



Лукашин Юрий Павлович
доктор экономических наук, профессор,
заведующий сектором экономического моделирования
Института мировой экономики и международных
отношений Российской академии наук (ИМЭМО РАН),
заместитель директора Института статистики
и эконометрики Московского государственного
университета экономики, статистики и информатики (МЭСИ),
заведующий кафедрой высших финансовых
вычислений МЭСИ,
заведующий кафедрой естественно-научных и математических
дисциплин Московской международной высшей школы
бизнеса "МИРБИС" (Институт).
Выполнял исследовательские работы по заказу
Центрального банка РФ, Московской межбанковской
валютной биржи, Внешэкономбанка,
Центральной расчетной палаты.
Имеет около 90 опубликованных работ,
в том числе изданных за рубежом.
Основные монографии:
"Адаптивные методы краткосрочного прогнозирования" (1979);
"Линейная регрессия с переменными параметрами" (1992);
"Проверка гипотез в эконометрике" (2002).

E-mail: loukashin@rambler.ru

ISBN 5-279-02740-5



9 785279 027408

Ю.П. ЛУКАШИН



Ю.П. ЛУКАШИН

АДАПТИВНЫЕ МЕТОДЫ КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ



Ю.П. ЛУКАШИН

АДАПТИВНЫЕ МЕТОДЫ КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

1881
Рекомендовано
Учебно-методическим объединением
по образованию в области статистики
в качестве учебного пособия для студентов
высших учебных заведений, обучающихся
по специальности 161700 "Статистика"
и другим экономическим специальностям



МОСКВА
"ФИНАНСЫ И СТАТИСТИКА"
2003

РЕЦЕНЗЕНТЫ:

кафедра статистики и эконометрики
Московского государственного университета
экономики, статистики и информатики — МЭСИ
(заведующий кафедрой доктор экономических наук,
профессор В.С. Мхитарян);

Е.М. Четыркин,
доктор экономических наук, профессор

Лукашин Ю.П.

Л84 Адаптивные методы краткосрочного прогнозирования
временных рядов: Учеб. пособие. — М.: Финансы и статисти-
ка, 2003. — 416 с.: ил.

ISBN 5-279-02740-5

Посвящено построению статистических моделей с переменными параметрами для прогнозирования нестационарных временных рядов. Рассмотрены адаптивные модели полиномиальных и стохастических трендов, сезонных и циклических колебаний, гистограмм, модели семейства ARIMA, ARCH. Приводятся примеры прогнозирования курсов акций, валют, цен на золото. Материалы пособия апробированы на занятиях в МЭСИ, МИРБИС и других вузах.

Для студентов, аспирантов, преподавателей экономических вузов, менеджеров и финансовых аналитиков.

Л 1602090000 — 149
010(01) — 2003 276 — 2003

УДК 338.27.015(075.8)
ББК 65.23я73

ISBN 5-279-02740-5

© Ю.П. Лукашин, 2003

ПРЕДИСЛОВИЕ

Данное учебное пособие посвящено одному из современных направлений статистического анализа и прогнозирования временных рядов. Важность этого направления не вызывает сомнения, так как необходимость решения соответствующих задач с помощью адаптивных методов возникает сравнительно часто. Адаптивные методы могут применяться для прогнозирования показателей фондового рынка, денежных потоков, изменений ежедневных остатков на складах, в инструментальных кладовых, магазинах. С помощью этих же методов удастся описать эволюцию изменения технико-экономических характеристик изделий и переменных параметров химических процессов, изучить поведение показателя частоты отказов оборудования в зависимости от его возраста. Наконец, названные методы полезны при анализе сезонных явлений. В ряде случаев эти методы могут с успехом применяться для прогнозирования макропоказателей. Методы адаптивного прогнозирования применяются там, где основной информацией для прогноза являются временные ряды.

Инструментом прогноза при адаптивном методе служит модель. Первоначальная оценка параметров этой модели основывается на данных базового (исходного) временного ряда. На основе новых данных, получаемых на каждом следующем шаге, происходит корректировка параметров модели во времени, их адаптация к новым, непрерывно изменяющимся условиям развития явления. Таким образом, модель постоянно «впитывает» новую информацию и приспосабливается к ней.

Адаптивные модели изолированных рядов при всей их простоте могут давать более надежные результаты, чем сложные эконометрические системы уравнений. Так, при существенной перестройке некоторой экономической структуры (например, под влиянием научно-технического прогресса, изменений социально-политических условий и т.п.) эконометрическая модель с постоянными параметрами будет экстраполировать существенно устаревшие зависимости. Адаптивная модель в таких же условиях перманентно приспосабливается и учитывает эти изменения.

Здесь уместно сослаться на эксперимент Ч. Нельсона (1972 г.), в котором сравнивалась точность прогнозов, полученных на основе эконометрической модели, состоящей из нескольких уравнений, и достаточно простых адаптивных моделей, применявшихся для прогнозирования нескольких временных рядов. Оказалось, что для периода, использованного при оценивании параметров эконометрической модели, последняя показала лучшее приближение к данным наблюдения, чем адаптивные модели. Однако за пределом периода наблюдения с помощью эконометрической модели (с экзогенными реальными данными) полученные результаты хуже, чем с помощью адаптивных моделей.

Рассматриваемые в данном пособии методы обладают следующими свойствами:

- они применимы для широкого круга задач;
- адаптивное прогнозирование не требует большого объема информации, оно базируется на интенсивном анализе информации, содержащейся в отдельных временных рядах;
- модель, описывающая структуру показателя и его динамику, как правило, отличается ясностью и простотой математической формулировки;
- неоднородность временных рядов и их связей находит отражение в адаптивной эволюции параметров или даже структуры моделей.

Вышеизложенное, однако, не означает, что адаптивные методы и модели могут заменить любые другие виды моделей. Эти методы пригодны лишь для обработки рядов с умеренными изменениями во времени. Они оказываются весьма грубыми при прогнозировании на большое число шагов вперед. Тем не менее следует отметить, что в последние годы они получили значительное развитие, в том числе и благодаря усилиям автора книги.

Начало развития методов адаптивного прогнозирования, по видимому, положено работой Ч. Хольта, появившейся в 1957 г. В ней рассматривалась проблема экспоненциального сглаживания. Наиболее полно этот вопрос был изучен Р.Г. Брауном и автором данного учебного пособия. В настоящее время имеется обширная литература, в которой предлагаются различные методы и модели, относящиеся к адаптивному прогнозированию.

Цель настоящего издания — обобщить материал, представленный в многочисленных монографиях и статьях, как отечественных, так и зарубежных. При этом автор выявляет их сильные и слабые стороны, сопоставляет их свойства, базируясь на экспериментальных расчетах.

Безусловный интерес вызывают предложенные автором некоторые новые подходы: адаптивные измерители корреляции (гл. 9); комбинирование одновременно нескольких моделей с различными структурами и переключение с одной на другую в зависимости от текущих свойств ряда (гл. 5); критерии и методы оптимизации адаптивной гистограммы (гл. 11); применение принципа адаптивной фильтрации для построения адаптивных авторегрессионных и многомерных моделей (гл. 3, 8 и 14).

Рассмотрены современные представления о детерминированном и стохастическом характере тренда и методы статистической идентификации типа тренда (критерии Дикки—Фуллера (гл. 12)); понятия интегрированности и коинтегрированности переменных (гл. 13), используемые при спецификации регрессионных уравнений модели. Большое внимание уделено моделям авторегрессии-скользящего среднего, разработанным Дж. Боксом и Г. Дженкинсом (гл. 7), современным методам их оценивания (гл. 12), их использованию для моделирования локальных колебаний дисперсии остаточного члена уравнения или ошибок прогнозов (модели семейства ARCH, GARCH и др. — гл. 14).

Положительным моментом является введение в учебное пособие примеров практического применения рассматриваемых подходов для прогнозирования реальных экономических показателей: курсов акций, цены на золото, курсов валют (гл. 1, 2, 5, 7 и 16), кассовых остатков в банковском учреждении (гл. 11), некоторых макроэкономических показателей (гл. 8). Этому способствовало сочетание автором научной и практической работы и, в частности, сотрудничество с Центральным банком Российской Федерации, Внешэкономбанком, Центральной расчетной палатой, Московской межбанковской валютной биржей.

В учебном пособии нашли отражение научные разработки автора разных лет (последнее десятилетие представлено в гл. 9–16), апробированные на международных и российских конференциях, в процессе преподавания таких дисциплин, как «Эконометрика», «Статистические методы прогнозирования», «Прогнозирование социальных и экономических процессов», «Финансовая математика», в МЭСИ, МИРБИС и других вузах Москвы.

Это учебное пособие поможет студенту освоить современные методы статистического прогнозирования, аспиранту и практику — найти наиболее эффективный метод прогнозирования, теоретику и разработчику моделей позволит сократить период ознакомления с достижениями данного направления.

Профессор Е.М. Четыркин

ВВЕДЕНИЕ

Экономическое прогнозирование характеризует будущее развитие, исходя из гипотезы, что основные факторы и тенденции прошлого периода сохранятся на период прогноза или что можно обосновать и учесть направление их изменений в рассматриваемой перспективе. Надежды здесь возлагаются на инерционность экономических систем. Между тем в большинстве случаев подвижность экономических явлений возрастает.

Наибольшей инерционностью обладают макроэкономические характеристики. Для параметров, описывающих процессы, происходящие на уровне отраслей, предприятий, цехов, характерна большая зависимость от местных условий. В современных условиях исследователь часто имеет дело с новыми экономическими явлениями с короткими статистическими рядами или со старыми явлениями, претерпевающими коренные изменения, поэтому при использовании информации для построения моделей встает вопрос о приемственности данных. Устаревшие данные при моделировании часто оказываются бесполезными и даже вредными. К тому же статистическое описание процесса редко может удовлетворить, потому что необходимо знать, не как развивается процесс в среднем, а как будет развиваться его тенденция, существующая в данный момент. Значит, надо строить модели, опираясь в основном на малое количество самых свежих данных. В этом случае альтернативой статистическому обоснованию модели может быть наделение ее адаптивными свойствами.

Направленное развитие экономической системы прокладывает себе путь через единичные явления в условиях столкновения противоборствующих сил и поэтому обычно предполагается стохастическим процессом. Для прогнозирования стационарных процессов¹ математический аппа-

рат разработан довольно хорошо. Но экономические процессы, как правило, являются нестационарными. Чем больше период прогноза (упреждения), тем больше возможностей для изменения тенденций экономического развития, особенно в современных условиях. Кроме того, в исследуемом периоде могут произойти вообще непредвиденные, непредсказуемые события, существенно деформирующие изучаемый процесс. К непредсказуемым событиям относятся такие, для предвидения которых в момент составления прогноза нет необходимых данных, или такие, природа которых неопределенна.

В. Н. Тутубалин в работе [32] совершенно справедливо отмечает широко распространенное заблуждение, согласно которому если некоторая величина не является детерминированной (т. е. предопределенной), то она может рассматриваться как случайная. Это совершенно неверно, потому что случайность требует статистической устойчивости. Таким образом, кроме детерминированных и случайных величин бывают еще неопределенные. Между тем математическая модель может содержать либо детерминированные, либо случайные величины, либо те и другие вместе, но при нынешнем состоянии науки не может содержать неопределенных. Следовательно, искусство построения математической модели состоит в том, чтобы неопределенные величины, с которыми приходится иметь дело на практике, приближенно представить через детерминированные и случайные. Это значительно затрудняет экономическое прогнозирование.

В ряде случаев экономико-математические модели прогнозирования могут быть полезным инструментом исследования. При этом, конечно, для увеличения точности прогнозов экономического развития в изменяющихся условиях, в условиях неопределенности или неполной информации необходима работа по совершенствованию моделей. Важную роль в этом должны сыграть адаптивные методы прогнозирования. Отличие адаптивных моделей от других прогнозистических моделей состоит в том, что они отражают текущие свойства ряда и способны непрерывно учитывать эволюцию динамических характеристик изучаемых процессов. Цель адаптивных методов заключается в построении самокорректирующихся (самонастраивающихся) экономико-математических моделей, которые способны отражать изменяющиеся во времени условия, учитывать информацион-

¹ Процессы с неизменными основными характеристиками, такими, как математическое ожидание, дисперсия, автокорреляционная функция.

ную ценность различных членов временной последовательности и давать достаточно точные оценки будущих членов данного ряда. Именно поэтому такие модели предназначаются прежде всего для краткосрочного прогнозирования.

Отметим различие понятия краткосрочного прогноза в экономике и статистике. В экономике под краткосрочным прогнозом обычно понимают прогноз с периодом упреждения до одного года. В статистике информацию о процессе обычно получают в виде записей значений, наблюдаемых через равные промежутки времени. Соответственно под краткосрочным прогнозом, как правило, подразумевается прогноз на один интервал времени (в крайнем случае на несколько). Сам же интервал может быть любым. Статистика интересуется не величиной временного интервала, а объемом выборки, ее однородность и т. п. Именно в этом статистическом смысле нужно понимать название предлагаемой книги.

Основным средством анализа и прогноза временного ряда будет модель. Понятие модель используется в двух значениях: как модель временного ряда, выражающая закон генерирования членов ряда, и как прогнозная модель, или предиктор. Главное отличие этих двух типов моделей в том, что на выходе модели временного ряда фактические члены ряда, а на выходе прогнозной модели — оценки будущих членов ряда. Теоретически свойства предиктора исследуются в предположении, что он применен для получения прогнозов некоего процесса, генерируемого моделью, заданной аналитически.

На временной ряд воздействуют в разное время различные факторы. Одни из них по тем или иным причинам ослабляют свое влияние, другие воздействуют активнее. Таким образом, реальный процесс протекает в изменяющихся условиях, составляющих его внешнюю среду, к которой он приспосабливается, адаптируется. А модель, в свою очередь, адаптируется к ряду, представляющему этот процесс. Поскольку мы рассматриваем варьирующие, нестационарные ряды, т. е. ряды, у которых уровень, скорость линейного роста и прочие характеристики не остаются постоянными во времени, модель будет всегда находиться в движении. Образно говоря, процесс адаптации модели к ряду можно было бы назвать «гонкой за лидером».

По-видимому, трудно провести четкую грань, отделяющую адаптивные методы прогнозирования от неадаптивных. Уже прогнозирование методом экстраполяции обычных регрессионных кривых содержит некоторый элемент адаптации, когда с каждым новым получением фактических данных параметры регрессионных кривых пересчитываются, уточняются. Через достаточно большой промежуток времени может быть заменен даже тип кривой. Однако здесь степень адаптации весьма незначительна; к тому же с течением времени она падает вместе с увеличением общего количества наблюдаемых точек и соответственно с уменьшением в выборке удельного веса каждой новой точки.

Модификацией этого метода является метод кусочно-линейной аппроксимации, использование которого ведет к уменьшению «памяти» модели, к «забыванию» старых данных и построению линий регрессии на искусственно ограниченном количестве информации. Этот метод лучше учитывает новые тенденции, быстрее приспосабливается к изменившимся характеристикам процесса, но зато сильнее реагирует на помехи, случайные отклонения и искажения в связи с уменьшением доли «наследственности». Соотношение между «изменчивостью» и «наследственностью» в кусочно-линейном варианте регрессионного анализа определяется субъективным выбором интервалов аппроксимации. Недостатком является также то, что ценность информации в пределах интервала аппроксимации считается одинаковой независимо от возраста, а вне его пределов скачком падает до нуля.

Ценность информации в зависимости от возраста можно учесть с помощью геометрически убывающих весовых коэффициентов (взвешенная регрессия).

Введем веса для квадратов ошибок. Совокупность весовых коэффициентов представляет собой функцию ценности информации от времени. Тогда параметры регрессионных кривых будут отыскиваться из условия

$$\min_i \sum \beta' [x_{T-i} - y_{T-i}(a_1, a_2, \dots, a_n)]^2, \quad 0 < \beta < 1,$$

где

T — текущий момент времени;

x_{T-i} — фактическое значение процесса в момент $T - i$;

$y_{T-i}(a_1, \dots, a_n)$ — значение подбираемой функции в момент $T - i$.

Веса ошибок для более ранних моментов времени уменьшаются по закону убывающей геометрической прогрессии. Минимум этого выражения достигается при равенстве нулю частных производных функционала по a_j .

В данном случае функция ценности информации будет иметь экспоненциальную форму в отличие от прямоугольной в случае простого или кусочно-линейного метода построения регрессионных кривых. Такое взвешивание ошибок обеспечивает лучший подгон регрессионной кривой к более свежим данным.

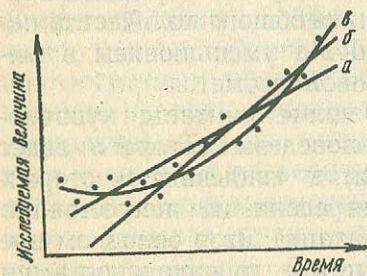


Рис. 1. Сравнение моделей:

- а) обычная линейная регрессия;
- б) экспоненциально-взвешенная линейная регрессия;
- в) обычная параболическая регрессия

Выбор величины β зависит прежде всего от характера моделируемого процесса, от его динамических свойств и статистических характеристик. Во многом выбор β зависит от опыта исследователя. Наилучшее ее значение можно установить экспериментально методом проб.

Положительным свойством такой модели является ее способность лучше приспосабливаться к динамике процесса. Следует подчеркнуть, что если при выборе типа кривой регрессии была допущена ошибка, то в данной модели она обычно частично компенсируется. Это иллюстрирует рис. 1.

Вместо параболы ошибочно была взята прямая. В случае обычной регрессионной модели процесс характеризовался бы прямой (а), в случае экспоненциально-взвешенной регрессии — прямой (б). Как видим, прогноз по прямой (б) будет точнее. Однако такой подход обладает рядом недостатков, главными из которых являются субъективность выбора структуры модели, ее чрезмерная жесткость и в силу этого автокоррелированность остатков. Громоздкий пересчет параметров регрессионных кривых с получением каждой новой точки требует повторного вовлечения в вычислительную процедуру всего объема информации. Адаптивное моделирование позволяет в известной мере избавиться от этих недостатков.

У истоков адаптивного направления лежит простейшая модель экспоненциального сглаживания. Модификации и обобщения этой модели привели к появлению целого семейства адаптивных моделей с различными свойствами.

Остановимся на основных особенностях предлагаемых вниманию читателя методов адаптивного моделирования. Прежде всего отметим, что все рассматриваемые модели имеют жесткий алгоритм поведения. Однако, как известно, система может быть механистической по своей природе и обнаруживать тем не менее адаптивные свойства. Адаптация в данных моделях складывается из небольших дискретных сдвигов. В основе процедуры адаптации лежит метод проб и ошибок, который совершенно справедливо считается универсальным путем выработки нового поведения. Такой процесс необходим, ибо только он позволяет извлечь нужную информацию, без которой успешная адаптация невозможна.

Последовательность процесса адаптации в основном выглядит следующим образом. Пусть модель находится в некотором исходном состоянии (т. е. определены текущие значения ее коэффициентов) и по ней делается прогноз. Выжидаем, пока истечет одна единица времени (шаг моделирования), и анализируем, насколько далек результат, полученный по модели, от фактического значения ряда. Ошибка прогнозирования через обратную связь поступает на вход системы и используется моделью в соответствии с ее логикой для перехода из одного состояния в другое с целью большего согласования своего поведения с динамикой ряда. На изменения ряда модель должна отвечать «компенсирующими» изменениями. Затем делается прогноз на следующий момент времени, и весь процесс повторяется.

Таким образом, адаптация осуществляется итеративно с получением каждой новой фактической точки ряда. Однако каковы должны быть правила перехода системы от одного состояния к другому? Другими словами, какова должна быть «логика механизма» адаптации? В сущности, этот вопрос решается каждым исследователем интуитивно. «Логика механизма» адаптации задается априорно, а затем проверяется эмпирически. При построении модели мы неизбежно наделяем ее «врожденными» свойствами и вместе с тем для большей гибкости должны позаботиться о механизмах «условных рефлексов», усваиваемых или утрачиваемых с определенной инерционностью. Их совокупность и составляет «логику механизма» адаптации.

Быстроту реакции модели на изменения в динамике процесса характеризует так называемый параметр адаптации. Процесс «обучения» модели состоит в выборе наилучшего параметра адаптации на основе проб на ретроспективном материале. Например, воздействие мощного «раздражителя» в виде процесса «белого шума»¹ с большой дисперсией вызывает у модели адекватную «оборонительную» реакцию. Модель не адаптируется к «белому шуму», а, наоборот, инертна по отношению к нему, отфильтровывает его. При наличии тенденции в стохастическом процессе наилучшей реакцией модели является определенный компромисс между двумя крайними ситуациями, обеспечивающий отражение тенденции и одновременно фильтрацию случайных отклонений от нее. По тому, насколько хорошо модель поддается «обучению», можно судить о ее способности адекватно отражать закономерности данного временного ряда. После выбора параметра адаптации самообучение модели происходит в процессе переработки новых статистических данных.

В силу простоты каждой отдельно взятой модели и ограниченности исходной (входной) информации, зачастую представленной единственным рядом, нельзя ожидать, что какая-либо одна адаптивная модель годится для прогнозирования любого ряда, любых вариаций поведения. Адаптивные модели достаточно гибки, однако на их универсальность рассчитывать не приходится. Поэтому при построении и объяснении конкретных моделей необходимо учитывать наиболее вероятные закономерности развития реального процесса, динамические свойства ряда соотносить с возможностями модели. Исследователь должен закладывать в модель те адаптивные свойства, которых, по его мнению, хватит для слежения модели за реальным процессом с заданной точностью. Вместе с тем нельзя надеяться на успешную самоадаптацию модели, более общей по отношению к той, которая необходима для отражения данного процесса, ибо увеличение числа параметров придает системе излишнюю чувствительность, приводит к ее «раскачке» и ухудшению получаемых по ней прогнозов.

Таким образом, при построении адаптивной модели приходится выбирать между общей и частной моделью и,

¹ «Белым шумом» называется последовательность независимых случайных величин ... $\varepsilon_{t-2}, \varepsilon_{t-1}, \varepsilon_t$..., имеющих фиксированное распределение, которое обычно предполагается нормальным, с нулевым математическим ожиданием и дисперсией σ_ε^2 .

взвешивая их достоинства и недостатки, отдавать предпочтение той, от которой можно ожидать наименьшей ошибки прогнозирования. Только при этом условии можно надеяться, что последовательность проб и ошибок постепенно приведет к наиболее эффективному предиктору. Поэтому исследователь должен иметь определенный запас специализированных моделей, разнообразных по структуре и функциональным свойствам.

Для сравнения возможных альтернатив необходим критерий полезности модели. Несмотря на то что в общем случае такой критерий является предметом спора, в случае краткосрочного прогнозирования признанным критерием обычно является средний квадрат ошибки прогнозирования. О качестве модели судят также по наличию автокорреляции в ошибках. В более развитых системах процесс проб и ошибок осуществляется в результате анализа как последовательных во времени, так и параллельных (конкурирующих) модификаций модели. Здесь используется принцип конкуренции или автоматического отбора (селекции) предиктора по заданному критерию.

Предлагаемая книга начинается с изучения простейших адаптивных моделей. Показаны пути усовершенствования их структуры и повышения адаптивных свойств, позволяющие более гибко учитывать особенности динамического ряда. Проведено эмпирическое сопоставление адаптивных методов. В заключение рассмотрены принципы построения адаптивных моделей многомерных временных рядов.

Глава 1

ПРОСТЕЙШИЕ АДАПТИВНЫЕ МОДЕЛИ И ИХ СВОЙСТВА

В главе вводятся некоторые общие понятия и изучаются свойства простейшей адаптивной модели, в основе которой лежит экспоненциальная средняя.

§ 1. ВРЕМЕННЫЕ РЯДЫ И СТОХАСТИЧЕСКИЕ ПРОЦЕССЫ

Временной ряд — это множество наблюдений, получаемых последовательно во времени. Если время изменяется дискретно, временной ряд называется дискретным. Мы будем рассматривать только дискретные временные ряды, в которых наблюдения делаются через фиксированный интервал времени, принимаемый за единицу счета. Переход от момента одного наблюдения к моменту следующего наблюдения будем называть шагом.

Если значения членов временного ряда точно определены какой-либо математической функцией, то временной ряд называется детерминированным. Если эти значения могут быть описаны только с помощью распределения вероятностей, временной ряд называется случайным.

Явление, развивающееся во времени согласно законам теории вероятностей, называется стохастическим процессом. В дальнейшем будем называть его просто процессом. Анализируемый отрезок временного ряда может рассматриваться как одна частная реализация (выборка) изучаемого стохастического процесса, генерируемого скрытым вероятностным механизмом.

Среди стохастических процессов выделяют класс процессов, называемых стационарными. Обозначим член временного ряда, наблюдаемый в момент t , через x_t . Стохастический процесс называется стационарным, если его свойства не изменяются во времени. В частности, он имеет постоянное математическое ожидание $\bar{x} = M(x_t)$ (т. е. среднее значение, относительно которого он варьирует), постоянную дисперсию $D(x) = M[(x_t - \bar{x})^2] = \sigma_x^2$, определяющую размах его колебаний относительно среднего значения, а также постоянную автоковариацию и коэффициенты автокорреляции¹. Ковариация между значениями x_t и x_{t+k} , отделенными интервалом в k единиц времени, называется автоковариацией с лагом (задержкой) k и определяется как

$$R_{xx}(k) = \text{cov}(x_t, x_{t+k}) = M[(x_t - \bar{x})(x_{t+k} - \bar{x})].$$

Для стационарных процессов автоковариация зависит только от лага k и $R_{xx}(0) = \sigma_x^2$. Автокорреляция с лагом k является лишь нормированной автоковариацией и равна:

$$\rho_k = \frac{M[(x_t - \bar{x})(x_{t+k} - \bar{x})]}{\sqrt{M[(x_t - \bar{x})^2] M[(x_{t+k} - \bar{x})^2]}} = \frac{M[(x_t - \bar{x})(x_{t+k} - \bar{x})]}{\sigma_x^2},$$

так как для стационарного процесса $\sigma_x^2 = \text{const}$. Таким образом, k -й коэффициент автокорреляции $\rho_k = \frac{R_{xx}(k)}{R_{xx}(0)}$. Он обладает тем свойством, что $-1 \leq \rho_k \leq 1$.

Для описания временных рядов используются математические модели. Представим, что временной ряд x_t , генерируемый некоторой моделью, можно представить в виде двух компонент

$$x_t = \xi_t + \varepsilon_t,$$

где величина ε_t генерируется случайным неавтокоррелированным процессом с нулевым математическим ожиданием и конечной (не обязательно постоянной) дисперсией, а величина ξ_t может быть генерирована либо детерминированной функцией, либо случайным процессом, либо какой-нибудь их комбинацией. Величины ε_t и ξ_t различаются характером воздействия на значения последующих членов ряда. Переменная ε_t влияет только на значение синхронного ей члена ряда, в то время как величина ξ_t в известной степени определяет значение нескольких или всех последующих членов ряда. Через величину ξ_t осуществляется

¹ Такие процессы называют стационарными процессами второго порядка, но, так как другие классы стационарных процессов рассматриваться не будут, мы называем их просто стационарными.

взаимодействие членов ряда; таким образом, в ней содержится информация, необходимая для получения прогнозов.

Назовем величину ξ_t уровнем ряда в момент t , а закон эволюции уровня во времени — трендом. Таким образом, тренд может быть выражен как детерминированной, так и случайной функциями, либо их комбинацией. Стохастические тренды имеют, например, ряды со случайным уровнем или случайным скачкообразным характером роста.

Приведем пример детерминированного тренда:

$$\xi_t = a_1 + a_2 t + a_3 t^2,$$

где a_1, a_2, a_3 — постоянные коэффициенты;
 t — время.

Пример случайного тренда:

$$\xi_t = \xi_{t-1} + u_t = \xi_0 + \sum_{i=1}^t u_i,$$

где ξ_0 — некоторое начальное значение;
 u_t — случайная переменная.

Пример тренда смешанного типа:

$$\xi_t = a_1 + a_2 t + u_t + q u_{t-1} + b \sin \omega t,$$

где a_1, a_2, q, b, ω — постоянные коэффициенты;
 u_t — случайная переменная.

Известно множество определений уровня и тренда ряда (см. [34, с. 16]), отличных от принятых нами. Существующие понятия тренда противоречивы и имеют условный характер. Каждое из этих определений скорее указывает на частный способ оценки тренда, а не на его сущность. Очень часто под трендом понимают детерминированную составляющую процесса, что значительно обедняет содержание термина и препятствует его применению для анализа временных рядов в общем случае.

Компоненты временного ряда ξ_t и ε_t ненаблюдаемы. Они являются теоретическими величинами. Их выделение и составляет предмет анализа временного ряда в задаче прогнозирования. Оценку будущих членов ряда обычно делают по прогнозной модели. Прогнозная модель — это модель, аппроксимирующая тренд. Прогнозы — это оценки будущих уровней ряда, а последовательность прогнозов для различных периодов упреждения $\tau = 1, 2, \dots, k$ составляет оценку тренда.

При построении прогнозной модели выдвигается гипотеза о динамике величины ξ_t , т. е. о характере тренда. Однако в связи с тем, что уверенность в гипотезе всегда относительна, рассматриваемые нами модели наделяются адаптивными свойствами, способностью к корректировке исходной гипотезы или даже к замене ее другой, более адекватно (с точки зрения точности прогнозов) отражающей поведение реального ряда.

Простейшая адаптивная модель основывается на вычислении так называемой экспоненциальной средней, к изучению которой мы переходим.

§ 2. ЭКСПОНЕНЦИАЛЬНОЕ СГЛАЖИВАНИЕ

Предположим, что исследуется временной ряд x_t . Выявление и анализ тенденции динамического ряда часто производится с помощью его выравнивания или сглаживания. Экспоненциальное сглаживание — один из простейших и распространенных приемов выравнивания ряда. В его основе лежит расчет экспоненциальных средних.

Экспоненциальное сглаживание ряда осуществляется по рекуррентной формуле

$$S_t = \alpha x_t + \beta S_{t-1}, \quad (1.1)$$

где S_t — значение экспоненциальной средней в момент t ;
 α — параметр сглаживания, $\alpha = \text{const}$, $0 < \alpha < 1$;
 $\beta = 1 - \alpha$.

Выражение (1.1) можно переписать следующим образом:

$$S_t = \alpha x_t + (1 - \alpha) S_{t-1} = S_{t-1} + \alpha (x_t - S_{t-1}). \quad (1.2)$$

Экспоненциальная средняя на момент t здесь выражена как экспоненциальная средняя предшествующего момента плюс доля α разницы текущего наблюдения и экспоненциальной средней прошлого момента.

Если последовательно использовать рекуррентное соотношение (1.1), то экспоненциальную среднюю S_t можно выразить через значения временного ряда x :

$$\begin{aligned} S_t &= \alpha x_t + \beta S_{t-1} = \alpha x_t + \alpha \beta x_{t-1} + \beta^2 S_{t-2} = \dots = \\ &= \alpha x_t + \alpha \beta x_{t-1} + \alpha \beta^2 x_{t-2} + \dots + \alpha \beta^{t-1} x_1 + \beta^t S_0 = \\ &= \alpha \sum_{i=0}^{t-1} \beta^i x_{t-i} + \beta^t S_0, \end{aligned} \quad (1.3)$$

где N — количество членов ряда;

S_0 — некоторая величина, характеризующая начальные условия для первого применения формулы (1.1) при $t = 1$.

Так как $\beta < 1$, то при $N \rightarrow \infty$ $\beta^N \rightarrow 0$, а сумма коэффициентов $\alpha \sum_{i=0}^{N-1} \beta^i \rightarrow 1$. Тогда

$$S_t = \alpha \sum_{i=0}^{\infty} \beta^i x_{t-i}.$$

Таким образом, величина S_t оказывается взвешенной суммой всех членов ряда. Причем веса падают экспоненциально в зависимости от давности («возраста») наблюдения. Это и объясняет, почему величина S_t названа экспоненциальной средней. Если, например, $\alpha = 0,3$, то текущее наблюдение будет иметь вес 0,3, а веса предшествующих данных составят соответственно 0,21; 0,147; 0,1029 и т. д.

Рассмотрим, ряд, генерированный моделью

$$x_t = a_1 + \varepsilon_t,$$

где $a_1 = \text{const}$;

ε_t — случайные неавтокоррелированные отклонения, или шум, со средним значением 0 и дисперсией σ^2 .

Применим к нему процедуру экспоненциального сглаживания (1.1). Тогда

$$\begin{aligned} S_t &= \alpha \sum_{i=0}^{\infty} \beta^i x_{t-i} = \alpha \sum_{i=0}^{\infty} \beta^i (a_1 + \varepsilon_{t-i}) = \\ &= a_1 + \alpha \sum_{i=0}^{\infty} \beta^i \varepsilon_{t-i}. \end{aligned}$$

Найдем математическое ожидание

$$M(S_t) = M(x_t) = a_1$$

и дисперсию

$$\begin{aligned} D(S_t) &= M[(S_t - a_1)^2] = M\left[\left(\alpha \sum_{i=0}^{\infty} \beta^i \varepsilon_{t-i}\right)^2\right] = \\ &= \alpha^2 \sum_{i=0}^{\infty} \beta^{2i} \sigma^2 = \frac{\alpha}{2-\alpha} \sigma^2. \end{aligned} \quad (1.4)$$

Так как $0 < \alpha < 1$, $D(S_t) < D(x_t) = \sigma^2$.

Таким образом, экспоненциальная средняя S_t имеет то же математическое ожидание, что и ряд x , но меньшую дисперсию. Как видно из (1.4), при высоком значении α дисперсия экспоненциальной средней незначительно отличается от дисперсии ряда x . Чем меньше α , тем в большей степени сокращается дисперсия экспоненциальной средней. Следовательно, экспоненциальное сглаживание можно представить как фильтр, на вход которого в виде потока последовательно поступают члены исходного ряда, а на выходе формируются текущие значения экспоненциальной средней. И чем меньше α , тем в большей степени фильтруются, подавляются колебания исходного ряда.

После появления работ Р. Брауна [47, 48] экспоненциальная средняя часто используется для краткосрочного прогнозирования. В этом случае предполагается, что ряд генерируется моделью

$$x_t = a_{1,t} + \varepsilon_t,$$

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 имеет вид

$$\hat{x}_\tau(t) = \hat{a}_{1,t},$$

где $\hat{x}_\tau(t)$ — прогноз, сделанный в момент t на τ единиц времени (шагов) вперед;
 $\hat{a}_{1,t}$ — оценка $a_{1,t}$ (знак $\hat{}$ над величиной здесь и далее будет означать оценку).

Средством оценки единственного параметра модели служит экспоненциальная средняя $\hat{a}_{1,t} = S_t$. Таким образом, все свойства экспоненциальной средней распространяются на прогнозную модель. В частности, если S_{t-1} рассматривать как прогноз на 1 шаг вперед, то в выражении (1.2) величина $(x_t - S_{t-1})$ есть погрешность этого прогноза, а новый прогноз S_t получается в результате корректировки предыдущего прогноза с учетом его ошибки. В этом и состоит существо адаптации.

При краткосрочном прогнозировании желательно как можно быстрее отразить изменения $a_{1,t}$ и в то же время как можно лучше «очистить» ряд от случайных колебаний.

Таким образом, с одной стороны, следует увеличивать вес более свежих наблюдений, что может быть достигнуто повышением α (см. (1.3)), с другой стороны, для сглаживания случайных отклонений величину α нужно уменьшить. Как видим, эти два требования находятся в противоречии. Поиск компромиссного значения α составляет задачу оптимизации модели.

Для уяснения процедуры расчета экспоненциальной средней и ее свойств рассмотрим числовой пример сглаживания ряда курса акций фирмы ИБМ¹ (см. табл. 1.1).

Таблица 1.1

Экспоненциальные средние*

№ точки (время)	Члены ряда**	$\alpha=0,1$	$\alpha=0,5$	$\alpha=0,9$	№ точки (время)	Члены ряда**	$\alpha=0,1$	$\alpha=0,5$	$\alpha=0,9$
1	510	506,4	508,0	509,6	16	512	505,7	513,3	513,1
2	497	505,5	502,5	498,3	17	510	506,1	511,7	510,3
3	504	505,3	503,2	503,4	18	506	506,1	508,8	506,4
4	510	505,8	506,6	509,3	19	515	507,0	511,9	514,1
5	509	506,1	507,8	509,0	20	522	508,5	517,0	521,2
6	503	505,8	505,4	503,6	21	523	509,9	520,0	522,8
7	500	505,2	502,7	500,4	22	527	511,6	523,5	526,6
8	500	504,7	501,4	500,0	23	523	512,8	523,2	523,4
9	500	504,2	500,7	500,0	24	528	514,3	525,6	527,5
10	495	503,3	497,8	495,5	25	529	515,8	527,3	528,9
11	494	502,4	495,9	494,2	26	538	518,0	532,7	537,1
12	499	502,0	497,5	498,5	27	539	520,1	525,8	538,8
13	502	502,0	499,7	501,2	28	541	522,2	538,4	540,8
14	509	502,7	504,4	508,3	29	543	524,3	540,7	542,8
15	525	505,0	514,7	523,3	30	541	525,9	540,9	541,2

* Во всех случаях начальное значение экспоненциальной средней было принято равным $\frac{1}{5} \sum_{i=1}^5 x_i = 506$.

** См. приложение № 4.

¹ Свойства сглаживания особенно наглядно проявляются при значительной колеблемости исходных данных. В связи с этим для иллюстрации взят один из показателей, отражающий конъюнктурные колебания американской экономики, — курс акций фирмы ИБМ, производящей электронно-вычислительные машины. Этот временной ряд уже использовался для испытания некоторых адаптивных моделей Р.Г. Брауном, из работы которого [47] он и взят.

Определим S_0 как $\frac{1}{5} \sum_{i=1}^5 x_i = \frac{1}{5} (510 + 497 + 504 + 510 + 509) = 506$. Дальнейшие вычисления при $\alpha = 0,1$ выглядят следующим образом:

$$S_1 = \alpha x_1 + (1 - \alpha) S_0 = 0,1 \cdot 510 + 0,9 \cdot 506 = 506,4;$$

$$S_2 = \alpha x_2 + (1 - \alpha) S_1 = 0,1 \cdot 497 + 0,9 \cdot 506,4 = 505,46;$$

$$S_3 = \alpha x_3 + (1 - \alpha) S_2 = 0,1 \cdot 504 + 0,9 \cdot 505,46 = 505,31 \text{ и т. д.}$$

Результаты вычислений экспоненциальных средних при $\alpha = 0,1$, $\alpha = 0,5$ и $\alpha = 0,9$ приведены в табл. 1.1.

На рис. 1.1 изображен график динамики временного ряда и экспоненциальных средних при $\alpha = 0,1$ и $\alpha = 0,5$. На графике наглядно проявляется влияние величины α на подвижность экспоненциальной средней.

