

Т.А. ДУБРОВА

СТАТИСТИЧЕСКИЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ

*Рекомендовано Учебно-методическим объединением
по образованию в области «Статистики и антикризисного управления»
Министерства образования РФ в качестве учебного пособия
для студентов высших учебных заведений, обучающихся по специальности
061700 «Статистика» и другим экономическим специальностям*

*Рекомендовано Учебно-методическим центром
«Профессиональный учебник» в качестве учебного пособия
для студентов высших учебных заведений*



Москва • 2003

УДК 338.27:311(075.8)

ББК 65.23в647я73

Д79

Рецензенты:

кафедра экономической кибернетики

Московской сельскохозяйственной академии им. К.А. Тимирязева
(зав. кафедрой д-р экон. наук, проф., член-корр. РАСХН А.М. Гатаулин);

д-р экон. наук, проф. Г.Л. Громыко

(Московский государственный университет им. М.В. Ломоносова)

Главный редактор издательства

доктор экономических наук Н.Д. Эриашвили

Дуброва Т.А.

Д79. Статистические методы прогнозирования: Учеб. пособие для вузов. — М.: ЮНИТИ-ДАНА, 2003. — 206 с.

ISBN 5-238-00497-4

В систематизированном виде изложены статистические методы анализа и прогнозирования одномерных временных рядов. Подробно рассмотрены наиболее часто применяемые в экономической практике методы и особенности их реализации в современных пакетах прикладных программ. Видное место отводится процедурам сглаживания временных рядов, адаптивным методам прогнозирования, методологии построения моделей ARIMA. Особое внимание уделяется реализации рассмотренных классов моделей в среде SPSS.

Структура изложения соответствует логической последовательности основных этапов анализа и прогнозирования временных рядов. Приведены примеры решения практических задач; большое внимание уделяется анализу полученных результатов. В основе книги — курс, читаемый в МЭСИ.

Для студентов экономических специальностей вузов, а также аспирантов, преподавателей и специалистов, занимающихся применением статистических методов в финансовой и социально-экономической сферах.

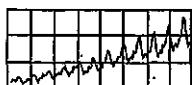
ББК 65.23в647я73

ISBN 5-238-00497-4

© Т.А. Дуброва, 2003

© ИЗДАТЕЛЬСТВО ЮНИТИ-ДАНА, 2003.

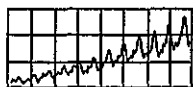
Воспроизведение всей книги или любой ее части запрещается без письменного разрешения издательства



ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ	5
ГЛАВА 1. ВРЕМЕННЫЕ РЯДЫ И ИХ ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ	8
1.1. Классификация рядов динамики и основные правила их построения	8
1.2. Важнейшие показатели изменения уровней рядов динамики	13
1.3. Компоненты временных рядов	24
1.4. Стационарные временные ряды и их основные характеристики	27
Практикум к главе 1	31
ГЛАВА 2. СГЛАЖИВАНИЕ ВРЕМЕННЫХ РЯДОВ С ПОМОЩЬЮ СКОЛЬЗЯЩИХ СРЕДНИХ	33
2.1. Применение простых скользящих средних	33
2.2. Использование взвешенных скользящих средних	37
2.3. Восстановление краевых значений	43
2.4. Эффект Слуцкого—Юла	45
2.5. Использование скользящих средних в техническом анализе	47
Практикум к главе 2	49
ГЛАВА 3. АНАЛИТИЧЕСКОЕ ВЫРАВНИВАНИЕ ВРЕМЕННЫХ РЯДОВ. МОДЕЛИРОВАНИЕ ТЕНДЕНЦИИ РАЗВИТИЯ	51
3.1. Проверка гипотезы о существовании тренда	51
3.2. Применение моделей кривых роста для анализа и прогнозирования основной тенденции развития	55
3.3. Методы выбора кривых роста	68
3.4. Оценка адекватности и точности выбранных моделей	72
Практикум к главе 3	82
ГЛАВА 4. СТАТИСТИЧЕСКИЙ АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ПЕРИОДИЧЕСКИХ КОЛЕБАНИЙ	84
4.1. Статистические методы оценки уровня сезонности. Прогнозирование с помощью тренд-сезонных моделей	84
4.2. Сезонная декомпозиция и корректировка временных рядов	92

4.3. Моделирование сезонных колебаний с помощью фиктивных переменных	97
4.4. Исследование периодических колебаний методами спектрального анализа	101
Практикум к главе 4	110
Глава 5. ИСПОЛЬЗОВАНИЕ АДАПТИВНЫХ МЕТОДОВ ПРОГНОЗИРОВАНИЯ В ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЯХ	113
5.1. Сущность адаптивных методов	113
5.2. Экспоненциальное сглаживание	116
5.3. Адаптивные полиномиальные модели	123
5.4. Адаптивные модели сезонных явлений	131
Практикум к главе 5	138
Глава 6. ПРОГНОЗИРОВАНИЕ С ПОМОЩЬЮ МОДЕЛИ АВТОРЕГРЕССИИ — ПРОИНТЕГРИРОВАННОГО СКОЛЬЗЯЩЕГО СРЕДНЕГО	141
6.1. Модели стационарных временных рядов	141
6.1.1. Авторегрессионные модели	141
6.1.2. Модели скользящего среднего	151
6.1.3. Авторегрессионные модели со скользящими средними в остатках	153
6.2. Методология Бокса—Дженкинса	156
Практикум к главе 6	163
Глава 7. АНАЛИЗ ВРЕМЕННЫХ РЯДОВ И ПРОГНОЗИРОВАНИЕ В СРЕДЕ SPSS	166
7.1. Особенности современного статистического программного обеспечения	166
7.2. Реализация моделей кривых роста в SPSS	167
7.3. SPSS: построение моделей, основанных на экспоненциальном сглаживании	171
7.4. Прогнозирование с помощью моделей авторегрессии — проинтегрированного скользящего среднего (ARIMA) в среде SPSS	178
7.5. Сезонная декомпозиция временных рядов с использованием SPSS	184
Практикум к главе 7	188
БИБЛИОГРАФИЧЕСКИЙ СПИСОК	190
ПРИЛОЖЕНИЯ	192



ПРЕДИСЛОВИЕ

В настоящее время статистические методы прогнозирования заняли видное место в экономической практике. С развитием компьютерной техники, совершенствованием информационных технологий, распространением пакетов прикладных программ (ППП) эти методы вышли за стены учебных и научно-исследовательских институтов. Они стали важным инструментом в деятельности плановых, аналитических, маркетинговых отделов производственных предприятий и объединений, торговых, страховых компаний, банков, правительственных учреждений. В условиях перехода экономики на рыночные отношения существенно меняются информационные запросы управляющих структур по объему, составу, достоверности и оперативности информации. В связи с этим для руководителей различных уровней возрастает роль прогнозов в принятии обоснованных управленческих решений.

Стремительное распространение пакетов прикладных программ позволило сделать доступными и наглядными современные методы и подходы статистического прогнозирования. При этом применение эконометрического программного обеспечения позволяет создать для пользователя уникальную среду, в которой статистическая обработка данных становится увлекательным исследованием, позволяющим получать многовариантные решения.

Процесс анализа и прогнозирования временных рядов с помощью таких статистических систем, как правило, включает следующие этапы:

- ввод данных в систему;
- визуализацию данных с помощью различных типов графиков;
- преобразование данных, адекватное выбранным статистическим методам;
- реализацию алгоритмов статистических методов;
- вывод результатов анализа в виде графиков и таблиц с числовой и текстовой информацией;
- интерпретацию полученных результатов.

Пользователь освобождается от всей черновой работы (проведение трудоемких расчетов, построение таблиц и графиков), на его долю остается исследовательская, творческая работа: постановка задачи, выбор методов прогнозирования, оценка качества полученных моделей, интерпретация результатов. Для этого необходимо иметь определенную подготовку в области прикладной

статистики, знать методы и подходы статистического анализа и прогнозирования временных рядов.

Статистические методы прогнозирования стали неотъемлемой частью блока эконометрических дисциплин, изучаемых в экономических вузах. Однако в настоящее время ощущается нехватка доступных учебников и учебных пособий, посвященных системному, комплексному рассмотрению методологии применения статистических методов прогнозирования в экономике. Ощущается острая потребность в издании учебников, как ориентированных на начальный, базовый уровень, так и на углубленное изучение этой тематики. Тем более что в последнее время появились новые направления в анализе одномерных и многомерных временных рядов, например, авторегрессионные условно гетероскедастичные модели (ARCH- и GARCH- модели), ставшие популярными при описании финансовых рынков, модели, связанные с коинтеграцией временных рядов, модели векторной авторегрессии (VAR) и др.

Сейчас наблюдается устойчивый интерес студентов, аспирантов, преподавателей экономических вузов, специалистов в области прикладной статистики к такой литературе. В данном учебном пособии предпринята попытка частично восполнить имеющийся пробел в этой области. Здесь подробно рассмотрены наиболее часто применяемые в экономической практике методы. Структура изложения соответствует логической последовательности основных этапов анализа и прогнозирования временных рядов, начинающейся с первичной обработки исходных данных и заканчивающейся анализом и интерпретацией полученных результатов.

В МЭСИ курс «Статистические методы прогнозирования» включен в учебные планы всех экономических специальностей. Опыт преподавания убедительно показывает необходимость использования в рамках этого курса современных статистических пакетов программ. В учебном процессе в МЭСИ на протяжении ряда лет успешно использовались отечественные пакеты прикладных программ «Олимп» (МЭСИ) и «Мезозавр» (ЦЭМИ РАН), а также англоязычные статистические системы STATISTICA, SPSS, имеется некоторый опыт применения SAS. Эти пакеты позволяют проводить самостоятельные исследования с реальными данными, стимулирующие интерес студентов к теоретическому материалу. Постепенно у студентов формируются устойчивые навыки работы со статистическими пакетами, расширяется кругозор в области методов прикладной статистики и их программного обеспечения. Это позволяет им в дальнейшем самостоятельно осуществлять обоснованный выбор математического обеспечения для решения

конкретных социально-экономических задач, исходя из собственного опыта и содержания задачи.

В этом учебном пособии обсуждение статистических методов прогнозирования привязано к их реализации в пакетах программ, в ходе изложения материала обсуждаются особенности тех или иных процедур в различных программных продуктах.

Книга состоит из основного учебного материала (гл. 1—7), содержащего изложение статистических методов анализа и прогнозирования одномерных временных рядов, и Приложений. Видное место отводится процедурам сглаживания временных рядов, адаптивным методам прогнозирования, методологии построения моделей ARIMA. Материал каждой из гл. 1—6 завершается практикумом, включающим контрольные вопросы и упражнения. Выполнение упражнений, содержащих вычисления, предполагает использование компьютера. В частности, для уяснения алгоритмов соответствующих методов удобно при решении задач использовать Excel.

Глава 7 посвящена вопросам проведения анализа и прогнозирования временных рядов в среде SPSS. Чем объясняется выбор именно этой статистической системы?

На протяжении ряда лет на рынке труда наблюдается устойчивый спрос на специалистов, свободно владеющих этим пакетом. МЭСИ — официально зарегистрированный пользователь SPSS. Кроме того, все рассмотренные в этом пособии классы моделей реализованы в SPSS. Знание этой статистической системы облегчит пользователю работу и с другими пакетами обработки данных, а также позволит приобрести навыки работы с английской статистической лексикой.

В заключение отметим, что при решении практических задач не должно быть чисто формальных механических подходов к выбору методов и моделей прогнозирования. Успех их применения зависит от глубины содержательного экономического анализа, от сочетания знаний в области прикладной статистики с профессиональным знанием моделируемого процесса.

Автор выражает глубокую благодарность д.э.н., проф. А.М. Гаутаину и д.э.н., проф. Г.Л. Громыко за работу по рецензированию рукописи и сделанные ими ценные замечания.



Глава

ВРЕМЕННЫЕ РЯДЫ И ИХ ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ

1.1. Классификация рядов динамики и основные правила их построения

Среди основных задач статистики видное место занимает описание изменений показателей во времени, изучение динамики развития социально-экономических процессов.

Как изменяется уровень оплаты труда? Каковы колебания курса доллара? Какая тенденция прослеживается в изменении важнейших макроэкономических показателей? Ответы на эти и аналогичные вопросы могут быть получены с помощью специальных статистических методов, анализирующих ряды динамики.

Рядом динамики (динамическим рядом, временным рядом) называется последовательность значений статистического показателя (признака), упорядоченная в хронологическом порядке, т.е. в порядке возрастания временного параметра. Отдельные наблюдения временного ряда называются *уровнями* этого ряда.

В англоязычной литературе для временных рядов используется термин «*time series*».

Каждый ряд динамики содержит два элемента:

- значения времени;
- соответствующие им значения уровней ряда.

В качестве показателя времени в рядах динамики могут указываться либо определенные моменты (даты) времени, либо отдельные периоды (сутки, месяцы, кварталы, полугодия, годы и т.д.). В зависимости от характера временного параметра ряды делятся на *моментные* и *интервальные*.

В *моментных* рядах динамики уровни характеризуют значения показателя по состоянию на определенные моменты времени. Например, моментными являются временные ряды цен на определенные виды товаров, ряды курсов акций, уровни которых фиксируются для конкретных чисел. Примерами моментных рядов динамики могут служить также ряды численности населения или стоимости основных фондов, так как значения уровней этих рядов определяются ежегодно на одну и ту же дату.

В *интервальных* рядах уровни характеризуют значение показателя за определенные интервалы (периоды) времени. Примерами могут

служить ряды годовой (месячной, квартальной) динамики производства продукции в натуральном или стоимостном выражении.

В табл. 1.1, 1.2 приведены моментные ряды динамики, а в табл. 1.3, 1.4 — интервальные, причем у последнего ряда динамики, в отличие от других примеров, уровни не равноотстоящие во времени.

Таблица 1.1. Цены акций промышленной компании

<i>Дата</i>	<i>06.09.99</i>	<i>07.09.99</i>	<i>08.09.99</i>	<i>09.09.99</i>	<i>10.09.99</i>	<i>13.09.99</i>
Цена акций промышленной компании на момент открытия торгов, долл. США	383	392	391	395	393	393

Таблица 1.2. Остатки вкладов населения в банках на начало месяца (1995 г.)

<i>Месяц</i>	<i>Январь</i>	<i>Февраль</i>	<i>Март</i>	<i>Апрель</i>	<i>Май</i>	<i>Июнь</i>
Остаток вкладов населения в банках на начало каждого месяца, млрд руб.	28317	30624	33408	36408	40524	45416

Таблица 1.3. Фонд заработной платы в первом полугодии 2000 г.

<i>Месяц</i>	<i>Январь</i>	<i>Февраль</i>	<i>Март</i>	<i>Апрель</i>	<i>Май</i>	<i>Июнь</i>
Фонд заработной платы работников предприятия, тыс. руб.	79,5	84,1	85,5	88,5	89,9	90,0

Таблица 1.4. Объем экспорта продукции предприятия

<i>Год</i>	<i>1990</i>	<i>1995</i>	<i>1997</i>	<i>1998</i>	<i>1999</i>
Объем экспорта продукции предприятия, тыс. долл. США	1200	1350	1400	1370	1325

Уровни рядов динамики могут представлять собой абсолютные, относительные и средние значения. Если уровни ряда представляют собой непосредственно не наблюдаемые значения, а производные — средние или относительные, то такие ряды называ-

ются производными. Уровни этих рядов получаются с помощью некоторых вычислений на основе абсолютных показателей. Примером производного ряда динамики может служить ряд среднесуточного производства промышленной продукции (табл. 1.5).

Таблица 1.5. Среднесуточное производство продукции

Месяц	Производство продукции, тыс. шт.	Число рабочих дней в месяце	Среднесуточное производство, тыс. шт.
1	2	3	4*
Январь	195,0	25	7,8
Февраль	204,0	24	8,5
Март	210,6	26	8,1
Апрель	195,0	26	7,5
Май	207,5	25	8,3
Июнь	205,4	26	7,9

* Данные гр. 4 получаются с помощью деления данных гр. 2 на данные гр. 3.

Важная особенность интервальных рядов динамики абсолютных величин заключается в возможности суммирования их уровней. В результате этой процедуры получают накопленные итоги, имеющие осмысленное содержание благодаря отсутствию повторного счета. Например, суммируя фонд заработной платы работников предприятия за три первых и три последующих месяца (см. табл. 1.3), получаем соответственно фонд заработной платы за первый и второй кварталы, а сумма этих квартальных данных дает фонд заработной платы за полугодие.

Суммирование уровней моментного ряда динамики не практикуется, так как полученные накопленные итоги лишены всякого смысла. Например, уровни моментного ряда «остатки вкладов населения в банках» (см. табл. 1.2) содержат элементы повторного счета. Второй уровень частично содержит вклады населения, учтенные первым уровнем, и т.д. Таким образом, моментные ряды динамики в отличие от интервальных не обладают свойством аддитивности (от англ. *to add* — добавлять).

При исследовании моментного ряда динамики определенный смысл имеет расчет разностей уровней, характеризующих изменение показателя за некоторый отрезок времени. Например, за январь 1995 г. остатки вкладов населения в банках увеличились на 2307 млрд руб.

На практике часто требуется проанализировать динамику показателя не только за данный отрезок времени, но и с учетом ряда

предшествующих периодов. Для этого строится ряд динамики с нарастающими итогами, уровни которого дают обобщающий результат развития показателя с начала отчетного периода (квартала, полугодия, года и т.д.). В качестве примера рассмотрим данные о производстве холодильников и морозильников на предприятии в первом полугодии 2000 г. (табл. 1.6). Данные гр. 3 получены последовательным суммированием смежных уровней.

Таблица 1.6. Производство холодильников и морозильников (2000 г.)

Месяц	Произведено холодильников и морозильников, шт.	
	за месяц	с начала года
Январь	146	146
Февраль	181	$146 + 181 = 327$
Март	174	$327 + 174 = 501$
Апрель	152	$501 + 152 = 653$
Май	160	$653 + 160 = 813$
Июнь	179	$813 + 179 = 992$

Уровни ряда могут принимать детерминированные или случайные значения. Примером ряда с детерминированными значениями уровней служит ряд последовательных данных о количестве дней в месяцах. Естественно, анализу, а в дальнейшем и прогнозированию, подвергаются ряды со случайными значениями уровней.

Временные ряды имеют характерные отличия от пространственных выборок.

Во-первых, в отличие от пространственных данных уровни временного ряда, как правило, не являются статистически независимыми.

Во-вторых, члены временного ряда не являются одинаково распределенными.

Очевидно, что эти особенности должны быть учтены в исследовательской работе.

Успешность статистического анализа развития процессов во времени во многом зависит от правильного построения рядов динамики. Так, большое значение имеет выбор интервалов между соседними уровнями ряда. Удобнее всего иметь дело с равноотстоящими друг от друга уровнями ряда. При этом нежелательно брать слишком большой интервал времени, так как можно упустить существенные закономерности в динамике показателя. Например, по квартальным данным невозможно судить о месячных

сезонных колебаниях, но информация может оказаться и, слишком «короткой» для использования некоторых методов анализа и прогнозирования динамики, предъявляющих «жесткие» требования к длине рядов. К тому же слишком малые интервалы между наблюдениями увеличивают объем вычислений, а также могут приводить к появлению ненужных деталей в динамике процесса, засоряющих общую тенденцию. Разумеется, вопрос о выборе интервала времени между уровнями ряда должен решаться исходя из целей каждого конкретного исследования.

Одно из важнейших условий, необходимых для правильного отражения временным рядом реального процесса развития, — *сопоставимость уровней ряда*. Для несопоставимых величин неправомерно проводить исследование динамики.

Между тем, появление несопоставимых уровней может быть вызвано разными причинами: изменением методики расчета показателя, изменением классификации, терминологии и т.д. Например, уровни временного ряда, характеризующие число малых предприятий, могут оказаться несопоставимыми из-за изменения самого понятия «малое предприятие». Подразумевается, что это понятие должно быть одинаковым для всего исследуемого периода.

Чаще всего несопоставимость встречается в стоимостных показателях, что вызвано изменением цен в разные периоды времени, поэтому на практике осуществляют пересчет уровней в сопоставимые цены (цены одного периода). Несопоставимость может возникнуть и вследствие территориальных изменений, например как результат изменения границ области, района, страны. При этом следует иметь в виду, что вопрос о сопоставимости будет зависеть от целей исследования. Например, при описании военной, экономической мощи страны следует учитывать данные в изменяющихся границах территории, а при сопоставлении темпов развития промышленности следует сравнение проводить в одних и тех же территориальных границах.

Несопоставимость может возникнуть и в результате структурных изменений. Например, произошло укрупнение нескольких ведомств путем слияния их в единое целое или укрупнение производства в результате слияния нескольких предприятий в одно объединение.

В большинстве случаев удастся устранить несопоставимость, вызванную указанными причинами, путем пересчета более ранних значений показателей. Хотя иногда проведение такой обработки не обеспечивает требуемой точности, приводит к снижению ценности исходной информации, а следовательно, и к затруднению дальнейшего анализа.

Для успешного изучения динамики процесса важно, чтобы информация была полной, чтобы временной ряд имел достаточную длину (с учетом конкретных целей исследования). Например, при изучении периодических колебаний желательно иметь информацию не менее чем за три полных периода колебания. Поэтому при анализе сезонных колебаний на базе рядов месячной или квартальной динамики желательно иметь информацию, как правило, не менее чем за 3 года.

Применение определенного математического аппарата также накладывает ограничение на допустимую длину временных рядов. Например, использование более тонкого статистического аппарата для исследования периодичности (гл. 4—6) требует большей длины информации — не менее 5—6 полных периодов колебания. При построении регрессионных моделей требуется иметь временные ряды, длина которых в несколько раз превосходит число независимых переменных.

Уровни временных рядов могут содержать аномальные значения или «выбросы». Часто появление таких значений может быть вызвано ошибками при сборе, записи и передаче информации. Возможными источниками появления ошибочных значений могут стать: сдвиг запятой при перенесении информации из документа, занесение данных в другую графу и т.д.

Выявление, исключение таких значений, замена их истинными или расчетными — необходимый этап первичной обработки данных, так как применение математических методов к «засоренной» информации приводит к искажению результатов анализа. Однако аномальные значения могут отражать и реальное развитие процесса, например скачок курса доллара в «черный вторник». Как правило, эти значения также заменяются расчетными при построении моделей, но учитываются при расчете возможных отклонений фактических значений от полученных по модели.

Соответствие исходной информации всем указанным требованиям проверяется на этапе предварительного анализа временных рядов. Лишь после этого переходят к расчету и анализу основных показателей динамики развития, построению моделей прогнозирования, получению прогнозных оценок.

1.2. Важнейшие показатели изменения уровней рядов динамики

При анализе изменений явления во времени на практике часто определяют средние показатели, в том числе *средний уровень ряда*. Средний уровень — это важная обобщающая характеристика для рядов динамики, изменение которых стабилизировалось в ис-

следуемом периоде и при этом подвержено ощутимым случайным колебаниям. Например, средний уровень урожайности за ряд лет лучше опишет урожайность, чем уровень одного года, значение которого формируется под действием множества случайных факторов.

В то же время, если в исследуемом периоде выделяются неоднородные этапы, в течение которых условия развития существенно менялись, то нецелесообразно рассчитывать общую среднюю, следует построить анализ динамики по отдельным этапам.

Средний уровень ряда определяется по-разному для моментных и интервальных рядов, при этом следует обратить внимание на то, какие — равноотстоящие или не равноотстоящие во времени — уровни наблюдаются в ряду динамики. Для *интервальных рядов динамики с равноотстоящими во времени уровнями* расчет среднего уровня проводится по формуле простой средней арифметической:

$$\bar{y} = \frac{\sum_{t=1}^n y_t}{n}, \quad (1.1)$$

где n — число уровней или длина ряда;

y_t — уровень ряда динамики ($t = 1, 2, \dots, n$).

В качестве примера рассмотрим определение среднего уровня для интервального ряда динамики, представленного в табл. 1.3. Среднемесячный фонд заработной платы в первом полугодии составил:

$$\bar{y} = \frac{75,5 + 84,1 + 85,5 + 88,5 + 89,9 + 90,0}{6} = 86,25 \text{ тыс. руб.}$$

В случае *интервальных рядов динамики с не равноотстоящими во времени уровнями* для расчета среднего уровня используется формула взвешенной средней арифметической, где в качестве весовых коэффициентов используется продолжительность интервалов времени между уровнями (число периодов времени, при которых значение уровня не изменяется).

Для *моментных рядов динамики с равноотстоящими во времени уровнями* средний уровень (так называемая средняя хронологическая) находится по формуле:

$$\bar{y} = \frac{\frac{y_1 + y_n}{2} + \sum_{t=2}^{n-1} y_t}{n-1}, \quad (1.2)$$

где $y_1, y_2, \dots, y_n$ — уровни ряда динамики;

y_1 и y_n — соответственно начальный и конечный уровни ряда;

n — число уровней или длина ряда.

Проиллюстрируем использование формулы на примере расчета средней за неделю цены акций (см. табл. 1.1):

$$\bar{y} = \frac{\frac{383+393}{2} + 392 + 391 + 395 + 393}{5} = 391,8 \text{ долл.}$$

В случае *моментных рядов динамики с не равноотстоящими во времени уровнями* средний уровень определяется по формуле средней хронологической взвешенной:

$$\bar{y} = \frac{(y_1 + y_2)t_1 + (y_2 + y_3)t_2 + \dots + (y_{n-1} + y_n)t_{n-1}}{2 \sum_{i=1}^{n-1} t_i}, \quad (1.3)$$

где $y_1, y_2, \dots, y_n$ — уровни ряда динамики;

t_i — продолжительность интервала времени между соседними уровнями.

Рассмотрим применение формулы (1.3) на следующем примере.

Требуется определить средний уровень для моментного ряда динамики, характеризующего численность официально зарегистрированных безработных в городе (табл. 1.7).

Таблица 1.7. Численность официально зарегистрированных безработных на начало месяца

Дата	01.01.99	01.03.99	01.06.99	01.09.99	01.01.00
Численность безработных	1800	1950	1899	1600	1850

Ряд динамики имеет не равноотстоящие во времени уровни, поэтому применим формулу (1.3).

Средний уровень ряда равен:

$$\bar{y} = \frac{(1800+1950) \cdot 2 + (1950+1899) \cdot 3 + (1899+1600) \cdot 3 + (1600+1850) \cdot 4}{2(2+3+3+4)} = 1806 \text{ чел.}$$

На практике для количественной оценки динамики явлений широко применяются следующие основные аналитические показатели:

- абсолютные приросты;
- темпы роста;
- темпы прироста.

Причем каждый из указанных показателей может быть трех видов:

- цепной;
- базисный;
- средний.

В основе расчета этих показателей динамики лежит сравнение уровней временного ряда. Если сравнение осуществляется с одним и тем же уровнем, принятым за базу сравнения, то эти показатели называются *базисными*. В качестве базы сравнения выбирается либо начальный уровень динамического ряда, либо уровень, с которого начинается новый этап развития. Например, при анализе динамики производства основных видов промышленной продукции в натуральном выражении часто за базу сравнения выбирают 1990 г. Это объясняется тем, что до этого года во многих отраслях российской промышленности наблюдался замедлявшийся подъем, перешедший затем в спад производства. Поэтому наметившийся в последнее время рост производства желательнее оценить не только по отношению к предыдущему году, но и в сравнении с 1990 г.

Если сравнение осуществляется при переменной базе и каждый последующий уровень сравнивается с предыдущим, то вычисленные таким образом показатели называются *цепными*.

Абсолютный прирост равен разности двух сравниваемых уровней и характеризует изменение показателя за определенный промежуток времени. В общем случае абсолютный прирост может быть представлен в виде:

$$\Delta y_t = y_t - y_{t-k} \quad (1.4)$$

где y_t — текущий уровень ряда динамики;
 $t = 2, 3, \dots, n$; $k = 1, 2, \dots, n - 1$.

При $k = 1$ из текущего уровня y_t вычитается предыдущий уровень y_{t-1} и получается формула для расчета цепного абсолютного прироста:

$$\Delta y_t = y_t - y_{t-1} \quad (1.5)$$

При $k = t - 1$ из формулы (1.4) вытекает выражение для базисного абсолютного прироста, определяемого относительно начального уровня ряда:

$$\Delta y_t^b = y_t - y_1 \quad (1.6)$$

Для записи формулы базисного абсолютного прироста в более общем виде уровень y_1 в (1.6) может быть заменен на уровень временного ряда, принятый за базу сравнения, — y_0 :

$$\Delta y_t^b = y_t - y_0 \quad (1.7)$$

Средний абсолютный прирост — это обобщающая характеристика скорости изменения исследуемого показателя во времени (скоростью будем называть прирост в единицу времени). Для его определения за весь период наблюдения используется формула простой средней арифметической:

$$\Delta \bar{y} = \frac{\sum_{t=2}^n \Delta y_t}{n-1}, \quad (1.8)$$

где Δy_t — цепной абсолютный прирост;
 n — длина временного ряда.

Подставив в числитель выражения для цепных абсолютных приростов, получим более удобную форму записи для среднего абсолютного прироста:

$$\Delta \bar{y} = \frac{y_2 - y_1 + y_3 - y_2 + \dots + y_n - y_{n-1}}{n-1} = \frac{y_n - y_1}{n-1}, \quad (1.9)$$

где y_n и y_1 — соответственно конечный и начальный уровни ряда динамики.

Темп роста T характеризует отношение двух сравниваемых уровней ряда, как правило, выраженное в процентах.

Темп роста может быть представлен в виде:

$$T_t = \frac{y_t}{y_{t-k}} 100\%, \quad (1.10)$$

где y_t — текущий уровень ряда динамики;
 $t = 2, 3, \dots, n; k = 1, 2, \dots, n-1$.

Отметим, что индекс уровня y_{t-k} , находящегося в знаменателе, определяется так же, как и в случае абсолютного прироста. Следовательно, из выражения (1.10) в зависимости от значений индекса k получаются формулы для расчета цепных и базисных темпов роста.

Цепной темп роста равен:

$$T_t = \frac{y_t}{y_{t-1}} 100\%. \quad (1.11)$$

Базисный темп роста может быть представлен в виде:

$$T_t = \frac{y_t}{y_0} 100\%, \quad (1.12)$$

где y_0 — уровень временного ряда, принятый за базу сравнения.

Темп роста всегда положителен. Если темп роста равен 100%, то значение уровня не изменилось, если меньше 100%, то значение уровня понизилось, больше 100% — повысилось.

Средний темп роста — обобщающая характеристика динамики, отражающая интенсивность изменения уровней ряда. Он показывает, сколько в среднем процентов последующий уровень составляет от предыдущего на всем периоде наблюдения. Этот показатель рассчитывается по формуле средней геометрической из цепных темпов роста:

$$\bar{T} = \sqrt[n]{T_2 T_3 \dots T_n}. \quad (1.13)$$

Выразив цепные темпы роста $T_2, T_3, \dots, T_n$ через соответствующие уровни ряда, получим

$$\bar{T} = \sqrt[n]{\frac{y_2}{y_1} \cdot \frac{y_3}{y_2} \cdot \frac{y_4}{y_3} \cdot \dots \cdot \frac{y_n}{y_{n-1}}} 100\% = \sqrt[n]{\frac{y_n}{y_1}} 100\%. \quad (1.14)$$

Темп прироста K характеризует абсолютный прирост в относительных величинах. Определенный в процентах темп прироста показывает, на сколько процентов изменился сравниваемый уровень по отношению к уровню, принятому за базу сравнения. Темп прироста есть выраженное в процентах отношение абсолютного прироста к уровню, принятому за базу сравнения:

$$K_t = \frac{y_t - y_{t-k}}{y_{t-k}} 100\%, \quad (1.15)$$

где y_t — текущий уровень ряда динамики;
 $t = 2, 3, \dots, n$; $k = 1, 2, \dots, n - 1$.

Очевидно, что темп прироста может быть положительным, отрицательным или равным нулю.

При $k = 1$ получаем цепной темп прироста:

$$K_t = \frac{y_t - y_{t-1}}{y_{t-1}} 100\%. \quad (1.16)$$

Преобразовав выражение (1.16), можно показать зависимость цепного темпа прироста от соответствующего темпа роста:

$$K_t = \frac{y_t}{y_{t-1}} 100\% - 100\% = T_t - 100\%, \quad (1.17)$$

где T_t — цепной темп роста.

Базисный темп прироста равен отношению базисного абсолютного прироста к уровню ряда, принятому за базу сравнения:

$$K_t^6 = \frac{\Delta y_t^6}{y_6}. \quad (1.18)$$

По аналогии с (1.17) получаем

$$K_t^6 = T_t^6 - 100\%, \quad (1.19)$$

где T_t^6 — базисный темп роста.

Соответственно *средний темп прироста* может быть выражен через средний темп роста:

$$\bar{K} = \bar{T} - 100\%. \quad (1.20)$$

Сравнение абсолютного прироста и темпа прироста за одни и те же периоды времени показывает, что в реальных экономических процессах замедление темпов прироста не всегда сопровождается уменьшением абсолютных приростов. Поэтому на практике часто проводят сопоставление этих показателей. Для этого рассчитывают *абсолютное значение одного процента прироста*, определяемое как отношение абсолютного прироста к соответствующему темпу прироста:

$$|\%| = \frac{y_t - y_{t-1}}{\frac{y_t - y_{t-1}}{y_{t-1}} 100} = 0,01 y_{t-1}. \quad (1.21)$$

В табл. 1.8 приведены выражения для вычисления рассмотренных аналитических показателей динамики.

Таблица 1.8. Основные показатели динамики

Вид показателя	Абсолютный прирост	Темп роста, %	Темп прироста, %
Цепной	$\Delta y_t = y_t - y_{t-1}$	$T_t = \frac{y_t}{y_{t-1}} 100$	$K_t = T_t - 100$
Базисный	$\Delta y_t^6 = y_t - y_6$	$T_t^6 = \frac{y_t}{y_6} 100$	$K_t^6 = T_t^6 - 100$
Средний	$\Delta \bar{y} = \frac{y_n - y_1}{n - 1}$	$\bar{T} = \sqrt[n-1]{\frac{y_n}{y_1}} 100$	$\bar{K} = \bar{T} - 100$

Проиллюстрируем расчет и анализ статистических показателей динамики на следующем примере.

Пример 1.1

По данным о вводе в действие жилых домов (табл. 1.9) рассчитайте цепные, базисные и средние:

а) абсолютные приросты;

б) темпы роста;

в) темпы прироста.

В качестве базисного уровня возьмите начальный уровень ряда.

Таблица 1.9. Ввод в действие жилых домов, млн м²

Текущий номер года <i>t</i>	1	2	3	4	5
Общая площадь, млн м ²	7,0	6,5	5,9	5,5	4,9

Решение

Представим расчет цепных и базисных абсолютных приростов, темпов роста, темпов прироста в табл. 1.10.

Для получения обобщающих показателей динамики развития определим средние характеристики: средний абсолютный прирост, средний темп роста и прироста.

Средний абсолютный прирост равен:

$$\Delta \bar{y} = \frac{y_n - y_1}{n - 1} = \frac{y_5 - y_1}{4} = \frac{4,9 - 7,0}{4} = -0,525 \text{ млн м}^2,$$

т.е. в среднем ежегодно общая площадь вводимого жилья уменьшалась на 0,525 млн м². Определим средний темп роста:

$$\bar{T} = n \sqrt[n]{\frac{y_n}{y_1}} 100\% = 4 \sqrt[4]{\frac{4,9}{7,0}} 100\% = 91,47\%.$$

т.е. в среднем ежегодно строительство жилья составляло 91,47% уровня предыдущего года.

Средний темп прироста $\bar{K} = \bar{T} - 100\% = -8,53\%$, т.е. в среднем ежегодно строительство жилья снижалось на 8,53%.

Наибольший интерес для статистического анализа представляют средний абсолютный прирост, средний темп роста, средний темп прироста, так как эти показатели — обобщающие характеристики динамики. С их помощью можно строить прогнозы исследуемых показателей. Однако необходимо отметить, что их применение требует определенной осторожности.

Описание динамики ряда с помощью среднего абсолютного прироста соответствует его представлению в виде прямой, проведенной через две крайние точки. В этом случае, чтобы получить прогноз на *L* шагов вперед (*L* — период упреждения), достаточно воспользоваться следующей формулой:

Таблица 1.10. Статистические показатели динамики

t	y, млн м ²	Абсолютный прирост, млн м ²		Темп роста, %		Темп прироста, %	
		Цепной	Базисный	Цепной	Базисный	Цепной	Базисный
1	7,0	—	—	—	—	—	—
2	6,5	$6,5 - 7,0 = -0,5$	$6,5 - 7,0 = -0,5$	$\frac{6,5}{7,0}100 = 92,86$	$\frac{6,5}{7,0}100 = 92,86$	$92,86 - 100 = -7,14$	$92,86 - 100 = -7,14$
3	5,9	$5,9 - 6,5 = -0,6$	$5,9 - 7,0 = -1,1$	$\frac{5,9}{6,5}100 = 90,77$	$\frac{5,9}{7,0}100 = 84,29$	$90,77 - 100 = -9,23$	$84,29 - 100 = -15,71$
4	5,5	$5,5 - 5,9 = -0,4$	$5,5 - 7,0 = -1,5$	$\frac{5,5}{5,9}100 = 93,22$	$\frac{5,5}{7,0}100 = 78,57$	$93,22 - 100 = -6,78$	$78,57 - 100 = -21,43$
5	4,9	$4,9 - 5,5 = -0,6$	$4,9 - 7,0 = -2,1$	$\frac{4,9}{5,5}100 = 89,09$	$\frac{4,9}{7,0}100 = 70,00$	$89,09 - 100 = -10,91$	$70,00 - 100 = -30,00$

$$\hat{y}_{n+L} = y_n + L\bar{\Delta y}, \quad (1.22)$$

где y_n — фактическое значение в последней n -й точке ряда (конечный уровень ряда);

$\hat{y}_{n+L}$ — прогнозное значение $(n + L)$ -го уровня ряда;

$\bar{\Delta y}$ — значение среднего абсолютного прироста, рассчитанное для временного ряда $y_1, y_2, \dots, y_n$.

Очевидно, что такой подход к получению прогнозного значения корректен, если характер развития близок к линейному. На такой равномерный характер развития могут указывать примерно одинаковые значения цепных абсолютных приростов. Продемонстрируем использование рассмотренного приема на следующем примере.

Пример 1.2

Данные таблицы описывают изменение процентной ставки банка в течение семи кварталов.

Текущий номер квартала t	1	2	3	4	5	6	7
Процентная ставка банка y_t , %	17,0	16,5	15,9	15,5	14,9	14,5	13,8

Требуется:

1) обосновать правомерность использования среднего абсолютного прироста для получения прогнозного значения процентной ставки в восьмом квартале;

2) рассчитать прогнозное значение процентной ставки банка в восьмом квартале с помощью среднего абсолютного прироста.

Решение

1. Рассчитаем цепные абсолютные приросты:

$$\Delta y_2 = 16,5 - 17,0 = -0,5\%; \quad \Delta y_3 = 15,9 - 16,5 = -0,6\%;$$

$$\Delta y_4 = 15,5 - 15,9 = -0,4\%; \quad \Delta y_5 = 14,9 - 15,5 = -0,6\%;$$

$$\Delta y_6 = 14,5 - 14,9 = -0,4\%; \quad \Delta y_7 = 13,8 - 14,5 = -0,7\%.$$

Легко заметить, что цепные абсолютные приросты примерно одинаковы. Они незначительно варьируют от $-0,7$ до $-0,4$, что свидетельствует о близости процесса развития к линейному. Поэтому для определения прогнозного значения показателя в восьмом квартале ($\hat{y}_8$) используем средний абсолютный прирост.

2. Найдем значение среднего абсолютного прироста, воспользовавшись формулой (1.9):

$$\Delta \bar{y} = \frac{y_7 - y_1}{6} = \frac{13,8 - 17,0}{6} = -0,5\%.$$

Рассчитаем прогнозное значение процентной ставки банка с помощью выражения (1.22):

$$\hat{y}_8 = y_7 + \Delta \bar{y} = 13,8 - 0,5 = 13,3\%.$$

Применение среднего темпа роста (и среднего темпа прироста) для описания динамики ряда соответствует его представлению в виде показательной или экспоненциальной кривой, проведенной через две крайние точки. Поэтому использование этого показателя в качестве обобщающего целесообразно для тех процессов, изменение динамики которых происходит с примерно постоянным темпом роста. В этом случае прогнозное значение на L шагов вперед может быть получено по формуле:

$$\hat{y}_{n+L} = y_n \bar{T}^L, \quad (1.23)$$

где $\hat{y}_{n+L}$ — прогнозная оценка значения $(n + L)$ -го уровня ряда;

y_n — фактическое значение в последней n -й точке ряда (конечный уровень ряда);

$\bar{T}$ — средний темп роста, рассчитанный для ряда $y_1, y_2, \dots, y_n$ (не в %-ном выражении).

Данный подход иллюстрирует следующий пример.

Пример 1.3

Изменение ежеквартальной динамики процентной ставки банка происходило с примерно постоянным темпом роста в течение семи кварталов. Процентная ставка банка в первом квартале равнялась 8,3%, а в седьмом квартале — 14%. Рассчитайте прогнозное значение процентной ставки банка в восьмом квартале, используя средний темп роста.

Решение

Известно, что изменение процентной ставки банка происходило с примерно постоянным темпом роста в течение семи кварталов. Следовательно, правомерно использовать средний темп роста для расчета прогнозного значения этого показателя.

Средний темп роста, согласно (1.14), равен:

$$\bar{T} = \sqrt[n-1]{\frac{y_n}{y_1}} 100\% = \sqrt[6]{\frac{14}{8,3}} 100\%; \quad \bar{T} = 109,1\%.$$

Рассчитаем прогнозное значение процентной ставки банка в восьмом квартале, в соответствии с (1.23):

$$\hat{y}_8 = \bar{y}_7 \bar{T},$$

где $\bar{T}$ — не в процентном выражении;

$$\hat{y}_8 = 14 \cdot 1,091 = 15,3\%.$$

К недостаткам среднего прироста и среднего темпа роста следует отнести то, что они учитывают лишь конечный и начальный уровни ряда, исключая влияния промежуточных уровней. Тем не менее, эти показатели имеют весьма широкую область применения, что объясняется чрезвычайной простотой их вычисления. Они могут быть использованы как приближенные, простейшие способы прогнозирования, предшествующие более глубокому количественному и качественному анализу.

1.3. Компоненты временных рядов

В практике исследования динамики явлений и прогнозирования принято считать, что значения уровней временных рядов могут содержать следующие компоненты (составные части или структурообразующие элементы):

- тренд;
- сезонную компоненту;
- циклическую компоненту;
- случайную компоненту.

Под трендом понимают изменение, определяющее общее направление развития, основную тенденцию временного ряда. Это систематическая составляющая долговременного действия.

Наряду с долговременными тенденциями во временных рядах часто возникают более или менее регулярные колебания — периодические составляющие рядов динамики.

Если период колебаний не превышает одного года, то их называют *сезонными*. Чаще всего причиной их возникновения считаются природно-климатические условия. Примером могут служить колебания цен на сельскохозяйственную продукцию, в частности на картофель. Из года в год наблюдается снижение цен в период после уборки урожая и последующее повышение цен, связанное с необходимостью хранения продукции. Своего «пика» цены достигают перед следующим урожаем. Таким образом, в колебаниях цен прослеживается устойчивая годовая периодичность.

Иногда причины сезонных колебаний имеют социальный характер, например, увеличение закупок в предпраздничный период, увеличение платежей в конце квартала и т.д.

При большом периоде колебания считают, что во временных рядах имеет место *циклическая* составляющая. Примерами могут служить циклы деловой активности, исследованные Кондратьевым, демографические, инвестиционные и другие циклы.

В экономических временных рядах редко представляется возможность для выделения и дальнейшего анализа циклической компоненты, так как ряды динамики экономических показателей часто оказываются слишком «короткими» для проведения такого исследования.

Хорошо изучены циклические составляющие в рядах динамики, относящихся к естественным наукам. Например, по многолетним наблюдениям установлена цикличность солнечной активности (с периодом колебаний примерно в 11 лет).

Если из временного ряда удалить тренд и периодические составляющие, то останется *нерегулярная* компонента.

Экономисты разделяют факторы, под действием которых формируется нерегулярная компонента, на два вида:

- факторы резкого, внезапного действия;
- текущие факторы.

Факторы первого вида (например, стихийные бедствия, эпидемии, война, кризис и т.д.), как правило, вызывают более значительные отклонения. Иногда такие отклонения называют катастрофическими колебаниями.

Факторы второго вида вызывают *случайные* колебания, являющиеся результатом действия большого числа побочных причин. Влияние каждого из текущих факторов незначительно, но ощущается их суммарное воздействие.

Если временной ряд представляется в виде суммы соответствующих компонент, то полученная модель носит название *аддитивной* (1.24), если в виде произведения — *мультипликативной* (1.25); также можно выделить еще один вид модели — *смешанного типа* (1.26):

$$y_t = u_t + s_t + v_t + \varepsilon_t \quad (1.24)$$

$$y_t = u_t \cdot s_t \cdot v_t \cdot \varepsilon_t \quad (1.25)$$

$$y_t = u_t \cdot s_t \cdot v_t + \varepsilon_t \quad (1.26)$$

где y_t — уровни временного ряда;

u_t — трендовая компонента;

s_t — сезонная компонента;

v_t — циклическая компонента;

ε_t — случайная компонента.

На рис. 1.1 и 1.2 приведены примеры временных рядов, иллюстрирующие присутствие в них указанных компонент. Графики

месячных временных рядов производства промышленной продукции наглядно демонстрируют устойчивые сезонные колебания при снижающемся тренде, причем на последнем участке темпы падения производства заметно снижаются.

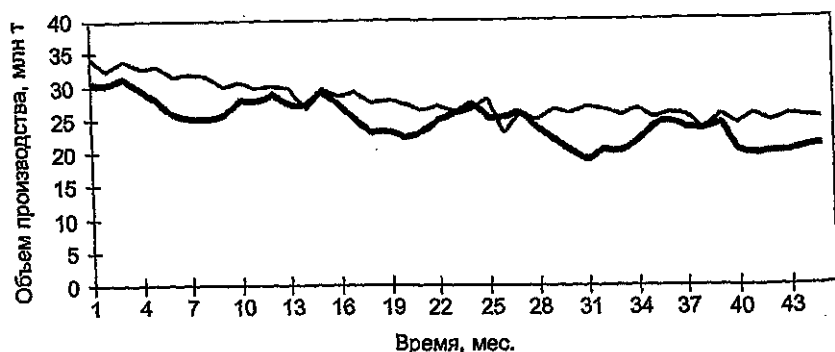


Рис. 1.1. Месячная динамика производства отдельных видов промышленной продукции в натуральном выражении:

— — нефть; — — уголь

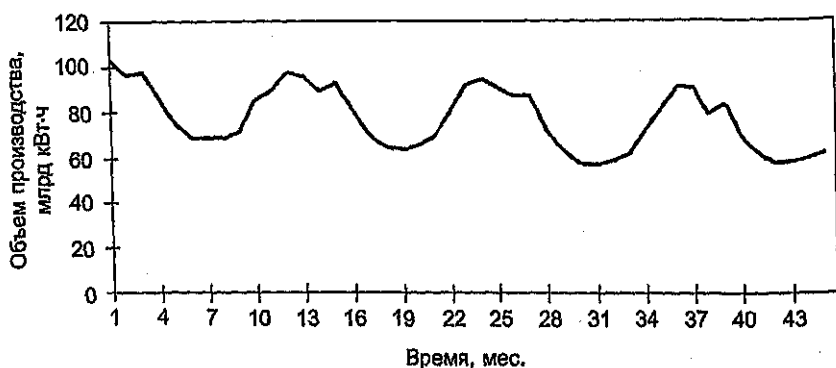


Рис. 1.2. Месячная динамика производства электроэнергии

Отметим, что не обязательно в процессе формирования значений уровней каждого временного ряда должны участвовать одновременно все компоненты. В изменении значений одного показателя может отсутствовать трендовая, другого — периодические составляющие, динамика третьего показателя может описываться лишь случайной составляющей. Однако наличие случайной, нерегулярной составляющей предполагается во всех случаях.

Решение любой задачи по анализу и прогнозированию временных рядов начинается с построения графика исследуемого показателя.

теля, тем более что современные программные средства предоставляют для этого пользователю большие возможности. Иногда на стадии графического анализа можно определить характер сезонных колебаний: аддитивный или мультипликативный. Отличительная особенность аддитивной модели заключается в том, что амплитуда сезонных колебаний, отражающая отклонения от тренда или среднего, остается примерно постоянной, неизменной во времени.

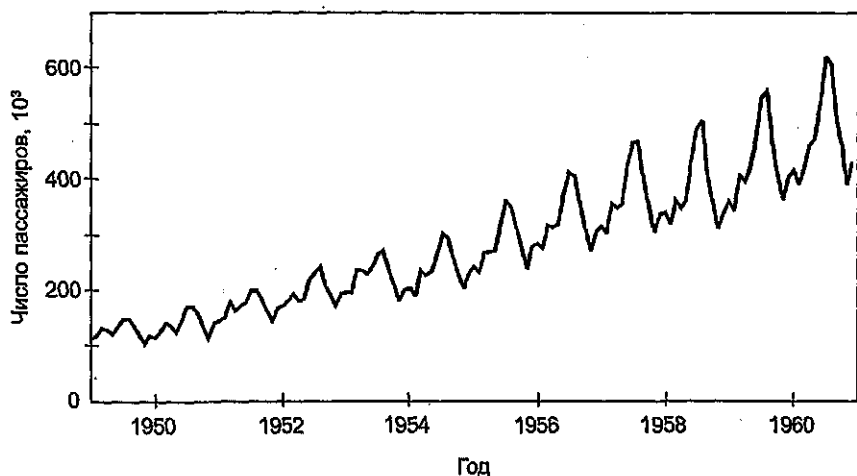


Рис. 1.3. Месячная динамика перевозок авиапассажиров за период с 1949 по 1960 гг.

На рис. 1.3 приведена динамика объема пассажирских перевозок, выполненных на авиалиниях Великобритании. На графике отчетливо прослеживаются сезонные колебания, наслаивающиеся на монотонно возрастающий тренд. Ежегодно повторяющиеся пики активности в авиаперевозках приходятся на праздники (рождественские и пасхальные каникулы), а также на время летних отпусков. Амплитуда сезонных колебаний возрастает с ростом объемов перевозок, что приводит к выводу о мультипликативном характере сезонности.

1.4. Стационарные временные ряды и их основные характеристики

Анализируемый временной ряд может рассматриваться как одна частная реализация стохастического процесса. Стационарные процессы представляют специальный случай стохастических процессов, обладая свойством эргодичности. При этом каждая отдельная

реализация случайного процесса служит «полномочным представителем» всей совокупности возможных реализаций. Согласно теореме Биркгофа—Хинчина [12] для эргодичных процессов основные характеристики можно приближенно вычислять не по нескольким реализациям, как в общем случае, а по одной.

Рассмотрим формальное определение стационарности.

Ряд называется *строго стационарным* (*strictly stationary*) или *стационарным в узком смысле*, если совместное распределение m наблюдений $y_{t_1}, y_{t_2}, \dots, y_{t_m}$ такое же, как и для m наблюдений

$$y_{t_1+\tau}, y_{t_2+\tau}, \dots, y_{t_m+\tau} \text{ при любых } m, t_1, t_2, \dots, t_m, \tau.$$

Таким образом, свойства строго стационарного временного ряда не зависят от начала отсчета времени.

Исследователя, как правило, интересует не все распределение, а средние значения и ковариации. Поэтому на практике чаще используется понятие *слабой стационарности* (*weak stationary*) или *стационарности в широком смысле*.

В этом случае стационарность временного ряда связывается с требованиями того, чтобы он имел среднее, дисперсию и ковариацию, не зависящие от момента времени t :

$$\begin{aligned} M(y_t) &= M(y_{t+\tau}) = \mu; \\ D(y_t) &= M(y_t - \mu)^2 = M(y_{t+\tau} - \mu)^2 = \gamma(0); \\ \text{cov}(y_t, y_{t+\tau}) &= M[(y_t - \mu)(y_{t+\tau} - \mu)] = \gamma(\tau). \end{aligned} \quad (1.27)$$

В дальнейшем изложении под термином «стационарность» будем иметь в виду *слабую стационарность*.

Из (1.27) следует, что автоковариация будет зависеть только от сдвига по времени τ и не будет зависеть от t .

Здесь ковариация называется *автоковариацией*, так как характеризует статистическую связь между уровнями одного и того же временного ряда, отстоящими на τ временных тактов.

При анализе изменения $\gamma(\tau)$ в зависимости от временного сдвига τ принято говорить об *автоковариационной функции* (*autocovariation function*).

Очевидно, что при $\tau = 0$ автоковариационная функция равна дисперсии временного ряда (см. (1.27)).

