] \Rightarrow D_Y[1] = a^2 D_0 + b^2, \quad D_Y[2] = a^4 D_0 + a^2 b^2 + b^2, \dots,$$

$$D_Y[n+1] = a^{2(n+1)} D_0 + b^2 \sum_{i=0}^n a^{2i} = a^{2(n+1)} D_0 + (a^{2(n+1)} - 1) \frac{b^2}{a^2 - 1}.$$

Так как $a^2 < 1$, то при $n \rightarrow \infty$, т.е. в установившемся режиме, $D_Y = -\frac{b^2}{a^2 - 1} = \sigma^2$. Потребуем, чтобы это равенство выполнялось при всех n :

$$a^{2(n+1)}D_0 + (a^{2(n+1)} - 1) \frac{b^2}{a^2 - 1} = -\frac{b^2}{a^2 - 1} \Rightarrow D_0 = -\frac{b^2}{a^2 - 1} = \sigma^2.$$

Таким образом, если начальное условие в (4.38) является не произвольным, а представляет собой центрированную случайную величину с дисперсией $D_0 = \sigma^2$, то множество реализаций, полученных из белого шума с помощью формирующего фильтра (4.38), образует последовательность с постоянной при всех n дисперсией σ^2 . Покажем, что и ковариационная функция этой последовательности будет удовлетворять условиям стационарности. Для этого «свернем» алгоритм (4.38), записав $y_n = a^n y_0 + b \sum_{i=0}^{n-1} a^i x_{n-1-i}$, и найдем

$$M\{Y_n Y_k\} = a^{n+k} D_0 + b^2 \sum_{i=0}^{n-1} \sum_{j=0}^{k-1} a^{i+j} M\{X_{n-1-i} X_{k-1-j}\}.$$

Так как X_i — дискретный белый шум, то $M\{X_{n-1-i} X_{k-1-j}\}$ равняется единице при $n-1-i = k-1-j$ и нулю в остальных случаях. Поэтому при $n > k$ имеем $i = n - k + j$, $j = 0, 1, \dots, k-1$ и

$$\begin{aligned} M\{Y_n Y_k\} &= a^{n+k} D_0 + b^2 a^{n-k} \sum_{j=0}^{k-1} a^{2j} = \\ &= -a^{n+k} \frac{b^2}{a^2 - 1} + a^{n-k} (a^{2k} - 1) \frac{b^2}{a^2 - 1} = \sigma^2 a^{n-k}. \end{aligned}$$

Аналогичным образом при $k > n$ выразим $j = k - n + i$, $i = 0, 1, \dots, n-1$ и

$$\begin{aligned} M\{Y_n Y_k\} &= a^{n+k} D_0 + b^2 a^{k-n} \sum_{i=0}^{n-1} a^{2i} = \\ &= -a^{n+k} \frac{b^2}{a^2 - 1} + a^{k-n} (a^{2n} - 1) \frac{b^2}{a^2 - 1} = \sigma^2 a^{k-n}. \end{aligned}$$

Последние два соотношения в совместной записи приобретают вид:

$$M\{Y_n Y_k\} = K_Y[n-k] = \sigma^2 a^{|n-k|} = \sigma^2 e^{-\alpha|n-k|},$$

т.е. ковариационная функция последовательности Y_i , $i = 0, 1, \dots$, построенной с помощью формирующего фильтра, определяется только расстоянием между сечениями, и сама последовательность действительно при всех i является стационарной с заданной ковариационной функцией.

4.10. Типовые модели стохастических временных рядов эконометрики

Традиционно в эконометрических приложениях используют определенный класс моделей стохастических временных рядов, получивших широкое распространение в практике обработки временных рядов и включенных во многие пакеты прикладных программ. Все эти модели можно рассматривать как частные случаи линейного разностного уравнения (4.32) при определенном выборе параметров и порождающем дискретном белом шуме X_n , $n = 0, 1, \dots$, т.е. как своеобразные формирующие фильтры. Для удобства запишем еще раз это уравнение:

$$\begin{aligned} a_0 y_n + a_1 y_{n+1} + a_2 y_{n+2} + \dots + a_k y_{n+k} = \\ = b_0 x_n + b_1 x_{n+1} + b_2 x_{n+2} + \dots + b_m x_{n+m}, \quad k \geq m. \end{aligned} \quad (4.39)$$

Конкретизацию этого уравнения удобно осуществить, перейдя к операторному представлению уравнения. Для этого введем в рассмотрение оператор сдвига ζ , такой, что $\zeta y_n = y_{n+1}$, $\zeta^2 y_n = y_{n+2}$ и т.д. Тогда уравнение (4.39) примет вид $(a_0 + a_1 \zeta + a_2 \zeta^2 + \dots + a_k \zeta^k) y_n = (b_0 + b_1 \zeta + b_2 \zeta^2 + \dots + b_m \zeta^m) x_n$ или же

$$y_n = \frac{B(\zeta)}{A(\zeta)} x_n = \frac{b_0 + b_1 \zeta + b_2 \zeta^2 + \dots + b_m \zeta^m}{a_0 + a_1 \zeta + a_2 \zeta^2 + \dots + a_k \zeta^k} x_n,$$

что и будет операторной формой записи разностного уравнения. Положим $b_0 = b_1 = \dots = b_{p-1} = b_{p+2} = \dots = b_m = 0$, $a_{p+1} = a_{p+2} = \dots = 0$

и обозначим $c_i = -\frac{a_{p-i}}{a_p}$ ($i = 1, \dots, p$), $b = \frac{b_p}{a_p}$. Тогда, возвратившись к обычной форме записи уравнения и изменив нумерацию аргументов переменных переходом от $n+p$ к n , получим уравнение

$$y_n = c_1 y_{n-1} + c_2 y_{n-2} + \dots + c_p y_{n-p} + b x_n. \quad (4.40)$$

Соотношение (4.40) принято называть *авторегрессионной* p -го порядка моделью стохастического временного ряда и обозначать символом $AR(p)$ или в зарубежном варианте $AR(p)$. Рассмотренную ранее модель (4.38) можно рассматривать как частный вариант модели $AR(1)$. Иногда в этой модели предусматривают еще одно слагаемое, моделирующее средний уровень ряда, и тогда модель записывают в форме $y_n = c_0 + c_1 y_{n-1} + c_2 y_{n-2} + \dots + c_k y_{n-p} + b x_n$. Это слагаемое можно интерпретировать как ненулевое математическое ожидание порождающего шума x_n .

Второй не менее распространенной в эконометрических приложениях моделью является так называемая модель *скользящего среднего*. Ее также можно получить, исходя из уравнения (4.32), если положить $a_0 = a_1 = \dots = a_{q-1} = a_{q+1} = \dots = a_k = 0$, $b_{q+1} = b_{q+2} = \dots = b_m = 0$. Обозначая $d_i = \frac{b_{q-i}}{a_q}$, $i = 0, 1, \dots, q$, и заменяя в нумерации переменных $n + q$ на n , получаем модель скользящего среднего порядка q :

$$y_n = d_0 x_n + d_1 x_{n-1} + \dots + d_q x_{n-q}, \quad (4.41)$$

которую в литературе часто сокращенно обозначают как $CC(q)$ или $MA(q)$.

Если в уравнении (4.32) положить $a_{p+1} = a_{p+2} = \dots = a_k = 0$, $b_0 = b_1 = \dots = b_{p-q-1} = 0$, $b_{p+1} = b_{p+2} = \dots = b_m = 0$, ввести обозначения $c_i = -\frac{a_{p-i}}{a_p}$ ($i = 1, 2, \dots, p$), $d_i = \frac{b_{m-i}}{a_m}$ ($i = 0, 1, \dots, p - q$) и опять же изменить нумерацию аргументов переменных, заменив $n + p$ на n , получим модель, известную под названием смешанной модели *авторегрессии — скользящего среднего* порядка p, q (сокращенно $ARCC(p, q)$ или $ARMA(p, q)$) и имеющую, таким образом, вид

$$y_n = c_1 y_{n-1} + c_2 y_{n-2} + \dots + c_p y_{n-p} + d_0 x_n + d_1 x_{n-1} + \dots + d_q x_{n-q}. \quad (4.42)$$

Модель (4.32) определяет как стационарные, так и нестационарные временные стохастические ряды. Можно показать, что если все корни алгебраического уравнения

$$A(z) = a_0 + a_1 z + a_2 z^2 + \dots + a_k z^k = 0, \quad (4.43)$$

называемого, как и в аналогичном непрерывном случае, *характеристическим*, находятся в комплексной плоскости строго внутри окружности единичного радиуса, т.е. по модулю меньше единицы, то определяемый уравнением (4.39) процесс в установившемся режиме оказывается стационарным. Если же хотя бы один из корней этого уравнения находится за пределами единичной окружности или принадлежит ей, последовательность Y_n будет нестационарной. Проиллюстрируем эту мысль на примере уравнения (4.38). Соответствующее ему характеристическое уравнение $z - a = 0$ имеет единственный корень $z_1 = a$. Дисперсия $D_Y[n]$ определяемого этим уравнением временного ряда Y_n , как было показано выше, имеет вид

$$D_Y[n] = a^{2n} D_0 + b^2 \sum_{i=0}^{n-1} a^{2i} = a^{2n} D_0 + (a^{2n} - 1) \frac{b^2}{a^2 - 1}.$$

Легко обнаружить, что при $a > 1$ дисперсия $D_Y[n]$ оказывается возрастающей функцией n , и это является признаком нестационарности ряда. При $a = 1$ получаем $D_Y[n] = D_0 + b^2 n$, что также свидетельствует о нестационарности ряда. И только при $a < 1$ дисперсия стремится к установившемуся значению, а ковариационная функция, что также было уже показано, перестает зависеть от «адресов» сечений, т.е. последовательность Y_n становится стационарной. Подобное свойство проявляется и при многочленах $A(z)$ произвольной структуры, если только корни многочлена $A(z)$ не компенсируются аналогичными корнями многочлена $B(z)$. В эконометрических приложениях среди всех возможных вариантов расположения корней характеристического уравнения особо выделяют случай, когда небольшая часть корней (один, два) оказывается равной единице, а остальные корни — внутри единичной окружности. Эта ситуация приводит к нестационарному ряду Y_n , но такому, что разность некоторого порядка этого ряда оказывается стационарной. Покажем это.

Прежде всего воспользуемся введенным оператором сдвига ζ и дополнительно определим оператор *первой разности* $\Delta y_n = y_{n+1} - y_n = (\zeta - 1)y_n$. Аналогично второй разностью $\Delta^2 y_n$ назовем первую разность первых разностей $\Delta^2 y_n = \Delta y_{n+1} - \Delta y_n = y_{n+2} - 2y_{n+1} + y_n = (\zeta - 1)^2 y_n$. Подобным образом можно ввести разности произвольных порядков. Функцию Y_n назовем *первообразной* функции y_n , если выполняется условие $\Delta Y_n = Y_{n+1} - Y_n = y_n$. Спра-

ведливо легко проверяемое условие $Y_n = \Delta^{-1} y_n = \sum_{i=0}^{n-1} y_i$. Поэтому

операцию суммирования часто интерпретируют как обратную по отношению к операции взятия разности (подобно соотношению между операциями дифференцирования и интегрирования). В алгебраизированном виде уравнение (4.32) можно записать так: $A(\zeta)y_n = B(\zeta)x_n$, где $A(\zeta) = a_0 + a_1\zeta + a_2\zeta^2 + \dots + a_k\zeta^k$, $B(\zeta) = b_0 + b_1\zeta + b_2\zeta^2 + \dots + b_m\zeta^m$. Пусть теперь характеристическое уравнение (4.43) имеет d равных единиче корней и, следовательно, можно представить $A(z) = A_1(z)(z-1)^d$, причем все корни многочлена $(k-d)$ -го порядка $A_1(z)$ находятся внутри окружности единичного радиуса. Но этому представлению многочлена $A(z)$ соответствует запись уравнения (4.32) в виде $A_1(\zeta)(\zeta-1)^d y_n = B(\zeta)x_n$. Теперь чис-

то формально запишем $y_n = \frac{B(\zeta)}{A_1(\zeta)(\zeta-1)^d} x_n$ и обозначим $q_n = B(\zeta)x_n$, что соответствует модели скользящего среднего; $A_1(\zeta)v_n = q_n$, что соответствует модели авторегрессии; $y_n = \frac{1}{(\zeta-1)^d} v_n = \Delta^{-d} v_n$, что

соответствует оператору d -кратного суммирования. Из последнего равенства следует $\Delta^d y_n = v_n$. Но последовательность v_n в установленном режиме является стационарной, так как корни многочлена $A_1(z)$ расположены внутри окружности единичного радиуса. Следовательно, и d -я разность временного ряда Y_n также образует стационарный процесс. Нестационарные временные ряды, порожденные стохастическим разностным уравнением (4.32), характеристическое уравнение которого имеет d равных единиче корней, а остальные корни по модулю меньше единицы, принято называть процессами авторегрессии — проинтегрированного скользящего среднего (АРПСС) порядка (k, d, m) или ARIMA (k, d, m) . Важная особенность модели, таким образом, проявляется в том, что она задает процесс нестационарный, но со стационарной разностью d -го порядка. Названия всех перечисленных моделей не очень благозвучны для русского восприятия и не являются безупречно содержательными, но они устоялись и подвергать их ревизии нецелесообразно.