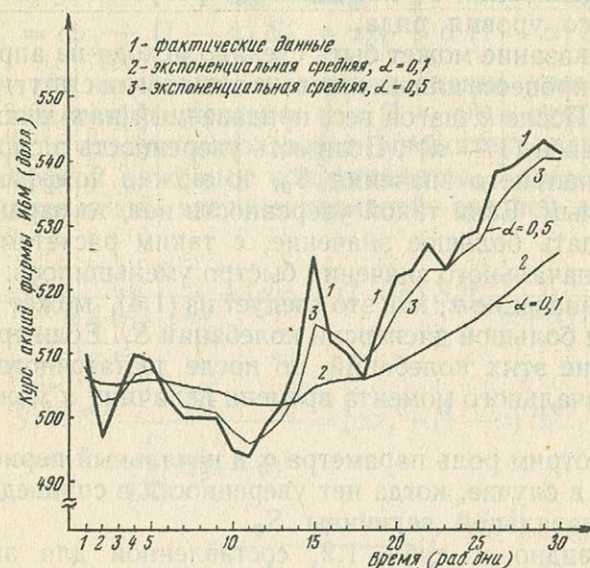


Рис. 1.1. Экспоненциальное сглаживание временного ряда

Экспоненциальное сглаживание является простейшим вариантом самообучающейся модели. Вычисления просты и выполняются итеративно. Они требуют даже меньше

Изменение весов в начальный период времени
при экспоненциальном сглаживании с $\alpha = 0,1$

Итерация	Вес началь- ной величины	Вес первого члена ряда	Вес второго члена ряда	Вес третьего члена ряда	Вес чет- вертого члена ряда
1	0,900	0,100			
2	0,810	0,090	0,100		
3	0,729	0,081	0,090	0,100	
4	0,656	0,073	0,081	0,090	0,100

§ 3. НАЧАЛЬНЫЕ УСЛОВИЯ ЭКСПОНЕНЦИАЛЬНОГО СГЛАЖИВАНИЯ

Экспоненциальное выравнивание всегда требует предыдущего значения экспоненциальной средней. Когда процесс только начинается, должна быть некоторая величина S_0 , которая может быть использована в качестве значения, предшествующего S_1 . Если есть прошлые данные к моменту начала выравнивания, то в качестве начального значения S_0 можно использовать арифметическую среднюю всех имеющихся точек или какой-то их части. Когда для такого оценивания S_0 нет данных, требуется предсказание начального уровня ряда.

Предсказание может быть сделано исходя из априорных знаний о процессе или на основе его аналогии с другими процессами. После k шагов вес, придаваемый начальному значению, равен $(1 - \alpha)^k$. Если есть уверенность в справедливости начального значения S_0 , то можно коэффициент α взять малым. Если такой уверенности нет, то параметру α следует дать большое значение, с таким расчетом, чтобы влияние начального значения быстро уменьшилось. Однако большое значение α , как это следует из (1.4), может явиться причиной большой дисперсии колебаний S_t . Если требуется подавление этих колебаний, то после достаточного удаления от начального момента времени величину α можно убавить.

Рассмотрим роль параметра α в начальный период сглаживания в случае, когда нет уверенности в справедливости выбора начальной величины S_0 .

Как видно из табл. 1.2, составленной для значения $\alpha = 0,1$, начальная величина S_0 в течение длительного времени имеет чрезмерный вес. Даже после 20 итераций вес S_0 равен 0,122, что означает, что ему дается все еще больший вес, чем любому другому члену ряда. Таким образом, в этом случае получение прогнозов по экспоненциальной средней, построенной на малом отрезке ряда (выборке), чревато большими ошибками. Для того чтобы элиминиро-

вать избыточный вес, приданный начальной величине, Р. Вейд [106] предлагает модифицировать процедуру сглаживания следующим образом.

Для исходного момента времени запишем:

$$S'_0 = \alpha S_0;$$

$$S'_1 = \alpha x_1 + (1 - \alpha) S'_0 = \alpha x_1 + \alpha (1 - \alpha) S_0,$$

где S_0 — как и раньше, начальная оценка уровня ряда.

Так как коэффициенты α и $\alpha(1 - \alpha)$ в сумме теперь не дают 1, то следует использовать множитель, равный единице, деленной на сумму коэффициентов. Таким образом, модифицированной экспоненциальной средней для $t = 1$ будет

$$\tilde{S}_1 = S'_1 \frac{1}{\alpha + \alpha(1 - \alpha)} = [\alpha x_1 + (1 - \alpha) S'_0] \frac{1}{\alpha + \alpha(1 - \alpha)}$$

и вообще

$$\tilde{S}_t = S'_t \frac{1}{\sum_{i=0}^t \alpha(1 - \alpha)^i} = [\alpha x_t + (1 - \alpha) S'_{t-1}] \times \frac{1}{\sum_{i=0}^t \alpha(1 - \alpha)^i}.$$

По табл. 1.3 можно видеть, что сущность этого метода состоит в том, чтобы убрать избыточный вес от веса, даваемого начальному значению S_0 , и распределить его пропорционально по всем членам ряда. Прогнозы, получаемые

Таблица 1.3

Изменение весов в начальный период времени
при $\alpha = 0,1$ в модифицированной модели

Итерация	Вес начальной величины	Вес первого члена ряда	Вес второго члена ряда	Вес третьего члена ряда	Вес четвертого члена ряда
1	0,474	0,526			
2	0,299	0,332	0,369		
3	0,212	0,236	0,262	0,291	
4	0,160	0,178	0,198	0,220	0,244

по соответствующей модифицированной модели, основываются в большей степени на фактических данных, чем на предварительной оценке S_0 даже при малых выборках. Для того чтобы сократить время вычислений, целесообразно вернуться к обычному экспоненциальному сглаживанию,

когда сумма коэффициентов $\sum_{i=0}^t \alpha (1 - \alpha)^i$ приближается к 1. На основе эмпирического анализа Р. Вейд рекомендует осуществлять такой переход при сумме коэффициентов 0,995. При заданном значении α можно заранее определить, на каком шаге следует вернуться к обычной модели.

§ 4. ВЫБОР ПОСТОЯННОЙ СГЛАЖИВАНИЯ

Выбору величины постоянной сглаживания следует уделять особое внимание. Поиски должны быть направлены на отыскание оснований для выбора наилучшего значения. Нужно учитывать условия, при которых эта величина должна принимать значения, близкие то одному крайнему значению, то другому. Нетрудно заметить, что при $\alpha = 0$ $S_t = S_0$ представляет случай абсолютной фильтрации и полного отсутствия адаптации, а при $\alpha = 1$ приходим к так называемой наивной модели $\hat{x}_t(t) = S_t = x_t$, в соответствии с которой прогноз на любой срок равен текущему фактическому значению ряда. На практике эта модель из-за простоты пользуется особой популярностью.

В § 2 уже отмечалось, что постоянная сглаживания характеризует скорость реакции модели $\hat{x}_t(t) = S_t$ на изменения уровня процесса, но одновременно определяет и спо-

собность системы сглаживать случайные отклонения. Поэтому величине α следует давать то или иное промежуточное значение между 0 и 1 в зависимости от конкретных свойств динамического ряда.

В качестве удовлетворительного компромисса Р. Браун рекомендует брать α в пределах от 0,1 до 0,3. Эта рекомендация некритически повторена в ряде работ. Между тем в [5] показано, что даже при прогнозировании ряда, использованного Брауном для иллюстрации, наилучшие результаты получаются при $\alpha = 0,9$. Наш опыт работы с экономическими рядами показывает, что наибольшая точность прогнозирования может быть достигнута при любых допустимых значениях α . Однако, как правило, если в результате испытаний обнаружено, что наилучшее значение константы α близко к 1, следует проверить законность выбора модели данного типа. Часто к большим значениям α приводит наличие в исследуемом ряде ярко выраженных тенденций или сезонных колебаний. В этом случае для получения эффективных прогнозов требуется другая модель.

Ясно, что наилучшее значение α в общем случае должно зависеть от срока прогнозирования τ . Для конъюнктурных прогнозов в большей мере должна учитываться свежая информация. При увеличении периода упреждения τ более поздняя информация, отражающая последнюю конъюнктуру, должна, по-видимому, иметь несколько меньший вес, чем в случае малых τ . Для того чтобы сгладить конъюнктурные колебания, следует в большей мере учитывать информацию за прошлые периоды времени. Для проведения подобного анализа вводят понятие среднего возраста данных. Возраст текущего наблюдения равен 0, возраст предыдущего наблюдения равен 1 и т. д. Средний возраст — это сумма взвешенных возрастов данных, использованных для подсчета сглаженной величины. Причем возраста имеют те же веса, что и соответствующая информация. При экспоненциальном выравнивании вес, даваемый точке с возрастом k , равен $\alpha\beta^k$, где $\beta = 1 - \alpha$ и средний возраст информации равен:

$$k = 0 \cdot \alpha + 1 \cdot \alpha\beta + 2 \cdot \alpha\beta^2 + \dots = \alpha \sum_{k=0}^{\infty} k\beta^k = \frac{\beta}{\alpha}.$$

Таким образом, чем меньше α , тем больше средний возраст информации. Для конъюнктурных прогнозов значение

α , как правило, надо брать большим, а для более долгосрочных — малым. Это положение иллюстрирует рис. 1.2, на котором отображена зависимость стандартной ошибки прогнозирования, обычно принимаемой за показатель точности, от α . Однако характер зависимостей, аналогичных

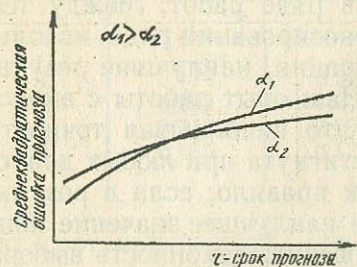


Рис. 1.2. Примерная зависимость среднеквадратической ошибки прогноза от величины постоянной сглаживания α и периода упреждения τ

тем, что отражены на рисунке, следует изучать специально в каждом конкретном случае.

Теоретический анализ проблемы выбора постоянной сглаживания при применении простейшей экспоненциальной модели для прогнозирования стационарного процесса с автокорреляционной функцией вида $\rho_k = \rho_1^k$, где ρ_1 — коэффициент автокорреляции при лаге 1, k — лаг, проведен Д. Р. Коксом [53] и Дж. Д. Кохеном [52].

Они показали, что минимум среднего квадрата ошибки при прогнозировании такого ряда на 1 шаг вперед ($\tau = 1$) будет при

$$\alpha_{\text{опт}} = \begin{cases} \frac{3\rho_1 - 1}{2\rho_1} & 1/3 < \rho_1 \leq 1; \\ 0 & -1 \leq \rho_1 \leq 1/3. \end{cases} \quad (1.5)$$

Соответствующая дисперсия ошибки прогноза равна:

$$D_e = \begin{cases} \frac{8\rho_1(1-\rho_1)}{(1+\rho_1)^2} \sigma_x^2 & 1/3 < \rho_1 \leq 1; \\ \sigma_x^2 & -1 \leq \rho_1 \leq 1/3. \end{cases}$$

Табл. 1.4 показывает соотношения между ρ_1 , $\alpha_{\text{опт}}$ и точностью прогнозирования на 1 шаг вперед.

Табл. 1.5 показывает, что для данной ρ_1 величина D_e при $\tau = 1$ слабо зависит от α , так что точность прогноза в некоторой окрестности $\alpha_{\text{опт}}$ нечувствительна к выбору постоянной сглаживания.

Результат (1.5) означает, что если $\rho_1 > 1/3$, то при соответствующем выборе величины α экспоненциальная средняя в определенной степени отражает колебания, связан-

Таблица 1.4

Соотношения между ρ_1 , $\alpha_{\text{опт}}$, D_e/σ_x^2 при прогнозировании стационарного процесса с $\rho_k = \rho_1^k$ по модели экспоненциального сглаживания ($\tau = 1$)

ρ_1	$\alpha_{\text{опт}}$	D_e/σ_x^2	ρ_1	$\alpha_{\text{опт}}$	D_e/σ_x^2
$\leq 1/3$	0	1	0,7	0,786	0,581
0,4	0,250	0,980	0,8	0,875	0,395
0,5	0,500	0,889	0,9	0,944	0,199
0,6	0,667	0,750	0,95	0,974	0,100

Таблица 1.5

D_e/σ_x^2 как функция от α при $\rho_1 > 1/3$

α	$\rho_1 = 0,4$	$\rho_1 = 0,7$	$\rho_1 = 0,9$	α	$\rho_1 = 0,4$	$\rho_1 = 0,7$	$\rho_1 = 0,9$
1	1,200	0,600	0,200	0,4	0,987	0,647	0,272
0,9	1,136	0,587	0,200	0,3	0,980	0,692	0,318
0,8	1,087	0,581	0,203	0,2	0,980	0,758	0,397
0,7	1,049	0,584	0,211	0,1	0,987	0,853	0,554
0,6	1,020	0,595	0,223	0	1,000	1,000	1,000
0,5	1,000	0,615	0,242	$\alpha_{\text{опт}}$	0,980	0,581	0,200

ные сильной автокорреляцией. С другой стороны, если $\rho_1 \leq 1/3$, то наибольшее, что может дать простейшая модель, это оценка среднего уровня, вокруг которого варьирует процесс. Но и при $\rho_1 \leq 1/3$ на практике не следует брать α слишком малым, иначе предиктор окажется нечувствительным к изменениям среднего уровня.

Определенным руководством при этом может служить табл. 1.6, которая характеризует дисперсии ошибок, получаемых при прогнозировании стационарных процессов с $\rho_k = \rho_1^k$, где $\rho_1 \leq 1/3$. Из таблицы видно, что при $\rho_1 < 0$ можно добиться немногого, полагая α меньше 0,1—0,2. Вообще говоря, очевидно, что если $\rho_1 < 0$, то простейшая модель экспоненциального типа не является хорошим предиктором.

Если $\tau > 1$, то существенно повышается критическая величина $\rho_{1\text{крит}}$, ниже которой оптимальное значение α равно 0. Этот факт иллюстрирует табл. 1.7.

Таблица 1.6

Дисперсия ошибки прогноза
для стационарного процесса с $\rho_k = \rho_1^k$,
где $\rho_1 \leq 1/3$, $\tau = 1$

α	D_e / σ_x^2				
	$\rho_1 = 1/3$	$\rho_1 = 1/10$	$\rho_1 = 0$	$\rho_1 = -1/4$	$\rho_1 = -1/2$
0	1,000	1,000	1,000	1,000	1,000
0,05	1,001	1,020	1,026	1,036	1,043
0,10	1,002	1,041	1,053	1,074	1,089
0,20	1,011	1,087	1,111	1,157	1,190
0,30	1,022	1,139	1,176	1,252	1,307
0,40	1,042	1,197	1,250	1,359	1,442
0,50	1,067	1,263	1,333	1,481	1,600

Таблица 1.7

Зависимость ρ_1 крит от τ

τ	1	2	3
ρ_1 крит	0,333	0,516	0,821

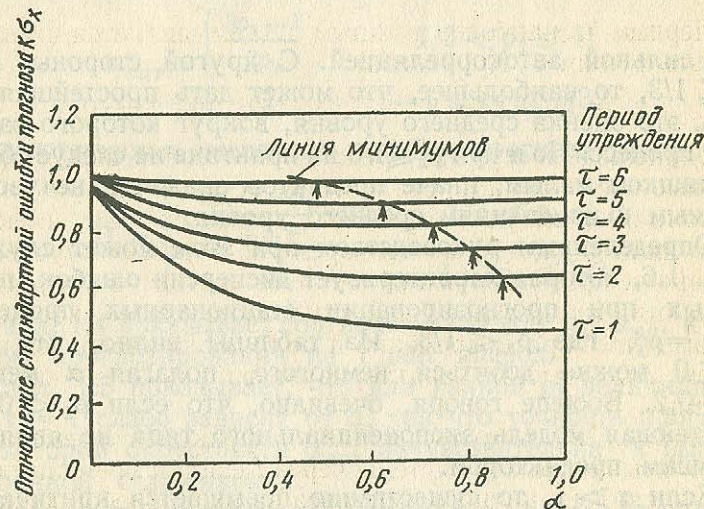
Рис. 1.3. Влияние α на точность прогнозирования при однократном экспоненциальном сглаживании данных с $\rho_k = 0,9^{|k|}$

Рис. 1.3 показывает стандартную ошибку прогнозирования для всех значений постоянной сглаживания в случае стационарного процесса с сильной автокорреляцией $\rho_k = 0,9^{|k|}$, т. е. автоковариацией $R_{xx}(k) = \sigma_x^2 (0,9)^{|k|}$.

Пунктирная линия выделяет геометрическое место точек решений, которые минимизируют ошибку прогнозирования. Отсюда можно сделать вывод, что если данные сильно коррелированы и период упреждения τ мал, то сглаживать не стоит. Целесообразно в качестве прогноза использовать наиболее позднее наблюдение.

§ 5. РЕАКЦИЯ МОДЕЛИ НА НЕКОТОРЫЕ СТАНДАРТНЫЕ ВХОДНЫЕ ПОТОКИ ДАННЫХ

Рассмотрим, каковы реакции модели экспоненциального сглаживания на некоторые стандартные входные потоки, содержащие типичные нарушения стационарности, и как сказывается на этих реакциях величина постоянной сглаживания. Такие входные потоки носят абстрактный характер, однако проводимый с их помощью анализ позволяет глубже изучить адаптивные свойства модели.

Реакция на импульс

Первым и наиболее важным тестом является импульс (дельта-функция Кронекера):

$$x_t = \delta(t) \text{ или } x_0 = 1; x_t = 0 \text{ при } t \neq 0.$$

Одиночные импульсы представляют собой события, вызванные посторонними для изучаемого явления причинами. В экономике импульс отражает разовое явление, имеющее место лишь в данный момент времени. В общем случае требование к прогнозирующей системе состоит в том, чтобы ее реакция на импульс была как можно слабее, ибо действие импульса кратковременно и не будет проявляться в будущем. Эта реакция характеризует фильтрующие свойства системы.

Реакция на импульс является функцией времени h_t . Она описывает выход системы через t единиц времени после поступления единичного импульса на вход. Любой дискретный временной ряд можно рассматривать как серию импульсов соответствующей амплитуды. Если реакция линейной

системы на единичный импульс есть h_i , то ее реакция на последовательность импульсов x в момент t определяется как результат сложения реакций на каждый импульс, т. е. как сумма:

$$y_t = \sum_{i=0}^{\infty} h_i x_{t-i}. \quad (1.6)$$

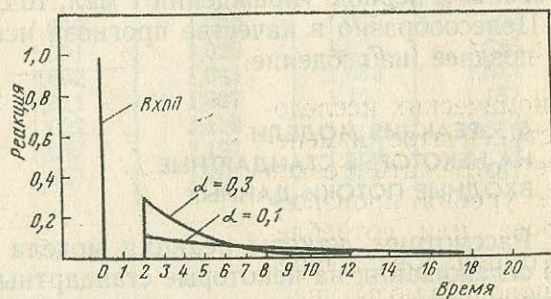


Рис. 1.4. Реакции простейшей прогнозной модели экспоненциального типа на единичный импульс в зависимости от постоянной сглаживания α при $\tau=2$ (дискретные значения соединены плавной линией)

Реакция модели экспоненциального сглаживания, как известно, является линейной функцией членов временного ряда и имеет вид

$$S_t = \sum_{i=0}^{\infty} \alpha \beta^i x_{t-i}. \quad (1.7)$$

Из сопоставления (1.6) и (1.7) следует, что $h_i = \alpha \beta^i$. Таким образом, можно говорить или о весах, с которыми взвешиваются члены ряда, или о реакции на единичный импульс. Любая дискретная система, выражаемая с помощью линейных уравнений в конечных разностях с постоянными коэффициентами, может быть исчерпывающе описана с помощью ее реакции на единичный импульс.

На рис. 1.4 изображены реакции на единичный импульс прогнозной модели экспоненциального типа с различными постоянными сглаживания при $\tau = 2$. В этой модели текущее значение экспоненциальной средней относится на момент $t + \tau$, для которого составляется прогноз. На рисунке видно, что модель с большей постоянной сглаживания реагирует на импульс сильнее, а этот эффект нежелателен.

Реакция на ступенчатое изменение

При наличии существенных долговременных изменений структуры ряда необходимо добиваться, чтобы модель учитывала их как можно быстрее. Рассмотрим прежде всего ступенчатое изменение уровня входного потока. Запишем его аналитически:

$$x_t = a_1 \text{ для } t < t_1;$$

$$x_t = a_2 \text{ для } t \geq t_1 \quad a_1 \neq a_2.$$

В экономических исследованиях ступенчатое изменение может отражать резкое изменение уровня производства, спроса или потребления какого-либо товара в результате поворота моды, внедрения значительного научного открытия, изменения политической или внешнеэкономической ситуации.

Единичный ступенчатый скачок является следующей функцией:

$$x_t = 0 \text{ для } t < 0;$$

$$x_t = 1 \text{ для } t \geq 0.$$

Так как рассматриваемая процедура сглаживания является линейной, то можно определить реакцию на такого рода изменения с помощью реакции на единичный импульс. Реакция модели будет

$$S_t = \sum_{i=0}^{\infty} h_i x_{t-i} = \sum_{i=0}^t \alpha \beta^i = 1 - \beta^{t+1}.$$

С ростом t член β^{t+1} становится незначительным и величина S_t приближается к уровню x_t .

На рис. 1.5 показана реакция модели экспоненциальной средней на ступенчатый скачок при различных значениях постоянной сглаживания ($\tau = 3$). Здесь, как видим, для скорейшего отражения произошедших изменений выгоднее брать большее значение α .

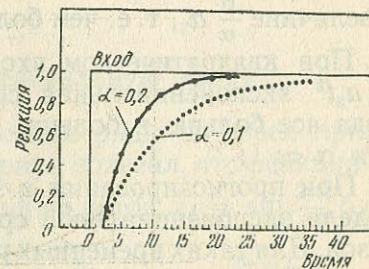


Рис. 1.5. Влияние α на скорость реакции прогнозной модели при $\tau=3$

Реакция на линейную и квадратическую функции

Временные экономические ряды часто имеют тенденцию линейного или параболического роста. Можно показать, что при линейно-нарастающем входном потоке $a_1 + a_2 t$ экспоненциальная средняя будет постоянно отставать от временного ряда и что это отставание в конце концов стремится к величине $\frac{\beta}{\alpha} a_2$, т. е. чем больше α , тем меньше отставание.

При квадратическом входном потоке типа $a_1 + a_2 t + a_3 t^2$ экспоненциальная средняя отстает от временного ряда все больше и больше. Наименьшее отставание будет при $\alpha \approx 1$.

При прогнозировании в этих обоих случаях нельзя от модели экспоненциальной средней ожидать хороших прогнозов. Для таких временных рядов требуются другие модели, которые мы рассмотрим ниже.

В сводной табл. 1.8 приведены реакции экспоненциальной средней на рассмотренные входные потоки.

Таблица 1.8

Реакции экспоненциальной средней
на стандартные входные потоки

Входной поток x_t	Экспоненциальная средняя S_t	Установившаяся ошибка $x_\infty - S_\infty$
Импульс $x_t = \delta(t)$	$\alpha \beta^t$	0
Ступенчатый скачок $x_t = \begin{cases} 0 & t < 0, \\ 1 & t \geq 0 \end{cases}$	$1 - \beta^{t+1}$	0
Линейный рост $x_t = t, t \geq 0$	$t - \frac{\beta}{\alpha} (1 - \beta^t)$	$\frac{\beta}{\alpha}$
Парабола $x_t = t^2, t \geq 0$	$t^2 - \frac{2\beta}{\alpha} t + \frac{\beta(1+\beta)}{\alpha^2} (1 - \beta^t)$	$\frac{2\beta}{\alpha} t - \frac{\beta(1+\beta)}{\alpha^2}$

Приведенные реакции простейшей модели на входные ряды, представляющие типичные нарушения стационарности, говорят о противоречивых требованиях к постоян-

ной сглаживания в задаче прогнозирования. Это свидетельствует о существовании оптимального значения α , которое зависит от того, какой тип изменений в процессе встречается наиболее часто. Кроме того, из табл. 1.8 следует, что даже когда адаптивная модель становится неадекватной исследуемому ряду, она все же в какой-то мере учитывает реальные изменения, хотя в этом случае лучше всего заменить модель.

Реакция на синусоидальную волну

Во многих случаях временные ряды по своей природе являются периодическими. Браун показал, что реакция экспоненциальной средней на входной поток $x_t = \sin \frac{2\pi t}{12}$ будет:

$$S_t = \frac{\alpha \beta^{t+1}}{2(1 - \sqrt{3}\beta + \beta^2)} + \frac{\alpha}{\sqrt{1 - \sqrt{3}\beta + \beta^2}} \sin \frac{2\pi(t - \varphi)}{12},$$

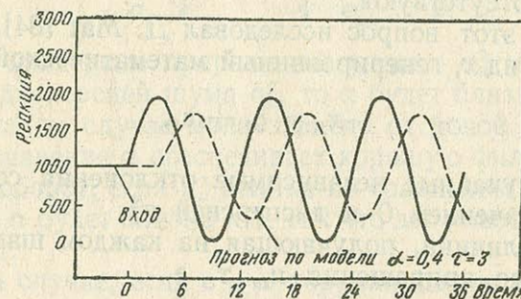


Рис. 1.6. Прогнозирование синусоидальных колебаний с помощью экспоненциального сглаживания

где фазовый угол φ определяется из соотношения

$$\operatorname{tg} \varphi = \frac{\beta}{\sqrt{3}\beta - 2}.$$

Первое слагаемое стремится к нулю с течением времени, и в результате экспоненциальная средняя будет также изменяться по синусоиде того же самого периода, но с амплитудой и угловым сдвигом, зависящими от α . Реакция на синусоидальную волну показана на рис. 1.6.

При определенных значениях τ и α ряд x_t и его прогнозы по экспоненциальной средней могут оказаться колеблющимися в противофазе, что приведет к очень большим ошибкам. Как предиктор синусоидальных процессов данная модель совершенно неудовлетворительна.

§ 6. СВОЙСТВО ОПТИМАЛЬНОСТИ

Главное достоинство прогнозной модели, основанной на экспоненциальной средней, которое мы установили, состоит в том, что она способна последовательно адаптироваться к новому уровню процесса без значительного реагирования на случайные отклонения. Однако целесообразно определить статистические свойства таких временных рядов, по которым этот метод прогнозирования работает особенно хорошо. В результате можно будет лучше судить о сфере применения этого метода или модифицировать его в том случае, когда необходимые свойства у временного ряда отсутствуют.

Первым этот вопрос исследовал Д. Мат [84]. Он взял временной ряд x , генерированный математической моделью

$$x_t = \xi_t + \varepsilon_t, \quad (1.8)$$

где ε_t — случайные независимые отклонения со средним значением 0 и дисперсией σ_ε^2 ;

ξ_t — величина, получающаяся на каждом шаге некоторые приращения u_t , т. е.

$$\xi_t = \xi_{t-1} + u_t = \sum_{i=1}^t u_i,$$

где величины u_i независимы, имеют среднее значение 0 и дисперсию σ_u^2 .

Значения величин ε и u пока предполагаются независимыми. Такой временной ряд можно рассматривать как случайное движение уровня процесса ξ_t , на которое наложен шум ε_t .

Д. Мат поставил задачу отыскать оптимальные веса ω_k в предикторе вида

$$\hat{x}_1(t) = \sum_{k=0}^{\infty} \omega_k x_{t-k}, \quad (1.9)$$

которые минимизируют дисперсию ошибки прогноза ряда (1.8). Он пришел к выводу, что оптимальными весами в модели (1.9) будут:

$$\omega_k = (1 - \lambda) \lambda^k, \quad k = 0, 1, 2, \dots;$$

$$\lambda = 1 + \frac{1}{2} \frac{\sigma_u^2}{\sigma_\varepsilon^2} - \frac{\sigma_u}{\sigma_\varepsilon} \sqrt{1 + \frac{1}{4} \frac{\sigma_u^2}{\sigma_\varepsilon^2}} < 1.$$

Веса имеют ту же самую форму, что и в модели экспоненциальной средней, определяемой выражением

$$\hat{x}_1(t) = \alpha \sum_{k=0}^{\infty} (1 - \alpha)^k x_{t-k},$$

если положить $\alpha = 1 - \lambda$. Следовательно, модель экспоненциальной средней в данном случае является оптимальной при

$$\alpha = \frac{\sigma_u}{\sigma_\varepsilon} \sqrt{1 + \frac{1}{4} \frac{\sigma_u^2}{\sigma_\varepsilon^2}} - \frac{1}{2} \frac{\sigma_u^2}{\sigma_\varepsilon^2}.$$

Если дисперсия приращений уровня σ_u^2 мала по сравнению с дисперсией шума σ_ε^2 , то α будет близка к 0. Прогнозы в таком случае мало зависят от новой информации. Низкое значение α обеспечивает хорошую фильтрацию шума. И наоборот, если σ_u^2 велика по сравнению с дисперсией шума, то α будет близка к 1, так что вес новой информации возрастет.

В том случае, если ε и u коррелированы и

$$M(\varepsilon_t u_t) = \sigma_{\varepsilon u};$$

$$M(\varepsilon_t u_{t'}) = 0 \quad t \neq t',$$

необходимо только в выражении для α заменить отношение

$$\frac{\sigma_u^2}{\sigma_\varepsilon^2} \quad \text{на} \quad \frac{\sigma_u^2}{\sigma_\varepsilon^2 + \sigma_{\varepsilon u}}.$$

§ 7. МОДЕЛИ ЛИНЕЙНОГО РОСТА

В § 5 было показано, что экспоненциальная средняя приводит к смещенным прогнозам, т. е. дает систематическую ошибку, когда временной ряд имеет тенденцию линейного роста. Для этого случая разработано несколько вариантов адаптивных моделей, также использующих про-

цедуру экспоненциального сглаживания. В основе моделей лежит гипотеза о том, что прогноз может быть получен по уравнению

$$\hat{x}_t(t) = \hat{a}_{1,t} + \tau \hat{a}_{2,t},$$

где $\hat{a}_{1,t}$, $\hat{a}_{2,t}$ — текущие оценки коэффициентов адаптивного полинома первого порядка.

Одной из первых моделей этого типа была двухпараметрическая модель Ч. Хольта [68], в которой оценка коэффициентов производится следующим образом:

$$\hat{a}_{1,t} = \alpha_1 x_t + (1 - \alpha_1) (\hat{a}_{1,t-1} + \hat{a}_{2,t-1}); \quad (1.10)$$

$$\hat{a}_{2,t} = \alpha_2 (\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \alpha_2) \hat{a}_{2,t-1},$$

где α_1 и α_2 — параметры экспоненциального сглаживания ($0 < \alpha_1, \alpha_2 < 1$), которые мы также будем называть параметрами адаптации.

Эти уравнения могут быть переписаны в виде:

$$\hat{a}_{1,t} = \hat{a}_{1,t-1} + \hat{a}_{2,t-1} + \alpha_1 e_t;$$

$$\hat{a}_{2,t} = \hat{a}_{2,t-1} + \alpha_1 \alpha_2 e_t,$$

где $e_t = x_t - \hat{x}_1(t-1)$ — ошибка прогноза.

Частным случаем модели Хольта является модель линейного роста Брауна:

$$\hat{a}_{1,t} = \hat{a}_{1,t-1} + \hat{a}_{2,t-1} + (1 - \beta^2) e_t; \quad (1.11)$$

$$\hat{a}_{2,t} = \hat{a}_{2,t-1} + (1 - \beta)^2 e_t,$$

где параметр β — коэффициент дисконтирования, характеризующий обесценение данных наблюдения за единицу времени, $0 < \beta < 1$.

Если модель Хольта усовершенствовать путем включения разности ошибок, то получим полную трехпараметрическую модель прогнозирования Дж. Бокса и Г. Дженкинса [44]:

$$\hat{x}_t(t) = \hat{a}_{1,t} + \tau \hat{a}_{2,t};$$

$$\hat{a}_{1,t} = \alpha_1 x_t + (1 - \alpha_1) (\hat{a}_{1,t-1} + \hat{a}_{2,t-1}) + \alpha_3 (e_t - e_{t-1});$$

$$\hat{a}_{2,t} = \alpha_2 (\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \alpha_2) \hat{a}_{2,t-1},$$

где $\alpha_1, \alpha_2, \alpha_3$ являются параметрами модели, $0 < \alpha_1, \alpha_2, \alpha_3 < 1$; $e_t = x_t - \hat{x}_1(t-1)$ — ошибка прогнозирования.

На основе практических испытаний модели на многих экономических рядах Бокс и Дженкинс пришли к выводу, что включение в модель разности ошибок не является необходимым. Коэффициент α_3 всегда оказывался близким к нулю. П. Харрисон [65] пришел к такому же заключению. Это объясняется стохастическим характером данных, и, в частности, тем, что корреляция ошибок в подобных случаях неустойчива.

Харрисон провел эмпирическое сравнение однопараметрической модели Брауна с многопараметрическими моделями. Многопараметрические модели ни в одном случае не дали заметного преимущества. Поэтому на практике для прогнозирования рядов с линейной тенденцией предпочтительнее использовать более простую модель Брауна. Из теоретического сопоставления различных моделей, проведенного П. Харрисоном [67] и Д. Вардом [108], следует аналогичный вывод. К положительным чертам метода Брауна можно отнести следующие: логичная, ясная и легко понимаемая концепция; оптимальное значение единственного параметра можно быстро найти эмпирическим путем; коэффициенты модели прогнозирования оцениваются совместно таким образом, чтобы уменьшить автокорреляцию в остатках. Все это делает модель Брауна легко применимой. Мы еще вернемся к ней в следующей главе.

§ 8. СТОХАСТИЧЕСКИЙ ПРОЦЕСС ТЕЙЛА И ВЕЙДЖА

Г. Тейл и С. Вейдж [103] аналогично тому, как это сделал Д. Мат [84] при изучении экспоненциальной средней, в целях дальнейшего изучения свойств адаптивных моделей предложили применить двухпараметрический предиктор Хольта (1.10) для прогнозирования некоторого вероятностного процесса, характеризующегося стохастическим трендом. Они вывели выражения для определения оптимальных параметров адаптации, минимизирующих средний квадрат ошибки прогнозирования.

Процесс Тейла—Вейджа аналитически записывается как:

$$\begin{aligned} x_t &= a_{1,t} + e_t; \\ a_{1,t} &= a_{1,t-1} + a_{2,t}; \\ a_{2,t} &= a_{2,t-1} + v_t, \end{aligned} \quad (1.12)$$

где $a_{1,t}$ — значение уровня исследуемого временного ряда x_t в момент t ;
 $a_{2,t}$ — прирост уровня от момента $t-1$ к моменту t ;
 ε_t, v_t — временные последовательности с нулевым математическим ожиданием, постоянными дисперсиями и отсутствием ковариации, т. е.

$$M(\varepsilon_t) = M(v_t) = 0;$$

$$M(\varepsilon_t \varepsilon_{t'}) = \begin{cases} \sigma_\varepsilon^2 & t = t'; \\ 0 & t \neq t'; \end{cases}$$

$$M(v_t v_{t'}) = \begin{cases} \sigma_v^2 & t = t'; \\ 0 & t \neq t'; \end{cases}$$

$$M(v_t \varepsilon_{t'}) = 0 \text{ для любой пары } (t, t').$$

Временной ряд x_t не является стационарным и не имеет строго определенной автоковариационной функции. Однако М. Нерлов и С. Вейдж [89] показали, что из уравнений (1.12) следует стационарность вторых разностей процесса x_t , которые мы обозначим через y_t :

$$y_t = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = \nabla^2 x_t = \nabla^2 a_{1,t} + \nabla^2 \varepsilon_t = v_t + \nabla^2 \varepsilon_t,$$

где ∇ — разностный оператор, $\nabla x_t = x_t - x_{t-1}$, $\nabla^2 x_t = \nabla(\nabla x_t)$.

Вторые разности имеют вполне определенную автоковариационную функцию

$$\text{cov}_{yy}(k) = M(y_t y_{t-k}) = \begin{cases} (g^2 + 6) \sigma_\varepsilon^2 & k = 0; \\ -4\sigma_\varepsilon^2 & |k| = 1; \\ \sigma_\varepsilon^2 & |k| = 2; \\ 0 & |k| > 2, \end{cases} \quad (1.13)$$

где

$$g^2 = \frac{\sigma_v^2}{\sigma_\varepsilon^2}.$$

Эти свойства могут быть использованы для решения вопроса о возможности адекватного представления наблюдаемого временного ряда процессом Тейла—Вейджа. При этом не следует упускать из виду то обстоятельство, что оценки автоковариационной функции являются довольно грубы-

ми и коррелированными и точные равенства (1.13) на практике будут выполняться лишь приближенно.

Схема составления прогноза в соответствии с (1.10) выглядит следующим образом:

$$\hat{a}_{1,t} = \alpha_1 x_t + (1 - \alpha_1) (\hat{a}_{1,t-1} + \hat{a}_{2,t-1}); \quad (1.14)$$

$$\hat{a}_{2,t} = \alpha_2 (\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \alpha_2) \hat{a}_{2,t-1}; \quad (1.15)$$

$$\hat{x}_t(t) = \hat{a}_{1,t} + \tau \hat{a}_{2,t};$$

$$0 < \alpha_1, \alpha_2 < 1.$$

Если ошибку прогноза, сделанного в момент t на 1 шаг вперед, обозначить через $e_1(t)$, то уравнения адаптации (1.14) и (1.15) можно записать в виде:

$$\hat{a}_{1,t} = \hat{a}_{1,t-1} + \hat{a}_{2,t-1} + \alpha_1 e_1(t-1);$$

$$\hat{a}_{2,t} = \hat{a}_{2,t-1} + \gamma e_1(t-1);$$

$$\gamma = \alpha_1 \alpha_2.$$

Ошибка прогноза:

$$e_1(t) = x_{t+1} - \hat{a}_{1,t} - \hat{a}_{2,t} = (a_{1,t} + a_{2,t} + v_{t+1} + \varepsilon_{t+1}) - \hat{a}_{1,t} - \hat{a}_{2,t} = (a_{1,t} - \hat{a}_{1,t}) + (a_{2,t} - \hat{a}_{2,t}) + v_{t+1} + \varepsilon_{t+1}.$$

Следовательно, ошибка прогноза является суммой трех компонент: ошибки оценки уровня процесса в момент t , ошибки оценки прироста уровня в момент t и комбинации случайных компонент v и ε в момент $t+1$.

Очевидно, что определение оптимальных α_1 и γ эквивалентно определению оптимальных α_1 и α_2 . Оптимум обычно отыскивается путем минимизации среднего квадрата ошибки прогноза. Но когда рассматриваются нестационарные временные ряды, то в общем случае не очевидно, что средний квадрат ошибок прогнозирования адаптивным методом является величиной устойчивой, которая может быть минимизирована. Используя соотношение

$$\hat{y}_{t+1} = \hat{x}_{t+1} - 2x_t + x_{t-1},$$

Нерлов и Вейдж показали, что проблема прогнозирования x_{t+1} эквивалентна задаче прогнозирования второй разности y_{t+1} и что при ограничениях, наложенных на параметры адаптации α_1 и α_2 , ошибка прогноза является

линейной комбинацией текущего и прошлых значений стационарного ряда y_t :

$$e_1(t) = y_{t+1} - \sum_{k=0}^{\infty} \omega_k y_{t-k},$$

где $\{\omega\}$ — сходящийся ряд весов.

Это означает, что ошибки прогноза стационарны и их средний квадрат вполне определен.

В результате минимизации дисперсии ошибки прогноза на 1 шаг вперед $D_e(1)$ Тейл и Вейдж получили следующие результаты:

$$\alpha_1 = \frac{2h}{1+h}; \quad \alpha_2 = h; \quad \gamma = \frac{2h^2}{1+h};$$

$$h = \sqrt{\left(-\frac{1}{8}g^2 + \frac{1}{2}g\sqrt{1 + \frac{1}{16}g^2}\right)}; \quad g^2 = \frac{\sigma_y^2}{\sigma_e^2};$$

$$D_e(1) = \frac{1+h}{1-h} \sigma_e^2.$$

Для практического применения этого, однако, недостаточно, и нами в приложении № 1 выведено более общее выражение для дисперсии ошибки прогнозирования на $\tau \geq 1$ по той же схеме.

Грубая оценка соотношения дисперсий g^2 может быть получена из соотношений (1.13) по подсчитанным автоковариациям процесса y_t . Уточнение g^2 производится экспериментально методом проб на имеющемся отрезке ряда. Давая значения в окрестностях грубой оценки, находят g^2 , минимизирующее дисперсию ошибки $D_e(1)$.