С понятием автоковариационной функции тесно связана *автокорреляционная функция*, АКФ (*autocorrelation function*, ACF):

$$\rho(\tau) = \frac{\text{cov}(y_t, y_{t+\tau})}{D(y_t)} = \frac{\gamma(\tau)}{\gamma(0)}. \quad (1.28)$$

Значения АКФ также характеризуют тесноту (степень) статистической связи между уровнями временного ряда, разделенными τ временными тактами. Однако в отличие от значений автоковариационной функции они безразмерны, не зависят от масштаба измерения уровней исследуемого временного ряда и подчиняются ограничению

$$|\rho(\tau)| \leq 1 \text{ (очевидно, что } \rho(0) = 1\text{)}.$$

Из условия стационарности следует, что $\rho(\tau) = \rho(-\tau)$, поэтому при рассмотрении поведения АКФ ограничиваются лишь областью положительных значений τ .

На практике значения АКФ статистически оцениваются по имеющимся уровням временного ряда. Выборочная оценка коэффициента автокорреляции $r(\tau)$ может быть определена следующим образом [1]:

$$r(\tau) = \frac{\frac{1}{n-\tau} \sum_{t=1}^{n-\tau} (y_t - \bar{y})(y_{t+\tau} - \bar{y})}{\frac{1}{n} \sum_{t=1}^n (y_t - \bar{y})^2}, \quad (1.29)$$

где n — длина временного ряда $y_1, y_2, \dots, y_n, \dots, y_n$;

τ — временной сдвиг ($\tau = 1, 2, \dots, n-1$);

$\bar{y}$ — оценка среднего значения, найденная по формуле:

$$\bar{y} = \frac{\sum_{t=1}^n y_t}{n}.$$

Отметим, что числитель выражения (1.29) представляет выборочную оценку коэффициента автоковариации. Иногда график АКФ, отражающий изменение $r(\tau)$ в зависимости от сдвига τ , называют *коррелограммой* (*correlogram*).

Очевидно, что с увеличением значения лага τ число пар наблюдений ($n - \tau$), используемых для расчета в (1.29), уменьшается. Поэтому в практических руководствах рекомендуется поддерживать соотношение $\tau \leq n/4$.

Для стационарного временного ряда с увеличением τ автокорреляционная функция должна демонстрировать свойство монотонного убывания по абсолютной величине, так как взаимосвязь между уровнями ряда с ростом τ ослабевает. Однако это условие может нарушаться для выборочных АКФ.

Идея перенесения частной корреляции, «очищенной» взаимосвязи, на временные ряды находит свое выражение в *частной автокорреляционной функции*, ЧАКФ (*partial autocorrelation function*, *PACF*). С помощью ЧАКФ измеряется корреляция между уровнями ряда y_t и $y_{t+\tau}$, разделенными τ временными тактами, при исключении влияния на эту взаимосвязь всех промежуточных уровней ряда $y_{t+1}, y_{t+2}, \dots, y_{t+\tau-1}$.

Например, коэффициент частной автокорреляции ρ_τ при $\tau = 2$ будет определять корреляцию между уровнями временного ряда, разделенными двумя тактами времени, при условии, что значения промежуточных уровней зафиксированы на среднем уровне:

$$\rho_\tau(2) = \rho(y_t, y_{t+2} | y_{t+1} = \mu).$$

В эконометрических пакетах присутствует возможность построения графика ЧАКФ, на котором показаны выборочные оценки коэффициентов частной автокорреляции в зависимости от лагов τ .

Очевидно, что коэффициент частной автокорреляции $\rho_\tau(1)$ для лага $\tau = 1$ будет равен коэффициенту автокорреляции $\rho(1)$, так как при этом значении τ нет промежуточных лагов. Но при $\tau > 1$ уже появятся отличия в значениях этих коэффициентов.

Примером стационарности служит *белый шум* (*white noise*).

Временные ряды, образующие белый шум, удовлетворяют следующим условиям:

$$y_t = \varepsilon_t, \quad (1.30)$$

$$\text{где } \begin{cases} M\varepsilon_t = 0; \\ M(\varepsilon_t \varepsilon_{t+\tau}) = \begin{cases} \sigma_0^2 & \text{при } \tau = 0; \\ 0 & \text{при } \tau \neq 0. \end{cases} \end{cases}$$

Таким образом,

$$\gamma(\tau) = 0 \quad (\rho(\tau) = 0) \quad \text{при } \tau > 0,$$

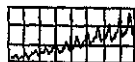
$$\gamma(0) = \sigma_0^2 \quad (\text{причем постоянная дисперсия } \sigma_0^2 \text{ не зависит от } t).$$

В качестве примера белого шума можно привести остатки, рассматриваемые в классической линейной регрессионной модели. В случае нормального распределения остатков они образуют гауссовский белый шум.

Следует подчеркнуть, что белый шум не обязательно должен подчиняться нормальному распределению, в то же время нормальный (гауссовский) белый шум удовлетворяет условию

$$\begin{cases} \varepsilon_t \in N(0, \sigma_0^2); \\ M(\varepsilon_t, \varepsilon_{t+\tau}) = \begin{cases} \sigma_0^2 & \text{при } \tau = 0; \\ 0 & \text{при } \tau \neq 0. \end{cases} \end{cases} \quad (1.31)$$

В заключение отметим, что вид АКФ и ЧАКФ оказывает существенную помощь в выборе моделей, описывающих поведение анализируемых временных рядов.



Практикум к главе 1

Контрольные вопросы

1. Какие виды временных рядов вы знаете? Приведите примеры.
2. Поясните, в чем состоят характерные отличия временных рядов от пространственных выборок.
3. Какие требования предъявляются к временным рядам как к исходной информации при прогнозировании?
4. Как рассчитываются средний абсолютный прирост, средний темп роста, средний темп прироста? Когда правомерно использовать средний абсолютный прирост и средний темп роста для расчета прогнозов?
5. Как на стадии графического анализа динамики временного ряда можно определить характер сезонности (аддитивный или мультипликативный)?
6. Поясните понятие «белый шум». Каковы его свойства?
7. Что такое автокорреляционная функция? Как можно рассчитать выборочную оценку коэффициента автокорреляции?
8. Охарактеризуйте компоненты временных рядов. Что такое мультипликативная (аддитивная) модель временного ряда?
9. Дайте определение стационарного временного ряда в узком и широком смысле. Можно ли утверждать, что стационарный в узком смысле временной ряд является одновременно и стационарным в широком смысле? Будет ли справедливым обратное утверждение?
10. При работе на фондовом рынке используются термины «бычий тренд», «медвежий тренд», «боковой тренд». Поясните, в каком случае применяется каждый из них.

Упражнения

- 1.1. В таблице представлены данные о численности безработных, (Российский статистический ежегодник: Стат. сб. — М.: Госкомстат России, 2001).

Год	1995	1996	1997	1998	1999	2000
Численность, тыс. чел.	6712	6732	8058	8902	9094	6999

Рассчитайте цепные, базисные и средние:

а) абсолютные приросты;

б) темпы роста;

в) темпы прироста.

В качестве базисного следует взять уровень 1995 г.

Дайте экономическую интерпретацию полученным результатам.

- 1.2. Изменение жилищного фонда города происходило с примерно постоянным темпом роста в течение пяти лет (с 1997 по 2001 г.). Средний темп роста $\bar{T}$ составил 102,7%. Рассчитайте прогнозное значение жилищного фонда города в 2002 г. (время упреждения $L = 1$), если в 2000 г. он составил 2600 тыс. м².
- 1.3. Для временного ряда из упражнения 1 рассчитайте оценки коэффициентов автокорреляции для значения лага $\tau = 1, 2$.
- 1.4. Анализ ежеквартальной динамики прибыли фирмы, проведенный за шесть последних кварталов, показал, что значения цепных абсолютных приростов примерно одинаковы. Средний абсолютный прирост $\Delta_{\bar{y}}$ составил 9,5 тыс. долл. Рассчитайте прогнозное значение прибыли фирмы в следующем полугодии, если в последнем, шестом квартале она равнялась 65 тыс. долл.



СГЛАЖИВАНИЕ ВРЕМЕННЫХ РЯДОВ С ПОМОЩЬЮ СКОЛЬЗЯЩИХ СРЕДНИХ

2.1. Применение простых скользящих средних

При анализе рядов динамики возникает важная задача: определение основной тенденции в развитии исследуемого явления. В некоторых случаях общая тенденция ясно прослеживается в динамике показателя, в других ситуациях она может не просматриваться из-за ощутимых случайных колебаний. Например, в отдельные моменты времени сильные колебания в курсах акций могут заслонить наличие тенденции к росту или снижению этого показателя.

На практике для обнаружения общей тенденции часто используют простой прием — *укрупнение интервалов*. Например, ряд недельных данных можно преобразовать в ряд месячной динамики, ряд квартальных данных заменить годовыми уровнями. Уровни нового ряда могут быть получены суммированием уровней исходного ряда либо могут представлять средние значения.

При выявлении тенденции развития используется распространенный прием — *сглаживание временного ряда*. Суть различных приемов сглаживания сводится к замене фактических уровней временного ряда расчетными, которые в меньшей степени подвержены колебаниям. Это способствует более четкому проявлению тенденции развития.

Методы сглаживания можно условно разделить на два класса, опирающиеся на различные подходы: *аналитический и алгоритмический*.

Аналитический подход основан на допущении, что исследователь может задать общий вид функции, описывающей регулярную, неслучайную составляющую. Например, на основе визуального и содержательного экономического анализа динамики временного ряда предполагается, что трендовая составляющая может быть описана с помощью показательной функции $\hat{y}_t = ab^t$.

Тогда на следующем этапе будет проведена статистическая оценка неизвестных коэффициентов модели, а затем определены сглаженные значения уровней временного ряда путем подстановки соответствующего значения временного параметра t в полученное уравнение (заданное в явном аналитическом виде). Процедуры моделирования, опирающиеся на этот подход, рассматриваются в гл. 3.

В *алгоритмическом подходе* отказываются от ограничительного допущения, свойственного аналитическому. Процедуры этого класса

не предполагают описания динамики неслучайной составляющей с помощью единой функции, они предоставляют исследователю лишь алгоритм расчета неслучайной составляющей в любой заданный момент времени t . Методы сглаживания временных рядов с помощью скользящих средних относятся к этому подходу.

Скользящие средние позволяют сгладить как случайные, так и периодические колебания, выявить имеющуюся тенденцию в развитии процесса и поэтому служат важным инструментом при фильтрации компонент временного ряда.

Алгоритм сглаживания *по простой скользящей средней* может быть представлен в виде следующей последовательности шагов.

1. Определяют длину интервала сглаживания l , включающего в себя l последовательных уровней ряда ($l < n$). При этом надо иметь в виду, что чем шире интервал сглаживания, тем в большей степени поглощаются колебания, и тенденция развития носит более плавный, сглаженный характер. Чем сильнее колебания, тем шире должен быть интервал сглаживания.

2. Разбивают весь период наблюдений на участки, при этом интервал сглаживания «скользит» по ряду с шагом, равным 1.

3. Рассчитывают средние арифметические из уровней ряда, образующих каждый участок.

4. Заменяют фактические значения ряда, стоящие в центре каждого участка, на соответствующие средние значения.

При этом удобно брать длину интервала сглаживания l в виде нечетного числа $l = 2p + 1$, так как в этом случае полученные значения скользящей средней приходятся на средний член интервала.

Наблюдения, которые берутся для расчета среднего значения, называются *активным участком сглаживания*.

При нечетном значении $l = 2p + 1$ все уровни активного участка могут быть представлены в виде:

$$y_{t-p}, y_{t-p+1}, \dots, y_{t-1}, y_t, y_{t+1}, \dots, y_{t+p-1}, y_{t+p},$$

где y_t — центральный уровень активного участка;
 $y_{t-p}, y_{t-p+1}, \dots, y_{t-1}$ — последовательность из p уровней активного участка, предшествующих центральному;
 $y_{t+1}, \dots, y_{t+p-1}, y_{t+p}$ — последовательность из p уровней активного участка, следующих за центральным.

Тогда скользящая средняя может быть определена по формуле:

$$\bar{y}_t = \frac{\sum_{i=t-p}^{t+p} y_i}{2p+1} = \frac{y_{t-p} + y_{t-p+1} + \dots + y_{t+p-1} + y_{t+p}}{2p+1}, \quad (2.1)$$

где y_i — фактическое значение i -го уровня;
 $\hat{y}_t$ — значение скользящей средней в момент t ;
 $2p + 1$ — длина интервала сглаживания.

Процедуры скользящих средних опираются на известную теорему Вейерштрасса, согласно которой «любая гладкая функция при самых общих допущениях может быть локально (т.е. в ограниченном интервале изменения ее аргумента t) представлена алгебраическим полиномом подходящей степени» [1, с. 805].

При реализации простой скользящей средней выравнивание на каждом активном участке проводится по прямой (по полиному первого порядка). Таким образом, осуществляется аппроксимация неслучайной составляющей с помощью линейной функции времени: $\hat{y}_t = a_0 + a_1 t$.

Рассмотрим активный участок с длиной интервала сглаживания $l = 2p + 1$. Перенесем начало координат в середину временного интервала, т.е. будем рассматривать моменты времени

$$t = -p, -p + 1, \dots, -1, 0, 1, \dots, p - 1, p.$$

Неизвестные коэффициенты линейной модели подбираются таким образом, чтобы минимизировать критерий метода наименьших квадратов (МНК):

$$Q = \sum_{t=-p}^p (y_t - a_0 - a_1 t)^2 \Rightarrow \min.$$

Находим частные производные по a_0, a_1 и приравниваем их к нулю:

$$\frac{\partial Q}{\partial a_j} = 0, \quad j = 0; 1.$$

Отсюда получаем систему нормальных уравнений:

$$\begin{cases} \sum_{t=-p}^p y_t = a_0(2p+1) + a_1 \sum_{t=-p}^p t; \\ \sum_{t=-p}^p y_t t = a_0 \sum_{t=-p}^p t + a_1 \sum_{t=-p}^p t^2. \end{cases} \quad (2.2)$$

Очевидно, что за счет выполненного переноса начала координат $\sum_{t=-p}^p t = 0$, а сглаженное значение в центральной точке активного участка определяется коэффициентом a_0 . Из первого уравнения системы (2.2) получаем выражение для этого коэффициента:

$$a_0 = \frac{\sum_{t=-p}^p y_t}{2p+1}.$$

Таким образом, в качестве сглаженного значения в центральной точке активного участка следует брать среднее арифметическое из уровней ряда, образующих этот участок. Полученный вывод — обоснование ранее рассмотренного алгоритма сглаживания.

Процедура сглаживания приводит к устранению периодических колебаний во временном ряду, если длина интервала сглаживания берется равной или кратной периоду колебаний.

Для устранения сезонных колебаний на практике часто требуется использовать скользящие средние с длиной интервала сглаживания, равной 4 или 12; но при этом не будет выполняться условие нечетности. Поэтому при четном числе уровней принято первое и последнее наблюдение на активном участке брать с половинными весами:

$$\begin{aligned} \hat{y}_t &= \frac{\frac{1}{2}y_{t-p} + y_{t-p+1} + \dots + y_{t-1} + y_t + y_{t+1} + \dots + y_{t+p-1} + \frac{1}{2}y_{t+p}}{2p} = \\ &= \frac{\frac{1}{2}y_{t-p} + \sum_{i=t-p+1}^{t+p-1} y_i + \frac{1}{2}y_{t+p}}{2p}. \end{aligned} \quad (2.3)$$

Тогда для сглаживания сезонных колебаний при работе с временными рядами квартальной или месячной динамики можно использовать 4- (2.4) и 12-членную (2.5) скользящую среднюю:

$$\hat{y}_t = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4}; \quad (2.4)$$

$$\hat{y}_t = \frac{\frac{1}{2}y_{t-6} + y_{t-5} + \dots + y_t + \dots + y_{t+5} + \frac{1}{2}y_{t+6}}{12}. \quad (2.5)$$

В (2.4) каждый активный участок содержит 5 уровней, в (2.5) — 13, при этом крайние уровни имеют половинные весовые коэффициенты.

Метод простой скользящей средней применим, если графическое изображение динамического ряда напоминает прямую. Если для процесса характерно нелинейное развитие, то простая скользящая средняя может привести к существенным искажениям. Когда тренд

выравниваемого ряда имеет изгибы и для исследователя желательно сохранить мелкие волны, то целесообразно использовать взвешенную скользящую среднюю.

2.2. Использование взвешенных скользящих средних

При построении *взвешенной скользящей средней* на каждом активном участке значение центрального уровня заменяется на расчетное, определяемое по формуле средней арифметической взвешенной:

$$\hat{y}_t = \frac{\sum_{i=t-p}^{t+p} y_i w_i}{\sum_{i=t-p}^{t+p} w_i}, \quad (2.6)$$

где w_i — весовые коэффициенты.

Простая скользящая средняя учитывает все уровни ряда, входящие в активный участок сглаживания, с равными весами (w_i), а взвешенная средняя приписывает каждому уровню вес, зависящий от удаления данного уровня до уровня, стоящего в середине активного участка. Это вызвано тем, что при простой скользящей средней выравнивание на каждом активном участке проводится по прямой (полиному первого порядка), а при сглаживании по взвешенной скользящей средней используются полиномы более высоких порядков, чаще всего второго или третьего. Поэтому метод простой скользящей средней может рассматриваться как частный случай метода взвешенной скользящей средней.

На рис. 2.1 показаны результаты применения скользящих средних для активного участка с длиной интервала сглаживания $l = 9$. Значение центрального уровня, соответствующее после переноса начала координат времени $t = 0$, заменяется на расчетное. Для простой скользящей средней оно определяется с помощью выравнивания по прямой, для взвешенной — с помощью выравнивания по квадратичному полиному.

Весовые коэффициенты определяются с помощью МНК, причем нет необходимости каждый раз вычислять их заново при уровнях ряда, входящих в активный участок сглаживания, так как они будут одинаковыми для каждого активного участка.

Пусть длина интервала сглаживания $l = 7$, а локальное поведение сглаженного временного ряда внутри каждого активного участка описывается с помощью полинома 3-го порядка. Перенесем начало координат в середину временного интервала, т.е. будем рассматривать моменты времени $t = -3, -2, -1, 0, 1, 2, 3$.

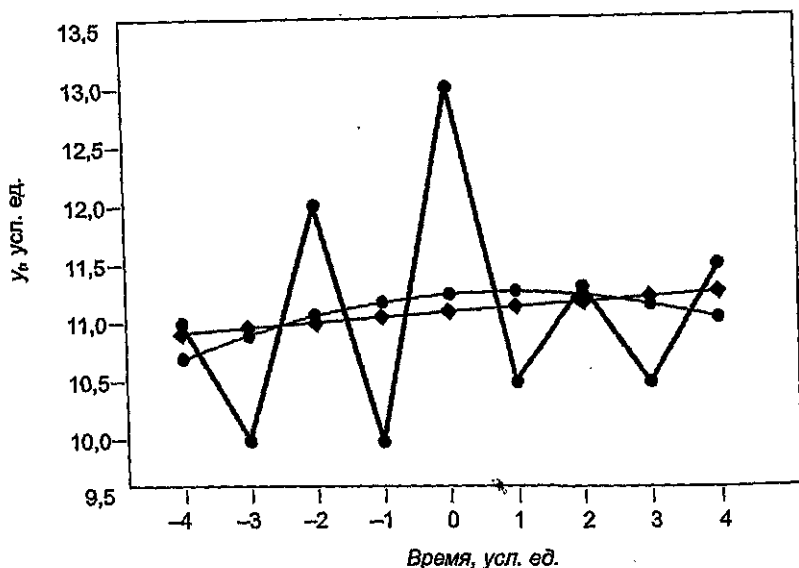


Рис. 2.1. Сглаживание с помощью простой и взвешенной скользящей средней:

- фактические уровни активного участка;
- ◆— выравнивание по полиному первой степени;
- выравнивание по полиному второй степени

Неизвестные коэффициенты полинома 3-й степени оцениваются с помощью МНК, т.е. находятся коэффициенты, минимизирующие функционал

$$Q = \sum_{t=-3}^3 (y_t - a_0 - a_1 t - a_2 t^2 - a_3 t^3) \Rightarrow \min.$$

Находим частные производные и приравняем их к нулю:

$$\frac{\partial Q}{\partial a_j} = 0, \quad j = 0; 1; 2; 3.$$

Отсюда получаем четыре уравнения следующего вида:

$$\sum_{t=-3}^3 t^j y_t = a_0 \sum_{t=-3}^3 t^j + a_1 \sum_{t=-3}^3 t^{j+1} + a_2 \sum_{t=-3}^3 t^{j+2} + a_3 \sum_{t=-3}^3 t^{j+3},$$

где $j = 0; 1; 2; 3$.

Учитывая, что после переноса начала координат в середину временного интервала

$$\sum_{t=-3}^3 t^k = 0,$$

где k — нечетное число,

получим упрощенную систему нормальных уравнений:

$$\begin{cases} \sum_{t=-3}^3 y_t = 7a_0 + 28a_2; \\ \sum_{t=-3}^3 y_t t = 28a_1 + 196a_3; \\ \sum_{t=-3}^3 y_t t^2 = 28a_0 + 196a_2; \\ \sum_{t=-3}^3 y_t t^3 = 196a_1 + 1588a_3. \end{cases} \quad (2.7)$$

Сглаженное значение в центральной точке активного участка определяется коэффициентом a_0 , который входит в первое и третье уравнения системы (2.7).

Умножим левую и правую части первого уравнения на 7, а затем вычтем из него третье уравнение:

$$7 \sum_{t=-3}^3 y_t - \sum_{t=-3}^3 y_t t^2 = 21a_0.$$

Следовательно,

$$\begin{aligned} a_0 &= \frac{1}{21} \left(7 \sum_{t=-3}^3 y_t - \sum_{t=-3}^3 y_t t^2 \right) = \\ &= \frac{1}{21} [7(y_{-3} + y_{-2} + y_{-1} + y_0 + y_1 + y_2 + y_3) - (9y_{-3} + 4y_{-2} + y_{-1} + y_1 + 4y_2 + 9y_3)] = \\ &= \frac{1}{21} (-2y_{-3} + 3y_{-2} + 6y_{-1} + 7y_0 + 6y_1 + 3y_2 - 2y_3). \end{aligned}$$

Таким образом, оценка сглаженного значения в центральной точке активного участка определяется как взвешенная средняя арифметическая из семи уровней, образующих этот участок (см. (2.6)).

Соответствующие весовые коэффициенты равны

$$-\frac{2}{21}; \frac{3}{21}; \frac{6}{21}; \frac{7}{21}; \frac{6}{21}; \frac{3}{21}; -\frac{2}{21}.$$

Учитывая симметрию относительно центрального значения, их можно представить с помощью символической записи (табл. 2.1):

$$\frac{1}{21} [-2; 3; 6; 7]$$

Отметим, что использование в качестве аппроксимирующей функции полинома 2-й степени (вместо полинома 3-й степени) не привело бы к изменению весовых коэффициентов. Это объясняется тем, что первое и третье уравнения в системе (2.7), содержащие коэффициент a_0 , остались бы без изменения.

Процедура определения весовых коэффициентов, проиллюстрированная на простом примере, носит общий характер. Если для каждого активного участка с длиной интервала сглаживания $l = 2p + 1$ подбирается полином порядка m , то согласно МНК необходимо минимизировать функционал

$$Q = \sum_{t=-p}^p (y_t - a_0 - a_1 t - \dots - a_m t^m)^2.$$

При этом весовые коэффициенты, найденные для сглаживания по полиномам четной степени $m = 2k$, будут неизменными при использовании полиномов степени $m' = 2k + 1$ (т.е. для полиномов на единицу большей нечетной степени).

В табл. 2.1 представлены весовые коэффициенты в зависимости от длины интервала сглаживания (при сглаживании по полиному второго или третьего порядка).

Таблица 2.1. Весовые коэффициенты для взвешенной скользящей средней

Длина интервала сглаживания	Весовые коэффициенты
5	$\frac{1}{35}[-3; +12; +17]$
7	$\frac{1}{21}[-2; +3; +6; +7]$
9	$\frac{1}{231}[-21; +14; +39; +54; +59]$
11	$\frac{1}{429}[-36; +9; +44; +69; +84; +89]$
13	$\frac{1}{143}[-11; 0; +9; +16; +21; +24; +25]$

Так как веса симметричны относительно центрального уровня, то в таблице использована символическая запись: приведены веса для половины уровней активного участка; выделен вес, относящийся к уровню, стоящему в центре участка сглаживания. Для оставшихся уровней веса не приводятся, так как они могут быть симметрично отражены.

Отметим важные свойства весовых коэффициентов:

- 1) они симметричны относительно центрального уровня;
- 2) сумма весов с учетом общего множителя, вынесенного за скобки, равна единице;
- 3) наличие как положительных, так и отрицательных весов позволяет сглаженной кривой сохранять различные изгибы кривой тренда.

Проиллюстрируем использование табл. 2.1 на примере вычисления 5-членной взвешенной скользящей средней. В этом случае центральное значение на каждом активном участке $y_{t-2}, y_{t-1}, y_t, y_{t+1}, y_{t+2}$ будет оцениваться по формуле:

$$\bar{y}_t = \frac{1}{35}(-3y_{t-2} + 12y_{t-1} + 17y_t + 12y_{t+1} - 3y_{t+2}),$$

где соответствующие весовые коэффициенты уровней $-3/35, 12/35, 17/35$ взяты из первой строки табл. 2.1.

Пример 2.1

По данным об урожайности (табл. 2.2) за 16 лет рассчитайте:

- 1) трех-, семилетние скользящие средние и графически сравните результаты;
- 2) пятилетнюю взвешенную скользящую среднюю.

Таблица 2.2. Урожайность пшеницы, ц/га

Текущий номер года t	1	2	3	4	5	6	7	8
y_t	10,3	14,3	7,7	15,8	14,4	16,7	15,3	20,2
Текущий номер года t	9	10	11	12	13	14	15	16
y_t	17,1	7,7	15,3	16,3	19,9	14,4	18,7	20,7

Решение

1. Результаты расчетов представлены в табл. 2.3.

Таблица 2.3. Расчет скользящих средних

t	y _t	Скользящие средние		Взвешенная скользящая средняя l = 5
		l = 3	l = 7	
1	2	3	4	5
1	10,3	—	—	—
2	14,3	10,8	—	—
3	7,7	12,6	—	11,9
4	15,8	12,6	13,5	12,6
5	14,4	15,6	14,9	16,2
6	16,7	15,5	15,3	15,2
7	15,3	17,4	15,3	17,4
8	20,2	17,5	15,2	18,8
9	17,1	15,0	15,5	15,2
10	7,7	13,4	16,0	11,7
11	15,3	13,1	15,8	12,5
12	16,3	17,2	15,6	18,1
13	19,9	16,9	16,1	17,3
14	14,4	17,7	—	17,1
15	18,7	17,9	—	—
16	20,7	—	—	—

При трехлетней скользящей средней (гр. 3 табл. 2.3)

$$\hat{y}_2 = \frac{10,3 + 14,3 + 7,7}{3} = 10,8;$$

$$\hat{y}_3 = \frac{14,3 + 7,7 + 15,8}{3} = 12,6 \text{ и т.д.}$$

При семилетней скользящей средней (гр. 4)

$$\hat{y}_4 = \frac{10,3 + 14,3 + 7,7 + 15,8 + 14,4 + 16,7 + 15,3}{7} = 13,5;$$

$$\hat{y}_5 = \frac{14,3 + 7,7 + 15,8 + 14,4 + 16,7 + 15,3 + 20,2}{7} = 14,9 \text{ и т.д.}$$

Графический анализ показывает (рис. 2.2), что ряд, сглаженный по семилетней скользящей средней, носит более гладкий характер. Это объясняется тем, что чем больше длина интервала сглаживания, тем более гладкий ряд получается на выходе модели.

2. Для вычисления значений пятилетней взвешенной скользящей средней воспользуемся табл. 2.1. Тогда

$$\bar{y}_3 = \frac{1}{35}(-3 \cdot 10,3 + 12 \cdot 14,3 + 17 \cdot 7,7 + 12 \cdot 15,8 - 3 \cdot 14,4) = 11,9;$$

$$\hat{y}_4 = \frac{1}{35}(-3 \cdot 14,3 + 12 \cdot 7,7 + 17 \cdot 15,8 + 12 \cdot 14,4 - 3 \cdot 16,7) = 12,6 \text{ и т.д.}$$

Гр. 5 отражает результаты дальнейших расчетов.

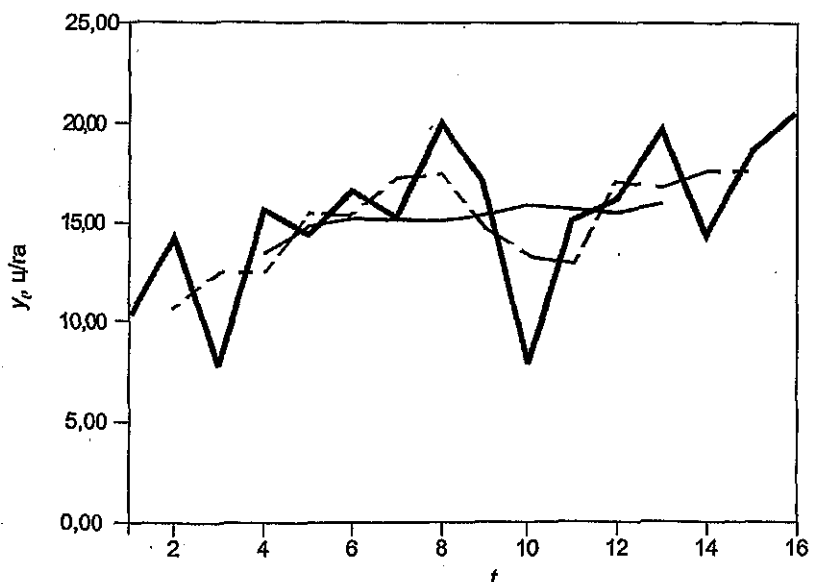


Рис. 2.2. Сглаживание ряда урожайности с помощью скользящих средних: — фактические уровни y_t ; - - $l=3$; — $l=7$

2.3. Восстановление краевых значений

При использовании скользящей средней с длиной активного участка $l = 2p + 1$ первые и последние p уровней ряда сгладить нельзя, их значения теряются. Очевидно, что потеря значений последних точек — существенный недостаток, так как для исследователя «свежие» данные обладают наибольшей информационной ценностью.

Рассмотрим один из приемов, позволяющих восстановить потерянные значения временного ряда при использовании простой скользящей средней. Для этого необходимо:

- вычислить средний абсолютный прирост на последнем активном участке;
- получить p сглаженных значений в конце временного ряда путем последовательного прибавления среднего абсолютного прироста к последнему сглаженному значению.

Аналогичную процедуру можно реализовать для оценивания первых уровней временного ряда.

Рассмотрим еще один из возможных способов восстановления краевых значений. Для определения p первых и p последних потерянных уровней анализируемого временного ряда можно использовать расчетные значения, полученные с помощью аппроксими-

рующих полиномов той же степени, что и для сглаживания остальных членов ряда. Причем неизвестные коэффициенты полиномов определяются соответственно по $l = 2p + 1$ первым и последним уровням временного ряда.

Пример 2.2

Пусть сглаживание осуществляется по пятичленной скользящей средней ($l = 5$), причем аппроксимация осуществляется квадратичным полиномом ($m = 2$). Требуется определить весовые коэффициенты для восстановления двух последних уровней ряда.

Решение

Осуществим перенос начала координат в середину активного участка:

$$t = -2; -1; 0; 1; 2.$$

После этого система нормальных уравнений примет вид:

$$\begin{cases} \sum_{t=-2}^2 y_t = 5a_0 + 10a_2; \\ \sum_{t=-2}^2 ty_t = 10a_1; \\ \sum_{t=-2}^2 t^2 y_t = 10a_0 + 34a_2. \end{cases} \quad (2.8)$$

Из 1 и 3-го уравнений системы (2.8) определим выражение для коэффициента a_0 :

$$a_0 = \frac{1}{35}(-3y_{-2} + 12y_{-1} + 17y_0 + 12y_1 - 3y_2),$$

или в символической записи (см. табл. 2.1):

$$a_0 = \frac{1}{35}[-3; 12; 17].$$

Выразим теперь и остальные неизвестные параметры из системы (2.8):

$$a_1 = \frac{\sum_{t=-2}^2 ty_t}{10} = \frac{1}{10}(-2y_{-2} - y_{-1} + y_1 + 2y_2);$$

$$a_2 = \frac{\sum_{t=-2}^2 t^2 y_t - 2 \sum_{t=-2}^2 ty_t}{14} = \frac{1}{14}(2y_{-2} - y_{-1} - 2y_0 - y_1 + 2y_2).$$

Полученные выражения для коэффициентов a_0, a_1, a_2 подставим в уравнение сглаживающего квадратичного полинома:

$$\hat{y}_t = \frac{1}{35}(-3y_{-2} + 12y_{-1} + 17y_0 + 12y_1 - 3y_2) + \frac{t}{10}(-2y_{-2} - y_{-1} + y_1 + 2y_2) + \frac{t^2}{14}(2y_{-2} - y_{-1} - 2y_0 - y_1 + 2y_2).$$

Последовательно подставляя в это выражение $t = 1; 2$, получим весовые коэффициенты для восстановления последних уровней ряда:

- при $t = 1$ (восстановление предпоследнего уровня ряда)

$$\hat{y}_1 = \frac{1}{35}[-5; 6; 12; 13; 9];$$

- при $t = 2$ (восстановление последнего уровня ряда)

$$\hat{y}_2 = \frac{1}{35}[3; -5; -3; 9; 31].$$

Если последними пятью уровнями ряда были бы 0; 1; 4; 9; 16, то восстановление двух последних значений осуществлялось бы следующим образом:

- при $t = 1$

$$\hat{y}_1 = \frac{1}{35}[(-5 \cdot 0) + (6 \cdot 1) + (12 \cdot 4) + (13 \cdot 9) + (9 \cdot 16)] = 9;$$

- при $t = 2$

$$\hat{y}_2 = \frac{1}{35}[(3 \cdot 0) + (-5 \cdot 1) + (-3 \cdot 4) + (9 \cdot 9) + (31 \cdot 16)] = 16.$$

Полученные результаты точно совпали с исходными уровнями, так как мы использовали квадратичный полином для выравнивания последовательности квадратов.

2.4. Эффект Служкого—Юла

Временной ряд, полученный после применения процедуры скользящих средних, также содержит случайную составляющую $\tilde{\varepsilon}_t$, однако ее влияние будет сглажено и поэтому выражено не так явно. Очевидно, что уменьшение дисперсии случайных колебаний зависит от интервала сглаживания l .

Рассмотрим аддитивную модель исходного временного ряда, в которой случайная компонента ε_t удовлетворяет следующим свойствам:

$$M\varepsilon_t = 0;$$

$$M(\varepsilon_t \varepsilon_{t \pm \tau}) = \begin{cases} \sigma^2 & \text{при } \tau = 0; \\ 0 & \text{при } \tau \neq 0, \end{cases} \quad (2.9)$$

т.е. случайные остатки взаимно некоррелированы и гомоскедастичны.

Пусть длина интервала сглаживания $l = 2p + 1$, тогда весовые коэффициенты для каждого активного участка могут быть представлены в виде:

$$w_{t-p}, w_{t-p+1}, \dots, w_{t-1}, w_t, w_{t+1}, \dots, w_{t+p-1}, w_{t+p}.$$

Эти же весовые коэффициенты будут использованы для определения оценки случайной составляющей в центральной точке активного участка:

$$\tilde{\varepsilon}_t = \sum_{i=t-p}^{t+p} w_i \varepsilon_i.$$

Учитывая (2.9), можно записать:

$$M\tilde{\varepsilon}_t = 0; \quad (2.10)$$

$$D\tilde{\varepsilon}_t = \sigma^2 \sum_{i=t-p}^{t+p} w_i^2. \quad (2.11)$$

Тогда

$$\begin{aligned} \text{cov}(\tilde{\varepsilon}_t, \tilde{\varepsilon}_{t+\tau}) = M[(w_{t-p}\varepsilon_{t-p} + w_{t-p+1}\varepsilon_{t-p+1} + \dots + w_{t+p-1}\varepsilon_{t+p-1} + \\ + w_{t+p}\varepsilon_{t+p}) \times (w_{t-p+\tau}\varepsilon_{t-p+\tau} + w_{t-p+1+\tau}\varepsilon_{t-p+1+\tau} + \dots + \\ + w_{t+p-1+\tau}\varepsilon_{t+p-1+\tau} + w_{t+p+\tau}\varepsilon_{t+p+\tau})]. \end{aligned}$$

Учитывая взаимную некоррелированность исходных случайных остатков (см. (2.9)), последнее выражение можно представить в виде:

$$\text{cov}(\tilde{\varepsilon}_t, \tilde{\varepsilon}_{t+\tau}) = \sigma^2 \sum_{i=t-p}^{t+p-\tau} w_i w_{i+\tau}. \quad (2.12)$$

Перейдем к коэффициенту автокорреляции порядка τ (для случайных остатков $\tilde{\varepsilon}_t$, отстоящих друг от друга на τ тактов времени). С учетом (2.12) и (2.11) получим

$$\rho_\tau = \frac{\text{cov}(\tilde{\varepsilon}_t, \tilde{\varepsilon}_{t+\tau})}{\sigma^2 \sum_{i=t-p}^{t+p} w_i^2} = \frac{\sum_{i=t-p}^{t+p-\tau} w_i w_{i+\tau}}{\sum_{i=t-p}^{t+p} w_i^2}, \quad (2.13)$$

где $\tau = 1; 2; \dots, 2p$.

Из (2.13) вытекает важное следствие: коэффициенты автокорреляции для производного, сглаженного ряда будут отличны от нуля вплоть до порядка $\tau = 2p$, а коэффициенты более высоких

порядков будут равны нулю. В [13, с. 49] отмечается, что для скользящих средних, наиболее часто применяемых на практике, ρ_1 будет положительным и может принимать довольно высокие значения.

Таким образом, производные ряды, будучи более гладкими по сравнению с исходными, могут содержать систематические колебания, вызванные лишь усреднением случайных составляющих. Этот вывод, ставший в свое время сенсацией, получил название *эффекта Слущкого—Юла*, по имени ученых, впервые обративших внимание на этот факт¹.

С помощью скользящих средних Е.Е. Слущкий получил подобие цикла, напоминавшее типичное поведение некоторых экономических рядов. Взяв последние цифры облигаций из таблиц выигрышного займа, он в своей работе сумел показать, что «сложение случайных причин порождает волнообразные ряды, имеющие тенденцию на протяжении большего или меньшего числа волн имитировать гармонические ряды, сложенные из небольшого числа синусоид».

2.5. Использование скользящих средних в техническом анализе

Весь комплекс скользящих средних (простых, взвешенных, экспоненциальных, см. гл. 5) представляет собой важное аналитическое средство технического анализа. К их бесспорным достоинствам можно отнести то, что они легко вычисляются, объясняются и интерпретируются.

До середины 70-х годов в финансовом анализе господствовали две традиционные школы, жестко отстаивающие свои преимущества при исследовании фондового рынка: это представители фундаментального анализа и технического. Впоследствии это противостояние сменилось сочетанием двух подходов, а также появлением новых теорий и методов.

Сторонники фундаментального анализа уделяли пристальное внимание изучению финансовой отчетности, считая, что поведение курса ценных бумаг есть следствие финансово-хозяйственной деятельности корпорации, состояния исследуемой отрасли и экономики в целом.

Представители технического анализа делали упор на изучение внутренней информации о фондовом рынке, а именно, на ценовых показателях и информации об объемах сделок. Они считали, что действие всех внешних факторов, оказывающих влияние на рынок, сказывается в перечисленных показателях.

¹ Е.Е. Слущкий. Сложение случайных величин как источник циклических процессов // Вопросы конъюнктуры. 1927. Т. III. Вып. 1; Yule G. Udny (1927) On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers // Phil. Trans. V.A. — P. 226, 267.

На биржевом графике по вертикальной шкале отражаются ценовые изменения, а по горизонтальной — изменение временного параметра. Для каждого дня торгов традиционно могут быть отмечены: максимальная и минимальная цена, цена открытия и цена последней сделки (цена закрытия). Часто в нижней части графика показывают объемы торгов, используя для этого вторую шкалу ординат.

Скользящие средние позволяют обнаруживать и отслеживать тренды в ценовых показателях, определять направления существующих трендов, а также выявлять происходящие в них изменения.

В техническом анализе, как правило, отказываются от ранее рассмотренных центрированных скользящих средних. Хотя такой вид сглаживания статистически более обоснован, он имеет очевидный недостаток: сигналы, подаваемые скользящими средними, будут поступать с запаздыванием. Поэтому в техническом анализе результат усреднения часто относится к последней точке активного участка.

Самый популярный подход к использованию процедур сглаживания состоит в сравнении взаимного расположения графиков скользящей средней, построенной по цене закрытия, и собственно цены закрытия. (Аналогичный подход может быть применен и к остальным ценовым характеристикам.) Если кривая скользящей средней находится ниже кривой исходных данных, т.е. на рынке преобладает тенденция к повышению, то ценовой тренд называют «бычьим». Если же кривая скользящей средней расположена над ценовой кривой, т.е. курс имеет тенденцию к понижению, то ценовой тренд называют «медвежьим».

Изменение направленности тренда происходит при пересечении графиков скользящих средних и исходных данных. Сигнал к продаже поступает после того, как цена опускается ниже кривой своей скользящей средней, сигнал к покупке — после того, как цена поднимается выше своей кривой скользящей средней. Такой подход позволяет принимать решения таким образом, чтобы покупать недалеко от «дна» цены, а продавать вблизи от «вершины».

Отметим, что в практике технического анализа широко используются скользящие средние не только для ценовых характеристик, но и для многих индикаторов. Интерпретация скользящих средних для индикаторов сходна с ранее описанным подходом. Если индикатор пересекает кривую его скользящей средней снизу вверх, то это подтверждает движение индикатора вверх, и наоборот. В практических руководствах рекомендуется использовать скользящие средние применительно к таким индикаторам, как Momentum, Stochastics, Rate of Change и др.¹

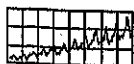
¹ Ратай И.С. Работа с пакетом WINDOW ON WALLSSTREET. — М.: Торцентр, 1997.

Некоторые индикаторы, например, краткосрочный Stochastics, часто изменяются ошибочно, подавая ложные сигналы. Построив кривую скользящей средней для данного индикатора, можно увидеть общую тенденцию вместо его дневных колебаний.

Для анализа рыночных трендов часто используют комбинации скользящих средних. Например, если короткая скользящая средняя находится выше более длинной, то это подтверждает «бычий» тренд.

Во время сильного верхнего («бычьего») тренда более короткая скользящая средняя поднимается быстрее, чем длинная. Спрэд между двумя средними будет расширяться. Когда спрэд между ними начинает сужаться, это обычно служит предупреждением о том, что тренд теряет «скорость». Скользящие средние — это индикаторы, следующие за трендом, поэтому они лучше работают при наличии тренда¹.

Важная практическая задача состоит в нахождении оптимальных сглаживающих окон. Чем больше длина интервала сглаживания, тем больше запаздывают подаваемые скользящей средней сигналы, тем менее чувствительной становится эта кривая. В свою очередь, использование скользящей средней очень маленького порядка будет приводить к частому появлению ложных сигналов. Очевидно, что целесообразно выбирать длину интервала сглаживания в зависимости от того, на какой промежуток времени требуется получить прогноз поведения ценовых характеристик. Отметим, что «размер окон» для наиболее часто используемых на конкретных рынках скользящих средних (краткосрочных, среднесрочных и долгосрочных), как правило, получают эмпирическим путем.



Практикум к главе 2

Контрольные вопросы

1. Объясните назначение скользящих средних. Влияние каких компонент временного ряда устраняется с их помощью?
2. Поясните, когда целесообразно использовать простые скользящие средние, а для каких временных рядов предпочтительнее применение взвешенных.
3. Приведите алгоритм расчета простых скользящих средних.
4. В чем отличие алгоритма расчета взвешенных скользящих средних от простых?
5. Сколько значений теряется при использовании скользящей средней с длиной интервала сглаживания $l = 2p + 1$? Какие приемы восстановления потерянных уровней после реализации процедур сглаживания используются на практике?

¹ Жаколюк Ю.В. Внутрдневная торговля на рынке ФОРЕКС. — СПб.: Питер, 2000.

6. В чем суть эффекта Служцкого—Юла?
7. Поясните, каким образом используются процедуры сглаживания с помощью скользящих средних в техническом анализе.
8. Как рассчитываются простые скользящие средние при четной длине интервала сглаживания?
9. Каким образом определены весовые коэффициенты, приведенные в табл. 2.1?
10. Объясните, почему при реализации взвешенных скользящих средних весовые коэффициенты будут неизменными при сглаживании по полиному как второго порядка, так и третьего.

Упражнения

- 2.1. Выведите весовые коэффициенты для сглаживания временного ряда:
 - а) по полиному третьего порядка при длине интервала сглаживания $l = 5$;
 - б) по полиному четвертого порядка при длине интервала сглаживания $l = 7$.
- 2.2. Докажите, что весовые коэффициенты при сглаживании временного ряда по полиному четвертого и пятого порядков будут одинаковыми.
- 2.3. В таблице представлены данные об изменении урожайности озимой пшеницы за 16 лет. Требуется:
 - а) провести сглаживание временного ряда, используя:
 - пятилетнюю простую скользящую среднюю;
 - пятилетнюю взвешенную скользящую среднюю (выравнивание проводить по полиному второго порядка).
 - б) восстановить потерянные уровни.

Порядковый номер года	1	2	3	4	5	6	7	8
Урожайность, ц/га	15,3	17,2	18,1	17,3	18,9	17,6	20,9	16,9
Порядковый номер года	9	10	11	12	13	14	15	16
Урожайность, ц/га	17,8	18,9	19,2	18,5	21,6	20,1	18,9	19,7

В заданиях 4, 5 выберите правильный вариант ответа.

- 2.4. При сглаживании временного ряда с помощью 11-членной скользящей средней теряются:
 - а) первые и последние 5 уровней временного ряда;
 - б) первые и последние 11 уровней временного ряда;
 - в) только первые 5 уровней;
 - г) только первые 11 уровней.
- 2.5. Более гладкий ряд получится при использовании:
 - а) 5-членной скользящей средней;
 - б) 11-членной скользящей средней;
 - в) 9-членной скользящей средней;
 - г) 21-членной скользящей средней.



АНАЛИТИЧЕСКОЕ ВЫРАВНИВАНИЕ ВРЕМЕННЫХ РЯДОВ. МОДЕЛИРОВАНИЕ ТЕНДЕНЦИИ РАЗВИТИЯ

3.1. Проверка гипотезы о существовании тренда

Решение любой задачи по анализу и прогнозированию временных рядов начинается с построения графика исследуемого показателя, тем более что современные программные средства предоставляют для этого пользователю большие возможности. При этом не всегда четко прослеживается присутствие тренда во временном ряду. В этих случаях прежде чем перейти к определению тенденции и выделению тренда, нужно выяснить, существует ли вообще тенденция в исследуемом процессе.

Основные подходы к решению этой задачи основаны на статистической проверке гипотез. Критерии выявления компонент ряда основаны на проверке гипотезы о случайности ряда, т.е. по существу на статистической проверке гипотезы

$$H_0 : My(t) = a = \text{const.} \quad (3.1)$$

Рассмотрим критерий серий [1], часто используемый на практике для проверки «наличия — отсутствия» тренда. Он имеет две модификации:

- критерий серий, основанный на медиане выборки;
- критерий «восходящих и нисходящих» серий.

Применение критерия серий, основанного на медиане выборки, может быть представлено в виде следующей последовательности шагов.

1. Из исходного ряда с уровнями $y_1, y_2, \dots, y_n$ образуется ранжированный (вариационный) ряд $y_1^1, y_2^1, \dots, y_n^1$, где y_1^1 — наименьшее значение из уровней исходного ряда $y_1, y_2, \dots, y_n$.

2. Определяется медиана (Me) этого вариационного ряда. В случае нечетного значения длины ряда n ($n=2m+1$) $Me = y_{m+1}^1$, в противном случае ($n=2m$) $Me = (y_m^1 + y_{m+1}^1) : 2$.

3. Образуется последовательность δ_t из плюсов и минусов по следующему правилу:

$$\delta_t = \begin{cases} +, & \text{если } y_t > Me, \quad t=1, 2, \dots, n; \\ -, & \text{если } y_t < Me, \quad t=1, 2, \dots, n. \end{cases} \quad (3.2)$$

Если значение уровня исходного ряда y , равно медиане, то это значение пропускается. Очевидно, что общее число знаков «+» и «-» заранее не известно. Индекс i может принимать значения 1, 2, ..., k , где $k \leq n$.

4. Подсчитывается $v(n)$ — число серий в совокупности δ , где под серией понимается последовательность подряд идущих плюсов или минусов. Один плюс или один минус тоже будет считаться серией.

Определяется $\tau_{\max}(n)$ — протяженность самой длинной серии.

5. Проверка гипотезы основывается на том, что при условии случайности ряда (при отсутствии систематической составляющей) протяженность самой длинной серии не должна быть слишком большой, а общее число серий — слишком маленьким. Поэтому для того чтобы не была отвергнута гипотеза о случайности исходного ряда (об отсутствии систематической составляющей), должны выполняться следующие неравенства:

$$\begin{aligned} v(n) &> \left\lceil \frac{1}{2}(n+1-1,96\sqrt{n-1}) \right\rceil; \\ \tau_{\max}(n) &< [1,43 \ln(n+1)], \end{aligned} \quad (3.3)$$

где n — длина временного ряда;

$v(n)$ — число серий;

$\tau_{\max}(n)$ — число подряд идущих плюсов или минусов в самой длинной серии.

Если хотя бы одно из неравенств (3.3) нарушается, то гипотеза (3.1) отвергается с вероятностью ошибки α , заключенной между 0,05 и 0,0975 (следовательно, подтверждается наличие зависящей от времени неслучайной составляющей).

Квадратные скобки в правой части неравенства означают целую часть числа. Напомним, что целая часть числа A , т.е. $[A]$ — это целое число, ближайшее к A и не превосходящее его.

Рассмотрим теперь применение критерия «восходящих и нисходящих» серий для проверки гипотезы (3.1).

1. На первом шаге, как и в предыдущем варианте критерия, образуется последовательность знаков — плюсов и минусов, однако в данном случае правила образования этой последовательности иные.

Для временного ряда с уровнями $y_1, y_2, \dots, y_n$ определяется вспомогательная последовательность, исходя из следующих условий:

$$\delta_i = \begin{cases} +, & \text{если } y_{i+1} - y_i > 0 \text{ для } i=1, 2, \dots, n-1; \\ -, & \text{если } y_{i+1} - y_i < 0 \text{ для } i=1, 2, \dots, n-1. \end{cases} \quad (3.4)$$

В случае, когда последующее наблюдение окажется равным предыдущему, учитывается только одно наблюдение. Таким образом, элементы этой последовательности принимают значение «+», если последующее значение уровня ряда больше предыдущего, и «-» — если меньше. Общее число знаков «+» и «-» заранее не известно. Индекс i может принимать значения 1, 2, ..., k , где $k \leq n - 1$.

2. Подсчитывается общее число серий $v(n)$ и протяженность самой длинной серии $\tau_{\max}(n)$ аналогично тому, как это делалось в предыдущем варианте критерия. Очевидно, что при этом каждая серия, состоящая из плюсов, соответствует возрастанию уровней ряда («восходящая» серия), а последовательность минусов — их убыванию («нисходящая» серия).

3. Для того чтобы не была отвергнута гипотеза (3.1), должны выполняться следующие неравенства (при уровне значимости α , заключенном между 0,05 и 0,0975):

$$v(n) > \left[\frac{1}{3}(2n-1) - 1,96 \sqrt{\frac{16n-29}{90}} \right]; \quad (3.5)$$

$$\tau_{\max}(n) < \tau_0(n),$$

где $\tau_0(n)$ — табличное значение, зависящее от n — длины временного ряда (табл. 3.1).

Таблица 3.1. Значение $\tau_0(n)$ для критерия «восходящих и нисходящих» серий

n	$n \leq 26$	$26 < n \leq 153$	$153 < n \leq 1170$
$\tau_0(n)$	5	6	7

Если хотя бы одно из неравенств (3.5) нарушается, то нулевая гипотеза (3.1) отвергается (следовательно, подтверждается наличие зависящей от времени неслучайной составляющей).

Кроме рассмотренных подходов в литературе описаны и другие критерии, отличающиеся друг от друга мощностью, сложностью математического аппарата, например, критерий квадратов последовательных разностей (критерий Аббе) [1], метод Фостера — Стюарта [29], метод проверки разностей средних уровней [31] и др.

Пример 3.1

В табл. 3.2 представлены годовые данные об урожайности зерновых культур.

Исследователя может интересовать вопрос: носит ли изменение урожайности случайный характер, можно ли считать неизменной среднюю урожайность на всем анализируемом временном интервале. В то же время характер изменения уровней ряда может породить альтернативное предположение: динамика временного ряда характеризуется наличием систематической составляющей, зависящей от t , — в данном случае, тенденцией к увеличению урожайности. Для проверки нулевой гипотезы (3.1) воспользуемся критерием «восходящих и нисходящих» серий.