Проведенный формальный переход от уравнения (4.32) к частным моделям AR(p) (4.40), MA(q) (4.41), ARMA (4.42), ARIMA

упростится, если предварительно провести следующее изменение в уравнении (4.32): обозначим $n+k=i$ и перепишем уравнение в виде

$$a_0 y_{i-k} + a_1 y_{i-k+1} + a_2 y_{i-k+2} + \dots + a_k y_i = b_0 x_{i-k} + b_1 x_{i-k+1} + b_2 x_{i-k+2} + \dots + b_m x_{i-k+m}, \quad k \geq m$$

или же, в целях унификации символики обозначив $i=n$ и положив $k=m$, в виде

$$a_k y_n + a_{k-1} y_{n-1} + a_{k-2} y_{n-2} + \dots + a_0 y_{n-k} = b_k x_n + b_{k-1} x_{n-1} + b_{k-2} x_{n-2} + \dots + b_0 x_{n-k}.$$

Тогда легко обнаруживается, что это уравнение превращается в уравнение AR(p) (4.40), если положить $a_0 = a_1 = \dots = a_{k-p-1} = 0$,

$b_0 = b_1 = \dots = b_{k-1} = 0$ и обозначить $c_1 = -\frac{a_{k-1}}{a_k}$, $c_2 = -\frac{a_{k-2}}{a_k}$, ...,

$c_p = -\frac{a_{k-p}}{a_k}$, $b = \frac{b_k}{a_k}$. Если в этом же уравнении принять $a_0 = a_1 =$

$= \dots = a_{k-1} = 0$, $b_0 = b_1 = \dots = b_{k-q-1} = 0$ и обозначить

$d_0 = \frac{b_k}{a_k}$, $d_1 = \frac{b_{k-1}}{a_k}$, ..., $d_q = \frac{b_{k-q}}{a_k}$, то получим модель MA(q) (4.41).

Наконец, при $a_0 = a_1 = \dots = a_{k-p-1} = 0$ и $b_0 = b_1 = \dots = b_{k-q-1} = 0$ и сохранении введенных обозначений для параметров c_i , d_j получим модель ARMA(p, q) (4.42). Если дополнительно среди корней уравнения $a_k + a_{k-1}z^{-1} + a_{k-2}z^{-2} + \dots + a_{k-p}z^{-p} = 0$ «предусмотреть» d корней, равных 1, получим модель ARIMA(p, d, q).

При выборе модели стохастического временного ряда возникают те же проблемы, что и при построении квазидетерминированной модели: какой вид модели обоснованно предпочесть, как определить наилучший порядок модели, как по экспериментальным данным оценить параметры модели. Ответы на эти вопросы частично созвучны ранее найденным применительно к задачам регрессионного анализа и достаточно полно освещены в литературе по исследованию типовых моделей стохастических рядов (например, [2, 4]).

4.11. Стохастический вектор состояния. Обобщенная матрично-векторная модель временного ряда

Характерная особенность современных методов обработки случайных последовательностей проявляется в широком применении рекуррентных алгоритмов, подобных ранее рассмотренному рекуррентному методу наименьших квадратов. Поэтому попытаемся все ранее рассмотренные модели стохастических рядов, названные нами типовыми, представить в форме одной модели, соответствующей основным принципам построения рекуррентных алгоритмов обработки экспериментальных данных.

Обратимся к разностному уравнению (4.32) и перепишем его в форме

$$a_k y_{n+k} - b_k x_{n+k} + a_{k-1} y_{n+k-1} - b_{k-1} x_{n+k-1} + \dots + a_1 y_{n+1} - b_1 x_{n+1} + a_0 y_n - b_0 x_n = 0$$

или же, используя оператор сдвига,

$$\zeta^k (a_k y_n - b_k x_n) + \zeta^{k-1} (a_{k-1} y_n - b_{k-1} x_n) + \dots + \zeta (a_1 y_n - b_1 x_n) + (a_0 y_n - b_0 x_n) = 0.$$

Введем новые переменные

$$\begin{aligned} z_{1,n} &= a_k y_n - b_k x_n, \\ z_{2,n} &= z_{1,n+1} + a_{k-1} y_n - b_{k-1} x_n, \\ z_{3,n} &= z_{2,n+1} + a_{k-2} y_n - b_{k-2} x_n, \\ &\dots \\ z_{k,n} &= z_{k-1,n+1} + a_1 y_n - b_1 x_n, \\ z_{k,n+1} + a_0 y_n - b_0 x_n &= 0. \end{aligned}$$

Из первого уравнения выразим

$$y_n = \frac{1}{a_k} z_{1,n} + \frac{b_k}{a_k} x_n \quad (4.44)$$

и, подставив это значение в оставшиеся уравнения, найдем:

$$\begin{aligned} z_{1,n+1} &= -\frac{a_{k-1}}{a_k} z_{1,n} + z_{2,n} + \left(-\frac{a_{k-1} b_k}{a_k} + b_{k-1} \right) x_n, \\ z_{2,n+1} &= -\frac{a_{k-2}}{a_k} z_{1,n} + z_{3,n} + \left(-\frac{a_{k-2} b_k}{a_k} + b_{k-2} \right) x_n, \\ z_{3,n+1} &= -\frac{a_{k-3}}{a_k} z_{1,n} + z_{4,n} + \left(-\frac{a_{k-3} b_k}{a_k} + b_{k-3} \right) x_n, \\ &\dots \\ z_{k-1,n+1} &= -\frac{a_1}{a_k} z_{1,n} + z_{k,n} + \left(-\frac{a_1 b_k}{a_k} + b_1 \right) x_n, \\ z_{k,n+1} &= -\frac{a_0}{a_k} z_{1,n} + \left(-\frac{a_0 b_k}{a_k} + b_0 \right) x_n. \end{aligned} \quad (4.45)$$

В результате проведенной операции единственное разностное уравнение (4.32) k -го порядка заменяется системой из k разностных уравнений 1-го порядка, разрешенных относительно новых переменных в момент $n+1$. Решение исходного уравнения (4.32) выражается через решение системы (4.45) в соответствии с (4.44).

Уравнения (4.44), (4.45) удобно переписать в матрично-векторной форме:

$$y_n = C^T Z_n + b x_n, \quad (4.46)$$

$$Z_{n+1} = A Z_n + B x_n, \quad (4.47)$$

где использованы естественным образом следующие из представлений (4.44), (4.45) обозначения:

$$b = \frac{b_k}{a_k}, \quad Z_n = \begin{bmatrix} z_{1,n} \\ z_{2,n} \\ \dots \\ z_{k-1,n} \\ z_{k,n} \end{bmatrix}, \quad A = \begin{bmatrix} -\frac{a_{k-1}}{a_k} & 1 & 0 & 0 & \dots & 0 \\ -\frac{a_{k-2}}{a_k} & 0 & 1 & 0 & \dots & 0 \\ -\frac{a_{k-3}}{a_k} & 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ -\frac{a_1}{a_k} & 0 & 0 & 0 & \dots & 1 \\ -\frac{a_0}{a_k} & 0 & 0 & 0 & \dots & 0 \end{bmatrix}, \quad (4.48)$$

$$B = \frac{1}{a_k} \begin{bmatrix} a_k b_{k-1} - a_{k-1} b_k \\ a_k b_{k-2} - a_{k-2} b_k \\ a_k b_{k-3} - a_{k-3} b_k \\ \dots \\ a_k b_1 - a_1 b_k \\ a_k b_0 - a_0 b_k \end{bmatrix}, \quad C = \begin{bmatrix} 1 \\ a_k \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}.$$

Вектор Z_n , определяемый стохастическим матричным уравнением (4.47), по сложившейся традиции назовем *стохастическим вектором состояния*. Этот вектор путем линейного преобразования (4.46) формирует решение y_n уравнения (4.32). Как из уравнения (4.32) в качестве частных случаев были получены модели AR, MA, ARMA, ARIMA, так и из системы (4.46), (4.47) путем введения соответствующих ограничений на параметры характеристик (4.48) можно сформировать те же модели, но в терминах вектора состояния.

Положив в (4.48) $k = p$, $b_0 = b_1 = \dots = b_{p-1} = 0$, $b_p \neq 0$, т.е. $B^T = -b[a_{p-1} \ a_{p-2} \ \dots \ a_0]$, и сохранив A , C в виде (4.48), получим авторегрессионную модель AR(p). Покажем это для случая $p = 2$, воспользовавшись чисто формальной процедурой. Обозначим:

$$c_1 = -\frac{a_1}{a_2}, \quad c_2 = -\frac{a_0}{a_2}, \quad \text{так что}$$

$$A = \begin{bmatrix} c_1 & 1 \\ c_2 & 0 \end{bmatrix}, \quad B = b_2 \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \quad C = \frac{1}{a_2} \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

и, воспользовавшись оператором сдвига, запишем $(\zeta E_2 - A)Z_n = Bx_n$ или же $Z_n = (\zeta E_2 - A)^{-1} Bx_n$, где E_2 — единичная матрица размерности два. Тогда $y_n = C^T (\zeta E_2 - A)^{-1} x_n + bx_n$. Осуществив формальные операции, получим

$$y_n = \frac{b_2(c_1 \zeta + c_2)}{a_2(\zeta^2 - c_1 \zeta - c_2)} x_n + bx_n.$$

Умножим обе части этого выражения на $\zeta^2 - c_1 \zeta - c_2$ и учтем смысл оператора ζ . В результате находим $y_{n+2} - c_1 y_{n+1} - c_2 y_n = bc_1 x_{n+1} + bc_2 x_n + bx_{n+2} - bc_1 x_{n+1} - bc_2 x_n$, что после замены $n + 2$ на n приводит к окончательному результату (4.40):

$$y_n = c_1 y_{n-1} + c_2 y_{n-2} + bx_n.$$

Если в выражениях (4.48) положить $k = q$, $a_0 = a_1 = \dots = a_{q-1} = 0$, т.е.

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}, \quad B = \begin{bmatrix} b_{q-1} \\ b_{q-2} \\ \dots \\ b_1 \\ b_0 \end{bmatrix}, \quad C = \frac{1}{a_q} \begin{bmatrix} 1 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix},$$

то получим модель скользящего среднего MA(q), в чем можно убедиться подобным предыдущему образом.

Положив в соотношениях (4.48) $k = p$, $b_0 = b_1 = \dots = b_{p-q-1} = 0$, $b_{p-q} \neq 0$, $b_{p-q+1} \neq 0$, ..., $b_p \neq 0$, получим модель авторегрессии — скользящего среднего ARMA(p , q). Наконец, если представить $A(z) = A_1(z)(z - 1)^d$, выразить коэффициенты многочлена $A(z)$ через коэффициенты многочлена $A_1(z)$ и построить соответствующую матрицу A , получим матрично-векторную модель ARIMA(p , q). Таким образом, уравнения (4.46), (4.47) при надлежащем определении структурных параметров (4.48) можно рассматривать как обобщенную матрично-векторную стохастическую модель временного ряда, отражающую современные тенденции описания случайных последовательностей.

Модель (4.46), (4.47) является достаточно гибкой в том смысле, что легко адаптируется к рядам, содержащим не только стохастические составляющие, но и детерминированные. Так, если наблюдения (4.46) дополнительно содержат гармоническую составляющую $f_n = u \sin(\omega n)$ с неизвестной амплитудой u и известной частотой ω , имитирующую сезонные колебания, то можно, расширив вектор состояния, свести описание ряда к той же модели (4.46), (4.47), но увеличенной размерности. С этой целью зададим амплитуду u как решение уравнения $u_n = u_{n-1}$, добавив его к системе (4.45), и величину u_n включим в качестве $(k + 1)$ -го компонента в состав вектора Z_n . Как следствие, это приведет к возрастанию размерностей всех матричных величин (4.48), которые окажутся равными:

$$Z_n = \begin{bmatrix} z_{1,n} \\ z_{2,n} \\ \dots \\ z_{k-1,n} \\ z_{k,n} \\ u_n \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & 1 & 0 & 0 & \dots & 0 & 0 \\ a_{21} & 0 & 1 & 0 & \dots & 0 & 0 \\ a_{31} & 0 & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{k-1,1} & 0 & 0 & 0 & \dots & 1 & 0 \\ a_{k1} & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix},$$

$$B = \begin{bmatrix} b_1^* \\ b_2^* \\ \dots \\ b_{k-1}^* \\ b_k^* \\ 0 \end{bmatrix}, \quad C = C_n = \begin{bmatrix} c_1^* \\ 0 \\ 0 \\ \dots \\ 0 \\ \sin(\omega n) \end{bmatrix}.$$

Здесь для упрощения записей использованы более лаконичные обозначения элементов матриц A , B , C , но их смысл тот же, что и в (4.48). Аналогичным образом можно поступить и в случае совместно неизвестных u и w , введя дополнительное уравнение для частоты $w_n = w_{n-1}$, что еще на единицу увеличит размерность модели и сделает ее нелинейной. Воспользовавшись аналогичным подходом, можно надлежащим выбором матрицы A предусмотреть присутствие в составе ряда полиномиального тренда и иных детерминированных составляющих.