Нерлов и Вейдж провели теоретический анализ чувствительности дисперсии ошибки прогноза $D_e(1)$ к ошибке в определении g^2 . Оказалось, что процентное изменение дисперсии $D_e(1)$ пропорционально квадрату относительной ошибки оценки g^2 с коэффициентом пропорциональности

$$h(1-h)^2(6+8h+3h^2)/[4(2-h^2)^3].$$

Для наиболее реальных, малых значений g^2 даже 50%-ная ошибка в оценке g^2 дает увеличение в $D_e(1)$ менее чем на 1,5%.

Наши эксперименты говорят о том, что несмотря на довольно жесткую структуру процесса Тейла—Вейджа его

можно использовать для аппроксимации некоторых реальных временных рядов. Сопоставительный анализ, проведенный в § 3 гл. 7, показывает, что процесс Тейла—Вейджа занимает особое место среди моделей стохастических рядов.

§ 9. ПРИМЕРЫ

Сравним свойства нескольких отрезков реальных временных рядов со свойствами стохастического процесса Тейла—Вейджа и в тех случаях, когда они будут приблизительно одинаковыми, построим предикторы и получим прогнозы и их доверительные границы.

Для исследования возьмем сильно колеблющиеся ряды курса акций фирмы ИБМ и цен¹ на золото, приведенные в приложении № 4. Разобьем первый ряд на 3 отдельных участка по 48 точек, второй ряд разделим на 6 участков по 60 точек и проанализируем автокорреляционную функцию для вторых разностей соответствующих данных (см. табл. 1.9).

Таблица 1.9

Автокорреляционные функции вторых разностей

Ряд	№ участка	Лег k					
		1	2	3	4	5	6
Курс акций фирмы ИБМ	1	—0,36	—0,05	—0,16	—0,01	0,16	—0,11
	2	—0,32	—0,45	0,31	0,03	—0,15	0,07
	3	—0,57	0,08	0,08	—0,19	0,04	0,16
Цена на золото	1	—0,59	0,01	0,14	—0,03	—0,10	0,04
	2	—0,47	—0,01	0,04	—0,15	0,12	—0,05
	3	—0,62	0,07	0,04	0,08	—0,11	0,04
	4	—0,44	—0,16	0,09	0,08	—0,13	—0,03
	5	—0,20	—0,40	0,08	—0,07	—0,06	0,30
	6	—0,61	0,08	0,15	—0,21	0,13	—0,01

Для рядов, описываемых процессом Тейла—Вейджа, автокорреляционные функции вторых разностей должны

¹ Используется ряд цен на золото на Лондонском рынке на начало рабочего дня с 1 апреля 1971 г. по 14 сентября 1972 г.

иметь следующие свойства:

$$-\frac{2}{3} < \rho_1 < 0;$$

$$\rho_1 : \rho_2 = (-4) : 1;$$

$$\rho_k = 0 \quad |k| > 2.$$

Поэтому при моделировании можно выдвинуть гипотезу что участок 3 ряда курса акций и участки 1, 3, 6 ряда цен на золото генерированы процессом Тейла—Вейджа.

Методом перебора для этих отрезков были найдены значения g^2 , минимизирующие средний квадрат ошибки прогнозирования на 1 шаг. Начальные условия для уравнений адаптации находились по обычной линейной регрессии. Затем каждый отрезок экстраполировался на $\tau = 10$.

Экстраполяция производилась по следующим моделям:

$$\text{ИБМ}_3 \quad \hat{x}_\tau(48) = 646,5 + 1,0\tau \quad (\text{рис. 1.7});$$

$$\text{золото}_1 \quad \hat{x}_\tau(60) = 40,2 - 0,07\tau \quad (\text{рис. 1.8});$$

$$\text{золото}_3 \quad \hat{x}_\tau(60) = 43,7 - 0,00\tau \quad (\text{рис. 1.9});$$

$$\text{золото}_6 \quad \hat{x}_\tau(60) = 66,9 - 0,06\tau \quad (\text{рис. 1.10}).$$

Доверительные границы рассчитывались по формуле, выведенной в приложении № 1. Результаты представлены на рис. 1.7—1.10.

Для отрезков ИБМ₃, золото₁, золото₃ прогнозы и доверительные уровни достаточно хорошо описывают будущее поведение процесса. При экстраполяции отрезка золото₆, в предположении что временной ряд цен генерируется процессом Тейла—Вейджа, получается слишком большой размах доверительных уровней, лишаящий прогноз ценности.

Отметим, что прогнозы по адаптивной модели могут подсчитываться формально по одной и той же схеме, каким бы ни был временной ряд. Однако качество прогноза зависит от динамики процесса. Поэтому при моделировании прежде всего делаются предположения относительно истинной структуры ряда, выбирается наиболее подходящая модель и в соответствии с этим рассчитываются доверительные границы полученного прогноза. Эти границы будут тем уже, чем лучше принятая гипотеза отражает реальные свойства ряда. Альтернативных гипотез может быть множество, в частности, стохастический нестационарный процесс Тейла—Вейджа является одной из них.

В завершение следует подчеркнуть, что конечная цель в данном случае состояла не столько в изучении выбранных рядов, сколько в исследовании самого метода. Отобранные ряды характеризуются большой изменчивостью, по-

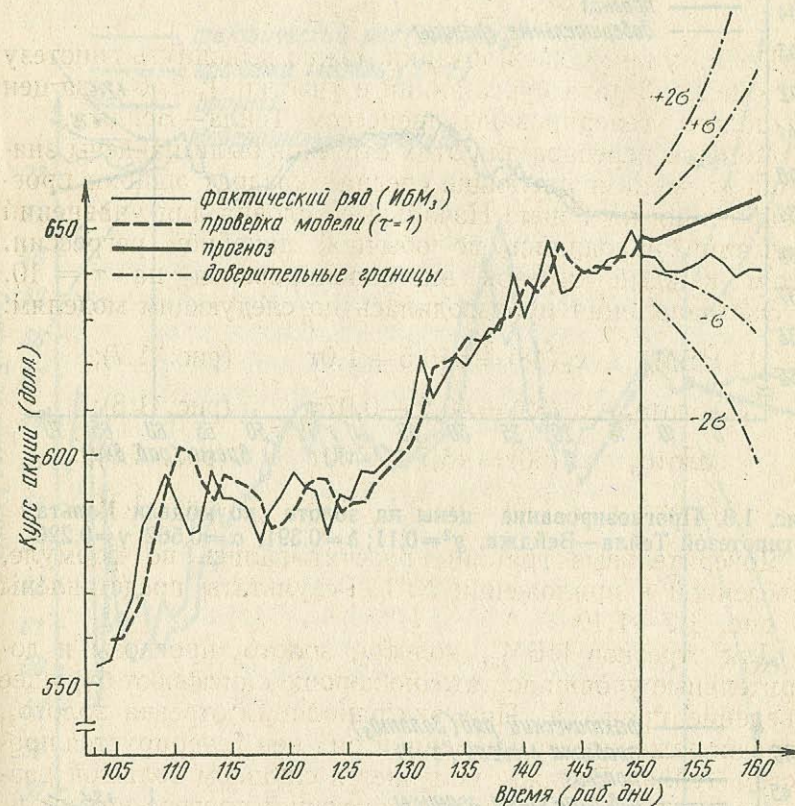


Рис. 1.7. Прогнозирование курса акций фирмы ИБМ по модели Хольта с гипотезой Тейла—Вейджа, $g^2=0,045$; $h=0,317$; $\alpha=0,482$; $\gamma=0,153$

этому их использование для испытания адаптивных свойств рассматриваемых моделей представляется оправданным. К этим рядам будем обращаться и в дальнейшем. Однако нельзя утверждать, что адаптивные методы являются наилучшим способом прогнозирования данных рядов.

Сделаем некоторые выводы. В гл. 1 рассмотрены простейшие модели экспоненциального сглаживания, линейные адаптивные модели, стохастический процесс Тейла и Вейд-

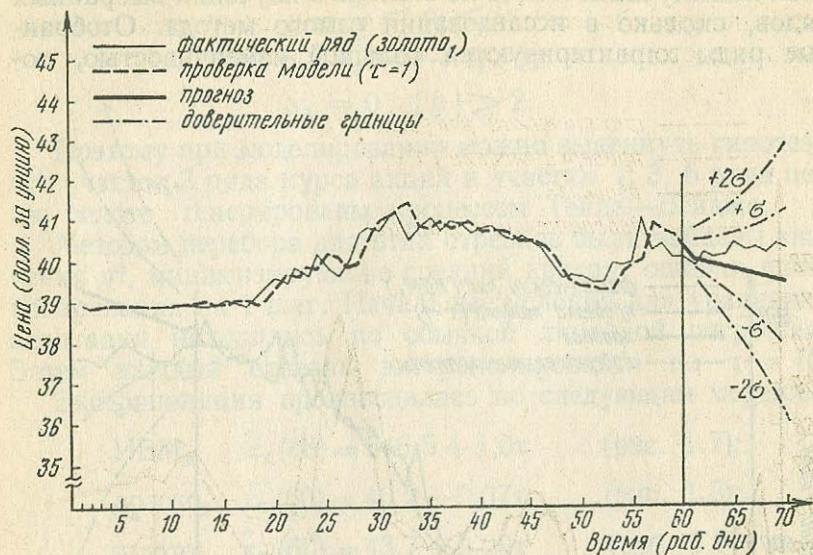


Рис. 1.8. Прогнозирование цены на золото по модели Хольта с гипотезой Тейла—Вейджа, $g^2=0,11$; $h=0,391$; $\alpha=0,562$; $\gamma=0,220$

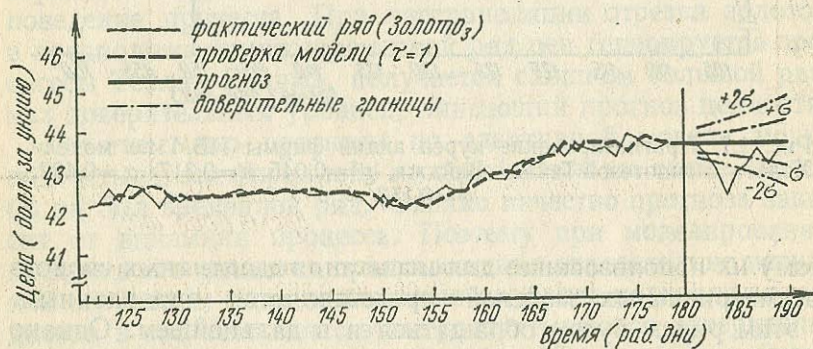


Рис. 1.9. Прогнозирование цены на золото по модели Хольта с гипотезой Тейла—Вейджа, $g^2=0,18$; $h=0,255$; $\alpha=0,406$; $\gamma=0,103$

жа. Несмотря на то что эти модели в дальнейшем получили эмпирическое или теоретическое развитие, они и без усовершенствования сохраняют практическое значение. Главное их достоинство — простота и возможность построения

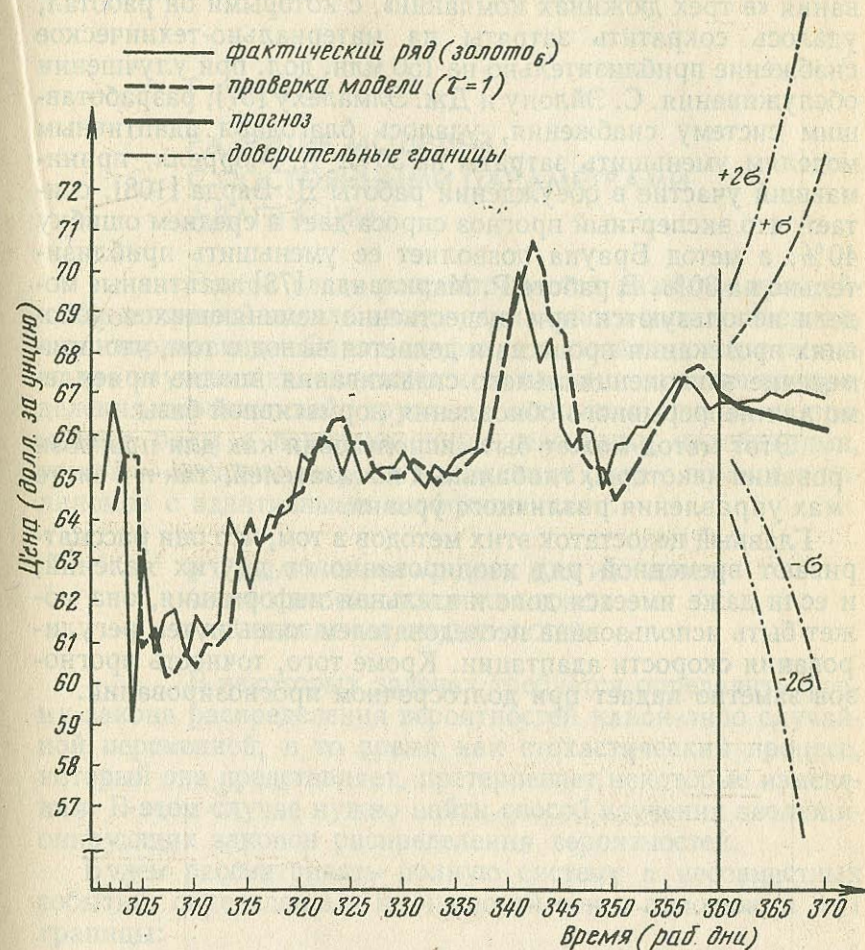


Рис. 1.10. Прогнозирование цены на золото по модели Хольта с гипотезой Тейла—Вейджа, $g^2=0,08$; $h=0,363$; $\alpha=0,533$; $\gamma=0,193$

и использования для прогнозирования при наличии небольшого количества фактических точек. Эти модели могут рассчитываться с помощью ЭВМ любого класса.

Методы экспоненциального сглаживания благодаря работам Хольта и Брауна стали эффективным и удобным ин-

струментом прогнозирования в экономике. В основном экспоненциальное сглаживание используется для прогнозирования рядов спроса и для управления ресурсами. Р. Браун [46] указывает, что благодаря этим методам прогнозирования «в трех дюжинах компаний», с которыми он работал, удалось сократить затраты на материально-техническое снабжение приблизительно на 150 млн. дол. при улучшении обслуживания. С. Эйлону и Дж. Элмалеху [57], разработавшим систему снабжения, удалось благодаря адаптивным моделям уменьшить затраты на 30%. А. Моррелл, принимавший участие в обсуждении работы Д. Варда [108], считает, что экспертный прогноз спроса дает в среднем ошибку 40%, а метод Брауна позволяет ее уменьшить приблизительно на 30%. В работе Р. Маркланда [73] адаптивные модели используются при существенно изменяющихся условиях протекания процесса и делается вывод о том, что применение экспоненциального сглаживания вполне приемлемо для непрерывного обновления нормативной базы.

Этот метод может быть использован как для прогнозирования некоторых глобальных показателей, так и в системах управления различного уровня.

Главный недостаток этих методов в том, что они рассматривают временной ряд изолированно от других явлений, и если даже имеется дополнительная информация, она может быть использована исследователем лишь путем регулирования скорости адаптации. Кроме того, точность прогнозов заметно падает при долгосрочном прогнозировании.

Глава 2

РАЗВИТИЕ МОДЕЛЕЙ С ПОСТОЯННЫМИ ПАРАМЕТРАМИ АДАПТАЦИИ

В этой главе рассматривается развитие простейшей адаптивной модели экспоненциального типа по нескольким направлениям. Одни модификации позволяют применять ее для изучения изменяющихся во времени законов распределения вероятностей, другие — адекватно описывать временные ряды с периодическими сезонными колебаниями, третьи — аппроксимировать тенденции ряда с помощью полиномов с адаптивными коэффициентами.

§ 1. АДАПТИВНАЯ МОДЕЛЬ ДЛЯ ИЗУЧЕНИЯ ЭВОЛЮЦИОНИРУЮЩИХ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

В некоторых задачах требуется определить форму закона распределения вероятностей какой-либо случайной переменной, в то время как стохастический процесс, который она представляет, претерпевает некоторые изменения. В этом случае нужно найти способ изучения эволюционирующих законов распределения вероятностей.

Будем рассматривать полную систему n несовместных событий, определенных на числовой оси с помощью $n + 1$ границы:

$$X_0 < X_1 < X_2 < \dots < X_n.$$

Если исследуются объемы заказов покупателей на какой-либо продукт, то могут быть, например, определены три события: 1) объем заказа менее 5 штук; 2) объем заказа не менее 5 штук, но менее 20; 3) объем заказа не менее 20 штук.

Независимо от того, в каком объеме продукт был заказан, мы можем рассматривать этот заказ как осуществле-

ние одного и только одного из указанных событий. Событие, связанное с наблюдением x_t , соответствует номеру интервала, в котором оказывается наблюдение. Любое возможное наблюдение должно либо быть равным одной из границ, либо оказаться между двумя соседними границами, т. е. имеется только одно k , такое, что $X_{k-1} < x_t \leq X_k$. Поэтому мы связываем с наблюдением x_t событие k . Это означает, что первая граница событий X_0 должна быть меньше, чем любое наблюдение, которое может иметь место, и последняя граница X_n должна быть больше, чем любое возможное наблюдение. Так как нельзя быть абсолютно уверенным, что наблюдаемые величины будут ограничены, то возможны две альтернативы конструирования системы. Одна состоит в том, чтобы положить $X_0 = -\infty$ и $X_n = +\infty$. Другой путь — положить X_0 столь малым, а X_n столь большим конечным числом, чтобы можно было ожидать, что реальные наблюдения будут находиться в этих пределах. Тогда для случая, когда появляется наблюденное значение, выходящее за установленные пределы, следует обеспечить выработку особого сигнала для вмешательства исследователя, который должен проанализировать неожиданные наблюдения, прежде чем обрабатывать их автоматически.

Рассмотрим простой метод оценки вероятностей $P_k(t)$ наступления различных событий

$$X_{k-1} < x_t \leq X_k,$$

предложенный Р. Г. Брауном [47]. Так как рассматривается полная система n несовместных событий, то степеней свободы будет $n - 1$, ибо сумма вероятностей должна быть равна единице, т. е. не все n значений P_k независимы.

Анализ прошлых данных или суждение о будущем делает возможным установить границы событий и сделать начальные, хотя бы грубые, оценки вероятностей различных событий $\hat{P}_k(0)$, $k = 1, 2, \dots, n$. Пусть наблюдение x_t в момент t свидетельствует о наступлении события k . Построим n -мерный вектор-столбец $U(t)$, который имеет $n - 1$ нулевую компоненту, а k -я компонента равна единице. Предыдущие оценки n вероятностей можно рассматривать как n -компонентный вектор-столбец $\hat{P}(t-1)$. Процесс пересмотра этих оценок с учетом текущей информации является экспоненциальным сглаживанием по правилу

$$\hat{P}(t) = \alpha U(t) + (1 - \alpha) \hat{P}(t-1), \quad 0 < \alpha < 1.$$

Каждая компонента вектора модифицируется простым экспоненциальным сглаживанием нуля или единицы. Например, если исследуется система из 3 событий и наблюдение означает наступление второго, то

$$\hat{P}_1(t) = \alpha \cdot 0 + (1 - \alpha) \hat{P}_1(t-1);$$

$$\hat{P}_2(t) = \alpha \cdot 1 + (1 - \alpha) \hat{P}_2(t-1);$$

$$\hat{P}_3(t) = \alpha \cdot 0 + (1 - \alpha) \hat{P}_3(t-1).$$

Так как $\hat{P}(t-1)$ является вектором вероятностей, то все его компоненты должны быть неотрицательны и их сумма должна быть точно равна 1.

Рассмотренный процесс векторного сглаживания не может сделать компоненту отрицательной, и сумма итоговых компонент та же, что сумма их прежних значений. Следовательно, если $\hat{P}(t-1)$ является вероятностным вектором, то им является и $\hat{P}(t)$.

Рассмотрим событие i . Если закон распределения наблюдаемых значений x_t не меняется, то математическое ожидание значения i -й компоненты вектора $U(t)$, подлежащего сглаживанию, точно равно действительной вероятности P_i наступления события i и математическое ожидание оценки равно действительной вероятности

$$M[\hat{P}_i(t)] = P_i.$$

Вероятность того, что придется сглаживать единицу, равна P_i , а вероятность того, что будем сглаживать нуль, равна $1 - P_i$. Легко подсчитать, что i -я компонента вектора $U(t)$ характеризуется дисперсией $P_i(1 - P_i)$.

Выражая дисперсию результата экспоненциального сглаживания через дисперсию на входе, получим дисперсию оценки вероятности наступления i -го события:

$$\sigma_i^2 = \frac{\alpha}{2 - \alpha} P_i(1 - P_i),$$

где α — постоянная сглаживания.

Таким образом, имеются два способа конструирования системы, которые наиболее предпочтительны. Границы событий целесообразно выбирать так, чтобы P_k была очень большой (близкой к 1) или очень маленькой (почти 0). Это обеспечит малую дисперсию оценок компонент вектора вероятностей.

Если форма распределения со временем меняется, то может быть использована большая постоянная сглаживания, для того чтобы быстро уменьшить влияние прежних данных. И наоборот, если вероятностное распределение постоянно во времени, то нет нужды уменьшать влияние старых данных, и для уменьшения дисперсии оценок можно использовать меньшую постоянную сглаживания.

Модель предназначена для изучения, например, спроса на отдельные виды товаров по возрастным группам, эволюции распределения семейных бюджетов по статьям расходов, распределения сумм текущих вкладов и других явлений экономической жизни.

§ 2. СЕЗОННЫЕ МОДЕЛИ

В экономике многие явления характеризуются периодически повторяющимися сезонными эффектами. Соответственно временные ряды, их отражающие, содержат периодические сезонные колебания. Эти ряды и их колебания можно представить как генерируемые моделями двух основных типов: моделями с мультипликативными и с аддитивными коэффициентами сезонности.

Модели первого типа имеют вид:

$$x_t = \xi_t + \varepsilon_t;$$

$$\xi_t = a_{1,t} f_t,$$

где динамика величины $a_{1,t}$ характеризует тенденцию развития процесса;

$f_t, f_{t-1}, \dots, f_{t-l+1}$ — коэффициенты сезонности;

l — количество фаз в полном сезонном цикле (если ряд представляет месячные наблюдения, то в экономике обычно $l = 12$, при квартальных данных $l = 4$ и т. п.);

ε_t — неавтокоррелированный шум с нулевым математическим ожиданием.

Модели второго типа записываются как:

$$x_t = \xi_t + \varepsilon_t;$$

$$\xi_t = a_{1,t} + g_t,$$

где величина $a_{1,t}$ описывает тенденцию развития процесса;

$g_t, g_{t-1}, \dots, g_{t-l+1}$ — аддитивные коэффициенты сезонности;

l — количество фаз в полном сезонном цикле;

ε_t — неавтокоррелированный шум с нулевым математическим ожиданием.

Адаптивная модель с мультипликативной сезонностью была предложена П. Р. Уинтерсом [113]. Аддитивная модель рассмотрена Г. Тейлом и С. Вейджем [103].

Уинтерс поставил задачу разработать модель для прогнозирования объемов сезонных продаж с использованием ЭВМ.

Модель должна быть такой, чтобы: а) прогнозы рассчитывались на основе одних и тех же программ для большого количества продуктов; б) вычисления производились быстро и дешево; в) использовался минимальный объем памяти для информации; г) учитывались изменяющиеся условия.

Прогнозы объемов продаж предназначаются для систем управления запасами и планирования производства. Такие системы предполагают наличие правил принятия решений, которые определяют, когда и сколько производить или заказывать отдельных видов товаров. Правила однообразно применяются ко многим продуктам, часто к десяткам тысяч или даже сотням тысяч изделий. Прогнозы должны делаться часто (ежемесячно или еженедельно). Метод прогнозирования должен быть четко формализован, что необходимо для автоматической обработки на ЭВМ. Необходимо иметь возможность легко вводить свежую информацию о фактических продажах.

Для прогнозирования продаж отдельных видов товаров могут быть применены несколько методов. Модель Уинтерса базируется на анализе изолированных временных рядов о продажах. Единственной используемой информацией является предыстория продаж данного товара. Модель Уинтерса является моделью экспоненциального типа. Эта схема, очевидно, имеет необходимые характеристики.

Для некоторых продуктов, характеризующихся стабильной интенсивностью продаж и малыми сезонными колебаниями, уже простая экспоненциальная модель является вполне удовлетворительной. Многие продукты, однако, имеют заметную тенденцию роста или падения продаж, особенно когда они производятся впервые или когда появляются конкурирующие товары. Для некоторых продуктов существенны сезонные изменения уровня продаж.

Поэтому целесообразно в прогностических моделях учитывать конкретный характер тенденции и сезонных колебаний. Это и сделал Уинтерс с помощью экспоненциальной схемы. Модель при этом становится сложнее, зато и точность

прогнозов для большинства товаров существенно возрастает.

Прежде чем переходить к полной модели Уинтерса, отражающей и сезонность, и линейную тенденцию роста, рассмотрим более простой вариант, который содержит только сезонный эффект.

Прогнозирование с коэффициентами сезонности

Модель имеет вид:

$$\hat{a}_{1,t} = \alpha_1 \frac{x_t}{\hat{f}_{t-l}} + (1 - \alpha_1) \hat{a}_{1,t-1} \quad 0 < \alpha_1 < 1;$$

$$\hat{f}_t = \alpha_2 \frac{x_t}{\hat{a}_{1,t}} + (1 - \alpha_2) \hat{f}_{t-l} \quad 0 < \alpha_2 < 1.$$

Как видим, $\hat{a}_{1,t}$ является взвешенной суммой текущей оценки $\frac{x_t}{\hat{f}_{t-l}}$, полученной путем очищения от сезонных колебаний фактических данных x_t и предыдущей оценки $\hat{a}_{1,t-1}$. В качестве коэффициента сезонности $\hat{f}_t$ берется его наиболее поздняя оценка, сделанная для аналогичной фазы цикла. Затем величина $\hat{a}_{1,t}$, полученная по первому уравнению, используется для определения новой оценки коэффициента сезонности по второму уравнению.

Прогноз следующего значения ряда:

$$\hat{x}_1(t) = \hat{a}_{1,t} \hat{f}_{t-l+1}.$$

Более общим выражением для прогноза на τ шагов вперед будет:

$$\hat{x}_\tau(t) = \hat{a}_{1,t} \hat{f}_{t-l+\tau}, \quad \tau \leq l.$$

Величины $\hat{a}_{1,t}$ и $\hat{f}_t$ могут быть записаны через прошлые данные и начальные условия:

$$\hat{a}_{1,t} = \alpha_1 \sum_{n=0}^t (1 - \alpha_1)^n \frac{x_{t-n}}{\hat{f}_{t-l-n}} + (1 - \alpha_1)^{t+1} \hat{a}_{1,0};$$

$$\hat{f}_t = \alpha_2 \sum_{n=0}^J (1 - \alpha_2)^n \frac{x_{t-nl}}{\hat{a}_{1,t-nl}} + (1 - \alpha_2)^{J+1} \hat{f}_{1,0}.$$

¹ Если прогнозы составляются на отрезок времени, больший чем l , то повторно используется l наиболее свежих оценок коэффициентов сезонности $\hat{f}_{t+1-l}, \dots, \hat{f}_t$.

где $\hat{a}_{1,0}$ — начальное значение a_1 ;

$\hat{f}_{1,0}$ — начальное значение f в соответствующей i фазе (месяце) цикла (года);

J — наибольшая целая часть $\frac{t}{l}$.

Следовательно, прогноз является функцией всех прошлых значений фактического ряда, параметров α_1 и α_2 и начальных условий $\hat{a}_{1,0}, \hat{f}_{1,0}, \hat{f}_{2,0}, \dots, \hat{f}_{l,0}$.

Влияние начальных условий на прогноз зависит от величины весов и длины ряда, предшествующего текущему моменту t . Влияние $\hat{a}_{1,0}$ обычно будет уменьшаться быстрее, чем влияние начальных значений $\hat{f}_{i,0}$, так как $\hat{a}_1$ пересматривается на каждом шаге, а $\hat{f}_i$ только один раз за цикл.

Если эта сезонная модель прогнозирования, структура которой не содержит элементов для отражения какой-либо тенденции роста, применяется для прогнозирования ряда, характеризующегося ярко выраженной тенденцией, то коэффициенты $\hat{f}_i$ перестают быть простыми коэффициентами сезонности и вскоре вбирают в себя в определенной мере эффект роста. Например, при обработке ряда ежемесячных наблюдений с длительной тенденцией к повышению будет получена сумма двенадцати $\hat{f}_i$, превышающая 12. Именно это и компенсирует отсутствие в структуре модели соответствующих элементов.

Если ряд имеет тенденцию, а многие ряды, по-видимому, ее имеют, то в модель необходимо ввести специфический член, учитывающий эту тенденцию.

Модель сезонных явлений с линейным ростом

Полная сезонная модель Уинтерса с линейным ростом аналогична только что рассмотренной:

$$\hat{a}_{1,t} = \alpha_1 \frac{x_t}{\hat{f}_{t-l}} + (1 - \alpha_1) (\hat{a}_{1,t-1} + \hat{a}_{2,t-1});$$

$$\hat{f}_t = \alpha_2 \frac{x_t}{\hat{a}_{1,t}} + (1 - \alpha_2) \hat{f}_{t-l};$$

$$\hat{a}_{2,t} = \alpha_3 (\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \alpha_3) \hat{a}_{2,t-1} \quad 0 < \alpha_1, \alpha_2, \alpha_3 < 1;$$

$$\hat{x}_\tau(t) = (\hat{a}_{1,t} + \tau \hat{a}_{2,t}) \hat{f}_{t-l+\tau}.$$

Единственным изменением в выражении для $\hat{a}_{1,t}$ является добавление $\hat{a}_{2,t-1}$ — наиболее поздней оценки аддитивного фактора роста, характеризующего изменение среднего за полный сезонный цикл уровня процесса за единицу времени (месяц). Выражение для обновления коэффициента сезонности остается тем же, что и раньше. Оценки $\hat{a}_{2,t}$ модифицируются по аналогичной процедуре экспоненциального сглаживания. Прогноз является здесь функцией прошлых и текущих данных, параметров $\alpha_1, \alpha_2, \alpha_3$ и первоначальных значений $\hat{a}_{1,0}, \hat{a}_{2,0}, \hat{f}_{i,0}$. Качество и точность прогнозов зависит от этих факторов.

Оптимальные параметры $\alpha_1, \alpha_2, \alpha_3$ Уинтерс предлагает находить экспериментальным путем. Критерием сравнения он берет стандартное отклонение ошибки. При этом предполагается, что прогноз не смещен. Поиск осуществлялся с помощью сетки значений $\alpha_1, \alpha_2, \alpha_3$. Функция стандартной ошибки вблизи минимума предполагалась достаточно плоской.

У Уинтерса и других исследователей это положение нашло экспериментальное подтверждение. В связи с этим предполагалось использовать один набор весов для широкого класса продуктов. Уинтерс использовал данные за 5—7 лет. При этом рядами были: данные о продажах кухонной утвари, о продажах краски, о котлованах для изготовленных заводским способом сооружений. Первая часть рядов (2—3 года) использовалась для построения модели, а на остальных данных проверялась точность прогнозирования.

Для получения представления о влиянии различных наборов величин ($\alpha_1, \alpha_2, \alpha_3$) на все три ряда сразу Уинтерс предложил следующую комбинированную оценку. В качестве меры потерь в точности прогнозирования для каждого ряда в отдельности при данном наборе ($\alpha_1, \alpha_2, \alpha_3$) он брал возрастание дисперсии ошибки, выраженное в процентах, по сравнению с дисперсией, соответствующей оптимальным параметрам сглаживания. Например, при (0,2; 0,4; 0,2) для первого ряда превышение над минимумом составило 2%, для второго — 2%, а для третьего — 22%. Комбинированная оценка равна их сумме — 26%. Наилучшей комбинированной оценкой у Уинтерса были 24% при (0,2; 0,4; 0,1).

Определение оптимальных параметров путем минимизации комбинированной оценки является попыткой найти

универсальные веса, подходящие для широкого класса рядов. Несмотря на то что полученный результат несет отпечаток специфических свойств использованных рядов, он является полезным ориентиром при работе с недостаточной информацией.

График ретроспективных прогнозов добычи газа в бывшем СССР, полученных по модели Уинтерса (рис. 2.1), свидетельствует об увеличении несоответствия прогнозов

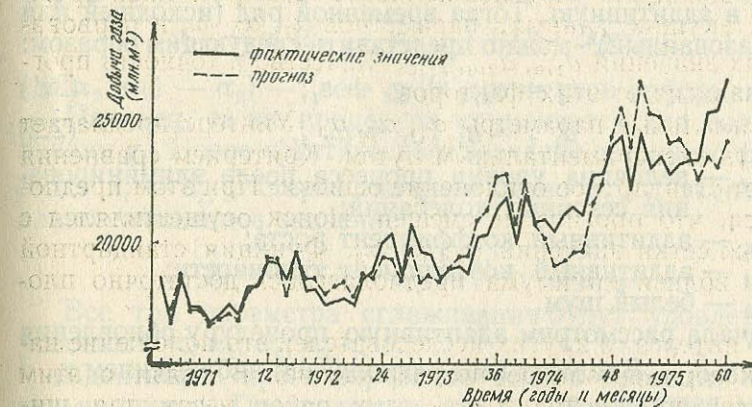


Рис. 2.1. Ретроспективное прогнозирование добычи газа по модели Уинтерса, $\tau = 1$

фактическим данным с ростом годового уровня добычи. Легко заметить более значительный размах сезонных колебаний прогнозов по сравнению с колебанием фактических данных при общем росте кривых. Все это свидетельствует о неадекватности мультипликативной модели реальному процессу. По-видимому, рост добычи газа в большей степени обеспечивается за счет разработки месторождений, продуктивность которых в меньшей мере подвержена сезонным колебаниям, чем это следует из гипотезы о мультипликативном характере сезонности.

Аддитивная модель сезонных явлений

Несмотря на то что для экономических временных рядов мультипликативная модель обычно оказывается наиболее подходящей, иногда требуется аддитивная модель. Рассмотрим аддитивную модель сезонных явлений с линейным ростом, предложенную Г. Тейлом и С. Вейджем [103].

Построение такой модели имеет целью упрощение процедуры прогнозирования, поскольку комбинация мультипликативной сезонной модели с линейным ростом математически громоздка. Кроме того, на практике чаще встречаются экспоненциальные тенденции, чем линейные. Поэтому замена значений первоначального временного ряда их логарифмами преобразует экспоненциальную тенденцию в линейную и одновременно мультипликативную сезонную модель в аддитивную. Тогда временной ряд (исходный или преобразованный)¹ можно представить следующим образом:

$$x_t = a_{1,t} + g_t + \varepsilon_t; \quad (2.1)$$

$$a_{1,t} = a_{1,t-1} + a_{2,t}, \quad (2.2)$$

где $a_{1,t}$ — величина уровня процесса после элиминирования сезонных колебаний;

$a_{2,t}$ — аддитивный коэффициент роста;

g_t — аддитивный коэффициент сезонности;

ε_t — белый шум.

Сначала рассмотрим адаптивную процедуру обновления значения $\hat{a}_{1,t}$. В момент t мы располагаем наблюдением x_t , о котором известно, что

$$x_t = a_{1,t} + g_t + \varepsilon_t.$$

Однако о шуме и сезонном факторе g_t никакой информации нет. Величину ε_t заменим нулем, а в качестве заменителя для g_t возьмем самую последнюю оценку сезонного фактора g_{t-l} , где l — период сезонного цикла.

Величину $x_t - g_{t-l}$ будем рассматривать как новое «фактическое» значение $a_{1,t}$. Последней оценкой уровня a_1 является $\hat{a}_{1,t-1}$, но она соответствует моменту $t-1$, а не t . Поэтому необходимо к $\hat{a}_{1,t-1}$ добавить еще $\hat{a}_{2,t}$ (см. 2.2). Но так как оценку $\hat{a}_{2,t}$ мы еще не можем получить, то вместо нее берем оценку $\hat{a}_{2,t-1}$, полученную на предыдущем шаге. Это приводит к следующей процедуре адаптации:

$$\hat{a}_{1,t} = \alpha_1 (x_t - \hat{g}_{t-l}) + (1 - \alpha_1) (\hat{a}_{1,t-1} + \hat{a}_{2,t-1}), \quad (2.3)$$

которая при данных весах α_1 и $(1 - \alpha_1)$ оценивает $a_{1,t}$ через наиболее свежее наблюдение x_t и ранее подсчитанные величины $\hat{a}_{1,t-1}$, $\hat{a}_{2,t-1}$, $\hat{g}_{t-l}$.

¹ Об особенностях прогнозирования временных рядов после логарифмического преобразования см. приложение № 3.

Учитывая (2.2) и вычитая из полученного $\hat{a}_{1,t}$ прежнюю оценку $\hat{a}_{1,t-1}$, можем получить оценку $\hat{a}_{2,t}$. Однако поскольку вычисления $\hat{a}_{1,t}$ не являются совершенными, в частности потому, что не принимались в расчет остатки из (2.1), то, очевидно, лучше не полагаться на эту разность полностью, а считать ее «фактическим» свидетельством динамики ряда и объединить со старым значением $\hat{a}_{2,t-1}$ по известной формуле экспоненциального сглаживания

$$\hat{a}_{2,t} = \alpha_2 (\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \alpha_2) \hat{a}_{2,t-1}, \quad (2.4)$$

где α_2 и $(1 - \alpha_2)$ — веса двух источников информации.

Наконец, та же процедура применяется для получения оценки g_t . Новое «фактическое» значение сезонного фактора будет $x_t - \hat{a}_{1,t}$, старое значение равно $\hat{g}_{t-l}$, экспоненциально-сглаженное значение

$$\hat{g}_t = \alpha_3 (x_t - \hat{a}_{1,t}) + (1 - \alpha_3) \hat{g}_{t-l}. \quad (2.5)$$

Все три параметра сглаживания будут удовлетворять условию $0 < \alpha_1, \alpha_2, \alpha_3 < 1$.

Адаптивное прогнозирование теперь провести сравнительно просто. Предположим, что t — текущий момент времени, так что $\hat{a}_{1,t}$, $\hat{a}_{2,t}$, $\hat{g}_t$, $\hat{g}_{t-1}$, ... имеются в нашем распоряжении. Предположим также, что мы хотим получить прогноз величины $x_{t+\tau}$ (прогноз на τ шагов вперед). Экстраполируем тенденцию линейного роста, используя самое последнее значение коэффициента $\hat{a}_{2,t}$, добавляем самую свежую оценку сезонного члена для этой фазы цикла и пренебрегаем шумом. В результате получаем

$$\hat{x}_\tau(t) = \hat{a}_{1,t} + \tau \hat{a}_{2,t} + \hat{g}_{t-l+\tau}$$

при условии, что $0 < \tau \leq l$. Если $l < \tau \leq 2l$, то необходимо $\hat{g}_{t-l+\tau}$ заменить на $\hat{g}_{t-2l+\tau}$ и т. д.

Модель готова. Однако на практике удобнее осуществлять адаптивное регулирование $\hat{a}_{1,t}$, $\hat{a}_{2,t}$ и $\hat{g}_t$ с помощью уравнений, связывающих эти величины с ошибкой прогноза, сделанного в конце периода $t-1$ на один шаг вперед.

Уравнение (2.3) можно переписать следующим образом:

$$\begin{aligned} \hat{a}_{1,t} - (\hat{a}_{1,t-1} + \hat{a}_{2,t-1}) &= \alpha_1 [x_t - (\hat{a}_{1,t-1} + \hat{a}_{2,t-1} + \hat{g}_{t-l})] = \\ &= \alpha_1 [x_t - \hat{x}_1(t-1)] = \alpha_1 e_1(t-1); \end{aligned} \quad (2.6)$$

из уравнений (2.3) — (2.5) легко получить:

$$\hat{a}_{2,t} - \hat{a}_{2,t-1} = \alpha_1 \alpha_2 e_1 (t-1); \quad (2.7)$$

$$\hat{g}_t - \hat{g}_{t-1} = (1 - \alpha_1) \alpha_3 e_1 (t-1). \quad (2.8)$$

Корректировки всех параметров модели совпадают по знаку и пропорциональны по величине ошибке последнего прогноза на один шаг вперед. Коэффициентами пропорциональности будут α_1 , $\alpha_1 \alpha_2$ и $(1 - \alpha_1) \alpha_3$ соответственно.