Таблица 3.2. Урожайность зерновых культур, ц/га

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
y_t	6,7	7,3	7,6	7,9	7,4	8,6	7,8	7,7	7,9	8,2	8,4	9,1	8,3	8,7	8,9	9,1	9,5	10,4	10,5	10,2	9,3

Решение

Процесс формирования серий показан в табл. 3.3. Во второй строке этой таблицы в соответствии с (3.4) указан «+», если последующее значение уровня временного ряда больше предыдущего, «—» — если меньше.

Таблица 3.3. Формирование серий

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
δ_i	+	+	+	—	+	—	—	+	+	+	+	—	+	+	+	+	+	+	—	—

Анализ полученной последовательности знаков позволил установить:

- число серий $v(21) = 8$,
- протяженность самой длинной серии $\tau_{\max}(21) = 6$.

Табличное значение (см. табл. 3.1) $\tau_0(21) = 6$.

В соответствии с (3.5) делаем проверку. Для этого сначала определим значение для правой части первого неравенства:

$$\left[\frac{1}{3}(2 \cdot 21 - 1) - 1,96 \sqrt{\frac{16 \cdot 21 - 29}{90}} \right] = 10. \quad (3.6)$$

Тогда проверка выполнения условий (3.5) показывает, что оба неравенства не выполняются. Следовательно, нулевая гипотеза (3.1) отвергается, динамика временного ряда характеризуется наличием систематической составляющей — в изменении урожайности присутствует тенденция. Очевидно, что этот же вывод можно было получить, опираясь лишь на второе неравенство в (3.5), не проводя вычислений (3.6).

3.2. Применение моделей кривых роста для анализа и прогнозирования основной тенденции развития

На практике для описания тенденции развития явления широко используются *модели кривых роста*, представляющие собой различные функции времени $y = f(t)$. При таком подходе изменение исследуемого показателя связывают лишь с течением времени; считается, что влияние других факторов несущественно или косвенно сказывается через фактор времени.

Правильно выбранная модель кривой роста должна соответствовать характеру изменения тенденции исследуемого явления. Кривая роста позволяет получить выровненные или теоретические значения уровней динамического ряда. Это те уровни, которые наблюдались бы в случае полного совпадения динамики явления с кривой.

Прогнозирование на основе модели кривой роста базируется на экстраполяции, т.е. на продлении в будущее тенденции, наблюдавшейся в прошлом. При этом предполагается, что во временном ряду присутствует тренд, характер развития показателя обладает свойством инерционности, сложившаяся тенденция не должна претерпевать существенных изменений в течение периода упреждения.

Процедура разработки прогноза с использованием кривых роста включает в себя следующие этапы:

- выбор одной или нескольких кривых, форма которых соответствует характеру изменения временного ряда;
- оценку параметров выбранных кривых;
- проверку адекватности выбранных кривых прогнозируемому процессу, оценку точности моделей и окончательный выбор кривой роста;
- расчет точечного и интервального прогнозов.

В настоящее время в литературе описано несколько десятков кривых роста. Эти модели условно могут быть разделены на три класса в зависимости от того, какой тип динамики развития они хорошо описывают.

К I классу относятся функции, используемые для описания процессов с монотонным характером тенденции развития и отсутствием пределов роста. Эти условия справедливы для многих экономических показателей, например, для большинства показателей промышленного производства в натуральном выражении.

Ко II классу относятся кривые, описывающие процесс, который имеет предел роста в исследуемом периоде. С такими процессами часто сталкиваются в демографии, при изучении потребностей в товарах и услугах (в расчете на душу населения), при исследовании

довании эффективности использования ресурсов и т.д. Примерами показателей, для которых могут быть указаны пределы роста, могут служить среднедушевое потребление определенных продуктов питания, расход удобрений на единицу площади и т.п.

Функции, относящиеся ко II классу, называются *кривыми насыщения*. Если кривые насыщения имеют точки перегиба, то они относятся к III классу кривых роста — к *S-образным кривым*.

Эти кривые описывают два последовательных лавинообразных процесса (когда прирост зависит от уже достигнутого уровня): один — с ускорением развития, другой — с замедлением.

S-образные кривые находят применение в демографических исследованиях, в страховых расчетах, при решении задач прогнозирования научно-технического прогресса, при определении спроса на новый вид продукции.

При использовании подхода, основанного на кривых роста, уровни временного ряда представляются следующей моделью:

$$y_t = f(t; \Theta) + \varepsilon_t, \quad t = 1, 2, \dots, n.$$

При этом предполагается, что вид функции $f(t; \Theta)$ известен, значения параметров $\Theta = (\theta_0, \theta_1, \dots, \theta_k)^T$ неизвестны и подлежат оцениванию, а возмущения ε_t удовлетворяют условиям (1.30) (см. § 1.4).

В подавляющем большинстве случаев расчет оценок параметров трендовых моделей $\hat{\Theta}$ осуществляется с помощью метода наименьших квадратов в рамках регрессионных моделей, в которых в качестве значений зависимой переменной выступают фактические уровни ряда y_t , а в роли независимой переменной — время t .

Для нелинейных трендовых моделей применяются процедуры линеаризации. В том случае, если уравнение тренда преобразовать к линейному виду невозможно, применяют нелинейные методы оценивания коэффициентов.

Вопрос о выборе кривой — основной при выравнивании ряда.

Существует несколько подходов к решению этой задачи, однако все они предполагают знакомство с основными свойствами используемых кривых роста.

Поэтому остановимся на характеристике отдельных типов кривых, наиболее часто применяемых на практике.

Среди кривых роста I класса прежде всего следует выделить класс полиномов:

$$\hat{y}_t = a_0 + a_1 t + a_2 t^2 + \dots + a_p t^p, \quad (3.7)$$

где a_i ($i = 0, 1, \dots, p$) — параметры многочлена,
 t — независимая переменная (время), $t = 1, 2, \dots, n$.

$$\hat{\Theta} = \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_p \end{pmatrix} = (X^T X)^{-1} X^T Y. \quad (3.10)$$

В этом выражении

- матрица X (по аналогии с регрессионным анализом — матрица наблюдаемых значений объясняющих переменных) имеет вид:

$$X_{[n \times (p+1)]} = \begin{pmatrix} 1 & 1 & 1^2 & \dots & 1^p \\ 1 & 2 & 2^2 & \dots & 2^p \\ \dots & \dots & \dots & \dots & \dots \\ 1 & n & n^2 & \dots & n^p \end{pmatrix};$$

- вектор-столбец фактических, наблюдаемых значений зависимой переменной

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}.$$

Системы для оценивания полиномов невысоких степеней выглядят намного проще. Например, нормальные уравнения для оценивания параметров линейного тренда $\hat{y}_t = a_0 + a_1 t$ имеют вид:

$$\begin{cases} \sum_{t=1}^n y_t = a_0 n + a_1 \sum_{t=1}^n t; \\ \sum_{t=1}^n y_t t = a_0 \sum_{t=1}^n t + a_1 \sum_{t=1}^n t^2. \end{cases} \quad (3.11)$$

Решение этой системы относительно искомых параметров дает следующие выражения:

$$a_1 = \frac{n \sum_{t=1}^n y_t t - \sum_{t=1}^n y_t \cdot \sum_{t=1}^n t}{n \sum_{t=1}^n t^2 - \left(\sum_{t=1}^n t \right)^2}; \quad (3.12)$$

$$a_0 = \frac{\sum_{t=1}^n y_t}{n} - a_1 \frac{\sum_{t=1}^n t}{n}.$$

Для параболического тренда (полинома второго порядка) $\hat{y}_t = a_0 + a_1 t + a_2 t^2$ получим аналогичную систему нормальных уравнений:

$$\begin{cases} \sum_{t=1}^n y_t = a_0 n + a_1 \sum_{t=1}^n t + a_2 \sum_{t=1}^n t^2; \\ \sum_{t=1}^n y_t t = a_0 \sum_{t=1}^n t + a_1 \sum_{t=1}^n t^2 + a_2 \sum_{t=1}^n t^3; \\ \sum_{t=1}^n y_t t^2 = a_0 \sum_{t=1}^n t^2 + a_1 \sum_{t=1}^n t^3 + a_2 \sum_{t=1}^n t^4. \end{cases} \quad (3.13)$$

Эта система содержит три уравнения, позволяющих найти оценки трех неизвестных коэффициентов a_0, a_1, a_2 .

Составление нормальных уравнений можно упростить, воспользовавшись тем, что величины $\sum t, \sum t^2, \dots$ не зависят от конкретных уровней динамического ряда. Эти суммы являются функциями только числа членов в динамическом ряду. Для них получены следующие формулы:

$$\begin{aligned} \sum_{t=1}^n t &= \frac{n(n+1)}{2}; & \sum_{t=1}^n t^2 &= \frac{n(n+1)(2n+1)}{6}; \\ \sum_{t=1}^n t^3 &= \frac{n^2(n+1)^2}{4}; & \sum_{t=1}^n t^4 &= \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30}. \end{aligned}$$

Другой подход к упрощению расчетов заключается в переносе начала координат в середину ряда динамики. Это позволяет упростить сами нормальные уравнения, а также уменьшить абсолютные значения величин, участвующих в расчете. Если до переноса начала координат t было равно 1, 2, 3, ..., то после переноса:

- для четного числа членов ряда $t = \dots, -5; -3; -1; 1; 3; 5; \dots$;
- для нечетного числа членов ряда $t = \dots, -3; -2; -1; 0; 1; 2; 3; \dots$

В формулах (3.14)–(3.17) суммирование проводится по t , полученному после переноса начала координат. При этом учитывается, что в этом случае $\sum t^k = 0$, где k — нечетное число. Такой подход существенно упрощает систему (3.9).

В частности, система нормальных уравнений для прямой (3.11) за счет того, что $\sum t = 0$, примет вид:

$$\begin{cases} \sum y_t = a_0 n; \\ \sum y_t t = a_1 \sum t^2. \end{cases} \quad (3.14)$$

Для параболы система (3.13) так же существенно упрощается, так как после переноса начала координат в середину ряда динамики $\sum t = 0$ и $\sum t^3 = 0$:

$$\begin{cases} \sum y_t = a_0 n + a_2 \sum t^2; \\ \sum y_t t = a_1 \sum t^2; \\ \sum y_t t^2 = a_0 \sum t^2 + a_2 \sum t^4. \end{cases} \quad (3.15)$$

В этом случае оценки параметров соответствующих полиномов имеют вид:

$$a_0 = \frac{\sum y_t}{n};$$

для прямой

$$a_1 = \frac{\sum y_t t}{\sum t^2}; \quad (3.16)$$

$$a_0 = \frac{\sum y_t}{n} - \frac{\sum t^2}{n} \left[\frac{n \sum y_t t^2 - \sum t^2 \sum y_t}{n \sum t^4 - (\sum t^2)^2} \right];$$

для параболы

$$a_1 = \frac{\sum y_t t}{\sum t^2}; \quad (3.17)$$

$$a_2 = \frac{n \sum y_t t^2 - \sum t^2 \sum y_t}{n \sum t^4 - (\sum t^2)^2}.$$

Для класса экспоненциальных кривых, в отличие от полиномов, характерна зависимость приростов от значений самой функции. Эти кривые хорошо описывают процессы, имеющие «лавинообразный» характер, когда прирост зависит от достигнутого уровня функции.

Простая экспоненциальная (показательная) кривая имеет вид:

$$y_t = ab^t. \quad (3.18)$$

Если $b > 1$, то кривая растет вместе с ростом t , и падает, если $b < 1$.

Параметр a характеризует начальные условия развития, а параметр b — постоянный темп роста.

Действительно, темп роста равен $T_t = \frac{y_t}{y_{t-1}} \cdot 100\%$.

В данном случае

$$T_t = \frac{ab^t}{ab^t - 1} \cdot 100\% = b \cdot 100\% = \text{const.}$$

Соответственно и темпы прироста постоянны:

$$K_t = T_t - 100\% = \text{const.}$$

Можно показать, что логарифм ординаты этой функции линейно зависит от t ; для этого прологарифмируем выражение (3.18):

$$\log y_t = \log a + t \log b.$$

Пусть

$$\log a = A; \quad \log b = B.$$

Тогда

$$\log y_t = A + tB.$$

Теперь, если тенденция ряда описывается с помощью модели $\hat{y}_t = ab^t$, то для оценивания неизвестных параметров можем использовать систему нормальных уравнений для прямой (3.11).

Иначе говоря, нормальные уравнения строятся исходя из минимизации:

$$\sum_{t=1}^n (\log y_t - \log \hat{y}_t)^2 \rightarrow \min.$$

Соответственно в нормальных уравнениях вместо фактических уровней выступают их логарифмы:

$$\begin{aligned} \sum_{t=1}^n \log y_t &= nA + B \sum_{t=1}^n t; \\ \sum_{t=1}^n (\log y_t) t &= A \sum_{t=1}^n t + B \sum_{t=1}^n t^2. \end{aligned} \tag{3.19}$$

Найдем неизвестные параметры A и B . Зная значения $A = \log a$ и $B = \log b$, определим значения a и b с помощью потенцирования и получим показательную функцию, служащую для выравнивания ряда.

Такой подход к оцениванию неизвестных параметров привлекает своей универсальностью.

Все рассмотренные типы кривых используются для описания процессов без «насыщения», с монотонно возрастающим или убывающим характером тенденции.

Когда процесс характеризуется «насыщением», его следует описывать при помощи кривой, имеющей отличную от нуля асимптоту. Примером такой кривой может служить модифицированная экспонента

$$\hat{y}_t = k + ab^t, \quad (3.20)$$

где $y = k$ — горизонтальная асимптота.

Если параметр a отрицателен, то асимптота находится выше кривой, если a положителен, то ниже. При решении экономических задач чаще всего приходится иметь дело с кривой, у которой $a < 0$, $b < 1$. В этом случае рост уровней происходит с замедлением и стремится к некоторому пределу.

При решении экономических задач часто можно определить значение асимптоты исходя из свойств прогнозируемого процесса (например, коэффициент использования оборудования не может превышать 1). Иногда значение асимптоты задается экспертным путем. Например, главный инженер предприятия указывает, что производственные мощности не позволят наращивать объемы производства выше определенного уровня k . Этот уровень является оценкой значения асимптоты k при прогнозировании производства продукции.

В этих случаях другие параметры кривой определяют с помощью МНК после приведения уравнения к линейному виду. При этом для оценивания оставшихся параметров можно использовать систему нормальных уравнений (3.19), заменив $\log(y_t)$ на $\log(y_t - k)$. Таким образом, модифицированная экспонента хорошо описывает процесс, на развитие которого воздействует ограничивающий фактор, причем влияние этого воздействия растет вместе с ростом достигнутого уровня.

Если воздействие ограничивающего фактора начинает сказываться только после определенного момента (точки перегиба), до которого процесс развивался по некоторому экспоненциальному закону, то для выравнивания используют S-образные кривые. Наиболее известны из них — кривая Гомперца и логистическая кривая (кривая Перла—Рида).

Уравнение кривой Гомперца имеет вид:

$$\hat{y}_t = k + ab^t. \quad (3.21)$$

Кривая несимметрична.

Если $\log a < 0$, кривая имеет S-образный вид, при этом асимптота, равная k , проходит выше кривой. Если $\log a > 0$, асимптота, равная k , лежит ниже кривой, а сама кривая изменяется моно-

тонно. При $b < 1$ она монотонно убывает; а при $b > 1$ — монотонно возрастает.

Для решения экономических задач наибольший интерес представляет вариант этой кривой, когда $\log a < 0$ и $b < 1$ (рис. 3.1).

Уравнение логистической кривой получается путем замены в модифицированной экспоненте y , обратной величиной $\frac{1}{y_t}$:

$$\frac{1}{\hat{y}_t} = k + ab^t. \quad (3.22)$$

Используется и другая форма записи уравнения логистической кривой:

$$\hat{y}_t = \frac{k}{1 + be^{-at}}.$$

При $t \rightarrow -\infty$ ордината стремится к нулю, а при $t \rightarrow \infty$ — к асимптоте, равной значению параметра k . Кривая симметрична относительно точки перегиба с координатами $t = \ln b : a; y_t = k : 2$.

Как видно из графика (см. рис. 3.1), логистическая функция возрастает сначала ускоренным темпом, затем темп роста замедляется и наконец рост почти полностью прекращается, о чем свидетельствует то, что кривая асимптотически приближается к некоторой прямой, параллельной оси абсцисс.

С помощью этой функции хорошо описывается развитие новой отрасли (нового производства). Сначала технические методы производства еще не достаточно разработаны, издержки производства высоки и спрос на рынке на данный товар еще очень мал, поэтому производство развивается медленно. В дальнейшем, благодаря усовершенствованию технических методов изготовления, переходу к массовому производству и увеличению емкости рынка для данного товара производство растет быстрее. Затем наступает период насыщения рынка, рост производства все более замедляется и наконец почти прекращается. Наступает стабилизация производства на определенном уровне. Однако выявленные закономерности развития следует обобщать с определенной осторожностью, причем для коротких периодов. Выявленная тенденция развития производства может быть нарушена, например, вследствие технического переворота в данной отрасли или связанной с нею.

Мы рассмотрели наиболее часто используемые виды кривых роста. Выявленные особенности и свойства этих кривых могут существенно помочь при решении задачи выбора типа кривой.

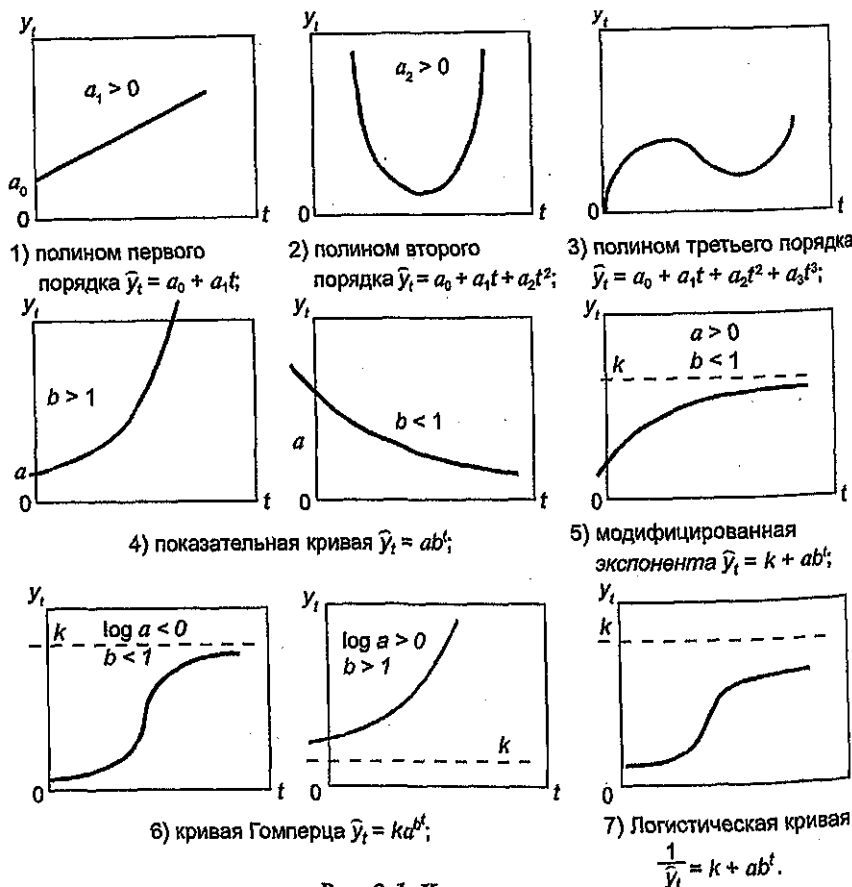


Рис. 3.1. Кривые роста

Пример 3.2

В табл. 3.4 представлены данные об остатках вкладов населения в банках за 15 месяцев. Остатки вкладов указаны на начало каждого месяца.

Таблица 3.4. Остатки вкладов населения в банках, млрд руб.

Порядковый номер месяца	y_t	Порядковый номер месяца	y_t	Порядковый номер месяца	y_t
1	14717	6	23342	11	40524
2	16642	7	28317	12	45416
3	18504	8	30624	13	50857
4	20376	9	33408	14	56024
5	21321	10	36505	15	59381

Необходимо рассчитать прогнозные значения остатков вкладов населения в банках на начало 16-го месяца, исходя из предположения, что тенденция ряда может быть описана:

- а) линейной моделью $\hat{y}_t = a_0 + a_1 t$;
- б) параболической моделью $\hat{y}_t = a_0 + a_1 t + a_2 t^2$;
- в) показательной моделью $\hat{y}_t = ab^t$.

Решение

а) Для расчета коэффициентов линейного тренда воспользуемся выражениями, полученными из системы нормальных уравнений после переноса начала координат в середину ряда (3.16). Так как число уровней ряда динамики нечетное ($n = 15$), то центральный уровень (восьмой) принимается за начало отсчета, ему соответствует $t = 0$. Вышестоящие уровни нумеруются с шагом -1 , нижестоящие — с шагом $+1$ (гр. 3 табл. 3.5, в которой представлены необходимые вспомогательные вычисления).

Таблица 3.5. Расчет параметров линейной и параболической моделей

№	y_t	t	yt	t^2	yt^2	t^4
1	2	3	4	5	6	7
1	14717	-7	-103019	49	721133	2401
2	16642	-6	-99852	36	599112	1296
3	18504	-5	-92520	25	462112	625
4	20376	-4	-81504	16	326016	256
5	21321	-3	-63963	9	191889	81
6	23342	-2	-46684	4	93368	16
7	28317	-1	-28317	1	28317	1
8	30624	0	0	0	0	0
9	33408	1	33408	1	33408	1
10	36505	2	73010	4	146020	16
11	40524	3	121572	9	364716	81
12	45416	4	181664	16	726656	256
13	50857	5	254285	25	1271425	625
14	56024	6	336144	36	2016864	1296
15	59381	7	415667	49	2909669	2401
Σ	495958		899891	280	9891193	9352

В соответствии с (3.16)

$$a_0 = \frac{\sum y_t}{n} = \frac{495958}{15} = 33\,063,866;$$

$$a_1 = \frac{\sum y_t t}{\sum t^2} = \frac{899\,891}{280} = 3213,896.$$

Следовательно, уравнение линейного тренда имеет вид:

$$\hat{y}_t = 33\,063,866 + 3213,896t.$$

Согласно этой модели оценка среднего уровня ряда при $t=0$ равна 33063,9 млрд руб., а среднемесячный прирост остатков вкладов населения составляет 3213,9 млрд руб.

Для прогнозирования на базе полученной модели на одну точку вперед необходимо в нее подставить соответствующее значение временного параметра, т.е. $t=8$. (Если бы оценки коэффициентов модели были получены без переноса начала координат в середину ряда, то следовало бы подставить в модель значение временного параметра $t=16$).

Определим прогнозное значение:

$$\hat{y}_8 = 33\,063,866 + 3213,896 \cdot 8; \quad \hat{y}_8 = 58\,775 \text{ млрд руб.}$$

б) Для расчета коэффициентов параболического тренда также воспользуемся выражениями, полученными из системы нормальных уравнений после переноса начала координат в середину ряда (3.17).

Промежуточные вычисления представлены в табл. 3.5:

$$\begin{aligned} a_1 &= \frac{899\,891}{280} = 3213,896; \\ a_2 &= \frac{15 \cdot 9\,891\,193 - 280 \cdot 495\,958}{15 \cdot 9352 - (280)^2} = 153,517; \\ a_0 &= 33\,063,866 - \frac{280}{15} \cdot 153,517 = 30\,198,16. \end{aligned}$$

Следовательно, уравнение параболического тренда примет вид:

$$\hat{y}_t = 30\,198,16 + 3213,896t + 153,517t^2.$$

Для определения прогнозного значения показателя надо подставить в полученную модель соответствующее значение временного параметра ($t=8$):

$$\begin{aligned} \hat{y}_8 &= 30\,198,16 + 3213,896 \cdot 8 + 153,517 \cdot 8^2; \\ \hat{y}_8 &= 65\,734 \text{ млрд руб.} \end{aligned}$$

в) Для определения параметров тренда, описываемого показательной функцией, воспользуемся (3.16), (3.19).

В табл. 3.6 представлены необходимые вспомогательные вычисления.

Таблица 3.6. Расчет параметров показательной модели

№	y_t	t	t^2	$\ln y_t$	$\ln(y_t) \cdot t$
1	14717	-7	49	9,5968	-67,1773
2	16642	-6	36	9,7197	-58,3181
3	18504	-5	25	9,8257	-49,1287
4	20376	-4	16	9,9221	-39,6885
5	21321	-3	9	9,9674	-29,9023
6	23342	-2	4	10,0580	-20,116
7	28317	-1	1	10,2512	-10,2512
8	30624	0	0	10,3295	0
9	33408	1	1	10,4166	10,4166
10	36505	2	4	10,5052	21,0104
11	40524	3	9	10,6097	31,829
12	45416	4	16	10,7236	42,8945
13	50857	5	25	10,8368	54,1839
14	56024	6	36	10,9335	65,6012
15	59381	7	49	10,9917	76,9421
Σ	495958		280	154,6876	28,2954

Тогда можно рассчитать

$$\ln a = \frac{154,6876}{15} = 10,3125; \quad \ln b = \frac{28,2954}{280} = 0,1011.$$

Проведя потенцирование, получаем: $a = 30\,106,61$; $b = 1,11$.

Следовательно, уравнение тренда примет вид

$$\hat{y}_t = 30\,106,61 \cdot 1,11^t.$$

Согласно этой модели среднемесячный темп роста остатков вкладов населения составлял 111%. В точке, принятой за начало отсчета ($t = 0$), значение тренда равно 30 106,61 млрд руб. Для определения прогнозного значения остатков вклада населения в банках на один месяц вперед подставляем в полученную модель значение $t = 8$:

$$\begin{aligned} \hat{y}_8 &= 30\,106,61 \cdot 1,11^8; \\ \hat{y}_8 &= 69\,382 \text{ млрд руб.} \end{aligned}$$

На рис. 3.2 изображен график исходного временного ряда и выровненные значения уровней, полученные на основе двух трендовых моделей: линейной и параболической.

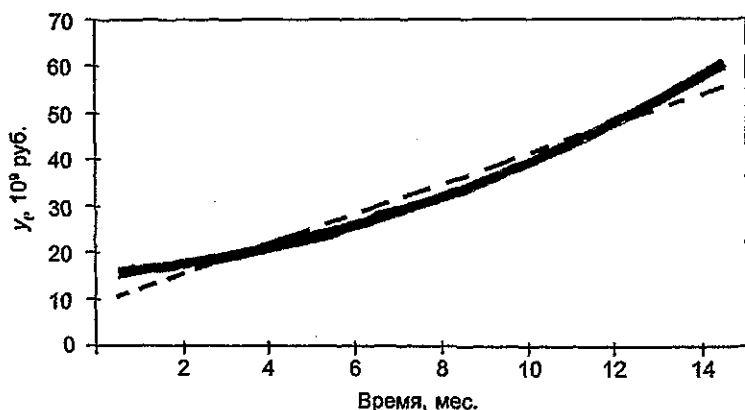


Рис. 3.2. Фактические (—) и расчетные уровни ряда динамики, полученные по линейной (---) и параболической (— · —) моделям

Графический анализ свидетельствует о том, что линейную модель нельзя признать удачной, она не подходит для описания тенденции этого временного ряда. Полученное же на ее основе прогнозное значение будет сильно занижено. Далека от реальности и показательная модель. Значительно ближе к фактическим данным ложатся уровни, выровненные по параболической модели, хотя прогнозное значение может быть несколько завышено. Дальнейшее исследование качества полученных моделей должно опираться на показатели, рассматриваемые в последующих параграфах.

3.3. Методы выбора кривых роста

Рассмотрим подробнее некоторые практические подходы, облегчающие процесс выбора кривой роста.

Наиболее простой путь — визуальный, опирающийся на графическое изображение временного ряда. Подбирают такую кривую роста, форма которой соответствует фактическому развитию процесса. Если на графике исходного ряда тенденция развития недостаточно четко просматривается, то можно провести некоторые стандартные преобразования ряда (например сглаживание), а потом подобрать функцию, отвечающую графику преобразованного ряда. В современных пакетах статистической обработки имеется богатый арсенал стандартных преобразований данных и широкие возможности для графического изображения, в том числе в различных масштабах. Все это позволяет существенно упростить для исследователя проведение данного этапа.

В статистической литературе описан метод последовательных разностей, помогающий при выборе кривых роста полиномиального типа [1, 2]. Предположим, что уровни временного ряда могут быть представлены в виде суммы систематической составляющей f_t и случайной компоненты ε_t :

$$y_t = f_t + \varepsilon_t.$$

При этом случайная компонента удовлетворяет условиям

$$M\varepsilon_t = 0;$$

$$M(\varepsilon_t \varepsilon_{t+\tau}) = \begin{cases} \sigma^2 & \text{при } \tau = 0; \\ 0 & \text{при } \tau \neq 0. \end{cases}$$

Если анализируемый временной ряд содержит в качестве неслучайной составляющей f_t алгебраический полином порядка p , то переход к последовательным разностям порядка $p + 1$ исключает регулярную составляющую, оставляя только элементы, выраженные через остаточную случайную компоненту¹. При этом уровни ряда разностей порядка $p + 1$ будут связаны с элементами случайной компоненты следующим образом:

$$\Delta^{p+1}y_t = \varepsilon_t - C_{p+1}^1 \varepsilon_{t-1} + C_{p+1}^2 \varepsilon_{t-2} - \dots + (-1)^{p+1} \varepsilon_{t-p-1}, \quad (3.23)$$

где $t = p + 2, p + 3, \dots, n$;

C_m^j — число сочетаний из m элементов по j .

Определим среднее значение и дисперсию для $\Delta^{p+1}y_t$:

$$\begin{aligned} M(\Delta^{p+1}y_t) &= 0; \\ D(\Delta^{p+1}y_t) &= Dy_t [1 + (C_{p+1}^1)^2 + (C_{p+1}^2)^2 + \dots + (C_{p+1}^p)^2 + 1] = \sigma^2 C_{2(p+1)}^{p+1}. \end{aligned} \quad (3.24)$$

Доказательство равенства суммы в квадратных скобках значению $C_{2(p+1)}^{p+1}$ приводится в [1].

Из (3.24) следует, что

$$\sigma^2 = \frac{D(\Delta^{p+1}y_t)}{C_{2(p+1)}^{p+1}}. \quad (3.25)$$

¹ Поясним, что для временного ряда $y_1, y_2, \dots, y_n$ последовательные разности первого порядка определяются следующим образом: $\Delta y_t = y_t - y_{t-1}$, $t = 2, \dots, n$. Последовательные разности второго порядка — это разности от последовательных разностей первого порядка $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$, $t = 3, \dots, n$.

Аналогично последовательные разности порядка $k \geq 3$ можно представить в виде: $\Delta^k y_t = \Delta^{k-1} y_t - \Delta^{k-1} y_{t-1}$, $t = k + 1, \dots, n$.

Таким образом, если осуществлять последовательный переход к разностям $\Delta^k y_t$, $k = 1, 2, \dots, p + 1$, то, начиная с порядка разностей $k = p + 1$ (и последующих порядков), выполняются условия (3.24). При этом с учетом равенства нулю среднего значения для последовательностей, образованных разностями этих порядков ($k \geq p + 1$), выборочную оценку дисперсии можно оценить как сумму квадратов элементов последовательности, деленную на $n - k$. В то же время такое оценивание дисперсии для разностей порядка $k < p + 1$ будет приводить к завышенным оценкам. Это объясняется тем, что их уровни будут выражены не только через случайную составляющую ϵ_t , но и через коэффициенты полинома и степени t , следовательно, средние значения для разностей порядка $k < p + 1$ будут отличны от нуля.

Приведенные рассуждения позволяют сформулировать следующее правило подбора порядка p аппроксимирующего полинома.

Для исходного ряда $y_1, y_2, \dots, y_n$ следует вычислить последовательные разности порядка $k = 1, 2, \dots$, а также определить оценки

$$\hat{\sigma}^2(k) = \frac{1}{n-k} \frac{\sum_{t=1}^{n-k} (\Delta^k y_t)^2}{C_{2k}^k}. \quad (3.26)$$

Далее следует проанализировать зависимость разностей и $\hat{\sigma}^2(k)$ от k . Начиная с некоторого значения $k = k_0$ величина (3.26) стабилизируется, оставаясь примерно на одном уровне при дальнейшем росте k . Тогда оценка порядка сглаживающего полинома $p = k_0 - 1$. При этом уровни разностей порядка $k_0 - 1$ также будут демонстрировать стабилизацию.

Как отмечается в [1], этот метод позволяет оценить верхний предел порядка полинома p , который целесообразно использовать для элиминирования неслучайной составляющей.

Этот подход, привлекательный своей простотой, также может быть использован в дальнейшем при приведении ряда к стационарному виду при построении модели ARIMA (см. гл. 6).

Однако на практике чаще всего к выбору формы кривой подходят исходя из значений критерия, в качестве которого принимают сумму квадратов отклонений фактических значений уровней от расчетных, получаемых выравниванием. Из рассматриваемых кривых предпочтение будет отдано той, которой соответствует минимальное значение критерия, так как чем меньше значение критерия, тем ближе к кривой ложатся данные наблюдений.

Используя этот подход, следует иметь в виду ряд моментов. Во-первых, к ряду, состоящему из m точек, можно подобрать мно-

гочлен (полином) степени $m - 1$, проходящий через все m точек. Во-вторых, существует множество многочленов более высоких степеней, также проходящих через все эти точки. Для этих многочленов значение критерия будет равно 0, однако очевидно, что такая кривая не слишком пригодна как для выделения тенденции, так и для целей прогнозирования.

Использование этого подхода должно проходить в два этапа. На первом происходит ограничение приемлемых функций, исходя из содержательного анализа задачи. На втором осуществляется расчет значений критерия и выбор на его основе наиболее подходящей кривой роста.

Необходимость содержательного анализа изучаемого процесса развития может быть проиллюстрирована следующими примерами.

Предположим, что на ретроспективном участке ряд динамики может быть хорошо описан с помощью экспоненциальной кривой. Однако первая половина логистической кривой также представлена экспонентой. Поэтому принять гипотезу об экспоненциальной тенденции ряда в будущем можно только после проведения содержательного анализа, в ходе которого следует дать ответ на вопрос: возможно ли наступление «насыщения» при данной совокупности условий. Например, процесс производства может быть ограничен материальными ресурсами или производственными мощностями.

Возможна ситуация, когда наилучшей функцией по данному критерию будет признана прямая, однако полученное на ее основе прогнозное значение будет отрицательным. Если из экономической сути показателя вытекает невозможность отрицательных значений (например, при прогнозировании объема выпускаемой продукции), то, естественно, следует отказаться от этой функции, выбрав менее «удачную» по данному критерию, но соответствующую содержательному смыслу показателя. Например, более подходящей в этом случае может оказаться показательная кривая при значении параметра $b < 1$ (см. рис. 3.1).

Иногда в качестве критерия выбирается средняя квадратическая ошибка

$$\hat{S} = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n - k}}, \quad (3.27)$$

где y_t — фактическое значение уровня ряда;

$\hat{y}_t$ — расчетное значение уровня ряда, полученное по модели;

k — число оцениваемых коэффициентов модели;

n — длина ряда.

В заключение отметим, что для выбора кривых роста нет жестких рекомендаций. Особенно осторожно следует подходить к решению этой задачи при использовании полученной функции для экстраполяции найденных закономерностей в будущее. Применение кривых роста должно базироваться на предположении о сохранении выявленной тенденции в прогнозируемом периоде. Рассмотренные в данном разделе различные статистические приемы и методы могут помочь исследователю при осуществлении выбора подходящей кривой роста.

3.4. Оценка адекватности и точности выбранных моделей

Вопрос о возможности применения построенных моделей в целях анализа и прогнозирования явления может быть решен только после проверки адекватности, т.е. соответствия модели исследуемому процессу.

Проверка адекватности выбранных моделей реальному процессу (в частности, адекватности полученной кривой роста) строится на анализе остаточной компоненты. Остаточная компонента получается после выделения из исследуемого ряда систематической составляющей (тренда и периодической составляющей, если она присутствует во временном ряду).

Предположим, что исходный временной ряд описывает процесс, не подверженный сезонным колебаниям, т.е. примем гипотезу об аддитивной модели ряда, содержащей трендовую и случайную компоненты.

Тогда ряд остатков будет получен как отклонения фактических уровней временного ряда (y_t) от расчетных ($\hat{y}_t$):

$$e_t = y_t - \hat{y}_t. \quad (3.28)$$

При использовании кривых роста $\hat{y}_t$ вычисляют, подставляя в уравнения выбранных кривых соответствующие последовательные значения времени.

Принято считать, что модель адекватна описываемому процессу, если остаточная последовательность (ряд остатков) представляет собой случайную компоненту ряда. В отношении случайной компоненты ε_t , как правило, предполагают, что

$$\begin{aligned} \varepsilon_t &\in N(0, \sigma_0^2); \\ M(\varepsilon_t \varepsilon_{t+\tau}) &= \begin{cases} \sigma_0^2 & \text{при } \tau = 0; \\ 0 & \text{при } \tau \neq 0. \end{cases} \end{aligned}$$

Поэтому при оценке «качества» модели проверяют, удовлетворяет ли остаточная последовательность следующим свойствам:

- случайности колебаний уровней ряда;
- соответствию распределения остаточной компоненты нормальному закону с нулевым математическим ожиданием;
- независимости значений уровней ряда остатков между собой.

При проверке первого свойства исследователю полезно провести графический анализ остаточной последовательности, а также на этом этапе может быть использован статистический аппарат, рассмотренный в § 3.1.

В современных эконометрических пакетах имеется набор графических средств, позволяющих судить о том, насколько распределение остатков согласуется с нормальным распределением. Например, полезными могут оказаться рис. 3.3—3.4, построенные в системе STATISTICA. На рис. 3.3 показана гистограмма остатков с наложенной нормальной плотностью, позволяющая исследователю оценить симметричность распределения остатков и близость к нормальному закону.

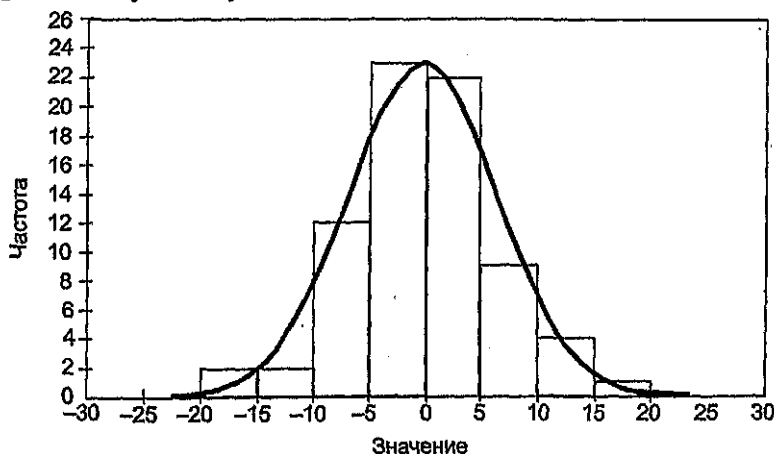


Рис. 3.3. Гистограмма остатков:
— — — ожидаемое нормальное распределение

На рис. 3.4 показаны остатки на нормальной вероятностной бумаге. На основе этого графика можно судить о том, насколько хорошо остатки ложатся на прямую. Если остатки не являются нормально распределенными, то будут наблюдаться значительные отклонения от прямой. Также на этом рисунке с помощью визуального анализа часто удается обнаружить выбросы. Кроме графических средств, в современных пакетах прикладных программ

представлены и статистические критерии, позволяющие проводить проверку гипотезы о нормальности распределения остатков, например критерий Пирсона и др. Однако на практике использование этих средств зачастую затруднено из-за небольшой длины временных рядов экономических показателей ($n < 50$). Поэтому проверка на нормальность может быть проведена приближенно, например на основе подхода, опирающегося на рассмотрение показателей асимметрии и эксцесса [26, 31].

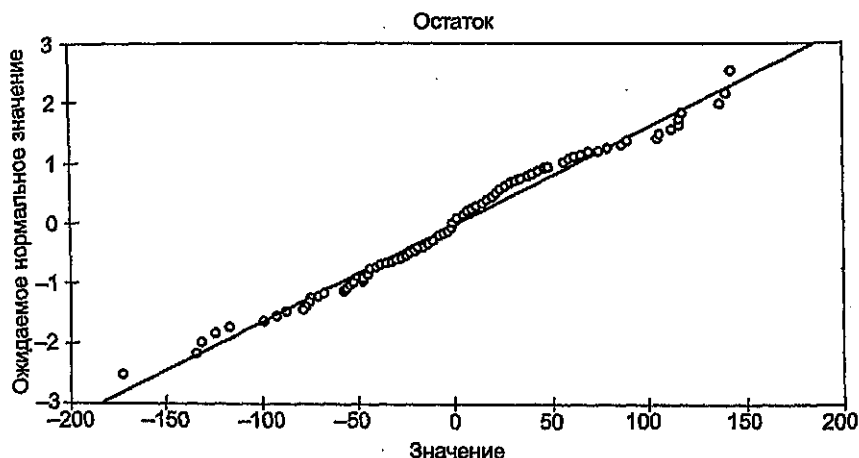


Рис. 3.4. График остатков на нормальной вероятностной бумаге

Как известно, при нормальном распределении показатели асимметрии и эксцесса равны нулю. Так как мы предполагаем, что отклонения от тренда представляют собой выборку из некоторой генеральной совокупности, то можно определить выборочные характеристики асимметрии (A) и эксцесса (Θ), а также оценить их средние квадратические ошибки, зависящие от длины ряда n :

$$A = \frac{\frac{1}{n} \sum_{t=1}^n e_t^3}{\sqrt{\left(\frac{1}{n} \sum_{t=1}^n e_t^2 \right)^3}};$$

$$\Theta = \frac{\frac{1}{n} \sum_{t=1}^n e_t^4}{\sqrt{\left(\frac{1}{n} \sum_{t=1}^n e_t^2 \right)^3}} - 3. \quad (3.29)$$

Если одновременно выполняются следующие неравенства:

$$\begin{aligned} |A| &< 1,5 \cdot \sqrt{\frac{6(n-2)}{(n+1)(n+3)}}; \\ \left| \Xi + \frac{6}{n+1} \right| &< 1,5 \cdot \sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}}, \end{aligned} \quad (3.30)$$

то гипотеза о нормальном характере распределения случайной компоненты не отвергается.

Если выполняется хотя бы одно из неравенств

$$\begin{aligned} |A| &\geq 2 \cdot \sqrt{\frac{6(n-2)}{(n+1)(n+3)}}; \\ \left| \Xi + \frac{6}{n+1} \right| &\geq 2 \cdot \sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}}, \end{aligned} \quad (3.31)$$

то гипотеза о нормальном характере распределения отвергается.

Другие случаи требуют дополнительной проверки с помощью более мощных критериев.

Рассмотрим подробнее последнее свойство. Если вид функции, описывающей систематическую составляющую, выбран неудачно, то последовательные значения ряда остатков могут не обладать свойствами независимости, так как они могут коррелировать между собой. В этом случае говорят, что существует *автокорреляция* остатков.

Существует несколько приемов обнаружения автокорреляции. Наиболее распространен подход, опирающийся на критерий Дарбина—Уотсона. Тест Дарбина—Уотсона связан с проверкой гипотезы об отсутствии автокорреляции первого порядка, т.е. автокорреляции между соседними остаточными членами ряда. При этом критическая статистика определяется по формуле

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-2})^2}{\sum_{t=1}^n e_t^2}. \quad (3.32)$$

Можно показать, что

$$d \approx 2(1 - r_1), \quad (3.33)$$

где r_1 — коэффициент автокорреляции первого порядка (т.е. парный коэффициент корреляции между двумя последовательностями остатков $e_1, e_2, \dots, e_{n-1}$ и $e_2, e_3, \dots, e_n$).

Действительно,

$$\begin{aligned}
 d &= \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} = \frac{\sum_{t=2}^n e_t^2 + \sum_{t=2}^n e_{t-1}^2 - 2 \sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} = \\
 &= \frac{\sum_{t=1}^n e_t^2 - e_1^2 + \sum_{t=1}^n e_t^2 - e_n^2 - 2 \sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} = \\
 &= 2 \left(1 - \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} \right) - \frac{e_1^2 + e_n^2}{\sum_{t=1}^n e_t^2}.
 \end{aligned}$$

Отсюда следует приближенное равенство (3.33), так как при большом числе наблюдений сумма $e_1^2 + e_n^2$ значительно меньше $\sum_{t=1}^n e_t^2$, а в силу условия $\sum_{t=1}^n e_t^2 = 0$ выражение для коэффициента автокорреляции 1-го порядка можно представить в виде:

$$r_1 \approx \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2}.$$

Из (3.33) видно, что близость значения статистики d к нулю означает наличие высокой положительной автокорреляции (коэффициент r_1 близок к 1); близость значения статистики d к 4 означает наличие высокой отрицательной автокорреляции (коэффициент r_1 близок к -1). Естественно, в случае отсутствия автокорреляции значение статистики d будет близким к 2 (коэффициент r_1 не сильно отличается от нуля).

Применение на практике критерия Дарбина—Уотсона основано на сравнении расчетного значения статистики d с пороговыми, граничными значениями d_u и d_l (u (upper) — индекс верхней границы, l (low) — индекс нижней границы).

Граничные значения d_u и d_l , зависящие от числа наблюдений n , числа объясняющих переменных в модели, уровня значимости α , находятся по таблицам (авторами критерия составлены таблицы для $\alpha = 0,05$, $\alpha = 0,025$ и $\alpha = 0,01$). Фрагмент таблицы Дарбина—Уотсона с критическими значениями d_u и d_l при 5%-ном уровне значимости приведен в Приложении 1, табл. 5.

Алгоритм выявления автокорреляции остатков на основе критерия Дарбина—Уотсона следующий.

Выдвигается гипотеза H_0 об отсутствии автокорреляции остатков. Пусть альтернативная гипотеза состоит в наличии в остатках положительной автокорреляции первого порядка.

Тогда при сравнении расчетного значения статистики d ($d < 2$) с d_u и d_l возможны следующие варианты.

1. Если $d < d_l$, то гипотеза H_0 об отсутствии автокорреляции отвергается (с вероятностью ошибки, равной α) в пользу гипотезы о положительной автокорреляции.

2. Если $d > d_u$, то гипотеза H_0 не отвергается.

3. Если $d_l \leq d \leq d_u$, то нельзя сделать определенного вывода по имеющимся исходным данным (значение d попало в область неопределенности).

Если альтернативной является гипотеза о наличии в остатках отрицательной автокорреляции первого порядка, то с пороговыми, граничными значениями d_u и d_l сравнивается значение $4 - d$ (при $d > 2$).

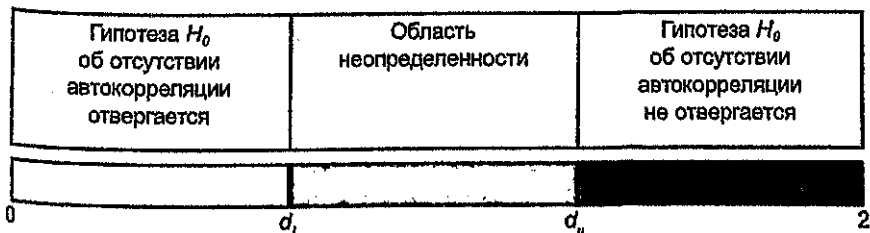
При этом возможны следующие варианты.

1. Если $4 - d < d_l$, то гипотеза H_0 об отсутствии автокорреляции отвергается (с вероятностью ошибки, равной α) в пользу гипотезы об отрицательной автокорреляции.

2. Если $4 - d > d_u$, то гипотеза H_0 не отвергается.

3. Если $d_l \leq 4 - d \leq d_u$, то нельзя сделать определенного вывода по имеющимся исходным данным.

Этот критерий нельзя использовать, если среди объясняющих переменных содержатся лагированные значения результативного показателя (например, он не применим к моделям авторегрессии).



Таким образом, будем считать, что в случае отсутствия автокорреляции в остатках расчетное значение статистики (3.32) «не слишком отличается» от 2 (рис. 3.5).

Пример 3.3

Для временного ряда в примере 3.2 (см. табл. 3.4) проверить гипотезу об отсутствии автокорреляции в остатках, полученных после построения линейной модели тренда (уровень значимости $\alpha = 0,05$).

Решение

В табл. 3.7 представлены вспомогательные вычисления, необходимые для расчета значения статистики Дарбина—Уотсона.

Таблица 3.7. Расчет статистики Дарбина—Уотсона

№	y_i	t	$\hat{y}_i$	e_i	e_i^2	$(e_i - e_{i-1})^2$
1	2	3	4	5	6	7
1	14717	-7	10566,59	4150,406	17225870	—
2	16642	-6	13780,49	2861,51	8188239	1661253
3	18504	-5	16994,39	1509,614	2278934	1827623
4	20376	-4	20208,28	167,718	28129,33	1800685
5	21321	-3	23422,18	-2101,18	4414949	5147889
6	23342	-2	26636,07	-3294,07	10850924	1423001
7	28317	-1	29849,97	-1532,97	2349997	3101487
8	30624	0	33063,87	-2439,87	5952946	822460,4
9	33408	1	36277,76	-2869,76	8235534	184810,6
10	36505	2	39491,66	-2986,66	8920126	13664,67
11	40524	3	42705,55	-2181,55	4759178	648192,5
12	45416	4	45919,45	-503,45	253461,9	2816033
13	50857	5	49133,35	1723,654	2970983	4959992
14	56024	6	52347,24	3676,758	13518549	3814615
15	59381	7	55561,14	3819,862	14591346	20478,75
Σ					$1,05 \cdot 10^8$	28242185

В примере 3.2 получено уравнение линейного тренда $\hat{y}_i = 33\,063,866 + 3213,896t$, причем коэффициенты модели оценивались после переноса начала координат в середину ряда динамики. Об этом следует помнить при определении расчетных уровней $\hat{y}_i$. В графах 6—7 табл. 3.7 приведен расчет сумм, необходимых для вычисления значения статистики по формуле (3.32).

Таким образом,

$$d = \frac{28\,242\,185}{1,05 \cdot 10^8} = 0,270.$$

Вывод очевиден, в данном случае и не требуется обращения к критическим табличным значениям: гипотеза H_0 об отсутствии автокорреляции отвергается в пользу гипотезы о положительной автокорреляции.

Это подтверждает ранее полученный вывод о том, что линейная модель плохо согласуется с динамикой анализируемого временного ряда.

Пример 3.4

Для прогнозирования численности промышленно-производственного персонала была выбрана модель $\hat{y}_t = a_0 + a_1 t$.

Оценка параметров трендовой модели осуществлялась по квартальным данным за период с I квартала 1998 г. по II квартал 2002 г. Значение статистики Дарбина—Уотсона для ряда остатков $d = 2,80$.

Проверить гипотезу об отсутствии автокорреляции первого порядка в остатках (уровень значимости $\alpha = 0,05$).

Решение

Из табл. 5 в Приложении 1 берем значения критических границ d_l и d_u для критерия Дарбина—Уотсона при $n = 17$ и $k = 1$:

$$d_l = 1,13; \quad d_u = 1,38.$$

При $d > 2$ с пороговыми значениями d_u и d_l сравнивается значение $4 - d$, равное в данном случае 1,2.

Так как $d_l \leq 4 - d \leq d_u$ ($1,13 < 1,2 < 1,38$), то нельзя сделать определенного вывода по имеющимся исходным данным (значение d попало в область неопределенности).

Важнейшие характеристики качества модели, выбранной для прогнозирования, — показатели ее точности. Они описывают значения случайных ошибок, полученные при использовании модели. Таким образом, чтобы судить о качестве выбранной модели, необходимо проанализировать систему показателей, характеризующих как адекватность модели, так и ее точность.

О точности можно судить по значению ошибки (погрешности) прогноза.

Ошибка прогноза — величина, характеризующая расхождение между фактическим и прогнозным значением показателя.

Абсолютная ошибка прогноза определяется по формуле

$$\Delta_t = \hat{y}_t - y_t, \quad (3.34)$$

где $\hat{y}_t$ — прогнозное значение показателя;

y_t — фактическое значение.

Эта характеристика имеет ту же размерность, что и прогнозируемый показатель, и зависит от масштаба измерения уровней временного ряда.

На практике широко используется относительная ошибка прогноза, выраженная в процентах относительно фактического значения показателя:

$$\delta_t = \frac{\hat{y}_t - y_t}{y_t} \cdot 100\%. \quad (3.35)$$

Из (3.34), (3.35) видно, что если абсолютная и относительная ошибки больше 0, то это свидетельствует о «завышенной» прогнозной оценке, если меньше 0, то прогнозное значение было занижено.