4.12. Рекуррентный алгоритм прогнозирования стохастических временных рядов (калмановский фильтр)

Уточним постановку задачи. Будем полагать, что наблюдается временной ряд $y_1, y_2, \dots, y_N$. Элементы (уровни) этого ряда представляют собой аддитивную смесь полезной составляющей ряда и сопутствующих любому эксперименту измерительных ошибок (шумов, возмущений). Полезная составляющая ряда линейным образом выражается через ненаблюдаемый стохастический вектор состояния. Математическая модель ряда, таким образом, имеет подобный (4.46), (4.47) вид:

$$y_n = C^T Z_n + p_n, \quad n = 1, 2, \dots, N, \quad (4.49)$$

$$Z_n = AZ_{n-1} + Bx_{n-1}. \quad (4.50)$$

Здесь обозначены: $p_n \sim N(0, \sigma_p^2)$ — экспериментальные ошибки типа гауссовского белого шума; $x_n \sim N(0, \sigma_x^2)$ — порождающий белый шум, причем $M\{p_i x_j\} = 0$ при $\forall i, j$; $Z_0 \sim N(m_{z0}, K_{z0})$ — определенное в вероятностном смысле начальное состояние, не зависящее от x_0 ; параметры модели C , A , B предполагаются известными. Отличие модели (4.49), (4.50) от ее прототипа (4.46), (4.47) проявляется в том, что в (4.49) включены экспериментальные шумы, но отсутствует слагаемое $b x_n$, фигурирующее в (4.46). Формально это означает, что в уравнении (4.32) $m < k$ и $b_k = 0$, что характерно для многих приложений. В последующем мы это ограничение снимем. Обозначим $y_n^* = C^T Z_n$. Решаемая далее задача заключается в следующем: располагая наблюдениями в объеме (4.49), требуется найти наилучшую в некотором смысле оценку $\hat{y}_m^*$ величины y_m^* , где $m > N$. Эта оценка ищется в виде $\hat{y}_m^* = C^T \hat{Z}_m$, где $\hat{Z}_m$ — прогнозируемое значение стохастического вектора состояния. В свою очередь, как будет показано, $\hat{Z}_m = A^{m-N} \hat{Z}_N$, где $\hat{Z}_N$ — оценка вектора состояния Z_N , найденная по наблюдениям $y_1, y_2, \dots, y_N$. Таким образом, задача прогнозирования сводится к поиску наилучшей в определенном смысле оценки $\hat{Z}_N$ вектора Z_N .

Обозначим символом y_1^n вектор наблюдений с компонентами $y_1, y_2, \dots, y_n$, т.е. $y_1^n = [y_1 \ y_2 \ \dots \ y_n]^T$, и по этим наблюдениям найдем оценку $\hat{Z}_n(y_1^n)$ вектора Z_n как функцию наблюдений y_1^n , максимизирующую апостериорную плотность вероятностей $\omega(Z_n | y_1^n)$ этого вектора. Таким образом, в качестве наилучшей принимается оценка

$$\hat{Z}_n = \arg \max_{Z_n \in R^k} \omega(Z_n | y_1^n). \quad (4.51)$$

Напомним, при рассмотрении байесовского метода оценивания параметров регрессионной модели было показано, что для линейной гауссовской модели наблюдений и квадратичной функции стоимости байесовские оценки совпадают с оценками по критерию максимума апостериорной плотности вероятностей. Это справедливо и в данном случае. Поэтому задача (4.51) эквивалентна условию

$$M\{\|Z_n - \hat{Z}_n\|^2 | y_1^n\} \rightarrow \min_{\hat{Z}_n},$$

т.е. решение задачи (4.51) обеспечивает наилучшее приближение оценки $\hat{Z}_n$ к оцениваемому вектору Z_n , но аналитически оперировать с (4.51) удобнее.

Для решения задачи (4.51) прежде всего установим структуру апостериорной плотности. Так как модель (4.49), (4.50) является линейной и гауссовской, то и условная плотность $\omega(Z_n | y_1^n)$ будет также гауссовской. Как известно, условная гауссовская плотность определяется двумя параметрами: условным математическим ожиданием $M\{Z_n | y_1^n\}$ и условной ковариационной матрицей $M\{Z_n^\circ (Z_n^\circ)^T | y_1^n\}$, где, как и ранее, $^\circ$ — символ центрирования. Но математическое ожидание гауссовской величины соответствует максимуму ее плотности вероятностей, в силу чего $M\{Z_n | y_1^n\} = \hat{Z}_n$. Дополнительно обозначим $M\{Z_n^\circ (Z_n^\circ)^T | y_1^n\} = M\{(Z_n - \hat{Z}_n)(Z_n - \hat{Z}_n)^T | y_1^n\} = R_n$. Эту матрицу, характеризующую точность оценивания вектора Z_n по данным y_1^n , принято обычно называть *апостериорной ковариационной матрицей*. Таким образом, $\omega(Z_n | y_1^n) = N(\hat{Z}_n, R_n)$.

Аналогичным образом и по тем же причинам выразим $\omega(Z_n | y_1^{n-1}) = N(\hat{Z}_{n,n-1}, R_{n,n-1})$, где $\hat{Z}_{n,n-1} = M\{Z_n | y_1^{n-1}\}$, $R_{n,n-1} = M\{(Z_n - \hat{Z}_{n,n-1})(Z_n - \hat{Z}_{n,n-1})^T | y_1^{n-1}\}$. Матрицу $R_{n,n-1}$ принято называть *априорной ковариационной матрицей*. Вектор $\hat{Z}_{n,n-1}$ является оценкой вектора состояния Z_n , найденной по наблюдениям $y_1^{n-1} = [y_1 y_2 \dots y_{n-1}]^T$, что принципиально отличает его от оценки $\hat{Z}_n$ того же вектора, но найденной по наблюдениям y_1^n . По существу, величина $\hat{Z}_{n,n-1}$ является прогнозированным значением вектора Z_{n-1} на один шаг вперед, найденным по наблюдениям $y_1, y_2, \dots, y_{n-1}$.

Установим связь между условными плотностями $\omega(Z_n | y_1^n)$ и $\omega(Z_n | y_1^{n-1})$. Для этого предположим, что найдена условная плотность $\omega(Z_{n-1} | y_1^{n-1})$, и рассмотрим ряд соотношений, основанных на общих свойствах плотностей вероятностей:

$$\omega(Z_n, y_n | y_1^{n-1}) = \omega(Z_n | y_1^{n-1}) \omega(y_n | Z_n, y_1^{n-1}) = \omega(y_n | y_1^{n-1}) \omega(Z_n | y_1^n).$$

Из этого равенства следует интересующее нас соотношение

$$\omega(Z_n | y_1^n) = c_n \omega(Z_n | y_1^{n-1}) \omega(y_n | Z_n, y_1^{n-1}) = c_n \omega(Z_n | y_1^{n-1}) \omega(y_n | Z_n), \quad (4.52)$$

где $c_n = 1/\omega(y_n | y_1^{n-1})$ и учтено, что при фиксированном векторе Z_n величина y_n не зависит от предшествующих наблюдений y_1^{n-1} , в силу чего $\omega(y_n | Z_n, y_1^{n-1}) = \omega(y_n | Z_n)$. Входящая в выражение (4.52) условная плотность $\omega(y_n | Z_n)$ находится из (4.49): $\omega(y_n | Z_n) = N(C^T Z_n, \sigma_p^2)$. Так как c_n не зависит от Z_n , а обе плотности $\omega(Z_n | y_1^{n-1})$ и $\omega(y_n | Z_n)$ из (4.52) являются гауссовскими, то общий показатель экспоненты произведения $\omega(Z_n | y_1^{n-1}) \times \omega(y_n | Z_n)$ имеет вид: $-\frac{1}{2}(Z_n - \hat{Z}_{n,n-1})^T R_{n,n-1}^{-1} (Z_n - \hat{Z}_{n,n-1}) - \frac{1}{2} \sigma_p^{-2} (y_n - C^T Z_n)^2$. Но тогда задача (4.51) в силу (4.52) эквивалентна задаче

$$J = (Z_n - \hat{Z}_{n,n-1})^T R_{n,n-1}^{-1} (Z_n - \hat{Z}_{n,n-1}) + \sigma_p^{-2} (y_n - C^T Z_n)^2 \rightarrow \min_{Z_n}. \quad (4.53)$$

Запишем необходимое условие минимума:

$$\nabla J = 2R_{n,n-1}^{-1} (Z_n - \hat{Z}_{n,n-1}) + 2\sigma_p^{-2} C(y_n - C^T Z_n) = 0_k,$$

где 0_k — нулевой k -вектор. Решение этого уравнения и будет представлять собой оценку $\hat{Z}_n$. Таким образом,

$$\hat{Z}_n = (R_{n,n-1}^{-1} + \sigma_p^{-2} C C^T)^{-1} (R_{n,n-1}^{-1} \hat{Z}_{n,n-1} + \sigma_p^{-2} C y_n). \quad (4.54)$$

Воспользовавшись леммой об обращении матрицы, представим

$$\begin{aligned} (R_{n,n-1}^{-1} + \sigma_p^{-2} C C^T)^{-1} &= \\ &= R_{n,n-1} - R_{n,n-1} C (C^T R_{n,n-1} C + \sigma_p^2)^{-1} C^T R_{n,n-1}, \end{aligned} \quad (4.55)$$

а на основе (4.50) запишем

$$\begin{aligned} \hat{Z}_{n,n-1} &= M\{Z_n | y_1^{n-1}\} = M\{A Z_{n-1} + B x_{n-1} | y_1^{n-1}\} = \\ &= A M\{Z_{n-1} | y_1^{n-1}\} = A \hat{Z}_{n-1}, \end{aligned}$$

откуда следует важное соотношение

$$\hat{Z}_{n,n-1} = A \hat{Z}_{n-1}. \quad (4.56)$$

Выражения (4.55), (4.56) позволяют следующим образом отредактировать формулу (4.54):

$$\begin{aligned}\hat{Z}_n &= (R_{n,n-1} - R_{n,n-1}C(C^T R_{n,n-1}C + \sigma_p^2)^{-1}C^T R_{n,n-1})(R_{n,n-1}^{-1}A\hat{Z}_{n-1} + \\ &\quad + \sigma_p^{-2}Cy_n) = A\hat{Z}_{n-1} + R_{n,n-1}\sigma_p^{-2}Cy_n - R_{n,n-1}C(C^T R_{n,n-1}C + \\ &\quad + \sigma_p^2)^{-1}C^T R_{n,n-1}(R_{n,n-1}^{-1}A\hat{Z}_{n-1} + \sigma_p^{-2}Cy_n) = \\ &= A\hat{Z}_{n-1} + \sigma_p^{-2}R_{n,n-1}C(C^T R_{n,n-1}C + \sigma_p^2)^{-1}(C^T R_{n,n-1}C + \sigma_p^2)y_n - \\ &\quad - R_{n,n-1}C(C^T R_{n,n-1}C + \sigma_p^2)^{-1}C^T R_{n,n-1}(R_{n,n-1}^{-1}A\hat{Z}_{n-1} + \sigma_p^{-2}Cy_n) = \\ &= A\hat{Z}_{n-1} + R_{n,n-1}C(C^T R_{n,n-1}C + \sigma_p^2)^{-1}(\sigma_p^{-2}(C^T R_{n,n-1}C + \sigma_p^2)y_n - \\ &\quad - C^T R_{n,n-1}(R_{n,n-1}^{-1}A\hat{Z}_{n-1} + \sigma_p^{-2}Cy_n)).\end{aligned}$$

Обозначив

$$K_n = R_{n,n-1}C(C^T R_{n,n-1}C + \sigma_p^2)^{-1}, \quad (4.57)$$

из последнего выражения получим

$$\hat{Z}_n = A\hat{Z}_{n-1} + K_n(y_n - C^T A\hat{Z}_{n-1}), \quad n = 1, 2, \dots, N. \quad (4.58)$$

Выражение (4.58) позволяет последовательно вычислять оценки $\hat{Z}_1, \hat{Z}_2, \hat{Z}_3, \dots$, причем каждая последующая оценка выражается через предыдущую и очередное наблюдение, т.е. алгоритм носит рекуррентный характер. Это выражение часто называют *уравнением фильтрации*.

Для завершения алгоритма необходимы дополнительные соотношения, определяющие матрицу $R_{n,n-1}$ в выражении (4.57). Займемся их поиском.

Обратимся к выражению (4.52) и рассмотрим его правую часть, т.е. функцию $\omega(Z_n|y_1^{n-1})\omega(y_n|Z_n)/\omega(y_n|y_1^{n-1})$. Дополнительно к уже найденным определим условную плотность $\omega(y_n|y_1^{n-1})$. При линейной гауссовской модели (4.49), (4.50) эта плотность также является гауссовской и определяется двумя параметрами — условными математическим ожиданием и дисперсией. Найдем их:

$$\begin{aligned}M\{y_n|y_1^{n-1}\} &= M\{C^T Z_n + p_n|y_1^{n-1}\} = C^T Z_{n,n-1}; \\ M\{(y_n - C^T Z_{n,n-1})^2|y_1^{n-1}\} &= M\{(C^T(Z_n - Z_{n,n-1}) + p_n)^2|y_1^{n-1}\} = \\ &= C^T R_{n,n-1}C + \sigma_p^2\end{aligned}$$

и, таким образом, $\omega(y_n|y_1^{n-1}) = N(C^T \hat{Z}_{n,n-1}, C^T R_{n,n-1}C + \sigma_p^2)$. Теперь, используя определение гауссовской плотности, построим

функцию $\rho(Z_n)$, являющуюся показателем экспоненты в выражении $\omega(Z_n|y_1^{n-1})\omega(y_n|Z_n)/\omega(y_n|y_1^{n-1})$:

$$\begin{aligned}\rho(Z_n) &= -\frac{1}{2}(Z_n - \hat{Z}_{n,n-1})^T R_{n,n-1}^{-1}(Z_n - \hat{Z}_{n,n-1}) - \\ &\quad - \frac{1}{2}\sigma_p^{-2}(y_n - C^T Z_n)^2 + \frac{1}{2}(C^T R_{n,n-1}C + \sigma_p^2)^{-1}(y_n - C^T \hat{Z}_{n,n-1})^2.\end{aligned}$$

Разложим эту функцию в ряд Тейлора в окрестности точки $\hat{Z}_n$. Так как функция квадратичная, разложение будет содержать лишь три слагаемых:

$$\rho(Z_n) = \rho(\hat{Z}_n) + \frac{\partial \rho(\hat{Z}_n)}{\partial Z_n}(Z_n - \hat{Z}_n) + \frac{1}{2}(Z_n - \hat{Z}_n)^T \frac{\partial^2 \rho(\hat{Z}_n)}{\partial Z_n^2}(Z_n - \hat{Z}_n).$$