Можно заметить, что процедура (2.3) — (2.5) является рекурсивной в том смысле, что для определения $\hat{a}_{1,t}$ используются прошлые данные о сезонных колебаниях и коэффициенте роста. Оценка $\hat{a}_{1,t}$ используется для получения новых значений $\hat{a}_{2,t}$ и $\hat{g}_t$. Но совершенно очевидно, что прошлые значения сезонного фактора и коэффициента роста при вычислении $\hat{a}_{1,t}$ являются лишь суррогатом; не лучше ли было использовать более свежие, текущие оценки этих компонент? Рассмотрим такой вариант.

Вместо выражения (2.3) имеем:

$$\hat{a}_{1,t} = \alpha_1 (x_t - \hat{g}_t) + (1 - \alpha_1) (\hat{a}_{1,t-1} + \hat{a}_{2,t}). \quad (2.9)$$

Выражения (2.4) и (2.5) остаются прежними, и путем их подстановки в (2.9) получаем:

$$\begin{aligned} \hat{a}_{1,t} = & \alpha_1 [x_t - \alpha_3 (x_t - \hat{a}_{1,t}) - (1 - \alpha_3) \hat{g}_{t-1}] + \\ & + (1 - \alpha_1) [\hat{a}_{1,t-1} + \alpha_2 (\hat{a}_{1,t} - \hat{a}_{1,t-1}) + \\ & + (1 - \alpha_2) \hat{a}_{2,t-1}] = [\alpha_1 \alpha_3 + (1 - \alpha_1) \alpha_2] \hat{a}_{1,t} + \\ & + \alpha_1 (1 - \alpha_3) (x_t - \hat{g}_{t-1}) + \\ & + (1 - \alpha_1) (1 - \alpha_2) (\hat{a}_{1,t-1} + \hat{a}_{2,t-1}); \end{aligned}$$

отсюда:

$$\begin{aligned} \hat{a}_{1,t} = & \{ \alpha_1 (1 - \alpha_3) / [1 - \alpha_1 \alpha_3 - (1 - \alpha_1) \alpha_2] \} (x_t - \hat{g}_{t-1}) + \\ & + \{ (1 - \alpha_1) (1 - \alpha_2) / [1 - \alpha_1 \alpha_3 - (1 - \alpha_1) \alpha_2] \} (\hat{a}_{1,t-1} + \hat{a}_{2,t-1}) = \\ & = \alpha'_1 (x_t - \hat{g}_{t-1}) + (1 - \alpha'_1) (\hat{a}_{1,t-1} + \hat{a}_{2,t-1}). \end{aligned}$$

При этом $0 < \alpha'_1 < 1$. Общая форма выражения, как видим, та же, что и в (2.3). В выражениях (2.6), (2.7) и (2.8) лишь α_1 меняется на α'_1 , в остальном они остаются без изменений. Отметим, что $\alpha'_1 = \alpha_1$ при $\alpha_2 = \alpha_3$.

Альтернативы моделей

Вообще возможно множество комбинаций различных типов тенденций и циклических явлений аддитивного и мультипликативного вида. В работе [93] представлены девять возможных моделей, которые обобщенно выражены одной формулой. Поскольку всегда необходимо использовать модель, наиболее точно отражающую динамику процесса, то целесообразно остановиться на этом вопросе подробнее.

Девять упомянутых моделей составляют три группы по три варианта в каждой. Графическое отображение первой группы представлено на рис. 2.2, а. Ее образуют модель без тренда 1 — А, модель с аддитивным линейным трендом 1 — В и модель с мультипликативным (экспоненциальным) трендом 1 — С. Вторую группу (рис. 2.2, б) составляют три модели из первой группы с наложенным на них аддитивным сезонным эффектом. В третью группу (рис. 2.2, в) входят три модели из первой группы, но с наложенным на них мультипликативным сезонным эффектом.

Из девяти моделей большинство представляются практически полезными и лишь модель 3 — А, по-видимому, нереалистична. Графики на рис. 2.2 позволяют быстро выбрать в конкретном случае наиболее подходящую модель по имеющимся прошлым данным или на основе предположений о будущем поведении ряда.

Все девять моделей могут быть отражены в одной общей записи

$$\hat{a}_{1,t} = \alpha_1 d_1 + (1 - \alpha_1) d_2,$$

где $\hat{a}_{1,t}$ — текущий уровень ряда после элиминирования сезонных колебаний;

α_1 — параметр сглаживания, $0 < \alpha_1 < 1$;

значения d_1 и d_2 даны в табл. 2.1, каждая клетка которой характеризует ту или иную модель.

Например, модель В — 2 с аддитивным трендом и аддитивным сезонным эффектом записана в клетке, находящейся на пересечении строки В и колонки 2.

Рассмотрим величины, соответствующие в обобщенной формуле символам d_1 и d_2 :

x_t — фактическое наблюдение;

Значения символов обобщающей формулы

Характер сезонности		Значения символов обобщающей формулы			
		1	2	3	
Тенденция роста	Отсутствие тенденции роста	Отсутствие сезонного эффекта	Аддитивный сезонный эффект	Мультипликативный сезонный эффект	
A	Отсутствие тенденции роста	d_1	x_t	$x_t - \hat{g}_{t-1}$	$x_t / \hat{f}_{t-1}$
		d_2	$\hat{a}_{1,t-1}$	$\hat{a}_{1,t-1}$	$\hat{a}_{1,t-1}$
B	Аддитивный рост	d_1	x_t	$x_t - \hat{g}_{t-1}$	$x_t / \hat{f}_{t-1}$
		d_2	$\hat{a}_{1,t-1} + \hat{a}_{2,t-1}$	$\hat{a}_{1,t-1} + \hat{a}_{2,t-1}$	$\hat{a}_{1,t-1} + \hat{a}_{2,t-1}$
C	Экспоненциальный рост	d_1	x_t	$x_t - \hat{g}_{t-1}$	$x_t / \hat{f}_{t-1}$
		d_2	$\hat{a}_{1,t-1} \hat{r}_{t-1}$	$\hat{a}_{1,t-1} \hat{r}_{t-1}$	$\hat{a}_{1,t-1} \hat{r}_{t-1}$

$\hat{a}_{2,t}$ — коэффициент линейного роста, оценивается по формуле

$$\hat{a}_{2,t} = \alpha_2 (\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \alpha_2) \hat{a}_{2,t-1}, \quad 0 < \alpha_2 < 1;$$

$\hat{r}_t$ — коэффициент экспоненциального роста, его получают по формуле

$$\hat{r}_t = \alpha_r \frac{\hat{a}_{1,t}}{\hat{a}_{1,t-1}} + (1 - \alpha_r) \hat{r}_{t-1}, \quad 0 < \alpha_r < 1;$$

$\hat{g}_t$ — аддитивная сезонная компонента, вычисляется по формуле

$$\hat{g}_t = \alpha_g (x_t - \hat{a}_{1,t}) + (1 - \alpha_g) \hat{g}_{t-1}, \quad 0 < \alpha_g < 1;$$

l — периодичность сезонного явления;

$\hat{f}_t$ — мультипликативный коэффициент сезонности, подсчитывается по формуле

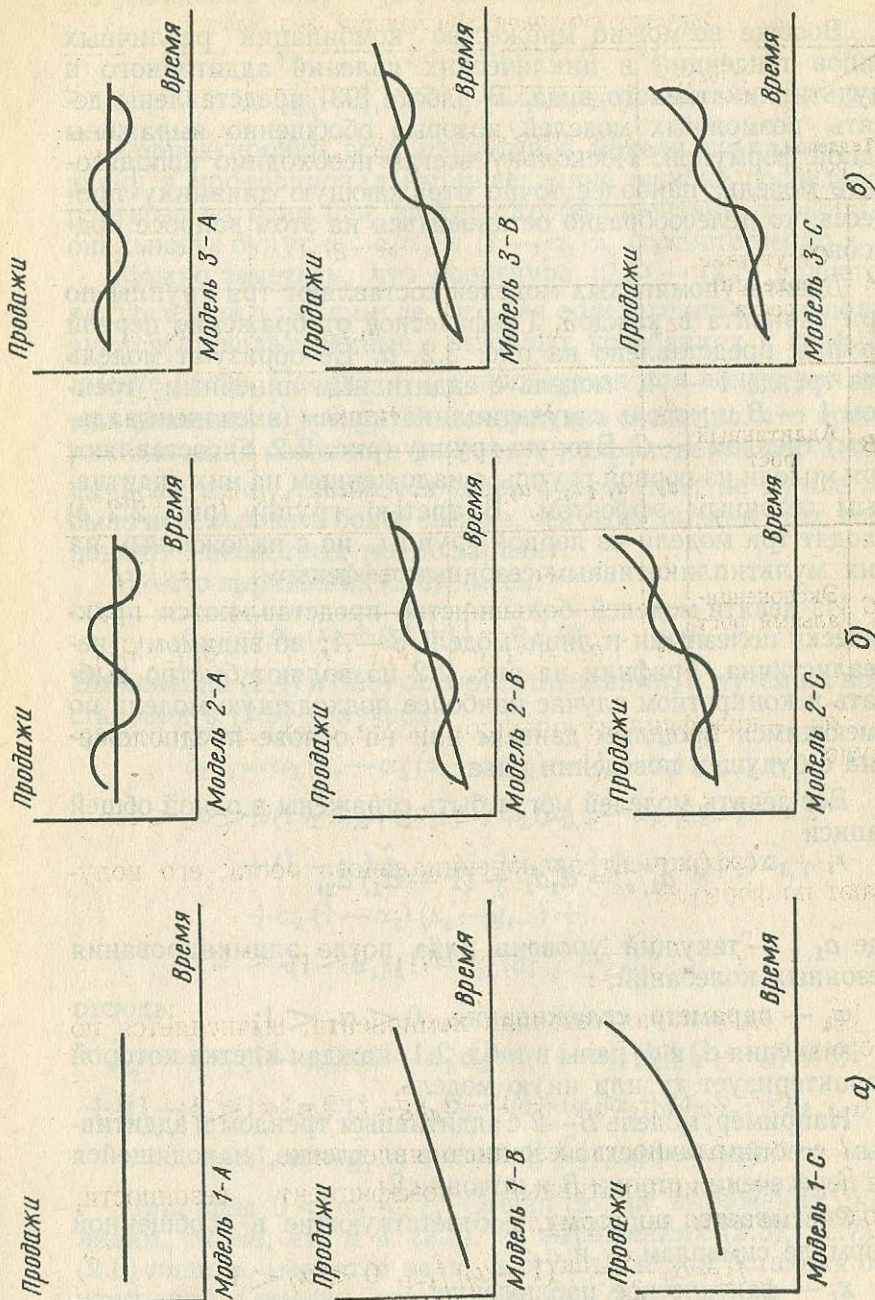
$$\hat{f}_t = \alpha_f \frac{x_t}{\hat{a}_{1,t}} + (1 - \alpha_f) \hat{f}_{t-1}, \quad 0 < \alpha_f < 1.$$

Рис. 2.2. Графическое отображение типов моделей:

а) модели без сезонного эффекта;

б) модели с мультипликативным сезонным эффектом;

в) модели с аддитивным сезонным эффектом.



Прогнозы фактических значений, ожидаемых в момент $t + \tau$, могут быть получены по общей формуле

$$\hat{x}_\tau(t) = P_\tau,$$

где P_τ подсчитывается для каждой модели способом, указанным в табл. 2.2

Таблица 2.2

Формулы для прогнозов

Характер сезонности Тенденция роста		1	2	3
		Отсутствие сезонного эффекта	Аддитивный сезонный эффект	Мультипликативный сезонный эффект
A	Отсутствие тенденции роста	$\hat{a}_{1,t}$	$\hat{a}_{1,t} + \hat{g}_{t-l+\tau}$	$\hat{a}_{1,t} \hat{f}_{t-l+\tau}$
B	Аддитивный рост	$\hat{a}_{1,t} + \hat{a}_{2,t}$	$\hat{a}_{1,t} + \tau \hat{a}_{2,t} + \hat{g}_{t-l+\tau}$	$(\hat{a}_{1,t} + \tau \hat{a}_{2,t}) \hat{f}_{t-l+\tau}$
C	Экспоненциальный рост	$\hat{a}_{1,t} \hat{r}_t^\tau$	$\hat{a}_{1,t} \hat{r}_t^\tau + \hat{g}_{t-l+\tau}$	$\hat{a}_{1,t} \hat{f}_{t-l+\tau} \hat{r}_t^\tau$

§ 3. АППРОКСИМАЦИЯ ПОЛИНОМИАЛЬНЫХ ТРЕНДОВ С ПОМОЩЬЮ МНОГОКРАТНОГО СГЛАЖИВАНИЯ

Многократное сглаживание

В главе 1 было рассмотрено экспоненциальное сглаживание для оценки текущего значения единственного коэффициента в прогнозной модели; мы познакомились с одно-, двух- и трехпараметрическими моделями линейного роста. Наиболее предпочтительной оказалась однопараметрическая модель Брауна, отличающаяся рядом достоинств. Р. Г. Брауном и Р. Ф. Майером [48] эта модель была развита для более сложных процессов путем включения в нее большего количества полиномиальных членов. Оценка коэффициентов полиномиальных членов производится путем многократного сглаживания. Здесь также предполагается,

что значения коэффициентов будут меняться медленно и оценки этих величин будут основываться на последних наблюдениях $x_t, x_{t-1}, \dots$

Понятие экспоненциальной средней S_t для временного ряда x_t можно обобщить на случай экспоненциальных средних более высоких порядков. Экспоненциальная средняя произвольного p -го порядка определяется как

$$S_t^{[p]} = \alpha S_t^{[p-1]} + \beta S_t^{[p]},$$

где $\beta = 1 - \alpha$; $p = 1, 2, \dots, n$; $S_t^{[0]} = x_t$;

$$S_0, S_0^{[2]}, \dots, S_0^{[n]} —$$

начальные значения экспоненциальных средних соответствующего порядка, т. е. выравнивание p -го порядка является простым экспоненциальным сглаживанием, примененным к результатам сглаживания $(p - 1)$ -го порядка.

Если в качестве гипотезы тренда некоторого процесса принимается полином степени n , то метод экспоненциального сглаживания и прогнозирования позволяет вычислить коэффициенты предсказывающего полинома через экспоненциальные средние соответствующих порядков.

В общем случае принимается гипотеза, что исследуемый процесс является параболой n -го порядка, а прогноз на τ шагов вперед выражается формулой

$$\hat{x}_\tau(\tau) = \hat{a}_1 + \hat{a}_2 \tau + \frac{1}{2!} \hat{a}_3 \tau^2 + \dots + \frac{1}{n!} \hat{a}_{n+1} \tau^n,$$

где параметры $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_{n+1}$ нужно определить.

Идея экспоненциального предсказания основана на том, что прогноз осуществляется полиномом, представляющим первые $n + 1$ членов разложения процесса x_t в ряд Тейлора:

$$x_{t+\tau} = \sum_{k=0}^{\infty} \frac{\tau^k x_t^{(k)}}{k!} = \sum_{k=0}^n \frac{\tau^k x_t^{(k)}}{k!} = \sum_{k=0}^n \frac{\tau^k a_{k+1}}{k!}.$$

Если наблюдаемый процесс может быть представлен полиномом степени n , то все производные порядка $n + 1$ и выше будут равны нулю.

Фундаментальная теорема метода экспоненциального сглаживания и прогнозирования, впервые доказанная Р. Брауном и Р. Майером [48], говорит о том, что коэффи-

циенты $a_1, a_2, \dots, a_{n+1}$ предсказывающего полинома связаны с экспоненциальными средними

$$S_t^{[p]}, \quad p = 1, 2, \dots, n+1$$

соотношением

$$S_t^{[p]} = \sum_{k=0}^n (-1)^k \frac{x_t^{(k)}}{k!} \frac{\alpha^p}{(p-1)!} \sum_{j=0}^{\infty} j^k \beta^j \frac{(p-1+j)!}{j!};$$

$$\beta = 1 - \alpha.$$

Таким образом, имеется $n+1$ уравнение, дающее сглаженные значения $S_t^{[p]}$ через линейные комбинации производных $x_t^{[k]}$.

Линейные уравнения для сглаженных величин можно записать в компактной форме, используя матрицы.

Пусть S_t будет вектором $S_t = \begin{bmatrix} S_t \\ S_t^{[2]} \\ \vdots \\ S_t^{[p]} \end{bmatrix}$

и пусть b_t будет вектором $b_t = \begin{bmatrix} b_{1,t} \\ b_{2,t} \\ \vdots \\ b_{n+1,t} \end{bmatrix} =$

$$= \begin{bmatrix} \frac{1}{0!} x_t^{(0)} \\ \frac{1}{1!} x_t^{(1)} \\ \vdots \\ \frac{1}{n!} x_t^{(n)} \end{bmatrix}$$

коэффициентов в разложении Тейлора.

Тогда содержание фундаментальной теоремы может быть выражено так:

$$S_t = M b_t,$$

где M — матрица $(n+1) \cdot (n+1)$ с элементами, содержащими бесконечные суммы степеней постоянной сглаживания:

$$m_{ik} = \frac{\alpha^i}{(i-1)!} \sum_{j=0}^{\infty} j^k \beta^j \frac{(i-1+j)!}{j!}.$$

Система $n+1$ линейных уравнений может быть разрешена относительно коэффициентов предсказывающего полинома по формуле

$$b_t = M^{-1} S_t.$$

Для экстраполяции трендов обычно используются полиномы не выше второго порядка. Для этих случаев элементы обратной матрицы M^{-1} определяются аналитически как функции постоянной сглаживания α .

Д.А.Д'Эзопо [58] доказал, что для любой последовательности наблюдений полином P степени n , полученный с помощью многократного сглаживания, является решением, которое минимизирует взвешенную сумму квадратов ошибок:

$$\alpha \sum_{i=0}^{\infty} \beta^i (x_{t-i} - P_{t-i})^2.$$

Экспоненциальное сглаживание очень просто. Однако надо отметить, что для сглаженных величин $S_t^{[p]}$, $p = 1, 2, \dots, n+1$ должны быть установлены начальные значения. Простейший путь задания исходных условий — оценка коэффициентов полинома разложения.

Если имеются выборочные данные, то эти оценки могут быть получены методом наименьших квадратов, как это делается в регрессионном анализе. Для новых временных рядов коэффициенты должны быть предсказаны или на основе аналогии с другими подобными рядами, или исходя из ожидаемых характеристик моделируемого процесса. В том и другом случае должны быть некие начальные оценки этих коэффициентов.

При выборе порядка полинома обычно решается вопрос о том, окупается ли сложность расчетов по выбранной кривой повышением точности прогнозирования. На практике используются полиномы не выше второго порядка. Приведем формулы расчета по этим моделям.

Во всех рассматриваемых ниже случаях предполагается, что тренд может быть представлен полиномом и наблюдается при наличии шума, т. е.

$$x_t = \xi_t + e_t;$$

$$\xi_t = \sum_{i=0}^n \frac{a_{i+1}}{i!} t^i,$$

где e_t — белый шум.

Ожидаемое значение x_t в момент $T + \tau$ будем обозначать $\hat{x}_\tau(T)$, где τ — время упреждения, на которое делается прогноз; T — текущий момент времени.

Адаптивная полиномиальная модель нулевого порядка: $n = 0$.

Гипотеза: $\xi_t = a_1$.

Экспоненциальная средняя:

$$S_t = \alpha x_t + \beta S_{t-1}, \quad \beta = 1 - \alpha.$$

Начальное условие:

$$S_0 = \hat{a}_{1,0}.$$

Оценка коэффициента:

$$\hat{a}_{1,T} = S_T.$$

Прогноз:

$$\hat{x}_\tau(T) = \hat{a}_{1,T} = S_T.$$

Адаптивная полиномиальная модель первого порядка: $n = 1$.

Гипотеза: $\xi_t = a_1 + a_2 t$.

Экспоненциальные средние:

$$S_t = \alpha x_t + \beta S_{t-1};$$

$$S_t^{[2]} = \alpha S_t + \beta S_{t-1}^{[2]}.$$

Начальные условия:

$$S_0 = \hat{a}_{1,0} - \frac{\beta}{\alpha} \hat{a}_{2,0};$$

$$S_0^{[2]} = \hat{a}_{1,0} - \frac{2\beta}{\alpha} \hat{a}_{2,0}.$$

Оценки коэффициентов:

$$\hat{a}_{1,T} = 2S_T - S_T^{[2]};$$

$$\hat{a}_{2,T} = \frac{\alpha}{\beta} (S_T - S_T^{[2]}).$$

Прогноз:

$$\hat{x}_\tau(T) = \hat{a}_{1,T} + \tau \hat{a}_{2,T} = \left(2 + \frac{\alpha}{\beta} \tau\right) S_T - \left(1 + \frac{\alpha}{\beta} \tau\right) S_T^{[2]}.$$

Адаптивная полиномиальная модель второго порядка: $n = 2$.

Гипотеза: $\xi_t = a_1 + a_2 t + \frac{1}{2} a_3 t^2$.

Экспоненциальные средние:

$$S_t = \alpha x_t + \beta S_{t-1};$$

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

$$S_t^{[3]} = \alpha S_t^{[2]} + \beta S_{t-1}^{[3]}.$$

Начальные условия:

$$S_0 = \hat{a}_{1,0} - \frac{\beta}{\alpha} \hat{a}_{2,0} + \frac{\beta(2-\alpha)}{2\alpha^2} \hat{a}_{3,0};$$

$$S_0^{[2]} = \hat{a}_{1,0} - \frac{2\beta}{\alpha} \hat{a}_{2,0} + \frac{\beta(3-2\alpha)}{\alpha^2} \hat{a}_{3,0};$$

$$S_0^{[3]} = \hat{a}_{1,0} - \frac{3\beta}{\alpha} \hat{a}_{2,0} + \frac{3\beta(4-3\alpha)}{2\alpha^2} \hat{a}_{3,0}.$$

Оценка коэффициентов:

$$\hat{a}_{1,T} = 3S_T - 3S_T^{[2]} + S_T^{[3]};$$

$$\hat{a}_{2,T} = \frac{\alpha}{2\beta^2} [(6-5\alpha) S_T - 2(5-4\alpha) S_T^{[2]} + (4-3\alpha) S_T^{[3]}];$$

$$\hat{a}_{3,T} = \frac{\alpha^2}{\beta^2} [S_T - 2S_T^{[2]} + S_T^{[3]}].$$

Прогноз:

$$\hat{x}_\tau(T) = \hat{a}_{1,T} + \tau \hat{a}_{2,T} + \frac{1}{2} \tau^2 \hat{a}_{3,T} =$$

$$= [6\beta^2 + (6-5\alpha)\alpha\tau + \alpha^2\tau^2] \frac{S_T}{2\beta^2} -$$

$$- [6\beta^2 + 2(5-4\alpha)\alpha\tau + 2\alpha^2\tau^2] \frac{S_T^{[2]}}{2\beta^2} +$$

$$+ [2\beta^2 + (4-3\alpha)\alpha\tau + \alpha^2\tau^2] \frac{S_T^{[3]}}{2\beta^2}.$$

Итак, прогнозы являются линейными комбинациями выравненных величин. Целью выравнивания данных является получение рекурсивных процедур пересмотра первоначальных оценок коэффициентов полинома.

Многократное экспоненциальное сглаживание является линейной дискретной системой с постоянными коэффициентами и может быть поэтому полностью описано своей реакцией на импульс. Изучив реакцию на импульс, легко получить реакции на другие стандартные входные потоки.

Раньше мы рассмотрели реакцию на импульс системы однократного экспоненциального сглаживания и использовали ее для изучения динамических свойств модели. Это было полезно для практического выбора компромиссного значения константы сглаживания α .

Можно провести подобный же анализ для системы многократного сглаживания. Для этого вводят понятие эквивалентной постоянной сглаживания, которое позволяет выводы, сделанные для однократного экспоненциального сглаживания, распространить на случай сглаживания любого порядка.

Предположим, что достаточно долго входной поток $x = 0$ и экспоненциальные средние всех порядков также равны нулю. Затем в момент времени $t = 0$ на вход подается единичный скачок $x_0 = 1$. Это может быть импульс, ступенчатое изменение или начало какой-либо другой временной функции. Экспоненциальная средняя первого порядка примет значение α . Реакция экспоненциальной средней второго порядка будет α^2 . Реакция экспоненциальной средней k -го порядка будет α^k . На основании фундаментальной теоремы оценка свободного члена $\hat{a}_{1,t}$ для полиномиальной модели n -порядка тогда будет выражена так:

$$\hat{a}_{1,t} = 1 - (1 - \alpha)^p, \quad p = n + 1.$$

Для анализа моделей, использующих экспоненциальные средние высокого порядка, удобно определить эквивалентную постоянную сглаживания таким образом, чтобы оценки свободного члена для различных моделей были одинаковыми.

Пусть при однократном выравнивании использовалась постоянная сглаживания α_1 . Тогда для модели линейного роста, использующей двойное сглаживание, значение эквивалентной постоянной α_2 будет такое, что

$$(1 - \alpha_2)^2 = 1 - \alpha_1 \quad \text{или} \quad \beta_2^2 = \beta_1.$$

Вообще для выравнивания порядка p эквивалентная постоянная сглаживания будет такая, что

$$(1 - \alpha_p)^p = 1 - \alpha_1 \quad \text{или} \quad \beta_p^p = \beta_1.$$

Это определение эквивалентной постоянной связано также с другой эквивалентностью. Для любой полиномиальной модели свободный член $a_{1,t}$ может быть оценен рекурсивно:

$$\hat{a}_{1,t} = x_t + \beta^p [\hat{x}_1(t-1) - x_t],$$

где $\hat{x}_1(t-1)$ — ожидаемое значение наблюдения в момент t , полученное на основе информации, имевшейся к моменту $t-1$.

Следовательно, эквивалентная постоянная сглаживания дает одинаковую коррекцию свободного члена при одинаковой ошибке прогноза на 1 шаг вперед.

Реакция адаптивных полиномиальных моделей нулевого, первого и второго порядков. На рис. 2.3, 2.4, 2.5, 2.6 адаптивные полиномиальные модели нулевого, первого и второго порядков с эквивалентными постоянными сглаживания сравниваются по их реакции на импульс, на ступенчатое изменение уровня, на входной поток с линейным ростом, на параболу при периоде упреждения $\tau = 3$. Наиболее сильно на импульс реагирует полиномиальная модель второго порядка (см. рис. 2.3).

На рис. 2.4 видно, что прогноз на три шага вперед в конце концов стремится к уровню ступенчатого изменения, но более сложные модели требуют на это больше времени. Реакция может быть ускорена увеличением постоянной сглаживания.

Стоит отметить, что прогнозы по полиномиальной модели первого порядка превышают уровень ступенчатого скачка и стремятся к нему сверху, в то время как прогнозы по квадратической модели превышают новый уровень входного потока, затем опускаются ниже и стремятся к нему снизу.

Рис. 2.5 показывает, что после первоначального переходного периода прогнозы полиномиальной модели нулевого порядка отстают от линейно растущего входного потока на постоянную величину; прогнозы полиномиальной модели первого порядка в конце концов приближаются к действительным значениям ряда; прогнозы полиномиальной моде-

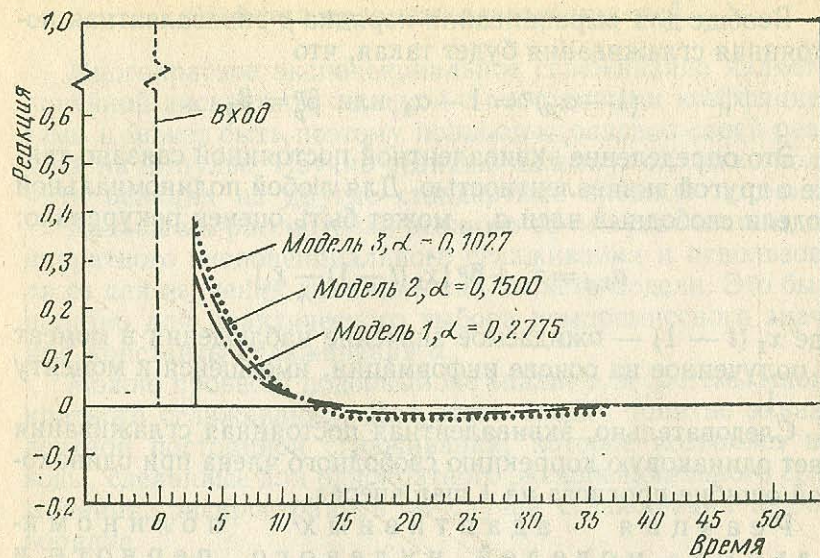


Рис. 2.3. Реакции на импульс прогнозных моделей экспоненциального типа, $\tau=3$:

модель 1 — полиномиальная модель нулевого порядка.
модель 2 — полиномиальная модель первого порядка.
модель 3 — полиномиальная модель второго порядка.

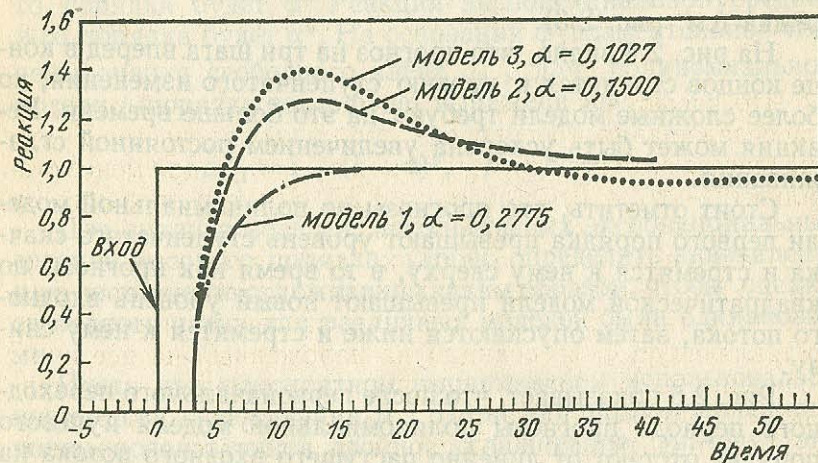


Рис. 2.4. Реакции на ступенчатое изменение, $\tau=3$

ли второго порядка превышают значения входного ряда, а затем приближаются к ним.

При прогнозировании параболы (рис. 2.6) первая модель все больше отстает от ее фактических значений; второй модели присуще постоянное смещение. Все более точные прогнозы дает квадратическая модель.

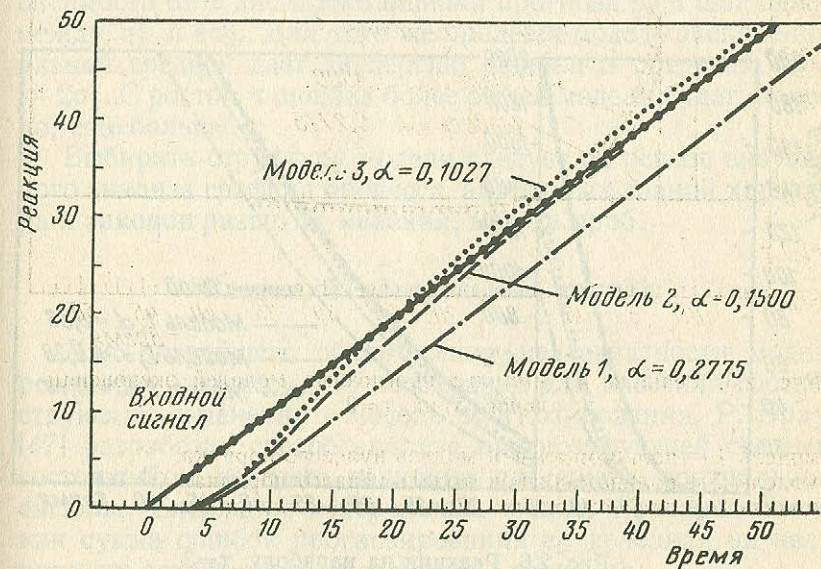


Рис. 2.5. Реакции на линейно-растающий входной поток, $\tau=3$

Что касается конкретных численных значений постоянной сглаживания, то для очень стабильных процессов Браун предлагает выбирать $\alpha_1 = 0,05$ или $\alpha_1 = 0,1$; для менее стабильных — $\alpha_1 = 0,1$ или $\alpha_1 = 0,25$. Эти три значения он и использует для моделирования.

Дисперсия оценок параметров полинома. Выход рассматриваемых моделей является суммой двух типов реакций: реакции на детерминированную составляющую ξ_t и реакции на наложенный на нее шум. Если тип модели соответствует порядку полинома, представляющего детерминированную составляющую, то для определения дисперсии оценок параметров достаточно рассмотреть случай, когда на входе только шум $x_t = \varepsilon_t$ с нулевым математическим ожиданием и дисперсией σ_ε^2 и автокорреляция отсутствует.

Такой анализ приводит к следующим результатам. При прогнозировании процесса $x_t = a_1 + e_t$ по полиномиальной модели нулевого порядка дисперсия оценки единственного параметра будет равна:

$$\sigma_{a_1}^2 = \frac{\alpha}{2-\alpha} \sigma_e^2.$$

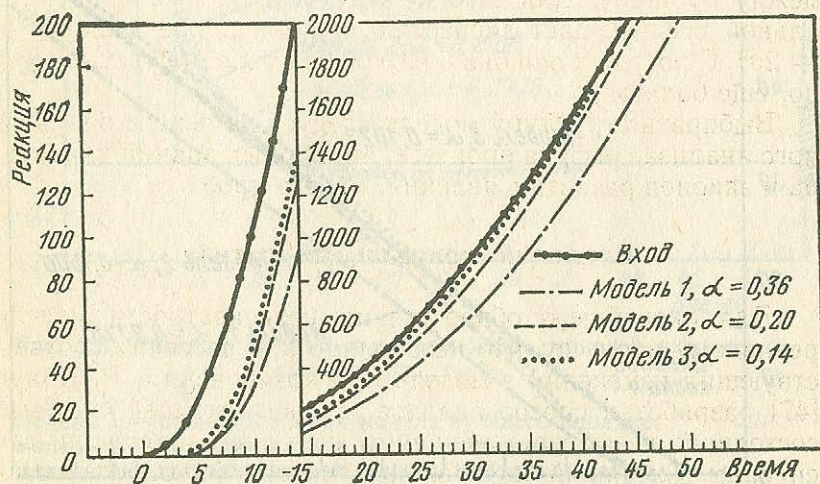


Рис. 2.6. Реакции на параболу, $\tau=3$

При прогнозировании процесса $x_t = a_1 + a_2 t + e_t$ по соответствующей модели получаем:

$$\sigma_{a_1}^2 = \frac{\alpha(1+4\beta+5\beta^2)}{(1+\beta)^3} \sigma_e^2; \quad (2.10)$$

$$\sigma_{a_2}^2 = \frac{2\alpha^3}{(1+\beta)^3} \sigma_e^2. \quad (2.11)$$

Для квадратической модели получить такие соотношения затруднительно. Поэтому оставим этот вопрос до рассмотрения обобщенной модели Брауна.

Выбор порядка полинома. Выбор порядка полинома — важная проблема, решение которой не всегда очевидно. Встает вопрос: не предпочтительнее ли брать «на всякий случай» более сложную модель и полагаться на ее свойства адаптировать свои параметры? В случае ошибки можно ожидать, что оценки соответствующих параметров

полинома будут стремиться к нулю. Оказывается, что поступать так довольно опрометчиво. Дж.Д. Кохен [52] рассмотрел случай, когда для стационарного процесса с постоянным уровнем и нулевой корреляцией для лагов, отличных от нуля, была ошибочно выбрана адаптивная полиномиальная модель первого порядка. В этом случае в зависимости от α дисперсия ошибки прогноза на 1 шаг лежит между σ_x^2 и $4\sigma_x^2$. Для того же процесса модель экспоненциальной средней дает дисперсию ошибки в пределах $\sigma_x^2 \div 2\sigma_x^2$. С ростом τ ошибка более общей модели будет, конечно, еще больше.

Выбирать структуру модели следует на основе визуального анализа графика процесса, априорных знаний характера и законов развития явления, метода проб.

Следящий контрольный сигнал

Для скорейшего обнаружения неадекватности модели реальному процессу, что необходимо для внесения соответствующих изменений в модель прогнозирования, Р. Браун [47] разработал способ анализа прогнозирующей системы, состоящий в подсчете величины следящего контрольного сигнала. Следящий контрольный сигнал K_t определяется как сумма ошибок прогнозирования e_t , деленная на величину их сглаженного абсолютного значения:

$$\tilde{e}_t = (1-\gamma)\tilde{e}_{t-1} + \gamma|e_t|,$$

т.е.

$$K_t = \frac{\sum_{i=0}^t e_i}{(1-\gamma)\tilde{e}_{t-1} + \gamma|e_t|},$$

где $0 < \gamma < 1$ — постоянная сглаживания.

Браун указывает значения критических уровней контрольного сигнала, превышение которых говорит о необходимости дополнительного изучения и изменения модели.

Такой прием имеет два недостатка. Во-первых, в случае когда контрольный сигнал вышел за установленные пределы, он не обязательно вернется в эти же пределы, даже если рассматриваемый процесс вновь будет развиваться по прежним законам и прогнозирующая модель окажется адекватной реальному процессу. Следовательно, потребуется вме-

шательство, чтобы сделать сумму ошибок снова близкой к нулю и избежать таким образом ложных сигналов тревоги.

Во-вторых, возможна и обратная ситуация, когда контрольный сигнал выходит из указанных пределов, а система начинает давать более точные прогнозы. Например, если с некоторого момента имеют место совершенные прогнозы, то среднее абсолютное отклонение будет стремиться к нулю, в то время как сумма ошибок остается неизменной. Таким образом, контрольный сигнал будет стремиться к бесконечности.

Д. Тригг [104] предложил простую модификацию правила Брауна, преодолевающую эти недостатки. Вместо суммы ошибок он использовал сглаженную ошибку $\hat{e}_t$:

$$\hat{e}_t = (1 - \gamma) \hat{e}_{t-1} + \gamma e_t.$$

Следящий контрольный сигнал K_t определяется отношением

$$K_t = \frac{\hat{e}_t}{\tilde{e}_t}.$$

Если прогнозирующая система окажется настолько неадекватной изучаемому процессу, что все ошибки будут одного знака, то контрольный сигнал будет стремиться к +1 или -1 (выйти за эти пределы он не может). Если известно, что прогнозирующая система адекватна реальному процессу и полученные ошибки образуют неавтокоррелированную, нормально распределенную случайную последовательность с нулевым средним и стандартным отклонением σ , то для контрольного сигнала могут быть определены доверительные интервалы.

Уравнение для сглаженной ошибки можно переписать в виде:

$$\hat{e}_t = \sum_{i=0}^t \gamma (1 - \gamma)^{t-i} e_{t-i},$$

ее дисперсия будет равна сумме дисперсий отдельных членов:

$$\sum_{i=0}^t [\gamma (1 - \gamma)^{t-i} \sigma]^2.$$

Так как $0 < (1 - \gamma) < 1$, то при $t \rightarrow \infty$ этот ряд сходится и его сумма равна:

$$\frac{\gamma^2 \sigma^2}{1 - (1 - \gamma)^2} = \frac{\gamma \sigma^2}{2 - \gamma}.$$

Пределы для сглаженной ошибки, определяемые величиной $2\sigma_{\hat{e}_t}$, равны, поэтому

$$\pm 2\sigma \sqrt{\frac{\gamma}{2 - \gamma}}.$$

Известно, что $\sigma \approx 1,2$ среднего абсолютного отклонения. Если γ достаточно мало, то можно принять, что локальная оценка среднего абсолютного отклонения относительно постоянна и приблизительно равна истинному среднему абсолютному отклонению, т. е. равна $\frac{\sigma}{1,2}$.