Также используются средние ошибки по модулю (абсолютные и относительные):

$$\begin{aligned} |\bar{\Delta}| &= \frac{\sum_{t=1}^n |\hat{y}_t - y_t|}{n}; \\ |\bar{\delta}| &= \frac{1}{n} \sum_{t=1}^n \left| \frac{\hat{y}_t - y_t}{y_t} \right| \cdot 100\%, \end{aligned} \quad (3.36)$$

где n — число уровней временного ряда, для которых определялось прогнозное значение.

Последняя характеристика, называемая в англоязычных пакетах Mean Absolute Percentage Error (MAPE), широко используется для сравнения точности прогнозов разнородных объектов прогнозирования. При этом в литературе часто встречаются указания на то, что значение $|\bar{\delta}| < 10\%$ свидетельствует о высокой точности модели, при значениях этой характеристики в диапазоне 10—20% точность можно признать хорошей, при $20\% < |\bar{\delta}| < 50\%$ — удовлетворительной.

Однако такая универсальность, не учитывающая специфики временных рядов и прогнозных расчетов (характера динамики, назначения прогнозов, предполагаемого времени упреждения и других характеристик) делает этот подход достаточно «механическим».

Очевидно, что все указанные характеристики могут быть вычислены после того, как период упреждения уже закончился, и имеются фактические данные о прогнозируемом показателе или при рассмотрении показателя на ретроспективном участке.

В последнем случае имеющаяся информация делится на две части: по первой — оцениваются параметры модели, а данные второй части рассматриваются в качестве фактических. Ошибки прогнозов, полученные ретроспективно (на втором участке) характеризуют точность применяемой модели.

На практике при проведении сравнительной оценки моделей может использоваться такая характеристика качества, как средняя квадратическая ошибка (S):

$$S = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}}, \quad (3.37)$$

где n — длина временного ряда;

y_t — фактическое значение уровня ряда;

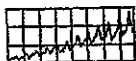
$\hat{y}_t$ — расчетное значение уровня, полученное по модели.

Часто знаменателем подкоренного выражения служит величина $(n - k)$, где k — число оцениваемых коэффициентов модели (см. (3.27)).

Иногда в эконометрических пакетах при выборе лучшего метода прогнозирования используют суммы квадратов ошибок (Sums of Squares Errors — SSE) или средние квадраты ошибок (Mean Squares Errors — MSE). Очевидно, что предпочтение в этом случае будет отдано той же модели, что и при проведении сравнительной оценки с помощью (3.37).

Чем меньше значения всех рассмотренных характеристик, тем выше точность модели. О точности модели нельзя судить по одному значению ошибки прогноза. Например, если прогнозная оценка месячного уровня производства в июне совпала с фактическим значением, то это не является достаточным доказательством высокой точности модели. Надо учитывать, что хорошее единичное прогнозное значение может быть получено и по плохой модели, и наоборот. Следовательно, о качестве применяемых моделей можно судить лишь по совокупности сопоставлений прогнозных значений с фактическими.

В заключение отметим, что не может быть чисто формальных подходов к выбору методов и моделей прогнозирования. Успешное применение статистических методов прогнозирования на практике возможно лишь при сочетании знаний в области самих методов с глубоким знанием объекта исследования, с содержательным анализом изучаемого явления.



Практикум к главе 3

Контрольные вопросы

1. Охарактеризуйте основные типы кривых роста, наиболее часто используемые на практике при построении трендовых моделей.
2. Назовите важнейшие характеристики точности моделей прогнозирования.
3. Объясните суть метода последовательных разностей.
4. Каким образом определяется значение критической статистики в тесте Дарбина—Уотсона?
5. Опишите алгоритм проверки гипотезы об отсутствии автокорреляции первого порядка в остатках модели с помощью критерия Дарбина—Уотсона.
6. Поясните, почему при отсутствии автокорреляции в остатках расчетное значение статистики Дарбина—Уотсона «не слишком отличается» от 2.
7. Запишите систему нормальных уравнений для определения параметров полиномиальной модели третьего порядка.
8. Какова интерпретация коэффициентов линейной трендовой модели?
9. Какова интерпретация коэффициентов показательной трендовой модели $\hat{y}_t = ab^t$?
10. Для каких целей может быть использован критерий серий?

Упражнения

- 3.1. По данным об остатках вкладов населения в банках (см. пример 3.2) рассчитайте значение критической статистики Дарбина—Уотсона для остаточной последовательности, полученной при построении параболической трендовой модели. Проверьте гипотезу об отсутствии автокорреляции первого порядка (при уровне значимости $\alpha = 0,05$).
- 3.2. По данным примера 3.2 найдите коэффициенты линейной трендовой модели, воспользовавшись формулой (3.10), представляющей решение системы нормальных уравнений в матричном виде. На основе полученной модели сделайте прогноз на два месяца вперед (для периода упреждения $L = 2$). Укажите, как можно использовать формулу (3.10) для оценивания коэффициентов показательной трендовой модели $\hat{y}_t = ab^t$.

В заданиях 3, 4 выберите правильный вариант ответа.

- 3.3. Тенденция изменения численности промышленно-производственного персонала предприятия за 10 лет описывается показательной функцией $\hat{y}_t = 431 \cdot 1,092^t$.

Из этой модели следует, что среднегодовой темп роста численности составил:

а) 109,2%; б) 431%; в) 92%; г) 9,2%.

3.4. На основе годовых данных об изменении урожайности картофеля в регионе были оценены коэффициенты линейного тренда: $\hat{y}_t = 172,2 + 4,418t$.

В соответствии с этой моделью среднегодовой прирост урожайности составляет: а) 4,418 ц/га; б) 172,2 ц/га; в) $(172,2 + 4,418)$ ц/га; г) 4,418%.

3.5. Динамика прибыли производственного объединения представлена в таблице.

t	1	2	3	4	5	6	7
y_t , млн руб.	70	76	83	88	94	101	106

Рассчитайте коэффициенты линейного тренда $\hat{y}_t = a_0 + a_1t$, используя перенос начала координат в середину ряда динамики. На основе полученной модели определите прогнозное значение прибыли производственного объединения в точке $t = 8$ (время упреждения прогноза $L = 1$).

3.6. Ежеквартальная динамика объемов реализованной продукции объединения представлена в таблице.

t	1	2	3	4	5	6	7	8	9
y_t , млн руб.	95	100	108	113	124	135	149	160	169

Оцените коэффициенты параболического тренда $\hat{y}_t = a_0 + a_1t + a_2t^2$, используя перенос начала координат в середину ряда динамики. На основе полученной модели, определите прогноз реализованной продукции объединения в точке $t = 10$ (время упреждения прогноза $L = 1$).



4.1. Статистические методы оценки уровня сезонности. Прогнозирование с помощью тренд-сезонных моделей

Исследования периодических колебаний в экономике уходят корнями далеко в прошлое. Еще К. Жюгляр (1819—1905) при изучении экономических временных рядов в целях выделения бизнес-циклов установил цикличность инвестиций (период цикла 7—11 лет). Позднее этой проблематикой занимались С. Китчин, С. Кузнец, Н. Кондратьев. Были выявлены циклы в обновлении оборотных средств (период 3—5 лет), циклы в строительстве (период 15—20 лет), большие «волны Кондратьева» [30].

В 20-х годах XX в. начали развиваться исследования, связанные с изучением сезонных колебаний в различных секторах народного хозяйства, отраслях промышленности.

При создании Гарвардского барометра под руководством У. Персона и У. Митчела рассчитывались три производных временных ряда, отражающих динамику развития фондового, товарного и денежного рынков, рассматривались взаимосвязи в их развитии. При этом в статистически обрабатываемых временных рядах исключалась тенденция, периодическая компонента, проводился всесторонний анализ колебаний.

Успех Гарвардского барометра способствовал развитию этого направления в других странах. Однако впоследствии Е.Е. Слуцкий в своей работе¹ показал возможность имитации действительной периодичности с помощью скользящих средних. Эта работа косвенно подвергла сомнению периодические закономерности, выявленные с помощью экономических барометров, а также в ряде других исследований.

Потребности экономической практики послужили мощным стимулом к совершенствованию статистической методологии в области выявления, измерения, моделирования и прогнозирования сезонных колебаний.

Первые методы исследования сезонности (метод простых средних, метод Персона, способ относительных чисел и другие) предполагали неизменность амплитуды и формы сезонной волны [21, 26].

¹ См. сноску к с. 47.

В последующие годы получили развитие итеративные методы фильтрации компонент временных рядов (например, методы Четверикова, Ферстера, Шискина—Эйзенпресса и др.) [21, 31]. Многие из этих подходов уже отказывались от предпосылки неизменности сезонной волны, делая процесс исследования периодических колебаний более гибким. Однако, опираясь на процедуры скользящих средних, эти методы приводили к потере части информации на концах временных рядов. Этого удалось избежать в современных подходах к сезонной декомпозиции (корректировке) (см. § 4.2). Именно эти процедуры представлены в эконометрических пакетах, широко используемых в настоящее время в мировой практике.

При моделировании периодических колебаний в экономике также стали использовать гармонический анализ, пришедший из естественных наук. Первые экономические работы в этой области связаны с именами Г. Мура и Бэвэриджа [30].

Обобщение разложения в ряд Фурье в частотной области привело к быстрому развитию аппарата спектрального анализа. Однако значительно позднее, чем в приложениях, связанных с естественными науками, в экономике для выявления и фильтрации периодических колебаний стали применяться спектральные методы (см. § 4.4).

В настоящее время при описании и прогнозировании тренд-сезонных процессов используются подходы, связанные с применением индексов сезонности в сочетании с кривыми роста, процедуры, опирающиеся на широкий спектр адаптивных моделей, сезонный вариант модели ARIMA, а также разрабатываются специализированные подходы, учитывающие специфику конкретных временных рядов.

Остановимся более подробно на методологических вопросах расчета *сезонной составляющей*. Предположим, что динамика ряда характеризуется неизменными во времени сезонными эффектами.

Очевидно, что процедуры расчета зависят от принятой модели временного ряда, содержащей сезонность в аддитивной или мультипликативной форме. При этом для аддитивной модели характеристики сезонности будут измеряться в абсолютных величинах, а для мультипликативной — в относительных.

Рассмотрим алгоритм расчета для случая *аддитивной сезонности*.

1. Для описания тенденции воспользуемся процедурой скользящей средней при четной длине интервала сглаживания $l = 2r$. Тогда для временных рядов месячной динамики скользящая средняя при $l = 12$ на каждом активном участке будет определяться выражением

$$y'_t = \frac{\frac{1}{2}y_{t-6} + y_{t-5} + \dots + y_t + \dots + y_{t+5} + \frac{1}{2}y_{t+6}}{12}. \quad (4.1)$$

Аналогичное выражение можно записать для определения скользящей средней при $l = 4$, часто используемой для рядов квартальной динамики:

$$y'_t = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4}. \quad (4.2)$$

2. Рассчитаем отклонения фактических значений от уровней сглаженного ряда:

$$x_t = y_t - y'_t. \quad (4.3)$$

Уровни вновь полученного ряда отражают эффект сезонности и случайности.

3. Для элиминирования влияния случайных факторов определим предварительные значения сезонной составляющей как средние значения из уровней x_t для одноименных месяцев (кварталов). Например, для временных рядов месячной динамики процедура усреднения может быть описана выражением:

$$\bar{x}_i = \begin{cases} \frac{1}{k} \sum_{j=1}^k x_{12j+i} & \text{при } i=1, 2, \dots, 6; \\ \frac{1}{k} \sum_{j=0}^{k-1} x_{12j+i} & \text{при } i=7, 8, \dots, 12, \end{cases} \quad (4.4)$$

где k — число целых периодов (циклов) во временном ряду, полученном на втором шаге.

Разные пределы суммирования объясняются тем, что при использовании скользящей средней с четным значением длины интервала сглаживания ($l = 2p$) p первых и p последних уровней ряда будут потеряны. (В нашем случае потери составят по 6 уровней в начале и в конце ряда.)

4. Проведем корректировку первоначальных значений сезонной составляющей, вызванную тем, что суммарное воздействие сезонности на динамику предполагается нейтральным. Для аддитивного случая сумма значений сезонной составляющей для полного сезонного цикла должна быть равна нулю. Поэтому окончательные, скорректированные оценки сезонной компоненты определим с помощью следующего выражения:

$$S_i = \bar{x}_i - \bar{x} \quad (i=1, 2, \dots, m), \quad (4.5)$$

где $\bar{x} = \frac{1}{m} \sum_{i=1}^m \bar{x}_i$;

m — число фаз в полном сезонном цикле (как правило, $m = 12$ для рядов месячной динамики и $m = 4$ для квартальных данных).

Для мультипликативной формы сезонности меняется содержание 2 и 4-го этапов алгоритма. Очевидно, что на втором шаге знак вычитания необходимо заменить делением:

$$x_t = y_t / y'_t. \quad (4.6)$$

После получения предварительных оценок сезонности усреднением на шаге 3 следует выполнить процедуру их корректировки. Взаимопогашаемость сезонных колебаний в мультипликативной форме выражается в том, что средняя арифметическая из значений коэффициентов сезонности для полного сезонного цикла должна быть равна 1. Поэтому окончательные оценки коэффициентов сезонности получим с помощью следующего выражения:

$$S_i = \bar{x}_i \bar{x} \quad (i=1, 2, \dots, m), \quad (4.7)$$

где $\bar{x} = \frac{m}{\sum_{i=1}^m \bar{x}_i}$, m — число фаз в полном сезонном цикле.

Процедуру построения тренд-сезонных моделей можно описать в виде следующей последовательности шагов.

1. Оценивание сезонной составляющей по описанной схеме (4.1)—(4.7) с учетом характера сезонности (аддитивной или мультипликативной).

2. Сезонная корректировка (десезонализация) исходных данных.

3. Расчет параметров тренда на основе временного ряда, полученного на втором шаге.

4. Моделирование динамики исходного ряда с учетом трендовой и сезонной составляющих.

5. Оценка точности и адекватности полученной модели.

6. Использование построенной модели для прогнозирования.

Рассмотрим реализацию этой методики на следующем примере.

Пример 4.1

В табл. 4.1 представлены квартальные данные об объемах продаж продукции фирмы за последние четыре года.

Требуется:

1) на основе графического анализа провести исследование компонентного состава временного ряда объема продаж;

2) рассчитать прогнозную оценку объемов продаж в первом полугодии следующего года.

Решение

1. Графический анализ исходного временного ряда свидетельствует о наличии трендовой компоненты: имеется устойчивая, ярко выраженная тенденция роста объемов продаж в течение последних четырех лет. Для исследования характера тенденции на практике часто пользуются приемом агрегирования уровней за счет перехода к укрупненным интервалам динамического ряда.

Рис. 4.1 показывает, что характер тенденции близок к линейному развитию, также отчетливо видны сезонные колебания (период колебаний равен одному году). Наблюдается устойчиво повторяющееся увеличение объемов продаж в 4- и в 1-м кварталах по сравнению с 3- и 2-м кварталами, причем наиболее существенные «всплески» в динамике показателя просматриваются в 4-м квартале. Так как амплитуда сезонных колебаний остается примерно постоянной, неизменной с течением времени, то для описания и прогнозирования динамики временного ряда можно предложить аддитивную модель.

2. Проведем выравнивание, сглаживание исходного ряда с помощью простой скользящей средней в соответствии с выражением (4.2). Результаты расчетов показаны в табл. 4.1., гр. 5.

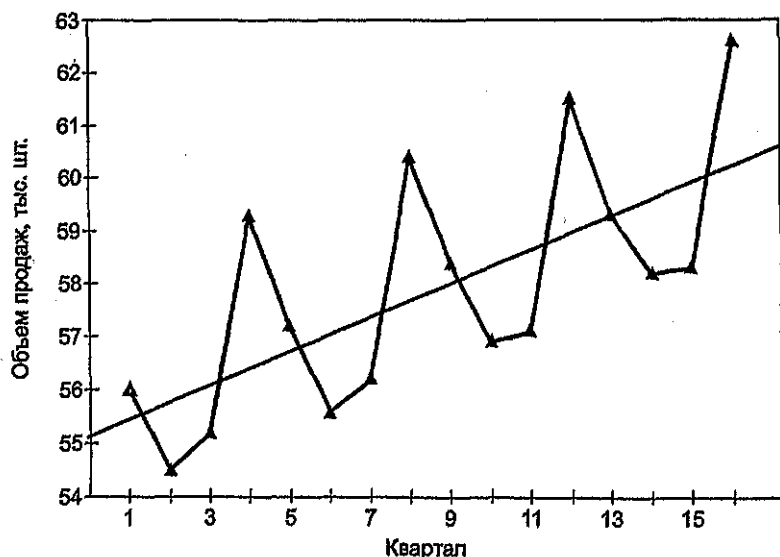


Рис. 4.1. Квартальная динамика объемов продаж

Вычитая из фактических уровней значения сглаженного ряда, получим временной ряд, уровни которого x_t отражают влияние случайных факторов и сезонности (гр.6).

Предварительную оценку сезонной компоненты получим усреднением уровней временного ряда x_t для одноименных кварталов. Например, значение сезонной составляющей для 1-го квартала:

$$\bar{x}_1 = \frac{0,250 + 0,312 + 0,125}{3} = 0,210.$$

Аналогично вычислены $\bar{x}_2, \bar{x}_3, \bar{x}_4$ (гр. 3).

Таблица 4.1. Сглаживание исходного ряда с помощью скользящей средней

№ года	№ квартала	t	Объем продаж y_t , тыс. шт.	Скользящая средняя y'_t	$x_t = y_t - y'_t$
1	2	3	4	5	6
1	I	1	56,0	—	—
	II	2	54,5	—	—
	III	3	55,2	56,400	-1,200
	IV	4	59,3	56,688	2,613
2	I	5	57,2	56,950	0,250
	II	6	55,6	57,213	-1,612
	III	7	56,2	57,500	-1,300
	IV	8	60,4	57,813	2,588
3	I	9	58,4	58,088	0,312
	II	10	56,9	58,338	-1,438
	III	11	57,1	58,588	-1,488
	IV	12	61,5	58,863	2,638
4	I	13	59,3	59,175	0,125
	II	14	58,2	59,463	-1,263
	III	15	58,3	—	—
	IV	16	62,6	—	—

Так как $\sum_{i=1}^4 \bar{x}_i = 0,075$ (отлична от нуля), то проведем корректировку значений сезонной составляющей. В соответствии с (4.5) найдем «поправку», на которую надо изменить предварительные оценки сезонности:

$$\bar{x} = \frac{0,075}{4} = 0,01875.$$

Скорректированные оценки сезонности приведены в табл. 4.2, гр. 4. Например, оценка сезонной компоненты для 1-го квартала

$$S_1 = 0,229 - 0,01875 \approx 0,210.$$

Таблица 4.2. Оценивание сезонной компоненты в аддитивной модели

№ квартала	<i>i</i>	Предварительная оценка сезонной компоненты $\bar{x}_i$	Скорректированные значения сезонной компоненты S_i
I	1	0,229	0,210
II	2	-1,438	-1,456
III	3	-1,329	-1,348
IV	4	2,613	2,594
Σ		0,075	0

Таким образом, табл. 4.1, 4.2 отражают процесс оценивания сезонной компоненты для аддитивной модели.

В табл. 4.3 показаны основные этапы процесса прогнозирования по аддитивной модели. Сначала из исходного ряда удаляется сезонная составляющая, т.е. осуществляется сезонная корректировка (десезонализация) объема продаж. Для описания тенденции воспользуемся моделью линейного тренда, так как это согласуется с результатами графического анализа динамики показателя.

Модель имеет вид:

$$\bar{y}_t^{(1)} = 55,543 + 0,28t.$$

Расчетные уровни, полученные подстановкой последовательных значений времени $t = 1, 2, \dots, 16$ в это уравнение, представлены в табл. 4.3, гр. 5.

На заключительном этапе определим расчетные уровни объема продаж по аддитивной модели, суммируя значения полученных оценок трендовой и сезонной составляющих (гр. 6).

Для прогнозирования объема продаж в первом полугодии следующего года оценим значения тренда при $t = 17$ и $t = 18$:

$$\hat{y}_{17}^{(1)} = 55,543 + 0,28 \cdot 17 = 60,303 \text{ тыс. шт.};$$

$$\hat{y}_{18}^{(1)} = 55,543 + 0,28 \cdot 18 = 60,583 \text{ тыс. шт.},$$

а затем прибавим к полученным значениям соответствующие оценки сезонной составляющей (для 1 и 2-го кварталов):

Таблица 4.3. Прогнозирование объема продаж с помощью аддитивной тренд-сезонной модели

t	Объем продаж y_t , тыс. шт.	Сезонная компо- нента S_t	Десезонали- зированный объем продаж $y_t^{(1)}$	Расчетные значения (линей- ный тренд) $\hat{y}_t^{(1)}$	Расчетные значения объема продаж $\hat{y}_t = \hat{y}_t^{(1)} + S_t$
1	56,0	0,210	55,790	55,822	56,032
2	54,5	-1,456	55,956	56,102	54,646
3	55,2	-1,348	56,548	56,381	55,034
4	59,3	2,594	56,706	56,661	59,255
5	57,2	0,210	56,990	56,940	57,151
6	55,6	-1,456	57,056	57,220	55,764
7	56,2	-1,348	57,548	57,499	56,152
8	60,4	2,594	57,806	57,779	60,373
9	58,4	0,210	58,190	58,059	58,269
10	56,9	-1,456	58,356	58,338	56,882
11	57,1	-1,348	58,448	58,618	57,270
12	61,5	2,594	58,906	58,897	61,491
13	59,3	0,210	59,090	59,177	59,387
14	58,2	-1,456	56,656	59,456	58,000
15	58,3	-1,348	59,648	59,736	58,388
16	62,6	2,594	60,006	60,015	62,609
17		0,210		60,303	60,513
18		-1,456		60,583	59,127

Таким образом, ожидаемый объем продаж в первом полугодии составляет 119,64 тыс. шт. Характеристики точности модели представлены в табл. 4.4. О высокой точности модели свидетельствуют рассчитанные абсолютные и относительные ошибки (максимальная относительная ошибка по модулю не превышает 0,5%, средняя относительная ошибка по модулю составила 0,18%).

В тоже время необходимо отметить, что рассмотренный подход имеет существенный недостаток — предположение о неизменности во времени сезонных эффектов, что далеко не всегда соответствует реальной динамике многих экономических процессов. Более гибкие методы оценивания сезонности будут рассмотрены в последующих главах.

**Таблица 4.4. Характеристики точности аддитивной
тренд-сезонной модели**

<i>t</i>	Объем продаж y_t , тыс. шт.	Расчетные значения объема продаж $\hat{y}_t$, тыс. шт.	Абсолютная ошибка, тыс. шт.	Относительная ошибка, %
1	56,0	56,033	0,033	0,059
2	54,5	54,646	0,146	0,267
3	55,2	55,034	-0,166	-0,302
4	59,3	59,255	-0,045	-0,076
5	57,2	57,151	-0,049	-0,086
6	55,6	55,764	0,164	0,294
7	56,2	56,152	-0,048	-0,086
8	60,4	60,373	-0,027	-0,045
9	58,4	58,269	0,169	0,291
10	56,9	56,882	-0,018	-0,032
11	57,1	57,270	0,170	0,297
12	61,5	61,491	-0,009	-0,015
13	59,3	59,387	0,287	0,486
14	58,2	58,000	-0,200	-0,344
15	58,3	58,388	0,088	0,150
16	62,6	62,609	0,090	0,014

4.2. Сезонная декомпозиция и корректировка временных рядов

На практике часто требуется выделить во временных рядах сезонную, случайную, трендовую компоненту, т.е. провести декомпозицию ряда, разложение его на составные части. Кроме того, сезонные и случайные колебания зачастую затрудняют анализ тенденции развития показателя, и их влияние должно быть элиминировано.

Развитие популярных методов сезонной декомпозиции связано с деятельностью Бюро переписи США (U.S. Bureau of Census). На основе одного из этих методов в 1954 г. была создана компьютерная программа сезонной корректировки, направленная на очищение временных рядов от сезонных эффектов. Она получила известность как «Метод Бюро переписи». В следующем году появилась новая, переработанная версия этой программы, названная «Метод Бюро переписи II». Впоследствии появилось много модификаций этой процедуры, причем каждая новая вер-

сия получала название, в котором присутствовал символ X и номер версии.

В настоящее время в мировой практике для сезонной корректировки широко используется процедура «II-X-11» или просто «X-11», представляющая собой развитие классической сезонной декомпозиции. Она получила широкую известность в сфере деятельности как правительственных учреждений, так и частного бизнеса. Эта процедура содержит множество усовершенствований, к ее достоинствам относятся: возможность учитывать различное влияние дней недели, отдельных месяцев, улавливать сложные календарные биения, возможность корректировки экстремальных наблюдений, более широкий выбор способов вычисления скользящих средних и другие. При использовании этого подхода временной ряд последовательно пропускается через несколько стадий обработки, причем на каждом новом этапе предыдущие оценки пересчитываются. Пользователь имеет возможность проанализировать компоненты ряда, полученные на разных ступенях обработки.

Дальнейшее развитие этой процедуры связано с исследованиями специалистов в Бюро статистики Канады (Statistics Canada), разработавшими важную модификацию основного подхода. Ими было установлено, что в некоторых случаях при добавлении к ряду новых данных происходит достаточно сильное изменение оцениваемых сезонных факторов, особенно заметное в конце участка моделирования. Эта проблема, получившая название «проблемы виляющего хвоста», может создать определенные трудности при прогнозировании. Она бывает вызвана особенностями реализации процедур скользящих средних, используемых методами сезонной корректировки (краевыми эффектами).

Новый подход, направленный на стабилизацию оценивания сезонных факторов, был назван «X-11-ARIMA». Он состоит из следующих укрупненных блоков:

- подбора модели ARIMA¹ к временному ряду;
- последующего добавления расчетных уровней к концам исходного ряда, так что ряд оказывается продолженным в конце и «продлен» в начале;
- сезонной корректировки расширенного ряда с помощью процедуры X-11.

Экспериментальные расчеты показали, что прогностическое расширение временного ряда позволило улучшить учет сезонных колебаний. В настоящее время в практической деятельности орга-

¹ Методологии построения модели ARIMA рассматривается в § 6.2.

нов статистики уже получила распространение следующая версия программы «X-12-ARIMA».

Процедуры классической сезонной декомпозиции состоят из следующих этапов.

1. Предварительная обработка исходных данных. На этом шаге по решению исследователя может быть исключено влияние устойчивых, предсказуемых воздействий, например, связанных с календарным фактором. В частности, при изучении динамики объемов производства важнейших видов промышленной продукции в натуральном выражении для элиминирования влияния различного числа рабочих дней в месяцах переходят к среднесуточным объемам производства для каждого месяца. Это преобразование проводят с учетом характера производства (прерывного или непрерывного).

2. Исходный ряд или ряд, полученный после преобразования на первом шаге (Y_t), сглаживается для получения предварительной оценки тренда $U_t^{(0)}$.

Для этой цели при работе с временными рядами месячной динамики может быть использована скользящая средняя с длиной активного участка $l = 12$:

$$U_t^{(0)} = \frac{\frac{1}{2}y_{t-6} + y_{t-5} + \dots + y_t + \dots + y_{t+5} + \frac{1}{2}y_{t+6}}{12} \quad (4.8)$$

3. Находится отношение соответствующих уровней ряда Y_t к уровням ряда $U_t^{(0)}$ для получения предварительной оценки сезонной и случайной составляющих (SE_t):

$$SE_t = Y_t / U_t^{(0)} \quad (4.9)$$

4. Для элиминирования влияния случайной составляющей уровня временного ряда SE , усредняются для каждого «сезона» (месяца, квартала — в зависимости от того, какой вид сезонной декомпозиции реализуется). В современных ППП пользователь может сам определить параметры процедуры сглаживания либо выбрать режим их автоматического задания.

Затем значения, полученные после сглаживания, корректируются для того, чтобы суммарное воздействие сезонности на динамику было нейтральным. В результате на этом шаге определяется оценка сезонной составляющей $S_t^{(1)}$.

5. Осуществляется поправка исходного ряда на сезонную компоненту:

$$UE_t = Y_t / S_t^{(1)} \quad (4.10)$$

6. Пересмотренная оценка тренда ($U_t^{(1)}$) получается после сглаживания ряда UE_t . Нечетная длина фильтров $l = 2p + 1$ может задаваться пользователем или выбираться используемой статистической системой автоматически.

7. Случайная компонента получается путем деления ряда, скорректированного на сезонные колебания (UE_t), на последнюю оценку тренда ($U_t^{(1)}$):

$$E_t^{(1)} = UE_t / U_t^{(1)}. \quad (4.11)$$

Необходимо отметить, что описанная схема представляет укрупненный вариант процедуры декомпозиции. Современные эконометрические пакеты для чистоты фильтрации компонент после 6-го шага повторяют шаги, начиная с 3-го. Теперь вместо $U_t^{(0)}$ на этом этапе будет использована более поздняя оценка тренда $U_t^{(1)}$. Окончательные оценки компонент будут получены после повторной реализации соответствующих этапов. Сезонная составляющая $S_t^{(2)}$ будет оценена на 4-м шаге, тренд $U_t^{(2)}$ — на 6-м, случайная составляющая $E_t^{(2)}$ — на 7-м.

Описанный вариант процедуры декомпозиции относится к мультипликативным моделям временных рядов. Для аддитивных моделей на 3, 5, 7-м шагах деление будет заменено вычитанием.

Пример 4.2

Рассмотрим результаты сезонной корректировки временного ряда индекса промышленного производства в топливной промышленности¹. Для реализации сезонных корректировок была использована программа «X-12-ARIMA». Сначала уровни временного ряда были преобразованы в индексы среднесуточного производства — тем самым было исключено влияние календарного фактора. Полученный временной ряд с исключенной календарной компонентой с помощью системы цепных индексов был пересчитан к одному базовому периоду (к декабрю 1997 г.). Такая операция проводится, чтобы все значения временного ряда получили одну «точку отсчета». Анализировать и строить прогнозы по индексам, рассчитанным по отношению к соответствующему месяцу прошлого года, было бы методологически неверно, так как на построенный

¹ Любавина О.Н. Статистический анализ и прогнозирование развития топливно-энергетических отраслей Российской Федерации: Автореф. дис. ... канд. экон. наук. — М., 2001.

таким образом временной ряд сильное влияние оказывал бы «эффект базы сравнения» прошлого года.

При осуществлении сезонной декомпозиции использовалась мультипликативная модель. Для прогнозного расширения полученного временного ряда применялась модель ARIMA, причем в программе предусмотрен автоматический выбор вида модели (ее идентификация). В то же время пользователь имеет возможность самостоятельно задать вид модели, опираясь на методологию, рассмотренную в гл. 6.

Результаты проведения сезонной корректировки показаны на рис. 4.2. Анализ показал, что последствия августовского кризиса в меньшей степени задели отрасли топливной промышленности, чем другие промышленные отрасли. В дальнейшем общепромышленное оживление, удачно сложившаяся конъюнктура на мировых рынках повлекли за собой рост производства в топливной промышленности.

Оценки показали, что с сентября 1999 г. был превышен уровень производства декабря 1997 г. В марте 2001 г. уровень производства превышал уровень декабря 1997 г. на 2%, а уровень августа 1998 г. — на 8%.

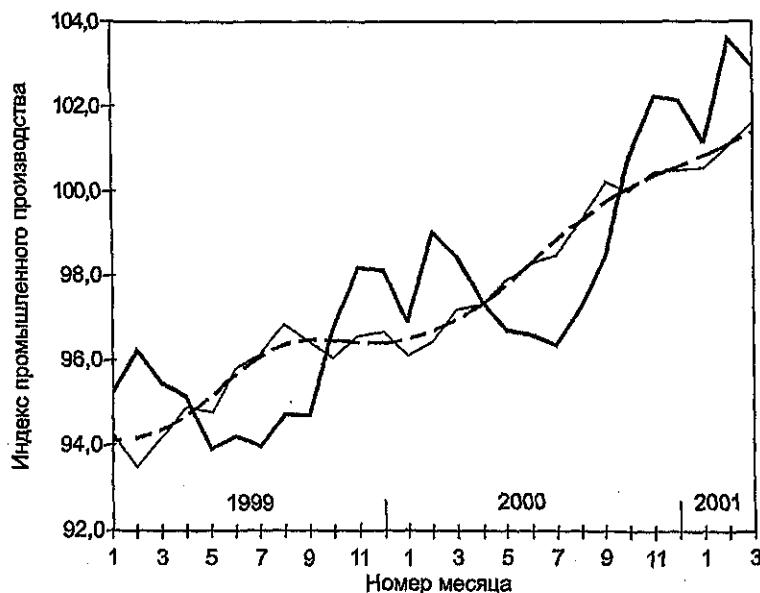


Рис. 4.2. Сезонная корректировка индекса промышленного производства в топливной промышленности (% к декабрю 1997 г.):

— с исключением календарного фактора; - - с исключением сезонной компоненты; — тренд

В настоящее время при исследовании динамики важнейших показателей развития промышленного комплекса России задача сезонной корректировки считается одной из наиболее актуальных. Причем она возникает на микро-, макро- и мезоуровнях. Такая корректировка должна осуществляться для индекса промышленного производства в целом, индексов производства для отдельных отраслей и отраслевых групп промышленности. Большой практический интерес вызывает элиминирование сезонных колебаний для многих временных рядов производства промышленной продукции в натуральном выражении.

4.3. Моделирование сезонных колебаний с помощью фиктивных переменных

Фиктивные переменные существенно расширяют сферу применения линейных регрессионных моделей (а также класса нелинейных моделей, приводимых к линейному виду с помощью соответствующих преобразований).

Они позволяют строить модели в условиях неоднородности структуры наблюдений, причем неоднородность может носить пространственный или временной характер, может объясняться влиянием качественных признаков. Как отмечается в [1, с. 738], этот прием «используется обычно при работе с неоднородными (в регрессионном смысле) исходными статистическими данными, в ситуациях, когда изменение значений сопутствующих переменных влияет на изменение значений только части коэффициентов регрессий».

Наибольшее распространение получили модели, в которых результативный признак рассматривается в зависимости от ряда экономических факторов x_j ($j = 1, 2, \dots, k$) и фиктивных переменных d_i ($i = 1, 2, \dots, m$). Последние отражают различие в формировании результативного признака по отдельным классам наблюдений. В англоязычной эконометрической литературе фиктивные переменные называют *dummy variables*, *манекены*, иногда в отечественной литературе используется термин *структурные переменные* [15, 30]. Фиктивность этих переменных состоит только в том, что они описывают качественные признаки количественным образом. Во всем остальном это обычные регрессоры, и вся процедура регрессионного анализа при их использовании не претерпевает никаких изменений при оценивании параметров, проверке значимости и т.д.

Можно выделить два основных достоинства регрессионной процедуры с использованием фиктивных переменных.

1. Повышается надежность статистических выводов по сравнению с моделями, коэффициенты которых оценивались бы отдельно по каждой однородной подвыборке. (Имеются в виду подвыборки, полученные путем разбиения исходных данных на однородные группы по значениям сопутствующих переменных.)

Известно, что надежность методов оценивания существенно зависит от соотношения длины выборки n и числа параметров модели $p + 1$. Применение фиктивных переменных приводит к увеличению отношения $n/(p + 1)$, следовательно, способствует повышению точности соответствующих оценок модели.

2. Появляется возможность в ходе построения регрессионной модели с фиктивными переменными проводить проверку гипотез о существовании статистически значимого влияния сопутствующих (вспомогательных) переменных на результативный признак и структуру модели.

Построение модели осуществляется с помощью аддитивного добавления в линейной форме в правую часть регрессионного уравнения дихотомических переменных (бинарных переменных, которые могут принимать одно из двух значений: нуль или единица).

Для отражения влияния качественной сопутствующей переменной, имеющей m состояний, необходимо включить в уравнение только $m - 1$ фиктивную переменную. Это объясняется следующим обстоятельством. Предположим, что в регрессионную модель со свободным членом были включены m фиктивных переменных. Тогда в матрице исходных наблюдений X сумма m столбцов, соответствующих этим дихотомическим переменным, будет равна первому единичному столбцу, соответствующему свободному члену в уравнении. Такая ситуация получила название *dummy trap* или *ловушки* [19]. При этом будет существовать линейная зависимость между столбцами, следовательно, ранг матрицы X будет меньше числа оцениваемых параметров, матрица $X'X$ окажется вырожденной, и оценивание параметров модели с помощью МНК (см. (3.19)) будет невозможным. Причем при комбинировании «обычных» экономических переменных с фиктивными из-за ошибок округления в программных расчетах определитель матрицы $X'X$ может оказаться отличным от нуля, и вычисления будут выполнены, однако полученные оценки не будут соответствовать реальным экономическим процессам.

Рассмотрим возможности применения фиктивных переменных для моделирования сезонных колебаний. Пусть временной ряд квартальной динамики представлен в виде аддитивной модели, содержащей трендовую, сезонную и случайную компоненты. Пред-

положим, что тренд может быть описан линейной моделью. Для описания сезонных колебаний используем фиктивные переменные.

Тогда общий вид модели может быть записан следующим образом:

$$y_t = a + bt + c_2d_2 + c_3d_3 + c_4d_4 + \varepsilon_t, \quad (4.12)$$

где a, b, c_2, c_3, c_4 — коэффициенты модели;

$$d_i = \begin{cases} 1, & \text{если наблюдение принадлежит } i\text{-му кварталу;} \\ 0 & \text{в остальных случаях;} \end{cases}$$

$$i = 2, 3, 4.$$

В (4.12) 1-й квартал взят в качестве эталонной категории, а фиктивные переменные позволят оценить разницу в уровнях сезонности между эталонным кварталом и остальными. Предположив, что регрессионные остатки ε удовлетворяют всем необходимым требованиям, найдем с помощью МНК оценки коэффициентов a, b, c_2, c_3, c_4 .

Тогда полученное уравнение регрессии

$$\hat{y}_t = \hat{a} + \hat{b}t + \hat{c}_2d_2 + \hat{c}_3d_3 + \hat{c}_4d_4 \quad (4.13)$$

после соответствующей проверки значимости может быть использовано для описания сезонных колебаний в каждом квартале.

Регрессионная модель, описывающая динамику уровней ряда, относящихся к эталонному 1-му кварталу, примет вид:

$$\hat{y}_t = \hat{a} + \hat{b}t;$$

соответственно для наблюдений

$$\text{2-го квартала} \quad \hat{y}_t = \hat{a} + \hat{b}t + \hat{c}_2;$$

$$\text{3-го квартала} \quad \hat{y}_t = \hat{a} + \hat{b}t + \hat{c}_3;$$

$$\text{4-го квартала} \quad \hat{y}_t = \hat{a} + \hat{b}t + \hat{c}_4.$$

Переход из одного квартала в другой будет отражаться лишь в изменении свободного члена регрессионного уравнения и не будет касаться значения параметра $\hat{b}$, определяющего угол наклона линейного тренда и характеризующего средний абсолютный прирост уровней ряда под воздействием тенденции.

Найденные значения коэффициентов $\hat{c}_2, \hat{c}_3, \hat{c}_4$ позволяют оценить «сдвиги» в уровнях за счет фактора сезонности относительно 1-го, эталонного квартала. Можно усреднить четыре полученные линии регрессии:

$$\hat{y}_t = \hat{a} + \hat{b}t.$$

Тогда расстояние между отдельной регрессионной прямой для любого квартала и усредненной моделью даст оценку сезонных отклонений в этом квартале. Очевидно, что для аддитивной модели сумма сезонных отклонений будет равна нулю.

Пример 4.3

Рассмотрим использование этого подхода для моделирования сезонных колебаний временного ряда среднесуточного производства электроэнергии в Российской Федерации длиной $n = 20$. Показатель представлен данными квартальной динамики за период с 1994 по 1998 г. Графический анализ компонентного состава временного ряда свидетельствует о наличии тренда (в этот период наблюдалось устойчивое падение производства), сезонности и случайной составляющей. Сезонные колебания демонстрируют периодические подъемы в 1 и 4-м кварталах (осенне-зимний сезон) и спады во 2- и 3-м кварталах (весенне-летний сезон). Визуальный анализ динамики показателя позволяет предположить аддитивный характер сезонности, а в качестве модели — принять выражение (4.12).

Оценив коэффициенты модели, можем записать регрессионную зависимость производства электроэнергии от времени и фиктивных переменных:

$$\hat{y}_t = 2953,23 - 8,899t - 755,995d_2 - 904,973d_3 - 185,301d_4;$$

(с.о.)	(35,48)	(1,74)	(27,85)	(28,01)	(28,28)
t_n	(83,24)	(-5,12)	(-27,14)	(-32,30)	(-6,55)

$R^2 = 0,99; F(4,15) = 381,44.$

Под уравнением приведены статистические характеристики, свидетельствующие о значимости уравнения и его коэффициентов (с.о. — стандартные ошибки коэффициентов).

Запишем четыре регрессионных уравнения, характеризующих динамику уровней ряда для различных кварталов:

для наблюдений 1-го квартала $\hat{y}_t = 2953,23 - 8,899t$;
 для наблюдений 2-го квартала $\hat{y}_t = 2197,235 - 8,899t$;
 для наблюдений 3-го квартала $\hat{y}_t = 2048,257 - 8,899t$;
 для наблюдений 4-го квартала $\hat{y}_t = 2767,929 - 8,899t$.

Результаты моделирования наглядно представлены на рис. 4.3.

Оценка сезонных отклонений по сравнению с усредненной регрессией

$$\hat{y}_t = 2491,663 - 8,899t$$

в каждом квартале соответственно составила: в 1-м: +461,567; во 2-м: -294,428; в 3-м: -443,405; в 4-м: +276,266 (сумма сезонных отклонений равна нулю).

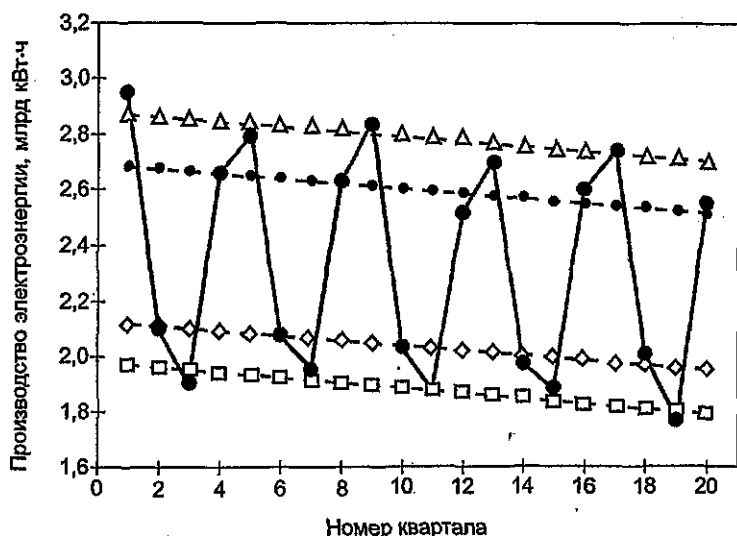


Рис. 4.3. Моделирование сезонных колебаний ряда среднесуточного производства электроэнергии в Российской Федерации с помощью фиктивных переменных:

● — фактический уровень; Δ, ◇, □, • — соответственно модели для уровней 1, 2, 3, 4-го кварталов

Несмотря на хорошее «качество» построенной модели на ретроспективном участке, она оказалась не совсем пригодной для целей прогнозирования. Это объяснялось инерционностью модели кривой роста, использованной для описания трендовой составляющей. Следующий год (1999) оказался переломным для производства электроэнергии, завершилось падение процесса производства, начался подъем. Однако выявленные особенности в динамике сезонных колебаний сохранились и в последующей динамике анализируемого ряда.

Таким образом, фиктивные переменные существенно расширяют возможности регрессионных моделей для описания динамики сложных экономических процессов.

4.4. Исследование периодических колебаний методами спектрального анализа

Важный подход к изучению структуры временных рядов, к анализу периодических составляющих — спектральный анализ. В настоящее время аппарат спектрального анализа базируется на фундаментальной математической теории, в основе которой лежит обоб-

щение разложения Фурье. Развитие этого подхода связано с трудами А.Н. Колмогорова, Н. Винера, Дж. Тьюки, А.Я. Хинчина и др.

Можно выделить три основных направления практического использования спектрального анализа [8]:

- получение полезных описательных статистик;
- важное средство диагностики;
- проверка постулируемых теоретических моделей.

Однако особенно эффективен спектральный метод при обобщении подхода на случай двух и более рядов. Метод позволяет оценить характеристики, отражающие тесноту связи между периодическими компонентами рядов, определить взаимные сдвиги (упреждение или запаздывание) для этих компонент, а также проследить во времени за изменением данных характеристик (если такие изменения значительны).

В [5, вып. 1] проводится аналогия между гистограммой, облегчающей выбор подходящего закона распределения, и построением выборочных оценок корреляционной функции и спектра, являющихся начальным этапом анализа временных рядов, полезным при дальнейшей идентификации нужного типа модели. Однако выборочный спектр легче поддается интерпретации, чем корреляционная функция, так как оценки спектра на соседних частотах приближенно независимы. При одновременном возникновении разнотипных колебаний коррелограмма теряет наглядность, потому что происходит суммирование или вычитание соответствующих автокорреляционных коэффициентов (коэффициенты автокорреляции наслаиваются друг на друга).

По сравнению с традиционными методами изучения сезонности спектральный анализ обладает рядом серьезных преимуществ. Спектральный анализ позволяет одновременно определить период колебания (частоту) различных периодических колебаний и интенсивность (амплитуду) этих колебаний. Многие же традиционные методы основаны на предположении, что характерные параметры колебания (например, период колебания) уже известны.

Кроме того, традиционные методы, в отличие от спектрального анализа, как правило, предполагают неизменность сезонной волны.

В настоящее время спектральный анализ стал полезным инструментом исследования во многих областях науки и техники. Рассмотрим некоторые практические приложения этого аппарата.

В [8, с. 17] отмечается: «Если под спектральным анализом понимать изучение индивидуальных частотных компонент интересующих нас временных рядов, то можно считать, что первое серь-

езное применение этой техники состоялось в 1664 г., когда Ньютон расщепил солнечный свет на отдельные компоненты, пропустив его через призму».

Действительно, после пропускания луча белого солнечного света через призму удастся разделить составляющие его волны разной длины. Аналогично применение спектрального анализа к временным рядам позволяет часто обнаружить периодические колебания разной длины, которые, на первый взгляд, выглядят как случайный шум.

Практические приложения методов спектрального анализа в физике, геофизике, биологии, медицине и других науках разнообразны. Наиболее известный пример — исследование солнечной активности (выявлена 11-летняя периодичность). Применялись эти методы для расшифровки записей землетрясений, подземных взрывов, для предсказания штормов, изучения структуры Луны, исследования биотоков мозга. Эти методы оказались полезны при конструировании моделей человеческого слуха, речевых сигналов. Таков далеко не полный перечень практических приложений методов спектрального анализа в прикладных областях естественных наук.

Имеются примеры успешного использования спектрального анализа в экономике (изучение колебаний в товарных ценах, в курсах ценных бумаг, периодических колебаний деловой активности и др.). Однако в экономике эти методы стали применяться значительно позднее и в недостаточном объеме, так как на пути их практического внедрения имелись следующие трудности:

- требовалась переоценка многих экономико-теоретических и методологических концепций, так как свойства временных рядов необходимо было рассматривать в непривычном для экономистов частотном, а не временном аспекте;

- большинство экономических временных рядов нестационарны в отличие от стационарных рядов, встречающихся в технике и естественных науках;

- в экономике в отличие от технических и естественных приложений спектрального анализа приходится иметь дело с единственной реализацией процесса, как правило, небольшой длины.

Дополнительным тормозом применения этого подхода долгое время оставалось отсутствие доступного программного обеспечения, ориентированного на решение широкого класса экономических задач. В современных ППП по статистической обработке данных (STATISTICA, SPSS и других) реализован представительный набор процедур спектрального анализа, позволяющих работать с временными рядами различной длины.

Перед применением процедур спектрального анализа к экономическим временным рядам часто встает проблема приведения их к стационарному виду. Как правило, в качестве методов фильтрации, позволяющих получить стационарность, используют последовательное дифференцирование (взятие разности), модели кривых роста (часто полиномиальный тренд) в сочетании со скользящими средними.

Исходный пункт применения спектрального анализа — оценивание выборочной спектральной плотности временных рядов. С технической точки зрения процесс получения оценок спектра наиболее трудоемок. Спектральная (автоспектральная) плотность или функция плотности спектра мощности $S_y(f)$ определяет распределение дисперсии временной последовательности y_t по частотам. Понятие спектральной плотности алогично понятию дисперсии в классической математической статистике.

Можно выделить три основных метода вычисления спектральных характеристик: *косвенный, прямой и смешанный*.

Косвенный метод вычисления основан на соотношениях Винера—Хинчина, позволяющих получить спектр как преобразование Фурье корреляционной (ковариационной) функции. В виде практической цифровой процедуры этот подход впервые был реализован Блэкменом и Тьюки¹.

Опираясь на дискретный вариант теоремы Винера—Хинчина и воспользовавшись четностью ковариационной функции, можно определить спектральную плотность как косинус-преобразование Фурье.

В качестве первоначальной оценки спектральной плотности $S''_y(f_i)$ для выборки длиной $N(N = 2n)$ рассмотрим выражение

$$S''_y(f_i) = 2 \left[C_y(0) + 2 \sum_{k=1}^{N-1} C_y(k) \cos(2\pi f_i k) \right], \quad (4.14)$$

где f_i — частота, $i = 1, 2, \dots, n$; $C_y(k)$ — ковариационная функция для лага k .

Однако Бартлетт (M.S. Bartlett)² показал, что (4.14) не является состоятельной оценкой спектральной плотности. Дисперсия этой оценки практически не зависит от объема выборки, и погрешность оценивания нельзя уменьшить путем увеличения объ-

¹ Blackman R.B., Tukey I.W. The measurement of power spectra from the point of view of communications engineering. — NY. Dover, 1958.

² Bartlett M.S. Smoothing periodograms from times series with continuous spectra // *Natura*. 1948. Vol. 161. — P. 686—687.

ема выборки. Состоятельную оценку получают применением процедуры сглаживания.

Сглаженную оценку можно записать в виде:

$$S_y(f_i) = 2 \left[C_y(0) + 2 \sum_{k=1}^{N-1} W(k) C_y(k) \cos(2\pi f_i k) \right]. \quad (4.15)$$

Весовую функцию $W(k)$ Блэкмен и Тьюки называли корреляционным окном, а ее преобразование Фурье — $H(f)$ -спектральным окном. Корреляционное окно $W(k)$ удовлетворяет следующим условиям:

$$0 \leq W(k) \leq 1; W(0) = 1; W(k) = W(-k); W(k) = 0, |k| > N. \quad (4.16)$$

На практике последнее условие заменяют на $W(k) = 0, |k| \geq L$, $L < N$, при этом нужно вычислять ковариации лишь до запаздывания L .

Процедура сглаживания искажает оценку, вносит в нее смещение, причем уменьшение значения точки отсечения L ведет к уменьшению дисперсии оценки и одновременному увеличению смещения. То есть, если взять L малым, то получим гладкую функцию спектральной плотности с незначительной дисперсией, при этом могут быть пропущены какие-то периодические составляющие. С ростом L уменьшается смещение пика от истинного значения частоты, но растет дисперсия («зазубренность») спектральной плотности.

На практике при выборе ширины полосы частот часто используют эмпирический метод оценивания спектров, получивший название «метод стягивания окна» (window closing) [10].

Суть метода состоит в вычислении сглаженных спектральных оценок сначала для широкой полосы частот (для малых значений точек отсечения), а затем для более узких полос частот. Основной вопрос состоит в том, когда остановить процесс стягивания окна. Практические результаты показывают, что L должно составлять от 10 до 40% длины временного ряда; чаще всего L принимают равным 20% N [6].

Прямой метод вычисления предполагает расчет спектральной плотности с помощью прямого преобразования Фурье исходной реализации. Долгое время этот метод не использовался на практике из-за громоздкости вычислений при получении дискретного преобразования Фурье:

$$Y(f_k) = T \sum_{i=0}^{N-1} y(i) \exp(-j \cdot 2\pi f_k i T), \quad (4.17)$$

где $Y(f_i)$

— преобразование Фурье временной последовательности $y(i)$;

$i = 0, 2, \dots, N-1$;

T

— временной интервал.

Однако появление в 1965 г. статьи Дж. Кули и Дж. Тьюки¹ о быстром методе вычисления дискретного преобразования Фурье открыло путь к широкому применению прямого подхода при вычислении спектров. Позже метод Кули и Тьюки был развит и стал широко известен как быстрое преобразование Фурье (БПФ). Ценность этого метода заключается в сокращении времени вычисления дискретного преобразования Фурье для большинства практических задач на один—два порядка.

Быстрое преобразование Фурье основывается на представлении N в виде ряда сомножителей и выполнении обычного преобразования Фурье для более коротких последовательностей, число членов в которых определяется соответствующими сомножителями. Для расчета всех значений по формуле (4.17) необходимо выполнить примерно N^2 операций умножения и сложения комплексных чисел. Отметим, что одна такая комплексная операция эквивалентна четырём операциям умножения и сложения действительных чисел.