В силу оптимальности оценки $\hat{Z}_n$ выполняется условие $\frac{\partial \rho(\hat{Z}_n)}{\partial Z_n} = 0$. Так как в (4.52) $\omega(Z_n|y_1^n) = N(\hat{Z}_n, R_n)$, то показатель экспоненты плотности $\omega(Z_n|y_1^n)$ при $Z_n = \hat{Z}_n$ обращается в нуль. Но тогда и функция $\rho(Z_n)$ в этой точке обращается в нуль, т.е. $\rho(\hat{Z}_n) = 0$, и, следовательно,

$$\rho(Z_n) = \frac{1}{2}(Z_n - \hat{Z}_n)^T \frac{\partial^2 \rho(\hat{Z}_n)}{\partial Z_n^2}(Z_n - \hat{Z}_n).$$

Показатель экспоненты функции $\omega(Z_n|y_1^n) = N(\hat{Z}_n, R_n)$ равняется $-\frac{1}{2}(Z_n - \hat{Z}_n)^T R_n^{-1}(Z_n - \hat{Z}_n)$, и так как должно выполняться равенство (4.52), он должен равняться функции $\rho(Z_n)$. Но это равенство достигается, если $\frac{\partial^2 \rho(\hat{Z}_n)}{\partial Z_n^2} = (-R_n^{-1})$, что после дифферен-

цирования приводит к соотношению $R_n^{-1} = R_{n,n-1}^{-1} + \sigma_p^{-2}CC^T$. Обращая обе части этого равенства и применяя к правой части лемму об обращении матрицы (4.55), с учетом (4.57) получаем

$$R_n = (E - K_n C^T)R_{n,n-1}. \quad (4.59)$$

Теперь установим связь между матрицами $R_{n, n-1}$ и R_{n-1} . Для этого рассмотрим последовательность равенств

$$\begin{aligned} R_{n, n-1} &= M\{(Z_n - \hat{Z}_{n, n-1})(Z_n - \hat{Z}_{n, n-1})^T | y_1^{n-1}\} = \\ &= M\{[A(Z_{n-1} - \hat{Z}_{n-1}) + Bx_{n-1}][A(Z_{n-1} - \hat{Z}_{n-1}) + Bx_{n-1}]^T | y_1^{n-1}\} = \\ &= AR_{n-1}A^T + \sigma_x^2 BB^T, \end{aligned}$$

из которых следует

$$R_{n, n-1} = AR_{n-1}A^T + \sigma_x^2 BB^T. \quad (4.60)$$

Совокупность соотношений (4.57) — (4.60) образует *рекуррентный калмановский* алгоритм оценивания стохастического вектора состояния. Вычисления по этому алгоритму организуются следующим образом.

1. В соответствии с априорной информацией $Z_0 \sim N(m_{z0}, K_{z0})$ принимаются начальные условия для алгоритма $\hat{Z}_0 = m_{z0}$, $R_0 = K_{z0}$.

2. Пусть $n = 1$.

2.1. В соответствии с (4.60) вычисляется матрица $R_{1,0}$.

2.2. На основе (4.57) находится вектор K_1 .

2.3. Используя (4.58), находят оценку $\hat{Z}_1$ вектора Z_1 , соответствующую наблюдению y_1 .

2.4. Из выражения (4.59) находится матрица R_1 .

3. Пусть $n = 2$.

3.1. На основе (4.60) вычисляется матрица $R_{2,1}$.

3.2. В соответствии с (4.57) вычисляется вектор K_2 .

3.3. Из (4.58) находят оценку $\hat{Z}_2$, соответствующую наблюдениям y_1, y_2 .

3.4. Вычисляют матрицу R_2 , используя (4.59).

Последующие расчеты проводятся аналогичным образом при $n = 3, 4, \dots, N$, результатом чего будет оценка $\hat{Z}_N$. Прогнозированное значение $\hat{Y}_m = C^T Z_m$ ряда обосновывается результатом (4.56) и принимает вид $\hat{Y}_m = C^T A^{m-N} \hat{Z}_N$. Апостериорная ковариационная матрица R_N определяет точность оценивания вектора состояния Z_N , соответствующая величина $C^T A^{m-N} - R_N (A^{m-N})^T C$ характеризует точность прогнозирования.

4.13. Нелинейная параметрическая идентификация модели стохастического временного ряда

При построении калмановского алгоритма прогнозирования предполагалось, что параметры A, B, C модели (4.49), (4.50) и входящих в нее белых шумов известны. Однако одна из особенностей эконометрических задач проявляется в том, что это предположение оказывается излишне оптимистическим, и параметры, прежде чем решать задачу прогнозирования, еще следует определить. Источником соответствующей информации при этом является сам временной ряд $y_1, y_2, \dots, y_N$. Ранее уже отмечалось, что для традиционных моделей ряда вида AR, MA, ARMA разработаны алгоритмы оценивания их параметров, широко представленные в литературе. Поэтому, воспользовавшись этими методами, можно идентифицировать неизвестные параметры, а затем построить модель ряда в терминах стохастического вектора состояния, но уже с известными параметрами, и решить задачу прогнозирования калмановскими средствами. Покажем, что калмановская идеология позволяет избежать этой двухэтапной процедуры и совместить в рамках общего алгоритма решение задач идентификации и прогнозирования.

Обратимся к соотношениям (4.49), (4.50). Неизвестными параметрами этой модели, что следует из (4.48), являются первый столбец матрицы A , вектор B , первый элемент вектора C , дисперсия σ_p^2 . Введем для них обозначения:

$$Z_1 = \begin{bmatrix} a_{11} \\ a_{12} \\ \dots \\ a_{1k} \end{bmatrix}, \quad Z_2 = B, \quad Z_3 = \begin{bmatrix} c_1 \\ \sigma_p \end{bmatrix}.$$

Так как эти параметры являются стационарными и во времени не меняются, условно их можно задать как решения простейших разностных уравнений:

$$Z_{1,n} = Z_{1,n-1}, \quad Z_{2,n} = Z_{2,n-1}, \quad Z_{3,n} = Z_{3,n-1}. \quad (4.61)$$

Введем в рассмотрение блочные векторы и матрицу соответственно:

$$Z^* = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z \end{bmatrix} \in R^{3k+2}, \quad B^* = \begin{bmatrix} 0_{2k+2} \\ Z_2 \end{bmatrix} \in R^{3k+2},$$

$$A^* = \begin{bmatrix} E_k & 0_{k \times k} & 0_{k \times 2} & 0_{k \times k} \\ 0_{k \times k} & E_k & 0_{k \times 2} & 0_{k \times k} \\ 0_{2 \times k} & 0_{2 \times k} & E_2 & 0_{2 \times k} \\ 0_{k \times k} & 0_{k \times k} & 0_{k \times 2} & A(Z_1) \end{bmatrix} \in R^{(3k+2)(3k+2)},$$

где $0_{i \times j}$ — нулевая матрица размерностью i на j , 0_i — нулевой i -вектор и подчеркнута зависимость матрицы A от вектора Z_1 . Воспользовавшись этими обозначениями, уравнения (4.50), (4.61) можно представить в виде одного $(3k+2)$ -мерного уравнения $Z_n^* = A^*(Z_{1,n-1})Z_{n-1}^* + B^*(Z_{2,n-1})x_{n-1}$ или же

$$Z_n^* = \Phi(Z_{n-1}^*) + B^*(Z_{n-1}^*)x_{n-1}, \quad \Phi(Z_{n-1}^*) = A^*(Z_{1,n-1})Z_{n-1}^*. \quad (4.62)$$

Вектор Z_n^* назовем *расширенным стохастическим вектором* состояния. Особенность описывающего его динамику разностного уравнения (4.62) проявляется в том, что это уравнение, в отличие от (4.50), является нелинейным.

Далее введем в рассмотрение $(3k+2)$ -мерную вектор-строку $G^{(s)}$, у которой на s -й позиции находится единица, а остальные элементы равны нулям. Тогда слагаемое $C^T Z_n$ из (4.49) с использованием новых обозначений можно записать как $G^{(2k+1)} Z_n^* G^{(2k+3)} Z_n^*$. Второе слагаемое p_n из этого же выражения, имеющее неизвестную дисперсию σ_p^2 , следующим образом выражается через стандартную гауссовскую величину p_n : $p_n = G^{(2k+2)} Z_n^* p_n$. Само уравнение (4.49) в новых обозначениях приобретает вид

$$y_n = h(Z_n^*) + G^{(2k+2)} Z_n^* p_n, \quad (4.63)$$

где нелинейная функция $h(Z_n^*) = G^{(2k+1)} Z_n^* G^{(2k+3)} Z_n^*$. Таким образом, и модель наблюдений (4.49) при неизвестных параметрах оказывается нелинейно зависящей от расширенного вектора состояния Z_n^* . Последующую задачу можем сформулировать так.

Временной ряд представлен уровнями, математически выражаемыми через ненаблюдаемый расширенный вектор состояния Z_n^* в соответствии с (4.63). Сам вектор состояния формируется из порождающего белого шума, как это предусмотрено разностным уравнением (4.62). Случайные составляющие x_n и p_n^* в обоих уравнениях являются независимыми гауссовскими белыми шумами с единичными дисперсиями. Задача, как и в предыдущем разделе, заключается в поиске оценки $\hat{Z}_n^*$ вектора состояния Z_n^* по вектору наблюдений y_1^n и в последующем использовании этой оценки в целях прогнозирования. Отличие этой задачи от предыдущей проявляется прежде всего в нелинейной структуре моделей (4.62), (4.63). Иным будет и результат решения задачи: в этом случае наряду с оцениванием вектора состояния Z_n проводится оценивание и неизвестных параметров модели (4.49), (4.50), включая дисперсии случайных составляющих. Платой за перспективу совместной параметрической идентификации и, как говорят, фильтрации оказывается размерность задачи.

Решение сформулированной задачи, как и выше, ищем в соответствии с критерием максимума апостериорной плотности вероятностей, аналогичным (4.51). Однако нелинейный характер модели существенно усложняет как сам процесс поиска точного решения, так и соответствующий алгоритм. Поэтому удобнее прибегать к помощи различных процедур линеаризации нелинейностей для получения более простых алгоритмов, подобных их линейному аналогу. Если, вооружившись идеей линеаризации, пройти путь, подобный приведенному к алгоритму (4.57)–(4.60), получим систему рекуррентных соотношений для совместной параметрической идентификации модели и фильтрации вектора состояния. В систематизированном виде эти уравнения таковы:

алгоритм фильтрации

$$\hat{Z}_n^* = \hat{Z}_{n,n-1}^* + K_n(y_n - h(\hat{Z}_{n,n-1}^*)); \quad (4.64)$$

алгоритм одношагового прогнозирования

$$\hat{Z}_{n,n-1}^* = \Phi(\hat{Z}_{n-1}^*); \quad (4.65)$$

априорная ковариационная матрица ошибок

$$R_{n,n-1} = \frac{\partial}{\partial Z^*} \Phi(Z^*) R_{n-1} \frac{\partial}{\partial Z^*} \Phi^T(Z^*) + B^*(Z^*) B^{*T}(Z^*) \quad (4.66)$$

при $Z^* = \hat{Z}_{n-1}^*$;

апостериорная ковариационная матрица ошибок

$$R_n = R_{n,n-1} - R_{n,n-1} \left(\frac{\partial}{\partial Z^*} h(Z^*) \right)^T \left(\frac{\partial}{\partial Z^*} h(Z^*) \right) R_{n,n-1} \left(\frac{\partial}{\partial Z^*} h(Z^*) \right)^T +$$

$$+ G^{(2k+2)} R_{n,n-1} (G^{(2k+2)})^T)^{-1} \frac{\partial}{\partial Z^*} h(Z^*) R_{n,n-1} \quad \text{при } Z^* = \hat{Z}_{n,n-1}^*; \quad (4.67)$$

коэффициент усиления

$$K_n = R_n \left(\frac{\partial}{\partial Z^*} h(Z^*) \right)^T (G^{(2k+2)} R_{n,n-1} (G^{(2k+2)})^T)^{-1} \quad (4.68)$$

при $Z^* = \hat{Z}_{n,n-1}^*$.

Входящие в выражения (4.66)–(4.68) производные могут быть конкретизированы. Так как в данном случае

$\Phi(Z^*) = A^*(Z_1)Z = [Z_1^T Z_2^T Z_3^T Z^T A^T(Z_1)]^T$, то в блочных обозначениях

$$\frac{\partial}{\partial Z^*} \Phi(Z^*) = \begin{bmatrix} E_k & 0_{k \times k} & 0_{k \times 2} & 0_{k \times k} \\ 0_{k \times k} & E_k & 0_{k \times 2} & 0_{k \times k} \\ 0_{2 \times k} & 0_{2 \times k} & E_2 & 0_{2 \times k} \\ z_1 E_k & 0_{k \times (k+2)} & Z_1 & E_{k-1}^* \end{bmatrix} \in R^{(3k+2) \times (3k+2)}, \quad (4.69)$$

где через z_1 обозначен первый компонент вектора Z , $s = [1 \ 1 \ \dots \ 1]^T \in R^k$, $E_{k-1}^* \in R^{k \times (k-1)}$ и представляет собой единичную $(k-1)$ -матрицу E_{k-1} , окаймленную снизу нулевой строкой.