Таким образом, при малых γ , при принятых допущениях относительно e_t величина $\hat{e}_t$ является случайной, нормально распределенной, а $\tilde{e}_t$ приблизительно постоянной величиной. Это дает возможность (несмотря на то, что $-1 \leq K_t \leq +1$) аппроксимировать распределение $K_t = \frac{\hat{e}_t}{\tilde{e}_t}$ нормальным распределением.

Следовательно, пределы для контрольного сигнала, определяемые величиной $2\sigma_K$, приблизительно равны

$$\pm 2,4 \sqrt{\frac{\gamma}{2 - \gamma}}.$$

Для $\gamma = 0,1$ получим $\pm 0,55$, т. е. с вероятностью 95% $-0,55 \leq K_t \leq +0,55$.

Пределы в $3\sigma_K$ (соответствующие вероятности 99%) при том же значении γ составят $-0,83 \leq K_t \leq +0,83$.

Для значений γ , которые не очень малы, эти рассуждения теряют справедливость, и доверительные интервалы целесообразно получать путем моделирования методом Монте-Карло.

Таблицы вероятностей 2.3 и 2.4 для однократного и двойного экспоненциального сглаживания были получены М. Бэтти [39].

Таблица 2.3

Функция распределения контрольного сигнала
в случае однократного экспоненциального сглаживания
(полиномиальная модель нулевого порядка)

Вероятность $P(K < K_T)$	K_T				
	$\gamma=0,1$	$\gamma=0,2$	$\gamma=0,3$	$\gamma=0,4$	$\gamma=0,5$
0,70	0,24	0,33	0,44	0,53	0,64
0,80	0,29	0,40	0,52	0,62	0,73
0,85	0,32	0,45	0,57	0,67	0,77
0,90	0,35	0,50	0,63	0,72	0,82
0,95	0,42	0,58	0,71	0,80	0,88
0,96	0,43	0,60	0,73	0,82	0,89
0,97	0,45	0,62	0,76	0,84	0,90
0,98	0,48	0,66	0,79	0,87	0,92
0,99	0,53	0,71	0,82	0,92	0,94

Таблица 2.4

Функция распределения контрольного сигнала
в случае двойного экспоненциального сглаживания
(полиномиальная модель первого порядка)

Вероятность $P(K < K_T)$	K_T				
	$\gamma=0,1$	$\gamma=0,2$	$\gamma=0,3$	$\gamma=0,4$	$\gamma=0,5$
0,70	0,21	0,30	0,39	0,48	0,58
0,80	0,25	0,37	0,47	0,57	0,66
0,85	0,28	0,41	0,52	0,62	0,71
0,90	0,32	0,46	0,57	0,67	0,76
0,95	0,38	0,52	0,65	0,74	0,82
0,96	0,39	0,54	0,67	0,76	0,83
0,97	0,41	0,56	0,69	0,78	0,85
0,98	0,43	0,59	0,72	0,81	0,87
0,99	0,46	0,65	0,76	0,86	0,90

Итак, следящий контрольный сигнал является мерой неадекватности модели реальному процессу. При превышении контрольным сигналом заданного доверительного уровня (обычно определяемого величиной $2\sigma_K$) модель может быть пересмотрена и заменена другой.

§ 4. ОБОБЩЕННАЯ МОДЕЛЬ БРАУНА

Р. Браун [47] еще более расширил возможности прогнозирования с использованием адаптивных моделей. Он рассмотрел процесс

$$x_i = \xi_i + \varepsilon_i,$$

где

$$\xi_t = a_1 f_1(t) + a_2 f_2(t) + \dots + a_n f_n(t) = \sum_{i=1}^n a_i f_i(t);$$

ε_t — случайная неавтокоррелированная величина со средним значением 0 и дисперсией σ_ε^2 ;

$f_i(t)$ — некоторые известные, выбранные заранее детерминированные функции от времени;

a_i — коэффициенты, подлежащие определению и адаптации.

Рассматриваемый процесс может иметь слабые случайные изменения одного или более коэффициента. Р. Браун задался целью разработать адаптивную процедуру для пересмотра оценок коэффициентов при каждом получении новой фактической точки ряда. Ему удалось построить компактную итеративную модель для случая, когда функции, входящие в модель, удовлетворяют соотношению

$$f(t) = Lf(t-1),$$

где

$$f(t) = \begin{bmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_n(t) \end{bmatrix} \text{ — вектор-столбец из подбираемых функций;}$$

L — матрица постоянных коэффициентов размерностью $(n \cdot n)$, названная Р. Брауном матрицей перехода.

Такие функции являются решениями линейных разностных уравнений. Ими могут быть лишь полиномы, экспоненты и синусоиды или их произведения. Имея матрицу перехода для соответствующего набора функций, использованных в модели, необходимо также определить значения функций в какой-либо начальный момент времени, обычно при $t = 0$ или $t = 1$. По вектору начальных значений и матрице L можно получить значения $f(t)$ для любого момента времени

$$f(t) = L^t f(0).$$

Для упрощения расчетов Р. Браун за начало отсчета времени принимает текущий момент T , т. е. момент составления прогноза. За критерий ошибки, который следует ми-

минимизировать, взята взвешенная сумма квадратов отклонений:

$$\sum_{j=0}^T \beta^j [x_{T-j} - \hat{a}'_T f(-j)]^2,$$

где

$$\hat{a}' = [\hat{a}_1, \hat{a}_2, \dots, \hat{a}_n],$$

β — дисконтирующий фактор, $0 < \beta < 1$.

Далее для подбираемых функций вводится $(n \cdot n)$ матрица $F(t)$ с элементами:

$$F_{ik} = \sum_{j=0}^t \beta^j f_i(-j) f_k(-j).$$

$F(t)$ может быть подсчитана рекурсивно:

$$F(t) = \sum_{j=0}^t \beta^j f(-j) f'(-j) = F(t-1) + \beta^t f(-t) f'(-t).$$

Здесь проявляется одно полезное свойство принятого способа отсчета времени. Когда подбираемые функции являются тригонометрическими или полиномами и $0 < \beta < 1$, то β^t стремится к нулю быстрее, чем они могут расти, так что матрица $F(t)$ стремится к стабильному значению и ее обратная матрица¹, которая понадобится позже, при достаточно большом t рассчитывается один раз и уже не пересматривается.

Для определения оценок коэффициентов модели в соответствии с выбранным критерием Р. Браун вывел n -компонентный вектор

$$h = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix} = F^{-1} f(0),$$

где F — матрица функций в стабильном состоянии.

¹ Так как ни одна из подбираемых функций $f_i(t)$ не является линейной комбинацией остальных (желательно, чтобы они были ортогональными, т. е. чтобы $\sum_{t=1}^T f_i(t) f_k(t) = 0$ для всех пар $i \neq k$), то F будет иметь обратную матрицу F^{-1} .

С его помощью адаптивные коэффициенты a_i должны обновляться по формуле

$$\hat{a}_t = L' \hat{a}_{t-1} + h e_1(t-1), \quad (2.12)$$

где L' — транспонированная матрица L ; $e_1(t-1)$ — ошибка прогноза, рассчитанного в момент $t-1$ на 1 шаг вперед.

Константы h_i зависят только от частного набора подбираемых функций, от вектора начальных значений $f(0)$ и от величины β .

Из (2.12) видно, что даже при отсутствии ошибок прогнозирования коэффициенты будут изменяться по закону

$$\hat{a}_t = L' \hat{a}_{t-1}.$$

Это связано с переносом начала отсчета времени на каждом шаге на 1 интервал вперед. Второе слагаемое в (2.12) дает правило корректирования коэффициентов в зависимости от ошибки прогноза.

Исследуя дисперсию прогнозов, Р. Браун установил, что модели, состоящие из тригонометрических функций и постоянной составляющей, характеризуются примерно одной и той же дисперсией прогнозов для всех периодов прогнозирования τ . При грубой прикидке для этих моделей можно использовать соотношение

$$D(\hat{x}_\tau) \approx \frac{1-\beta^n}{2} \sigma_e^2.$$

Здесь так же, как в § 3, вводится понятие эквивалентной постоянной сглаживания, определяемой соотношением

$$\alpha_1 = 1 - \beta^n = 1 - \beta_1 \quad \text{или} \quad \beta^n = \beta_1,$$

где n — число искомых коэффициентов.

Для моделей, содержащих возрастающие от времени полиномиальные члены, дисперсия прогнозов зависит от времени упреждения τ . Для небольших значений эквивалентной постоянной сглаживания $\alpha_1 = 1 - \beta^n$ дисперсия прогноза по полиномиальной модели первого порядка пропорциональна периоду прогнозирования

$$D(\hat{x}_\tau) \approx (1,25\alpha_1 + \alpha_1^2 \tau) \sigma_e^2.$$

а для прогноза по полиномиальной модели второго порядка она квадратично возрастает от τ :

$$D(\hat{x}_\tau) \approx (2\alpha_1 + 3\alpha_1^2 \tau + 3\alpha_1^3 \tau^2) \sigma_\varepsilon^2.$$

Отметим, что в предельном случае, при $T \rightarrow \infty$, обобщенная модель Р. Брауна, содержащая только полиномиальные функции, в точности совпадает с рассмотренными выше моделями многократного сглаживания того же порядка.

Для десяти моделей Р. Брауном разработаны подробные таблицы, 5 из них (табл. 2.5—2.9) здесь приводятся.

Дисперсии в таблицах измерены в единицах σ_ε^2 . Отметим, что для получения дисперсии ошибки прогноза нужно к дисперсии прогноза добавить σ_ε^2 . Все данные приведены для трех специфических значений эквивалентной постоянной. Низкое значение $\beta_1 = 0,75$, т. е. $\beta^n = 0,75$, рекомендует использовать при необходимости быстрой адаптации оценок коэффициентов. Среднее значение $\beta_1 = 0,90$. Высокое значение дисконтирующего фактора β соответствует $\beta_1 = 0,95$ и обычно используется при несильном изменении коэффициентов «генерирующего» процесса.

Под кумулятивным прогнозом в таблицах понимается сумма прогнозов для периодов упреждения от 1 до τ , т. е.

$$\hat{X}_\tau(T) = \sum_{i=1}^{\tau} \hat{x}_i(T).$$

Начальный вектор коэффициентов $\hat{a}_0$ должен быть получен методом взвешенной множественной регрессии на основе достаточно большого числа данных о процессе, а выбор функций $f_i(t)$ осуществляется путем сопоставления свойств элементарных функций с особенностями реального процесса. Причем если подбираемые функции являются тригонометрическими, то для каждой гармоник следует включать как синус, так и косинус. Адаптация оценок коэффициентов осуществляется по уравнению (2.12).

Рассмотренный метод, строго говоря, правомерен только в установившемся состоянии. Заслуга Р. Брауна в том, что он первый показал, возможность построения адаптивных моделей, способных описывать периодические колебательные процессы.

В гл. 7, где изучаются процессы интегрированной авторегрессии — скользящего среднего, будет показано, что сейчас появился более обоснованный метод построения мо-

Таблица 2.5

Адаптивная полиномиальная модель нулевого порядка

Гипотеза: $\xi_t = a_1$.

Модель: $\hat{x}_\tau(t) = \hat{a}_{1,t}$.

Число коэффициентов: $n = 1$.

Матрица перехода L				Начальные условия			
				f (0)		f (1)	
1				f ₁	1	1	
Сглаживающий вектор h				Дисперсии оценок коэффициентов			
$\beta =$	0,75	0,90	0,95	$\beta =$	0,75	0,90	0,95
$1-\beta^n =$	0,25	0,10	0,05	$1-\beta^n =$	0,25	0,10	0,05
h_1	0,25000	0,10000	0,05000	$D(\hat{a}_1) =$	0,14286	0,05263	0,02564
Дисперсия прогноза				Среднеквадратическая ошибка кумулятивного прогноза			
период прогнози- рования	β			β			
	0,7500	0,9000	0,9500	0,7500	0,9000	0,9500	
1	0,1429	0,0526	0,0256	1,0000	1,0000	1,0000	
2	0,1429	0,0526	0,0256	1,6036	1,4868	1,4500	
3	0,1429	0,0526	0,0256	2,0702	1,8638	1,7974	
4	0,1429	0,0526	0,0256	2,5071	2,2005	2,1001	
5	0,1429	0,0526	0,0256	2,9277	2,5131	2,3751	
6	0,1429	0,0526	0,0256	3,3381	2,8098	2,6312	
7	0,1429	0,0526	0,0256	3,7417	3,0950	2,8734	
8	0,1429	0,0526	0,0256	4,1404	3,3717	3,1050	
9	0,1429	0,0526	0,0256	4,5356	3,6419	3,3282	
10	0,1429	0,0526	0,0256	4,9281	3,9068	3,5446	
11	0,1429	0,0526	0,0256	5,3184	4,1676	3,7554	
12	0,1429	0,0526	0,0256	5,7071	4,4248	3,9614	

Таблица 2.6

Адаптивная полиномиальная модель первого порядка

Гипотеза: $\xi_t = a_1 + a_2 t$.

Модель: $\hat{x}_\tau(t) = \hat{a}_{1,t} + \hat{a}_{2,t} \cdot \tau$.

Число коэффициентов: $n = 2$.

Матрица перехода L				Начальные условия			
1 0				f (0) f (1)			
1 1				f_1	1	1	
				f_2	0	1	
Сглаживающий вектор h				Дисперсии оценок коэффициентов			
$\beta =$	0,86603	0,94868	0,97468	$\beta =$	0,86603	0,94868	0,97468
$1 - \beta^n =$	0,25	0,10	0,05	$1 - \beta^n =$	0,25	0,10	0,05
h_1	0,25000	0,10000	0,05000	$D(\hat{a}_1)$	0,16937	0,06446	0,03173
h_2	0,01795	0,00263	$0,641 \times 10^{-3}$	$D(\hat{a}_2)$	$0,740 \times 10^{-3}$	$0,365 \times 10^{-4}$	$0,422 \times 10^{-5}$
Дисперсия прогноза				Среднеквадратическая ошибка кумулятивного прогноза			
период прогнози- рования	β			β			
	0,86603	0,94868	0,97468	0,86603	0,94868	0,97468	
1	0,1900	0,0672	0,0324	1,0000	1,0000	1,0000	
2	0,2121	0,0701	0,0331	1,5418	1,4609	1,4369	
3	0,2357	0,0730	0,0337	2,0402	1,8457	1,7875	
4	0,2607	0,0760	0,0344	2,5351	2,1970	2,0960	
5	0,2873	0,0791	0,0351	3,0390	2,5305	2,3794	
6	0,3153	0,0822	0,0358	3,5579	2,8540	2,6461	
7	0,3448	0,0854	0,0365	4,0950	3,1720	2,9011	
8	0,3758	0,0887	0,0372	4,6522	3,4872	3,1475	
9	0,4082	0,0921	0,0380	5,2309	3,8016	3,3876	
10	0,4422	0,0955	0,0387	5,8319	4,1166	3,6228	
11	0,4776	0,0990	0,0394	6,4558	4,4331	3,8543	
12	0,5145	0,1026	0,0402	7,1030	4,7519	4,0831	

Таблица 2.7

Адаптивная полиномиальная модель второго порядка

Гипотеза: $\xi_t = a_1 + a_2 t + \frac{1}{2} a_3 t(t-1)$.

Модель: $\hat{x}_\tau(t) = \hat{a}_{1,t} + \hat{a}_{2,t} \cdot \tau + \frac{1}{2} \hat{a}_{3,t} \cdot \tau(\tau-1)$.

Число коэффициентов: $n = 3$.

Матрица перехода L				Начальные условия			
				$f(0) \quad f(1)$			
1	0	0		f_1	1	1	
1	1	0		f_2	0	1	
0	1	1		f_3	0	0	
Сглаживающий вектор h				Дисперсии оценок коэффициентов			
β	0,90856	0,96549	0,98305	$\beta =$	0,90856	0,96549	0,98305
$1-\beta^n =$	0,25	0,10	0,05	$1-\beta^n =$	0,25	0,10	0,05
h_1	0,25000	0,10000	0,05000	$D(\hat{a}_1)$	0,18273	0,07038	0,03478
h_2	0,02394	0,00351	$0,855 \times 10^{-3}$	$D(\hat{a}_2)$	0,00154	$0,758 \times 10^{-4}$	$0,875 \times 10^{-5}$
h_3	$0,765 \times 10^{-3}$	$0,411 \times 10^{-4}$	$0,487 \times 10^{-5}$	$D(\hat{a}_3)$	$0,151 \times 10^{-5}$	$0,100 \times 10^{-7}$	$0,274 \times 10^{-9}$
Дисперсия прогноза				Среднеквадратическая ошибка кумулятивного прогноза			
период прогнози- рования	β			β			
	0,90856	0,96549	0,98305	0,90856	0,96549	0,98305	
1	0,2137	0,0745	0,0358	1,0000	1,0000	1,0000	
2	0,2489	0,0789	0,0368	1,5622	1,4666	1,4394	
3	0,2884	0,0835	0,0378	2,0921	1,8600	1,7937	
4	0,3326	0,0882	0,0389	2,6322	2,2229	2,1071	
5	0,3818	0,0932	0,0399	3,1963	2,5708	2,3964	
6	0,4364	0,0984	0,0410	3,7912	2,9117	2,6700	
7	0,4966	0,1039	0,0422	4,4215	3,2501	2,9328	
8	0,5628	0,1095	0,0433	5,0901	3,5889	3,1881	
9	0,6354	0,1154	0,0445	5,7995	3,9301	3,4380	
10	0,7147	0,1215	0,0457	6,5516	4,2752	3,6841	
11	0,8010	0,1279	0,0469	7,3483	4,6251	3,9275	
12	0,8948	0,1345	0,0482	8,1909	4,9808	4,1692	

Таблица 2.8

Адаптивная модель синусоидальных колебаний

$$\text{Гипотеза: } \xi_t = a_1 + a_2 \sin \frac{2\pi t}{12} + a_3 \cos \frac{2\pi t}{12}.$$

$$\text{Модель: } \hat{x}_\tau(t) = \hat{a}_{1,t} + \hat{a}_{2,t} \sin \frac{2\pi \tau}{12} + \hat{a}_{3,t} \cos \frac{2\pi \tau}{12}.$$

Число коэффициентов: $n = 3$.

Матрица перехода L				Начальные условия			
				$f(0)$		$f(1)$	
1	0	0		f_1	1	1,00000	
0	0,86603	0,50000		f_2	0	0,50000	
0	-0,50000	0,86603		f_3	1	0,86603	
Сглаживающий вектор h				Дисперсии оценок коэффициентов			
$\beta =$	0,90856	0,96549	0,98305	$\beta =$	0,90856	0,96549	0,98305
$1-\beta^n =$	0,25	0,10	0,05	$1-\beta^n =$	0,25	0,10	0,05
h_1	0,08593	0,03347	0,01668	$D(\hat{a}_1)$	0,04512	0,01740	0,00853
h_2	0,04360	0,00640	0,00156	$D(\hat{a}_2)$	0,09012	0,03475	0,01706
h_3	0,16408	0,06654	0,03333	$D(\hat{a}_3)$	0,09456	0,03511	0,01711
Дисперсия прогноза				Среднеквадратическая ошибка кумулятивного прогноза			
период прогнози- рования	β			β			
	0,90856	0,96549	0,98305	0,90856	0,96549	0,98305	
1	0,1505	0,0530	0,0257	1,0000	1,0000	1,0000	
2	0,1471	0,0527	0,0257	1,5617	1,4696	1,4416	
3	0,1395	0,0523	0,0256	1,9894	1,8316	1,7817	
4	0,1330	0,0520	0,0256	2,3529	2,1407	2,0705	
5	0,1309	0,0520	0,0256	2,6606	2,4101	2,3237	
6	0,1322	0,0521	0,0256	2,9212	2,6473	2,5495	
7	0,1333	0,0521	0,0256	3,1460	2,8591	2,7537	
8	0,1322	0,0520	0,0256	3,3480	3,0522	2,9418	
9	0,1309	0,0520	0,0256	3,5411	3,2339	3,1185	
10	0,1331	0,0522	0,0256	3,7388	3,4111	3,2883	
11	0,1396	0,0526	0,0257	3,9525	3,5894	3,4547	
12	0,1472	0,0529	0,0257	4,1885	3,7719	3,6199	

Таблица 2.9

Адаптивная модель синусоидальных колебаний с тенденцией линейного роста

$$\text{Гипотеза: } \xi_t = a_1 + a_2 t + a_3 \sin \frac{2\pi t}{12} + a_4 \cos \frac{2\pi t}{12}.$$

$$\text{Модель: } \hat{x}_\tau(t) = \hat{a}_{1,t} + \hat{a}_{2,t} \tau + \hat{a}_{3,t} \sin \frac{2\pi \tau}{12} + \hat{a}_{4,t} \cos \frac{2\pi \tau}{12}.$$

Число коэффициентов: $n = 4$.

Матрица перехода L				Начальные условия		
1	0	0	0		f (0)	f (1)
1	1	0	0	f ₁	1	1,00000
0	0	0,86603	0,50000	f ₂	0	1,00000
0	0	-0,50000	0,86603	f ₃	0	0,50000
				f ₄	1	0,86603

Сглаживающий вектор h				Дисперсии оценок коэффициентов			
$\beta =$	0,93061	0,97400	0,98726	$\beta =$	0,93061	0,97400	0,98726
$1 - \beta^n =$	0,25	0,10	0,05	$1 - \beta^n =$	0,25	0,10	0,05
h_1	0,12949	0,05024	0,02503	$D(\hat{a}_1)$	0,08510	0,03245	0,01593
h_2	0,00457	0,660 × 10 ⁻³	0,160 × 10 ⁻³	$D(\hat{a}_2)$	0,866 × 10 ⁻⁴	0,452 × 10 ⁻⁵	0,526 × 10 ⁻⁵
h_3	0,04113	0,00605	0,00148	$D(\hat{a}_3)$	0,06879	0,02614	0,01281
h_4	0,12052	0,04977	0,02499	$D(\hat{a}_4)$	0,07074	0,02633	0,01283

Дисперсия прогноза				Среднеквадратическая ошибка кумулятивного прогноза		
период прогнози- рования	β			β		
	0,93060	0,97400	0,98726	0,93060	0,97400	0,98726
1	0,1720	0,0602	0,0291	1,0000	1,0000	1,0000
2	0,1798	0,0613	0,0293	1,5188	1,4545	1,4342
3	0,1825	0,0621	0,0296	1,9656	1,8224	1,7769
4	0,1824	0,0626	0,0297	2,3690	2,1440	2,0715
5	0,1822	0,0629	0,0298	2,7361	2,4325	2,3336
6	0,1827	0,0630	0,0298	3,0732	2,6951	2,5713
7	0,1836	0,0631	0,0298	3,3884	2,9379	2,7903
8	0,1853	0,0632	0,0298	3,6923	3,1666	2,9953
9	0,1898	0,0637	0,0299	3,9968	3,3872	3,1905
10	0,1997	0,0646	0,0301	4,3135	3,6054	3,3796
11	0,2152	0,0660	0,0304	4,6518	3,8255	3,5656
12	0,2332	0,0677	0,0308	5,0166	4,0504	3,7503

Влияние величины периода упреждения τ
оптимальное значение параметра адаптации α и средний квадрат

№ экспериментов	Выборка из ряда	Число точек	$\tau = 1$		
			№ модели	α	Средний квадрат ошибки
1	2	3	4	5	6
1	Курс акций фирмы ИБМ, дол.	144	1	0,9600	41,427
2	Курс акций фирмы ИБМ (часть 1)	48	1	0,9450	32,956
3	Курс акций фирмы ИБМ (часть 2)	48	1	0,9992	46,294
4	Курс акций фирмы ИБМ (часть 3)	48	1	0,9120	43,419
5	Цена на золото, дол. за унцию	360	1	0,8410	0,50123
6	Цена на золото (часть 1)	60	1	0,7510	0,08265
7	Цена на золото (часть 2)	60	1	0,9991	0,10480
8	Цена на золото (часть 3)	60	2	0,2711	0,02884
9	Цена на золото (часть 4)	60	1	0,9709	0,14521
10	Цена на золото (часть 5)	60	1	0,9999	0,84637
11	Цена на золото (часть 6)	60	1	0,5709	1,6928
12	Поставка минеральных удобрений сельскому хозяйству, тыс. т усл. ед.	30	2	0,8010	$28,148 \times 10^5$
13	Производство электроэнергии, млрд. кВт·ч	31	2	0,8990	26,593
14	Производство гидроэлектроэнергии, млрд. кВт·ч	31	2	0,5121	27,456
15	Сумма вкладов населения в сберкассы, млн. руб.	26	2	0,9909	$10,920 \times 10^5$
16	Численность городского населения, млн. чел.	27	2	0,9191	0,32741
17	Протяженность нефте- и нефтепродуктопроводов, тыс. км	21	2	0,6029	1,4529
18	Грузооборот нефте- и нефтепродуктопроводов, ткм	21	3	0,7298	168,732
19	Урожайность зерна, ц с га	30	2	0,0191	2,2653
20	Урожайность картофеля, ц с га	30	1	0,3060	170,716
21	Урожайность овощей, ц с га	30	2	0,0668	78,732
22	Производство яиц, млрд. шт.	30	2	0,7291	1,9412

на выбор типа модели Брауна,
ошибки ретроспективных прогнозов

Таблица 2.10

$\tau = 2$			$\tau = 3$			$\tau = 4$			$\tau = 5$		
№ модели	α	Средний квадрат ошибки	№ модели	α	Средний квадрат ошибки	№ модели	α	Средний квадрат ошибки	№ модели	α	Средний квадрат ошибки
7	8	9	10	11	12	13	14	15	16	17	18
1	0,8801	80,153	1	0,9480	112,390	1	0,9109	149,867	1	0,9610	184,634
1	0,6200	60,628	1	0,5830	76,934	1	0,6779	89,011	1	0,8500	102,58
1	0,6740	89,277	1	0,8320	108,44	1	0,9210	138,34	1	0,9490	179,08
1	0,9900	80,414	2	0,1350	120,91	2	0,1080	147,54	2	0,0990	171,05
1	0,7440	0,86646	1	0,6990	1,17043	1	0,7160	1,44257	1	0,7701	1,71068
1	0,7900	0,12906	1	0,9200	0,17592	1	0,8720	0,25169	1	0,9898	0,31361
1	0,9999	0,21768	1	0,9995	0,34643	1	0,9518	0,47156	1	0,9990	0,57996
2	0,2801	0,03570	2	0,2980	0,04503	1	0,7998	0,05862	1	0,7998	0,07811
2	0,2201	0,26517	2	0,2020	0,34595	2	0,870	0,43443	2	0,1781	0,54055
2	0,2035	1,2007	2	0,2070	1,3898	2	0,2229	1,5265	2	0,2519	1,6260
1	0,5780	2,0442	1	0,7190	2,4167	1	0,2390	3,0749	1	0,1760	3,4759
2	0,6529	$96,117 \times 10^5$	2	0,6520	$184,265 \times 10^5$	2	0,6909	$313,713 \times 10^5$	2	0,7200	$539,805 \times 10^5$
2	0,8900	113,153	2	0,9390	293,255	2	0,9620	674,750	2	0,9890	1367,83
2	0,3900	59,887	2	0,3372	98,412	2	0,3248	133,206	2	0,3481	160,083
2	0,9916	$54,069 \times 10^5$	2	0,9990	$135,774 \times 10^5$	2	0,9990	$291,646 \times 10^5$	2	0,9990	$544,420 \times 10^5$
2	0,8001	1,46924	2	0,0708	0,22443	2	0,0622	2,58524	2	0,0589	2,88108
2	0,6699	3,3538	2	0,4900	6,5018	2	0,4020	8,0100	2	0,3602	9,4980
3	0,6192	829,823	3	0,5490	1723,930	3	0,4907	3002,625	3	0,3709	4870,605
2	0,0210	2,2758	2	0,0089	2,3475	2	0,0031	2,4393	2	0,0031	2,5337
1	0,3500	166,539	1	0,3700	190,810	1	0,3271	219,417	1	0,3780	238,950
2	0,0710	76,065	2	0,1000	68,214	2	0,0701	75,964	2	0,0770	78,838
2	0,5611	6,1755	2	0,5030	12,0070	2	0,0010	10,608	2	0,0010	11,0246

делей, частным случаем которого являются модели, рассмотренные в данном параграфе. Однако на практике метод Р. Брауна благодаря своей простоте может быть полезен на этапе грубых прикидок или для обработки большого количества сходных рядов, когда проведение полной процедуры идентификации и оценки параметров модели, предусмотренной в гл. 7, для каждого ряда невозможно. Но и в этих случаях метод Р. Брауна лучше использовать не в чистом виде, а с модификациями, рассмотренными в гл. 4 и 5.

§ 5. ПРИМЕРЫ

Пример 2.1

Проведен эксперимент, целью которого было выявление влияния величины периода упреждения τ на выбор порядка полиномиальной модели и величину оптимального α в моделях многократного сглаживания. Результаты представлены в табл. 2.10. В этой таблице номера моделей означают: № 1 — полиномиальная модель нулевого порядка, № 2 — полиномиальная модель первого порядка, № 3 — полиномиальная модель второго порядка.

Как видно из таблицы, в экспериментах № 4, 8, 9, 10 с увеличением τ произошла замена одной модели другой. В экспериментах № 11, 14, 16, 17, 18, 22 значение оптимального α с ростом τ заметно падает. Это означает, что с увеличением периода упреждения веса, приписываемые членам ряда, перераспределяются в пользу более ранних наблюдений. В остальных случаях величина $\alpha_{\text{опт}}$ остается приблизительно постоянной.

Пример 2.2

На рис. 2.7 изображены прогнозы курса акций ИБМ, полученные по полиномиальной модели Брауна второго порядка с $\alpha = 0,1$. Доверительные уровни прогнозов, нанесенные на график, определялись адаптивно методом экспоненциального сглаживания квадратов ошибок прогнозов с постоянной сглаживания $\alpha = 0,1$. Анализ графика показывает, что реальный процесс укладывается в полученные таким образом доверительные уровни. Однако для лучшего выравнивания квадратов ошибок целесообразно использовать $\alpha < 0,1$.

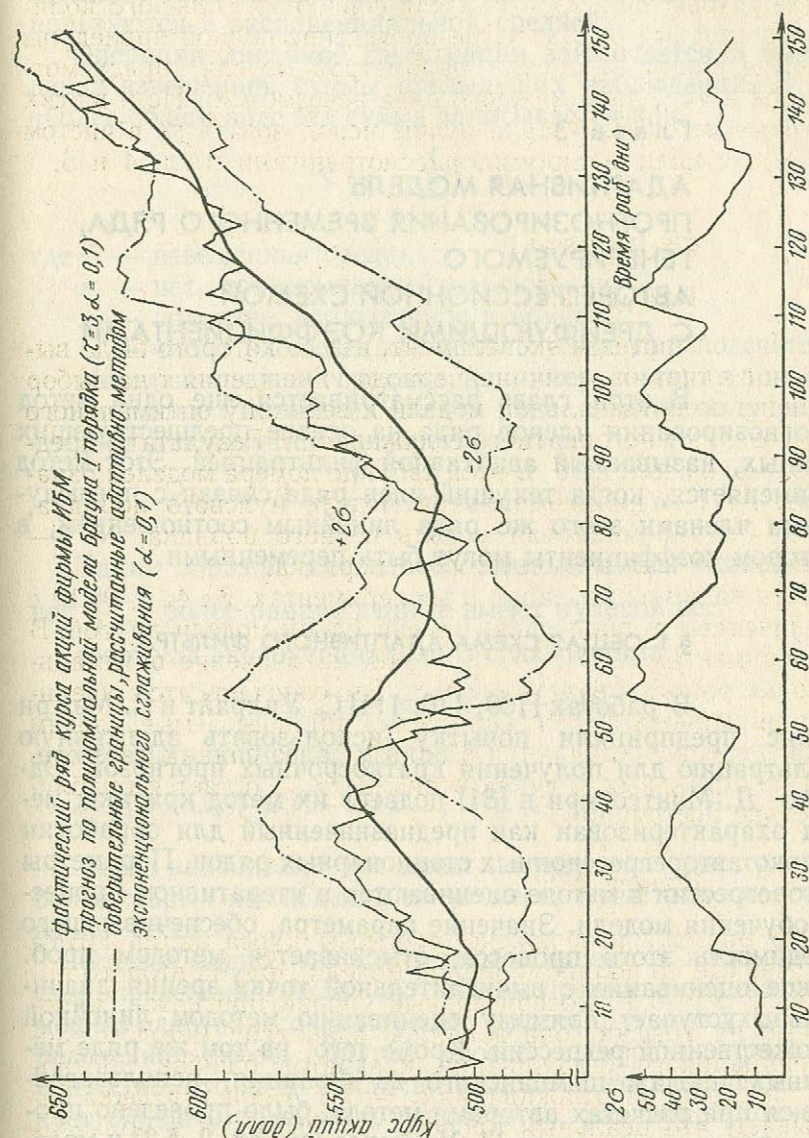


Рис. 2.7. Адаптивные доверительные границы прогнозов

Глава 3

АДАПТИВНАЯ МОДЕЛЬ ПРОГНОЗИРОВАНИЯ ВРЕМЕННОГО РЯДА, ГЕНЕРИРУЕМОГО АВТОРЕГРЕССИОННОЙ СХЕМОЙ С ДРЕЙФУЮЩИМИ КОЭФФИЦИЕНТАМИ

В этой главе рассматривается еще один метод прогнозирования членов ряда на основе предшествующих данных, называемый адаптивной фильтрацией. Этот метод применяется, когда текущий член ряда связан с предыдущими членами этого же ряда линейным соотношением, в котором коэффициенты могут быть переменными.

§ 1. ОБЩАЯ СХЕМА АДАПТИВНОГО ФИЛЬТРА

В работах [109, 110, 111] С. Уилрайт и С. Макридакис предприняли попытку использовать адаптивную фильтрацию для получения краткосрочных прогнозов. Однако Д. Монтгомери в [81] подверг их метод критике; метод охарактеризован как предназначенный для обработки только авторегрессионных стационарных рядов. Параметры авторегрессии в методе оцениваются в итеративном процессе обучения модели. Значение параметра, обеспечивающего сходимость этого процесса, отыскивается методом проб. Такое оценивание с вычислительной точки зрения значительно уступает прямому оцениванию методом линейной множественной регрессии. Кроме того, на том же ряде месячных продаж шампанского во Франции, использовавшемся при расчетах авторами метода, было проведено прогнозирование по модели П. Уинтерса (см. гл. 2, § 2) и модели Дж. Бокса — Г. Дженкинса ([5], вып. 1, гл. 9). Преимущество осталось за двумя последними моделями.

Учитывая опыт первых попыток построения моделей, основанных на идее адаптивной фильтрации, мы предлагаем способ организации адаптации моделей авторегрессионного типа на принципах, сходных с теми, которые используются в экспоненциальной средней.

Операция линейной фильтрации заключается в вычислении взвешенной суммы предыдущих наблюдений. В наиболее общем виде эта сумма записывается как

$$S = \sum_{i=t-l+1}^t \omega_i x_i, \quad (3.1)$$

где S — взвешенная средняя;

ω_i — вес, приписываемый наблюдению i ;

x_i — значение, наблюденное в момент i ;

l — число наблюдений, используемое при подсчете S .

Метод скользящих средних, например, состоит в том, что усредняются l последних членов ряда, а затем полученное среднее используется как прогноз, т. е.

$$\hat{x}_1(t) = S = \frac{1}{l} (x_t + x_{t-1} + \dots + x_{t-(l-1)}). \quad (3.2)$$

Таким образом, последние l данных имеют одинаковый вес $\frac{1}{l}$, а более ранние данные имеют нулевой вес.

Формула экспоненциального сглаживания

$$S_{t+1} = \alpha x_t + (1 - \alpha) S_t$$

может быть записана в виде:

$$S_{t+1} = \alpha x_t + \alpha (1 - \alpha) x_{t-1} + \alpha (1 - \alpha)^2 x_{t-2} + \dots \quad (3.3)$$

Здесь наибольший вес имеет текущее значение x_t , а для более старых наблюдений имеет место геометрическое убывание весов.

Отсюда видно, что эти методы отличаются лишь правилом определения весов ω_i . Метод прогнозирования с помощью адаптивного фильтра это просто другой подход к определению весов, другой способ построения фильтра. Очевидно, что из-за различий в весах от этих методов следует ожидать различных результатов, различной точности прогнозов.

Первая работа, посвященная синтезу фильтра, была опубликована Н. Винером [112] в 40-е годы. Винер рассмот-

рел вопросы построения линейных фильтров с постоянными коэффициентами для элиминирования шума и для сглаживания и прогнозирования стационарных процессов. Разработанные им процедуры дают оптимальные с точки зрения критерия наименьших квадратов результаты для случаев, когда ряд действительно стационарен.

Продолжая работу Винера, различные авторы, включая Р. Кальмана и Р. Бьюси [70], разработали процедуры, дающие оптимальные изменяющиеся во времени линейные



Рис. 3.1. Общая схема определения весов в прогнозе временного ряда

фильтры для нестационарных временных рядов. Для таких рядов метод Кальмана — Бьюси может дать лучшие результаты, чем обычный метод Винера.

Недостаток процедур Винера и Кальмана — Бьюси в том, что фильтры должны быть построены на основе априорных данных или допущений относительно статистик рассматриваемого временного ряда. На практике эти два метода фильтрации дают минимальные ошибки прогноза только при условии, что статистические характеристики ряда действительно соответствуют априорной информации, на основе которой были построены фильтры.

В описываемом здесь методе адаптивной фильтрации статистики ряда не измеряются, но учитываются неявно в процессе построения фильтра и итеративного обновления весов в ходе его адаптации.

Для того чтобы понять процедуру адаптивной фильтрации, рассмотрим рис. 3.1, на котором представлена общая схема определения весов в прогнозе временного ряда. Нижняя часть рисунка отражает то, что происходит в действительности. Реальные события взаимодействуют друг с другом в некой сложной динамической системе, в результате на выходе переменная принимает определенное значение.

Блоки и линии связи, расположенные выше, отражают процедуру адаптивной фильтрации. Это итеративная процедура. Ее цикл состоит в том, что берется данный набор весов, по ним делается прогноз, затем подсчитывается ошиб-

ка прогноза как разница между прогнозом и фактическим наблюдением, наконец, осуществляется корректировка весов для уменьшения среднего квадрата ошибки.

Задача состоит в том, чтобы найти подходящую процедуру корректировки (адаптации) весов, позволяющую переобстраивать модель (фильтр) в соответствии с текущей динамикой ряда. В следующем параграфе рассмотрим такую процедуру.

§ 2. АДАПТАЦИЯ КОЭФФИЦИЕНТОВ МОДЕЛИ АВТОРЕГРЕССИИ

Итак, ключом к эффективности адаптивной фильтрации является правило, используемое для адаптации весов на каждом шаге. Это правило может быть выведено путем анализа ошибки прогноза.

В модели адаптивного фильтра предполагается, что прогноз следующего члена ряда может быть получен с помощью взвешенной суммы l прошлых членов ряда

$$\hat{x}_1(t) = S, \quad (3.4)$$

где S — взвешенная средняя, определяемая выражением (3.1).

Выражение (3.4) имеет авторегрессионный характер. Поэтому эту модель будем также называть адаптивной моделью авторегрессии порядка l . После того как становится известным x_{t+1} — фактическое значение ряда в момент $t+1$, можно подсчитать ошибку прогноза $e_{t+1} = x_{t+1} - \hat{x}_1(t) = x_{t+1} - \sum \omega_i x_i$. Все суммирование здесь и далее будут производиться от $t-l+1$ до t .