Если N может быть представлено в виде произведения p целых и больших единицы чисел $N = r_1 r_2 \dots r_p$, то, как доказано в [4], последовательность $Y(f_k)$ в (4.17) может быть найдена итеративно, при этом будет выполнено только $4N \sum_{i=0}^p r_i$ операций с действительными числами.

С точки зрения программирования наиболее удобен случай $N = 2^p$ (p — целое), хотя сейчас имеются программы алгоритмов для N , равного произведению степеней 2, 3, 4 и 5, а также для обработки временных рядов произвольной длины N , причем скорость вычислений возрастает, если N не простое, а составное. На практике, как правило, ограничиваются программами для случая $N = 2^p$. Если длина ряда данных не равна степени 2, то ряд укорачивают или дополняют нулями.

Смешанный метод основан на использовании БПФ для получения ковариационной функции и дальнейшем применении метода Блэкмана—Тьюки. Такой подход основан на уже упоминавшейся теореме Винера—Хинчина, согласно которой спектральная плотность и ковариационная функция связаны между собой пре-

¹ Cooley J.W., Tukey J.M. An algorithm for the machine calculation of complex Fourier series. Mathematics of computation. 1965. Vol. 19. — P. 297—301.

образованием Фурье. Следовательно, ковариационную функцию можно определить с помощью быстрого обратного преобразования Фурье спектра, вычисленного приближенно.

Вычислительная эффективность прямого метода по сравнению с косвенным будет расти по мере увеличения длины исследуемых рядов. Большую экономию машинного времени можно получить при использовании этого метода для обработки рядов, состоящих из нескольких сотен, тысяч значений.

Согласно [4], вычисление спектральной плотности для ряда длиной N (где $N = 2^p$, p — целое) косвенным методом требует примерно NL операций (L — максимальное запаздывание), прямым методом — примерно $4Nr$ операций. Тогда коэффициент ускорения вычислений $K = L/4r$.

В табл. 4.5 проводится сравнение числа операций при расчетах спектральных плотностей прямым и косвенным методами для различных N . В соответствии с [6] L принято равным $0,2N$.

Таблица 4.5. Сравнение числа операций при расчетах спектральных плотностей прямым и косвенным методами

Длина ряда $N = 2^p$	Прямой метод ($4Nr$)	Косвенный метод ($0,2N^2$)	Коэффициент ускорения вычислений ($0,2N/4r$)
$2^6 = 64$	1536	819	0,5
$2^7 = 128$	3584	3277	0,9
$2^8 = 256$	8192	13107	1,6
$2^9 = 512$	18432	52429	2,8
$2^{10} = 1024$	40960	209715	5,1
$2^{11} = 2048$	90112	838861	9,3
$2^{12} = 4096$	196608	3355443	17,1

Смешанный метод вычисления спектров может оказаться более эффективным по сравнению с косвенным — это зависит от максимального запаздывания. Коэффициент ускорения вычислений при сравнении этих методов для нахождения спектральной плотности определен в [4] как $L/8r$. Например, при $N = 2^{13}$, $L = 0,2N$ использование БПФ позволяет ускорить расчет примерно в 16 раз. Приведенные сведения о коэффициентах ускорения вычислений носят приближенный характер. При использовании специальных программ можно добиться дополнительного ускорения вычислений.

Для длинных рядов использование прямого или смешанного метода безусловно эффективнее. Однако экономические временные ряды, представленные месячными, квартальными, годовыми данными, обладают, как правило, небольшой длиной. Для коротких же рядов обращение к сложной процедуре БПФ нецелесообразно. В то же время, например, временные ряды, описывающие динамику показателей развития фондовых рынков (фондовые индексы, курсы ценных бумаг и другие) могут иметь достаточно длинную историю.

Наличие хорошего программного обеспечения — важное условие успешного использования спектрального анализа. В современных ППП представлены различные вычислительные алгоритмы, позволяющие обрабатывать временные ряды произвольной длины. Например, в ППП STATISTICA легко реализуются алгоритмы спектрального анализа для рядов, имеющих даже более 100 000 наблюдений. При этом для рядов, не являющихся степенью 2, используется модификация алгоритма БПФ [7].

Аппарат спектрального анализа можно рассматривать как важный инструмент исследования периодических составляющих, тонкий индикатор, позволяющий выявлять их наличие в анализируемых временных рядах.

Пример 4.4

На рис. 4.4 показана спектральная плотность временного ряда производства электроэнергии.

Исходная информация была представлена данными месячной динамики за 8 лет (с 1992 г.). Перед применением спектрального анализа ряд был предварительно преобразован. Спектр демонстрирует острый выброс на частоте $1/12$ (0,0833), что соответствует годовой (12-месячной) периодичности. Наличие этого всплеска объясняется ежегодным увеличением производства в осенне-зимний период и снижением в весенне-летнее время. Для того чтобы изменчивость спектра могла быть выражена более удобным образом, часто спектральные оценки строят в логарифмическом масштабе. Логарифмирование растягивает графики спектральных плотностей там, где их значения малы, и позволяет лучше изучить поведение спектра на соответствующих частотах. На рис. 4.5 показан логарифм спектральной плотности. Дисперсия процесса концентрируется около сезонных частот $1/12$, $1/6$, $1/4$, $1/3$, $5/12$, $1/2$, соответствующих колебаниям с периодами 12, 6, 4, 3, $12/5$ и 2 месяца, наиболее существенный всплеск (максимальное значение спектральной плотности) — на частоте $1/12$, остальные колебания слабо выражены.

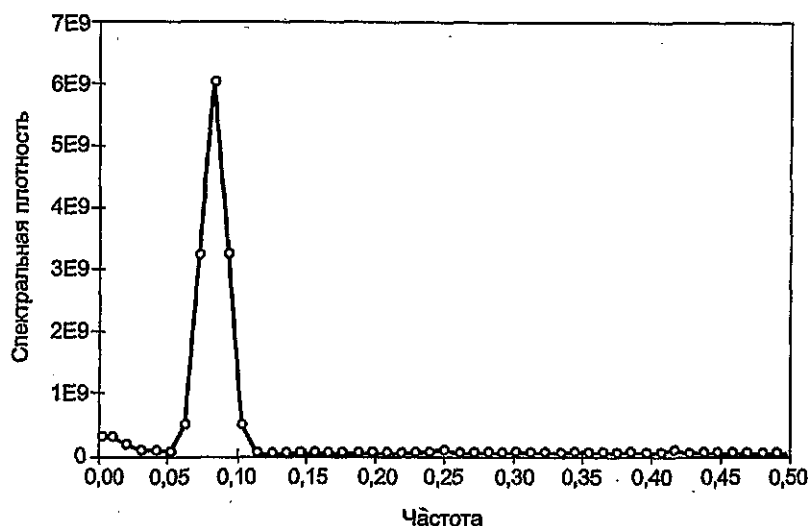


Рис. 4.4. Спектральная плотность временного ряда производства электроэнергии (тренд удален)

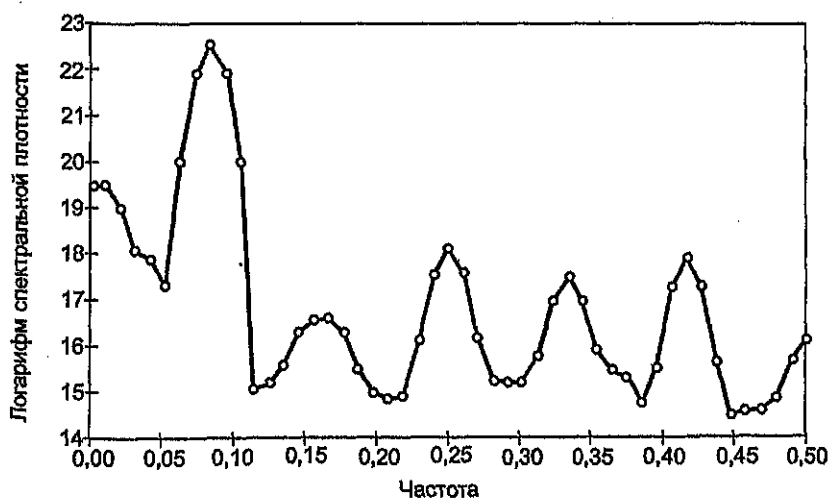


Рис. 4.5. Логарифм спектральной плотности временного ряда производства электроэнергии (тренд удален)

На рис. 4.6 показан спектр временного ряда производства электроэнергии после проведения сезонной корректировки. Отметим, что спектральный анализ может использоваться как для установления необходимости выполнения процесса корректировки, так и

для оценки качества различных методов снятия сезонности. В [27] отмечается: «хорошую сезонную корректировку можно определить как фильтр, который не пропускает мощности в окрестностях этих сезонных частот и пропускает мощность без суммирования в остальных значениях спектра». На рис. 4.6 видна концентрация мощности на низких частотах (близких к нулю), в то же время отсутствуют всплески на сезонных частотах (даже просматриваются небольшие провалы). Такие графики спектральных плотностей могут оказаться полезными при сравнительном анализе различных методов сезонной декомпозиции и корректировки.

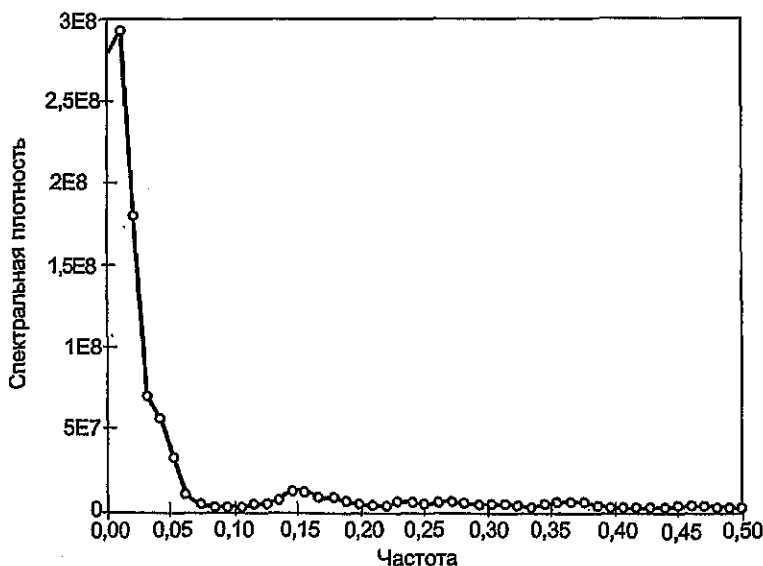
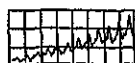


Рис. 4.6. Спектральная плотность временного ряда производства электроэнергии после сезонной корректировки



Практикум к главе 4

Контрольные вопросы

1. Поясните алгоритм оценивания сезонной составляющей для аддитивной модели.
2. В чем состоят отличия подходов к оцениванию сезонной составляющей в случае мультипликативного и аддитивного характера сезонности?
3. Чему равна сумма оценок коэффициентов сезонности для полного сезонного цикла (характер сезонности — мультипликативный)? Чему равна сумма оценок значений сезонной составляющей для полного сезонного цикла (характер сезонности — аддитивный)?

4. В чем суть десезонализации (сезонной корректировки) временных рядов?
5. Какие известны программы, реализующие процесс сезонной корректировки?
6. Для каких целей в регрессионные уравнения вводятся фиктивные переменные?
7. Поясните, каким образом можно моделировать сезонные колебания с помощью фиктивных переменных.
8. Для каких ситуаций используется термин *dummy trap* (ловушка)?
9. Укажите основные направления практического использования спектрального анализа.
10. Какие существуют подходы к вычислению спектральных характеристик?

Упражнения

- 4.1. В таблице представлены данные о квартальной динамике расходов потребителей на газ и электричество в США (в постоянных ценах, млрд долл.) [11].

Год	Расходы по кварталам			
	1	2	3	4
1977	7,33	4,70	5,10	5,46
1978	7,56	4,92	5,15	5,55
1979	7,96	5,01	5,05	5,59
1980	7,74	5,10	5,67	5,92
1981	8,04	5,27	5,51	6,04
1982	8,26	5,51	5,41	5,83

Рассчитайте прогноз показателя в 1-м квартале 1983 г., используя трендовую линейную модель в сочетании с фиктивными переменными для описания сезонных колебаний. Сравните точность моделирования с результатами, полученными при использовании адаптивных моделей (см. Практикум к гл. 5, упражнение 4).

- 4.2. В таблице представлены данные о квартальной динамике прибыли компании. Проведите графический анализ компонентного состава временного ряда. Рассчитайте оценки коэффициентов сезонности, предположив ее мультипликативный характер. Проведите интерпретацию полученных результатов.

Год	Прибыль по кварталам (тыс. долл.)			
	1	2	3	4
1	181	207	198	172
2	179	199	188	166
3	171	187	176	156
4	161	167	158	138

4.3. По данным упражнения 2 определите прогнозную оценку прибыли компании в первом полугодии следующего года с помощью тренд-сезонной модели.

Используйте оценки коэффициентов сезонности, рассчитанные в предыдущем упражнении.

4.4. На основе квартальных данных об объемах продаж продукции фирмы, тыс. шт., за пять лет была построена тренд-сезонная модель. Сезонность носила мультипликативный характер.

Оценки коэффициентов сезонности представлены в таблице.

Квартал	1	2	3	4
Коэффициент сезонности	?	1,15	1,25	0,71

Требуется:

а) определить значение коэффициента сезонности для первого квартала;

б) рассчитать прогнозную оценку уровня продаж в первом полугодии следующего года, если уравнение тренда имеет вид $\hat{y}_t = 15,2 + 0,15t$ ($t = 1, 2, \dots, 20$).

4.5. На основе квартальных данных об объемах продаж продукции фирмы, тыс. шт., за пять лет была построена тренд-сезонная модель. Сезонность носила аддитивный характер.

Оценки сезонной составляющей представлены в таблице.

Квартал	1	2	3	4
Сезонная составляющая	?	-0,12	0,17	0,20

Требуется:

а) определить значение сезонной составляющей для первого квартала;

б) рассчитать прогнозную оценку уровня продаж в первом полугодии следующего года, если уравнение тренда имеет вид: $\hat{y}_t = 11,56 + 0,32t$ ($t = 1, 2, \dots, 20$).

4.6. На основе квартальных данных о производстве продукции фирмы, тыс. шт., за шесть лет были оценены коэффициенты модели. Модель, содержащая линейный тренд и сезонные фиктивные переменные для 2-го (d_2), 3-го (d_3), 4-го (d_4) кварталов имела вид:

$$\hat{y}_t = 13,75 - 0,32t + 4,08d_2 + 3,02d_3 + 3,01d_4.$$

(с.о) (0,65) (0,04) (0,75) (0,73) (0,73)

В скобках указаны стандартные ошибки коэффициентов (с.о). Дайте интерпретацию полученной модели.



Глава

ИСПОЛЬЗОВАНИЕ АДАПТИВНЫХ МЕТОДОВ ПРОГНОЗИРОВАНИЯ В ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЯХ

5.1. Сущность адаптивных методов

При использовании традиционных подходов и методов для прогнозирования важнейших экономических показателей на макро-, мезо- и микроуровнях часто выдвигается гипотеза о том, что основные тенденции и факторы, выявленные на предыстории, сохраняются и для периода упреждения (на прогнозируемом периоде). Таким образом, процесс экстраполяции выявленных закономерностей, тенденций базируется на предположении об инерционности анализируемых экономических систем.

В последнее время в процессе коренных социально-экономических преобразований подвижность этих систем возрастает. Наблюдаются существенные изменения в развитии промышленного комплекса РФ, возрастает быстрота реакции на конъюнктуру внешнего и внутреннего рынков, на правительственные решения, на новые социально-экономические условия.

Очевидны структурные сдвиги по многим важнейшим показателям экономического развития. Даже наиболее инерционные макроэкономические характеристики становятся более подвижными. В связи с этим для прогнозирования таких сложных процессов требуется гибкий и современный статистический инструментарий.

В настоящее время одним из наиболее перспективных направлений исследования и прогнозирования одномерных временных рядов считаются адаптивные методы.

Термин *адаптация* происходит от лат. *adaptatio* — приспособление. В биологии это слово означает совокупность различных особенностей (морфологических, поведенческих и других) биологического вида, обеспечивающих приспособление к определенным условиям существования, к специфическим особенностям внешней среды. Адаптацией также называется и сам процесс выработки приспособлений. Применительно к прогнозированию процесс адаптации состоит в следующем.

При обработке временных рядов, как правило, наиболее ценной бывает информация последнего периода, так как необходимо

знать, как будет развиваться тенденция, существующая в данный момент, а не тенденция, сложившаяся в среднем на всем рассматриваемом периоде. Адаптивные методы позволяют учесть различную информационную ценность уровней временного ряда, степень «устаревания» данных.

Прогнозирование методом экстраполяции на основе кривых роста в какой-то мере тоже содержит элемент адаптации, поскольку с получением «свежих» фактических данных параметры кривых пересчитываются заново. Поступление новых данных может привести и к замене выбранной ранее кривой на другую модель. Однако степень адаптации в данном случае весьма незначительна, кроме того, она падает с ростом длины временного ряда, так как при этом уменьшается «весомость» каждой новой точки. В адаптивных методах различную ценность уровней в зависимости от их «возраста» можно учесть с помощью системы весов, придаваемых этим уровням.

Оценивание коэффициентов адаптивной модели обычно осуществляется на основе рекуррентного метода, который формально отличается от МНК, метода максимального правдоподобия и других методов тем, что не требует повторения всего объема вычислений при появлении новых данных.

Важнейшее достоинство адаптивных методов — построение самокорректирующихся моделей, способных учитывать результат прогноза, сделанного на предыдущем шаге. Пусть модель находится в некотором состоянии, для которого определены текущие значения ее коэффициентов. На основе этой модели делается прогноз. При поступлении фактического значения оценивается ошибка прогнозного значения (разница между этим значением и полученным по модели). Ошибка прогнозирования через обратную связь поступает в модель и учитывается в ней в соответствии с принятой процедурой перехода от одного состояния в другое. В результате вырабатываются «компенсирующие» изменения, состоящие в корректировании параметров в целях большего согласования поведения модели с динамикой ряда. Затем рассчитывается прогнозная оценка на следующий момент времени, и весь процесс повторяется вновь.

Таким образом, адаптация осуществляется итеративно с получением каждой новой фактической точки ряда. Модель постоянно «впитывает» новую информацию, приспосабливается к ней и поэтому отражает тенденцию развития, существующую в данный момент. На рис. 5.1 приведена общая схема построения адаптивных моделей прогнозирования [21].

Быстроту реакции модели на изменения в динамике процесса характеризует так называемый параметр адаптации. Параметр адаптации должен быть выбран таким образом, чтобы обеспечивалось адекватное отображение тенденции при одновременной фильтрации случайных отклонений. Значение параметра адаптации может быть определено на основе эмпирических данных, выведено аналитическим способом или получено на основе метода проб.

В качестве критерия оптимальности при выборе параметра адаптации обычно принимают критерий минимума среднего квадрата ошибок прогнозирования.

На основе рассмотренных особенностей дадим определение группы методов прогнозирования, объединенных общим названием адаптивные.

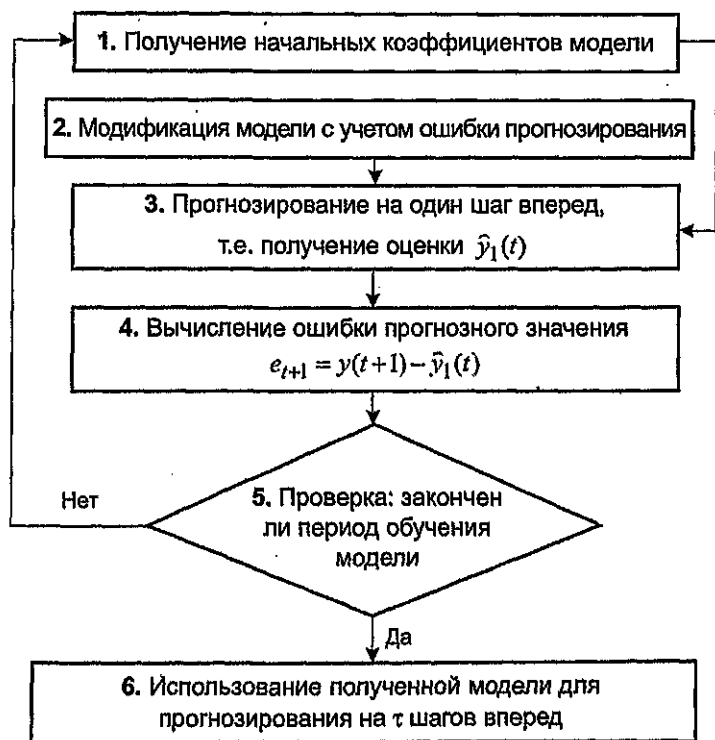


Рис. 5.1. Схема построения адаптивных моделей прогнозирования: $y(t)$ — фактические уровни временного ряда; $\hat{y}_\tau(t)$ — прогноз, сделанный в момент t на τ единиц времени (шагов) вперед; e_{t+1} — ошибка прогноза, полученная как разница между фактическим и прогнозным значением показателя в точке $(t+1)$

Адаптивными называются методы прогнозирования, позволяющие строить самокорректирующиеся (самонастраивающиеся) экономико-математические модели, которые способны оперативно реагировать на изменение условий путем учета результата прогноза, сделанного на предыдущем шаге, и учета различной информационной ценности уровней ряда. Благодаря отмеченным свойствам адаптивные методы особенно удачно используются при оперативном, краткосрочном прогнозировании. Указанное определение отражает основные характерные черты, присущие рассматриваемому подходу. В то же время деление на адаптивные и неадаптивные модели часто носит достаточно условный характер.

У истоков адаптивных методов лежит модель экспоненциального сглаживания.

5.2. Экспоненциальное сглаживание

Предположим, что модель временного ряда имеет вид:

$$y_t = a_1 + \varepsilon_t$$

где $a_1 = \text{const}$;

ε_t — случайные неавтокоррелированные отклонения с нулевым математическим ожиданием и дисперсией σ^2 .

Для экспоненциального сглаживания ряда используется рекуррентная формула:

$$S_t = \alpha y_t + \beta S_{t-1}, \quad (5.1)$$

где S_t — значение экспоненциальной средней в момент t ;

α — параметр сглаживания, $\alpha = \text{const}$, $0 < \alpha < 1$;

$\beta = 1 - \alpha$.

Если последовательно использовать соотношение (5.1), то экспоненциальную среднюю S_t можно выразить через предшествующие значения уровней временного ряда:

$$\begin{aligned} S_t &= \alpha y_t + \beta S_{t-1} = \alpha y_t + \beta(\alpha y_{t-1} + \beta S_{t-2}) = \\ &= \alpha y_t + \alpha \beta y_{t-1} + \beta^2 S_{t-2} = \dots = \alpha y_t + \alpha \beta y_{t-1} + \alpha \beta^2 y_{t-2} + \dots + \alpha \beta^n y_{t-n} + \dots + \beta^n S_0. \end{aligned}$$

Таким образом,

$$S_t = \alpha \sum_{i=0}^{n-1} \beta^i y_{t-i} + \beta^n S_0, \quad (5.2)$$

где n — длина ряда.

При $n \rightarrow \infty$ $\beta^n \rightarrow 0$, следовательно,

$$S_t = \alpha \sum_{i=0}^{\infty} \beta^i y_{t-i}. \quad (5.3)$$

Таким образом, величина S_t оказывается взвешенной суммой всех членов ряда. Причем веса отдельных уровней ряда убывают по мере их удаления в прошлое соответственно экспоненциальной функции (в зависимости от «возраста» наблюдений). Именно поэтому величина S_t названа экспоненциальной средней.

Например, пусть $\alpha = 0,3$. Тогда вес текущего наблюдения y_t будет равен $\alpha = 0,3$, вес предыдущего уровня y_{t-1} будет соответствовать $\alpha \cdot \beta = 0,3 \cdot 0,7 = 0,21$; для уровня y_{t-2} вес составит $\alpha \cdot \beta^2 = 0,147$; для $y_{t-3} - \alpha \cdot \beta^3 = 0,1029$ и т.д. [17].

Автор модели английский математик Р. Браун показал, что математические ожидания временного ряда и экспоненциальной средней совпадут, но в то же время дисперсия экспоненциальной средней $D[S_t]$ меньше дисперсии временного ряда σ^2 .

Представим выражение (5.3) в следующем виде:

$$S_t = \alpha \sum_{i=0}^{\infty} \beta^i y_{t-i} = \alpha \sum_{i=0}^{\infty} \beta^i (a_1 + \varepsilon_{t-i}) = a_1 + \alpha \sum_{i=0}^{\infty} \beta^i \varepsilon_{t-i}.$$

Отсюда очевидно, что математическое ожидание $M(S_t) = a_1$, так же как и математическое ожидание самого временного ряда.

Определим дисперсию экспоненциальной средней $D[S_t]$:

$$D[S_t] = M[(S_t - a_1)^2] = M \left[\left(\alpha \sum_{i=0}^{\infty} \beta^i \varepsilon_{t-i} \right)^2 \right].$$

Учитывая свойства ε_t , можно записать:

$$D[S_t] = \alpha^2 \sum_{i=0}^{\infty} \beta^{2i} \sigma^2 = \frac{\alpha}{2-\alpha} \sigma^2.$$

Таким образом,

$$D[S_t] = \frac{\alpha}{2-\alpha} \sigma^2. \quad (5.4)$$

Так как $0 < \alpha < 1$, то $D[S_t]$ меньше дисперсии временного ряда, равной σ^2 .

Из (5.4) видно, что при высоком значении α дисперсия экспоненциальной средней незначительно отличается от дисперсии ряда. С уменьшением α дисперсия экспоненциальной средней сокращается, возрастает ее отличие от дисперсии ряда. Тем самым экспоненциальная средняя начинает играть роль «фильтра», поглощающего колебания временного ряда.

Таким образом, с одной стороны, следует увеличивать вес более свежих наблюдений, что может быть достигнуто повышением

α , с другой стороны, для сглаживания случайных отклонений значение α нужно уменьшить. Эти два требования находятся в противоречии. Поиск компромиссного значения параметра сглаживания α составляет задачу оптимизации модели.

При расчете экспоненциальной средней в момент времени всегда требуется значение экспоненциальной средней в предыдущий момент времени, поэтому на первом шаге должно быть определено некоторое значение S_0 , предшествующее S_1 . Часто на практике в качестве начального значения S_0 используется среднее арифметическое значение из всех имеющихся уровней временного ряда или из какой-то их части. Из выражения (5.2) следует, что вес, приписываемый этому значению, уменьшается по экспоненциальной зависимости по мере удаления от первого уровня. Поэтому для длинных временных рядов влияние неудачного выбора S_0 погашается. Однако при малых выборках это может привести к существенным ошибкам, так как, например, при $\alpha = 0,1$ даже после 20 итераций вес начального значения S_0 превышает веса других уровней ряда. Для элиминирования избыточного веса, придаваемого S_0 , Р. Вейд предложил модифицировать процедуру¹.

Пусть $S'_0 = \alpha S_0$, тогда $S'_1 = \alpha y_1 + (1-\alpha)S'_0 = \alpha y_1 + (1-\alpha)\alpha S_0$.

Так как весовые коэффициенты в сумме теперь не дают единицу, то вводится дополнительный множитель, равный обратной величине суммы коэффициентов:

$$\tilde{S}_1 = S'_1 \frac{1}{\alpha + \alpha(1-\alpha)} = \frac{1}{\alpha + \alpha(1-\alpha)} [\alpha y_1 + (1-\alpha)\alpha S_0].$$

Тогда на первой итерации при $\alpha = 0,1$ вес текущего уровня y_1 определяется выражением $\frac{1}{2-\alpha} = 0,526$, а вес S_0 уже равен меньшему значению $\frac{1-\alpha}{2-\alpha} = 0,474$.

Если продолжить этот подход для следующей итерации, то получим

$$S'_2 = \alpha y_2 + (1-\alpha)S'_1 = \alpha y_2 + (1-\alpha)[\alpha y_1 + \alpha(1-\alpha)S_0];$$

$$\tilde{S}_2 = S'_2 \frac{1}{\alpha + (1-\alpha)\alpha + (1-\alpha)^2 \alpha}.$$

Соответственно вес текущего уровня y_2 определяется выражением

¹ Wade R.C. A technique for initializing exponential smoothing forecasts // Management Science. 1967. Vol.13. N. 7.

$$\frac{1}{1+(1-\alpha)+(1-\alpha)^2} = 0,369,$$

вес предыдущего уровня y_1 равен

$$\frac{1-\alpha}{1+(1-\alpha)+(1-\alpha)^2} = 0,332,$$

а вес S_0 равен меньшему значению:

$$\frac{(1-\alpha)^2}{1+(1-\alpha)+(1-\alpha)^2} = 0,299.$$

Таким образом, с помощью рассмотренного подхода удалось убрать избыточный вес, придаваемый начальному значению S_0 , и распределить его пропорционально по всем уровням ряда. При использовании модифицированной модели прогнозы даже при коротких временных рядах будут в большей степени основаны на фактических данных, чем на предварительной оценке S_0 . Р. Вейд разработал рекомендации, позволяющие при заданном значении параметра адаптации на определенной итерации возвращаться к обычной модели.

При использовании экспоненциальной средней для краткосрочного прогнозирования предполагается, что модель ряда имеет вид:

$$y_t = a_{1,t} + \varepsilon_t, \quad (5.5)$$

где $a_{1,t}$ — варьирующий во времени средний уровень ряда;

ε_t — случайные неавтокоррелированные отклонения с нулевым математическим ожиданием и дисперсией σ^2 .

Прогнозная модель определяется равенством

$$\hat{y}_\tau(t) = \hat{a}_{1,t}, \quad (5.6)$$

где $\hat{y}_\tau(t)$ — прогноз, сделанный в момент t на τ единиц времени (шагов) вперед;

$\hat{a}_{1,t}$ — оценка $a_{1,t}$.

Единственный параметр модели $\hat{a}_{1,t}$ определяется экспоненциальной средней:

$$\begin{aligned} \hat{a}_{1,t} &= S_t; \\ \hat{a}_{1,0} &= S_0. \end{aligned}$$

Выражение (5.1.) можно представить по-другому, перегруппировав члены:

$$S_t = S_{t-1} + \alpha(y_t - S_{t-1}). \quad (5.7)$$

Величину $(y_t - S_{t-1})$ можно рассматривать как погрешность прогноза. Тогда новый прогноз S_t получается в результате корректировки предыдущего прогноза с учетом его ошибки. В этом и состоит адаптация модели.

При краткосрочном прогнозировании желательно как можно быстрее отразить изменения ряда и в то же время очистить ряд, отфильтровав случайные колебания. Для этого величине α следует присваивать одно из промежуточных значений в интервале от 0 до 1. Если в результате экспериментальных расчетов получено наилучшее значение α , близкое к 1, то целесообразно проверить правомерность выбора модели данного типа. Р. Браун рекомендовал брать значения α в пределах 0,1—0,3. Во многих работах некритически повторяются эти рекомендации. Однако опыт практических исследований экономических временных рядов свидетельствует о том, что часто эти значения далеки от оптимальных.

Иногда поиск этого значения параметра осуществляется путем перебора на сетке значений. В этом случае в качестве оптимального выбирается то значение α , при котором получена наименьшая дисперсия ошибки. В большинстве эконометрических пакетов, например «Мезозавр», SPSS, STATISTICA и других, при построении этих моделей в меню предусмотрена ветвь «оптимизация», реализующая поиск значения по этой схеме.

Следует отметить, что выбор значения параметра α должен зависеть от периода упреждения прогноза. Для оперативных, конъюнктурных прогнозов в большей степени должна учитываться свежая информация, поэтому значение α следует брать большим. При увеличении срока прогнозирования более поздняя информация, последние данные должны иметь несколько меньший вес, конъюнктурные колебания должны быть сглажены, но учтены прошлые уровни. Для этих целей значение α следует уменьшить.

Экспоненциальное сглаживание — это пример самообучающейся модели. К ее безусловным достоинствам относится чрезвычайная простота вычислений, выполняемых итеративно, причем массив прошлой информации уменьшен до единственного значения S_{t-1} .

Пример 5.1

Рассчитайте экспоненциальную среднюю для временного ряда курса акций фирмы IBM (табл. 5.1) [17].

В качестве начального значения экспоненциальной средней возьмите среднее значение из пяти первых уровней ряда. Расчеты проведите для двух различных значений параметров адаптации α : а) $\alpha = 0,1$; б) $\alpha = 0,5$.

Сравните графически исходный временной ряд и ряды экспоненциальных средних, полученные при $\alpha = 0,1$ и $\alpha = 0,5$. Укажите, какой ряд носит более гладкий характер.

Таблица 5.1. Курс акций фирмы IBM, долл. США

t	y_t	t	y_t	t	y_t
1	510	11	494	21	523
2	497	12	499	22	527
3	504	13	502	23	523
4	510	14	509	24	528
5	509	15	525	25	529
6	503	16	512	26	538
7	500	17	510	27	539
8	500	18	506	28	541
9	500	19	515	29	543
10	495	20	522	30	541

Решение

1. Определим $S_0 = \frac{1}{5} \sum_{t=1}^5 y_t = \frac{1}{5} (510 + 497 + 504 + 510 + 509) = 506$.

Найдем значения экспоненциальной средней при $\alpha = 0,1$.

$$S_t = \alpha x_t + (1 - \alpha) S_{t-1}; \quad \alpha = 0,1 \text{ — по условию;}$$

$$S_1 = \alpha x_1 + (1 - \alpha) S_0; \quad S_1 = 0,1 \cdot 510 + 0,9 \cdot 506 = 506,4;$$

$$S_2 = \alpha x_2 + (1 - \alpha) S_1; \quad S_2 = 0,1 \cdot 497 + 0,9 \cdot 506,4 = 505,46;$$

$$S_3 = \alpha x_3 + (1 - \alpha) S_2; \quad S_3 = 0,1 \cdot 504 + 0,9 \cdot 505,46 = 505,31 \text{ и т.д.}$$

Результаты расчетов представлены в табл. 5.2.

Проведем аналогичные расчеты для $\alpha = 0,5$.

$$S_1 = \alpha x_1 + (1 - \alpha) S_0; \quad S_1 = 0,5 \cdot 510 + 0,5 \cdot 506 = 508;$$

$$S_2 = \alpha x_2 + (1 - \alpha) S_1; \quad S_2 = 0,5 \cdot 497 + 0,5 \cdot 508 = 502,5 \text{ и т.д.}$$

Результаты расчетов также представлены в табл. 5.2.

На рис. 5.2. наглядно проявляется влияние значения параметра адаптации α на характер сглаженного ряда.

При $\alpha = 0,1$ экспоненциальная средняя носит более гладкий характер, так как в этом случае в наибольшей степени поглощаются случайные колебания временного ряда.

Таблица 5.2. Экспоненциальные средние

t	Экспоненциальная средняя		t	Экспоненциальная средняя	
	$\alpha = 0,1$	$\alpha = 0,5$		$\alpha = 0,1$	$\alpha = 0,5$
1	506,4	508,0	16	505,7	513,3
2	505,5	502,5	17	506,1	511,7
3	505,3	503,2	18	506,1	508,8
4	505,8	506,6	19	507,0	511,9
5	506,1	507,8	20	508,5	517,0
6	505,8	505,4	21	509,9	520,0
7	505,2	502,7	22	511,6	523,5
8	504,7	501,4	23	512,8	523,2
9	504,2	500,7	24	514,3	525,6
10	503,3	497,8	25	515,8	527,3
11	502,4	495,9	26	518,0	532,7
12	502,0	497,5	27	520,1	525,8
13	502,0	499,7	28	522,2	538,4
14	502,7	504,4	29	524,3	540,7
15	505,0	514,7	30	525,9	540,9

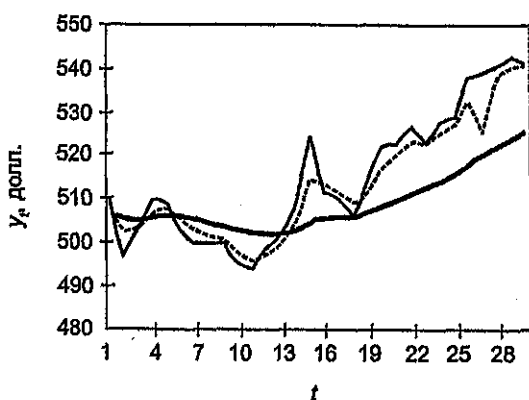


Рис. 5.2. Экспоненциальное сглаживание временного ряда курса акций:
 — фактические данные; — экспоненциальная средняя при $\alpha = 0,1$; — экспоненциальная средняя при $\alpha = 0,5$

5.3. Адаптивные полиномиальные модели

Если для прогнозирования временного ряда, имеющего ярко выраженную линейную тенденцию, использовать подход (5.6), опирающийся на модель экспоненциального сглаживания, то модель, как правило, будет давать смещенные прогнозы, т.е. систематическую ошибку. Для таких временных рядов целесообразно использовать модели линейного роста, также применяющие процедуру экспоненциального сглаживания.

В этих моделях прогноз может быть получен с помощью следующего выражения:

$$\hat{y}_\tau(t) = \hat{a}_{1,t} + \hat{a}_{2,t}\tau, \quad (5.8)$$

где $\hat{a}_{1,t}$ и $\hat{a}_{2,t}$ — текущие оценки коэффициентов;
 τ — время упреждения прогноза.

В табл. 5.3 представлены три модели данного типа: двухпараметрическая модель Ч. Хольта, однопараметрическая модель Р. Брауна и трехпараметрическая модель Дж. Бокса и Г. Дженкинса, отличающиеся рекуррентными выражениями для пересчета текущих оценок коэффициентов (параметры адаптации или параметры экспоненциального сглаживания $0 < \alpha_1, \alpha_2, \alpha_3, \beta < 1$).

Таблица 5.3. Модели линейного роста для расчета прогноза

$$\hat{y}_\tau(t) = \hat{a}_{1,t} + \hat{a}_{2,t}\tau$$

Название модели	Оценка коэффициентов
Модель Ч. Хольта	$\hat{a}_{1,t} = \alpha_1 y_t + (1 - \alpha_1)(\hat{a}_{1,t-1} + \hat{a}_{2,t-1})$ $\hat{a}_{2,t} = \alpha_2 (\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \alpha_2) \hat{a}_{2,t-1}$
Модель Р. Брауна	$\hat{a}_{1,t} = \hat{a}_{1,t-1} + \hat{a}_{2,t-1} + (1 - \beta^2) e_t$ $\hat{a}_{2,t} = \hat{a}_{2,t-1} + (1 - \beta^2) e_t$
Модель Дж. Бокса и Г. Дженкинса	$\hat{a}_{1,t} = \alpha_1 y_t + (1 - \alpha_1)(\hat{a}_{1,t-1} + \hat{a}_{2,t-1}) + \alpha_3 (e_t - e_{t-1})$ $\hat{a}_{2,t} = \alpha_2 (\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \alpha_2) \hat{a}_{2,t-1}$

В эконометрических пакетах чаще представлена модель Ч. Хольта с возможностью выбора оптимальных параметров по критерию минимума среднеквадратической ошибки путем перебора на сетке возможных значений. Рекуррентные формулы для оценки коэффициентов по этой модели могут быть преобразованы к следующему виду, явно показывающему зависимость «корректирующего воздействия» от величины ошибки:

$$\begin{aligned}\hat{a}_{1,t} &= \hat{a}_{1,t-1} + \hat{a}_{2,t-1} + \alpha_1 e_t; \\ \hat{a}_{2,t} &= \hat{a}_{2,t-1} + \alpha_1 \alpha_2 e_t,\end{aligned}\tag{5.9}$$

где $e_t = y_t - \hat{y}_1(t-1)$ — ошибка прогноза.

Из последних выражений видно, что частным случаем модели Ч. Хольта можно считать модель Р. Брауна, представленную в табл. 5.3. При этом единственный параметр β играет роль коэффициента дисконтирования наблюдений.

Трехпараметрическая модель Дж. Бокса и Г. Дженкинса представляет собой развитие модели Ч. Хольта, причем усовершенствование осуществляется путем включения в выражение для оценивания $\hat{a}_{1,t}$ разности ошибок.

В ряде работ проводилось теоретическое и экспериментальное сравнение различных моделей¹. В результате при обработке экономических временных рядов значения коэффициента α оказывались близкими к нулю, и авторы пришли к выводу о нецелесообразности включения в модель разности ошибок.

Пример 5.2

Рассмотрим в качестве примера использование двухпараметрической модели Хольта для прогнозирования среднегодовой численности промышленно-производственного персонала. Эти данные ежегодно публикуются Госкомстатом России как по всей промышленности, так и по основным отраслям. Исследуемый ряд годовой динамики относится ко всей промышленности в целом и охватывает временной период с 1980 г. Выбор такого интервала моделирования экономически оправдан, так как, хотя с начала 90-х годов начался новый этап экономического развития, но все предпосылки этих изменений формировались еще в 80-е годы. Практический интерес к таким прогнозным оценкам, получаемым по различным отраслям, вызван тем, что они служат своеобразными индикаторами изменений, структурных сдвигов, происходящих в развитии промышленного комплекса.

Как видно из табл. 5.3, для реализации этой модели требуется задание двух параметров, значения которых могут изменяться в диапазоне $0 < \alpha_1, \alpha_2 < 1$.

¹ Harrison P.J. Short-term sales forecasting. — Applied statistics // J. of the Royal Stat. Soc. 1965. Ser. C. Vol. 14. N 2, 3; Harrison P.J. Exponential smoothing and short-term sales forecasting // Management Science, 1967. Vol. 13. N 11; Ward D.H. Comparison of different systems of exponential weighted prediction // The Statistician. L. 1963. Vol. 13. N 3.

С помощью системы STATISTICA путем перебора возможных комбинаций на сетке значений были определены оптимальные значения этих параметров по критерию минимизации суммы квадратов ошибок.

Результаты моделирования представлены на рис 5.3. Фактические уровни обозначены сплошной линией, расчетные — пунктиром, два последних расчетных значения соответствуют полученному прогнозу.

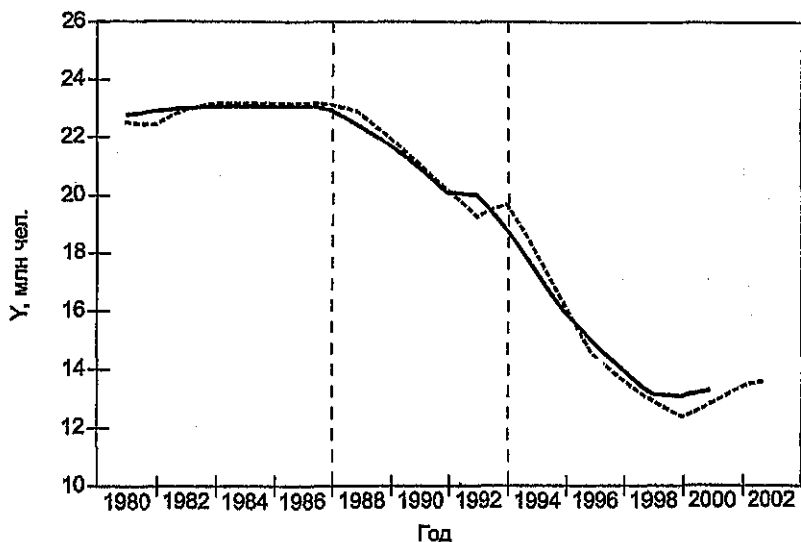


Рис. 5.3. Прогнозирование среднегодовой численности Y промышленно-производственного персонала с помощью модели Ч. Хольта

Модель достаточно удачно описывает динамику временного ряда. Значение средней относительной ошибки (по модулю) 1,7% говорит о высокой точности модели. Проведенный анализ ряда остатков свидетельствует о «хорошем качестве» выбранной модели. На рисунке видно, что значительная ошибка модели допущена в 1999 г., непосредственно следовавшем за серьезным финансовым кризисом 1998 г. Трудно было ожидать, что после столь глубокого спада, существенного снижения показателя в 1998 г. произойдет столь незначительное изменение уровня ряда в 1999 г., а затем последует рост. Это вызвано начавшимся ощутимым подъемом промышленности, связанным с процессом импортозамещения на внутреннем рынке, с ускоренным развитием экспортоориентированных отраслей и др. Поведение модели демонстрирует быструю адаптацию к произошедшим изменениям, в 2000 г. ошибка

модели уменьшилась, полученные же прогнозные оценки показывают дальнейший рост прогнозируемого показателя.

Понятие экспоненциальной средней можно обобщить в случае экспоненциальных средних более высоких порядков.

Выравнивание p -го порядка

$$S_t^{(p)} = \alpha S_t^{(p-1)} + \beta S_{t-1}^{(p)} \quad (5.10)$$

является простым экспоненциальным сглаживанием, примененным к результатам сглаживания $(p-1)$ -го порядка.

Если предполагается, что тренд некоторого процесса может быть описан полиномом степени n , то коэффициенты предсказывающего полинома могут быть вычислены через экспоненциальные средние соответствующих порядков.

В случае, когда исследуемый процесс, состоящий из детерминированной и случайной компоненты, описывается полиномом n -го порядка, прогноз на τ шагов вперед осуществляется по формуле:

$$\hat{y}_\tau(t) = \hat{a}_1 + \hat{a}_2\tau + \frac{1}{2}\hat{a}_3\tau^2 + \dots + \frac{1}{n!}\hat{a}_{n+1}\tau^n, \quad (5.11)$$

где $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_{n+1}$ — оценки параметров.

Фундаментальная теорема метода экспоненциального сглаживания и прогнозирования, впервые доказанная Р. Брауном и Р. Майером, говорит о том, что $(n+1)$ неизвестных коэффициентов полинома n -го порядка $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_{n+1}$ могут быть оценены с помощью линейных комбинаций экспоненциальных средних $S_t^{(i)}$, где $i = 1 \div n+1$. Следовательно, задача сводится к вычислению экспоненциальных средних, порядок которых изменяется от 1 до $n+1$, а затем к переходу через их линейные комбинации — к определению коэффициентов полинома.

На практике обычно используются полиномы не выше второго порядка.

В табл. (5.4) приведены формулы, необходимые для расчета по моделям нулевого, первого и второго порядков.

Процедура прогнозирования временных рядов на основе адаптивных полиномиальных моделей состоит из следующих этапов.

1. Выбирается вид модели экспоненциального сглаживания, задается значение параметра сглаживания α . При выборе порядка адаптивной полиномиальной модели могут использоваться различные подходы, например графический анализ, применение метода последовательных разностей и др.

Таблица 5.4. Основные формулы для прогнозирования по адаптивным полиномиальным моделям

Сте- пень моде- ли n	Начальные условия	Экспоненциальные средние	Оценка коэффициентов	Модель прогноза
1	2	3	4	5
0	$S_0^{(1)} = \hat{a}_{1,0}$	$S_t^{(1)} = \alpha y_t + \beta S_{t-1}^{(1)}$	$\hat{a}_{1,t} = S_t^{(1)}$	$\hat{y}_\tau(t) = \hat{a}_{1,t}$
1	$S_0^{(1)} = \hat{a}_{1,0} - \frac{\beta}{\alpha} \hat{a}_{2,0}$ $S_0^{(2)} = \hat{a}_{1,0} - \frac{2\beta}{\alpha} \hat{a}_{2,0}$	$S_t^{(1)} = \alpha y_t + \beta S_{t-1}^{(1)}$ $S_t^{(2)} = \alpha S_t^{(1)} + \beta S_{t-1}^{(2)}$	$\hat{a}_{1,t} = 2S_t^{(1)} - S_t^{(2)}$ $\hat{a}_{2,t} = \frac{\alpha}{\beta} [S_t^{(1)} - S_t^{(2)}]$	$\hat{y}_\tau(t) = \hat{a}_{1,t} + \tau \hat{a}_{2,t}$
2	$S_0^{(1)} = \hat{a}_{1,0} - \frac{\beta}{\alpha} \hat{a}_{2,0} + \frac{\beta \cdot (2-\alpha)}{2\alpha^2} \hat{a}_{3,0}$ $S_0^{(2)} = \hat{a}_{1,0} - \frac{2\beta}{\alpha} \hat{a}_{2,0} + \frac{\beta(3-2\alpha)}{\alpha^2} \hat{a}_{3,0}$ $S_0^{(3)} = \hat{a}_{1,0} - \frac{3\beta}{\alpha} \hat{a}_{2,0} + \frac{3\beta(4-3\alpha)}{2\alpha^2} \hat{a}_{3,0}$	$S_t^{(1)} = \alpha y_t + \beta S_{t-1}^{(1)}$ $S_t^{(2)} = \alpha S_t^{(1)} + \beta S_{t-1}^{(2)}$ $S_t^{(3)} = \alpha S_t^{(2)} + \beta S_{t-1}^{(3)}$	$\hat{a}_{1,t} = 3(S_t^{(1)} - S_t^{(2)}) + S_t^{(3)}$ $\hat{a}_{2,t} = \frac{\alpha}{2\beta^2} [(6-5\alpha)S_t^{(1)} - 2(5-4\alpha)S_t^{(2)} + (4-3\alpha)S_t^{(3)}]$ $\hat{a}_{3,t} = \frac{\alpha^2}{\beta^2} (S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)})$	$\hat{y}_\tau(t) = \hat{a}_{1,t} + \tau \hat{a}_{2,t} + \frac{1}{2} \tau^2 \hat{a}_{3,t}$

2. Определяются начальные условия. Например, для полиномиальной модели первого порядка необходимо определить $\bar{a}_{1,0}, \bar{a}_{2,0}$.

Чаще всего в качестве этих оценок берут коэффициенты соответствующих полиномов, полученные с помощью МНК. Начальные условия для модели нулевого порядка обычно получают усреднением нескольких первых уровней ряда. Зная эти оценки, с помощью указанных в табл. 5.4 формул находят начальные значения экспоненциальных средних (гр. 2).

3. Проводится расчет значений соответствующих экспоненциальных средних (гр. 3).

4. Находятся оценки коэффициентов модели (гр. 4).

5. Осуществляется прогноз на одну точку вперед (гр. 5 при $\tau = 1$); находится отклонение фактического значения временного ряда от прогнозируемого. Шаги с 3-го по 5-й данной процедуры повторяются для всех $t \leq n$, где n — длина ряда.

6. Окончательная прогнозная модель формируется на последнем шаге в момент $t = n$. Прогноз получается на базе выражения (5.11) путем подстановки в него последних значений коэффициентов и времени упреждения τ .

К положительным особенностям рассмотренных моделей следует отнести то, что при поступлении новой, «свежей» информации расчеты повторять не придется. Достаточно принять в качестве начальных условий последние значения функций сглаживания $S_t^{(i)}$ и продолжить вычисления.

Адаптация процедуры экспоненциального сглаживания с помощью следящего контрольного сигнала.

Рядом специалистов были предприняты попытки повысить адаптивные свойства ранее рассмотренных моделей путем ускорения реакции прогнозирующей системы на изменения динамики временного ряда. Для этого следовало определить вид индикатора, позволяющего постоянно следить за качеством прогнозных оценок. Этот индикатор должен был улавливать момент изменения поведения показателя и своевременно подавать сигнал об ухудшении качества модели и о том, что требуется модификация ее параметров. При моделировании индикатора следовало учитывать два требования:

- чувствительность к происходящим изменениям;
- способность отфильтровывать случайные колебания, чтобы не подавать ложных сигналов «тревоги».

Р. Браун¹ предложил в качестве такого индикатора использовать значения следящего контрольного сигнала (трекинг-сигнала), определяемого следующим образом:

¹ Brown R.G. Smoothing forecasting and prediction of discrete time series. — N.-Y., 1963.

$$K_t = \frac{\sum_{i=0}^t e_i}{\tilde{e}_t}, \quad (5.12)$$

где $\tilde{e}_t = (1-\gamma) \cdot \tilde{e}_{t-1} + \gamma \cdot |e_t|$, параметр сглаживания $0 < \gamma < 1$.

Таким образом, трэкинг-сигнал определяется как отношение суммы ошибок к сглаженному по схеме экспоненциальной средней абсолютному значению ошибки модели. Автор определил критические границы следящего контрольного сигнала, превышение которых говорит о необходимости корректировки модели.

Однако у данного подхода имеются следующие недостатки.

1. В случае выхода за установленные пределы трэкинг-сигнал не обязательно вернется обратно, даже если прогнозирующая система вновь окажется адекватной исследуемому процессу. Следовательно, потребуются дополнительное вмешательство, чтобы снова сделать числитель выражения (5.12) близким к нулю.

2. Возможна ситуация, когда модель начинает давать точные прогнозы, а значения следящего контрольного сигнала выходят за установленные границы. Пусть с определенного момента времени система идеально описывает поведение ряда, давая прогнозы с высокой точностью. При этом сумма в числителе практически остается неизменной, а знаменатель может стремиться к нулю, следовательно, сама характеристика (5.12) — к бесконечности.