Аналогичным образом несложно установить

$$\frac{\partial}{\partial Z^*} h(Z^*) = \frac{\partial}{\partial Z^*} G^{(2k+1)} Z^* G^{(2k+3)} Z^* = G^{(2k+1)} (G^{(2k+3)} Z^*) +$$

$$(G^{(2k+1)} Z^*) G^{(2k+3)} = [0_{2k}^T \ z_{2k+3}^* \ 0 \ z_{2k+1}^* \ 0_{k-1}^T] \in R^{3k+2}, \quad (4.70)$$

где z_i^* — i -й компонент вектора Z^* . Таким образом, в данной задаче матрица (4.69) и вектор-строка (4.70) существенно разрежены, что, несмотря на возросшую размерность задачи, способствует упрощению программной реализации алгоритма (4.64)–(4.68). Организация вычислений в соответствии с этим алгоритмом осуществляется так же, как и в случае (4.57)–(4.60). Отличие проявляется лишь в том, что вычислению оценки $\hat{Z}_n^*$ предшествует вычисление прогноза $\hat{Z}_{n,n-1}^*$ по правилу (4.65).

Это же правило используется для проведения одношагового прогнозирования после обработки всех уровней временного ряда:

$$\hat{Y}_{N+1} = h(\hat{Z}_{N+1,N}^*) = h(\Phi(\hat{Z}_N^*)).$$

Если принять $\hat{Z}_{N+2,N}^* = \Phi(\hat{Z}_{N+1,N}^*)$, то можно построить прогноз на два шага, и т. д. (см. приложение 2).

4.14. Обобщенный рекуррентный алгоритм прогнозирования стохастических временных рядов

При построении и последующем изучении модели (4.49), (4.50) было опущено слагаемое $b x_n$ в составе соотношения (4.49). Хотя во многих прикладных задачах это условие выполняется, модель (4.49), порожденная традиционными эконометрическими моделями типа AR, MA, ARMA, ARIMA, это слагаемое содержит. Поэтому целесообразно калмановский алгоритм прогнозирования обобщить и на этот случай.

Итак, пусть временной ряд представлен моделями

$$y_n = C^T Z_n + b x_n + p_n, \quad n = 1, 2, \dots, N, \quad (4.71)$$

$$Z_n = A Z_{n-1} + B x_{n-1}. \quad (4.72)$$

Отличительная особенность этого представления временного ряда проявляется в том, что теперь шумы $p_n^* = b x_n + p_n$ и $B x_{n-1}$ в обоих уравнениях оказываются коррелированными, а именно: $M\{p_n^* x_k\} = q \delta_{n,k}$, где $q = M\{p_n^* x_n\} = b \sigma_x^2$ и $\delta_{n,k}$ — дельта-символ Кронекера. Чтобы воспользоваться прежним способом вывода алгоритма прогнозирования, воспользуемся идеей «раскоррелирования» шумов [20]. Существо идеи заключается в следующем.

Перепишем уравнение (4.72) в таком виде:

$$Z_{n+1} = AZ_n + Bx_n + W(y_n - C^T Z_n - bx_n - p_n),$$

или же

$$Z_{n+1} = A^* Z_n + x_n^* + Wy_n, \quad A^* = A - WC^T, \quad x_n^* = (B - bW)x_n - Wp_n.$$

Найдем такой вектор W , при котором величины x_n^* и p_n^* окажутся некоррелированными, т.е. потребуем $M\{x_n^* p_n^*\} = M\{(B - bW)x_n - Wp_n\}(bx_n + p_n) = 0$. Усреднив, получим уравнение $(B - bW)b\sigma_x^2 - W\sigma_p^2 = 0$, из которого следует

$$W = \frac{b\sigma_x^2}{b^2\sigma_x^2 + \sigma_p^2} B. \quad (4.73)$$

Таким образом, при выполнении (4.73) временной ряд описывается уравнениями

$$y_n = C^T Z_n + p_n^*, \quad (4.74)$$

$$Z_n = A^* Z_{n-1} + Wy_{n-1} + x_n^*, \quad (4.75)$$

эквивалентными (4.71), (4.72), но содержащими некоррелированные шумы и в этом смысле подобными (4.49), (4.50). Присутствие известного слагаемого Wy_{n-1} в (4.75) не препятствует теперь по модели (4.71), (4.72) получить алгоритм прогнозирования тем же образом, что и в случае (4.49), (4.50). Опуская доказательства, приведем окончательную редакцию алгоритма в обозначениях выражений (4.71), (4.72):

$$\begin{aligned} \hat{Z}_n &= \hat{Z}_{n,n-1} + K_n(y_n - C^T \hat{Z}_{n,n-1}), \\ \hat{Z}_{n,n-1} &= A \hat{Z}_{n-1} + W(y_{n-1} - C^T \hat{Z}_{n-1}), \\ K_n &= R_{n,n-1} C (C^T R_{n,n-1} C + b^2 \sigma_x^2 + \sigma_p^2)^{-1}, \\ R_{n,n-1} &= (A - WC^T) R_{n-1} (A - WC^T)^T + \sigma_x^2 B B^T + (b^2 \sigma_x^2 + \sigma_p^2) W W^T, \\ R_n &= (E - K_n C^T) R_{n,n-1}. \end{aligned}$$

Это правило рекуррентных вычислений мы и называем обобщенным калмановским алгоритмом. Конкретная организация вычислений по этому алгоритму проводится точно так же, как и в случае (4.57)–(4.60). Отличие, как и в предыдущем случае, проявляется только в вычислении прогнозируемого значения $\hat{Z}_{n,n-1}$, предшествующего оценке $\hat{Z}_n$.

ПРИЛОЖЕНИЕ 1

Операционные методы исследования динамических систем

При исследовании различного рода явлений, процессов, закономерностей и т.п. (обобщенно-динамических систем), описываемых линейными дифференциальными или разностными уравнениями с постоянными параметрами, широкое распространение получили так называемые операционные методы, основанные на преобразовании Лапласа и z -преобразовании. Привлекательная сторона этих методов проявляется в возможности превращения дифференциального или разностного уравнения в алгебраическое, содержащее одну неизвестную функцию и легко разрешимое относительно нее. Правда, полученное решение алгебраического уравнения определено в ином пространстве, нежели решения исходных уравнений. Поэтому требуются дополнительные усилия на переход в пространство искомых решений. Но это окупается общим упрощением процесса поиска решений. Кратко изложим существо этих методов.

Определение П1.1. Функцию $y(t)$, где для определенности t — время, называют *оригиналом*, если

- 1) $y(t) \equiv 0$ при $\forall t < 0$;
- 2) $y(t) \neq 0$ при всех или некоторых $t \geq 0$ и является однозначной, непрерывной или кусочно-непрерывной функцией;
- 3) $\exists M, c > 0$ такие, что

$$|y(t)| \leq M e^{ct}. \quad (\text{П1.1})$$

Определение П1.2. Функция $\tilde{y}(s)$ комплексного аргумента s называется *изображением* или прямым преобразованием Лапласа оригинала $y(t)$, если

$$\tilde{y}(s) = \int_0^{\infty} y(t) e^{-st} dt. \quad (\text{П1.2})$$

Утверждение П1.1. Если $y(t)$ — оригинал, то несобственный интеграл в (П1.2) сходится, причем абсолютно, при всех s , для ко-

торых $\operatorname{Re} s > c_0$, где c_0 — точная нижняя грань множества значений параметра s , удовлетворяющих неравенству (П1.1).

Соотношение (П1.2) часто записывают лаконично $\tilde{y}(s) = L\{y(t)\}$, понимая под $L\{\dots\}$ стоящий в правой части (П1.2) оператор (интеграл) Лапласа. Прямое преобразование Лапласа устанавливает по оригиналу изображение. Существует обратная операция, устанавливающая по изображению оригинал и известная как обратное преобразование Лапласа:

$$Y(t) = L^{-1}\{\tilde{y}(s)\} = \frac{1}{2\pi j} \int_{-j\infty+\mu}^{j\infty+\mu} \tilde{y}(s) e^{st} ds, \quad (\text{П1.3})$$

где j — мнимая единица ($j^2 = -1$) и интегрирование проводится в комплексной плоскости по прямой, параллельной мнимой оси и отстоящей от нее на расстоянии $\mu = \operatorname{Re} s$. Для элементарных и многих неэлементарных функций составлены таблицы соответствия оригиналов и изображений, а также разработаны операции вычисления оригиналов по изображениям без непосредственного использования интегральной процедуры (П1.3), но при продуктивном использовании свойств преобразования Лапласа.

Приведем без доказательств, не вызывающих больших затруднений, основные свойства преобразования Лапласа, используя при этом обозначение $\tilde{y}(s) = L\{y(t)\}$.

1. Теорема линейности

$$L\left\{\sum_{i=1}^k \alpha_i y_i(t)\right\} = \sum_{i=1}^k \alpha_i L\{y_i(t)\}, \quad \alpha_i = \text{const}. \quad (\text{П1.4})$$

2. Теорема подобия

$$L\{y(\alpha t)\} = \frac{1}{\alpha} \tilde{y}\left(\frac{s}{\alpha}\right), \quad \alpha = \text{const} \neq 0.$$

3. Теорема смещения

$$L\{e^{\alpha t} y(t)\} = \tilde{y}(s - \alpha), \quad \alpha = \text{const}.$$

4. Теория запаздывания оригинала

$$L\{y(t - \tau)\} = e^{-s\tau} \tilde{y}(s).$$

5. Теорема о дифференцировании изображения

$$L^{-1}\left\{\frac{d^n}{ds^n} \tilde{y}(s)\right\} = (-1)^n t^n y(t), \quad n = 1, 2, 3, \dots$$

6. Теорема о дифференцировании оригинала

$$L\{y^{(k)}(t)\} = s^k \tilde{y}(s) - s^{k-1} y(0_+) - s^{k-2} y^{(1)}(0_+) - \dots - y^{(k-1)}(0_+), \quad (\text{П1.5})$$

где $k = 1, 2, 3, \dots$, $y^{(k)}(t)$ — оригинал и $y^{(i)}(0_+) - i$ -я правая производная оригинала в точке $t = 0$.

7. Теорема об интегрировании оригинала

$$L\left\{\int_0^t y(\tau) d\tau\right\} = \frac{1}{s} \tilde{y}(s).$$

8. Теорема об интегрировании изображения

$$L\left\{\frac{1}{t} y(t)\right\} = \int_s^\infty \tilde{y}(s) ds.$$

9. Теорема об изображении периодической функции: если $y(t) = y(t + T)$, то

$$\tilde{y}(s) = \frac{1}{1 - e^{-Ts}} \int_0^T y(t) e^{-st} dt.$$

10. Теорема о свертке

$$L^{-1}\{\tilde{y}_1(s) \tilde{y}_2(s)\} = \int_0^t y_1(\tau) y_2(t - \tau) d\tau = \int_0^t y_1(t - \tau) y_2(\tau) d\tau, \quad (\text{П1.6})$$

где $y_1(t)$, $y_2(t)$ — оригиналы, соответствующие изображениям $\tilde{y}_1(s)$, $\tilde{y}_2(s)$.

11. Теорема о начальном значении оригинала

$$\lim_{t \rightarrow 0, t > 0} y(t) = \lim_{s \rightarrow \infty} s \tilde{y}(s).$$

12. Теорема о предельном значении оригинала

$$\lim_{t \rightarrow \infty} y(t) = \lim_{s \rightarrow 0} s \tilde{y}(s). \quad (\text{П1.7})$$

13. Формула Хевисайда: если $\tilde{y}(s) = \frac{B_m(s)}{A_n(s)}$, где $B_m(s)$ и $A_n(s)$ —

многочлены по степеням s соответственно m -го и n -го порядков, причем $n > m$, то при простых корнях s_i , $i = 1, 2, \dots, n$, уравнения $A_n(s) = 0$ справедливо

$$y(t) = \sum_{i=1}^n \frac{B_m(s_i)}{A'_n(s_i)} e^{s_i t}.$$

Существуют обобщения этой формулы на случай кратных корней того же уравнения $A(s) = 0$.

Рассмотрим существо операционного метода в связи с решением следующей задачи Коши. Пусть задано линейное дифференциальное уравнение с постоянными параметрами

$$\begin{aligned} \alpha_n y^{(n)}(t) + \alpha_{n-1} y^{(n-1)}(t) + \dots + \alpha_0 y(t) = \\ = \beta_m x^{(m)}(t) + \beta_{m-1} x^{(m-1)}(t) + \dots + \beta_0 x(t), \quad n > m, \end{aligned} \quad (\text{П1.8})$$

где α_i, β_j — известные постоянные параметры, $x(t)$ — заданная функция-оригинал. Требуется найти решение уравнения (П1.8), удовлетворяющее начальным условиям

$$y^{(i)}(0_+) = y_{0i}, \quad i = 0, 1, \dots, n-1. \quad (\text{П1.9})$$

Для решения задачи обе части уравнения (П1.8) подвергнем преобразованию Лапласа. Воспользовавшись теоремами линейности и дифференцирования оригинала и сгруппировав слагаемые, содержащие $\tilde{y}(s) = L\{y(t)\}$, $\tilde{x}(s) = L\{x(t)\}$ и одни и те же начальные условия, получим

$$A_n(s) \tilde{y}(s) - \sum_{i=0}^{n-1} \varphi_i(s) y_{0i} = B_m(s) \tilde{x}(s) - \sum_{k=0}^{m-1} \psi_k(s) x^{(k)}(0_+), \quad (\text{П1.10})$$

где использованы обозначения:

$$\begin{aligned} A_n(s) &= \alpha_n s^n + \alpha_{n-1} s^{n-1} + \dots + \alpha_0, \\ B_m(s) &= \beta_m s^m + \beta_{m-1} s^{m-1} + \dots + \beta_0, \end{aligned}$$

$$\varphi_i(s) = \alpha_n s^{n-1-i} + \alpha_{n-1} s^{n-2-i} + \dots + \alpha_{1+i},$$

$$i = 0, 1, \dots, n-2; \quad \varphi_{n-1}(s) = \alpha_n,$$

$$\psi_k(s) = \beta_m s^{m-1-k} + \beta_{m-1} s^{m-2-k} + \dots + \beta_{1+k},$$

$$k = 0, 1, \dots, m-2; \quad \psi_{m-1}(s) = \beta_m.$$

Полученное в результате преобразования выражение (П1.10) является линейным алгебраическим уравнением, содержащим единственную неизвестную функцию $\tilde{y}(s)$, которая легко находится из этого уравнения:

$$\tilde{y}(s) = \frac{B_m(s)}{A_n(s)} \tilde{x}(s) + \frac{\sum_{i=0}^{n-1} \varphi_i(s) y_{0i} - \sum_{k=0}^{m-1} \psi_k(s) x^{(k)}(0_+)}{A_n(s)}. \quad (\text{П1.11})$$

Функция (П1.11) является изображением искомого решения задачи (П1.8), (П1.9). Для поиска оригинала следует осуществить переход из пространства изображений в пространство оригиналов:

$$y(t) = L^{-1}\{\tilde{y}(s)\},$$

что во многих случаях делается на основании формулы Хевисайда (П1.7) или путем разложения изображения (П1.11) на простейшие слагаемые с последующим установлением оригиналов этих слагаемых. В этом и заключается суть операционного метода интегрирования линейного дифференциального уравнения с постоянными параметрами.