Запишем выражение для квадрата ошибки:

$$e_{t+1}^2 = (x_{t+1} - \sum \omega_i x_i)^2 = x_{t+1}^2 - 2 \sum \omega_i x_i x_{t+1} + \sum_i \sum_j \omega_i \omega_j x_i x_j.$$

Это выражение показывает, что квадрат ошибки является квадратической функцией от весов ω_i . Геометрической интерпретацией этой функции в трехмерном пространстве является параболический цилиндр. При определенном сочетании весов (вдоль прямой $\sum \omega_i x_i = x_{t+1}$) функция принимает минимальное значение, равное нулю.

Предполагается, что в момент $t - 1$ были получены некоторые оценки весовых коэффициентов, по которым и подсчитана оценка x_t . Если использованные оценки весовых коэффициентов привели к ненулевой ошибке, то имеется возможность скорректировать веса так, чтобы уменьшить ошибку. Можно было бы выбрать такие веса, которые сводили бы ошибку к нулю. Но нужно иметь в виду, что по оценкам весов, полученным в момент t , будет сформирована оценка величины x_{t+1} . Между тем в момент $t + 1$ функция e_{t+1}^2 от весов изменится, так как ее коэффициенты (т. е. значения x -ов) будут другими. В трехмерном пространстве эту функцию будет отображать другой параболический цилиндр, принимающий минимальное значение вдоль новой прямой. Таким образом, процедуру адаптации весов в момент t нельзя рассматривать изолированно, так как корректировка весов на основе текущей ошибки производится с целью минимизации будущей ошибки. Поэтому процедура адаптации весов должна учитывать динамические особенности процесса и предусматривать сглаживание чисто случайных колебаний параметров изучаемой системы.

Очевидно, корректировку весов можно осуществлять различными способами. Мы рассмотрим один из них. Для адаптации весов воспользуемся методом наискорейшего спуска. Сущность его состоит в выборе начальной точки на исследуемой поверхности и последующем передвижении к низшей точке поверхности с применением итеративной процедуры. Для этого необходимо иметь возможность вычислять в каждой точке поверхности вектор, указывающий направление движения. Тогда можно корректировать веса таким образом, что новые веса будут представлять точку, которая ближе к оптимальному набору весов по сравнению со старыми весами. При использовании метода наискорейшего спуска корректировка весов осуществляется по следующему правилу:

$$W_H = W_C - k \text{ grad } (e_{t+1}^2),$$

где W_C — вектор старых весов;
 W_H — вектор новых весов;
 k — коэффициент ($k > 0$);
 $\text{grad } (e_{t+1}^2)$ — вектор, градиент e_{t+1}^2 .

Это уравнение показывает, что корректировка весов производится путем добавления к старому вектору весов поправки, получаемой умножением коэффициента k на гра-

диент, взятый с отрицательным знаком. Градиент с отрицательным знаком указывает наискратчайший путь достижения минимума исследуемой поверхности, а коэффициент k определяет, насколько мы продвигаемся в этом направлении.

Компоненты градиента находим дифференцированием квадрата ошибки по весам:

$$\frac{\partial e_{t+1}^2}{\partial \omega_i} = 2e_{t+1} \frac{\partial e_{t+1}}{\partial \omega_i} = -2e_{t+1} x_i.$$

В результате получаем, что градиент в целом

$$\text{grad } (e_{t+1}^2) = -2e_{t+1} X,$$

где X — вектор t последних наблюдений. Это определяет способ корректировки весов

$$W_H = W_C + 2ke_{t+1} X.$$

Неизвестным остается значение k , определяющее скорость движения в направлении, обратном градиенту. Чтобы определить характер влияния k на процедуру адаптации, поступим следующим образом. Вернемся в точку t и вновь сделаем прогноз, но уже с весами W_H . Получим новое значение ошибки:

$$(e_H)_{t+1} = x_{t+1} - \sum [(\omega_C)_i + 2ke_{t+1} x_i] x_i = x_{t+1} - \sum (\omega_C)_i x_i - \sum 2ke_{t+1} x_i^2 = e_{t+1} - 2ke_{t+1} \sum x_i^2 = e_{t+1} (1 - 2k \sum x_i^2),$$

где e_{t+1} — ошибка, полученная при старых весах $(\omega_C)_i$.

Теперь если положить

$$k = \frac{\alpha}{2 \sum x_i^2}, \quad \text{то } (e_H)_{t+1} = e_{t+1} (1 - \alpha)$$

и при $0 < \alpha < 2$ $|(e_H)_{t+1}| < |e_{t+1}|$.

Следовательно, α определяет реакцию модели на полученную ошибку и корректирует веса так, чтобы компенсировать ошибку на $(1 - |1 - \alpha|) \cdot 100\%$ или при $0 < \alpha < 1$ на $\alpha \cdot 100\%$. Вспомним простое экспоненциальное сглаживание (гл. 1, § 2), где экспоненциальная средняя корректируется аналогичным образом: $S_t = S_{t-1} + \alpha e_t$, где $e_t = x_t - S_{t-1}$.

Будем называть α параметром адаптации модели и считать его заданным и постоянным, k при этом будет переменной величиной. Оптимальное значение α , обеспечивающее минимум среднего квадрата ошибки, можно определить

методом проб на ретроспективном материале. Определение оптимального значения α можно назвать процедурой «обучения» адаптивного фильтра.

§ 3. ПРИМЕРЫ

Нами было проведено экспериментальное сравнение метода адаптивной фильтрации с другими методами, рассматриваемыми в этой книге. Модели сравнивались на одном и том же статистическом материале, приведенном в приложении № 4. Временные ряды экономического содержания имели различную динамику. Объем выборок от 21 до 360 точек.

Результаты сопоставления моделей на основе оценок точности ретроспективного прогнозирования приведены в табл. 6.7 и 7.4. Из этих таблиц видно устойчивое преимущество адаптивной модели авторегрессии перед моделями Брауна и их модификациями. Как известно, экономические ряды очень часто имеют авторегрессионный характер. Поэтому такие результаты получены, очевидно, не случайно. Лишь метод Бокса — Дженкинса, рассматриваемый в гл. 7, по-видимому, чаще дает несколько меньшую ошибку прогнозирования, чем метод адаптивной фильтрации. Однако простота построения и использования адаптивной модели авторегрессии позволяет выбирать структуру модели (определять величину l) и получать прогнозы на ЭВМ. Это делает ее намного привлекательнее моделей Бокса — Дженкинса, построение которых требует привлечения высококвалифицированных специалистов.

На рис. 3.2 приведен график, иллюстрирующий ретроспективное прогнозирование одного из рядов по адаптивной модели авторегрессии. Первые значения весов для начала итеративной процедуры адаптации как в этом, так и в остальных случаях моделирования определялись оцениванием методом наименьших квадратов соответствующей линейной множественной регрессии на базе первых 15 точек выборки.

Порядок модели авторегрессии, т. е. оптимальное значение параметра l , в экспериментах определялся методом перебора значений l от 1 до 5. В большинстве случаев $l_{\text{опт}}$ было равно 1 и ни разу не оказалось больше 3.

В заключение отметим, что использование статистических измерителей точности в методе адаптивной фильтрации остается областью, еще не исследованной. Практически, если выборка невелика (< 50 —60 точек), в качестве

статистического показателя, характеризующего доверительные интервалы прогноза, можно использовать среднеквадратическую ошибку ретроспективных прогнозов. Если же объем выборки составляет 100—200 или более точек,

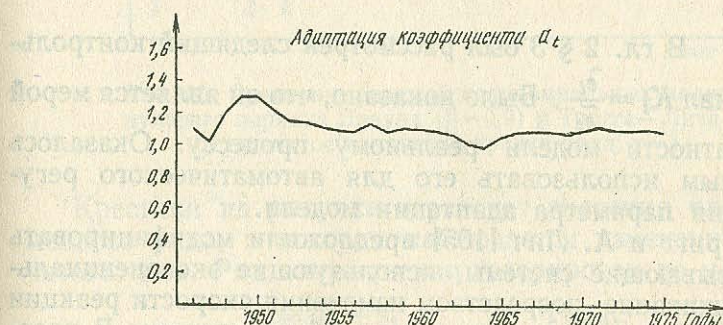
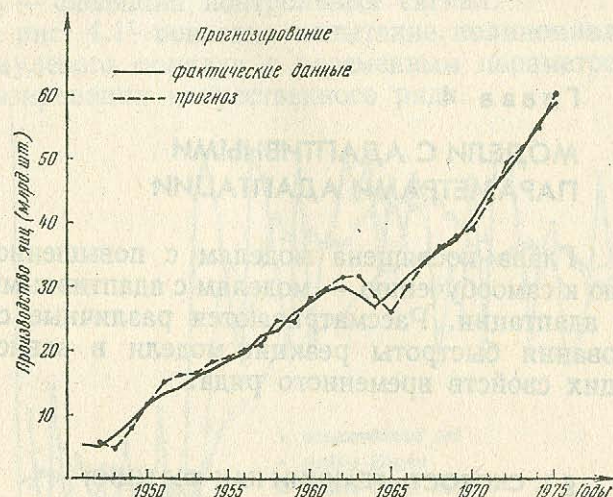


Рис. 3.2. Ретроспективное прогнозирование производства яиц по адаптивной модели авторегрессии $\hat{x}_t(t) = a_t x_t, \tau=1$

можно рекомендовать оценку доверительных интервалов по среднеквадратической ошибке, подсчитанной с помощью метода экспоненциального сглаживания.

В гл. 8 (§ 2) метод адаптивной фильтрации будет применен для построения адаптивной модели множественной регрессии общего вида.

Глава 4

МОДЕЛИ С АДАПТИВНЫМИ ПАРАМЕТРАМИ АДАПТАЦИИ

Глава посвящена моделям с повышенной способностью к самообучению — моделям с адаптивными параметрами адаптации. Рассматриваются различные способы регулирования скорости реакции модели в зависимости от текущих свойств временного ряда.

§ 1. СКОРОСТЬ РЕАКЦИИ КАК ФУНКЦИЯ СЛЕДЯЩЕГО КОНТРОЛЬНОГО СИГНАЛА (МОДЕЛЬ ТРИГГА — ЛИЧА)

В гл. 2 § 3 был рассмотрен следящий контрольный сигнал $K_t = \frac{\hat{e}_t}{e_t}$. Было показано, что он является мерой неадекватности модели реальному процессу. Оказалось возможным использовать его для автоматического регулирования параметра адаптации модели.

Д. Тригг и А. Лич [105] предложили модифицировать предсказывающие системы, использующие экспоненциальное сглаживание, посредством изменения скорости реакции в зависимости от величины контрольного сигнала. В простейшей модели это эквивалентно регулированию параметра сглаживания α . Наиболее очевидный способ заставить систему автоматически реагировать на расхождение прогнозов и фактических данных — это увеличение α с тем, чтобы придать больший вес свежим данным и, таким образом, обеспечить более быстрое приспособление модели к новой ситуации. Как только система приспособилась, необходимо опять уменьшить величину α для фильтрации шума.

Простой способ достижения такой адаптивной скорости реакции состоит в выборе

$$\alpha_t = |K_t|,$$

где K_t — следящий контрольный сигнал.

На рис. 4.1¹ показано испытание полиномиальной модели нулевого порядка с переменным параметром α при прогнозировании искусственного ряда.

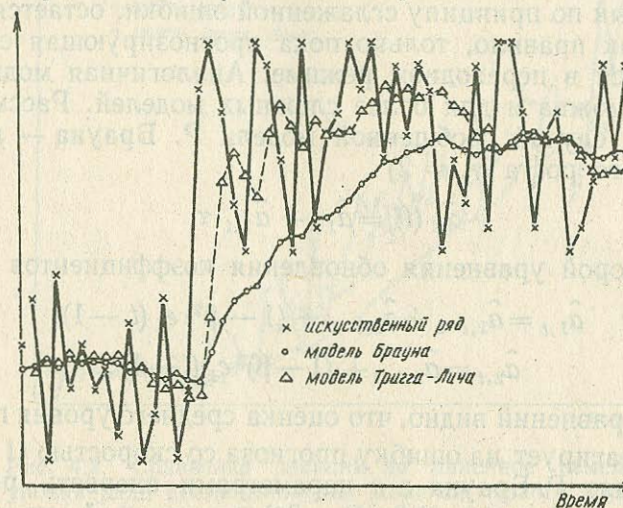


Рис. 4.1. Сравнение реакций полиномиальных моделей нулевого порядка Брауна ($\beta=0,9$) и Тригга-Лича на ступенчатое изменение уровня ряда, $\tau=1$

Крестики на рисунке отражают значения членов временного ряда, в котором наблюдается изменение ступенчатого типа. Ряд искусственно генерирован по модели

$$x_t = a'_1 + \varepsilon_t, \quad t < t_1;$$

$$x_t = a''_1 + \varepsilon_t, \quad t > t_1;$$

$$a'_1 = \text{const};$$

$$a''_1 = \text{const};$$

$$a'_1 \neq a''_1,$$

где ε_t — неавтокоррелированные случайные нормальные отклонения с нулевым математическим ожиданием и диспер-

¹ Рис. 4.1—4.4 заимствованы из работы Тригга и Лича.

сией σ^2 . Реакция простейшей модели экспоненциального типа с постоянным коэффициентом сглаживания $\alpha = 0,1$ отмечена кружками. Пунктирная линия характеризует реакцию подобной же системы, но с переменным α_t . Можно видеть, что система с адаптивным α приспосабливается к ступенчатым изменениям намного быстрее, а после отработки ступеньки размах ее колебаний не больше, чем у обычной системы, поскольку контрольный сигнал, построенный по принципу сглаженной ошибки, остается большим, как правило, только пока прогнозирующая система находится в переходном режиме. Аналогичная модификация возможна и для более сложных моделей. Рассмотрим частный случай обобщенной модели Р. Брауна — модель линейного роста ($n = 2$)

$$\hat{x}_t(t) = \hat{a}_{1,t} + \hat{a}_{2,t} \cdot t.$$

для которой уравнения обновления коэффициентов будут:

$$\hat{a}_{1,t} = \hat{a}_{1,t-1} + \hat{a}_{2,t-1} + (1 - \beta^2) e_1(t-1);$$

$$\hat{a}_{2,t} = \hat{a}_{2,t-1} + (1 - \beta)^2 e_1(t-1).$$

Из уравнений видно, что оценка среднего уровня процесса $\hat{a}_1$ реагирует на ошибку прогноза со скоростью $(1 - \beta^2)$. В моделях Р. Брауна с n параметрами скорость реакции определяется величиной $(1 - \beta^n)$, называемой эквивалентной постоянной сглаживания.

В многопараметрической модели Р. Брауна представляется естественным приравнять эквивалентную постоянную сглаживания модулю контрольного сигнала. В линейной модели мы могли бы положить

$$(1 - \beta^2) = |K_t|,$$

откуда

$$\beta_t = \sqrt{1 - |K_t|}.$$

Это означало бы, что каждый элемент вектора $\mathbf{h}$ как функции от β каждый раз претерпевает соответствующие изменения. Однако эксперименты показали, что зависимость всех элементов h_i от контрольного сигнала ухудшает прогноз, делая его неустойчивым. Если же ограничить модификацию вектора $\mathbf{h}$ только его первой составляющей h_1 , то эксперимент показывает, что во всех случаях такая система приводит к более стабильным результатам.

На рис. 4.2 показан ряд с линейной тенденцией роста, на который наложены те же случайные данные, что и на рис. 4.1. В этом примере среднеквадратическое отклонение шума взято пропорциональным среднему уровню ряда. Реакция, соответствующая прогнозу на один шаг вперед обычной модели линейного роста с эквивалентной постоян-

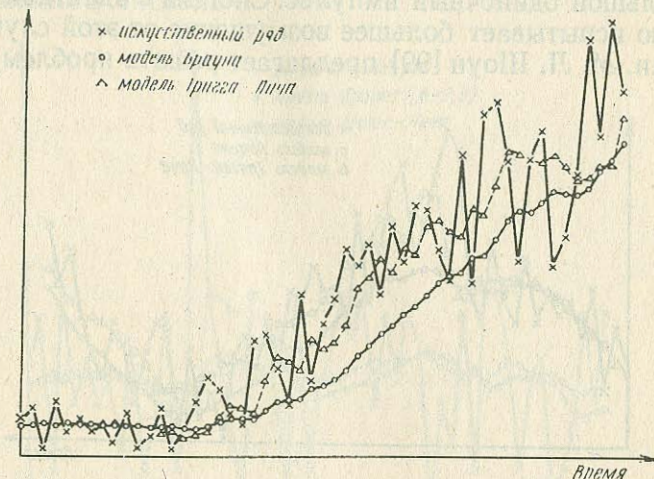


Рис. 4.2. Сравнение реакций на линейное изменение уровня ряда полиномиальных моделей первого порядка Брауна и Тригга—Лича, $\beta^2 = 0,9$, $\tau = 1$

ной $(1 - \beta^2) = 0,1$, отмечена кружками. Прогнозы аналогичной модели, но с $h_1 = |K_t|$ показаны пунктиром.

На рис. 4.3 линия, обозначенная крестиками, показывает те же случайные данные, но с синусоидальным колебанием с периодом 52 (52 недели в году).

Через 15 интервалов возникает ступенчатое изменение и отмеченная кружками линия показывает реакцию модели

$$\hat{x}_t(t) = \hat{a}_{1,t} + \hat{a}_{2,t} \tau + \hat{a}_{3,t} \sin \frac{2\pi t}{52} + \hat{a}_{4,t} \cos \frac{2\pi t}{52}$$

при $(1 - \beta^4) = 0,1$, $\tau = 1$.

Пунктирная линия обозначает реакцию той же модели, когда $h_1 = |K_t|$.

Практическое моделирование дало результаты, похожие на представленные на рис. 4.1—4.3, т. е. приравнивание h_1 модулю следящего контрольного сигнала дает значительно

большую скорость приспособления к новому характеру ряда.

Ясно, что не во всех реально возникающих ситуациях модели с адаптивным параметром α дают лучшие прогнозы по сравнению с обычными адаптивными системами. На рис. 4.4 показаны сравнительные реакции этих моделей на большой одиночный импульс. Система с адаптивной скоростью испытывает большее возмущение от этой случайной помехи. М. Л. Шоун [99] предлагает решать проблему чрез-



Рис. 4.3. Сравнение моделей Брауна и Тригга—Лича, $n=4$; $\beta^n=0,9$, $\tau=1$

мерной реакции на большой одиночный импульс путем задержки использования контрольного сигнала на один шаг. Если взять в качестве первого элемента вектора $\mathbf{h}$ для данного момента t модуль контрольного сигнала, подсчитанный в предыдущий момент $t-1$, то отмеченный недостаток преодолевается. Реакция на ступеньку или на линейный тренд станет при этом медленнее.

Метод приравнивания $h_1 = |K_t|$ прост и эффективен. Особенно он удобен там, где прогнозы делаются на ЭВМ для большого числа рядов. Если динамика ряда не изменяется, то шум фильтруется столь же эффективно, как и обычным методом с постоянным параметром α . Когда такие возмущения все же случаются, адаптация происходит более быстро.

Этот метод особенно ценен для моделирования рядов с короткой историей или на первоначальном этапе построения модели. Чтобы им воспользоваться, достаточно провести грубую оценку коэффициентов прогнозирующей мо-

дели. Если ошибки прогнозов по этой модели велики, то у контрольного сигнала проявляется тенденция к быстрому абсолютному росту и система приспособляется с соответствующей скоростью. Этот метод позволяет обойти проблему определения оптимального значения α . Однако возникает задача выбора наилучшего значения γ для подсчета контрольного сигнала.

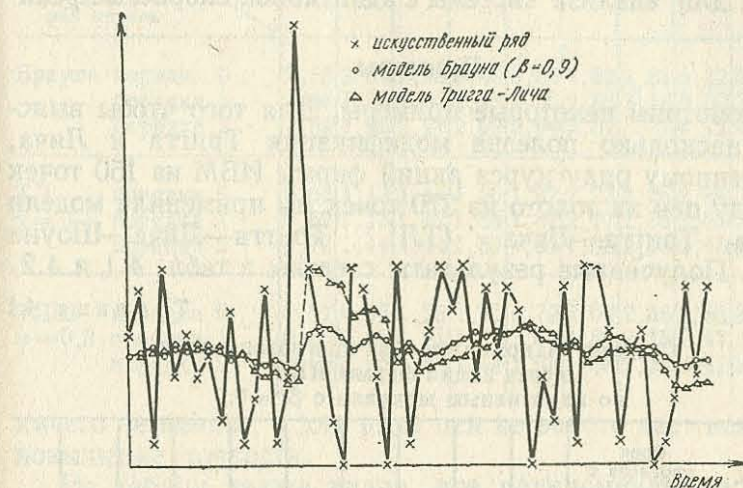


Рис. 4.4. Сравнение реакций простейших моделей Брауна и Тригга—Лича ($n=1$) на единичный импульс, $\tau=1$

В приведенных выше примерах моделирования величина $\gamma = 0,1$. Если необходима большая осторожность по отношению к скорости процесса адаптации, величину γ можно уменьшить и взять $\gamma = 0,05$. После того как γ задана, модель работает автоматически. Иногда все же необходимо вмешательство человека для корректировки прогнозов, даваемых системой с адаптивной скоростью реакции, так как нельзя ожидать, что система справится с любыми изменениями в структуре ряда. Кроме того, исследователь может располагать ценной дополнительной информацией.

Моделирование реакции простейшей модели на ступенчатые изменения показало, что хотя модель довольно быстро устанавливается на новом среднем уровне, контрольный сигнал еще некоторое время остается большим. Это мешает системе отфильтровывать случайный шум, что особенно проявляется, когда константа γ , использованная для полу-

чения сглаженной ошибки и среднего абсолютного отклонения, мала, скажем, 0,05. Большая величина γ увеличивает дисперсию экспоненциальной средней ошибки, до некоторой степени ухудшая возможности достижения цели, ради которой используется адаптивная скорость реакции.

Заметим, что функции распределения контрольного сигнала, указанные в табл. 2.3—2.4, не могут быть использованы для анализа системы с адаптивной скоростью реакции.

Примеры

Рассмотрим некоторые примеры. Для того чтобы выяснить, насколько полезна модификация Тригга и Лича, к временному ряду курса акций фирмы ИБМ из 150 точек и к ряду цен на золото из 370 точек мы применили модели Брауна, Тригга—Лича (ТЛ), Тригга—Лича—Шоуна (ТЛШ). Полученные результаты сведены в табл. 4.1 и 4.2.

Таблица 4.1

Средние квадраты ошибок прогнозирования
курса акций фирмы ИБМ
по адаптивным моделям с $\beta^n=0,9$

Срок прогноза τ										
	1	2	3	4	5	6	7	8	9	10
Полиномиальная модель										
Брауна порядка 0	248	297	349	402	458	518	579	643	707	772
порядка 1	180	217	255	297	343	388	435	477	512	559
порядка 2	182	221	263	320	388	436	471	493	613	716
ТЛ порядка 0	59	94	130	167	206	249	297	355	417	474
$\gamma=0,3$ порядка 1	60	94	128	166	208	253	296	338	366	405
порядка 2	61	93	131	195	297	361	373	311	422	584
ТЛШ порядка 0	63	102	134	169	205	248	303	362	421	474
$\gamma=0,3$ порядка 1	61	100	132	167	205	247	299	345	369	403
порядка 2	61	100	135	197	297	352	373	318	400	525

Сравнение этих результатов свидетельствует о преимуществах моделей Тригга—Лича перед соответствующими моделями Брауна. Задержка при использовании контрольного сигнала для обновления параметра сглаживания, предложенная Шоуном, для ряда курса акций ИБМ практически

Таблица 4.2

Средние квадраты ошибок прогнозирования цен на золото
по адаптивным моделям с $\beta^n=0,9$

Срок прогноза τ										
	1	2	3	4	5	6	7	8	9	10
Полиномиальная модель										
Брауна порядка 0	2,16	2,57	2,98	3,40	3,85	4,32	4,81	5,32	5,84	6,38
порядка 1	1,48	1,75	2,01	2,28	2,56	2,86	3,18	3,50	3,83	4,17
порядка 2	1,64	1,96	2,27	2,59	2,94	3,31	3,71	4,12	4,56	5,01
ТЛ порядка 0	0,68	1,13	1,45	1,70	1,97	2,22	2,55	2,89	3,18	3,47
$\gamma=0,3$ порядка 1	0,69	1,13	1,42	1,64	1,87	2,07	2,35	2,63	2,85	3,07
порядка 2	0,72	1,14	1,41	1,61	1,81	1,98	2,23	2,45	2,61	2,79
ТЛШ порядка 0	0,57	0,95	1,26	1,52	1,79	2,09	2,46	2,86	3,22	3,51
$\gamma=0,3$ порядка 1	0,60	0,95	1,21	1,42	1,62	1,85	2,15	2,47	2,75	2,94
порядка 2	0,62	0,96	1,22	1,43	1,63	1,85	2,14	2,42	2,67	2,84

ничего не меняет, а для ряда цен на золото дает некоторое повышение точности.

Из таблиц также видно, что полиномиальные модели ТЛШ нулевого, первого и второго порядков имеют примерно одинаковые дисперсии ошибок для небольших τ . Графики (мы их не приводим) для моделей ТЛШ при $\tau = 1, 2, 3$ подтверждают, что специфика полиномиальных моделей нулевого, первого и второго порядков здесь теряется и прогнозы этих предикторов располагаются весьма близко. Это же относится к моделям ТЛ. Такой результат объясняется тем, что корректировка прогнозов обеспечивается в основном за счет корректировки первого коэффициента полиномиальных моделей.

§ 2. РЕГУЛИРОВАНИЕ ПАРАМЕТРА АДАПТАЦИИ ПО ИЗМЕНЕНИЯМ СПЕКТРАЛЬНЫХ ХАРАКТЕРИСТИК

Модель Рао и Шапиро (спектральная модель)

Изменения в структуре рядов (острый пик или ступенчатое изменение) будут обнаруживаться в их спектральных характеристиках. Основная идея метода [96] заключается в том, чтобы использовать изменения в спектре

для управления величиной α в процедуре подсчета экспоненциальной средней. Спектр оценивается сначала для отрезка временного ряда, образуемого членами 1, 2, ..., N , затем для отрезка с членами ряда 2, 3, ..., $N+1$ и т. д. Ряд рассматривается как бы через движущееся вдоль оси времени окно постоянной ширины N . Таким образом, спектр последовательно оценивается для частично перекрывающихся отрезков временного ряда постоянной длины. Длина этих отрезков N должна быть достаточно велика, чтобы получать устойчивые оценки некоторого числа компонент спектра, но не слишком большой, иначе случай фундаментального изменения ряда будет потерян при усреднении.

Спектральная плотность случайного процесса характеризует распределение всей энергии стационарного ряда как функцию частоты. Математически это может быть выражено в виде преобразования Фурье автоковариационной функции ряда.

Оценки спектральной плотности как функции частоты получают, подсчитывая¹:

$$\hat{S}_x(h\Delta\nu) = \left[\hat{R}_{xx}(0) + 2 \sum_{p=1}^{m-1} \hat{R}_{xx}(p) \cos \frac{\pi p h}{m} + \hat{R}_{xx}(m) \cos \pi h \right],$$

где $h = 0, 1, \dots, m$; $\Delta\nu$ — частота первой гармоники;

$$\hat{R}_{xx}(p) = \frac{1}{N-p} \sum_{i=1}^{N-p} x_i x_{i+p} - \frac{1}{(N-p)^2} \sum_{i=1}^{N-p} x_i \sum_{i=1}^{N-p} x_{i+p},$$

где m — максимальная величина лага, для которого подсчитывается автоковариационная функция;

N — длина временного отрезка.

Первичные спектральные оценки затем выравниваются. Это предлагается делать методом Тьюки—Хеннинга:

$$\hat{f}_x(h\Delta\nu) = \frac{1}{4} \hat{S}_x[(h-1)\Delta\nu] + \frac{1}{2} \hat{S}_x(h\Delta\nu) + \frac{1}{4} \hat{S}_x[(h+1)\Delta\nu]$$

¹ Если m не простое число, то возможно быстрое преобразование Фурье, экономное с вычислительной точки зрения (см. [13, вып. 2, с. 68]).

для $h = 1, 2, \dots, m-1$;

$$\hat{f}_x(0) = \frac{1}{2} \hat{S}_x(0) + \frac{1}{2} \hat{S}_x(\Delta\nu);$$

$$\hat{f}_x(m\Delta\nu) = \frac{1}{2} \hat{S}_x[(m-1)\Delta\nu] + \frac{1}{2} \hat{S}_x(m\Delta\nu).$$

В экспериментах была выбрана длина временного отрезка $N = 36$ точек, при $m = 6$, т. е. всего семь компонент спектра. Таким образом, сначала оценивался спектр отрезка от точки x_1 до точки x_{36} , затем отрезка от x_2 до x_{37} и т. д. Эти спектры могут быть отражены в табл. 4.3.

Таблица 4.3

Последовательные спектры

Отрезок ряда	Спектральные характеристики			
	$k=0$	$k=1$	...	$k=6$
$x_1, \dots, x_{36}$	$\hat{F}_{10}$	$\hat{F}_{11}$		$\hat{F}_{16}$
$x_2, \dots, x_{37}$	$\hat{F}_{20}$	$\hat{F}_{21}$		$\hat{F}_{26}$
$x_3, \dots, x_{38}$	$\hat{F}_{30}$	$\hat{F}_{31}$		$\hat{F}_{36}$
...				

$\hat{F}_{ik} = \ln \hat{f}_{ik}$ здесь означает натуральный логарифм оценки спектральной плотности для k -й гармоники на i -м отрезке ряда. Ясно, что изменения в структуре ряда будут представлены изменениями в последовательных спектрах, т. е. векторами

$$\{\hat{F}_{ik} - \hat{F}_{i+1,k}\}.$$

Для сравнения этих изменений нужны элементы $\hat{F}_{ik}$, которые перед подсчетом разностей желательно выравнивать. Для выравнивания использовалась скользящая средняя для трех значений F и отклонения от этой скользящей средней подсчитывались следующим образом:

$$\delta_{ik} = \frac{1}{3} (\hat{F}_{i-2,k} + \hat{F}_{i-1,k} + \hat{F}_{ik}) - \hat{F}_{ik}.$$

Обозначим через λ_i максимальное по модулю изменение спектра, выравненного за три периода:

$$\lambda_i = \max_k |\delta_{ik}|, k = 0, 1, \dots, 6.$$

Если величина λ_i мала по сравнению со стандартным отклонением, то это означает, что никаких значительных перемен в структуре временного ряда не произошло. Таким образом, α следует оставлять малой. Когда λ_i велико, величину α следует повышать вплоть до максимума, т. е. до 1.

Для определения величины α значения λ_i могут быть использованы различными путями. Считалось, что новое значение α должно зависеть от $(\lambda_i)^2$, чтобы сделать модель более чувствительной к изменениям в структуре рядов.

Вводилась величина

$$d_i = b + c \left(\frac{\lambda_i}{\sigma} \right)^2,$$

где σ — стандартное отклонение δ_{ik} ;

b и c — соответствующим образом выбранные константы, рассматриваемые ниже.

Величина α определялась по формуле

$$\alpha_i = \max [0, 1; \min (\exp d_i - 1; 1)].$$

Константы b и c могут быть определены из условий:

$$b + r_1^2 c = 0,67; \quad (e^{0,67} - 1 = 0,95);$$

$$b + r_2^2 c = 0,095; \quad (e^{0,095} - 1 = 0,1),$$

где r_1 — то значение $\frac{\lambda_i}{\sigma}$, при котором α достигает значения $\alpha = 0,95$;

r_2 — то значение $\frac{\lambda_i}{\sigma}$, ниже которого α остается равным

0,1. Придя к заключению, что $\frac{\delta_{i0}^2}{\sigma^2}$, $\frac{\delta_{i1}^2}{\sigma^2}$, ..., $\frac{\delta_{im}^2}{\sigma^2}$ можно считать случайными величинами, асимптотически распределенными как χ^2 с одной степенью свободы¹, авторы метода предлагают следующую интерпретацию r_1 и r_2 .

¹ Величина, являющаяся суммой квадратов l независимых стандартизованных нормально распределенных переменных, как известно, имеет χ^2 распределение с l степенями свободы (см., например, [31], с. 49). Так как δ_{ik} по предположению распределены асимптотически нормально, то $\frac{\delta_{ik}^2}{\sigma^2}$ будут распределены как χ^2 с одной степенью свободы. Таблицы стандартных распределений можно найти в учебниках по статистике и справочниках по математике.

Для того чтобы найти такое значение A , при котором $P(\lambda_i < A) = 0,99$, необходимо, чтобы $n \cdot P(|\delta_{ik}| > A) = 0,01$, где n — количество частот в разложении, т. е. в нашем случае $n = 7$. Отсюда $P(|\delta_{ik}| > A) = 0,0014$, или $P\left(\frac{\delta_{ik}^2}{\sigma^2} > \frac{A^2}{\sigma^2}\right) = 0,0014$.

Но $\frac{\delta_{ik}^2}{\sigma^2}$ асимптотически распределено как χ^2 с одной степенью свободы. Отсюда $\frac{A^2}{\sigma^2} = 10,2$, или $\frac{A}{\sigma} = 3,1$.

Следовательно, $r_1 = 3,1$ означает, что α_i приравнивается 0,95, когда λ_i принимает такое значение, вероятность превышения которого при условии, что коренных изменений ряд не претерпел, равна 0,01; т. е. получение столь больших λ_i при отсутствии изменений в динамике ряда маловероятно. Поэтому есть основание полагать, что ряд претерпел существенные изменения и нужно обеспечить быструю адаптацию модели.

Интерпретацию r_2 можно получить аналогичным образом ($P(\lambda_i > A) = 0,99$ и т. д.). Остается определить стандартное отклонение σ , необходимое для вычисления d_i .

Авторы рассматриваемого метода получили следующую приближенную оценку дисперсии:

$$D(\delta_{ik}) = \sigma^2 = \frac{Km}{N} \left(\frac{2}{3} - \frac{2}{9} R_p^2 - \frac{4}{9} R_{2p}^2 \right) (1 + q_k),$$

где K — константа, зависящая от специфического используемого фильтра;

$$q_k = \begin{cases} 1, & \text{если } k = 0, \\ 0 & \text{— в других случаях;} \end{cases}$$

R_p^2 — когерентность (сохранено обозначение, принятое в [96]).

Когерентность на частоте $k\Delta\nu$ является показателем корреляции между соответствующими частотными составляющими двух отрезков ряда $\{x_j, \dots, x_{j+N-1}\}$ и $\{x_l, \dots, x_{l+N-1}\}$ по N членов в каждом.

Для стационарных процессов когерентность практически не зависит от частоты, а от j и l зависит только через разность $p = |j - l|$. Для R_p^2 авторы вывели следующее простое соотношение:

$$R_p^2 = \left(1 - \frac{p}{N} \right)^2.$$

Для рассматриваемого ими примера $N = 36$, $m = 6$, $K = 0,75$, $p = 1$. Отсюда

$$D(\delta_{ik}) = \sigma^2 = 0,0075 (1 + q_k).$$

Таким образом, алгоритм вычислений полностью определен.

Однако применение спектральной модели сопряжено еще с одной трудностью, которую можно проиллюстрировать следующим примером. Пусть $x_1, \dots, x_{100}$ будет временной ряд продаж. Предположим, что этот ряд характеризуется положительным ступенчатым скачком от x_{50} до x_{100} , т. е. $M(x_1) = \dots = M(x_{49}) < M(x_{50}) = \dots = M(x_{100})$, и что оценка спектра S_i подсчитывается для последовательных отрезков ряда

$$x_i - 35, \dots, x_i, i = 36, \dots, 100$$

(S_i , конечно, функция частоты). Никаких значительных изменений в спектре $S_{36}, \dots, S_{49}$ не должно быть. Однако S_{49} и S_{50} должны различаться существенно, так как первая величина не включает ступенчатое изменение, а вторая включает. Некоторые различия могут также иметь место для нескольких следующих значений спектра. Далее спектр практически не меняется. Однако между S_{84} и S_{85} опять будет значительная разница, так как S_{84} включает ступенчатый скачок, а S_{85} не включает. Таким образом, величина α_{85} будет большой, даже если структура ряда в точке 85 не изменилась.

Может показаться, что наипростейший способ устранения указанного недостатка состоит в том, чтобы в программе предусмотреть автоматическое установление $\alpha = 0,1$ через 36 точек (в данном примере) после того, как отмечено изменение структуры ряда. Однако этот путь может быть опасен. Во-первых, в данной точке может произойти действительное изменение и оно будет пропущено. Во-вторых, может иметь место целая серия промежуточных скачков в уровне продаж между первой и 36 точками.

Эта проблема была решена оценкой двух спектров: для 36-точечных и 18-точечных отрезков временного ряда. Для каждого из этих спектров определялось соответствующее α_i . Меньшее из них использовалось в качестве текущего значения сглаживающего коэффициента.

Авторами метода предлагается и другое, более простое решение данной проблемы. Перед подсчетом спектральных оценок начало 36-точечного отрезка преобразуется путем замены $x_t, t = 1, \dots, 6$, на $x'_t = x_t \sin \frac{2\pi(t-1)}{24}$.

Сравнение спектральной модели с моделью Тригга—Лича

Для сравнения рассмотренной адаптивной процедуры с методом Тригга—Лича авторами было промоделировано несколько рядов с различными типами изменений.

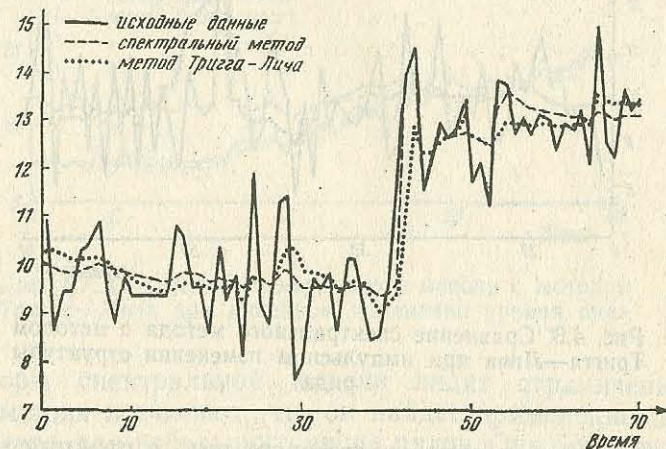


Рис. 4.5. Сравнение спектрального метода с методом Тригга—Лича при ступенчатом изменении уровня ряда

Первый ряд был сконструирован для проверки реакции спектрального метода на ступенчатое изменение. Ряд был образован из независимых наблюдений $x_1, \dots, x_{70}$, таких, что $x_1, \dots, x_{40}$ были распределены нормально со средним значением 10 и $\sigma_x = 1$, в то время как $x_{41}, \dots, x_{70}$ были распределены нормально со средним значением 13 и $\sigma_x = 1$.

Реакция спектральной модели и модели, использующей контрольный сигнал Тригга, для этого случая показана на рис. 4.5. Видно, что ни тот, ни другой метод не отражает первого скачка. Причиной является то, что величина скачка не превышает пределов 3σ стохастических колебаний относительно первоначального уровня и скачок восприни-

маются моделями как случайное отклонение. В следующей точке, однако, спектральная модель следует за рядом к его новому среднему уровню, в то время как модель Тригга—Лича слегка отстает. Спектральный метод в данном случае дает меньшее среднеквадратическое отклонение.