Для устранения выявленных недостатков Д. Тригг¹ предложил модификацию рассмотренного подхода, состоящую в замене числителя в (5.12) на сглаженную ошибку $\hat{e}_t$:

$$\hat{e}_t = (1-\gamma)\hat{e}_{t-1} + \gamma e_t. \quad (5.13)$$

Таким образом, модифицированная формула для расчета следящего контрольного сигнала определяется выражением

$$K_t = \frac{\hat{e}_t}{\tilde{e}_t}. \quad (5.14)$$

К достоинствам этого подхода следует отнести то, что он преодолел ранее указанные недостатки, а также то, что значения модифицированного трэкинг-сигнала принадлежат интервалу $(-1; 1)$. Крайние значения могут получаться только в случае, если все ошибки будут иметь один знак. При реализации этого метода Д. Тригг рекомендовал соблюдать соотношение $\gamma \leq \alpha$.

¹ Trigg D.W. Monitoring a Forecasting System // Oper. Reser. Quart. 1964. V.15. N. 3. P. 271—274, 451—468.

При использовании трэкинг-сигнала важно определить пороговые значения, выход за которые свидетельствует об ухудшении качества прогнозных оценок. Возможные подходы к решению этой задачи рассмотрены в ряде работ [17, 26].

Проанализированный подход также можно использовать для регулирования скорости реакции модели экспоненциального сглаживания на текущие свойства временного ряда. Такой подход был предложен в работе Д. Тригга и А. Лича¹.

Например, в модели нулевого порядка регулирование параметра адаптации осуществляется за счет присваивания ему значения модуля трэкинг-сигнала (отметим, что эти характеристики имеют одинаковые пределы изменения):

$$\alpha_t = |K_t|.$$

При этом в случае существенного расхождения прогнозных и фактических уровней будет происходить увеличение значения, следовательно, и весовости последних наблюдений. Это будет способствовать адаптации модели к текущим условиям. После приспособления модели значение параметра уменьшится, что повысит способность модели отфильтровывать случайные колебания. К достоинствам этой схемы относится то, что она работает по принципу обратной связи, а также снижает значимость удачного задания начальных значений коэффициентов и параметра сглаживания.

Аналогичный подход возможен и для более сложных моделей [17]. Например, в модели линейного роста Р. Брауна удалось достичь стабильных результатов при присваивании значения модуля трэкинг-сигнала величине, определяющей весовость ошибки при оценивании коэффициента $\hat{a}_{1,t}$:

$$(1 - \beta_t^2) = |K_t|.$$

Обобщая рассмотренный подход к модификации моделей экспоненциального сглаживания, его практическую реализацию можно представить в виде следующей последовательности шагов.

1. Выбор вида модели экспоненциального сглаживания.
2. Определение значений параметров и начальных условий.
3. Вычисление экспоненциальных средних.
4. Определение оценок коэффициентов модели.
5. Прогнозирование на одну точку вперед.

¹ Trigg D.W., Leech A.G. Exponential Smoothing with an adaptive response rate // Oper. Res. Quart. 1969. V. 18. N. 1. — P. 53—59.

6. Вычисление абсолютной ошибки — отклонения фактического значения от расчетного.

7. Определение значения трэкинг-сигнала и переход к 5-му шагу.

На этом этапе некоторые авторы рекомендуют всегда модифицировать значение параметра адаптации в соответствии со значением модуля трэкинг-сигнала, другие — только в том случае, если значение контрольного следящего сигнала выходит за пороговое.

Однако, как показывают экспериментальные расчеты, не во всех реально возникающих ситуациях в динамике экономических временных рядов такой подход дает лучшие результаты по сравнению с обычными адаптивными схемами. В то же время проанализированный подход достаточно часто обеспечивает гибкое мониторинговое наблюдение за качеством прогнозирующих систем, способствует повышению скорости реакции модели экспоненциального сглаживания на текущие свойства временного ряда.

5.4. Адаптивные модели сезонных явлений

Многие экономические временные ряды содержат периодические сезонные колебания. Как было сказано ранее, в зависимости от характера этих колебаний их часто делят на два класса: мультипликативные и аддитивные. При мультипликативных сезонных колебаниях предполагается, что амплитуда колебаний изменяется во времени пропорционально уровню тренда (текущему среднему уровню ряда). При аддитивном характере сезонности исходят из предположения о неизменности во времени, примерном постоянстве амплитуды периодических колебаний, ее независимости от уровня тренда. При этом для аддитивных колебаний характеристики сезонности будут измеряться в абсолютных величинах и отражаться в статистической модели в виде слагаемых, а для мультипликативных колебаний — в относительных величинах и представляться в моделях в виде сомножителей.

Таким образом, экономические временные ряды, содержащие периодические сезонные колебания, могут быть описаны моделями как с аддитивным характером сезонности (5.15), так и с мультипликативным (5.16):

$$y_t = a_{1,t} f_t + \varepsilon_t; \quad (5.15)$$

$$y_t = a_{1,t} + g_t + \varepsilon_t, \quad (5.16)$$

где $a_{1,t}$ — характеристика тенденции развития;
 $g_t, g_{t-1}, \dots, g_{t-t+1}$ — аддитивный сезонный фактор;
 $f_t, f_{t-1}, \dots, f_{t-t+1}$ — мультипликативный сезонный фактор;

l — число фаз в полном сезонном цикле (для ежемесячных наблюдений $l = 12$, для квартальных $l = 4$);

ε_t — неавтокоррелированный шум с нулевым математическим ожиданием.

Очевидно, что можно составить множество адаптивных сезонных моделей, перебирая различные комбинации типов тенденций в сочетании с сезонными эффектами аддитивного и мультипликативного вида. Выбор той или иной модели будет продиктован характером динамики исследуемого процесса. В качестве примера рассмотрим модель с линейным характером тенденции и мультипликативным сезонным эффектом. Эта модель представляет собой объединение двухпараметрической модели линейного роста Хольта и сезонной модели Уинтерса, поэтому ее чаще всего называют моделью Хольта—Уинтерса.

Прогноз по модели Хольта—Уинтерса на τ шагов вперед определяется выражением

$$\hat{y}_\tau(t) = (\hat{a}_{1,t} + \tau \hat{a}_{2,t}) \hat{f}_{t-l+\tau}. \quad (5.17)$$

Обновление коэффициентов осуществляется следующим образом:

$$\begin{aligned} \hat{a}_{1,t} &= \alpha_1 \frac{y_t}{\hat{f}_{t-l}} + (1 - \alpha_1)(\hat{a}_{1,t-1} + \hat{a}_{2,t-1}); \\ \hat{f}_t &= \alpha_2 \frac{y_t}{\hat{a}_{1,t}} + (1 - \alpha_2) \hat{f}_{t-l}; \\ \hat{a}_{2,t} &= \alpha_3 (\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \alpha_3) \hat{a}_{2,t-1}; \\ 0 &< \alpha_1, \alpha_2, \alpha_3 < 1. \end{aligned} \quad (5.18)$$

Из (5.18) видно, что $\hat{a}_{1,t}$ — взвешенная сумма текущей оценки $\frac{y_t}{\hat{f}_{t-l}}$, полученной путем очищения от сезонных колебаний фактических данных y_t , и суммы предыдущих оценок $\hat{a}_{1,t-1} + \hat{a}_{2,t-1}$. В качестве коэффициента сезонности f_t берется его наиболее поздняя оценка, сделанная для аналогичной фазы цикла ($\hat{f}_{t-l}$).

Затем значение $\hat{a}_{1,t}$, полученное по первому уравнению, используется для определения новой оценки коэффициента сезонности по второму уравнению. Оценки $\hat{a}_{2,t}$ модифицируются по процедуре, аналогичной экспоненциальному сглаживанию.

Оптимальные значения для $\alpha_1, \alpha_2, \alpha_3$ П. Уинтерс предлагает находить экспериментальным путем, перебирая возможные ком-

бинации этих параметров на сетке значений. Критерием сравнения при этом выступает средняя квадратическая ошибка.

Примером другого подхода — с аддитивной сезонностью — может служить модель сезонных явлений с линейным ростом, предложенная Г. Тейлом и С. Вейджем. Практическая значимость этой модели объясняется не только тем, что в экономических временных рядах довольно часто можно встретить этот тип динамики развития. Опыт проведения экспериментальных расчетов свидетельствует о том, что динамика многих экономических показателей может быть описана с помощью модели, сочетающей в себе экспоненциальную тенденцию с мультипликативным сезонным эффектом. Прологарифмировав исходный временной ряд, на практике часто преобразуют экспоненциальную тенденцию в линейную и одновременно мультипликативный сезонный эффект в аддитивный. Таким образом, динамику преобразованного показателя можно моделировать и прогнозировать с помощью модели Г. Тейла и С. Вейджа.

Рассмотрим подробнее адаптивную тренд-сезонную модель, сочетающую линейный рост с аддитивной сезонностью.

Прогноз по этой модели на τ шагов вперед определяется выражением:

$$\hat{y}_\tau(t) = \hat{a}_{1,t} + \hat{a}_{2,t}\tau + \hat{g}_{t-l+\tau}. \quad (5.19)$$

Обновление коэффициентов осуществляется следующим образом:

$$\begin{aligned} \hat{a}_{1,t} &= \alpha_1(y_t - \hat{g}_{t-l}) + (1 - \alpha_1)(\hat{a}_{1,t-1} + \hat{a}_{2,t-1}); \\ \hat{g}_t &= \alpha_2(y_t - \hat{a}_{1,t}) + (1 - \alpha_2)\hat{g}_{t-l}; \\ \hat{a}_{2,t} &= \alpha_3(\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \alpha_3)\hat{a}_{2,t-1}; \\ 0 &< \alpha_1, \alpha_2, \alpha_3 < 1. \end{aligned} \quad (5.20)$$

Прогнозные оценки на основе формул (5.17) и (5.19) получают экстраполяцией тенденции линейного роста на основе последних значений коэффициентов $\hat{a}_{1,t}$ и $\hat{a}_{2,t}$, а также добавлением (в виде сомножителя или слагаемого) самой свежей оценки сезонного эффекта для этой фазы цикла ($\hat{f}_{t-l+\tau}$ или $\hat{g}_{t-l+\tau}$). Это справедливо для случая, когда время упреждения удовлетворяет условию $0 < \tau \leq l$. Очевидно, что для $0 < \tau \leq 2l$ самой последней оценкой сезонного эффекта будут значения $\hat{f}_{t-2l+\tau}$ или $\hat{g}_{t-2l+\tau}$ и т.д.

Таким образом, в двух рассмотренных моделях прогнозные оценки есть функция прошлых и текущих уровней временного

ряда, параметров адаптации $\alpha_1, \alpha_2, \alpha_3$, а также начальных значений как коэффициентов $\hat{a}_{1,0}, \hat{a}_{2,0}$, так и сезонного фактора для каждой фазы цикла. В качестве $\hat{a}_{1,0}, \hat{a}_{2,0}$ на практике берут МНК-оценки коэффициентов линейного тренда $\hat{y}_t = a_1 + a_2 t$, определенные по исходному временному ряду или его части. Начальные значения сезонного фактора для аддитивной модели определяют усреднением отклонений фактических уровней от расчетных ($\hat{y}_t$) для каждой фазы цикла (например для одноименных месяцев, кварталов), для мультипликативной модели — усреднением частного от деления фактических уровней на расчетные ($\hat{y}_t$) для каждой фазы цикла. Этот этап оценивания соответствует схеме, рассмотренной в § 4.1.

Влияние начальных условий на прогнозную оценку зависит от весов и длины ряда. Причем, как справедливо отмечается в [17], влияние начальных значений сезонного фактора уменьшается медленнее, чем влияние начальных значений коэффициентов $\hat{a}_{1,0}, \hat{a}_{2,0}$. Это объясняется тем, что значения последних корректируются на каждом шаге процедуры, а оценки сезонности для каждой фазы пересматриваются лишь один раз за цикл.

Отметим, что по аналогичной схеме строятся модели с экспоненциальным и демпфирующим трендом в сочетании с сезонными эффектами обоих типов.

Для повышения адаптивных свойств моделей, их способности улавливать изменения в динамике ряда предлагались модификации алгоритмов, отказывающиеся от поиска неизменных значений параметров сглаживания. Например, для однопараметрической модели использовались три различных значения параметра: α — нормальное значение, $(\alpha - h)$ — низкое значение, $(\alpha + h)$ — высокое значение (h — шаг изменения параметра). В результате при реализации алгоритма на каждом шаге получается не одна, а три оценки следующего уровня.

Оценка, полученная при нормальном значении параметра, считается прогнозной, а две другие оценки носят контрольный характер.

При сравнении фактического уровня ряда с расчетными определяется значение параметра, давшего наименьшую ошибку. На следующем шаге это значение параметра берется в качестве нормального. Одновременно рассчитываются контрольные оценки следующего уровня для низкого и высокого значений параметра (скорректированных с учетом полученного нормального значения). Таким образом, первоначально выбранное значение параметра

сглаживания постоянно изменяется в целях выработки компенсирующих воздействий для уменьшения возникающих ошибок прогнозирования.

Этот подход был развит для двух- и трехпараметрических моделей. Например, для модели Уинтерса строятся оценки при различных сочетаниях значений параметров сглаживания, заданных из расчета по три значения на каждый параметр. При наличии в модели k управляемых параметров на каждом шаге вычисляется $(2^k + 1)$ оценок будущего уровня. Оценка, полученная при нормальных значениях всех параметров, считается прогнозной. После анализа точности пробных оценок может быть принято решение об изменении нормальных значений параметров. Очевидно, что при этом проводится и соответствующая корректировка набора низких и высоких значений параметров.

При рассмотрении данной схемы возникает аналогия с процессом биологической эволюции. Выбор лучших значений параметров аналогичен процессу естественного отбора, поэтому рассмотренный подход получил название эволюционного планирования [17].

Адаптивные сезонные модели важная составная часть современных статистических ППП, ориентированных на решение задач прогнозирования. На рис. 5.4 показан фрагмент меню, позволяющего реализовать 12 моделей этого класса в ППП STATISTICA (аналогичный набор моделей представлен в SPSS). Первый столбец образуют четыре модели, предназначенные для описания различного характера тенденции (отсутствие тренда, линейный и экспоненциальный тренд, демпфирующий (с насыщением) вид тенденции). Вторая из этих моделей представлена ранее рассмотренной моделью линейного роста Ч. Хольта.

	None	Additive	Multiplicative
No trend	 G single		
Linear trend	 Holt		 Winters
Exponential			
Damped trend			

Рис. 5.4. Фрагмент меню в ППП STATISTICA: адаптивные модели, основанные на процедуре экспоненциального сглаживания

Следующие два столбца составляют модели из первой группы с добавленным к ним сезонным эффектом в аддитивной (второй столбец) и мультипликативной форме (третий столбец). Безусловно,

графическое представление (5.4) облегчает пользователю выбор нужной модели в каждом конкретном случае. При этом исследователь опирается на визуальный анализ динамики временного ряда, на представления о будущем поведении показателя, учитывает содержательный экономический анализ имеющейся информации.

Выявленные достоинства сезонных адаптивных моделей, их преимущества перед традиционными подходами, предполагающими неизменность сезонной компоненты во времени, свидетельствуют о том, что этот класс моделей служит эффективным средством в практике краткосрочного прогнозирования развития российской экономики.

Пример 5.3

Рассмотрим в качестве примера временной ряд месячной динамики производства электроэнергии в Российской Федерации за период с января 1994 г. по декабрь 1999 г. Для краткосрочного прогнозирования этого показателя применим класс тренд-сезонных адаптивных моделей, реализованный в системе STATISTICA.

На первом этапе проведем визуальный анализ компонентного состава временного ряда. Для исследования общей тенденции используем прием укрупнения интервалов. Преобразуем суммированием ряд месячной динамики в ряд годовых данных.

На рис. 5.5 показан график среднегодовых значений, а также размах уровней для каждого года (вертикальной чертой соединены максимальное и минимальное значение уровней каждого года). На основании этого графика можно предположить, что тенденция ряда носит линейный характер, а амплитуда внутригодовых колебаний примерно постоянна.

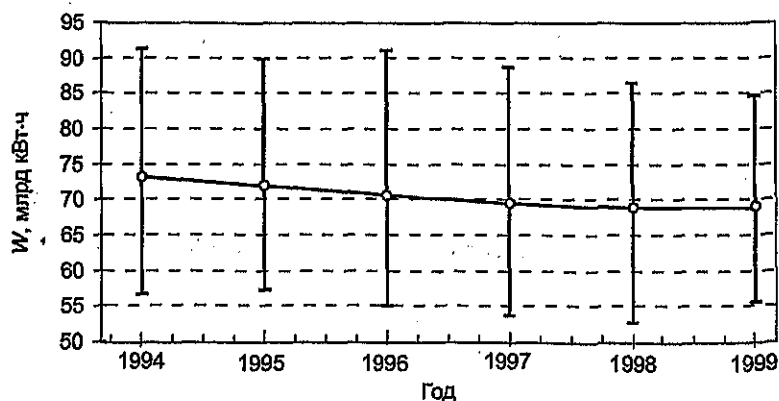


Рис. 5.5. Среднегодовые уровни производства электроэнергии W в Российской Федерации и размах уровней для каждого года

На рис. 5.6 сплошной линией показан график исходного временного ряда месячной динамики. На нем видны устойчивые сезонные колебания, имеющие годовую периодичность: очевидны повторяющиеся подъемы производства в осенне-зимний период, спады — в весенне-летний период.

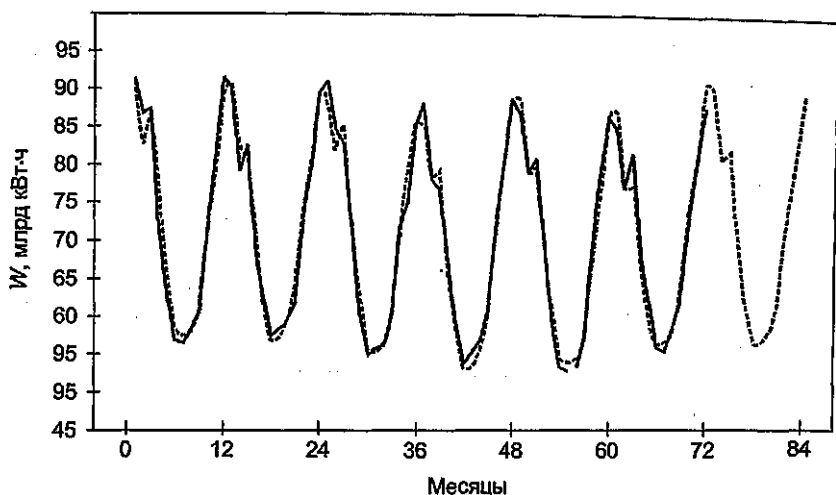


Рис. 5.6. Прогнозирование производства электроэнергии W в Российской Федерации с помощью тренд-сезонной модели, основанной на экспоненциальном сглаживании:

— — фактические данные; ---- — расчетные данные

Амплитуду периодических колебаний можно считать практически неизменной, не зависящей от уровня тренда. Таким образом, на основе проведения графического анализа можно в меню методов, показанном на рис. 5.4, выбрать модель линейного роста с аддитивным сезонным эффектом. Как видно из (5.19)—(5.20), для реализации этой модели требуется задание трех параметров, значения которых могут изменяться в диапазоне

$$0 < \alpha_1, \alpha_2, \alpha_3, < 1.$$

С помощью системы STATISTICA, перебирая возможные комбинации на сетке значений, определим оптимальные значения этих параметров по критерию минимизации суммы квадратов ошибок. Результаты моделирования представлены на рис. 5.6, значения за последние 12 месяцев, показанные пунктиром, — прогноз на год вперед.

Модель достаточно хорошо описывает динамику временного ряда. Графический анализ ряда остатков позволяет предположить

случайный характер колебаний около нуля, график автокорреляционной функции остатков также свидетельствует об адекватности выбранной модели (рис. 5.7). Значение средней относительной ошибки (по модулю) 1,8% говорит о высокой точности модели.

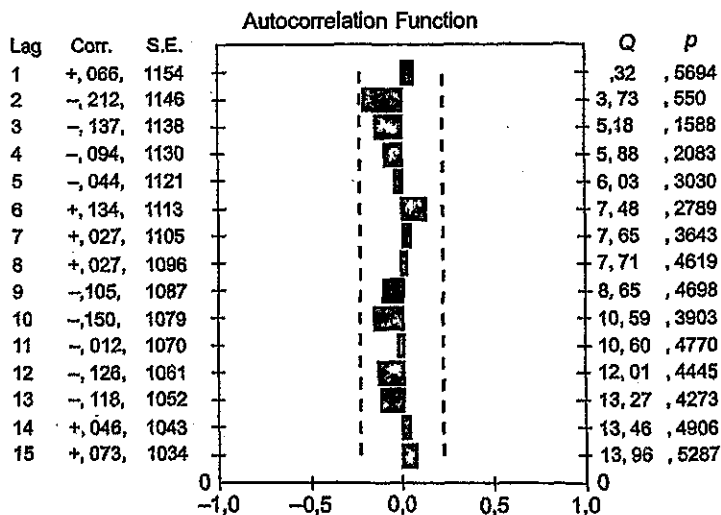
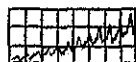


Рис. 5.7. Автокорреляционная функция остатков, полученных при прогнозировании производства электроэнергии в Российской Федерации с помощью тренд-сезонной модели

Экспериментальные расчеты показали, что при прогнозировании основных показателей промышленного производства в натуральном выражении часто удается повысить точность моделирования, а иногда и упростить вид модели исключением влияния фактора рабочего времени, исключением календарных биений. Этого эффекта достигают переходом к уровням среднесуточного производства для каждого анализируемого месяца. Это одно из возможных направлений совершенствования рассмотренного подхода.



Практикум к главе 5

Контрольные вопросы

1. Укажите характерные особенности адаптивных методов прогнозирования.
2. Какие типы адаптивных моделей вы знаете?
3. Чем объясняется название «экспоненциальная средняя»?
4. Какую роль играет параметр адаптации α в процедуре экспоненциального сглаживания?
5. Поясните суть модификации процедуры экспоненциального сглаживания, предложенной Р. Вейдом.

6. Как влияет значение параметра адаптации α на характер ряда, полученного после экспоненциального сглаживания?
7. Какие модели линейного роста вы знаете? В чем их особенности?
8. Как рассчитывается значение следящего контрольного сигнала?
9. Поясните, каким образом можно использовать следящий контрольный сигнал в процедуре экспоненциального сглаживания.
10. Для прогнозирования каких временных рядов используется модель Хольта—Уинтерса?
11. На рис. 5.4 показано меню ППП STATISTICA, предлагающее пользователю 12 адаптивных моделей, опирающихся на процедуру экспоненциального сглаживания. Какие из этих моделей можно было бы выбрать для описания динамики и прогнозирования временного ряда, представленного на рис. 1.3?

Упражнения¹

- 5.1. В таблице представлены данные о среднегодовой численности промышленно-производственного персонала в легкой промышленности (Российский статистический ежегодник: Стат. сб. — М.: Госкомстат России, 2001).

Год	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Численность, тыс. чел.	2288	2145	1845	1699	1600	1332	1133	1006	888	863	849

Рассчитайте прогнозное значение показателя в 2001 г., используя двухпараметрическую модель линейного роста Ч. Хольта.

В качестве $\hat{a}_{1,0}, \hat{a}_{2,0}$ возьмите МНК-оценки линейного тренда.

Примите $\tau = 1$; $\alpha_1 = 0,6$; $\alpha_2 = 0,5$. Оцените точность модели.

- 5.2. Рассчитайте экспоненциальную среднюю для временного ряда урожайности зерновых культур.

В качестве начального значения экспоненциальной средней S_0 возьмите среднее значение из пяти первых уровней ряда. Значение параметра адаптации α примите равным 0,3. Рассчитайте прогнозную оценку урожайности в 2000 г. с помощью модели нулевого порядка, основанной на экспоненциальном сглаживании ($\tau = 1$; $\alpha = 0,3$).

Урожайность зерновых культур, ц/га

Год	t	y_t	Год	t	y_t
1986	1	17,5	1993	8	15,9
1987	2	15,0	1994	9	14,4
1988	3	18,5	1995	10	16,2
1989	4	14,2	1996	11	18,0
1990	5	14,9	1997	12	18,3
1991	6	12,6	1998	13	17,0
1992	7	15,2	1999	14	18,8

¹ Рекомендуется для расчетов использовать Excel.

5.3. Укажите по данным упражнения № 2, какой ряд будет носить более гладкий характер:

а) исходный ряд;

б) ряд, полученный после экспоненциального сглаживания при $\alpha = 0,8$;

в) ряд, полученный после экспоненциального сглаживания при $\alpha = 0,3$.

5.4. В упражнении 1 практикума к гл. 4 представлены данные квартальной динамики о расходах потребителей на газ и электричество в США (в постоянных ценах, млрд долл.)

Рассчитайте прогноз показателя в 1-м квартале 1983 г., используя модель Тейла—Вейджа. В качестве $\hat{a}_{1,0}$, $\hat{a}_{2,0}$ следует взять МНК-оценки линейного тренда. Начальные значения сезонного фактора определите усреднением отклонений фактических уровней от расчетных трендовых значений для одноименных кварталов. Примите $\tau = 1$; $\alpha_1 = 0,1$; $\alpha_2 = 0,4$; $\alpha_3 = 0,3$. Сравните точность моделирования с результатами, полученными при использовании фиктивных переменных (см. практикум к гл. 4, упражнение 1).

5.5. В Приложении 2 приводятся данные месячной динамики пройденного авиалайнерами расстояния. Путем агрегирования уровней перейдите к данным квартальной динамики. По данным за пять лет (с 1963 по 1967 г.) рассчитайте прогнозное значение показателя в 1-м квартале 1968 г., используя модель Хольта—Уинтерса. В качестве $\hat{a}_{1,0}$, $\hat{a}_{2,0}$ следует взять МНК-оценки линейного тренда. Начальные значения сезонного фактора определите усреднением для одноименных кварталов частного от деления фактических уровней на расчетные трендовые значения. Примите $\tau = 1$; $\alpha_1 = 0,2$; $\alpha_2 = 0,4$; $\alpha_3 = 0,3$. Оцените точность моделирования.



ПРОГНОЗИРОВАНИЕ С ПОМОЩЬЮ МОДЕЛИ АВТОРЕГРЕССИИ — ПРОИНТЕГРИРОВАННОГО СКОЛЬЗЯЩЕГО СРЕДНЕГО

6.1. Модели стационарных временных рядов

6.1.1. Авторегрессионные модели

Часто экономические показатели, представленные временными рядами, имеют настолько сложную структуру, что моделирование таких рядов путем построения моделей тренда, сезонности и применения других традиционных подходов не приводит к удовлетворительным результатам. Во временном ряду ошибок остаются статистические зависимости, которые можно моделировать.

В последнее время большое внимание уделяется моделированию стационарных временных рядов. Это объясняется тем, что многие временные ряды могут быть приведены к стационарному виду после операции выделения тренда, фильтрации сезонной компоненты или взятия разности. Как правило, ряд ошибок — это стационарный ряд.

Наиболее распространенные модели стационарных рядов — *модели авторегрессии и модели скользящего среднего*.

Рассмотрим класс авторегрессионных моделей, сокращенно обозначаемых *AR(p)-модели* (число в скобках указывает порядок авторегрессии) или, как принято в специальной англоязычной литературе и мировой статистической практике, *AR(p)-models* (*AutoRegressive processes of order p*).

В авторегрессии каждое значение ряда находится в линейной зависимости от предыдущих значений. Если анализируемый динамический процесс зависит от значений, отстоящих до p временных лагов назад, то это авторегрессионный процесс порядка p , т.е. $AR(p)$:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t$$

или

$$(1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p) y_t = \Phi(B) y_t = \varepsilon_t, \quad (6.1)$$

где B — оператор сдвига, т.е. преобразование ряда, смещающее его на один временной такт;

$\Phi(B)$ — оператор авторегрессии.

При этом для выполнения условия стационарности все корни многочлена $\Phi(B)$ должны лежать вне единичного круга, т.е. все корни соответствующего характеристического уравнения должны быть по модулю больше 1 и различны.

Рассмотрим простейший вариант линейного авторегрессионного процесса — модель авторегрессии 1-го порядка — AR(1), или марковский процесс.

Эта модель может быть представлена в виде:

$$y_t = \alpha y_{t-1} + \varepsilon_t, \quad (6.2)$$

где α — числовой коэффициент, $|\alpha| < 1$,

ε_t — последовательность случайных величин, образующих белый шум.

Как известно, для белого шума

$$\begin{aligned} M\varepsilon_t &= 0; \\ M(\varepsilon_t \varepsilon_{t+\tau}) &= \begin{cases} \sigma_0^2 & \text{при } \tau = 0; \\ 0 & \text{при } \tau \neq 0. \end{cases} \end{aligned} \quad (6.3)$$

Перечислим основные свойства марковского процесса:

1) $My_t = 0$;

$$2) D(y_t) = \frac{\sigma_0^2}{1 - \alpha^2};$$

$$3) \text{cov}(y_t, y_{t+k}) = \alpha^k D(y_t);$$

$$4) \rho(y_t, y_{t+k}) = \alpha^k.$$

Рассмотрим 1-е свойство.

Представим $y_t = \alpha y_{t-1} + \varepsilon_t$ в виде:

$$\begin{aligned} y_t &= \alpha(\alpha y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \alpha^2 y_{t-2} + \alpha \varepsilon_{t-1} + \varepsilon_t = \alpha^2 (\alpha y_{t-3} + \varepsilon_{t-2}) + \\ &+ \alpha \varepsilon_{t-1} + \varepsilon_t = \alpha^3 y_{t-3} + \alpha^2 \varepsilon_{t-2} + \alpha \varepsilon_{t-1} + \varepsilon_t = \dots = \varepsilon_t + \alpha \varepsilon_{t-1} + \alpha \varepsilon_{t-2} + \dots = \\ &= \sum_{k=0}^{\infty} \alpha^k \varepsilon_{t-k}; \end{aligned} \quad (6.4)$$

(Очевидно, что y_t зависит от всех предыдущих ε_t). Отсюда с учетом (6.3) непосредственно следует, что $My_t = 0$.

Получим выражение для дисперсии марковского процесса. Рассмотрим AR(1) в виде (6.4):

$$D(y_t) = \sigma_0^2 (1 + \alpha^2 + \alpha^4 + \dots) = \frac{\sigma_0^2}{1 - \alpha^2}. \quad (6.5)$$

Таким образом,

$$D(y_t) = \frac{\sigma_0^2}{1 - \alpha^2}.$$

Отсюда видно, что при α близком к ± 1 дисперсия будет намного больше σ_0^2 — дисперсии белого шума. Следовательно, если последовательные значения ряда сильно коррелированы, то даже незначительные возмущения будут порождать размашистые колебания.

Остановимся подробнее на 3 и 4-м свойствах.

Рассмотрим уравнение

$$y_t = \alpha y_{t-1} + \varepsilon_t.$$

Умножим его на y_{t-1} и возьмем математическое ожидание:

$$M(y_t y_{t-1}) = \alpha M(y_{t-1} y_{t-1})$$

или

$$\text{cov}(y_t, y_{t-1}) = \alpha D(y_t).$$

Таким образом,

$$\alpha = \frac{\text{cov}(y_t, y_{t-1})}{D(y_t)},$$

т.е. α — коэффициент автокорреляции первого порядка (определяет значение коэффициента парной корреляции между соседними уровнями ряда):

$$\rho(1) = \alpha.$$

Можно показать, что

$$\text{cov}(y_t, y_{t+k}) = \alpha^k D(y_t),$$

а следовательно,

$$\rho(y_t, y_{t+k}) = \alpha^k = \rho(k).$$

Поэтому степень тесноты корреляционной связи между членами последовательности экспоненциально убывает по мере их взаимного удаления друг от друга во времени.

Все автокорреляции марковского процесса можно выразить через автокорреляцию первого порядка:

$$\rho(k) = \rho(1)\rho(k-1) = \rho(1)^2\rho(k-2) = \dots = \rho(1)^k = \alpha^k.$$

Значения *частной автокорреляционной функции* равны нулю для всех лагов $k \geq 2$, что может быть использовано при подборе модели. Этот результат верен для теоретической частной автокорреляционной функции и может не выполняться для выборочной автокорреляционной функции. Однако, если выборочные частные корреляции статистически незначимо отличаются от нуля при $k \geq 2$, то использование модели AR(1) не противоречит исходным данным.

Условие стационарности ряда для AR(1) определяется требованием к коэффициенту α : $|\alpha| < 1$ или, что то же самое, корень z_0 уравнения $1 - \alpha z = 0$, являющегося частным случаем характеристического уравнения для общего линейного процесса авторегрессии, должен быть по абсолютной величине больше 1.

На процесс AR(1) с $\alpha = 1$ внешне похож процесс «случайного блуждания» (*random walk*). Это название объясняется тем, что каждое значение определяется случайным отклонением от предыдущего:

$$y_t = y_{t-1} + \varepsilon_t \quad (6.6)$$

Однако свойства этого процесса существенно отличаются от стационарного процесса AR(1) (при $|\alpha| < 1$).

В ряде работ авторы доказывают, что многие экономические, финансовые показатели ведут себя как случайные блуждания [6, 16]. В частности, модель случайного блуждания часто используется для описания движения биржевых цен.

Учитывая свойства ε_t , а также некоррелированность ε_t и y_{t-1} , можно получить

$$M(y_t) = M(y_{t-1}) + 0;$$

$$D(y_t) = D(y_{t-1}) + \sigma^2,$$

т.е. случайное блуждание нестационарно, так как $D(y_t) \neq D(y_{t-1})$. При этом дисперсия может неограниченно возрастать со временем.

Процесс случайного блуждания отличается от стационарного процесса AR(1) тем, что влияние возмущений ε_t не затухает:

$$y_t = \varepsilon_t + \varepsilon_{t-1} + \dots,$$

в то время как в AR(1) их влияние падает:

$$y_t = \varepsilon_t + \alpha \varepsilon_{t-1} + \alpha^2 \varepsilon_{t-2} + \dots$$

Отметим, что процесс с $|\alpha| > 1$ тем более нестационарный. Такие ряды маловероятны в реальных финансово-экономических задачах, так как это подразумевает взрывные ряды, а давление экономической среды не позволяет показателям принимать бесконечно большие значения.

Некоторые теоретические проблемы с моделью случайного блуждания связаны с тем, что регрессионные модели для таких нестационарных временных рядов приводят к ложному результату. В этом случае теорема Гаусса—Маркова не подходит, так как случайное блуждание не имеет конечной дисперсии.

Используя процедуру дифференцирования (взятие первой разности), осуществим переход от уравнения $y_t = \alpha y_{t-1} + \varepsilon_t$ к следующему соотношению:

$$\Delta y_t = (\alpha - 1)y_{t-1} + \varepsilon_t,$$

где $\Delta y_t = y_t - y_{t-1}$; $\beta = \alpha - 1$.

Так как наличие *единичного корня* (*unit root*) у характеристического уравнения существенно влияет на свойства процесса, то большое практическое значение приобретает его тестирование по имеющимся наблюдениям. При этом не рекомендуется использовать традиционный *t*-критерий Стьюдента для проверки значимости β , так как регрессия для оценивания коэффициента β применяется в предположении, что $\beta < 0$ ($|\alpha| < 1$). Использование стандартной процедуры проверки, опирающейся на *t*-статистику, приведет к тому, что нулевая гипотеза существования единичного корня будет отвергаться (ошибочно) слишком часто.

В этом случае рекомендуется использовать распределение *t*-статистики, описанное Дики и Фуллером¹.

Дики и Фуллер рассмотрели следующие модификации модели:

- 1) $y_t = \alpha y_{t-1} + \varepsilon_t$ (нулевое среднее, отсутствие тренда);
- 2) $y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$ (в финансовых задачах обычно включают положительную среднюю, например, рискованные активы предполагают положительную норму прибыли);
- 3) $y_t = \alpha_0 + \alpha_2 t + \alpha_3 y_{t-1} + \varepsilon_t$ (наличие временного тренда).

Все три модели соответственно могут быть преобразованы:

$$\Delta y_t = \beta y_{t-1} + \varepsilon_t; \quad (1)$$

$$\Delta y_t = \alpha_0 + \beta y_{t-1} + \varepsilon_t; \quad (2)$$

$$\Delta y_t = \alpha_0 + \alpha_2 t + \beta y_{t-1} + \varepsilon_t. \quad (3)$$

Нулевая гипотеза $H_0: \beta = 0$ (при альтернативной $H_1: \beta < 0$), будет отвергнута, если наблюдаемое значение критерия t_n меньше критического значения $t_{кр}$, взятого из таблиц Дики и Фуллера. Значение $t_{кр}$ зависит от фиксированного уровня значимости, размера выборки и вида модели (1, 2, 3).

Если нулевая гипотеза отвергается, то ряд может быть отнесен к стационарным. Если нулевая гипотеза не отвергается, то проверка не противоречит тому, что имеется единичный корень и исследуемый ряд соответствует модели случайного блуждания.

¹ В [19, с. 244] приведены односторонние критические значения статистики Дики—Фуллера (DF).

Тест Дики—Фуллера включают многие современные эконометрические пакеты. Проблемы при проверке стационарности возникают при наличии автокорреляции остатков. Поэтому в ППП «EViews» реализован расширенный критерий Дики—Фуллера (Augmented Dickey—Fuller test, ADF). Это обобщение обычного теста DF для моделей, в правую часть которых добавляются в виде слагаемых лаговые значения приращений. Такая модель может быть записана в виде:

$$\Delta y_t = \alpha_0 + \beta y_{t-1} + \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \dots + \phi_p \Delta y_{t-p} + \varepsilon_t \quad (6.7)$$

Критические значения статистики Дики—Фуллера остаются справедливыми и для уравнения (6.7). Это позволяет осуществлять проверку единичного корня в AR-моделях более высокого порядка (выше 1-го порядка). Приращения добавляют в модель для того, чтобы устранить возможную автокорреляцию ошибок, так как критические значения статистики Дики—Фуллера справедливы только лишь в случае, если ошибки представляют собой белый шум. Однако включение чрезмерного числа лагов снижает мощность критерия.

Из авторегрессионных процессов выше первого порядка в экономической практике часто встречаются так называемые *процессы Юла*. Они описываются с помощью модели AR(2):

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t \quad (6.8)$$

где ε_t — белый шум.

$$M\varepsilon_t = 0;$$

$$M(\varepsilon_t \varepsilon_{t+\tau}) = \begin{cases} \sigma_0^2 & \text{при } \tau = 0; \\ 0 & \text{при } \tau \neq 0. \end{cases}$$

Определим выражение для вычисления любого значения АКФ $\rho(k)$. Для этого домножим обе части уравнения (6.8) на y_{t-k} :

$$y_t y_{t-k} = \alpha_1 y_{t-1} y_{t-k} + \alpha_2 y_{t-2} y_{t-k} + \varepsilon_t y_{t-k} \quad (6.9)$$

Возьмем математическое ожидание

$$\gamma(k) = \alpha_1 \gamma(k-1) + \alpha_2 \gamma(k-2). \quad (6.10)$$

Очевидно, что

$$M(\varepsilon_t y_{t-k}) = 0,$$

так как y_{t-k} зависит от ε_{t-k} и предшествующих значений, а не от будущих значений ε_t . Разделим обе части равенства (6.10) на $\gamma(0)$:

$$\rho(k) = \alpha_1 \rho(k-1) + \alpha_2 \rho(k-2). \quad (6.11)$$

Это выражение позволяет вычислить значение АКФ для различных значений лагов k . Подставим последовательно в (6.11) значение $k = 1, 2$.

С учетом того, что $\rho(0) = 1$, а $\rho(-1) = \rho(1)$, получим

$$\begin{cases} \rho(1) = \alpha_1 + \alpha_2 \rho(1); \\ \rho(2) = \alpha_1 \rho(1) + \alpha_2. \end{cases} \quad (6.12)$$

Эта система называется *системой Юла—Уокера (Yule—Walker)* для AR(2).

Если разрешить эту систему относительно α_1 и α_2 , то получим выражения:

$$\begin{aligned} \alpha_1 &= \frac{\rho(1)[1 - \rho(2)]}{1 - \rho^2(1)}; \\ \alpha_2 &= \frac{\rho(2) - \rho^2(1)}{1 - \rho^2(1)}. \end{aligned} \quad (6.13)$$

Выразим из системы (6.12) два первых значения АКФ:

$$\begin{aligned} \rho(1) &= \frac{\alpha_1}{1 - \alpha_2}; \\ \rho(2) &= \alpha_2 + \frac{\alpha_1^2}{1 - \alpha_2}. \end{aligned} \quad (6.14)$$

Определив $\rho(1)$ и $\rho(2)$, можно вычислить последующие значения АКФ с помощью (6.11).

Получим соотношение, связывающее между собой дисперсию ряда y_t и дисперсию белого шума ε_t , равную σ_0^2 . Для этого умножим уравнение AR(2) на y_t :

$$y_t^2 = \alpha_1 y_t y_{t-1} + \alpha_2 y_t y_{t-2} + \varepsilon_t (\alpha_1 y_{t-1} + \alpha_2 y_{t-2} y_{t-k} + \varepsilon_t).$$

Возьмем математическое ожидание:

$$\gamma(0) = \alpha_1 \gamma(1) + \alpha_2 \gamma(2) + \sigma_0^2.$$

Очевидно, что $M(\varepsilon y_{t-1}) = 0$, $M(\varepsilon y_{t-2}) = 0$, так как y_t не зависит от будущих значений ε_t .

Следовательно,

$$\sigma_0^2 = \gamma(0) - \alpha_1 \gamma(1) - \alpha_2 \gamma(2).$$

Перейдем от коэффициентов автоковариации к автокорреляционным коэффициентам, умножив и разделив правую часть уравнения на $\gamma(0)$:

$$\sigma_0^2 = \gamma(0)[1 - \alpha_1\rho(1) - \alpha_2\rho(2)]. \quad (6.15)$$

Подставим в (6.15) выражения (6.14) для $\rho(1)$ и $\rho(2)$ через α_1 и α_2 :

$$\gamma(0) = \frac{(1 - \alpha_2)\sigma_0^2}{(1 + \alpha_2)[(1 - \alpha_2)^2 - \alpha_1^2]}. \quad (6.16)$$

Отсюда, учитывая, что дисперсия должна быть положительна, получаем условия стационарности процесса AR(2). Условия стационарности ряда u_t также могут быть получены с учетом (6.14) из требований

$$|\rho(1)| < 1, |\rho(2)| < 1.$$

Отметим, что те же самые условия получаются из требования, чтобы все корни соответствующего характеристического уравнения $1 - \alpha_1 z - \alpha_2 z^2 = 0$ лежали вне единичного круга.

Условия стационарности процесса AR(2) могут быть записаны в виде

$$\begin{aligned} |\alpha_2| &< 1; \\ \alpha_1 + \alpha_2 &< 1; \\ \alpha_2 - \alpha_1 &< 1. \end{aligned} \quad (6.17)$$

ЧАКФ для процесса AR(p) будет иметь ненулевые значения лишь при $k \leq p$, а начиная с лага $k = p + 1$ теоретическая ЧАКФ равна нулю. Это свойство становится ключевым при подборе порядка p авторегрессионной модели для конкретных экономических временных рядов.

Проанализированные свойства авторегрессионных моделей позволяют сформулировать следующие *практические рекомендации* по их идентификации, опирающиеся на изучение АКФ (ACF) и ЧАКФ (PACF).

1. У моделей AR(p) значения коэффициентов АКФ экспоненциально затухают (либо монотонно, либо попеременно меняя знак).

2. ЧАКФ для модели AR(p) имеет выбросы (пики) на первых p лагах, а значения коэффициентов для лагов, больших порядка авторегрессии, статистически незначимы.

Типичные примеры корреляционных функций для двух процессов AR(1) показаны на рис. 6.1—6.4.

Autocorrelation Function
VAR1
(Standart errors are whaite-noise estimates)

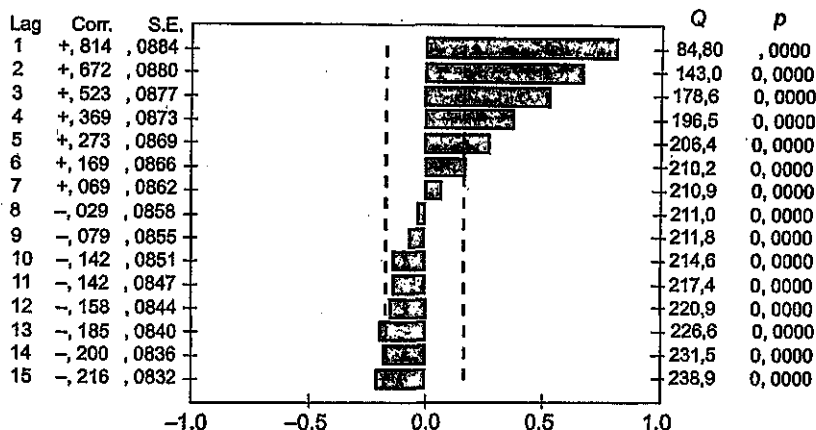


Рис. 6.1. АКФ для ряда, генерированного моделью AR(1)
 $y_t = 0,8y_{t-1} + \varepsilon_t$ Корень 1,25

Partial Autocorrelation Function
VAR1
(Standart errors are assume-AR order of k-1)

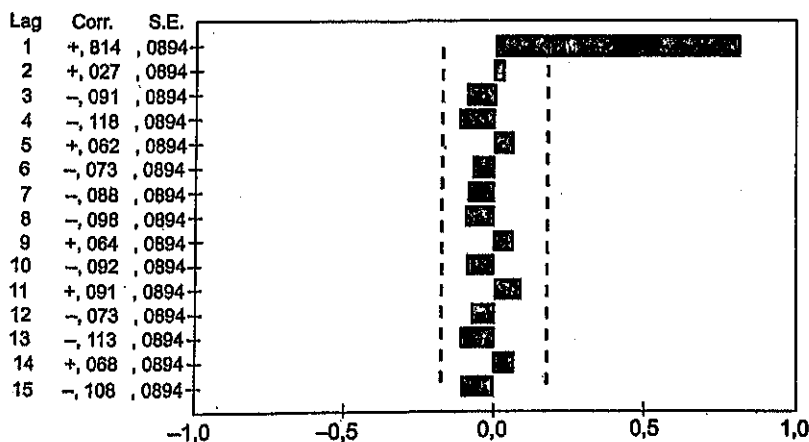


Рис. 6.2. ЧАКФ для ряда, генерированного моделью AR(1)
 $y_t = 0,8y_{t-1} + \varepsilon_t$

Autocorrelation Function
VAR2
(Standart errors are whaite-noise estimates)

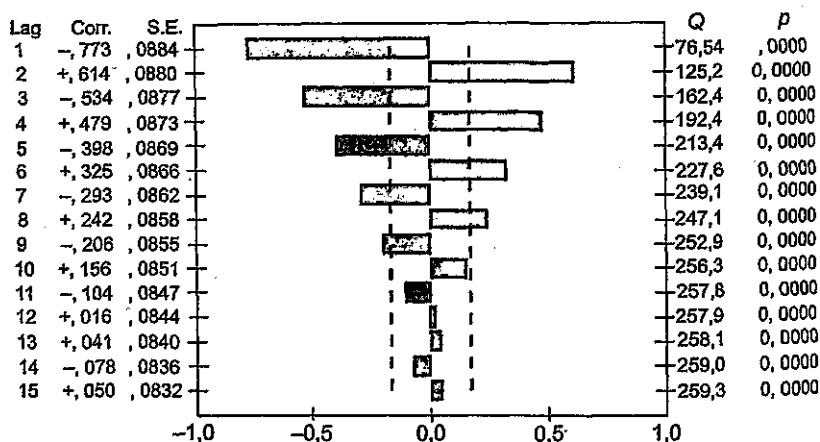


Рис. 6.3. АКФ для ряда, генерированного моделью AR(1)
 $y_t = -0,8y_{t-1} + \varepsilon_t$. Корень $-1,25$

Partial Autocorrelation Function
VAR2
(Standart errors are assume-AR order of k-1)

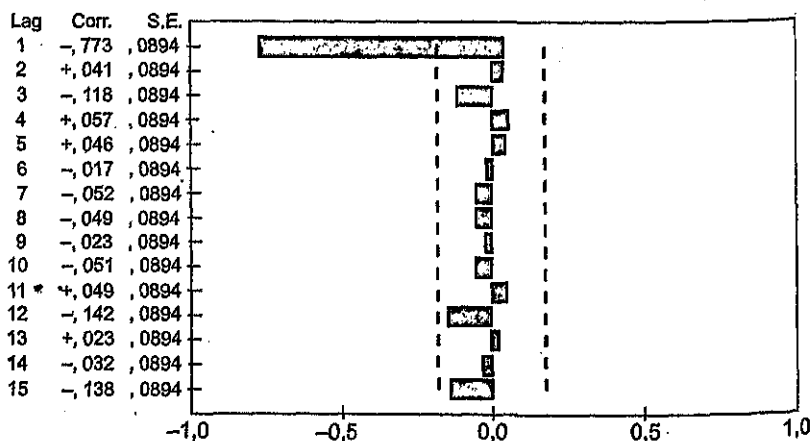


Рис. 6.4. ЧАКФ для ряда, генерированного моделью AR(1)
 $y_t = -0,8y_{t-1} + \varepsilon_t$

6.1.2. Модели скользящего среднего

Модель скользящего среднего предполагает, что в ошибках модели в предшествующие периоды сосредоточена информация обо всей предыстории ряда. В этой модели каждое новое значение — среднее между текущей флуктуацией и несколькими (в частности, одной) предыдущими ошибками.

Модели скользящего среднего порядка q , обозначаемые $CC(q)$, в англоязычной литературе $MA(q)$ (*Moving Average models*), имеют вид:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \quad (6.18)$$

где ε_t — белый шум.

Широко распространены в статистической практике модели скользящего среднего 1-го ($q = 1$) и второго порядка ($q = 2$):

$$MA(1): y_t = \varepsilon_t - \theta \varepsilon_{t-1}; \quad (6.19)$$

$$MA(2): y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}. \quad (6.20)$$

Рассмотрим модель скользящего среднего 1-го порядка — $MA(1)$.

Преобразуем (6.19), последовательно выражая ε_{t-1} , ε_{t-2} , ε_{t-3} и т.д.:

$$\begin{aligned} \varepsilon_t &= y_t + \theta \varepsilon_{t-1} = y_t + \theta(y_{t-1} + \theta \varepsilon_{t-2}) = \\ &= y_t + \theta y_{t-1} + \theta^2(y_{t-2} + \theta \varepsilon_{t-3}) = \\ &= y_t + \theta y_{t-1} + \theta^2 y_{t-2} + \theta^3(y_{t-3} + \theta \varepsilon_{t-4}) = \dots = \\ &= y_t + \theta y_{t-1} + \theta^2 y_{t-2} + \theta^3 y_{t-3} + \dots \end{aligned}$$

Это выражение можно переписать в виде:

$$y_t = \varepsilon_t - \sum_{k=1}^{\infty} \theta^k y_{t-k}. \quad (6.21)$$

Таким образом, ряд y_t , сгенерированный моделью $MA(1)$, может быть представлен также в виде модели авторегрессии бесконечного порядка. В моделях скользящего среднего $MA(q)$ не требуется накладывать никаких ограничений на параметры θ_1 , θ_2 , ..., θ_q для обеспечения стационарности ряда. Однако, если в модели $MA(1)$ параметр θ по абсолютной величине больше или равен 1, то текущее значение y_t в соответствии с (6.21) будет зависеть от своих прошлых значений y_{t-1} , y_{t-2} , ..., берущихся с весами, бесконечно растущими по мере удаления в прошлое. Чтобы избежать этого, надо, чтобы веса в (6.21) образовывали сходящийся ряд, т.е. чтобы $|\theta| < 1$.

Отметим, что подобно тому, как ряд, генерированный моделью скользящего среднего первого порядка $MA(1)$, может быть пред-

ставлен в виде модели авторегрессии бесконечного порядка $AR(\infty)$, также существует представление $AR(1)$ в виде $MA(\infty)$. При этом на параметры процесса $AR(p)$ не накладываются никакие условия для того, чтобы этот процесс был обратимым. Но для стационарности процесса корни его характеристического уравнения должны лежать вне единичного круга. В то же время параметры процесса $MA(q)$ не должны удовлетворять никаким условиям для стационарности, однако для обратимости корни его характеристического уравнения

$$1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_q z^q = 0$$

должны лежать вне единичного круга.

Найдем выражение для АКФ процесса $MA(q)$. Для этого представим y_{t-k} в виде соотношения (6.18):

$$y_{t-k} = \varepsilon_{t-k} - \theta_1 \varepsilon_{t-k-1} - \theta_2 \varepsilon_{t-k-2} - \dots - \theta_q \varepsilon_{t-k-q}. \quad (6.22)$$

Перемножим соответственно левые и правые части уравнений (6.18) и (6.22), а затем возьмем математическое ожидание от получившегося выражения. При этом следует учесть, что элементы белого шума ε_{t_1} и ε_{t_2} не коррелируют при $t_1 \neq t_2$.