Соотношение (П1.11) существенно упрощается, если на начальные условия $y^{(i)}(0_+)$ никаких ограничений типа (П1.9) не наложено. В этом случае начальные условия порождены непосредственно функцией $x(t)$, а это приводит к тому, что второе слагаемое в (П1.11) обращается в нуль, и, как следствие, оказывается, что

$$\tilde{y}(s) = \frac{B_m(s)}{A_n(s)} \tilde{x}(s). \quad (\text{П1.12})$$

Комплексный коэффициент пропорциональности $W(s) = \frac{B_m(s)}{A_n(s)}$, связывающий изображения $\tilde{y}(s)$ и $\tilde{x}(s)$, принято называть *передаточной функцией* динамической системы, описываемой диффе-

ренциальным уравнением (П1.8). Очевидно, передаточная функция сама является изображением решения уравнения (П1.8), но при такой функции $x(t)$, изображение которой $\tilde{X}(s) = 1$. Такой функцией является дельта-функция Дирака $\delta(t)$, обладающая свойствами:

$$\delta(t) = \begin{cases} \infty, & t = 0 \\ 0, & t \neq 0 \end{cases} \quad \int_{-\varepsilon}^{\varepsilon} \delta(t) dt = 1, \quad \forall \varepsilon > 0, \quad \int_{-\varepsilon}^{\varepsilon} \delta(t) z(t) dt = z(0).$$

В силу последнего равенства принимается

$$L\{\delta(t)\} = \int_0^{\infty} \delta(t) e^{-st} dt = 1.$$

Заметим, что иногда значение этого интеграла считают равным $\frac{1}{2}$. Решение уравнения (П1.8) при $x(t) = \delta(t)$ называют *весовой функцией*, или импульсной характеристикой динамической системы. Обозначим ее символом $\omega(t)$, т.е. $\omega(t) = L^{-1}\{W(s)\}$. Тогда оригинал изображения (П1.12) можно найти на основании теоремы о свертке

$$y(t) = \int_0^t \omega(\tau) x(t-\tau) d\tau = \int_0^t \omega(t-\tau) x(\tau) d\tau. \quad (\text{П1.13})$$

Каждое из интегральных выражений в этом соотношении называют *оператором свертки*. Оператор свертки, таким образом, позволяет при отсутствии ограничений на начальные условия найти решение уравнения (П1.8) при произвольной функции $x(t)$, если известно решение при $x(t) = \delta(t)$, представленное функцией $\omega(t)$.

Аналогичная методология решения распространяется на линейные разностные уравнения. Изложению соответствующего подхода, как и ранее, предположим ряд начальных определений.

Определение П1.3. Пусть $y(t)$, $t \in T$ — непрерывная или кусочно непрерывная функция, определенная на непрерывном множестве T , и $\{t_0, t_1, t_2, \dots\} \subset T$ — дискретное подмножество точек из T , причем $t_{n+1} - t_n = \text{const}$, $n = 0, 1, \dots$. Тогда функция

$$y[n] = \begin{cases} y(t), & t = t_n, \quad n = 0, 1, 2, \dots, \\ 0, & t \neq t_n \end{cases}$$

называется *решетчатой*.

С решетчатыми функциями связывают ряд понятий, уподобляющих их непрерывным. В частности, функцию

$$\Delta y[n] = y[n+1] - y[n]$$

называют первой разностью, или разностью первого порядка решетчатой функции и рассматривают как своеобразный аналог производной от непрерывной функции. Первую разность первых разностей принято называть второй разностью (второго порядка) и обозначать

$$\Delta^2 y[n] = \Delta y[n+1] - \Delta y[n].$$

Подобным образом первую разность $(k-1)$ -х разностей определяют как k -ю разность

$$\Delta^k y[n] = \Delta^{k-1} y[n+1] - \Delta^{k-1} y[n].$$

Воспользовавшись методом математической индукции, не сложно убедиться, что k -я разность выражается непосредственно через значения решетчатой функции по правилу

$$\Delta^k y[n] = \sum_{v=0}^k (-1)^v C_k^v y[n+k-v], \quad C_k^v = \frac{k!}{v!(k-v)!}, \quad (\text{П1.14})$$

и наоборот,

$$y[n+q] = \sum_{i=0}^q C_q^i \Delta^i y[n].$$

Функцию $Y[n]$, для которой $\Delta Y[n] = y[n]$, называют *первообразной функции* $y(t)$. Справедлив непосредственно проверяемый результат:

$$Y[n] = \sum_{i=0}^{n-1} y[i] + c,$$

где $c = \text{const}$, в частности $c = 0$.

Пусть $y[n]$ — неизвестная решетчатая функция, $x[n]$ — известная. Тогда алгебраическое соотношение, связывающее разности различных порядков функций $y[n]$ и $x[n]$, называют *разностным уравнением*. Если разности выразить в соответствии с (П1.14), уравнение окажется записанным в терминах самих решетчатых

функций. Распространенный вариант такого уравнения имеет вид

$$\alpha_0 y[n] + \alpha_1 y[n+1] + \alpha_2 y[n+2] + \dots + \alpha_k y[n+k] = \beta_0 x[n] + \beta_1 x[n+1] + \beta_2 x[n+2] + \dots + \beta_m x[n+m], \quad (\text{П1.15})$$

где α_i, β_j — заданные константы и $k \geq m$. В этом случае уравнение называют линейным разностным k -го порядка с постоянными параметрами. Поиск решения этого уравнения, удовлетворяющего начальным условиям:

$$\begin{aligned} y[0] &= y_0, \\ y[1] &= y_1, \\ &\dots\dots\dots, \\ y[k-1] &= y_{k-1}, \end{aligned} \quad (\text{П1.16})$$

называют по аналогии с дифференциальными уравнениями задач Коши. Хотя эта задача легко решается в «числе», используют различные аналитические подходы с целью получения решения в аналитической форме. Среди таких подходов широкое распространение получил операционный. Его сущность подобна основанному на преобразовании Лапласа варианту, но отражает особенности (в частности, недифференцируемость) решетчатых функций.

Определение П1.4. Решетчатую функцию $y[n]$ со свойствами:

- 1) $y[n] \equiv 0, n < 0$,
- 2) $y[n] \neq 0$ при всех или некоторых $n \geq 0$,
- 3) $\exists M, c > 0$ такие, что $|y[n]| \leq Me^{cn}$

называют *оригиналом*.

Определение П1.5. Функция $\tilde{y}(z)$ комплексного аргумента z называется *изображением*, или z -преобразованием оригинала $y[n]$, если

$$\tilde{y}(z) = Z\{y[n]\} = \sum_{n=0}^{\infty} y[n]z^{-n}. \quad (\text{П1.17})$$

Утверждение П1.2. Если $y[n]$ — оригинал, то ряд в (П1.17) сходится, причем абсолютно, при всех z , для которых $|z| > e^{c_0}$, где c_0 определено аналогичным утверждению П1.1 образом.

Изображение $\tilde{y}(z)$ отождествляется с суммой ряда (П1.17). Существует обратная операция, позволяющая установить оригинал по изображению и называемая обратным z -преобразованием:

$$y[n] = Z^{-1}\{\tilde{y}(z)\} = \frac{1}{2\pi j} \int_G \tilde{y}(z) z^{n-1} dz, \quad n \geq 0, \quad (\text{П1.18})$$

где G — любой несамопересекающийся контур, содержащий внутри себя окружность радиуса e^{c_0} . Интеграл (П1.18) может быть вычислен в соответствии с теоремой о вычетах. Как и для непрерывных функций, построены обширные таблицы соответствия решетчатых оригиналов и их изображений.

Опуская достаточно несложные доказательства, приведем наиболее характерные свойства z -преобразования.

1. $Z\left\{\sum_i c_i y_i[n]\right\} = \sum_i c_i Z\{y_i[n]\}, \quad c_i = \text{const.}$
2. $Z\{y[n+k]\} = z^k \tilde{y}(z) - \sum_{i=0}^{k-1} z^{k-i} y[i].$
3. $Z\{y[n-k]\} = z^{-k} \tilde{y}(z), \quad k \geq 0.$
4. $Z\{\Delta^k y[n]\} = (z-1)^k \tilde{y}(z) - Z \sum_{v=0}^{k-1} (z-1)^{k-1-v} \Delta^v y[0].$
5. $Z\left\{\sum_{i=0}^{n-1} y[i]\right\} = \frac{1}{z-1} \tilde{y}(z).$
6. $Z^{-1}\{\tilde{y}_1(z) \tilde{y}_2(z)\} = \sum_{i=0}^n y_1[i] y_2[n-i] = \sum_{i=0}^n y_1[n-1] y_2[i],$

где $y_1[n]$ и $y_2[n]$ — оригиналы, соответствующие изображениям $\tilde{y}_1(z)$ и $\tilde{y}_2(z)$.

7. $y[0] = \lim_{z \rightarrow \infty} \tilde{y}(z).$
8. $\lim_{n \rightarrow \infty} y[n] = \lim_{z \rightarrow 1} (z-1) \tilde{y}(z).$

Рассмотрим теперь существо операционного метода решения задачи (П1.15), (П1.16). Логическая направленность метода пол-

ностью соответствует предыдущему изложению, относящемуся к дифференциальным уравнениям.

Вооружившись свойствами 1 и 2 z -преобразования и начальными условиями (П1.16), подвергаем обе части уравнения (П1.15) z -преобразованию. Обозначим $\tilde{y}(z) = Z\{y[n]\}$, $\tilde{x}(z) = Z\{x[n]\}$ и сгруппируем вместе слагаемые, содержащие $\tilde{y}(z)$, $\tilde{x}(z)$ и одинаковые начальные условия, обусловленные ограничением (П1.16) и процессом $x[n]$. В результате получим подобное (П1.10) алгебраическое уравнение с единственной неизвестной функцией $\tilde{y}(z)$:

$$A_k(z)\tilde{y}(z) - \sum_{i=0}^{k-1} \varphi_i(z)y_i = B_m(z)\tilde{x}(z) - \sum_{j=0}^{m-1} \psi_j(z)x[j],$$

откуда следует очевидное

$$\tilde{y}(z) = \frac{B_m(z)}{A_k(z)} \tilde{x}(z) + \frac{\sum_{i=0}^{k-1} \varphi_i(z)y_i - \sum_{j=0}^{m-1} \psi_j(z)x[j]}{A_k(z)}. \quad (\text{П1.19})$$

Смысл обозначений в (П1.19) таков:

$$A_k(z) = \alpha_0 + \alpha_1 z + \alpha_2 z^2 + \dots + \alpha_k z^k,$$

$$B_m(z) = \beta_0 + \beta_1 z + \beta_2 z^2 + \dots + \beta_m z^m,$$

$$\varphi_i(z) = \alpha_k z^{k-i} + \alpha_{k-1} z^{k-i-1} + \dots + \alpha_{i+1} z, \quad i = 0, 1, \dots, k-2;$$

$$\varphi_{k-1}(z) = \alpha_k z,$$

$$\psi_j(z) = \beta_m z^{m-j} + \beta_{m-1} z^{m-1-j} + \dots + \beta_{j+1} z, \quad j = 0, 1, \dots, m-2;$$

$$\psi_{m-1}(z) = \beta_m z.$$

Подвергнув изображение (П1.19) обратному z -преобразованию, получим решение исходной задачи (П1.15), (П1.16). Выражение (П1.19) значительно упрощается, если, подобно предыдущему, снять ограничения (П1.16) на начальные условия и считать их естественным проявлением влияния функции $x[n]$. Это приводит к взаимной компенсации определяемых начальными условиями слагаемых и в результате оказывается

$$\tilde{y}(z) = \frac{B_m(z)}{A_k(z)} \tilde{x}(z). \quad (\text{П1.20})$$

Последующие понятия подобны ранее введенным. Функцию

$$W(z) = \frac{B_m(z)}{A_k(z)}$$

называют *передаточной функцией* динамической системы, математическая модель которой сведена к разностному уравнению (П1.15). Понятно, что эта функция сама является изображением решения уравнения (П1.15), если $\tilde{y}(z) = 1$. Функцией с таким изображением является дельта-символ Кронекера

$$\delta_{n,0} = \begin{cases} 1, & n=0, \\ 0, & n \neq 0. \end{cases}$$

Следовательно, передаточная функция $W(z)$ представляет собой изображение решения разностного уравнения (П1.15) при $x[n] = \delta_{n,0}$ и отсутствующих ограничениях на начальные условия. Соответствующий оригинал $\omega[n] = Z^{-1}\{W(z)\}$, т.е. само решение уравнения (П1.15) при $x[n] = \delta_{n,0}$, принято называть *весовой функцией* динамической системы. Вычислив эту функцию и воспользовавшись свойством 6 z -преобразования, решение уравнения (П1.15) при произвольной функции $x[n]$ и выполнении (П1.20) можно представить в виде

$$y[n] = \sum_{i=0}^n \omega[i]x[n-i] = \sum_{i=0}^n \omega[n-i]x[i].$$

Каждую из этих сумм, подобно (П1.13), принято называть *оператором свертки*.