Далее исследовалась реакция на острый пик. Пример такого пика показан на рис. 4.6 в точке 46. Спектральная

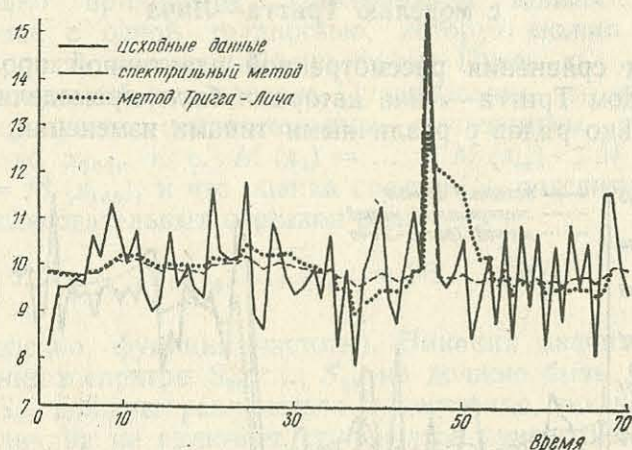


Рис. 4.6. Сравнение спектрального метода с методом Тригга—Лича при импульсном изменении структуры ряда

модель почти идеально отслеживает пик, а реакция модели с контрольным сигналом отражает лишь половину высоты пика; затем, как видно на рисунке, потребовалось 6 точек на возвращение назад.

Рао и Шапиро указывают, что спектральный метод в данном случае отражает значение ряда с меньшей ошибкой, чем метод Тригга—Лича. Но на рис. 4.6, взятом из работы Рао и Шапиро, эффект запаздывания прогноза спектральной модели по отношению к реальным изменениям не отражен ($\tau = 0$). Однако спектральная модель, как и любая адаптивная модель рассматриваемого класса, характеризуется запаздыванием реакции на изменения в реальном процессе по крайней мере на один шаг. Значит, в точке 45 нет информации для прогнозирования скачка в точке 46. В точке 46 составляется прогноз на следующий момент времени. Спектральная модель при этом уже в полной мере

учитывает изменения в точке 46, несмотря на то, что этот скачок имеет характер помехи. Поэтому прогноз по спектральной модели на момент времени 47 будет иметь большую погрешность, чем прогноз модели Тригга и Лича, хотя в дальнейшем она быстрее приспособится к фактическому уровню. Вообще сильная реакция на пик едва ли является положительным свойством модели прогнозирования.

На рис. 4.7 показан третий тип нарушения стационарности — линейный рост уровня ряда. Для этого случая точность обоих методов приблизительно одинакова.



Рис. 4.7. Сравнение спектрального метода с методом Тригга—Лича при линейном изменении уровня ряда

Авторы спектральной модели видят ограниченность своего метода и считают, что он найдет применение, вероятнее всего, для небольшого числа рядов. Они рекомендуют этот метод только для весьма колеблющихся процессов. Однако для таких рядов эффект запаздывания реакции модели будет особенно сказываться на точности прогнозирования. По-видимому, в подобном случае предпочтительнее использовать «наивную» модель $x_t(t) = x_t$.

Рао и Шапиро считают, что особенно эффективно применение спектральной модели для определения типа перемен в структуре ряда. Пусть $k\Delta\nu$ означает частоту, при которой $\lambda_i = |\delta_{ik}|$. Эксперименты показывают, что когда переменные в структуре ряда имеют ступенчатый характер, то k принимает малые значения, т. е. максимальные изменения спектра имеют место в низких частотах.

Когда переменные имеют форму пика, значение k соответствует высоким частотам. Линейный рост вызывает максимальные отклонения в середине частотного диапазона.

Эти особенности дают возможность использовать изменения в спектре для введения существенных изменений в структуру модели, а не только для регулирования параметра сглаживания. Мы же попытаемся решить эту проблему в следующей главе более простым способом.

§ 3. АДАПТАЦИЯ ПАРАМЕТРА МЕТОДОМ ЭВОЛЮЦИИ

Другой способ модификации однопараметрической экспоненциальной модели с коррекцией коэффициента линейного тренда предложен У. Чоу [50]. Он также отказывается от поиска какого-либо одного оптимального значения параметра сглаживания и предлагает процедуру, обеспечивающую адаптацию самого этого параметра к изменениям в динамике ряда. Метод комбинирует экспоненциальное сглаживание с методом эволюционного планирования [43, 26].

Предполагается, что используются три оценки следующего члена ряда, получаемые с помощью трех различных значений параметра α : высокого, нормального, низкого. Эти значения первоначально выбираются произвольно, а затем модифицируются по мере продвижения модели во времени. Оценка, сделанная при нормальном значении параметра, считается прогнозом, остальные две оценки являются контрольными. По получении новой фактической точки можно сделать вывод о том, какое из трех значений параметра привело к меньшей ошибке и оказалось наилучшим. Теперь это значение параметра рассматривается уже как нормальное. Прогноз на следующий период делается именно на основе данного наилучшего значения. Одновременно рассчитываются контрольные оценки будущего наблюдения по более высокому и более низкому (по сравнению с нормальным) значениям параметра.

В своих экспериментах Чоу использовал значения параметров

$$0,05 \leq \alpha - h, \alpha, \alpha + h \leq 0,95,$$

где $h = 0,05$ — шаг изменения (эволюции) параметра;

α — нормальное значение параметра;

$(\alpha - h)$ — низкое значение параметра;

$(\alpha + h)$ — высокое значение параметра.

В качестве начального значения он брал $\alpha = 0,1$, как наиболее широко употребляемое на практике. Всего было проанализировано 60 рядов месячных данных о сделках на различные виды продукции. Здесь были ряды с данными о перчатках, смазочных материалах, сальниках, подшипниках и т. д. Данные представляли собой разнообразные образцы поведения экономических временных рядов, включая циклическое движение. В 59 случаях предлагаемый метод показал преимущества перед стандартной процедурой и в одном случае результаты были почти одинаковы. Обобщая результаты исследований, Чоу делает вывод о том, что точность его схемы (если судить по относительной ошибке прогноза) более чем на 10% выше, чем у стандартного метода; ее преимущества становятся особенно явными при моделировании нестационарных рядов с изменяющимися автокорреляционными свойствами.

Идея Чоу получила дальнейшее развитие в работе С. Робертса и Р. Рида [97], в которой рассмотрены также двух- и трехпараметрические модели. Адаптационная оптимизация модели и в этом случае проводится методом эволюционного планирования. В данной схеме реализуются все возможные опыты (пробы) при различных сочетаниях значений параметров сглаживания, заданных из расчета по три значения на каждый параметр. Эти опыты образуют так называемый полный факторный эксперимент с добавлением центральной точки. Всего при k управляемых параметрах в каждый момент времени подсчитывается $2^k + 1$ оценок будущего наблюдения, и одна из них, соответствующая центральной точке плана, считается прогнозом.

После получения фактической точки анализируется точность пробных оценок (анализируются ошибки, являющиеся функциями рассматриваемых параметров). В результате может быть принято решение о переносе центральной точки в точку с минимальной ошибкой. Прогноз на следующий момент времени осуществляется при значениях параметров сглаживания, определяемых центральной точкой. Вследствие переноса центральной точки изменяется набор высоких и низких уровней рассматриваемых параметров. При этом для стабильности прогностической системы накладывают ограничение, чтобы значения параметров были в пределах 0,05—0,95.

Можно провести аналогию этой схемы с биологической эволюцией. Небольшие изменения независимых параметров

можно уподобить мутациям. Отбор лучших значений параметров аналогичен процессу естественного отбора. Отсюда и название метода — эволюционное планирование.

В качестве базовой модели авторы схемы взяли трехпараметрическую модель Уинтерса, рассмотренную в гл. 2, § 2. Это модель с аддитивным коэффициентом роста и мультипликативной сезонностью; более простые модели были получены путем наложения на нее соответствующих ограничений.

Рассматривались три случая: 1. Один параметр α . Отсутствие тенденции роста и сезонности. 2. а) два параметра α_1, α_2 . Наличие тенденции роста и отсутствие сезонности; б) два параметра α_1, α_2 . Наличие сезонности и отсутствие тенденции роста. 3. Три параметра $\alpha_1, \alpha_2, \alpha_3$. Наличие и сезонности, и тенденции роста. Полная модель Уинтерса.

Рис. 4.8. План факторного эксперимента для системы с одним параметром

Однопараметрическая система. Оценки будущего члена временного ряда получают по трем значениям единственного параметра α , как это показано на рис. 4.8. Точка 1 соответствует нормальному значению α , точка 2 — высокому, точка 3 — низкому значению α .

В гл. 5, § 3 приведены результаты, полученные нами при эмпирических испытаниях однопараметрических предикторов с регулированием параметра адаптации методом эволюции. Проведено их сравнение с однотипными моделями Брауна с оптимальными значениями α , определенными при помощи итеративной процедуры поиска.

Двухпараметрическая система. Рассмотрим случай 2а. Результаты в случае 2б аналогичны. Система имеет два параметра α_1, α_2 . Оценку будущего члена ряда получают в соответствии с планом факторного эксперимента: по два значения на каждый параметр плюс центральная точка (см. рис. 4.9).

Трехпараметрическая система. Система имеет параметры $\alpha_1, \alpha_2, \alpha_3$, для которых план факторного эксперимента, представленный на рис. 4.10, строится аналогичным образом. Всего точек плана $2^3 + 1$.

На основе статистических критериев нельзя установить существенное преимущество рассмотренной схемы перед обычной моделью Уинтерса. Однако для рядов, характери-

зующихся высокой автокорреляцией, предложенная модель дает более точные прогнозы, чем модель Уинтерса. С учетом того, что в экономических рядах часто присутствует автокорреляция, эта модель во многих случаях окажется полезной. Кроме того, реакция на стандартные входные потоки — единичный импульс, ступенчатое изменение уровня, линейно-нарастающая функция — у данной схемы лучше (с точки зрения точности прогнозов), чем у модели

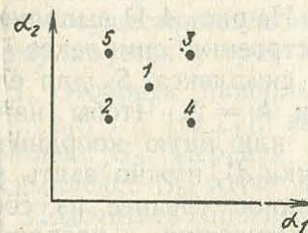


Рис. 4.9. План факторного эксперимента для системы с двумя параметрами

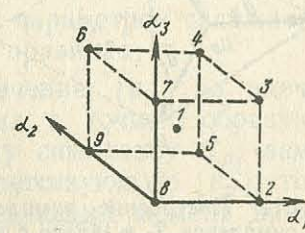


Рис. 4.10. План факторного эксперимента для системы с тремя параметрами

Уинтерса, и ее применение предпочтительно, когда о структуре исследуемого ряда имеется недостаточно полное представление.

Другой прием адаптивной оптимизации — симплекс-планирование — предложен У. Спиндлеем, Дж. Хечтом и Ф. Химсуорсом [101] и использован в адаптивных моделях прогнозирования Д. Монтомери [80]. Этот метод дает возможность заранее определить четкие правила принятия решений о том, куда и когда двигаться из данной точки параметрического пространства. В отличие от метода эволюционного планирования, требующего в каждой точке рассчитывать $2^k + 1$ пробу (k — количество управляемых параметров), метод симплекс-планирования требует в каждой точке делать один-единственный прогноз. В этом методе используются линейные симплекс-планы (см., например, [25]). Симплекс — простейшая фигура. На плоскости — это треугольник, в трехмерном пространстве — тетраэдр и т. д.¹

¹ Существо симплекс-планирования состоит в адаптивной оптимизации путем последовательного перехода от одного симплекса

Рассмотрим правильный симплекс S_0 с вершинами $d_1, d_2, \dots, d_{k+1}$ и центром c_0 . На каждой грани симплекса S_0 можно построить новый симплекс S_j с центром c_j и k вершинами $d_1, d_2, \dots, d_{j-1}, d_{j+1}, \dots, d_{k+1}$, принадлежащими множеству S_0 и одной новой вершиной d_j^* , являющейся зеркальным отображением точки d_j относительно грани, общей для обоих симплексов.

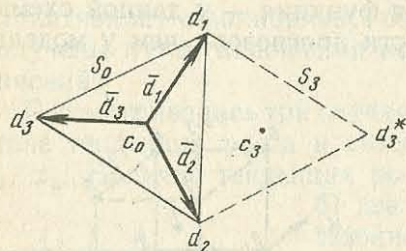


Рис. 4.11. Построение симплекса S_3 из симплекса S_0 в задаче с двумя независимыми параметрами

На рис. 4.11 выполнено построение симплекса S_3 из симплекса S_0 для случая $k = 2$. Чтобы найти ту или иную координату точки d_j^* , нужно взять удвоенное среднее из соответствующих координат точек $d_1, d_2, \dots, d_{j-1}, d_{j+1}, \dots, d_{k+1}$ и вычесть из

него соответствующую координату точки d_j . В векторном обозначении это запишется так:

$$d_j^* = \frac{2}{k}(d_1 + d_2 + \dots + d_{j-1} + d_{j+1} + \dots + d_{k+1}) - d_j.$$

Допустим теперь, что каждая точка d_i симплекса характеризует значения параметров сглаживания и каждой точке после накопления результатов (в каждый момент времени дается лишь один прогноз — одна проба) поставлена в соответствие абсолютная ошибка прогноза $|e_i|$. Отсюда следует, что до начала итеративной процедуры необходим $k + 1$ шаг для получения первых прогнозов и определения

их ошибок в вершинах исходного симплекса. Предположим далее, что $|e_p|$ — ошибка в точке d_p — наибольшее значение среди $k + 1$ значений, полученных для симплекса S_0 . Тогда ясно, что надо двигаться в направлении точки d_p^* . Можно доказать, что движение из центра симплекса за грань, находящуюся против точки d_p , будет совпадать с направлением крутого снижения, рассчитанным по результатам наблюдений в вершинах правильного симплекса.

Стратегия симплекс-планирования, в соответствии с которой осуществляется эволюция параметров сглаживания, формулируется в трех простых правилах:

1. Отобрать наибольшее значение $|e_p|$ из значений $|e_1|, |e_2|, \dots, |e_{k+1}|$, замеренных в точках, образующих симплекс S_0 . Перейти к новому симплексу S_p , заменив в симплексе S_0 точку d_p , соответствующую $|e_p|$, точкой d_p^* . Подсчитать прогноз на следующий момент времени, используя значения параметров, определяемые точкой d_p^* .

2. Если применение первого правила приводит к тому, что одна i -я точка (вершина) сохранилась в $k + 1$ последовательных симплексах, то следует прекратить применение этого правила и рассчитать прогноз на следующий момент времени, используя значения параметров, определяемые точкой d_i . Затем применять правило первое.

3. Если значение $|e_p|$ было наибольшим в предыдущем симплексе, а значение $|e_p^*|$ оказалось наибольшим в последующем симплексе, то, не возвращаясь к предыдущему симплексу, следует двигаться из последнего симплекса, отбросив (заменяв ее на зеркальную) точку, которой соответствует значение ошибки, ближайшее к наибольшему.

Таким образом, в этом методе четко указано, когда и куда двигаться. Это позволяет полностью возложить на ЭВМ процесс управления параметрами.

Д. Монтгомери экспериментировал с двухпараметрической моделью, учитывающей линейный тренд, и трехпараметрической моделью, аналогичной модели Уинтерса.

Исходный симплекс с произвольной ориентацией и длиной ребра обычно задают матрицей D , строки которой определяют вершины симплекса. Матрица D может быть сконструирована из последних k столбцов матрицы $\sqrt{k+1} \cdot N$, где N — любая ортогональная матрица размерности $(k+1) \cdot (k+1)$, элементы первого столбца которой равны.

са к другому, смежному, по направлению крутого снижения. Начинается процедура с вычисления в вершинах исходного симплекса значений целевой функции, на основании которых по установленному правилу принимается четкое решение о том, куда и когда двигаться.

В отличие от симплекс-планирования симплекс-метод в линейном программировании предлагает правила перехода от одной вершины симплекс-многогранника условий, определяемого в n -мерном пространстве наложенными линейными ограничениями, к соседней вершине того же симплекса, обеспечивающей возрастание линейной формы.

Как видим, эти два метода решают задачи, различающиеся по постановке.

В двухпараметрической модели, произвольно полагая $\alpha_1 = 0,100$ и $\alpha_2 = 0,075$, Монтгомери использовал матрицу:

$$D = \begin{bmatrix} \alpha_1 & \alpha_2 \\ 0,100 & 0,075 \\ 0,168 & 0,093 \\ 0,118 & 0,143 \end{bmatrix}.$$

Длина ребра в этом конкретном случае $l = [(d_i - d_j)'(d_i - d_j)]^{1/2} = 0,07$. Верхний и нижний пределы изменения параметров здесь также были установлены равными 0,95 и 0,05 соответственно.

В трехпараметрической модели, полагая $\alpha_1 = 0,100$, $\alpha_2 = 0,075$, $\alpha_3 = 0,100$, в качестве исходного он использовал симплекс, определяемый матрицей:

$$D = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ 0,100 & 0,075 & 0,100 \\ 0,166 & 0,092 & 0,117 \\ 0,117 & 0,141 & 0,117 \\ 0,117 & 0,092 & 0,166 \end{bmatrix}.$$

Здесь также $l = 0,07$, $0,95 \geq \alpha_1, \alpha_2, \alpha_3 \geq 0,05$.

Рассмотренный метод является альтернативным методу Робертса и Риды. Его преимущество в том, что он намного экономичнее с вычислительной точки зрения. К сожалению, сопоставление точности этих методов проведено не было.

Глава 5

АДАПТИВНЫЕ КОМБИНИРОВАННЫЕ МОДЕЛИ

В главе предпринята попытка построения более гибких комбинированных моделей, в состав (базовый набор) которых входит несколько более простых адаптивных моделей. В комбинированных моделях селективного типа на каждом шаге организован автоматический выбор по заданному критерию наилучшей модели из числа входящих в базовый набор. Таким образом, адаптация происходит на двух уровнях: по структуре или типу модели и по параметрам. В комбинированной гибридной модели прогноз формируется как взвешенная сумма прогнозов, полученных по альтернативным моделям. Веса при этом имеют адаптивный характер.

§ 1. АДАПТИВНАЯ СЕЛЕКТИВНАЯ МОДЕЛЬ

Вполне естественно предположить, что реальный процесс время от времени претерпевает коренные изменения. Изменяются уровень и динамические свойства ряда. Причем на одних участках сохраняется приблизительно постоянный уровень, на других ряд растет с определенной скоростью или характеризуется появлением ускорения. Поэтому всякая попытка подобрать какую-либо одну прогнозную модель для всего ряда приводит к некоей усредненной модели с чрезмерными дисперсиями оцениваемых параметров и ошибками прогнозирования. Когда изменение структуры происходит резким скачком, исследователю довольно просто отсечь устаревшие данные и строить модель только на свежей информации. Однако гораздо чаще изменения коренных свойств ряда происходят не сразу, а непрерывно. Причем ряд меняет тенденцию, делает зигзаги и бывает трудно, а при одновременном исследовании многих рядов

даже невозможно быстро анализировать такую динамику и заменять одну модель другой.

Мы предлагаем две процедуры адаптации структуры модели, которые могут выполняться на ЭВМ. Одна из них основывается на принципе непрерывной селекции.

Предположим, что рассматриваемый процесс генерируется моделью с постоянным уровнем, моделью с линейным ростом, квадратической моделью поочередно. Учитыв-

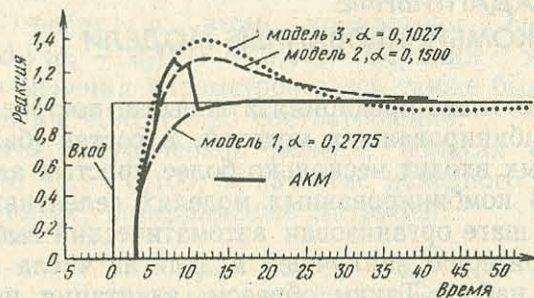


Рис. 5.1. Сравнение реакций на ступенчатое изменение уровня полиномиальных моделей Брауна нулевого, первого и второго порядков (модели 1, 2, 3 соответственно) и построенной на их основе адаптивной комбинированной модели (АКМ) селективного типа, $\tau=3$

вая это, построим адаптивную комбинированную модель (АКМ), которая включает в свой базовый набор три модели: экспоненциальной средней, линейного роста и квадратическую.

Вычисления будущих значений ряда осуществляются по каждой из них в отдельности, но в качестве прогноза выбирается расчетная величина, полученная по модели, наилучшим образом отражающей на данном временном интервале реальный процесс. Наилучшая модель избирается в соответствии с заданным критерием селекции. Наилучшей, естественно, считать ту модель, которая дала минимальную абсолютную ошибку прогноза текущего члена ряда при заданном периоде упреждения τ .

На рис. 5.1 изображена реакция на ступенчатое изменение АКМ, построенной по принципу селекции (выбора) той модели, которая имеет на текущий момент наименьшую ошибку прогноза. До ступенчатого скачка и через

$\tau - 1$ единиц времени после его появления все три модели, входящие в базовый набор предикторов, дают одинаковые τ -прогнозы. Это приводит к затруднению при выборе прогнозирующей модели. В такой ситуации для большей определенности можно взять модель линейного роста в качестве компромисса между моделью скользящей средней и квадратической моделью. В момент $t = 3$ становится ясно, что квадратическая модель дает результаты, более близкие к реальным точкам. АКМ переключается на нее. В дальнейшем из-за заметного превышения этой моделью уровня ступенчатого скачка АКМ переключается на модель линейного роста и модель экспоненциальной средней, но переключение это происходит не сразу, а с лагом $\tau = 3$. Такая задержка ухудшает общий результат. Ясно, что наиболее эффективно в данном случае критерий селекции должен работать при малых τ .

В рассмотренном примере АКМ испытывалась на входном потоке данных, содержащем детерминированный ступенчатый скачок. Если тот же эксперимент провести в более жизненных условиях, при наличии помех, то столь простой критерий селекции будет не всегда удовлетворителен, так как из-за случайной составляющей лучший результат будет то у одной, то у другой модели. Таким образом, нельзя в общем случае исходить только из последней ошибки прогнозирования. Необходимо учитывать некоторую их совокупность. Имеется довольно много вариантов построения критерия селекции. Предложим, например, два таких критерия.

Критерий К. Переключение на данную модель осуществлять тогда, когда K ее последних прогнозов являются наилучшими в сравнении с прогнозами по другим моделям, входящим в базовый набор АКМ.

Критерий В. Переключение на данную модель осуществлять тогда, когда ее экспоненциально сглаженный квадрат ошибки прогнозирования B минимален по сравнению с аналогичным показателем для остальных моделей в базовом наборе АКМ.

Критерий B формируется следующим образом:

$$B_t = (1 - \alpha_B) B_{t-1} + \alpha_B e_t^2(t - \tau),$$

где $0 < \alpha_B \leq 1$ — параметр сглаживания;
 $e_t(t - \tau)$ — ошибка прогноза, сделанного в момент $t - \tau$ на τ шагов вперед.

Легко заметить, что при $\alpha_B = 1$ критерий B эквивалентен критерию K при $K = 1$.

Если раньше отмечалось, что, в модели экспоненциального типа параметр сглаживания часто бывает близок к 0, то здесь этого сказать нельзя. Параметр B характеризует инерционность переключения, а мы уже видели, что промедление с переключением может привести к плохой работе АКМ, нарушить соответствие структуры модели динамике процесса. То же относится к критерию K .

По существу, величину B_i можно было бы назвать оценкой текущего значения дисперсии ошибки прогноза на t единиц времени вперед. Но учитывая, что α_B обычно велико, сглаживание квадратов ошибок будет недостаточным. Поэтому придавать величине B такой смысл в общем случае было бы неоправданным. При малых значениях параметра сглаживания эта интерпретация допустима и аналогичная процедура может использоваться для грубой оценки достоверных уровней прогнозов при прогнозировании нестационарных процессов.

Принцип непрерывной селекции весьма прост. Но в связи с тем, что переход с модели на модель зависит от предикторов, входящих в базовый набор АКМ, и специфических динамических свойств ряда, общий теоретический анализ эффективности АКМ затруднителен, и мы оставляем решающее слово за экспериментом (см. § 3).

§ 2. АДАПТИВНАЯ ГИБРИДНАЯ МОДЕЛЬ

Адаптивная селективная модель рассчитана на выбор одного предиктора из некоторого их множества. Но очевидно, что такой выбор однозначно и эффективно можно осуществлять, только если модели, входящие в базовый набор, существенно различны. Для тех случаев, когда в АКМ входят модели, дающие сравнительно близкие результаты, и селекция затруднена, можно предложить гибридную АКМ, прогноз по которой является взвешенной суммой прогнозов, полученных по входящим в нее предикторам. Веса прогнозов ω_i предлагаем взять адаптивными, обратно пропорциональными величине $B_{i,t}$, подсчитываемой, как и в селективной АКМ:

$$\omega_{i,t} = \frac{g_i}{B_{i,t}}.$$

Коэффициент пропорциональности g_i определяется из равенства суммы весов единице. Например, для базового набора из трех моделей веса будут определяться так:

$$\omega_{1,t} = \frac{B_{2,t} B_{3,t}}{B_{2,t} B_{3,t} + B_{1,t} B_{3,t} + B_{1,t} B_{2,t}};$$

$$\omega_{2,t} = \frac{B_{1,t} B_{3,t}}{B_{2,t} B_{3,t} + B_{1,t} B_{3,t} + B_{1,t} B_{2,t}};$$

$$\omega_{3,t} = \frac{B_{1,t} B_{2,t}}{B_{2,t} B_{3,t} + B_{1,t} B_{3,t} + B_{1,t} B_{2,t}}.$$

Получаемый в данном случае прогноз будет ближе к результату, получаемому то по одной, то по другой модели, являясь некоторой адаптивной равнодействующей.

В отличие от модели, рассмотренной ранее, гибридная модель осуществляет переключение более плавно, со множеством промежуточных положений. В результате получается непрерывный спектр возможных конструкций прогноза.

Для обеих процедур построения АКМ были составлены программы, которые содержат наборы по 5 моделей: экспоненциальной средней, линейного и квадратического роста (Брауна или Тригга—Лича) и две «наивные» модели:

$$\hat{x}_t(t) = x_t;$$

$$\hat{x}_t(t) = x_t + \tau(x_t - x_{t-1}).$$

Базовый набор АКМ формируется исследователем или автоматически из любого сочетания предикторов этого программного набора.

Автоматическое формирование основано на непрерывном сравнении средних квадратов ошибок прогнозов, полученных по различным моделям к текущему моменту времени. Плохие модели могут случайно приближаться к реальному процессу и кратковременно давать хорошие прогнозы, что приводит к переключению АКМ именно на них. В результате появляются большие ошибки, что ухудшает эффективность АКМ. Такие модели лучше сразу вывести из базового набора. Поэтому если базовый набор формируется автоматически, то из программного набора моделей в него включаются только те модели, средние квадраты ошибок которых не больше чем в m раз превосходят минимальную из них:

$$D_i(\tau) \leq m D_{\min}(\tau).$$

Обычно $m = 1,2—1,5$. Причем меньшее значение m соответствует малым τ , так как при увеличении периода упреждения разброс средних квадратов ошибок различных моделей, как правило, возрастает.

При автоматическом формировании базовый набор не является раз навсегда определенным. Если с течением времени меняется соотношение между средними квадратами ошибок, то меняется соответственно и базовый набор.

Таким образом, при прогнозировании модель подвергается двум испытаниям. Одно, основанное на анализе средних квадратов прошлых ошибок, дает ответ на вопрос о целесообразности включения модели в базовый набор предикторов. Другое, путем сравнения текущих свойств моделей по критерию B , определяет ту модель из базового набора, от которой можно ожидать наилучшего прогноза.

Принципы, изложенные в этом параграфе, являются достаточно общими и могут быть применены для построения АКМ с разнообразными базовыми наборами, в том числе включающими одинаковые по структуре, но отличающиеся значениями параметров модели. Число предикторов в наборе не целесообразно брать большим. Обычно достаточно 3—4 модели.

§ 3. ПРИМЕРЫ

Пример 5.1

Построим селективную АКМ. Включим в ее базовый набор предикторов полиномиальные модели Брауна нулевого, первого и второго порядков — модели многократного сглаживания с постоянной $\alpha = 0,1$. Испытаем эту модель на данных о курсе акций фирмы ИБМ и цене на золото. На рис. 5.2 представлены прогнозы курса акций на $\tau = 3$ по АКМ и по каждой из моделей, входящих в ее базовый набор.

Об эффективности процедуры селекции по критерию B можно судить по среднему квадрату ошибок прогнозирования в зависимости от τ (рис. 5.3, 5.4). Как видим, применение процедуры селекции для малых τ оказывается полезным и приводит к уменьшению среднего квадрата ошибок прогноза.

Из графиков также видно, что выбор моделей для базового набора довольно трудное дело. Казалось бы, что чем лучше работает модель отдельно, тем полезнее будет ее включение в базовый набор. Но из рис. 5.2 следует,

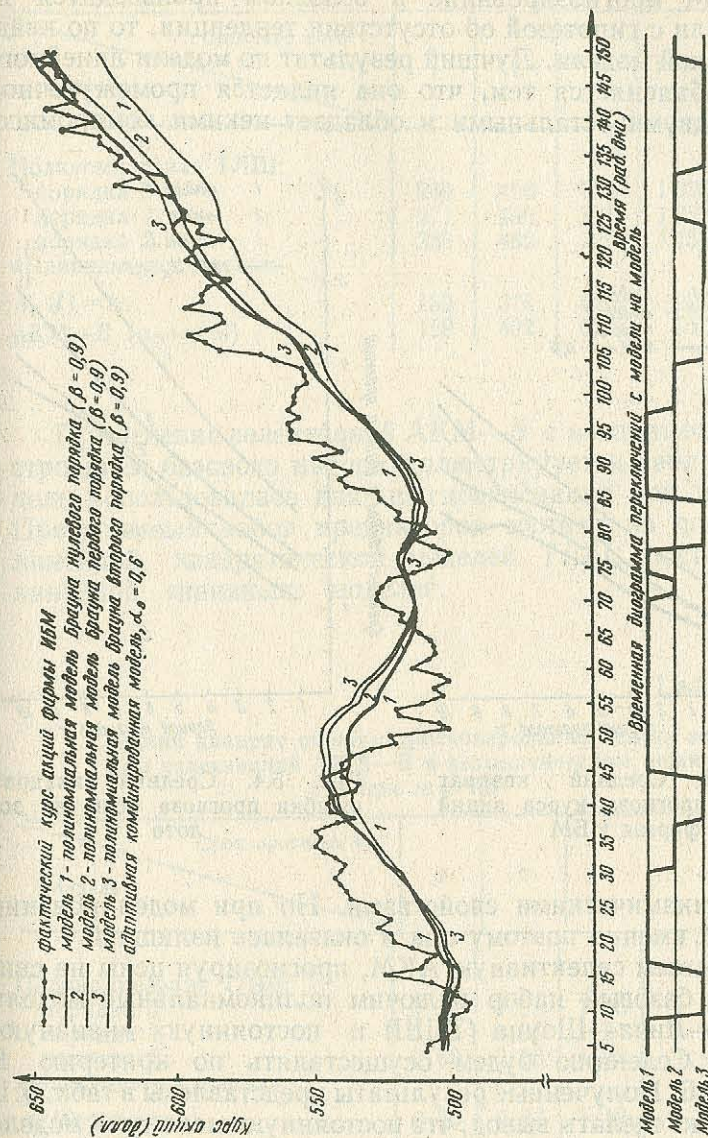


Рис. 5.2. Прогнозирование по селективной АКМ, $\tau = 3$

что, хотя модель линейного роста является более точной по сравнению с остальными двумя, при моделировании по АКМ прогнозирование в основном производится то по модели с гипотезой об отсутствии тенденции, то по квадратической модели. Лучший результат по модели линейного роста объясняется тем, что она является промежуточной между двумя остальными и обладает некими компромисс-

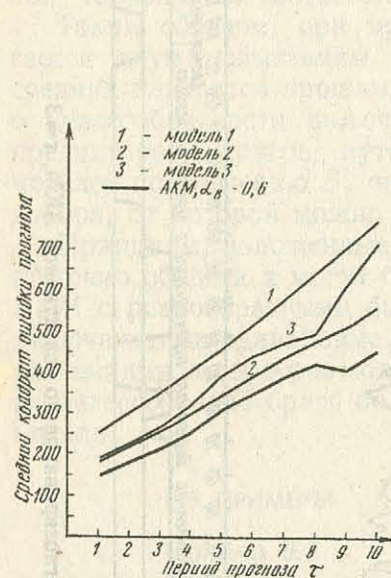


Рис. 5.3. Средний квадрат ошибки прогноза курса акций фирмы ИБМ

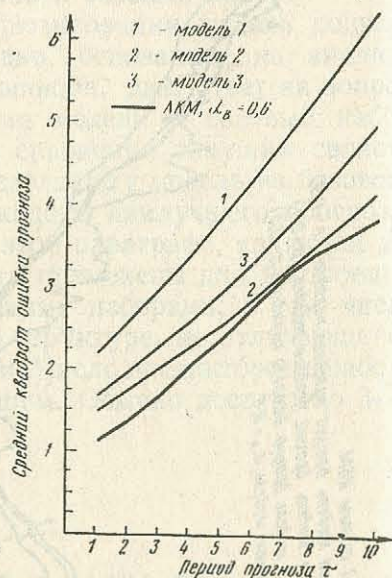


Рис. 5.4. Средний квадрат ошибки прогноза цены на золото

ными динамическими свойствами. Но при моделировании по АКМ именно поэтому она и оказалась излишней.

Испытаем селективную АКМ, прогнозируя цены на свинец¹. В базовый набор включим полиномиальные модели Тригга—Лича—Шоуна (ТЛШ) и постоянную «наивную» модель. Селекцию будем осуществлять по критерию B с $\alpha_B = 0,6$. Полученные результаты представлены в табл. 5.1.

Можно сделать вывод, что постоянную «наивную» модель в ряде случаев полезно включать в базовый набор АКМ.

¹ Все стохастические временные ряды, используемые для испытания моделей, приведены в приложении № 4.

Таблица 5.1

Средний квадрат ошибок прогнозирования цен на свинец

Срок прогноза τ					
	1	2	3	4	5
Модель					
Полиномиальная ТЛШ:					
порядка 0	269	496	734	1 025	1 367
порядка 1	271	496	731	1 023	1 373
порядка 2	258	482	729	1 152	1 686
«Наивная» постоянная					
$\hat{x}_\tau(t) = x_t$	139	374	626	922	1 253
АКМ—В ($\alpha_B = 0,6$)	159	402	623	915	1 233

Применение селективной АКМ—В с автоматическим построением базового набора иллюстрируется табл. 5.2. Модель использовалась для прогнозирования цен на золото. Программный набор предикторов состоял из постоянной, линейной, квадратической моделей ТЛШ, постоянной и линейной «наивных» моделей.

Таблица 5.2

Средний квадрат ошибки прогнозирования цен на золото по селективной АКМ—В в автоматическом режиме при $m=1,3$

Срок прогноза τ					
	1	2	3	4	5
Модель					
Полиномиальная ТЛШ:					
порядка 0	0,57	0,95	1,26	1,52	1,79
порядка 1	0,60	0,95	1,21	1,42	1,62
порядка 2	0,62	0,96	1,22	1,43	1,63
«Наивная» постоянная					
$\hat{x}_\tau(t) = x_t$	0,50	0,87	1,18	1,45	1,70
«Наивная» линейная					
$\hat{x}_\tau(t) = x_t + \tau(x_t - x_{t-1})$	1,14	3,27	6,41	10,5	15,5
АКМ—В ($\alpha_B = 0,6$)	0,77	0,86	1,14	1,43	1,69

Как видим, модель неплохо справилась с выбором полезных предикторов и прогнозов. Можно отметить, что при $\tau = 1$ «наивная» линейная модель явно портит общий результат АКМ. В таких случаях рекомендуются два пути. Во-первых, заведомо плохую модель можно сразу устранить и перейти на режим принудительного задания базового набора предикторов. Во-вторых, при $\tau = 1$ можно уменьшить m до 1,2. Это позволит осуществлять более строгий отбор предикторов в автоматическом режиме.

Пример 5.2

В качестве примера применения гибридной АКМ рассмотрим прогнозирование курса акций фирмы ИБМ. За базовые взяты линейная модель ТЛШ и постоянная «наивная» модель. Прогнозы характеризуются такими дисперсиями ошибок (табл. 5.3):

Таблица 5.3

Средние квадраты ошибок прогнозирования курса акций по гибридной АКМ

Модель \ τ	1	2	3	4	5	6	7	8	9	10
ТЛШ линейная	61,4	100	132	167	205	247	299	345	369	403
Постоянная «наивная»	40,9	79	111	148	182	219	267	323	379	433
АКМ ($\alpha_B = 0,5$)	40,9	82	117	155	190	227	281	336	366	400

Как видно из табл. 5.3, гибридная АКМ в случаях, когда трудно отдать предпочтение той или иной модели, помогает уйти от максимальной возможной ошибки.

Пример 5.3

На базе полиномиальных моделей нулевого, первого и второго порядков (модели № 1, 2, 3 соответственно) с параметром α , регулируемым методом эволюции, была построена селективная АКМ с выбором предиктора на каждом шаге по критерию $K = 1$. Алгоритм программы построен таким образом, что комбинированная модель рассматривается как одна из равноправных моделей наряду с предикторами,

входящими в ее базовый набор. Реальный прогноз делается по предиктору, отобранному после анализа работы этих четырех моделей на ретроспективных данных.

Результаты испытаний этих моделей при $\tau = 1$ и $\tau = 2$ сведены в табл. 5.4 и 5.5. Там же для сравнения приведены данные по моделям Брауна. Прямоугольником выделен наилучший результат по строке, а наилучший результат, полученный методом эволюции, выделен жирным шрифтом. При $\tau = 1$ незначительное преимущество имеют модели Брауна. При $\tau = 2$ это преимущество утрачивается и оба метода дают примерно равную точность. При этом с вычислительной точки зрения метод эволюции предпочтительнее метода Брауна, при котором требуется процедура оптимизации параметра α .

* * *

Сделаем некоторые выводы. В главах 4 и 5 рассмотрены попытки эмпирического совершенствования адаптивных моделей с помощью автоматической регулировки параметров сглаживания, предложена простая схема адаптации структуры модели методом селекции или путем симбиоза нескольких предикторов. Показана эффективность ряда моделей при краткосрочном прогнозировании. В силу простоты этих моделей нельзя ожидать от них большой точности. Видимо, почти в каждом случае можно отыскать способ получения более точных прогнозов. Однако надо иметь в виду, что основное назначение этих моделей — автоматическая обработка большого количества рядов. Поэтому решалась не столько задача поиска оптимальных систем, сколько проблема создания универсальных предикторов, пригодных для практической обработки рядов с различной динамикой. Предлагаемые методы могут быть также рекомендованы для случаев, когда нельзя раз и навсегда выбрать какую-либо одну определенную структуру модели. Именно поэтому они могут найти применение в ряде исследований.

Применение моделей с адаптивной структурой вызывает множество вопросов, ответить на которые пока трудно. Это прежде всего вопросы, касающиеся выбора исходных предикторов для формирования базового набора и регулирования инерционности переключения с модели на модель. Однако это направление, на наш взгляд, перспективно для дальнейших исследований.