Тогда выражение для ковариации $M(y_t y_{t-\tau}) = \gamma(\tau)$ примет вид:

$$\gamma(\tau) = \begin{cases} \sigma_0^2(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) & \text{при } \tau = 0; \\ \sigma_0^2(-\theta_\tau + \theta_1 \theta_{\tau+1} + \theta_2 \theta_{\tau+2} + \dots + \theta_{q-\tau} \theta_q) & \text{при } \tau = 1, 2, \dots, q; \\ 0 & \text{при } \tau > q. \end{cases} \quad (6.23)$$

АКФ получается путем деления (6.23) на дисперсию процесса $\gamma(0)$:

$$\rho(\tau) = \begin{cases} \frac{-\theta_\tau + \theta_1 \theta_{\tau+1} + \theta_2 \theta_{\tau+2} + \dots + \theta_{q-\tau} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & \text{при } \tau = 1, 2, \dots, q; \\ 0 & \text{при } \tau > q. \end{cases} \quad (6.24)$$

Таким образом, АКФ процесса $MA(q)$ равна нулю для всех значений τ , больших порядка q . Это важное характеристическое свойство модели.

Рассмотрим наиболее часто используемый на практике частный случай модели — модель скользящего среднего 1-го порядка $MA(1)$:

$$y_t = \varepsilon_t - \theta \varepsilon_{t-1},$$

где ε_t — белый шум.

Как уже было показано ранее, для обратимости процесса необходимо выполнение условия $|\theta| < 1$.

Очевидно, что $M(y_t) = 0$; $D(y_t) = \sigma_0^2(1 + \theta^2)$.

АКФ согласно (6.24) определяется выражением

$$\rho(\tau) = \begin{cases} \frac{-\theta}{1 + \theta^2} & \text{при } \tau = 1; \\ 0 & \text{при } \tau \geq 2. \end{cases} \quad (6.25)$$

ЧАКФ $\rho_q(\tau)$ задается выражением¹

$$\rho_q(\tau) = -\theta^\tau \frac{1 - \theta^2}{1 - \theta^{2(\tau+1)}}. \quad (6.26)$$

Поведение ЧАКФ определяется затухающей экспонентой. Если значение $\rho(1)$ положительно, то параметр $\theta < 0$, следовательно, $\rho_q(\tau)$ осциллирует с переменным знаком. Если значение $\rho(1)$ отрицательно, то параметр $\theta > 0$, следовательно, все значения $\rho_q(\tau)$ отрицательны.

Проанализированные свойства моделей скользящих средних позволяют сформулировать следующие *практические рекомендации* по их идентификации.

Для моделей МА(1):

- автокорреляционная функция имеет выброс (пик) при лаге, равном 1, а остальные значения статистически незначимы;
- частная автокорреляционная функция экспоненциально затухает (либо монотонно, либо осциллируя, т.е. меняя знак).

Для моделей МА(2):

- автокорреляционная функция имеет выбросы (пики) на лагах, равных 1 и 2, а остальные значения статистически незначимы;
- частная автокорреляционная функция имеет форму синусоиды или экспоненциально затухает.

6.1.3. Авторегрессионные модели со скользящими средними в остатках

В § 6.1.2 была показана двойственность в представлении авторегрессионных процессов (AR) и моделей скользящих средних (MA). Например, процесс МА(1) может быть записан как бесконечный авторегрессионный процесс, и наоборот, процесс AR(1) может быть сведен к процессу скользящих средних МА.

На практике в целях экономичного описания анализируемого процесса в модель могут быть включены как члены, описывающие авторегрессионные составляющие, так и члены, моделирующие

¹ Вывод см. в [5]. Вып. 1. С. 87.

щие остаток в виде процесса скользящих средних. Такой процесс называется *процессом авторегрессии скользящего среднего* — *APCC* (p, q) или, как принято в англоязычной литературе, *AutoRegressive-Moving Average* (*ARMA* (p, q)). Параметры p и q определяют соответственно порядок авторегрессионной составляющей и порядок скользящих средних.

Модель *ARMA*(p, q) имеет вид:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (6.27)$$

Такая модель может интерпретироваться как линейная модель множественной регрессии, в которой в качестве объясняющих переменных выступают прошлые значения самой зависимой переменной, а в качестве регрессионного остатка — скользящие средние из элементов белого шума.

Чтобы процесс (6.27) был стационарным, необходимо и достаточно, чтобы все корни характеристического уравнения *AR*(p)-процесса лежали вне единичного круга:

$$1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_p z^p = 0. \quad (6.28)$$

Аналогично, для обратимости процесса (6.27) необходимо и достаточно, чтобы все корни характеристического уравнения процесса *MA*(q) лежали вне единичного круга:

$$1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_q z^q = 0. \quad (6.29)$$

Рассмотрим простейший смешанный процесс *ARMA*(1,1):

$$y_t = \alpha y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}. \quad (6.30)$$

Уравнение (6.30) можно преобразовать к виду:

$$y_t - \alpha y_{t-1} = \varepsilon_t - \theta \varepsilon_{t-1}. \quad (6.31)$$

Стационарность процесса *ARMA*(1,1) обеспечивается условием $|\alpha| < 1$, а обратимость, в свою очередь, гарантируется выполнением условия $|\theta| < 1$.

Применив ранее описанный и использованный подход для определения автоковариационной функции, получим:

$$\gamma(0) = \frac{(1 + \theta^2 - 2\alpha\theta)\sigma_0^2}{1 - \alpha^2}; \quad (6.32)$$

$$\gamma(1) = \frac{(1 - \alpha\theta)(\alpha - \theta)\sigma_0^2}{1 - \alpha^2}. \quad (6.33)$$

Значение автоковариационной функции для лага больше 1 определяется следующим рекуррентным соотношением:

$$\gamma(\tau) = \alpha\gamma(\tau - 1), \quad \text{при } \tau > 1; \quad (6.34)$$

Следовательно, значения АКФ будут иметь вид

$$\rho(1) = \frac{(1 - \alpha\theta)(\alpha - \theta)}{1 + \theta^2 - 2\alpha\theta}; \quad (6.35)$$

$$\rho(\tau) = \alpha\rho(\tau - 1) = \alpha^{\tau-1}\rho(1) \quad \text{при } \tau > 1. \quad (6.36)$$

Из (6.35), (6.36) видно, что, хотя выражение для $\rho(1)$ отличается от соответствующего выражения для процесса $AR(1)$, соотношение между $\rho(1)$ и последующими значениями АКФ такое же. Таким образом, для процесса $ARMA(1,1)$ значения АКФ будут экспоненциально убывать от значения $\rho(1)$, причем если α положительно, — то монотонно, если отрицательно, — то знакопеременно.

Поведение ЧАКФ определяется начальным значением $\rho_4(1)$, после которого функция экспоненциально убывает. Если θ положительно, то функция убывает монотонно, если отрицательно, — то знакопеременно.

Исследования показывают, что при использовании в экономических задачах модели $ARMA(p, q)$, потребностям практики, как правило, удовлетворяют следующие пять видов этой модели, представленных в табл. 6.1.

Таблица 6.1. Свойства автокорреляционных и частных автокорреляционных функций

Функция	$ARMA(1,0)$	$ARMA(2,0)$	$ARMA(0,1)$	$ARMA(0,2)$	$ARMA(1,1)$
АКФ	Экспоненциально затухает (монотонно или знакопеременно)	Экспоненциально затухает или имеет форму синусоидальной волны	Выброс (пик) на лаге 1	Выбросы (пики) на лагах 1, 2	Экспоненциально затухает от значения $\rho(1)$ (монотонно или знакопеременно)
ЧАКФ	Выброс (пик) на лаге 1	Выбросы (пики) на лагах 1, 2	Экспоненциально затухает (монотонно или знакопеременно)	Экспоненциально затухает или имеет форму синусоидальной волны	Экспоненциально убывает от значения $\rho_4(1)$ (монотонно или знакопеременно)

В этой таблице указаны свойства АКФ и ЧАКФ для этих моделей. Очевидно, что модели ARMA(1,0) и ARMA(2,0) — это рассмотренные ранее модели AR(1), AR(2), а модели ARMA(0,1) и ARMA(0,2) соответствуют моделям скользящего среднего MA(1) и MA(2).

6.2. Методология Бокса—Дженкинса

Экономические временные ряды за редким исключением нестационарны. Нестационарность чаще всего проявляется в наличии зависящей от времени неслучайной составляющей $f(t)$. Если случайный остаток, полученный вычитанием из исходного ряда его неслучайной составляющей $f(t)$, представляет собой стационарный временной ряд, то исходный ряд называется нестационарным однородным.

Для описания таких рядов используется модель авторегрессии — проинтегрированного скользящего среднего АРПСС(p, d, q) или в англоязычном варианте *Autoregressive Integrated Moving Average model* (ARIMA-model). В специальной литературе она также известна как модель Бокса—Дженкинса, по имени авторов, разработавших этот подход и методологию.

Модель ARIMA используется для описания временных рядов, обладающих следующими свойствами:

- 1) ряд включает (аддитивно) составляющую $f(t)$, имеющую вид алгебраического полинома;
- 2) ряд, получившийся после применения к нему процедур последовательных разностей, может быть описан моделью ARMA(p, q).

Наиболее распространены в эконометрических исследованиях модели ARIMA(p, d, q) со значениями параметров, не превышающими 2. При этом параметры p и q определяют соответственно порядок авторегрессионной составляющей и порядок скользящего среднего (аналогично модели ARMA(p, q)), а параметр d — порядок разности (дискретной производной).

Пусть временной ряд после взятия последовательных разностей (после дифференцирования k -го порядка) стал стационарным, удовлетворяющим модели ARMA(p, q):

$$\Delta^k y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \quad (6.37)$$

где $\Delta^k y_t$ — k -я последовательная разность ряда с уровнями y_t .

Очевидно, что из модели для $\Delta^k y_t$ легко получить модель для исходного ряда. Например, для модели Δy_t это можно сделать с помощью выражения

$$y_t = y_{t-1} + \Delta y_t.$$

Методология построения ARIMA-модели для исследуемого временного ряда включает следующие основные этапы:

- идентификацию пробной модели;
- оценивание параметров модели и диагностическую проверку адекватности модели;
- использование модели для прогнозирования.

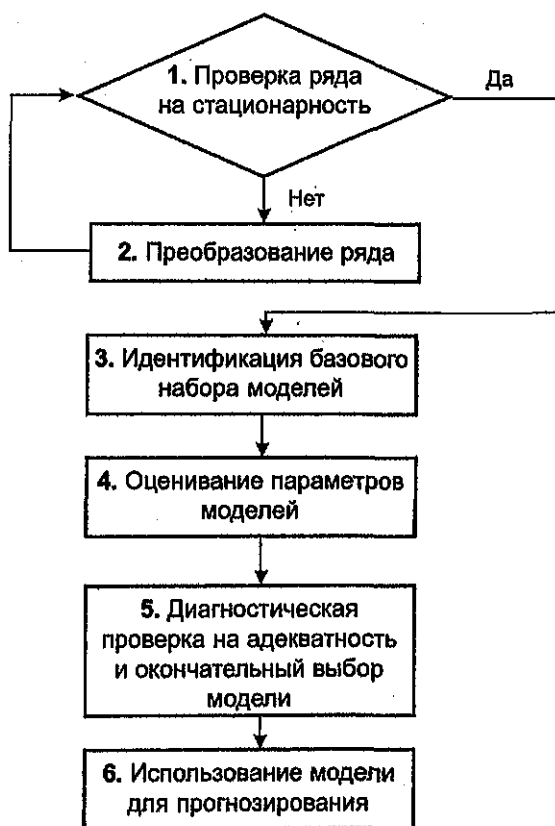


Рис. 6.5. Укрупненная структурная схема подбора модели ARIMA

Поясним содержание некоторых блоков, показанных на рис. 6.5.

Сначала (в блоке 1, 2) необходимо получить стационарный ряд. При тестировании исходных данных на стационарность прежде всего используется визуальный анализ графика. Например, уже на этом этапе можно обнаружить ярко выраженную трендовую или сезонную составляющую.

Также в методике Бокса—Дженкинса рекомендуется проводить анализ АКФ (ЧАКФ). Быстрое убывание значений выборочной АКФ — простой критерий стационарности (аналогичное поведение должна демонстрировать и ЧАКФ).

На рис. 6.6 показан пример АКФ, рассчитанной для временного ряда месячной динамики добычи нефти с газовым конденсатом. Визуальный анализ графика исходного ряда свидетельствовал о нестационарном характере его динамики (наблюдался тренд и периодичность). Значения коэффициентов АКФ демонстрируют медленное затухание, выделяется коэффициент на лаге 12.

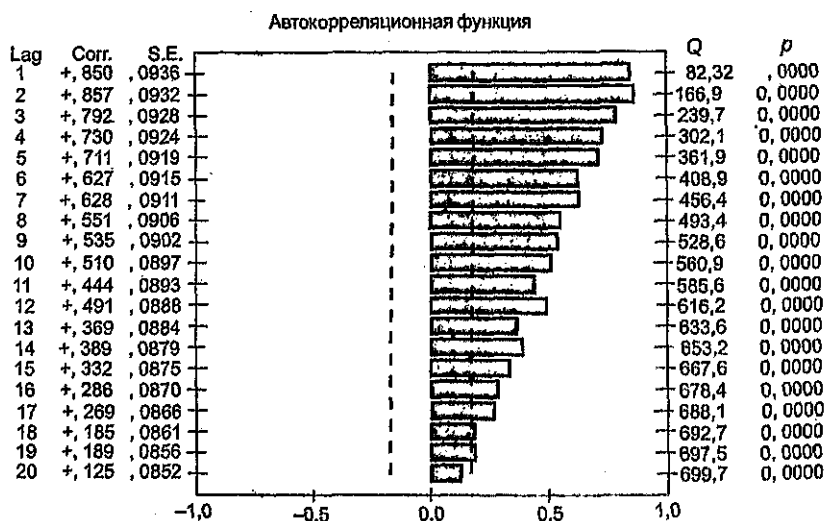


Рис. 6.6. Автокорреляционная функция временного ряда добычи нефти с газовым конденсатом

Также на этом этапе используются ранее описанные статистические тесты на наличие единичного корня (тест Дики—Фуллера, расширенный тест Дики—Фуллера).

Для перехода к стационарному ряду традиционно применяют оператор взятия последовательных разностей (процедуру дискретного дифференцирования). Для определения значения параметра d (порядка разности) может быть использован эвристический кри-

терий, описанный в § 3.3. Также о том, что необходимая для стационарности ряда степень разности достигнута, будет свидетельствовать быстрое затухание АКФ. В практических исследованиях d , как правило, не превышает 2.

В блоке 3 после получения стационарного ряда исследуется характер поведения выборочных АКФ и ЧАКФ и выдвигаются гипотезы о значениях параметров p (порядок авторегрессии) и q (порядок скользящего среднего). При этом следует иметь в виду, что выборочные корреляционные функции могут не демонстрировать детального сходства с теоретическими. Например, умеренно большие значения выборочной АКФ могут наблюдаться после затухания теоретической функции, а также могут наблюдаться всплески, не имеющиеся в теоретической функции. Поэтому для идентификации модели могут использоваться главные черты АКФ при расхождении более тонких деталей. На выходе блока 3 может формироваться базовый набор, включающий 1,2 или даже большее число моделей.

В блоке 4 после осуществления идентификации моделей необходимо оценить их параметры. В современных эконометрических ППП используются разные подходы (МНК, нелинейный МНК; метод максимального правдоподобия (ММП)). Все эти оценки при больших объемах выборок асимптотически эквивалентны [16, 19, 5].

В блоке 5 для проверки каждой пробной модели на адекватность анализируется ее ряд остатков. У адекватной модели остатки должны быть похожими на белый шум, т.е. их выборочные автокорреляции не должны существенно отличаться от нуля.

При проверке значимости коэффициентов АКФ используются два подхода:

- проверка значимости каждого коэффициента автокорреляции отдельно;
- проверка значимости множества коэффициентов автокорреляции как группы.

Первый подход опирается на работу Бартлетта, показавшего, что если модель адекватна исходным данным и ошибки представляют собой белый шум, то распределение коэффициентов автокорреляции приближается к нормальному с нулевым математическим ожиданием и дисперсией $1/n$, т.е. к $N(0; 1/n)$.

Поэтому если выборочный коэффициент автокорреляции r_k выходит за интервал $\pm t_{\gamma} / \sqrt{n}$, то нулевая гипотеза о равенстве нулю коэффициента r_k отвергается.

Второй подход опирается на Q -статистику Бокса—Пирса¹, позволяющую проверить равенство нулю сразу τ первых значений АКФ остатков.

Q -статистика определяется как

$$Q = n \sum_{k=1}^{\tau} r_k^2. \quad (6.38)$$

При нулевой гипотезе об отсутствии автокорреляции статистика Q имеет χ^2 -распределение с $\nu = \tau - p - q$ степенями свободы, где p, q — параметры ARMA-модели.

Если $Q > \chi^2_{\alpha, \nu}$, то как группа первые τ коэффициентов автокорреляции значимы (т.е. не все $\rho_1, \rho_2, \dots, \rho_\tau$ равны нулю).

В некоторых эконометрических пакетах включена модификация этого подхода — тест Бокса—Льюнга². Соответствующая статистика в этом случае определяется выражением:

$$\tilde{Q} = n(n+2) \sum_{k=1}^{\tau} \frac{r_k^2}{n-k}. \quad (6.39)$$

$\tilde{Q}$ имеет такое же асимптотическое распределение, как и Q , однако, как отмечается в [19], ее распределение ближе к χ^2 для конечных выборок. В практических руководствах рекомендуется рассматривать $\tau \approx n/4$ (но не более 50). Кроме того, при построении модели ARIMA необходимо проверить значимость коэффициентов (по t -критерию). При этом модель не должна содержать лишних параметров, т.е. уменьшение числа параметров будет способствовать появлению значимой автокорреляции остатков.

Если в результате проверки несколько моделей оказываются адекватны исходным данным, то при окончательном выборе следует учесть два требования:

- повышение точности (качество подгонки модели);
- уменьшение числа параметров модели.

Воедино эти требования сведены в *информационный критерий Акайка* (Akaike information criterion, AIC), определяемом формулой:

$$AIC = \frac{p+q}{n} + \ln \left(\frac{\sum_{t=1}^n e_t^2}{n} \right), \quad (6.40)$$

где e_t — уровни ряда остатков.

¹ Box G.E.P., Pierce D.A. Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models // J. of the Am. Statistic. Ass. 1970. Vol. 65. — P. 1509—1526.

² Luong G.M., Box G.E.P. On a Measure of Lack of Fit in Time Series Models // Biometrika. 1978. Vol. 65. — P. 297—303.

Очевидно, что выбор следует сделать в пользу модели с меньшим значением AIC.

Аналогичный характер имеет критерий Шварца (*Schwarz criterion*), усиливающий требование уменьшения количества параметров модели:

$$SIC = \frac{(p+q) \ln n}{n} + \ln \left[\frac{\sum_{t=1}^n e_t^2}{n} \right]. \quad (6.41)$$

С помощью выбранной модели в блоке 6 (см. рис.6.5) можно строить точечный и интервальный прогноз на L шагов вперед.

Рассмотренная методология построения моделей ARIMA может быть несколько модифицирована на стадии преобразования ряда для приведения его к стационарному виду (см. блок 2).

Вместо дифференцирования ряда возможно элиминирование трендовой составляющей с помощью моделей кривых роста. При этом выбор «лучшей» модели может осуществляться с помощью критерия минимума среднеквадратического отклонения. К полученному ряду остатков (после проверки его на стационарность) может быть применена ранее рассмотренная методология (блоки 3—6).

Сезонная модель Бокса—Дженкинса может быть представлена в виде:

$$ARIMA(p, d, q)(P_s, D_s, Q_s),$$

где к параметрам модели p, d, q добавлены:

P_s — сезонный параметр авторегрессии;

Q_s — сезонный параметр скользящего среднего;

D_s — порядок сезонной разности (сезонной производной).

Для оценивания сезонных параметров модели в различных эконометрических пакетах требуется информация не менее чем за пять-шесть полных периодов колебаний.

При наличии ярко выраженной сезонной компоненты целесообразно включение в модель сезонного дифференцирования. Однако при решении практических экономических задач не рекомендуется использовать сезонные производные больше первого порядка. Также лучше не применять модели, у которых сумма порядков разностей (сезонных и несезонных) больше 2, то есть желательно выдерживать соотношение

$$d + D_s \leq 2.$$

Определение значений параметров сезонной авторегрессии $SAR(P_s)$ и сезонного скользящего среднего $SMA(D_s)$ также опирается на исследование АКФ и ЧАКФ. Только теперь все типичные проявления, всплески будут удалены друг от друга на значение лага S , где S — период сезонности.

Например, для процесса сезонного скользящего среднего первого порядка $SMA(1)$ будут наблюдаться пики (выбросы) на лагах $S, 2S, 3S$ и т.д. в АКФ, а значения коэффициентов ЧАКФ после лага S не значимы.

Для процесса сезонной авторегрессии 1-го порядка $SAR(1)$, наоборот, будут выделяться выбросы на лагах $S, 2S, 3S$ и т.д. в значениях ЧАКФ, а коэффициенты АКФ после лага S будут не значимы.

Опыт практического использования моделей этого класса свидетельствует о том, что не следует строить смешанных моделей, лучше избегать одновременного включения SAR - и SMA -составляющих.

При идентификации полной сезонной модели $ARIMA(p, d, q)(P_s, D_s, Q_s)$ часто сначала используют логарифмирование исходного ряда (для снижения дисперсии процесса), затем берут одну несезонную и одну сезонную разности (дифференцирование для несезонной части может и не потребоваться). На следующем этапе исследуется поведение АКФ и ЧАКФ для полученного производного ряда.

В некоторых эконометрических пакетах (например, в «X-12-ARIMA», «Мезозавр») реализованы процедуры автоматического подбора структуры модели Бокса—Дженкинса.

Работа с системой «Эксперт» в ППП «Мезозавр» построена по следующей схеме. На первом шаге определяются необходимый порядок разности d и целесообразность включения константы в модель.

Затем происходящий перебор моделей отражается на экране в табличной форме: по вертикали отложен порядок модели $AR(p)$, по горизонтали — порядок модели $MA(q)$.

Рассмотренные модели разбиваются на четыре группы:

- хорошие;
- недоопределенные (предполагается, что следует включить в модель дополнительные параметры);
- переопределенные (предполагается, что модель содержит лишние параметры);
- непонятные.

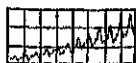
Работа с процедурой «Сезонный эксперт» строится по аналогичной схеме. После установления периода сезонности и опре-

деления порядка простой и сезонной разности осуществляется перебор моделей. Табличное представление перебираемых моделей организовано следующим образом. По вертикали отложен порядок SAR — P_s , по горизонтали — порядок SMA — Q_s . Внутри этих больших клеток маленькие клеточки определяют значения простых параметров AR — p (по вертикали) и MA — q (по горизонтали). В этом ППП предусмотрены определенные ограничения на значения параметров модели (p, q, P_s, Q_s).

Для оценки качества модели, так же как и в других эконометрических пакетах, используются АКФ и ЧАКФ остаточного ряда, статистические характеристики оценок параметров модели, представительный набор статистических критериев.

Однако окончательный выбор модели все же должен оставаться за исследователем-экспертом, хорошо представляющим предметную область. Иногда можно построить две модели, которые одинаково хорошо соответствуют данным на ретроспективном участке (например, модель с порядком дифференцирования 1 и AR-членами и модель с большим порядком дифференцирования и MA-членами). Однако предлагаемые прогнозы у этих моделей могут существенно различаться. Поэтому окончательный выбор между такими моделями должен опираться на представления исследователя о виде нестационарности исходного ряда, о характере его трендовой составляющей.

Успех применения мощного, гибкого, но в то же время сложного аппарата модели ARIMA во многом зависит от практического опыта и квалификации исследователя, а процедуры автоматического выбора вида модели призваны лишь облегчить его аналитическую деятельность.



Практикум к главе 6

Контрольные вопросы

1. Назовите виды моделей стационарных временных рядов.
2. Перечислите основные свойства марковского процесса — AR (1).
3. Что такое процесс случайного блуждания?
4. Для чего используется тест Дики—Фуллера?
5. Поясните, в каких случаях обращаются к расширенному тесту Дики—Фуллера.
6. Как выглядит модель, описывающая процесс Юла? Каковы условия стационарности процесса AR (2)?
7. Охарактеризуйте поведение АКФ и ЧАКФ для AR (2) и для MA (2).

8. Какие параметры имеются в модели «сезонная ARIMA»? Как на практике осуществляется подбор этих параметров?
9. Как выглядят уравнения авторегрессионных моделей со скользящими средними в остатках? Каковы для них условия стационарности и обратимости?
10. Раскройте суть основных этапов применения методологии Бокса—Дженкинса.
11. Для какой цели при построении модели ARIMA используется тест Бокса—Льюнга?
12. Что понимается под информационным критерием Акайка?

Упражнения

- 6.1. Временной ряд описывается следующей моделью:

$$y_t = 0,5y_{t-1} + \varepsilon_t, \quad (1)$$

где ε_t — белый шум.

- а) Как называется модель (1)?

Можно ли считать временной ряд y_t стационарным?

Чему равен корень характеристического уравнения для модели (1)?

- б) Нарисуйте АКФ временного ряда y_t , изобразив на графике значения коэффициентов автокорреляции для лагов $\tau = 1, 2, 3, 4$.

- в) Определите $D(y_t)$, если $D(\varepsilon_t) = \sigma_0^2 = 0,25$.

- 6.2. Временной ряд описывается следующей моделью:

$$y_t = 0,9y_{t-1} - 0,2y_{t-2} + \varepsilon_t, \quad (2)$$

где $M\varepsilon_t = 0$;

$$M(\varepsilon_t \varepsilon_{t+\tau}) = \begin{cases} \sigma_0^2 & \text{при } \tau = 0; \\ 0 & \text{при } \tau \neq 0. \end{cases}$$

- а) Как называется модель (2)?

Чему равны корни характеристического уравнения этой модели?

Можно ли считать временной ряд y_t стационарным?

- б) Нарисуйте АКФ временного ряда y_t , изобразив на графике значения коэффициентов автокорреляции для лагов $t = 1, 2, 3, 4$.

- в) Определите $D(y_t)$, если $\sigma_0^2 = 0,24$.

- 6.3. Временной ряд описывается следующей моделью:

$$y_t = -0,4y_{t-1} + \varepsilon_t - 0,5\varepsilon_{t-1}, \quad (3)$$

где $M\varepsilon_t = 0$;

$$M(\varepsilon_t \varepsilon_{t+\tau}) = \begin{cases} \sigma_0^2 & \text{при } \tau = 0; \\ 0 & \text{при } \tau \neq 0. \end{cases}$$

а) Как называется модель (3)?

Можно ли считать временной ряд y_t стационарным и обратимым?

б) Рассчитайте значения АКФ временного ряда y_t для лагов $\tau = 1, 2, 3$.

в) Определите $D(y_t)$, если $\sigma_0^2 = 0,3$.

6.4. Докажите, что для процесса МА(1) автокорреляционная функция $\rho(\tau) = 0$ при $\tau \geq 2$.

6.5. Временной ряд описывается следующей моделью:

$$y_t = \varepsilon_t - 0,9\varepsilon_{t-1} + 0,2\varepsilon_{t-2}, \quad (4)$$

где ε_t — белый шум.

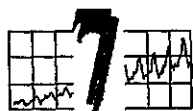
а) Как называется модель (4)?

Чему равны корни характеристического уравнения для этой модели?

Говорят ли их значения об обратимости ряда?

б) Рассчитайте значения АКФ временного ряда y_t для лагов $\tau = 1, 2$.

в) Определите $D(y_t)$, если $D(\varepsilon_t) = \sigma_0^2 = 0,25$.



7.1. Особенности современного статистического программного обеспечения

В последнее время наблюдается стремительное расширение рынка эконометрического программного обеспечения. Можно выделить две группы пакетов, активно применяемых сегодня в экономико-статистических исследованиях [19].

Первую группу образуют пакеты, написанные до появления персональных компьютеров и ориентированные на технологию работы с большими ЭВМ. Как правило, они представляли собой набор процедур, реализованных на языке Fortran. При переходе к персональным компьютерам в них был добавлен командный язык (или простая система меню, за которой скрывается этот язык) для управления вызовом процедур. Поэтому эти программные продукты уступают пакетам, специально написанным для РС по таким параметрам, как дизайн, пользовательский интерфейс, быстрое действие, графические возможности. Однако эти пакеты и ныне занимают прочные позиции на рынке статистических систем. Это объясняется тем, что они создавались и развивались в течение длительного периода большими коллективами разработчиков, содержат реализацию массы разнообразных процедур, тонко учитывающую их специфику, особенности.

Вторую группу образуют пакеты, специально написанные для РС, их программная реализация осуществлялась, как правило, на языке С. Это привело к появлению более гибкого интерфейса пользователя, расширению интерактивных возможностей графического интерфейса (например возможностей графического редактора).

В качестве примеров статистических систем первой группы можно указать SAS, SPSS, STATA, SYSTAT, второй группы — STATISTICA, EViews.

Опыт практических исследований показывает, что для работы часто бывает недостаточно одного пакета. Поэтому в процессе преподавания дисциплин эконометрического блока, в частности статистических методов прогнозирования, в МЭСИ используется комплекс статистических программных продуктов. В него входят отечественные ППП «Олимп» (МЭСИ) и «Мезозавр» (ЦЭМИ

РАН), а также англоязычные статистические системы STATISTICA, SPSS, SAS. В учебном процессе эти программные продукты не конкурируют, а дополняют друг друга.

Использование этого программного обеспечения позволяет создать уникальную среду, в которой статистическая обработка данных становится не рутинным занятием, а увлекательным исследованием, позволяющим получать многовариантные решения с использованием компьютерных технологий и современных методов. Знакомство с несколькими статистическими системами позволяет использовать достоинства каждой из них, почувствовать особенности реализации конкретных процедур в различных пакетах.

Конечно, окончательный выбор программного обеспечения зависит от постановки задачи, квалификации пользователя-аналитика, его опыта и вкусов.

В этой главе рассматривается SPSS — программный продукт, широко востребованный на практике, содержащий реализацию всех рассмотренных в данном пособии классов моделей.

7.2. Реализация моделей кривых роста в SPSS

Для построения моделей кривых роста в системе SPSS используются процедуры регрессионного анализа. Для реализации этих моделей надо из главного меню выполнить следующую последовательность действий: **Analyze** $\Rightarrow$ **Regression** $\Rightarrow$ **Curve Estimation** (Оценивание кривой). Появится диалоговое окно, показанное на рис. 7.1.

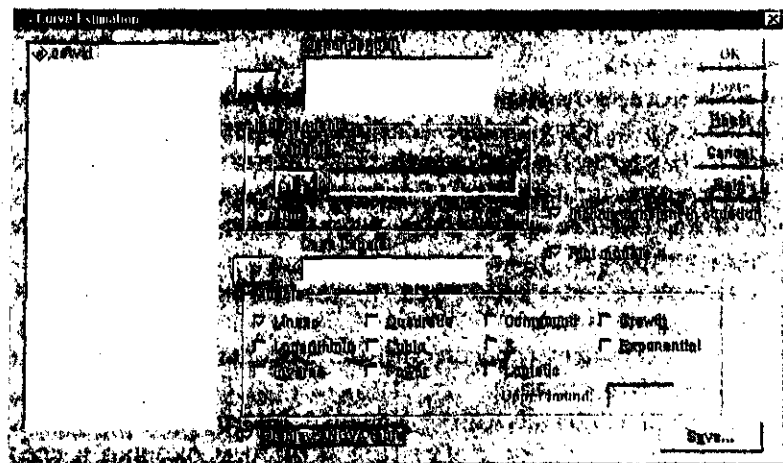


Рис. 7.1. Диалоговое окно «Оценивание кривой» (Curve Estimation)

В левой его части находится список существующих переменных в текущем файле данных. Из него необходимо выбрать переменную — временной ряд для анализа и прогнозирования и поместить ее в окно **Dependent (s)** (Зависимая). Для этого слева от него есть стрелка переноса. Ниже располагается окно **Independent** (Независимая). В него можно поместить любую независимую переменную, выбрав ее из списка, либо в качестве независимой переменной взять время (**Time**). При реализации кривых роста, естественно, выбирается второй вариант. Ниже находится окно **Case Labels** (Метки наблюдений). В него можно поместить переменную из списка, значения которой будут использованы для обозначений каждой точки графика. Далее располагается список трендовых моделей (**Models**). Следует отметить, что это — представительный перечень из одиннадцати моделей, охватывающих различные типы динамики. Например, для моделирования неограниченного в прогнозируемом периоде роста можно использовать линейную модель (**Linear**), параболическую (**Quadratic**), экспоненциальную (**Exponential**) и другие. Для процессов с насыщением в прогнозируемом периоде можно применить S-образные кривые, например логистическую (**Logistic**), причем при выборе этой модели пользователь строкой ниже должен будет задать оценку верхнего предела (**Upper bound**).

Для получения необходимой информации о виде любой модели следует щелкнуть на ней правой кнопкой мыши — появится встроенная в этом окне система помощи (**Help**), содержащая уравнение модели. Отметим, что пользователь с помощью флажков может осуществить выбор одновременно нескольких интересующих его моделей.

Кроме того, в этом диалоговом окне по умолчанию установлены флажки в опциях **Include constant in equation** (Включить константу в уравнение) и **Plot models** (Построить графики моделей).

Нажатием кнопки **Save** (Сохранить) вызывается диалоговое окно, позволяющее определить режим сохранения получаемых результатов (рис. 7.2).

В этом окне слева находится группа опций **Save Variables** (Сохранить переменные):

- **Predicted values** (предсказанные, модельные значения);
- **Residuals** (остатки);
- **Prediction intervals** (интервальный прогноз). При установке флажка в этой позиции становится активным окошко в следующей строке **% Confidence interval** (уровень вероятности для расчета интервального прогноза). По умолчанию интервальные оценки

определяются для 95% уровня, имеется возможность изменить это значение на 90% или 99%.

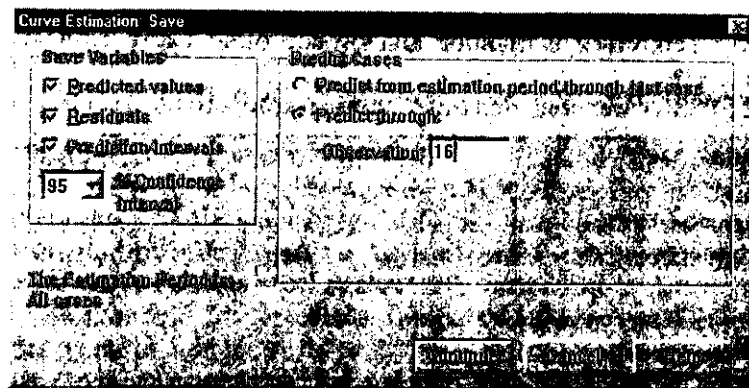


Рис. 7.2. Диалоговое окно сохранения результатов (Curve Estimation: Save)

После проведения расчетов новые созданные переменные будут помещены в рабочую область текущего файла данных. При этом их имена формируются следующим образом:

- имя переменной, содержащей модельные значения, — буквенная часть имени fit, знак подчеркивания, номер;
- имя переменной, содержащей остатки, — буквенная часть имени err, знак подчеркивания, номер;
- нижняя граница интервала — буквенная часть имени LCL, знак подчеркивания, номер;
- верхняя граница интервала — буквенная часть имени UCL, знак подчеркивания, номер.

Справа в диалоговом окне Curve Estimation: Save — группа опций Predict Cases (Предсказывать наблюдения). Пользователю предоставляется выбор следующих возможностей:

- Predict from estimation period through last case (Предсказывать до последнего наблюдения на периоде оценивания);
- Predict through (Предсказать до) — может использоваться для получения расчетных значений, следующих за последним наблюдением на периоде оценивания.

При установке флажка в этой позиции строкой ниже в окошке Observation (Наблюдение) следует определить конец периода моделирования. Способ задания зависит от того, каким образом ранее была определена переменная. Можно указать дату, соответствующую последнему моделируемому наблюдению (в соответствии с тем, как пользователь определил даты ранее в процедуре Define

Dates (Определить даты)), можно задать порядковый номер этого наблюдения.

Сделав все необходимые установки, с помощью кнопки **Continue** следует вернуться в главное диалоговое окно **Curve Estimation** и нажать **OK**.

В SPSS Viewer появится отчет, пример которого показан на рис.7.3.

MODEL: MOD_4.

Independent: Time

Dependent	Mth	Rsq	d.f.	F	Sigf	b0	b1	b2
Y	QUA	.997	19	2849,16	.000	4994,55	652,086	85,9560

The following new variables are being created:

Name Label

.FIT_1 Fit for Y from CURVEFIT, MOD_4 QUADRATIC

ERR_1 Error for Y from CURVEFIT, MOD_4 QUADRATIC

LCL_1 95% LCL for Y from CURVEFIT, MOD_4 QUADRATIC

UCL_1 95% UCL for Y from CURVEFIT, MOD_4 QUADRATIC

ОСТАТКИ ВКЛАДОВ НАСЕЛЕНИЯ В БАНКАХ

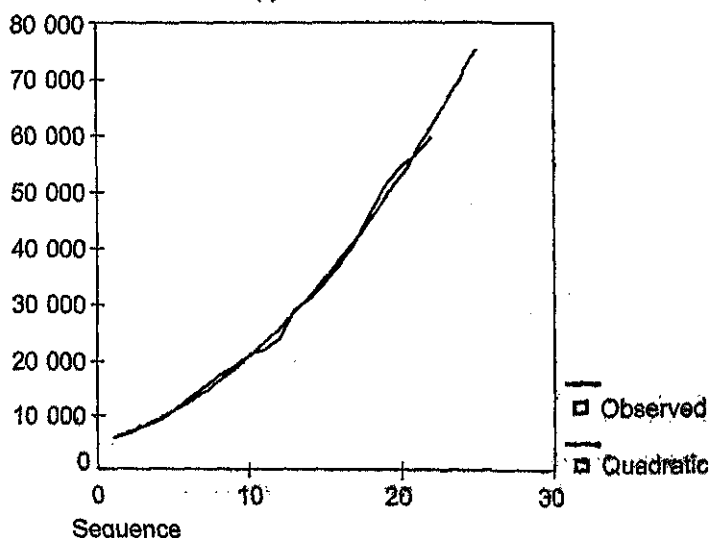


Рис 7.3. Результаты оценивания модели кривой роста с помощью процедуры *Curve Estimation*

Из приведенного на рис. 7.3 примера видно, что уравнение параболической модели (модель обозначена сокращенно **QUA** — квадратичная) имеет вид:

$$\hat{y}_t = 4994,55 + 652,086t + 85,956t^2,$$

где в качестве независимой переменной выступает время t .

Сообщение **The following new variables are being created** говорит о создании новых переменных:

FIT_1 — переменная, содержащая расчетные, модельные значения;

ERR_1 — ряд остатков;

LCL_1 и **UCL_1** — переменные, содержащие нижние и верхние границы доверительных интервалов.

Значения этих переменных можно увидеть в окне **SPSS Data Editor**.

Приведенный в отчете график отражает фактические, наблюдаемые уровни временного ряда (**Observed**) и расчетные, полученные по модели квадратичной параболы (**Quadratic**). При этом, как видно из графика, модельные значения приводятся как для периода оценивания, периода наблюдения, так и для последующих уровней.

Очевидно, что при исследовании качества модели недостаточно только проанализировать приведенный график, необходим всесторонний анализ ряда остатков. Пользователь может построить график остатков, проанализировать дескриптивные статистики, построить гистограмму распределения ряда остатков, изучить автокорреляционную функцию и частную автокорреляционную функцию остатков и т.д. Система SPSS предоставляет пользователю удобную среду для реализации всех этих этапов анализа полученной модели.

7.3. SPSS: построение моделей, основанных на экспоненциальном сглаживании

Для вызова процедур, основанных на экспоненциальном сглаживании, необходимо выбрать пункт **Analyze** главного меню, затем **Time Series** (Временные ряды) и **Exponential Smoothing** (Экспоненциальное сглаживание).

В левой части окна (рис. 7.4) находятся имена переменных текущего файла данных. Выбрав нужную числовую переменную, следует перенести ее имя в окно **Variables** (Переменные) с помощью верхней кнопки со стрелкой.

Группа опций **Model** (Модель) делает доступными четыре типа моделей.

Установив флажок в нужной позиции, пользователь может выбрать следующий вид модели:

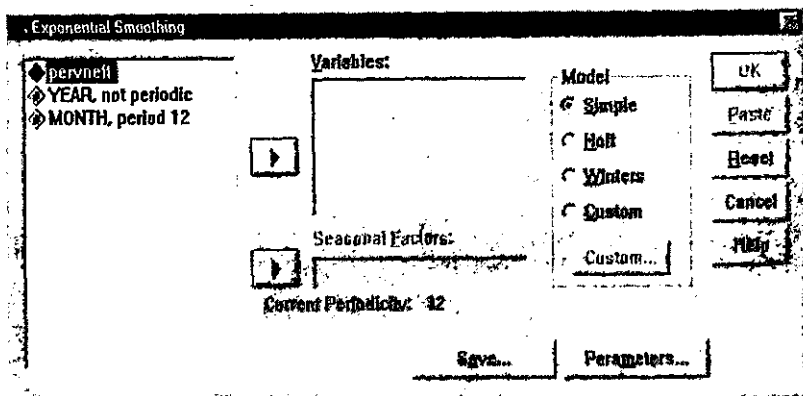


Рис. 7.4. Вид окна «Экспоненциальное сглаживание»
(Exponential Smoothing)

- **Simple** — модель простого экспоненциального сглаживания без трендовой и сезонной составляющих;
- **Holt** — модель Хольта (модель линейного роста без сезонности);
- **Winters** — модель Уинтерса (модель с мультипликативной сезонностью и линейным характером тренда);
- **Custom** — настраиваемый пользователем вид модели (предоставляется возможность выбора комбинации тренда и сезонности).

При активизации опции **Custom**, вызывается диалоговое окно выбора типа модели **Custom Model** (рис. 7.5).

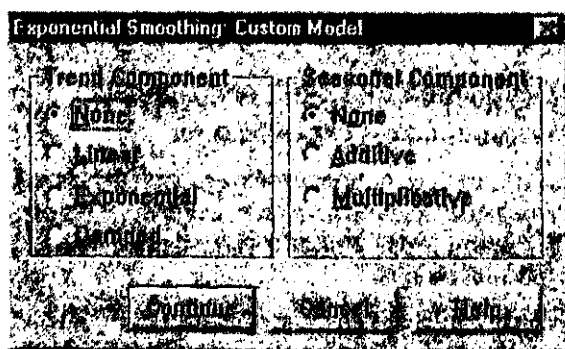


Рис. 7.5. Окно задания пользователем типа модели
(Exponential Smoothing: Custom Model)

Здесь находятся две группы опций, определяющих соответственно характер трендовой (Trend Component) и сезонной (Seasonal Component) составляющих.

В разделе **Trend Component** следует выбрать один из следующих вариантов:

- **None** — тренд отсутствует;
- **Linear** — линейный тренд;
- **Exponential** — экспоненциальный тренд;
- **Damped** — затухающий тренд.

Характер сезонности (**Seasonal Component**) может быть задан следующим образом:

- **None** — нет сезонной составляющей;
- **Additive** — аддитивная сезонная составляющая;
- **Multiplicative** — мультипликативная сезонная составляющая.

После выбора нужной комбинации нажатием кнопки **Continue** следует вернуться к предыдущему окну (**Exponential Smoothing**).

В нижней части окна **Exponential Smoothing** располагаются две кнопки **Save** (Сохранить) и **Parameters** (Параметры). Нажатием кнопки **Save** вызывается диалоговое окно, позволяющее определить режим сохранения получаемых результатов (рис. 7.6).

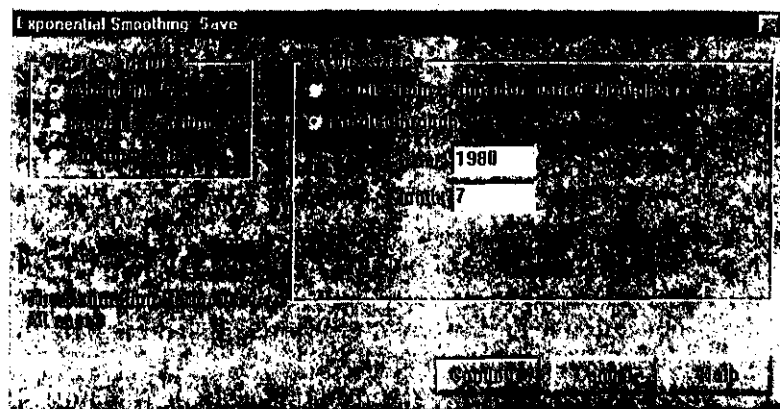


Рис. 7.6. Диалоговое окно сохранения результатов (*Exponential Smoothing: Save*)

Группа опций **Create Variables** (создать переменные) при создании новых переменных позволяет выбрать один из следующих вариантов:

- **Add to file** — добавить новые создаваемые процедурой экспоненциального сглаживания переменные в рабочий текущий файл данных;
- **Replace existing** — заменить существующие переменные;
- **Do not create** — не создавать новые переменные.

Первый вариант задается по умолчанию. При этом имя новой переменной состоит из трех букв, знака подчеркивания и номера. Во втором случае имя переменной вместо знака подчеркивания содержит символ #; при этом сама переменная рассматривается как временная.

С помощью опций **Predict Cases** (Предсказать наблюдения) можно задать следующие режимы:

- **Predict from estimation period through last case** — получить модельные значения только за оцениваемый период, т.е. до последнего наблюдения;
- **Predict through** (Предсказать до) — может использоваться для получения расчетных значений, следующих за последним фактическим наблюдением.

При установке флажка в этой позиции строкой ниже следует определить конец периода моделирования. Способ задания зависит от того, каким образом ранее была определена переменная: можно указать дату, соответствующую последнему моделируемому наблюдению (в соответствии с тем, как даты были определены ранее в процедуре **Define Dates** (Определить даты)), можно задать его порядковый номер.

На рис. 7.6 значение последнего уровня исходного ряда относилось к июню 1980 г. Выбор комбинации, приведенной в распечатке экрана, позволяет получить модельные значения за весь оцениваемый период и сделать прогноз на июль (на рисунке — 7. 1980). Кроме того, в рабочий текущий файл будет добавлена новая переменная, содержащая модельные значения.

После выбора нужной комбинации с помощью кнопки **Continue** следует вернуться к предыдущему окну (**Exponential Smoothing**).

Нажатием кнопки **Parameters** вызывается диалоговое окно установки параметров модели (рис. 7.7).

После выбора переменной и типа модели в главном окне **Exponential Smoothing** в этом окне активизируются те параметры, которые участвуют в реализации заданного вида модели. Всего предусмотрены четыре параметра:

- **General (Alpha)** (Общий (альфа)),
- **Trend (Gamma)** (Трендовый (гамма)),
- **Seasonal (Delta)** (Сезонный (дельта)),
- **Trend Mod. (Phi)** (Изменения тренда (фи)).

Не все из них могут быть активными при построении выбранного типа модели. Например, всегда будет доступен параметр **General (Alpha)**, параметр **Seasonal (Delta)** будет активен лишь для сезонных моделей.

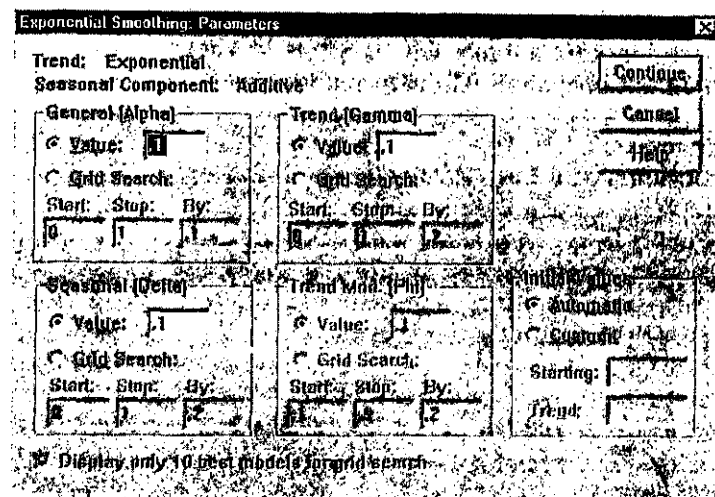


Рис. 7.7. Диалоговое окно «Параметры»
(*Exponential Smoothing: Parameters*)

В системе предусмотрены следующие режимы задания параметров:

- **Value** (Значение) — задание пользователем определенного значения параметра;
- **Grid Search** (Поиск по сетке) — поиск значения параметра на сетке. При выборе этого режима становится активной следующая строка, в которой следует указать границы возможного изменения параметра (**Start** и **Stop**), а также шаг изменения значения при поиске (**By**). Если этот режим установлен для нескольких параметров, то система переберет все возможные комбинации их значений.

Флажок в позиции **Display only 10 best models for grid search** обеспечивает показ в отчете 10 лучших комбинаций параметров по критерию минимизации суммы квадратов ошибок (**SSE**), а также соответствующее каждой из этих комбинаций значение критерия.

В этом же окне **Exponential Smoothing: Parameters** имеются опции **Initial Values** (начальные значения). Выбор режима **Automatic** (автоматически) позволяет системе на основе анализа исходных данных самой задавать начальные значения для реализации процедуры сглаживания. Выбор альтернативного режима **Custom** позволяет пользователю осуществлять задание начальных значений (ввод этих значений осуществляется строкой ниже).

После выбора нужной комбинации режимов, с помощью кнопки **Continue** следует вернуться в главное диалоговое окно (**Exponential Smoothing**).

В нем для реализации модели, после установки всех параметров, нужно нажать кнопку **OK**. В **SPSS — Viewer** появится отчет о выполненной операции. Пример такого отчета показан на рис. 7.8.

ExSmooth

MODEL: MOD_3.

Results of EXSMOOTH procedure for Variable Y

MODEL= WINTERS (Linear trend, multiplicative seasonality) Period= 12

Seasonal indices:

1	91,18565
2	88,22166
3	100,80678
4	97,29508
5	98,12681
6	111,33470
7	123,13503
8	121,46574
9	105,79046
10	92,16911
11	80,28828
12	90,18069

Results of EXSMOOTH procedure for Variable Y (CONTINUED)

MODEL= WINTERS (Linear trend, multiplicative seasonality) Period= 12

Initial values:	Series	Trend
	110,78030	2,64773

DFE = 131.

The 10 smallest SSE's are:

Alpha	Gamma	Delta	SSE
,8000000	,1000000	,1000000	17686,49798
,9000000	,1000000	,1000000	17796,39611
,9000000	,1000000	,2000000	18006,45637
,7000000	,1000000	,1000000	18130,22741
,9000000	,1000000	,3000000	18248,05906
,8000000	,1000000	,2000000	18272,94775
,9000000	,1000000	,4000000	18315,45012
,3000000	,1000000	,9000000	18544,03037
,3000000	,1000000	,8000000	18734,07933
,9000000	,1000000	,5000000	18804,75830

The following new variables are being created:

NAME	LABEL
------	-------

FIT_3	Fit for Y from EXSMOOTH, MOD_3 W/A,80 G,10 D,10
ERR_3	Error for Y from EXSMOOTH, MOD_3 W/A,80 G,10 D,10

12 new cases have been added.

Рис. 7.8. Вариант отчета экспоненциального сглаживания (линейный тренд, мультипликативная сезонность), полученного в SPSS — Viewer

Он содержит:

- указание на вид модели;
- **Seasonal indices** (сезонные индексы или коэффициенты сезонности);
- **Initial values** (начальные значения, использованные при сглаживании ряда);
- **The 10 smallest SSE's are:** Alpha Gamma Delta — 10 лучших комбинаций параметров модели, расположенных по возрастанию сумм квадратов ошибок (**SSE — Sums of Squares Errors**);
- сообщение **The following new variables are being created**, говорящее о создании двух новых переменных **fit_3** и **err_3**. Это соответственно модельные значения и остатки, добавленные в рабочий текущий файл данных;
- сообщение **12 new cases have been added** — добавление к расчетным значениям 12 наблюдений, следующих после периода оценивания.

Чтобы появилась возможность использования моделей с сезонной составляющей, необходимо предварительно соответствующим образом определить переменные. Для этого следует в основном меню **Data** выбрать **Define Dates** (Определить даты). Появится окно, показанное на рис. 7.9.

В примере на рис. 7.9 видно, что рассматривается временной ряд месячной динамики с годовой периодичностью, причем первый уровень ряда относится к январю 1970 г. После нажатия кнопки **OK** к существующим переменным в текущем файле данных добавятся новые переменные: год, номер месяца, дата (название месяца и номер года). В модуле **SPSS — Viewer** появится отчет о проведенной операции. После этого для исследуемого временного ряда станут доступными сезонные модели.