ПРИЛОЖЕНИЕ 2

Калмановское прогнозирование развития отдельных отраслей экономики России

Результаты прогнозирования, в частности краткосрочного (до 1 года), развития отдельных отраслей экономики России представляют большой интерес для многих управляющих органов страны. Это стимулирует разработку большого числа подходов к решению данной проблемы. Краткая аннотация этих подходов содержится в работе [31], где построена и успешно реализована авторская модель экономики России, представленная системой одновременных уравнений и идентифицированная на квартальных макроэкономических данных (4-й квартал 1994 г. — 1-й квартал 2001 г.) из официальных статистических источников. Покажем, что эти данные позволяют построить модели соответствующих временных рядов в терминах стохастического вектора состояния и использовать их для разработки алгоритмов прогнозирования калмановского типа, обладающих не худшими по сравнению с [31] характеристиками.

Данные, используемые нами далее для разработки стохастических моделей рядов, представлены на рис. П2.1 — П2.5. Здесь же указаны относительные ошибки прогнозирования, проведенного с помощью системы одновременных уравнений (СОУ) на 1 и 2 квартала 2001 г. [31] (2001:1 и 2001:2). Технология соответствующего анализа изложена в [31] и здесь не детализируется. Наша цель заключается в разработке по имеющимся апостериорным данным рекуррентного алгоритма прогнозирования калмановского типа и исследовании его свойств по аналогии с [31].

Визуальный анализ эмпирических данных позволяет интерпретировать их как аддитивную смесь квазидетерминированной и случайной составляющих. Первую из них в соответствии со структурой эмпирических данных аппроксимируем многочленом 3-го порядка. Случайную составляющую представляем в форме стохастического процесса, на который наложен независимый белый шум. Так как для всех 5 рядов техника формализации является общей, далее изложение ведется без конкретной адресации. Таким образом, если y_n — n -й уровень ряда, причем $n = 1, 2, \dots, 25$



Рис. П2.1. Эмпирические, модельные и прогнозные значения переменной CO



Рис. П2.2. Эмпирические, модельные и прогнозные значения переменной Y



Рис. П2.3. Эмпирические, модельные и прогнозные значения переменной X



Рис. П2.4. Эмпирические, модельные и прогнозные значения переменной N



Рис. П2.5. Эмпирические, модельные и прогнозные значения переменной M

в соответствии с объемом наблюдений, «пригодных» для разработки алгоритмов, то

$$y_n = f_n + \epsilon_n + p_n, n = 1, 25. \quad (\text{П2.1})$$

Детерминированная составляющая f_n описывается простейшей моделью $f_n = d_0 + d_1 n + d_2 n^2 + d_3 n^3$, или в векторных обозначениях

$$f_n = c_n^T d_n, \quad (\text{П2.2})$$

где

$$c_n^T = [1 \ n \ n^2 \ n^3], \ d_n = d_{n-1}, \ d_0^T = [d_0 \ d_1 \ d_2 \ d_3]. \quad (\text{П2.3})$$

Стохастическая составляющая ϵ_n отождествляется со случайной последовательностью типа ARMA и описывается разностным уравнением вида (4.32) $a_0 \epsilon_n + a_1 \epsilon_{n+1} + a_2 \epsilon_{n+2} = b_0 x_n + b_1 x_{n+1}$ с порождающим белым шумом x_n единичной интенсивности. В терминах состояния, как показано в разд. 4.11, это уравнение эквивалентно представлению

$$\varepsilon_n = \tilde{c}^T \tilde{Y}_n, \quad \tilde{Y}_n = \tilde{A} \tilde{Y}_{n-1} + \tilde{B} x_{n-1}, \quad (П2.4)$$

$$\tilde{A} = \begin{bmatrix} a_{11} & 1 \\ a_{21} & 0 \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix}, \quad \tilde{c} = \begin{bmatrix} c_1 \\ 0 \end{bmatrix},$$

причем элементы матриц $\tilde{A}$, $\tilde{B}$ и $\tilde{c}$ выражаются через параметры исходного уравнения. Введем теперь в рассмотрение следующие блочные конструкции:

$$Y_n = \begin{bmatrix} d_n \\ \tilde{Y}_n \end{bmatrix}, \quad A = \begin{bmatrix} E_4 & 0_{4 \times 2} \\ 0_{2 \times 4} & \tilde{A} \end{bmatrix}, \quad B = \begin{bmatrix} 0_4 \\ \tilde{B} \end{bmatrix}, \quad C_n = \begin{bmatrix} c_n \\ \tilde{c} \end{bmatrix}. \quad (П2.5)$$

Тогда ряд (П2.1), составляющие которого определяются выражениями (П2.2) – (П2.3), можем представить следующей обобщенной моделью:

$$y_n = C_n^T Y_n + p_n, \quad (П2.6)$$

$$Y_n = A Y_{n-1} + B x_{n-1}, \quad (П2.7)$$

в которой первое соотношение представляет собой уравнение наблюдения, а второе – уравнение стохастического состояния. Эта модель содержит ряд неизвестных параметров, именно: a_{11} , a_{21} , b_{11} , b_{21} , c_1 , σ_p , где σ_p – дисперсия помехи p_n . Поэтому задача прогнозирования должна сопровождаться решением задачи идентификации неизвестных параметров, т. е. в соответствии с разд. 4.13. Для построения подобной (4.62) и (4.63) модели формируется расширенный вектор состояния $Y^* = [a_{11} \ a_{21} \ b_{11} \ b_{21} \ c_1 \ \sigma_p \ d^T \ \tilde{Y}^T]^T \in R^{12}$, которому соответствуют структуры:

$$B^* = [0_{10}^T \ b_{11} \ b_{21}]^T, \quad A^* = \begin{bmatrix} E_{10} & 0_{10 \times 2} \\ 0_{2 \times 10} & \tilde{A} \end{bmatrix},$$

$$\Phi(Y^*) = [a_{11} \ a_{21} \ b_{11} \ b_{21} \ c_1 \ \sigma_p \ d^T \ (a_{11} \tilde{y}_1 + \tilde{y}_2) \ a_{21} \tilde{y}_1]^T,$$

$$h(Y^*, n) = c_n^T d + c_1 \tilde{y}_1,$$

где $\tilde{y}_1$, $\tilde{y}_2$ – компоненты вектора $\tilde{Y}$ и $d = [d_0 \ d_1 \ d_2 \ d_3]$. В терминах расширенного вектора состояния модель временного ряда приобретает вид:

$$y_n = h(Y_n^*) + G^{(6)} \dot{Y}_n^* p_n, \quad (П2.8)$$

$$Y_n^* = \Phi(Y_{n-1}^*) + B^*(Y_{n-1}^*) x_{n-1}. \quad (П2.9)$$

Соответствующий этой модели алгоритм совместной фильтрации, идентификации и прогнозирования определяется следующими вычислительными процедурами:

алгоритм фильтрации и идентификации

$$\hat{Y}_n^* = \hat{Y}_{n,n-1}^* + K_n(y_n - h(\hat{Y}_{n,n-1}^*));$$

алгоритм одношагового прогнозирования

$$\hat{Y}_{n,n-1}^* = \Phi(\hat{Y}_{n-1}^*);$$

априорная ковариационная матрица ошибок

$$R_{n,n-1} = \frac{\partial}{\partial Y^*} \Phi(Y^*) R_{n-1} \frac{\partial}{\partial Y^*} \Phi^T(Y^*) + B^*(Y^*) B^{*T}(Y^*) \text{ при } Y^* = \hat{Y}_{n-1}^*;$$

апостериорная ковариационная матрица ошибок

$$R_n = R_{n,n-1} - R_{n,n-1} \left(\frac{\partial}{\partial Y^*} h(Y^*) \right)^T \left(\frac{\partial}{\partial Y^*} h(Y^*) R_{n,n-1} \left(\frac{\partial}{\partial Y^*} h(Y^*) \right)^T + G^{(6)} R_{n,n-1} (G^{(6)})^T \right)^{-1} \frac{\partial}{\partial Y^*} h(Y^*) R_{n,n-1} \text{ при } Y^* = \hat{Y}_{n,n-1}^*;$$

коэффициент усиления

$$K_n = R_n \left(\frac{\partial}{\partial Y^*} h(Y^*) \right)^T \left(G^{(6)} R_{n,n-1} (G^{(6)})^T \right)^{-1} \text{ при } Y^* = \hat{Y}_{n,n-1}^*.$$

В соответствии с принятыми определениями в этом алгоритме

$$\frac{\partial}{\partial Y^*} \Phi(Y^*) = \begin{bmatrix} E_{10} & 0_{10 \times 2} \\ M(\tilde{y}_1) & \tilde{A} \end{bmatrix}, \quad M(\tilde{y}_1) = \begin{bmatrix} \tilde{y}_1 & 0 & 0 & \dots & 0 \\ 0 & \tilde{y}_1 & 0 & \dots & 0 \end{bmatrix} \in R^{2 \times 10},$$

$$\frac{\partial}{\partial Y^*} h(Y^*, n) = [0 \ 0 \ 0 \ 0 \ \tilde{y}_1 \ 0 \ 1 \ n \ n^2 \ n^3 \ c_1 \ 0].$$

Данный алгоритм совместно с определяющими его величинами положен в основу разработанной MATHCAD-программы (разработка программы и проведение соответствующих вычис-

ЛИТЕРАТУРА

1. Айвазян С. А., Мхитарян В. С. Прикладная статистика и основы эконометрики. — М.: ЮНИТИ, 1998.
2. Андерсон Т. Анализ временных рядов. — М.: Мир, 1976.
3. Афанасьев В. Н., Юзбашев М. М. Анализ временных рядов и прогнозирование. — М.: Финансы и статистика, 2001.
4. Бокс Дж., Дженкинс Г. Анализ временных рядов, прогноз и управление. — М.: Мир. Вып. 1, 2, 1974.
5. Боровиков В. П., Боровиков И. П. Statistica. Статистический анализ и обработка данных в среде Windows. — М.: Филинь, 1998.
6. Боровиков В. П., Ивченко Г. И. Прогнозирование в системе Statistica в среде Windows. — М.: Финансы и статистика, 2000.
7. Быков В. В. Цифровое моделирование в статистической радиотехнике. — М.: Сов. радио, 1971.
8. Воеводин В. В., Кузнецов Ю. А. Матрицы и вычисления. — М.: Наука, 1984.
9. Гантмахер Ф. Р. Теория матриц. — М.: Наука, 1988.
10. Гельфонд А. О. Исчисление конечных разностей. — М.: Наука, 1967.
11. Горчаков А. А., Орлова И. В. Компьютерные экономико-математические модели. — М.: ЮНИТИ, 1995.
12. Дуггерти К. Введение в эконометрику. — М.: Инфра-М, 2001.
13. Илюхин А. Г., Коваленко В. П. Численные методы обработки информации при исследовании динамических систем. — Киев.: Наукова думка, 1971.
14. Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. — М.: Наука, 1968.
15. Кремер Н. Ш., Путко Б. А. Эконометрика. — М.: ЮНИТИ, 2002.
16. Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования. — М.: Статистика, 1979.
17. Магнус Я. Р., Катышев П. К., Перецкий А. А. Эконометрика. Начальный курс. — М.: Дело, 1997.
18. Пакишин П. В. Дискретные системы со случайными параметрами и структурой. — М.: Физматлит, 1994.
19. Самарский А. А., Гулин А. В. Численные методы. — М.: Наука, 1989.
20. Сейдж Э., Мелс Дж. Теория оценивания и ее применение в связи и управлении. — М.: Связь, 1976.

лений выполнены экономистом-математиком А.Ю. Поляковым). Идентификация модели, сопровождаемая «подгонкой» модельных данных к каждому из 5 наблюдаемых рядов, осуществлялась циклическим применением алгоритма идентификации и фильтрации к каждому из рядов с использованием в качестве начальных условий очередного цикла конечных значений предыдущего. Для задач, не решаемых в натуральном масштабе времени, т. е. в процессе поступления данных, это обстоятельство не является обременительным. Начальные значения компонентов расширенного вектора состояния в 1-м цикле принимались равными 1, начальное значение апостериорной ковариационной матрицы задавалось в виде матрицы с единицами на главной диагонали и остальными элементами, равными 0,1.