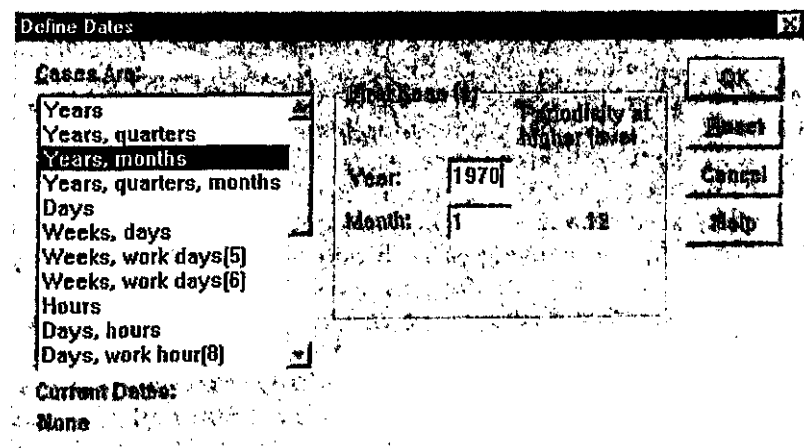


Рис 7.9. Вид окна определения дат (Define Dates)

7.4. Прогнозирование с помощью моделей авторегрессии — проинтегрированного скользящего среднего (ARIMA) в среде SPSS

Для реализации модели авторегрессии — проинтегрированного скользящего среднего в системе SPSS следует вызвать диалоговое окно этой процедуры, проделав следующую последовательность шагов из главного меню: **Analyze** $\Rightarrow$ **Time series** $\Rightarrow$ **ARIMA**.

Вид диалогового окна процедуры **ARIMA** показан на рис. 7.10.

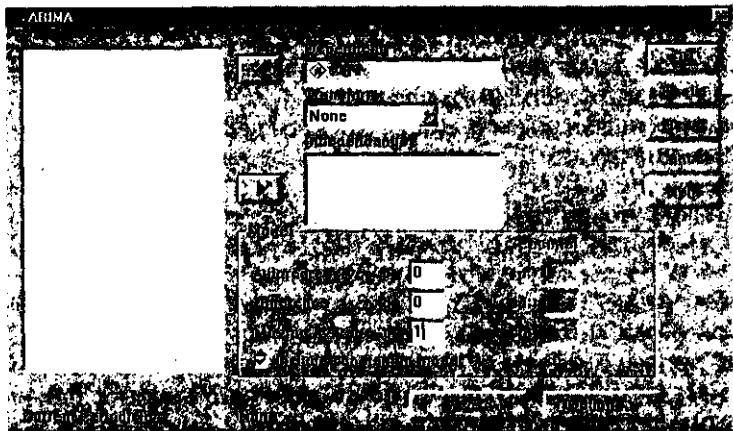


Рис. 7.10. Вид диалогового окна процедуры **ARIMA**

В списке переменных текущего файла следует выбрать анализируемую переменную и перенести ее в окно **Dependent** (Зависимая). Если эта переменная ранее была определена как сезонная с помощью процедуры **Define Dates**, то в группе опций **Model** (Модель) станут доступными окошечки, позволяющие задать параметры для полной сезонной модели **ARIMA** (p, d, q)(sp, sd, sq). В первом столбце задаются параметры несезонной части модели, во втором — аналогичные параметры для сезонной части (**Seasonal**):

- **Autoregressive** (Авторегрессия). p — порядок авторегрессии, sp — сезонный параметр авторегрессии;
- **Difference** (Разность или Дискретная производная). d — порядок разности или порядок дискретной производной, sd — порядок сезонной разности (сезонной производной);
- **Moving Average** (Скользящее среднее), q — порядок скользящего среднего, sq — сезонный параметр скользящего среднего.

Если анализируемая переменная не была заранее описана как переменная, содержащая сезонность, то для пользователя окажутся доступными только окошечки первого столбца. Идентификация модели, установка соответствующих значений параметров в груп-

пе опций **Model** проводится в соответствии с методологией, рассмотренной в гл.6. При реализации этого подхода полезны возможности системы SPSS, описанные в примечании.

В этой же группе по умолчанию установлен флажок в строке **Include constant in model** (Включить константу в модель), по желанию пользователя этот выбор может быть отменен. Кроме того, при необходимости, можно произвести анализ переменной в логарифмической шкале. Для этого следует выбрать вид нужного преобразования из выпадающего списка в окне **Transform** (Преобразовать):

- **None** (Никакого), обрабатывается исходная, непреобразованная переменная;

- **Natural log** (Натуральный логарифм);

- **Log base 10** (Десятичный логарифм).

Таким образом, при выборе логарифмирования обрабатывается преобразованная переменная, представляющая собой натуральный или десятичный логарифм исходной переменной. При этом расчетные, модельные значения (**fit**) и границы доверительных интервалов (**lcl** и **ucl**) будут преобразованы и представлены пользователю уже не в логарифмической шкале, а в исходной.

Далее необходимо установить параметры вычислений и режим сохранения результатов. Нажатие кнопки **Save** (Сохранить) открывает диалоговое окно (рис. 7.11), идентичное окну на рис. 7.6, в котором пользователь определяет режим создания новых переменных, период моделирования. Единственная особенность — при построении модели **ARIMA** (в отличие от процедур экспоненциального сглаживания) имеется возможность расчета доверительных интервалов при заданном уровне вероятности. В окошке **% Confidence intervals** (Уровень вероятности для расчета интервального прогноза) по умолчанию установлен 95-ный уровень, но имеется возможность изменить это значение на 90 или 99%.

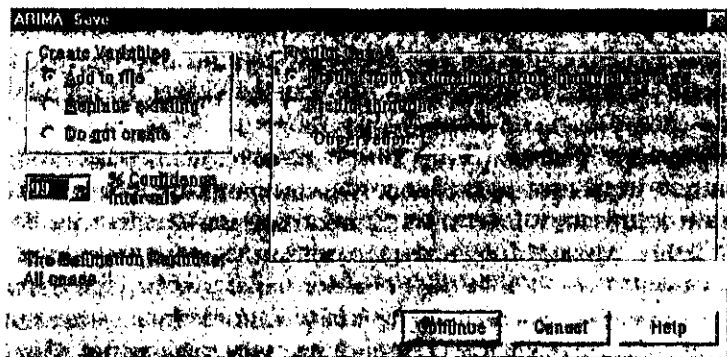


Рис. 7.11. Диалоговое окно сохранения результатов модели **ARIMA** (**ARIMA: Save**)

Нажатие кнопки **Continue** (Продолжить) возвращает в главное диалоговое окно **ARIMA**.

Для установки других параметров следует воспользоваться кнопкой **Options**, откроется диалоговое окно **ARIMA: Options** (рис. 7.12).

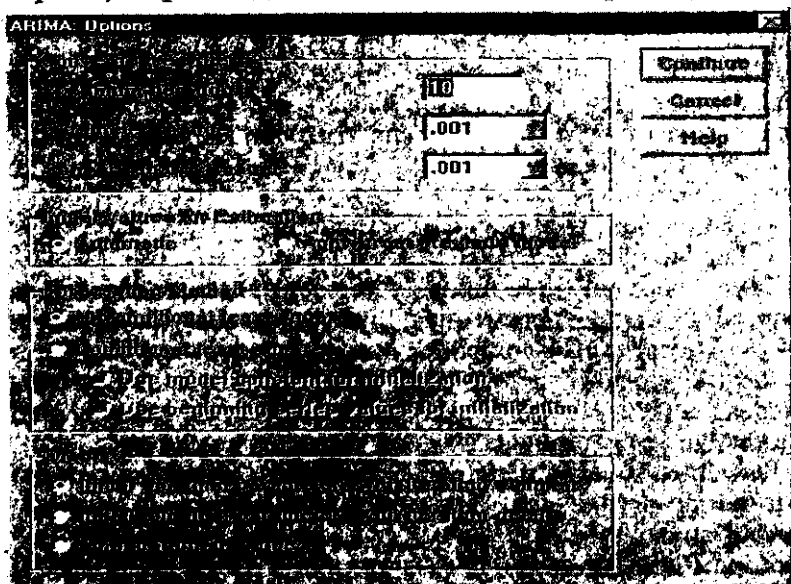


Рис. 7.12. Вид диалогового окна **ARIMA: Options**

Группа опций **Convergence Criteria** (Критерий сходимости) позволяет пользователю управлять процессом прекращения работы вычислительной процедуры.

- **Maximum iterations** (Максимальное число итераций). По умолчанию алгоритм останавливается после 10 итераций, даже если критерий сходимости не выполнен. Однако пользователь может поменять это значение по своему усмотрению, например, увеличить, если остановка алгоритма произошла после 10 итераций, а условия сходимости не были удовлетворены.

- **Parameter change** (Допустимое изменение параметров). По умолчанию итерации прекращаются, если при переходе от одной итерации к другой ни один из параметров не изменяется больше чем на 0,001. Можно изменить это значение, например, для того, чтобы улучшить точность оценивания параметров. При этом может потребоваться увеличение максимального числа итераций.

- **Sum of squares change** (Изменение сумм квадратов). По умолчанию вычислительный процесс прекращается, если при переходе от одной итерации к другой изменение суммы квадратов оши-

бок не превышает 0,001%. Так же как в предыдущем случае, можно изменить это значение, например, для того, чтобы улучшить точность оценивания параметров. При этом может потребоваться увеличение максимального числа итераций.

Группа **Initial Values for Estimation** (Начальные значения для оценивания):

- **Automatic** (Автоматически). Начальные значения подбираются процедурой автоматически;

- **Apply from previous model** (Взять из предыдущей модели). В качестве начальных оценок берутся значения, полученные при предыдущем выполнении процедуры ARIMA (в этой же рабочей сессии). Этот режим может дать выигрыш во времени вычислений, если модель и данные близки к предыдущим.

Группа **Forecasting Method** (Метод прогнозирования) определяет тип вычислительной процедуры для предсказания наблюдений:

- **Unconditional least squares** (Метод наименьших квадратов без дополнительных условий) — эта установка производится по умолчанию;

- **Conditional least squares** (Метод наименьших квадратов с условием). При выборе этого метода становятся активными две следующих опции:

- ♦ **Use model constant for initialization** (Использовать константу модели при инициализации),

- ♦ **Use beginning series values for initialization** (Использовать начальные значения ряда для инициализации).

Группа опций **Display** (Вывести):

- **Initial and final parameters with iteration summary** (Начальные и окончательные параметры и итог итераций). В результате работы процедуры выдаются начальные и окончательные оценки параметров, статистические характеристики, позволяющие судить о качестве оценок, число итераций и причина прекращения счета;

- **Initial and final parameters with iteration details** (Начальные и окончательные параметры и описание итераций). При выборе этого режима, кроме перечисленного, в результате работы процедуры сообщаются оценки параметров, полученные после каждой итерации;

- **Final parameters only** (Только окончательные параметры). ARIMA выдает окончательные оценки параметров, статистические характеристики, позволяющие судить о качестве оценок.

С помощью кнопки **Continue** (Продолжить) следует вернуться в главное диалоговое окно ARIMA и запустить вычислительную процедуру нажатием **OK**.

Фрагмент отчета показан на рис. 7.13.

Conclusion of estimation phase.

Estimation terminated at iteration number 3 because:

Sum of squares decreased by less than,001 percent.

FINAL PARAMETERS:

Number of residuals	368
Standard error	19,772476
Log likelihood	-1620,124
AIC	3242,2479
SBC	3246,156

Analysis of Variance:

DF Adj.	Sum of Squares	Residual Variance
Residuals 367	143683,13	390,95082

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
MA1	,61343664	,04130719	14,850803	,0000000

Рис. 7.13. Фрагмент отчета, полученного после реализации модели ARIMA (0, 1, 1)

Как следует из отчета, вычислительная процедура состояла из трех итераций. Дальнейшие вычисления были остановлены, так как на последней итерации сумма квадратов ошибок уменьшилась меньше, чем на 0,001%.

Среди приведенных статистических характеристик есть AIC (информационный критерий Акайка) и SBC (критерий Шварца—Байеса), которые используются при сравнении нескольких моделей, построенных для анализируемой переменной. Как отмечалось в гл. 6, эти статистики учитывают требования повышения точности и уменьшения числа параметров модели. Выбор делается в пользу модели с меньшим значением этих характеристик.

Также отчет содержит таблицу дисперсионного анализа, характеристики найденных оценок. В столбце B приведена оценка коэффициента модели (модели скользящего среднего первого порядка), в столбце SEB — оценка стандартной ошибки. Их отношение — t -статистика, указанная в третьем столбце, свидетельствует о значимости коэффициента модели (уровень значимости проставлен в четвертом столбце). Очевидно, что при исследова-

нии качества модели необходим всесторонний анализ ряда остатков. Сам ряд остатков, а также расчетные значения, границы доверительных интервалов после установки соответствующих режимов в окне **ARIMA: Save** пользователь может просмотреть с помощью окна **SPSS Data Editor**.

Пользователь может построить график остатков, проанализировать дескриптивные статистики, построить гистограмму распределения ряда остатков, изучить автокорреляционную функцию и частную автокорреляционную функцию остатков и т.д.

Система **SPSS** позволяет реализовать все эти этапы анализа полученной модели.

При построении модели **ARIMA** исследователь, как правило, не ограничивается только возможностями, предоставляемыми в меню **Time series** (Временные ряды), а обращается к дополнительным возможностям системы. Остановимся подробнее на тех функциях, которые наиболее часто востребованы пользователями.

При идентификации модели **ARIMA** на практике, как правило, используют методологию Бокса—Дженкинса, опирающуюся на анализ АКФ (ACF) и ЧАКФ (PACF). Графики АКФ и ЧАКФ оказываются полезными для изучения свойств остатков при анализе качества построенной модели.

Для построения графиков АКФ и ЧАКФ из основного меню **SPSS** следует проделать следующую последовательность действий: **Graphs** $\Rightarrow$ **Time series** $\Rightarrow$ **Autocorrelations**. Вид появляющегося после этого диалогового окна показан на рис. 7.14, а примеры АКФ и ЧАКФ — на рис. 7.15.

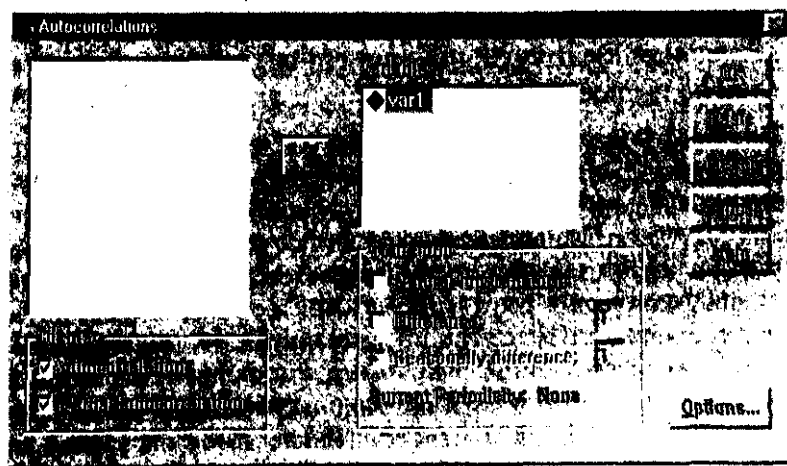


Рис. 7.14. Диалоговое окно процедуры «Автокорреляции»

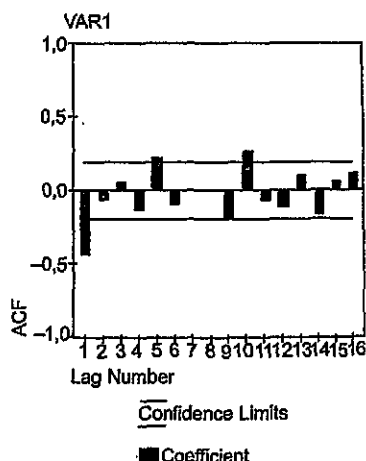
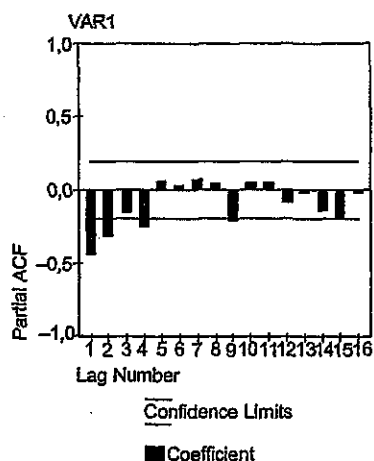


Рис. 7. 15. Графики АКФ (ACF) и ЧАКФ (PACF)

В гл. 6 подробно обсуждался процесс интерпретации подобных графиков, отмечалась их практическая значимость.

Мощный аппарат модели ARIMA занимает видное место среди современных средств анализа и прогнозирования временных рядов. Безусловно, успех применения этой модели во многом зависит от практического опыта и квалификации исследователя, мощные же статистические системы, подобные SPSS, призваны облегчить его аналитическую деятельность, предоставив удобную среду для проведения исследований.

7.5. Сезонная декомпозиция временных рядов с использованием SPSS

Процедура **Seasonal Decomposition** (Сезонная декомпозиция) позволяет выделить отдельные компоненты периодического временного ряда, провести его сезонную корректировку. Полученные на выходе этой процедуры результаты могут представлять самостоятельный интерес, а также использоваться в качестве входной информации при реализации других методов. Предварительно перед проведением сезонной декомпозиции в системе SPSS анализируемый временной ряд должен быть определен как периодический. Для этого используют процедуру **Define Dates** (Определить даты) в основном меню **Data**. После этого для вызова процедуры необходимо выбрать пункт **Analyze** главного меню, затем **Time Series** и **Seasonal Decomposition** (рис. 7.16).

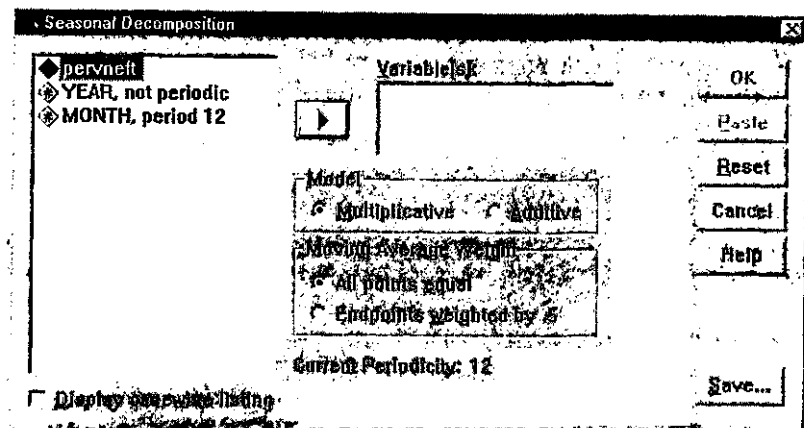


Рис. 7.16. Вид диалогового окна «Сезонная декомпозиция»
(Seasonal Decomposition)

В левой части окна находится список созданных переменных из текущего файла данных. Выбрав нужную числовую переменную, следует перенести ее имя в окно **Variable(s)** (Переменные) с помощью кнопки со стрелкой. Затем необходимо определить тип модели, установив один из двух вариантов в группе **Model** (Модель):

- **Multiplicative** (модель с мультипликативной сезонностью);
- **Additive** (модель с аддитивной сезонностью).

По умолчанию в системе задается первый режим.

Напомним, что при выборе характера сезонности часто оказывается полезным проведение графического анализа динамики временного ряда.

В группе **Moving Average Weight** (Веса скользящей средней) можно задать веса уровней при реализации процедуры скользящей средней:

- **All points equal** — все уровни равнозначны;
- **Endpoints weighted by 0,5** — задать крайним уровням веса 0,5.

Нажатием кнопки **Save** (Сохранить) вызывается диалоговое окно, позволяющее определить режим сохранения получаемых результатов (рис. 7.17).

В этом окне группа опций **Create Variables** (Создать переменные) позволяет выбрать один из следующих вариантов:

- **Add to file** — добавить новые создаваемые процедурой сезонной декомпозиции переменные в рабочий текущий файл данных;
- **Replace existing** — заменить существующие переменные. При этом новые создаваемые процедурой сезонной декомпозиции переменные будут добавлены в рабочий текущий файл данных в качестве временных переменных. Если до этого процедурой уже были

созданы такие временные переменные, то они будут заменены на последний расчетный вариант;

- **Do not create** — не создавать новых переменных в текущем рабочем файле данных.

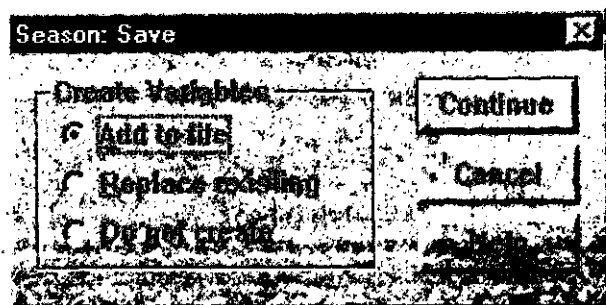


Рис. 7.17. Диалоговое окно сохранения результатов (Season: Save)

Первый вариант задается по умолчанию. При этом имя новой переменной состоит из трех букв, знака подчеркивания и номера. Во втором случае имя переменной вместо знака подчеркивания содержит символ #.

К существующим переменным в текущий файл данных будут добавлены четыре переменные, имена которых начинаются следующим образом:

- **saf** — сезонная составляющая в абсолютном выражении для аддитивной модели, представляющая относительные величины для мультипликативной периодичности;

- **sas** — временной ряд со скорректированной сезонностью;
- **stc** — тренд — циклическая составляющая;
- **err** — нерегулярная составляющая.

После выбора нужного варианта в окне **Season: Save** с помощью кнопки **Continue** следует вернуться к главному диалоговому окну (**Seasonal Decomposition**).

Если в нем установить флажок в **Display casewise listing** (Вывести листинг по наблюдениям), то в отчете, представленном в **SPSS — Viewer**, будет отображаться следующая информация:

- исходные значения анализируемой переменной;
- значения скользящей средней (**Moving averages**);
- разница между исходным значением переменной и значением скользящей средней в аддитивной модели или их отношение в случае мультипликативной сезонности (**Ratios**);
- сезонная составляющая (**Seasonal factors**);
- исходный ряд с поправкой на сезонность (**Seasonally adjusted series**) или, иными словами, сезонно скорректированный ряд. Он

получается из значений исходного ряда либо путем вычитания из него сезонной составляющей (аддитивная модель), либо путем деления его на сезонную составляющую (мультипликативный вариант сезонности);

- тренд-циклическая компонента — сглаженный исходный ряд с поправкой на сезонность (Smoothed trend-cycle);

- нерегулярная составляющая (Irregular component) — разница между исходным рядом с поправкой на сезонность и тренд-циклической компонентой (аддитивная модель) или их отношение в мультипликативном варианте.

После установки всех параметров нужно нажать кнопку **ОК** в главном диалоговом окне (**Seasonal Decomposition**). Вариант отчета, полученного в окне **SPSS — Viewer**, показан на рис. 7.18 (опция **Display casewise listing** не была отмечена).

MODEL: MOD_4.

Results of SEASON procedure for variable AIRPASS
Multiplicative Model. Centered MA method. Period = 12.

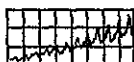
Period	Seasonal index (* 100)
1	91,064
2	88,120
3	100,811
4	97,307
5	98,291
6	111,446
7	122,619
8	121,636
9	105,982
10	92,187
11	80,386
12	90,151

The following new variables are being created:

Name	Label
ERR_1	Error for AIRPASS from SEASON, MOD_4 MUL CEN 12
SAS_1	Seas adj ser for AIRPASS from SEASON, MOD_4 MUL CEN 12
SAF_1	Seas factors for AIRPASS from SEASON, MOD_4 MUL CEN 12
STC_1	Trend-cycle for AIRPASS from SEASON, MOD_4 MUL CEN 12

Рис. 7.18. Вариант отчета процедуры сезонной декомпозиции, полученного в SPSS — Viewer

Процедура сезонной декомпозиции применялась к временно-му ряду месячной динамики с годовой периодичностью сезонных колебаний. Как следует из отчета, был выбран мультипликативный вариант модели. Отчет содержит значения 12 индексов сезонности, выраженные в процентах. Сезонные подъемы показателя приходятся на март, а также на летние месяцы и сентябрь (наибольший всплеск — июль, август). Самый сильный сезонный спад наблюдается в ноябре. Сообщение **The following new variables are being created** говорит о том, что после проведения расчетов в текущий файл данных добавлены к существующим переменным новые: **ERR_1**, **SAS_1**, **SAF_1**, **STC_1**, а созданные системой метки переменных поясняют их содержание.



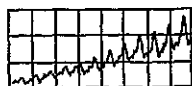
Практикум к главе 7

Упражнения

- 7.1. По данным о среднегодовой численности промышленно-производственного персонала в легкой промышленности (см. упражнение 1 практикума к гл. 5) рассчитайте с помощью SPSS прогнозное значение показателя в 2001 и 2002 г., используя двухпараметрическую модель линейного роста Хольта. Оцените «качество» модели.
- 7.2. Для моделирования среднегодовой численности промышленно-производственного персонала в легкой промышленности (см. упражнение 1 практикума к гл. 5) используйте кривые роста, реализованные в SPSS. Рассчитайте прогнозное значение показателя в 2001 и 2002 г., оцените «качество» выбранной модели, сравните с результатами, полученными в предыдущем упражнении.
- 7.3. Рассчитайте с помощью SPSS простую экспоненциальную среднюю для временного ряда урожайности зерновых культур (см. упражнение 3 практикума к гл. 5) для двух значений параметра адаптации α (0,2 и 0,8). На одном графике покажите три временных ряда: исходный и два сглаженных. Объясните, как повлияло значение параметра адаптации α на результат сглаживания.
- 7.4. В Приложении 2 приводятся данные о месячной динамике пройденного авиалайнерами расстояния. Используя представленные в SPSS тренд-сезонные модели, основанные на экспоненциальном сглаживании, рассчитайте прогнозные значения показателя на следующий год.

Сравните «качество» двух моделей: модели Хольта–Уинтерса и модели с экспоненциальным трендом и мультипликативной сезонностью.

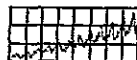
- 7.5. Используя методологию Бокса–Дженкинса, попытайтесь описать динамику временного ряда из предыдущего упражнения с помощью сезонного варианта модели ARIMA, реализованной в SPSS. Оцените «качество» модели, рассчитайте прогнозные значения показателя на следующий год, сравните с результатами, полученными в предыдущем упражнении.
- 7.6. В Приложении 2 приводятся данные об объемах перевозок грузов железнодорожным транспортом в России, млн т. Проанализируйте компонентный состав временного ряда, используя графические средства SPSS. В среде SPSS постройте модель, описывающую динамику временного ряда. Оцените «качество» модели, рассчитайте прогноз на следующий год.



БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики: — М.: ЮНИТИ, 1998.
2. Андерсон Т. Статистический анализ временных рядов. — М.: Мир, 1976.
3. Бендат Дж., Пирсол А. Измерение и анализ случайных процессов: Пер. с англ. — М.: Мир, 1974.
4. Бендат Дж., Пирсол А. Применения корреляционного и спектрального анализа: Пер. с англ. — М.: Мир, 1983.
5. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление: Пер. с англ. — М.: Мир, 1974. — Вып. 1, 2.
6. Болч Б., Хуань К.Дж. Многомерные статистические методы для экономики: Пер. с англ. — М.: Статистика, 1979.
7. Боровиков В.П., Иеченко Г.И. Прогнозирование в системе STATISTICA® в среде Windows. Основы теории и интенсивная практика на компьютере. — М.: Финансы и статистика, 1999.
8. Бриллинджер Д. Временные ряды. Обработка данных и теория. — М.: Мир, 1980.
9. Гренджер К., Хатанака М. Спектральный анализ временных рядов в экономике: Пер. с англ. — М.: Статистика, 1972.
10. Дженкинс Г., Ваттс Д. Спектральный анализ и его приложения: Пер. с англ. — М.: Мир, 1971. — Вып. 1; 1972. — Вып. 2.
11. Доугерти К. Введение в эконометрику: Пер. с англ. — М.: ИНФРА-М, 2001.
12. Кендалл М., Стьюарт А. Многомерный статистический анализ и временные ряды: Пер. с англ. — М.: Наука, 1976.
13. Кендэл М. Временные ряды: Пер. с англ. — М.: Финансы и статистика, 1981.
14. Кильдишев Г.С., Френкель А.А. Анализ временных рядов и прогнозирование. — М.: Статистика, 1973.
15. Кремер Н.Ш., Путко Б.А. Эконометрика / Под ред. Н.Ш. Кремера. — М.: ЮНИТИ-ДАНА, 2002.
16. Лугачев М.И., Ляпунов Ю.П. Методы социального прогнозирования. — М.: МГУ, ТЕИС, 1999.
17. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования. — М.: Статистика, 1979.
18. Льюис К.Д. Методы прогнозирования экономических показателей. — М.: Финансы и статистика, 1986.
19. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика: Начальный курс. — М.: Дело, 2000.

20. *Отнес Р., Эноксон Л.* Прикладной анализ временных рядов: Пер. с англ. — М.: Мир, 1982.
21. *Половников В.А.* Анализ и прогнозирование транспортной работы морского флота. — М.: Транспорт, 1983.
22. *Практикум по эконометрике / И.И. Елисеева, С.В. Курьшева, Н.М. Гордеенко и др. / Под ред. И.И. Елисеевой.* — М.: Финансы и статистика, 2001.
23. *Скучалина Л. Н., Крутова Т. А.* Организация и ведение базы данных временных рядов. Система показателей, методы определения, оценки прогнозирования информационных процессов. — М.: ГКС РФ, 1995.
24. *Статистическое моделирование и прогнозирование / Под ред. А.Г. Гранберга.* — М.: Финансы и статистика, 1990.
25. *Уотшем Т.Дж., Паррамоу К.* Количественные методы в финансах: Пер. с англ. / Под ред. М.Р. Ефимовой. — М.: Финансы, ЮНИТИ, 1999.
26. *Френкель А.А.* Прогнозирование производительности труда: методы и модели. — М.: Экономика, 1989.
27. *Хеннан Э.* Многомерные временные ряды. Пер. с англ. — М.: Мир, 1974.
28. *Хеннан Э.* Анализ временных рядов: Пер. с англ. М.: Наука, 1964.
29. *Четыркин Е.М.* Статистические методы прогнозирования. — М.: Статистика, 1977.
30. *Эконометрика / Под ред. И.И. Елисеевой.* — М.: Финансы и статистика, 2001.
31. *Экономико-математические методы и прикладные модели / Под ред. В.В. Федосеева.* — М.: ЮНИТИ, 1999.
32. *SPSS Trends™ 10.0.* SPSS Inc., 1999.



ПРИЛОЖЕНИЕ 1. Математико-статистические таблицы

Таблица 1. Нормальный закон распределения. Значение функции $\Phi(t) = P(|T| \leq t_{\text{табл}})$

Целые и десятич- ные доли t	Сотые доли t									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0080	0,0160	0,0239	0,0319	0,0399	0,0478	0,0558	0,0638	0,0717
0,1	0797	0876	0955	1034	1113	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2960	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4907	4971	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6679	6729	6778
1,0	0,6827	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7994	8029
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8789	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9090
1,7	9109	9127	9146	9164	9181	9199	9216	9233	9249	9265

[illegible]

Таблица 2. Распределение Стьюдента (t-распределение)

v	Вероятность $\alpha = SK(t) = P(T > t_{\alpha, n})$												
	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,043	6,859
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,327	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,583
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,833
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850

v	Вероятность $\alpha = SK(t) = P(T > t_{\text{табл}})$												
	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,868	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,402	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

**Таблица 3. Распределение Пирсона (χ^2 -распределение)
Значения $\chi^2_{\text{табл}}$ для вероятностей $P(\chi^2 > \chi^2_{\text{табл}})$**

v	Вероятность										
	0,999	0,995	0,99	0,98	0,975	0,95	0,90	0,80	0,75	0,70	0,50
1	0,05157	0,04393	0,03157	0,03628	0,03982	0,00393	0,0158	0,0642	0,102	0,148	0,455
2	0,00200	0,0100	0,0201	0,0404	0,0506	0,103	0,211	0,446	0,575	0,713	1,386
3	0,0243	0,0717	0,115	0,185	0,216	0,352	0,584	1,005	1,213	1,424	2,366
4	0,0908	0,207	0,297	0,429	0,484	0,711	1,064	1,649	1,923	2,195	3,357
5	0,210	0,412	0,554	0,752	0,831	1,145	1,610	2,343	2,675	3,000	4,351
6	0,381	0,676	0,872	1,134	1,237	1,635	2,204	3,070	3,455	3,828	5,348
7	0,598	0,989	1,239	1,564	1,690	2,167	2,833	3,822	4,255	4,671	6,346
8	0,857	1,344	1,646	2,032	2,180	2,733	3,490	4,594	5,071	5,527	7,344
9	1,152	1,735	2,088	2,532	2,700	3,325	4,168	5,380	5,899	6,393	8,343
10	1,479	2,156	2,558	3,059	3,247	3,240	4,865	6,179	6,737	7,267	9,342
11	1,834	2,603	3,053	3,609	3,816	4,575	5,578	6,989	7,584	8,148	10,341
12	2,214	3,074	3,571	4,178	4,404	5,226	6,304	7,807	8,438	9,034	11,340
13	2,617	3,565	4,107	4,765	5,009	5,892	7,042	8,634	9,299	9,926	12,340
14	3,041	4,075	4,660	5,368	5,629	6,571	7,790	9,467	10,165	10,821	13,339
15	3,483	4,601	5,229	5,985	6,262	7,261	8,547	10,307	11,036	11,721	14,339
16	3,942	5,142	5,812	6,614	6,908	7,962	9,312	11,152	11,912	12,624	15,338
17	4,416	5,697	6,408	7,255	7,564	8,672	10,085	12,002	12,892	13,531	16,338
18	4,905	6,265	7,015	7,906	8,231	9,390	10,865	12,857	13,675	14,440	17,338
19	5,407	6,844	7,633	8,567	8,907	10,117	11,651	13,716	14,562	15,352	18,338
20	5,921	7,434	8,260	9,237	9,591	10,871	12,443	14,578	15,452	16,266	19,337

v	Вероятность										
	0,999	0,995	0,99	0,98	0,975	0,95	0,90	0,80	0,75	0,70	0,50
21	6,447	8,034	8,897	9,915	10,283	11,591	13,240	15,445	16,344	17,182	20,337
22	6,983	8,643	9,542	10,600	10,982	12,338	14,041	16,314	17,240	18,101	21,337
23	7,529	9,260	10,196	11,293	11,688	13,091	14,848	17,187	18,137	19,021	22,337
24	8,035	9,886	10,856	11,992	12,401	13,848	15,659	18,062	19,037	19,943	23,337
25	8,649	10,520	11,524	12,697	13,120	14,611	16,173	18,940	19,939	20,887	24,337
26	9,222	11,160	12,198	13,409	13,844	15,379	17,292	19,820	20,843	21,792	25,336
27	9,803	11,808	12,879	14,125	14,573	16,151	18,114	20,703	21,749	22,719	26,136
28	10,391	12,461	13,565	14,847	15,308	16,928	18,937	21,588	22,657	23,617	27,336
29	10,986	13,121	14,256	15,574	16,047	17,708	19,768	22,475	23,567	24,577	28,336
30	11,588	13,787	14,953	16,306	16,791	18,493	20,599	23,364	24,478	25,508	29,336

v	Вероятность									
	0,30	0,25	0,20	0,10	0,05	0,025	0,02	0,01	0,005	0,001
1	1,074	1,323	1,642	2,706	3,841	5,024	5,412	6,635	7,879	10,827
2	2,408	2,773	3,219	4,605	5,991	7,378	7,824	9,210	10,597	13,815
3	3,665	4,108	4,642	6,251	7,815	9,348	9,837	11,345	12,838	16,268
4	4,878	5,385	5,989	7,779	9,488	11,143	11,668	13,277	14,860	18,465
5	6,064	6,626	7,289	9,236	11,070	12,839	13,388	15,086	16,750	20,517
6	7,231	7,841	8,558	10,645	12,592	14,449	15,033	16,812	18,548	22,457
7	8,383	9,037	9,803	12,017	14,067	16,013	16,622	18,475	20,278	24,322
8	9,524	10,219	11,030	13,362	15,507	17,535	18,168	20,090	21,955	26,125
9	10,656	11,389	12,242	14,684	16,919	19,023	19,679	21,666	23,589	27,877
10	11,781	12,549	13,412	15,987	18,307	20,483	21,161	23,209	25,188	29,588

v	Вероятность									
	0,30	0,25	0,20	0,10	0,05	0,025	0,02	0,01	0,005	0,001
11	12,899	13,701	14,631	17,275	19,675	21,920	22,618	24,725	26,757	31,264
12	14,011	14,845	15,812	18,549	21,026	23,337	24,054	26,217	28,300	32,909
13	15,119	15,984	16,985	19,812	22,362	24,736	25,472	27,688	29,819	34,528
14	16,222	17,117	18,151	21,064	23,685	26,119	26,873	29,141	31,319	36,123
15	17,322	18,245	19,311	22,307	24,996	27,488	28,259	30,578	32,801	37,697
16	18,418	19,369	20,465	23,542	26,296	28,845	29,633	32,000	34,267	39,252
17	19,511	20,489	21,615	24,769	27,587	30,191	30,995	33,409	35,718	40,790
18	20,601	21,605	22,760	25,989	28,869	31,526	32,346	34,805	37,156	42,312
19	21,689	22,718	23,900	27,204	30,144	32,852	33,687	36,191	38,582	43,820
20	22,775	23,828	25,038	28,412	31,410	34,170	35,020	37,566	39,997	45,315
21	23,858	24,935	26,171	29,615	32,671	35,479	36,343	38,932	41,401	46,797
22	24,939	26,039	27,301	30,813	33,924	36,781	37,659	40,289	42,796	48,268
23	26,018	27,141	28,429	32,007	35,172	38,076	38,968	41,638	44,181	49,728
24	27,096	28,241	29,553	33,196	36,415	39,364	40,270	42,980	45,558	51,170
25	28,172	29,339	30,675	34,382	37,652	40,046	41,566	44,314	46,928	52,620
26	29,246	30,434	31,795	35,563	38,885	41,923	42,856	45,642	48,290	54,052
27	30,319	31,528	32,912	36,741	40,113	43,194	44,140	46,963	49,645	55,476
28	31,391	32,620	34,027	37,916	41,337	44,461	45,419	48,278	50,993	56,893
29	32,461	33,711	35,139	39,087	42,557	45,722	46,693	49,588	52,336	58,302
30	33,530	34,800	36,250	40,256	43,773	46,979	47,962	50,892	53,672	59,703

Таблица 4. Распределение Фишера—Снедекора (F -распределение)

Значения $F_{\text{табл}}$, удовлетворяющие условию $P(F > F_{\text{табл}})$. Первое значение соответствует вероятности 0,05; второе — вероятности 0,01 и третье — вероятности 0,001; v_1 — число степеней свободы числителя; v_2 — знаменателя

$v_2 \backslash v_1$	1	2	3	4	5	6	8	12	24	∞	t
1	161,4 4052 406523	199,5 4999 500016	215,7 5403 536700	224,6 5625 562527	230,2 5764 576449	234,0 5859 585953	238,9 5981 598149	243,9 6106 610598	249,0 6234 623432	253,3 6366 636535	12,71 63,66 636,2
2	18,51 98,49 998,46	19,00 99,01 999,00	19,16 99,17 999,20	19,25 99,25 999,20	19,30 99,30 999,20	19,33 99,33 999,20	19,37 99,36 999,40	19,41 99,42 999,60	19,45 99,46 999,40	19,50 99,50 999,40	4,30 9,92 31,00
3	10,13 34,12 67,47	9,55 30,81 148,51	9,28 29,46 141,10	9,12 28,71 137,10	9,01 28,24 134,60	8,94 27,91 132,90	8,84 27,49 130,60	8,74 27,05 128,30	8,64 26,60 125,90	8,53 26,12 123,50	3,18 5,84 12,94
4	7,71 21,20 74,13	6,94 18,00 61,24	6,59 16,69 56,18	6,39 15,98 53,43	6,26 15,52 51,71	6,16 15,21 50,52	6,04 14,80 49,00	5,91 14,37 47,41	5,77 13,93 45,77	5,63 13,46 44,05	2,78 4,60 8,61
5	6,61 16,26 47,04	5,79 13,27 36,61	5,41 12,06 33,20	5,19 11,39 31,09	5,05 10,97 20,75	4,95 10,67 28,83	4,82 10,27 27,64	4,68 9,89 26,42	4,53 9,47 25,14	4,36 9,02 23,78	2,57 4,03 6,86
6	5,99 13,74 35,51	5,14 10,92 26,99	4,76 9,78 23,70	4,53 9,15 21,90	4,39 8,75 20,81	4,28 8,47 20,03	4,15 8,10 19,03	4,00 7,72 17,99	3,84 7,31 16,89	3,67 6,88 15,75	2,45 3,71 5,96

Продолжение табл. 4

$v_2 \backslash v_1$	1	2	3	4	5	6	8	12	24	∞	t
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23	2,36
	12,25	9,55	8,45	7,85	7,46	7,19	6,84	6,47	6,07	5,65	3,50
	29,22	21,69	18,77	17,19	16,21	15,52	14,63	13,71	12,73	11,70	5,40
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,99	2,31
	11,26	8,65	7,59	7,10	6,63	6,37	6,03	5,67	5,28	4,86	3,36
	25,42	18,49	15,83	14,39	13,49	12,86	12,04	11,19	10,30	9,35	5,04
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71	2,26
	10,56	8,02	6,99	6,42	6,06	5,80	5,47	5,11	4,73	4,31	3,25
	22,86	16,39	13,90	12,56	11,71	11,13	10,37	9,57	8,72	7,81	4,78
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54	2,23
	10,04	7,56	6,55	5,99	5,64	5,39	5,06	4,71	4,33	3,91	3,17
	21,04	14,91	12,55	11,28	10,48	9,92	9,20	8,45	7,64	6,77	4,59
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40	2,20
	9,65	7,20	6,22	5,67	5,32	5,07	4,74	4,40	4,02	3,60	3,11
	19,69	13,81	11,56	10,35	9,58	9,05	8,35	7,62	6,85	6,00	4,49
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30	2,18
	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,16	3,78	3,36	3,06
	18,64	12,98	10,81	9,63	8,89	8,38	7,71	7,00	6,25	5,42	4,32
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21	2,16
	9,07	6,70	5,74	5,20	4,86	4,62	4,30	3,96	3,59	3,16	3,01
	17,81	12,31	10,21	9,07	8,35	7,86	7,21	6,52	5,78	4,97	4,12

$v_1 \backslash v_2$	1	2	3	4	5	6	8	12	24	∞	t
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13	2,14
	8,86	6,51	5,56	5,03	4,69	4,46	4,14	3,80	3,43	3,00	2,98
	17,14	11,78	9,73	8,62	7,92	7,44	6,80	6,13	5,41	4,60	4,14
15	4,45	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07	2,13
	8,68	6,36	5,42	4,89	4,56	4,32	4,00	3,67	3,29	2,87	2,95
	16,59	11,34	9,34	8,25	7,57	7,09	6,47	5,81	5,10	4,31	4,07
16	4,41	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01	2,12
	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,55	3,18	2,75	2,92
	16,12	10,97	9,01	7,94	7,27	6,80	6,20	5,55	4,85	4,06	4,02
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96	2,11
	8,40	6,11	5,18	4,67	4,34	4,10	3,79	3,45	3,08	2,65	2,90
	15,72	10,66	8,73	7,68	7,02	6,56	5,96	5,32	4,63	3,85	3,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92	2,10
	8,28	6,01	5,09	4,58	4,25	4,01	3,71	3,37	3,01	2,57	2,88
	15,38	10,39	8,49	7,46	6,81	6,35	5,76	5,13	4,45	3,67	3,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88	2,09
	8,18	5,93	5,01	4,50	4,17	3,94	3,63	3,30	2,92	2,49	2,86
	15,08	10,16	8,28	7,26	6,61	6,18	5,59	4,97	4,29	3,52	3,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84	2,09
	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,23	2,86	2,42	2,84
	14,82	9,95	8,10	7,10	6,46	6,02	5,44	4,82	4,15	3,38	3,85

Продолжение табл. 4

$v_2 \backslash v_1$	1	2	3	4	5	6	8	12	24	∞	t
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,82	2,08
	8,02	5,78	4,87	4,37	4,04	3,81	3,51	3,17	2,80	2,36	2,83
	14,62	9,77	7,94	6,95	6,32	5,88	5,31	4,70	4,03	3,26	3,82
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78	2,07
	7,94	5,72	4,82	4,31	3,99	3,75	3,45	3,12	2,75	2,30	2,82
	14,38	9,61	7,80	6,81	6,19	5,76	5,19	4,58	3,92	3,15	3,79
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76	2,07
	7,88	5,66	4,76	4,26	3,94	3,71	3,41	3,07	2,70	2,26	2,81
	14,19	9,46	7,67	6,70	6,08	5,56	5,09	4,48	3,82	3,05	3,77
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73	2,06
	7,82	5,61	4,72	4,22	3,90	3,67	3,36	3,03	2,66	2,21	2,80
	14,03	9,34	7,55	6,59	5,98	5,55	4,99	4,39	3,74	2,97	3,75
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71	2,06
	7,77	5,57	4,68	4,18	3,86	3,63	3,32	2,99	2,62	2,17	2,79
	13,88	9,22	7,45	6,49	5,89	5,46	4,91	4,31	3,66	2,89	3,72
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69	2,06
	7,72	5,53	4,64	4,14	3,82	3,59	3,29	2,96	2,58	2,13	2,78
	13,74	9,12	7,36	6,41	5,80	5,38	4,83	4,24	3,59	2,82	3,71
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67	2,05
	7,68	5,49	4,60	4,11	3,78	3,56	3,26	2,93	2,55	2,10	2,77
	13,61	9,02	7,27	6,33	5,73	5,31	4,76	4,17	3,52	2,76	3,69

Окончание табл. 4

$v_2 \backslash v_1$	1	2	3	4	5	6	8	12	24	∞	t
28	4,19	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65	2,05
	7,64	5,54	4,57	4,07	3,75	3,53	3,23	2,90	2,52	2,06	2,76
	13,50	8,93	7,18	6,25	5,66	5,24	4,69	4,11	3,46	2,70	3,67
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64	2,05
	7,60	5,42	4,54	4,04	3,73	3,50	3,20	2,87	2,49	2,03	2,76
	13,39	8,85	7,12	6,19	5,59	5,18	4,65	4,05	3,41	2,64	3,66
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62	2,04
	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,84	2,47	2,01	2,75
	13,29	8,77	7,05	6,12	5,53	5,12	4,58	4,00	3,36	2,59	3,64
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39	2,00
	7,08	4,98	4,13	3,65	3,34	3,12	2,82	2,50	2,12	1,60	2,66
	11,97	7,76	6,17	5,31	4,76	4,37	3,87	3,31	2,76	1,90	3,36
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,03	1,96
	6,64	4,60	3,78	3,32	3,02	2,80	2,51	2,18	1,79	1,04	2,58
	10,83	6,91	5,42	4,62	4,10	3,74	3,27	2,74	2,13	1,05	3,29

Таблица 5. Значения d_1 и d_u критерия Дарбина—Уотсона на уровне значимости $\alpha = 0,05$ (n — число наблюдений, k^1 — число объясняющих переменных)

n	$k^1 = 1$		$k^1 = 2$		$k^1 = 3$	
	d_1	d_u	d_1	d_u	d_1	d_u
15	1,08	1,36	0,95	1,54	0,82	1,75
16	1,10	1,37	0,98	1,54	0,86	1,73
17	1,13	1,38	1,02	1,54	0,90	1,71
18	1,16	1,39	1,05	1,53	0,93	1,69
19	1,18	1,40	1,08	1,53	0,97	1,68
20	1,20	1,41	1,10	1,54	1,00	1,68
21	1,22	1,42	1,13	1,54	1,03	1,67
22	1,24	1,43	1,15	1,54	1,05	1,66
23	1,26	1,44	1,17	1,54	1,08	1,66
24	1,27	1,45	1,19	1,55	1,10	1,66
25	1,29	1,45	1,21	1,55	1,12	1,66
26	1,30	1,46	1,22	1,55	1,14	1,65
27	1,32	1,47	1,24	1,56	1,16	1,65
28	1,33	1,48	1,26	1,56	1,18	1,65
29	1,34	1,48	1,27	1,56	1,20	1,65
30	1,35	1,49	1,28	1,57	1,21	1,65
31	1,36	1,50	1,30	1,57	1,23	1,65
32	1,37	1,50	1,31	1,57	1,34	1,65
33	1,38	1,51	1,32	1,58	1,26	1,65
34	1,39	1,51	1,33	1,58	1,27	1,65
35	1,40	1,52	1,34	1,58	1,28	1,65
36	1,41	1,52	1,35	1,59	1,29	1,65
37	1,42	1,53	1,36	1,59	1,31	1,66
38	1,43	1,54	1,37	1,59	1,32	1,66
39	1,43	1,54	1,38	1,60	1,33	1,66
40	1,44	1,54	1,39	1,60	1,34	1,66
45	1,48	1,57	1,43	1,62	1,38	1,67
50	1,50	1,59	1,46	1,63	1,42	1,67
55	1,53	1,60	1,49	1,64	1,45	1,68
60	1,55	1,62	1,51	1,65	1,58	1,69
65	1,57	1,63	1,54	1,66	1,50	1,70
70	1,58	1,64	1,55	1,67	1,52	1,70
75	1,60	1,65	1,57	1,68	1,54	1,71
80	1,61	1,66	1,59	1,69	1,56	1,72
85	1,62	1,67	1,60	1,70	1,57	1,72
90	1,63	1,68	1,61	1,70	1,59	1,73
95	1,64	1,69	1,62	1,71	1,60	1,73
100	1,65	1,69	1,63	1,72	1,61	1,74



ПРИЛОЖЕНИЕ 2. Исходные данные для самостоятельной работы

Таблица 1. Расстояния, пройденные британскими авиалайнерами за месяц [1]

Месяц	Расстояние, тыс. миль							
	1963	1964	1965	1966	1967	1968	1969	1970
Январь	6827	7269	8350	8186	8334	8639	9491	10840
Февраль	6178	6775	7829	7444	7899	8772	8919	10436
Март	7084	7819	8829	8484	9994	10894	11607	13589
Апрель	8162	8371	9948	9864	10078	10455	8852	13402
Май	8462	9069	10638	10252	10801	11179	12537	13103
Июнь	9644	10248	11253	12282	12950	10588	14759	14933
Июль	10466	11030	11424	11637	12222	10794	13667	14147
Август	10748	10882	11391	11577	12246	12770	13731	14057
Сентябрь	9963	10333	10665	12417	13281	13812	15110	16234
Октябрь	8194	9109	9396	9637	10366	10857	12185	12389
Ноябрь	6848	7685	7775	8094	8730	9290	10645	11595
Декабрь	7027	7602	7933	9280	9614	10925	12161	12772

Таблица 2. Перевозки грузов железнодорожным транспортом¹

Год	Вес груза, млн т											
	Январь	Февраль	Март	Апрель	Май	Июнь	Июль	Август	Сентябрь	Октябрь	Ноябрь	Декабрь
1993	114,0	107,9	122,6	121,5	119,6	115,1	114,4	111,2	108,1	110,8	100,0	100,4
1994	91,5	83,1	92,8	91,7	88,7	86,8	84,8	87,9	85,3	89,6	85,6	86,3
1995	84,2	79,7	89,1	85,8	87,6	83,8	88,7	89,0	85,8	88,1	82,4	80,1
1996	75,8	71,0	79,1	78,9	77,2	73,9	74,4	76,4	76,2	79,8	74,4	72,4
1997	69,1	67,8	76,5	73,7	72,6	71,5	74,4	76,5	75,9	79,5	75,1	74,2
1998	69,3	64,3	73,2	71,7	70,2	67,4	68,5	69,5	66,2	73,4	68,1	72,4

¹ Айвазян С.А., Мхитарян В.С. Прикладная статистика в задачах и упражнениях. — М.: ЮНИТИ-ДАНА, 2001.

Учебное пособие

Дуброва Татьяна Абрамовна

СТАТИСТИЧЕСКИЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ

Редактор Т.М. Дубович

Корректор Т.А. Балашова

Оригинал-макет М.А. Бакаян

Оформление художника В.А. Лебедева

Лицензия серия ИД № 03562 от 19.12.2000 г.
Подписано в печать 25.12.2002. Формат 60х88 1/16

Усл. печ. л. 13,0. Уч.-изд. л. 9,5

Тираж 20 000 экз. (1-й завод – 5 000). Заказ 45

ООО «ИЗДАТЕЛЬСТВО ЮНИТИ-ДАНА»

Генеральный директор В.Н. Закаидзе

123298, Москва, ул. Ирины Левченко, 1
Тел. (095) 194-00-15. Тел/факс (095) 194-00-14
www.unity-dana.ru E-mail: unity@unity-dana.ru

Отпечатано во ФГУП ИПК «Ульяновский Дом печати»
432980, г. Ульяновск, ул. Гончарова, 14