Прогнозирование временного ряда на один шаг (квартал) проводится после обработки всех $N = 25$ его уровней по правилу

$$\hat{y}_{N+1} = h(\hat{Y}_{N+1, N}^*) = h(\Phi(\hat{Y}_N^*)).$$

Прогноз на два шага (квартала) сопровождается вычислением

$$\hat{y}_{N+2} = h(\hat{Y}_{N+2, N}^*), Y_{N+2, N}^* = \Phi(\hat{Y}_{N+1, N}^*).$$

Результаты соответствующего вычислительного эксперимента даны на рис. П2.1 — П2.5, где, помимо наблюдаемых уровней рядов и точностных характеристик прогноза по алгоритму [31], указаны модельные значения рядов, полученные после идентификации («подгонки») калмановских моделей, и относительные ошибки калмановского прогноза. Последние, как и в [31], понимаются в ретроспективном смысле, т. е. как отношение модуля отклонения прогнозируемого на данный квартал значения ряда от его реально наблюдаемого значения к этому наблюдаемому значению. Анализ полученных вычислительных материалов свидетельствует о полнейшей дееспособности построенных нелинейных алгоритмов совместной нелинейной фильтрации, идентификации и прогнозирования сложных экономических процессов. Полезно обратить внимание на то, что при их реализации удается ограничиться только апостериорными сведениями о самих прогнозируемых рядах и не прибегать к помощи дополнительной и весьма емкой информации, используемой в структуре одновременных уравнений.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Алгоритм

- градиентный 78
- оценивания 37
- калмановский 208
- скользящего среднего 13
- Аппроксимация регрессии 14

Вектор состояния

- стохастический 200
- — расширенный 210
- Величина смещения 40
- стандартная гауссовская случайная 19
- Время корреляции 175
- Выборка апостериорная 44

Гессиян 43

Градиент 42

- Дельта-символ Кронекера 129, 225
- Дельта-функция 171, 220
- Дисперсия случайного процесса 163

Задача

- корректно поставленная 60
- прогнозирования (предсказания, экстраполяции) 109
- Закон распределения вероятностей случайного процесса n -мерный 167

Идентификация и фильтрация совместная параметрическая 211

Информация априорная 37

Квантиль 20

Коэффициент

- детерминации 70
- корреляции 17
- — множественный 33
- — частный 32
- — эмпирический 22
- Коэффициенты Фурье 121

Критерий

- Дарбина—Уотсона 154
- идеального наблюдателя 24
- качества оценивания 37
- Неймана—Пирсона 25
- поворотных точек 153
- серий 153

Лемма об обращении матрицы 64

Математическое ожидание случайного процесса 163

Матрица

- апостериорная ковариационная 204
- априорная ковариационная 204
- Матрица
- Гессе 43
- идемпотентная 54
- ошибки оценивания ковариационная 40
- полного ранга 46
- Фишера информационная 41
- Якоби 43

Метод

- асимметрии и эксцесса 157
- наименьших квадратов рекуррентный 65
- неопределенных множителей Лагранжа 26, 48
- последовательной линеаризации 78
- МНК-оценка 46
- МНК-оценка обобщенная 89
- Многочлены
- Лежандра 119
- Чебышева второго рода 118
- Чебышева первого рода 118
- Модель
- авторегрессии скользящего среднего 194
- — проинтегрированного скользящего среднего 196
- авторегрессионная 194
- скользящего среднего 194
- Мощность критерия 23

21. Смирнов Н. В., Дунин-Барковский Н. В. Курс теории вероятностей и математической статистики для технических приложений. — М.: Наука, 1965.

23. Суэтин П. К. Классические ортогональные многочлены. — М.: Наука, 1976.

24. Тихонов А. Н., Уфимцев М. В. Статистическая обработка результатов эксперимента. — М.: МГУ, 1988.

25. Химмельблау Д. Прикладное нелинейное программирование. — М.: Мир, 1975.

26. Хорн Р., Джонсон Ч. Матричный анализ. — М.: Мир, 1989.

27. Четыркин Е. М. Статистические методы прогнозирования. — М.: Статистика, 1977.

28. Чуев Ю. В., Михайлов Ю. Б., Кузьмин В. И. Прогнозирование количественных характеристик процессов. — М., Сов. радио, 1975.

29. Чураков Е. П. Оптимальные и адаптивные системы. — М.: Энергоатомиздат, 1987.

30. Эконометрика/Под ред. И. И. Елисеевой. — М.: Финансы и статистика, 2002.

31. Эконометрическое моделирование: Учеб. пособие для вузов. — Вып. 2: Модель экономики России для целей краткосрочного прогноза и сценарного анализа/В. Л. Макаров, С. А. Айвазян, С. В. Борисов, Э. А. Лакалин. — М.: МЭСИ, 2002.

Неравенство
 — Бесселя 124
 — Коши—Буняковского 115
 — Минковского 115
 — Рао—Крамера 40
 Норма
 — вектора гильбертова 60
 — функции 114
 Нормы
 — векторные частные 60
 — матричные 61
 Оператор свертки 182, 185
 Оценка
 — байесовская 97
 — безусловная 37
 — достаточная 39
 — максимально правдоподобная 88
 — минимаксная 103
 — несмещенная 38
 — состоятельная 38
 — точечная 36
 — условная 37
 — эффективная 39
 — оптимальная по критерию максимума апостериорной плотности вероятностей 92
 Ошибка
 — аппроксимации 122
 — второго рода 23
 — оценивания 39
 — первого рода 23
 — прогнозирования 140
 — регрессионной модели относительная 157
 Переменная эндогенная 9
 Плотность
 — вероятностей апостериорная 92
 — случайного процесса n -мерная 167
 — — — одномерная 162
 — — случайных процессов совместная 172
 — — условная 166
 — непрерывного случайного процесса спектральная 177
 — случайной последовательности спектральная 179

Полиномы Лежандра — Чебышева
 дискретные 136
 Последовательность
 — Коши 112
 — минимизирующая 78
 — случайная 160
 — сходящаяся 78
 — фундаментальная 112
 Потери (риск) условные 103
 Преобразование Лапласа 215
 Проверка центрированности 153
 Прогноз калмановский 213
 Произведение функций скалярное 111, 114
 Пространство L_2 114
 — Банахово 113
 — Гильбертово 113
 — Евклидово 110
 — линейное 110
 — метрическое 110
 — нормированное 110
 — полное 113
 — функциональное 110
 Процесс случайный
 — дискретный 159
 — марковский 168
 — непрерывный 159
 — стационарный в узком смысле 173
 — — — в широком смысле 174
 — эргодический 176
 Равенство Парсеваля 127
 Разности решетчатой функции 221
 Раскоррелирование шумов 213
 Распределение
 — χ^2 19
 — вероятностей случайных процессов совместное 172
 — Гаусса 20
 — Стюдента 19
 — Фишера 20
 Реализация случайного процесса 159
 Регрессия гауссовская 16, 17
 Риск байесовский (средний) 97
 Ряд временной
 — стохастический 106
 — структурно детерминированный 106
 — Фурье обобщенный 121
 — формальный 122

Сглаживание
 — локальное 136
 — экспоненциальное 144
 Сечение случайного процесса 158
 Система
 — общего вида тригонометрическая 118
 — функций замкнутая 126
 — ортогональная 116
 — ортонормированная 116
 — основная тригонометрическая 117
 — полная 116
 Среднее апостериорное 99
 Статистика достаточная 39
 — Фишера 29
 Супермартингал 83
 Сходимость детерминированная 81
 — с вероятностью единица 83
 — среднеквадратичная 115
 — стохастическая 81
 Теорема Маркова 47
 Точка процентная 21
 Точность модели 157
 Уравнение
 — правдоподобия 89
 — разностное 185
 — характеристическое 183, 185
 Уровень значимости 23
 Условие нормировки 163, 165

Факторизация 187, 189
 Фильтр формирующий
 — дискретный 189
 — непрерывный 188
 Функции квадратично интегрируемые 113
 Функция
 — взаимная ковариационная 172
 — выпуска 51
 — Ляпунова 83
 — стохастическая 84
 — нормированная 116
 — потеря квадратичная 96
 — правдоподобия 88
 — производственная 50
 — распределения вероятностей случайного процесса одномерная 161
 — регрессии 12
 — решетчатая 220
 — случайного процесса ковариационная 169
 — — — корреляционная 170
 — стоимости (потерь) 88
 Число обусловленности 62
 Шум белый
 — дискретный 171
 — непрерывный 171
 Якобиан 43
 χ^2 -критерий Пирсона 156
 z -преобразование 222

ОГЛАВЛЕНИЕ

Предисловие	3
-------------------	---

Первая часть.

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ВОССТАНОВЛЕНИЯ ЗАВИСИМОСТЕЙ ПО ЭКСПЕРИМЕНТАЛЬНЫМ ДАННЫМ

Глава 1. Задачи регрессионного анализа и предварительный анализ данных

1.1. Проблема восстановления зависимостей по экспериментальным данным	7
1.2. Функция регрессии и регрессионная модель эндогенных переменных	11
1.3. Модели, аппроксимирующие функцию регрессии. Гауссовская регрессия	13
1.4. Некоторые специальные случайные величины и их свойства	19
1.5. Предварительный (дорегрессионный) анализ зависимости эндогенной и экзогенных переменных	21
1.5.1. Общие принципы	21
1.5.2. Критерий идеального наблюдателя	24
1.5.3. Критерий Неймана—Пирсона	25
1.5.4. Критерий проверки гипотезы H_0 при скалярной экзогенной переменной	27
1.5.5. Критерий проверки гипотезы H_0 при векторной экзогенной переменной	31

Глава 2. Методы оценивания параметров регрессионных моделей

2.1. Проблема оценивания и общие характеристики точечных оценок. Неравенство Рао—Крамера	35
2.2. Операции многомерного дифференцирования	41
2.3. Метод наименьших квадратов	44
2.3.1. МНК-оценки	44
2.3.2. Основные свойства МНК-оценок. Теорема Маркова	47
2.3.3. МНК-оценки параметров производственной функции Кобба—Дугласа	50

2.3.4. Идемпотентные матрицы	53
2.3.5. Несмещенная оценка дисперсии экспериментальных ошибок и ее свойства	55
2.3.6. Проблема обусловленности МНК-оценок. Векторные и матричные нормы	59
2.3.7. Рекуррентный метод наименьших квадратов	63
2.3.8. Коэффициент детерминации. Послерегрессионный анализ регрессионной модели	68
2.3.9. Нелинейные регрессионные модели. Проблема стохастической сходимости	77
2.4. Максимально правдоподобные оценки регрессионных параметров	86
2.5. Метод максимума апостериорной плотности вероятностей	91
2.6. Байесовские оценки регрессионных параметров	96
2.7. Минимаксные оценки регрессионных параметров ...	102

Вторая часть.

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ВРЕМЕННЫХ РЯДОВ

Глава 3. Структурно детерминированные временные ряды

3.1. Математические модели структурно детерминированных временных рядов	105
3.2. Ортонормированные системы функций	109
3.2.1. Банаховы и гильбертовы пространства	109
3.2.2. Пространство L_2	113
3.2.3. Ортогональные и ортонормированные системы функций	115
3.2.4. Обобщенные ряды Фурье	120
3.2.5. Минимальное свойство коэффициентов Фурье	122
3.2.6. Сходимость обобщенных рядов Фурье	125
3.2.7. Ортогональные (ортонормированные) системы дискретных функций	128
3.2.8. Ортогональные системы тригонометрических и полиномиальных дискретных функций	133
3.3. Предварительное (локальное) сглаживание временных рядов	136
3.4. Линейное прогнозирование структурно детерминированных рядов	140

3.5. Рекуррентное прогнозирование структурно детерминированных рядов	144
3.6. Анализ адекватности модели тренда временного ряда	152

Глава 4. Стохастические временные ряды

4.1. Случайные процессы (начальные определения и классификация)	158
4.2. Одномерные характеристики случайного процесса ..	160
4.3. Многомерные характеристики случайного процесса. Марковские процессы	164
4.4. Ковариационные и взаимные ковариационные функции случайных процессов. Белый шум	169
4.5. Стационарные и эргодические случайные процессы ..	173
4.6. Спектральная плотность случайного процесса	177
4.7. Преобразование случайного процесса линейным оператором	179
4.8. Преобразование случайного процесса оператором свертки	181
4.9. Формирующие фильтры	186
4.10. Типовые модели стохастических временных рядов эконометрики	193
4.11. Стохастический вектор состояния. Обобщенная матрично-векторная модель временного ряда	198
4.12. Рекуррентный алгоритм прогнозирования стохастических временных рядов (калмановский фильтр) ..	202
4.13. Нелинейная параметрическая идентификация модели стохастического временного ряда	209
4.14. Обобщенный рекуррентный алгоритм прогнозирования стохастических временных рядов	213
Приложение 1. Операционные методы исследования динамических систем	215
Приложение 2. Калмановское прогнозирование развития отдельных отраслей экономики России	226
Литература	233
Предметный указатель	235

Учебное издание

Чураков Евгений Павлович

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ В ЭКОНОМИКЕ

Заведующая редакцией Л. А. Табакова

Редактор В. В. Космин

Младший редактор Н. А. Федорова

Художественный редактор Ю. И. Артюхов

Технический редактор И. В. Завгородняя

Корректоры Т. К. Колпакова, Г. В. Хлопцева

Художественное оформление И. Н. Биксентеева

Компьютерная верстка И. А. Пиминовой

ИБ № 4642

Подписано в печать 15.10.2003 г.

Формат 60×88/16. Печать офсетная

Гарнитура «Таймс». Усл. п. л. 14,7. Уч.-изд. л. 12,84

Тираж 3000 экз. Заказ 865. «С» 238

Издательство «Финансы и статистика»

101000, Москва, ул. Покровка, 7

Телефон (095) 925-35-02, факс (095) 925-09-57

E-mail: mail@finstat.ru http://www.finstat.ru

ГУП «Великолукская городская типография»

Комитет по средствам массовой информации Псковской области,

182100, Великие Луки, ул. Полиграфистов, 78/12

Тел./факс: (811-53) 3-62-95

E-mail: VTL@MART.RU