Сравнение метода Брауна с методом эволюции
по средним квадратам ошибок ретроспективных прогнозов,
сделанных на один шаг вперед ($\tau=1$)

№ п/п	Выборка из ряда	Метод Брауна			Метод эволюции			
		Число точек	№ модели	средний квадрат ошибки	средний квадрат ошибки			комбинированная
					модель № 1	модель № 2	модель № 3	
1	2	3	4	5	6	7	8	9
1	Курс акций фирмы ИВМ, дол.	144	1	41,427	43,419	64,315	80,240	44,728
2	Курс акций фирмы ИВМ (часть 1)	48	1	32,956	37,565	47,913	73,710	34,863
3	Курс акций фирмы ИВМ (часть 2)	48	1	46,294	46,445	72,234	79,970	49,308
4	Курс акций фирмы ИВМ (часть 3)	48	1	43,419	53,252	58,787	68,305	49,516
5	Цена на золото, дол. за унцию	360	1	0,50123	0,51263	0,64924	0,76728	0,65877
6	Цена на золото (часть 1)	60	1	0,08265	0,09869	0,09877	0,14866	0,09719
7	Цена на золото (часть 2)	60	1	0,10480	0,10958	0,14941	0,27336	0,13601
8	Цена на золото (часть 3)	60	2	0,02884	0,02913	0,02902	0,03089	0,03224
9	Цена на золото (часть 4)	60	1	0,14521	0,16726	0,18879	0,30078	0,14726
10	Цена на золото (часть 5)	60	1	0,84637	0,96741	0,93089	0,98194	0,94999

11	Цена на золото (часть 6)	60	1	1,6928	1,4022	1,9341	2,6717	1,5401
12	Поставка минеральных удобрений сельскому хозяйству, тыс. т усл. ед.	30	2	28,148 · 10 ⁵	201,978 · 10 ⁵	41,999 · 10 ⁵	30,898 · 10 ⁵	32,579 · 10 ⁵
13	Производство электроэнергии, млрд. кВт·ч	31	2	26,593	1888,7	22,829	31,063	21,722
14	Производство гидроэлектроэнергии, млрд. кВт·ч	31	2	27,456	54,920	37,274	38,893	32,424
15	Сумма вкладов населения в сберкассы, млн. руб	26	2	10,92 · 10 ⁵	307,6 · 10 ⁵	19,04 · 10 ⁵	9,872 · 10 ⁵	11,86 · 10 ⁵
16	Численность городского населения, млн. чел.	27	2	0,32741	28,067	0,56278	0,81668	0,43933
17	Протяженность нефте- и нефтепродуктопроводов, тыс. км	21	2	1,4529	13,872	1,6304	1,4723	1,5126
18	Грузооборот нефте- и нефтепродуктопроводов, ткм	21	3	168,73	2757,0	322,11	296,64	298,66
19	Урожайность зерна, ц с га	30	2	2,2653	3,2946	3,1126	3,0164	3,2191
20	Урожайность картофеля, ц с га	30	1	170,716	162,68	186,43	202,33	198,12
21	Урожайность овощей, ц с га	30	2	78,732	121,31	90,85	100,58	117,68
22	Производство яиц, млрд. шт.	30	2	1,9412	9,2810	2,2714	2,2063	1,7782

Сравнение метода Брауна с методом эволюции по средним квадратам ошибок
ретроспективных прогнозов, сделанных на два шага вперед ($\tau=2$)

№ п/п	Выборка из ряда	Число точек	Метод Брауна		Метод эволюции			
			№ модели	средний квадрат ошибки	средний квадрат ошибки			
					модель № 1	модель № 2	модель № 3	
1	2	3	4	5	6	7	8	9
1	Курс акций фирмы ИБМ, дол.	144	1	80,153	81,478	96,944	144,213	85,298
2	Курс акций фирмы ИБМ (часть 1)	48	1	60,628	66,315	67,956	109,710	58,396
3	Курс акций фирмы ИБМ (часть 2)	48	1	89,277	87,494	119,665	117,536	95,678
4	Курс акций фирмы ИБМ (часть 3)	48	1	80,414	103,062	79,826	110,895	85,628
5	Цена на золото, дол. за унцию	360	1	0,86646	0,89830	1,10362	1,28570	1,03500
6	Цена на золото (часть 1)	60	1	0,12906	0,18163	0,19527	0,21458	0,15003
7	Цена на золото (часть 2)	60	1	0,21768	0,24762	0,35511	0,40347	0,25994
8	Цена на золото (часть 3)	60	2	0,03570	0,037352	0,033487	0,037917	0,040031
9	Цена на золото (часть 4)	60	2	0,26517	0,32079	0,31179	0,48198	0,36845
10	Цена на золото (часть 5)	60	2	1,2007	1,77632	1,32608	1,32086	1,63410
11	Цена на золото (часть 6)	60	1	2,0442	1,9197	3,1570	3,8169	2,0578
12	Поставка минеральных удобрений сельскому хозяйству, тыс. т усл. ед.	30	2	96,117·10 ⁵	684,509·10 ⁵	98,276·10 ⁵	189,329·10 ⁵	99,061·10 ⁵

Продолжение

п/п №	Выборка из ряда	Число точек	Метод Брауна		Метод эволюции			
			№ модели	средний квадрат ошибки	средний квадрат ошибки			
					модель № 1	модель № 2	модель № 3	
								комбинированная
1	2	3	4	5	6	7	8	9
13	Производство электроэнергии, млрд. кВт·ч	31	2	113,153	6937,3	82,446	51,701	95,061
14	Производство гидроэлектроэнергии, млрд. кВт·ч	31	2	59,887	158,68	60,268	51,062	51,592
15	Сумма вкладов населения в сберкаппы, млн. руб	26	2	54,069 · 10 ⁵	1014,2 · 10 ⁵	80,81 · 10 ⁵	44,59 · 10 ⁵	48,83 · 10 ⁵
16	Численность городского населения, млн. чел.	27	2	1,46924	86,821	0,95485	1,28914	1,27009
17	Протяженность нефте- и нефтепродуктопроводов, тыс. км	21	2	3,3538	44,1430	2,2234	1,7879	2,0119
18	Грузооборот нефте- и нефтепродуктопроводов, ткм	21	3	829,8	8836,7	1051,5	599,1	702,2
19	Урожайность зерна, ц с га	30	2	2,2758	3,4959	3,3335	3,9829	4,1992
20	Урожайность картофеля, ц с га	30	1	166,539	188,17	215,31	250,76	225,86
21	Урожайность овошей, ц с га	30	2	76,065	173,23	93,43	110,26	133,86
22	Производство яиц, млрд. шт.	30	2	6,1755	26,318	3,2831	4,1259	4,9086

Глава 6

БАЙЕСОВСКИЙ ПОДХОД К КРАТКОСРОЧНОМУ ПРОГНОЗИРОВАНИЮ

В основе подхода, рассматриваемого в этой главе, лежит гипотеза о том, что исследуемый временной ряд генерируется не одной, а несколькими простейшими вероятностными моделями поочередно, т. е. речь идет, как и в предыдущей главе, о построении модели с множеством состояний. Причем переключение с одного состояния на другое является вероятностным процессом и соответствует то появлению увеличенных случайных возмущений, то ступенчатым изменениям уровня ряда, то скачкам в динамике их роста. При таком подходе события, имеющие случайный характер, получают ясное отражение в модели, а последовательно поступающие данные используются для подсчета апостериорных вероятностей и анализа ситуаций.

§ 1. МОДЕЛЬ С МНОЖЕСТВОМ СОСТОЯНИЙ

Если ввести обозначения: x_t — текущее значение ряда; $a_{1,t}$ — текущий уровень после исключения сезонных колебаний; $a_{2,t}$ — текущее значение коэффициента линейного роста; f_t — текущее значение коэффициента сезонности, то обычно принимается гипотеза, что ряд генерируется следующим образом:

$$\begin{aligned} x_t &= a_{1,t} f_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, D_\varepsilon); \\ a_{1,t} &= a_{1,t-1} + a_{2,t} + u_t, \quad u_t \sim N(0, D_u); \\ a_{2,t} &= a_{2,t-1} + v_t, \quad v_t \sim N(0, D_v), \end{aligned} \quad (6.1)$$

где ε_t — шум;
 u_t — возмущение уровня;
 v_t — возмущение коэффициента линейного роста.

Случайные компоненты ε , u , v предполагаются независимо и нормально распределенными с нулевыми средними и известными дисперсиями D_ε , D_u , D_v .

Этот тип модели, в котором как уровень, так и параметр линейного роста подвержены непрерывным случайным возмущениям, часто оказывается полезным в прогнозировании практических временных рядов. Харрисоном было показано [67], что при генерировании ряда моделью типа (6.1) с постоянными дисперсиями D_ε , D_u , D_v и без эффекта сезонности оптимальный предиктор (в смысле минимума квадратов ошибок) в точности эквивалентен модели Хольта:

$$\begin{aligned} e_t &= x_t - (\hat{a}_{1,t-1} + \hat{a}_{2,t-1}); \\ \hat{a}_{1,t} &= \hat{a}_{1,t-1} + \hat{a}_{2,t-1} + \alpha_1 e_t; \\ \hat{a}_{2,t} &= \hat{a}_{2,t-1} + \alpha_2 e_t, \end{aligned}$$

где постоянные сглаживания α_1 и α_2 теоретически являются функциями отношений дисперсий D_u/D_ε и D_v/D_ε .

Широкое применение этой модели для краткосрочного прогнозирования, как и линейной модели Брауна (см. гл. 1), основанной на экспоненциально взвешенной регрессии и являющейся частным случаем модели Хольта, фактически свидетельствует о приемлемости гипотезы (6.1) для многих временных рядов.

Как уже отмечалось, важной проблемой при построении таких прогностических систем является выбор параметров α_1 и α_2 , который определяет ее чувствительность. Чувствительная система быстро реагирует на реальные изменения, а нечувствительная не реагирует на шум и случайные отклонения. Оба свойства полезны, но находятся в противоречии, и вследствие того, что фильтрация шума и случайных возмущений требуется чаще, чем адаптация к новому уровню или скорости роста, обычно на практике делают эту систему малочувствительной. Однако, когда такая система встречается с серьезными изменениями в динамике ряда, она не сигнализирует о их появлении, а лишь медленно двигается к новому уровню. В итоге в течение определенного периода ошибки прогнозов могут достигать значительной величины.

В гл. 4 были рассмотрены попытки сконструировать адаптивные системы, которые одновременно сигнализируют о появлении необычных ситуаций и повышают чувстви-

тельность модели так, что необходимость вмешательства человека значительно уменьшается. В частности, Тригг и Лич предложили систему, адаптивную к изменениям в уровне, но не обладающую такими свойствами по отношению к резким скачкам в коэффициенте роста. К тому же их система характеризуется нежелательной реакцией на импульсные возмущения. В данной главе рассматривается попытка преодолеть эти трудности с помощью байесовского подхода, использованного П. Харрисоном и К. Стевенсом [66] при разработке модели с множеством состояний.

Возвратимся к гипотезе (6.1). Генерирующая модель содержит случайную компоненту ε_t , которая оказывает воздействие только на текущее наблюдение и не отражается на будущих значениях a_1 и a_2 , а также возмущения u_t и v_t , значения которых влияют на последующее движение системы. Например, необычно большое значение u_t означало бы ступенчатый переход на новый уровень, а необычно большое значение v_t вызвало бы скачок в коэффициенте линейного роста; в то же время чрезмерно большое значение ε_t выражалось бы в появлении случайного импульсного отклонения, не оказывающего, однако, никакого влияния на будущее поведение ряда. Такие чрезмерно большие значения случайных компонент могут встречаться в реальных данных, хотя вероятность их появления обычно мала.

Может быть выдвинуто предположение, что для каждой из случайных величин ε_t , u_t , v_t имеется не одно, а некоторое число альтернативных распределений, в соответствии с которыми генерируются их текущие значения. Формально в этом случае генерирующая модель имеет следующий вид.

Пусть k — число состояний,

Pr_j — вероятность наступления j -го состояния, $j = 1, \dots, k$.

Тогда, если в момент t система находится в j -м состоянии, случайные компоненты ε_t , u_t , v_t генерируются процессами:

$$\begin{aligned}\varepsilon_t &\sim N(0, D_\varepsilon^{(j)}); \\ u_t &\sim N(0, D_u^{(j)}); \\ v_t &\sim N(0, D_v^{(j)}).\end{aligned}\quad (6.2)$$

Таким образом, система имеет параметры:

$$Pr_j, D_\varepsilon^{(j)}, D_u^{(j)}, D_v^{(j)}, j = 1, \dots, k.$$

В дальнейшем будем различать четыре состояния процесса, иллюстрируемые рис. 6.1, а именно: «отсутствие изменений», «ступенчатое изменение», «изменение в коэффициенте

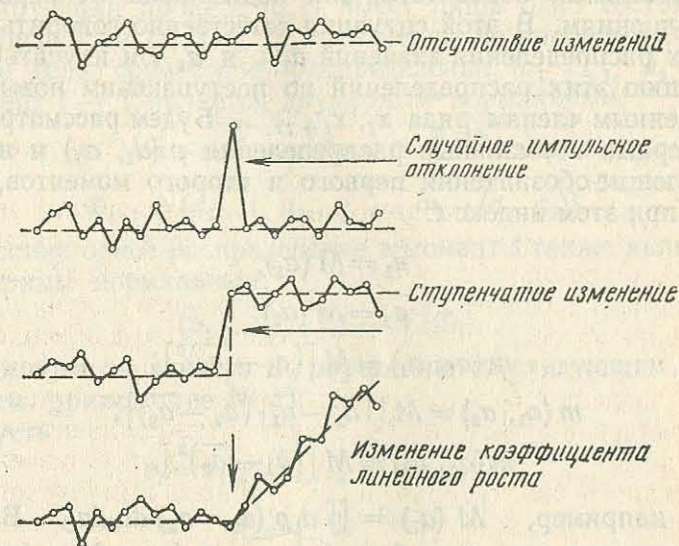


Рис. 6.1. Основные состояния процесса, генерирующего временной ряд

циенте линейного роста», «случайное импульсное отклонение». Качественная характеристика дисперсий, необходимых для описания этих состояний, приведена в табл. 6.1.

Таблица 6.1

Характеристика дисперсий генерирующего процесса с четырьмя состояниями

Состояние	D_ε	D_u	D_v
Отсутствие изменений	Нормальная	Ноль	Ноль
Ступенчатое изменение	Нормальная	Большая	Ноль
Изменение коэффициента линейного роста	Нормальная	Ноль	Большая
Случайное импульсное отклонение	Большая	Ноль	Ноль

Совершенно очевидно, что действительные значения $a_{1,t}$ и $a_{2,t}$ в нашем генерирующем процессе остаются неизвестными, более того, они подвержены непрерывным возмущениям. В этой ситуации естественно говорить о законах распределения значений $a_{1,t}$ и $a_{2,t}$ и изучать модификацию этих распределений по поступающим новым наблюдаемым членам ряда $x_t, x_{t+1}, \dots$. Будем рассматривать двумерные нормальные распределения $p(a_1, a_2)$ и введем следующие обозначения первого и второго моментов, опускаемая при этом индекс t :

$$\begin{aligned}\bar{a}_1 &= M(a_1); \\ \bar{a}_2 &= M(a_2); \\ m(a_1, a_1) &= M[(a_1 - \bar{a}_1)^2]; \\ m(a_1, a_2) &= M[(a_1 - \bar{a}_1)(a_2 - \bar{a}_2)]; \\ m(a_2, a_2) &= M[(a_2 - \bar{a}_2)^2],\end{aligned}\quad (6.3)$$

где, например, $M(a_1) = \iint a_1 p(a_1, a_2) da_1 da_2$. Введем также вектор моментов $\Phi = \{\bar{a}_1, \bar{a}_2, m(a_1, a_1), m(a_1, a_2), m(a_2, a_2)\}$.

Индексы при Φ будем понимать как относящиеся ко всем его компонентам, т. е.

$$\Phi^{(t,l)} = \{\bar{a}_{1,t}^{(l)}, \bar{a}_{2,t}^{(l)}, \dots, m_t^{(l,l)}(a_2, a_2)\}.$$

Запись $(a_1, a_2) \sim N(\Phi)$ будет означать, что пара величин (a_1, a_2) имеет совместное двумерное нормальное распределение с параметрами Φ .

Для обозначения смеси k двумерных нормально распределенных совокупностей с параметрами $\Phi^{(i)}, i = 1, \dots, k$, будем использовать запись

$$(a_1, a_2) \sim \sum_{i=1}^k v^{(i)} N(\Phi^{(i)}).$$

Эта запись указывает, что значения двух случайных величин могут с вероятностью $v^{(i)}$ принадлежать совокупности с i -м распределением.

Следует обратить внимание на то, что распределение $(a_{1,t}, a_{2,t})$ — это байесовское апостериорное (по отноше-

нию к текущему наблюдению x_t) распределение. В ряде случаев будем отмечать это записью $(a_{1,t}, a_{2,t} | x_t)$.

Перейдем к проблеме вычисления апостериорных вероятностей. Рассмотрим сначала простейший случай, когда генерирующий процесс (6.1) имеет лишь одно состояние, т. е. дисперсии D_e, D_u, D_v неизменны во времени. В этом случае если совместное распределение $(a_{1,t-1}, a_{2,t-1})$ в момент $t-1$ является двумерным нормальным:

$$(a_{1,t-1}, a_{2,t-1} | x_{t-1}) \sim N(\Phi_{t-1}),$$

где

$$\Phi_{t-1} = \{a_{1,t-1}, a_{2,t-1}, \dots, m_{t-1}(a_2, a_2)\},$$

то апостериорное распределение в момент t также является двумерным нормальным:

$$(\bar{a}_{1,t}, \bar{a}_{2,t} | x_t) \sim N(\Phi_t),$$

где компоненты вектора Φ_t определяются следующим образом (см. приложение № 2).

Пусть

$$e_t = x_t - (\bar{a}_{1,t-1} + \bar{a}_{2,t-1}) f_t; \quad (6.4)$$

$$R = \begin{pmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{pmatrix}, \quad (6.5)$$

где

$$\begin{aligned}r_{11} &= m_{t-1}(a_1, a_1) + 2m_{t-1}(a_1, a_2) + m_{t-1}(a_2, a_2) + D_u + D_v; \\ r_{12} &= m_{t-1}(a_1, a_2) + m_{t-1}(a_2, a_2) + D_v; \\ r_{22} &= m_{t-1}(a_2, a_2) + D_v.\end{aligned}\quad (6.6)$$

Далее, пусть

$$\left. \begin{aligned}D_e &= f_t^2 r_{11} + D_e; \\ \alpha_1 &= f_t r_{11} / D_e; \\ \alpha_2 &= f_t r_{12} / D_e.\end{aligned} \right\} \quad (6.7)$$

Тогда параметры совместного апостериорного распределения $(a_{1,t}, a_{2,t})$ в момент t можно вычислить по формулам:

$$\begin{aligned}\bar{a}_{1,t} &= \bar{a}_{1,t-1} + \bar{a}_{2,t-1} + \alpha_1 e_t; \\ \bar{a}_{2,t} &= \bar{a}_{2,t-1} + \alpha_2 e_t; \\ m_t(a_1, a_1) &= r_{11} - \alpha_1^2 D_e; \\ m_t(a_1, a_2) &= r_{12} - \alpha_1 \alpha_2 D_e; \\ m_t(a_2, a_2) &= r_{22} - \alpha_2^2 D_e.\end{aligned}\quad (6.8)$$

Введем для всего алгоритма пересчета моментов, заданного выражениями (6.4)—(6.8), обозначение:

$$\Phi_t = B(\Phi_{t-1}; D_e, D_u, D_v). \quad (6.9)$$

Функция B выражает вычисление байесовского апостериорного распределения в момент t через моменты апостериорного распределения в момент $t-1$ и дисперсии генерирующего процесса.

Теперь рассмотрим генерирующий процесс с множеством состояний, который описывается смесью априорных (относительно момента t) распределений:

$$(a_{1,t-1}, a_{2,t-1}) \sim \sum_{i=1}^k Q_{t-1}^{(i)} N(\Phi_{t-1}^{(i)}), \quad (6.10)$$

где $Q_{t-1}^{(i)}$ — вероятность (апостериорная по отношению к наблюдению x_{t-1}) того, что процесс в момент $t-1$ был в состоянии i ;

$\Phi_{t-1}^{(i)}$ — вектор моментов i -го распределения, описывающего состояние i в момент $t-1$.

Используя результаты, полученные для генерирующего процесса с одним состоянием, для описания каждого из k состояний $j = 1, \dots, k$ по очереди, получаем:

$$(a_{1,t}, a_{2,t} | x_t, s_t = j, s_{t-1} = i) \sim N(\Phi_t^{(i,j)}), \quad (6.11)$$

где s_t — номер состояния в момент t ;

$$\Phi_t^{(i,j)} = B(\Phi_{t-1}^{(i)}; D_e^{(j)}, D_u^{(j)}, D_v^{(j)}). \quad (6.12)$$

Здесь использована B -функция уравнения (6.9). Поэтому полное апостериорное распределение может быть записано как

$$(a_{1,t}, a_{2,t} | x_t) \sim \sum_{i,j} P_t^{(i,j)} N(\Phi_t^{(i,j)}), \quad (6.13)$$

где $\Phi_t^{(i,j)}$ задается уравнением (6.12), а $P_t^{(i,j)}$ — вероятность (апостериорная относительно наблюдаемого значения x_t) того, что процесс в момент $t-1$ был в i -м состоянии и теперь переходит в j -е состояние.

Возникает вопрос о том, как оценить вероятность перехода генерирующего процесса из состояния i в состояние j . Рассмотрим проблему в общем виде. Пусть событие z может наблюдаться лишь при условии, что произошло какое-нибудь событие H_i из числа несовместных событий $H_1, H_2, \dots$,

H_n , вероятности которых известны. События H_i обычно называют гипотезами относительно z . Вероятность реализации i -й гипотезы при условии, что наблюдается событие z , определяется соотношением Байеса:

$$P(H_i | z) = \frac{P(H_i) P(z | H_i)}{P(z)},$$

где $P(z)$ — полная вероятность наступления события z .

Используя эту формулу в нашем конкретном случае, можно получить оценку $P_t^{(i,j)}$ следующим образом:

$$\begin{aligned} P_t^{(i,j)} &= P(s_t = j, s_{t-1} = i | x(t)) = K p(x_t | s_t = j, \\ &\quad s_{t-1} = i, x(t-1)) P(s_t = j | s_{t-1} = i, \\ &\quad x(t-1)) P(s_{t-1} = i | x(t-1)) = \\ &= K p(x_t | s_t = j, s_{t-1} = i, x(t-1)) Pr_j Q_{t-1}^{(i)} = \\ &= K \sqrt{\frac{1}{2\pi D_e^{(i,j)}}} \exp \left\{ -\frac{[x_t - f_t(\bar{a}_{1,t-1}^{(i)} + \bar{a}_{2,t-1}^{(i)})]^2}{2D_e^{(i,j)}} \right\} Pr_j Q_{t-1}^{(i)}, \end{aligned} \quad (6.14)$$

где $x(t)$ — вектор всех наблюдений до t включительно;

$$K = 1/p(x_t | x(t-1));$$

$$D_e^{(i,j)} = f_t^2 r_{11}^{(i,j)} + D_e^{(i)};$$

$$\begin{aligned} r_{11}^{(i,j)} &= m_{t-1}^{(i)}(a_1, a_1) + 2m_{t-1}^{(i)}(a_1, a_2) + m_{t-1}^{(i)}(a_2, a_2) + \\ &\quad + D_u^{(i)} + D_v^{(j)}; \end{aligned}$$

буква P обозначает вероятность, буква p — плотность вероятности.

В этом выражении использовалась функция плотности вероятности для нормального распределения

$$p(x) = \frac{1}{\sqrt{2\pi D_x}} \exp \left[-\frac{(x - \bar{x})^2}{2D_x} \right],$$

а также то, что

$$\begin{aligned} P(x_t | s_t = j, s_{t-1} = i, x(t-1)) &= \\ &= \frac{p(x_t | s_t = j, s_{t-1} = i, x(t-1)) \Delta x_t}{p(x_t | x(t-1)) \Delta x_t}, \end{aligned}$$

где Δx_i — некоторый отрезок переменной x , содержащий точку x_i , в пределах которого плотность распределения вероятности можно считать постоянной.

Величину K в выражении (6.14) можно считать нормализующим коэффициентом и найти ее из условия $\sum_{i,j} P_i^{(i,j)} = 1$, что полностью определяет $P_i^{(i,j)}$ и апостериорное распределение (6.13).

Здесь возникает одно серьезное практическое затруднение. Начиная с k -компонентного априорного распределения, представленного в выражении (6.10), мы приходим к k^2 -компонентному апостериорному распределению в (6.13). Продолжая этот процесс, мы должны по получении последующих наблюдаемых значений $x_{t+1}, x_{t+2}, \dots$, переходить к апостериорным распределениям с k^3, k^4 и т. д. компонентами. Это очевидно неприемлемо с вычислительной точки зрения, так как требования к объему памяти и к машинному времени становились бы чрезмерными уже после небольших рядов наблюдений. Поэтому авторы рассматриваемого метода предлагают некоторые приближения для удержания требований к ресурсам компьютера в разумных рамках. Имеется несколько путей уменьшения числа компонент апостериорного распределения. Один из них заключается в следующем.

Все множество компонент смеси распределений в (6.13) с одинаковым текущим состоянием j в целом заменяется на единственное двумерное нормальное распределение, первые и вторые моменты которого получают соответствующей процедурой взвешивания, т. е.

$$(a_{1,t}, a_{2,t} | x_t) \sim \sum_{i=1}^k Q_i^{(j)} N(\Phi_i^{(j)}), \quad (6.15)$$

где

$$\left. \begin{aligned} Q_i^{(j)} &= \sum_l P_i^{(l,j)}; \\ \bar{a}_{1,t}^{(j)} &= \sum_l P_i^{(l,j)} \bar{a}_{1,t}^{(l,j)} / Q_i^{(j)}; \\ \bar{a}_{2,t}^{(j)} &= \sum_l P_i^{(l,j)} \bar{a}_{2,t}^{(l,j)} / Q_i^{(j)}. \end{aligned} \right\} \quad (6.16')$$

Выражения для вторых моментов более громоздки. В качестве примера приведем одно из них:

$$m_i^{(j)}(a_1, a_1) = \sum_l P_i^{(l,j)} [m_{1,t}^{(l,j)}(a_1, a_1) + (\bar{a}_{1,t}^{(l,j)} - \bar{a}_{1,t}^{(j)})^2] / Q_i^{(j)}. \quad (6.16'')$$

Рассмотренным способом удастся сохранить наиболее ценную информацию о текущем состоянии процесса и «сжать» апостериорное распределение $(a_{1,t}, a_{2,t} | x_t)$ в выражении (6.15), приведя его к тому же виду, что и в выражении (6.10) в начале цикла обновления.

Сделаем краткий обзор метода по этапам цикла обновления.

Этап 1. После получения наблюдения x_{t-1} предполагается, что апостериорное распределение

$$(a_{1,t-1}, a_{2,t-1} | x_{t-1})$$

является смесью двумерных нормальных распределений:

$$(a_{1,t-1}, a_{2,t-1} | x_{t-1}) \sim \sum_{i=1}^k Q_{i-1}^{(j)} N(\Phi_{i-1}^{(j)}),$$

где как

$$Q_{i-1}^{(j)} = P(s_{t-1} = i | x_{t-1}),$$

так и параметры $\Phi_{i-1}^{(j)}$ известны.

Этап 2. Априорно, до получения члена x_t , предполагается, что процесс в момент t может оказаться в любом из возможных состояний, каждое из которых характеризуется параметрами:

Pr_j — вероятность состояния j ;

$D_e^{(j)}, D_u^{(j)}, D_v^{(j)}$ — дисперсии генерирующего процесса в состоянии j .

Этап 3. Рассматривая все возможные переходы процесса из одного состояния в другое при движении от момента $t-1$ к моменту t и учитывая вероятности этих переходов, получаем выражение апостериорного распределения через смесь k^2 двумерных распределений:

$$(a_{1,t}, a_{2,t} | x_t) \sim \sum_{i,j=1}^k P_i^{(i,j)} N(\Phi_i^{(i,j)}),$$

где

$$\Phi_i^{(i,j)} = B(\Phi_{i-1}^{(j)}; D_e^{(j)}, D_u^{(j)}, D_v^{(j)})$$

и $P_i^{(i,j)}$ равно выражению (6.14).

Этап 4. Смесь распределений с k^2 компонентами, полученная на предыдущем этапе, сжимается в приближенно эквивалентное распределение с k компонентами:

$$(a_{1,t}, a_{2,t} | x_t) \sim \sum_{j=1}^k Q_t^{(j)} N(\Phi_t^{(j)}),$$

где $Q_t^{(j)} = \sum_i P_t^{(i,j)}$, а параметры $\Phi_t^{(j)}$ задаются уравнениями (6.16') и (6.16").

Этап 5. Апостериорное распределение теперь имеет тот же вид, что и на этапе 1, и следующая итерация может быть выполнена аналогичным образом.

Этап 6. Для получения прогнозов по модели

$$\hat{x}_\tau(t) = \bar{a}_{1,t} f_t + \bar{a}_{2,t} \tau$$

смесь распределений с k компонентами в свою очередь аппроксимируется двумерным нормальным распределением с параметрами:

$$\bar{a}_{1,t} = \sum_j Q_t^{(j)} \bar{a}_{1,t}^{(j)};$$

$$\bar{a}_{2,t} = \sum_j Q_t^{(j)} \bar{a}_{2,t}^{(j)};$$

$$m_t(a_1, a_1) = \sum_j Q_t^{(j)} [m_t^{(j)}(a_1, a_1) + (\bar{a}_{1,t}^{(j)} - \bar{a}_{1,t})^2];$$

$$m_t(a_1, a_2) = \sum_j Q_t^{(j)} [m_t^{(j)}(a_1, a_2) + (\bar{a}_{1,t}^{(j)} - \bar{a}_{1,t})(\bar{a}_{1,t}^{(j)} - \bar{a}_{1,t})];$$

$$m_t(a_2, a_2) = \sum_j Q_t^{(j)} [m_t^{(j)}(a_2, a_2) + (\bar{a}_{2,t}^{(j)} - \bar{a}_{2,t})^2].$$

Необходимо сделать несколько замечаний относительно сезонных явлений. Модель (6.1) включает сезонные коэффициенты как известные (а не как неизвестные, подобно $a_{1,t}$ и $a_{2,t}$). Разумеется, модель можно расширить таким образом, чтобы сезонные коэффициенты определять на той же байесовской основе. Однако делать это не рекомендуется, так как в связи с резким возрастанием числа неизвестных параметров эффект от полной байесовской обработки сезонного явления будет очевидно мал по сравнению с затратами, особенно принимая во внимание, что используется подход с множеством состояний.

Поскольку значения сезонных коэффициентов, в сущности, связаны с вращением Земли вокруг Солнца, то им

обычно присуща большая стабильность, чем уровню или коэффициенту линейного роста. Однако на практике и сезонные факторы подвержены изменениям, и в случае необходимости это предлагается учитывать вне байесовской системы.

§ 3. РЕАЛИЗАЦИЯ МЕТОДА

Полагаем, что дисперсия основной стохастической составляющей процесса в его «нормальном» состоянии (т. е. при отсутствии изменений) известна:

$$D_0 = D_0(a_{1,t}) = D_e^{(1)}.$$

Это требование не является чрезмерным. Приближенные значения характеристик основной стохастической компоненты необходимы в любой задаче практического прогнозирования, а байесовская система, как утверждают авторы метода, не боится даже больших ошибок в численных оценках этих значений.

Считаем также, что к началу процесса оценки коэффициентов сезонности известны. В общем случае как коэффициенты сезонности, так и характеристики случайных отклонений можно оценивать посредством предварительного анализа, например применением той или иной сглаживающей процедуры к прошлым выборочным данным. При этом, конечно, необходимо особое внимание к искажениям, вносимым самим процессом сглаживания. Иногда у исследователя имеется какая-то априорная информация. Тогда ее можно использовать непосредственно вместо анализа прошлых данных или в дополнение к нему.

Далее должны быть заданы остальные параметры системы:

k — количество состояний процесса;

Pr_j — вероятность наступления состояния j ;

$D_e^{(j)}, D_u^{(j)}, D_v^{(j)}$ — дисперсии стохастических компонент, характеризующие состояние j .

На практике удобно определять эти дисперсии в единицах дисперсии основных флуктуаций D_0 следующим образом:

$$D_e^{(j)} = R_e^{(j)} D_0;$$

$$D_u^{(j)} = R_u^{(j)} D_0;$$

$$D_v^{(j)} = R_v^{(j)} D_0.$$

Выбор значений параметров

Как отмечалось в § 1, будем различать четыре основные ситуации: отсутствие изменений, ступенчатое изменение, изменение в коэффициенте линейного роста, случайное импульсное отклонение, т. е. речь идет о четырех состояниях процесса, но если необходимо различать степени изменений, то их число может быть и большим. Авторы метода считают, однако, что четыре состояния системы дают хорошие результаты и нет необходимости вводить более тонкие различия. Таким образом, они берут $k = 4$.

Что касается остальных параметров, то при испытаниях с большим числом их различных численных значений авторами метода был сделан вывод о том, что работа системы мало чувствительна к значениям параметров, если при их выборе соблюдаются некоторые правила:

1. Состоянию «отсутствия изменений» приписывается высокая вероятность.

2. Состояние «ступенчатое изменение» должно характеризоваться большим значением R_u и очень малой вероятностью.

3. Состояние «изменение в коэффициенте линейного роста» должно иметь или а) $R_v \ll R_u$ и ту же вероятность, что и состояние «ступенчатое изменение», или б) $R_v = R_u$ и намного меньшую вероятность, чем вероятность «ступенчатого изменения».

4. Вероятность «случайного импульса» должна быть по крайней мере в 10 раз больше, чем вероятность «ступенчатого изменения» или «изменения в коэффициенте линейного роста».

5. Сумма относительных коэффициентов $R_e + R_u + R_v$ для состояния «случайного импульса» должны быть не меньше, чем соответствующая сумма для любого другого состояния.

Условие (1) является гипотезой того, что «нормальное» течение процесса лишь изредка прерывается скачкообразными изменениями в уровне или коэффициенте линейного роста.

Эффект условия (2) и (4) состоит в том, что получение первой большой ошибки будет рассматриваться системой как случайный импульс до тех пор, пока последующие данные не подтвердят, что произошло ступенчатое изменение уровня или коэффициента линейного роста.

Следствием условия (3) является то, что система становится менее чувствительной к изменениям в коэффициенте линейного роста, чем в уровне. Во многих практических ситуациях это свойство желательно: оно предотвращает «раскачку» системы.

Условие (5) обеспечивает более широкий диапазон амплитуд состоянию случайного импульсного возмущения.

В табл. 6.2 представлен рекомендуемый авторами метода набор значений параметров, удовлетворяющих перечисленным требованиям; этот набор давал у них хорошие результаты при проверке на искусственных и реальных данных.

Таблица 6.2

Рекомендуемые значения параметров ($k=4$)

Состояние	Вероятность	R_e	R_u	R_v
1. Изменения отсутствуют	0,900	1	0	0
2. Ступенчатое изменение	0,003	1	100	0
3. Изменение коэффициента линейного роста	0,003	1	0	1
4. Случайное импульсное отклонение	0,094	101	0	0

Несмотря на то что до сих пор речь шла о некотором числе различных состояний с известными постоянными вероятностями их наступления, нет оснований считать, что эти вероятности неизменны во времени. Поэтому в общую схему метода иногда целесообразно включать ту или иную процедуру модификации первоначально заданных вероятностей состояний с учетом фактических входных данных.

На рис. 6.2, 6.3, заимствованных нами из работы [66], применение метода демонстрируется на искусственно генерированном ряде. Этот ряд, имеющий нормальное стандартное отклонение 10 (т.е. $D_0 = 100$), содержит два случайных импульса в точках A и B , ступенчатое изменение в точке C величиной 10σ и в точке D изменение коэффициента линейного роста величиной в 1σ .

На рис. 6.2 представлены данные и их прогнозы на один шаг вперед. Видно, что случайным импульсам A и B система дает очень малый вес при формировании текущего прогноза.

Рис. 6.3 показывает, что неопределенность системы относительно значений уровня и коэффициента линейного

роста, выражаемая величинами σ_{a_1} и σ_{a_2} , резко возрастает. В точке C система сначала опять обрабатывает событие как случайный импульс тем же самым образом, что и в точках

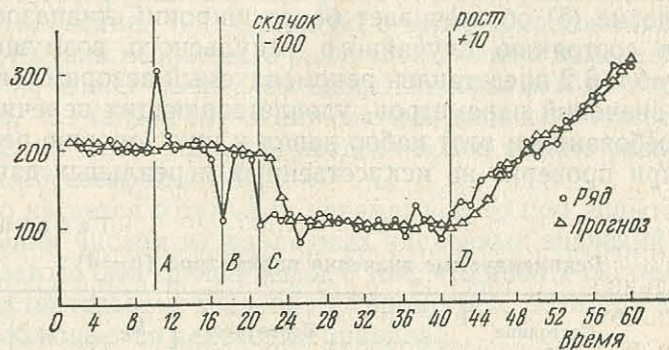


Рис. 6.2. Прогнозирование искусственного ряда по байесовской модели с четырьмя состояниями, $\tau=1$

А и В. После получения следующего значения ряда система «понимает», что эта гипотеза несостоятельна и что, вероятно, имело место большое ступенчатое изменение. Неопреде-

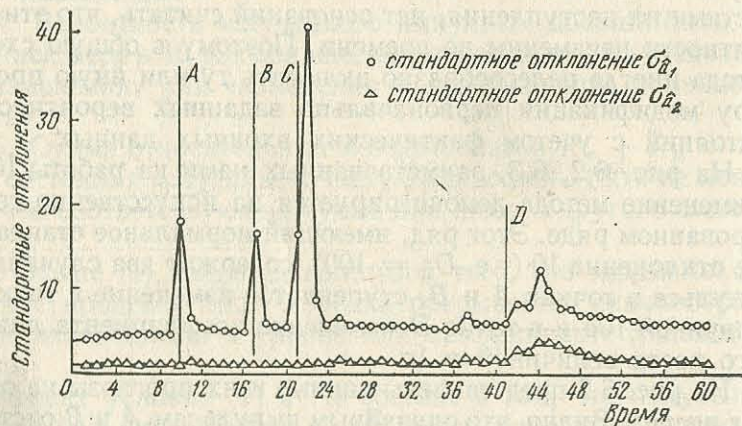


Рис. 6.3. Неопределенность оценок коэффициентов прогнозной модели σ_{a_1} и σ_{a_2}

ленность системы резко возрастает — прошлые данные теперь не дают информации относительно нового уровня, а затем быстро уменьшается с получением новых фактических данных. Поведение в точке D , где имеет место изменение коэффициента линейного роста, отличается своеобразием. Фактически истинное событие в течение некоторого времени скрыто нормальной колеблемостью. Если последовательно открывать на графике данные, то вплоть до $t = 44,45$ не ясно, было ли это изменение в коэффициенте роста, и даже после этого момента представление о его величине далеко от определенности. Байесовская система в целом оказывается весьма неопределенной относительно значений уровня и коэффициента линейного роста, но она становится более чувствительной, постепенно адаптируется и возвращается к нормальному состоянию по мере того, как последующие данные подтверждают новое направление развития ряда.

Исследование чувствительности этой системы показало, что ее эффективность (с точки зрения точности прогнозов) слабо зависит от четырехкратных ошибок в оценке D_0 . Столь большие ошибки в оценке этой величины на практике представляются почти невероятными.

Несмотря на то что описанный метод не получил распространения и требует большого объема вычислений, чем методы Хольта, Брауна или Бокса и Дженкинса (см. гл. 7), он представляет значительный интерес и заслуживает изуче-

Таблица 6.3
Варианты значений относительных дисперсий $R^{(j)} = \frac{D^{(j)}}{D_0}$

в байесовской модели с четырьмя состояниями, использовавшиеся при прогнозировании

Состояние	Вариант № 1			Вариант № 2			Вариант № 3		
	R_E	R_u	R_v	R_E	R_u	R_v	R_E	R_u	R_v
Отсутствие изменений	1	0	0	1	0	0	1	0	0
Ступенчатое изменение	1	10	0	1	50	0	1	100	0
Изменение скорости роста	1	0	1	1	0	1	1	0	1
Случайное импульсное отклонение	11	0	0	51	0	0	101	0	0

Таблица 6.4

Варианты априорных вероятностей четырех состояний
в байесовской модели,
использовавшиеся при прогнозировании

Состояние	Вероятность				
	вариант № 1	вариант № 2	вариант № 3	вариант № 4	вариант № 5
Отсутствие изменений	0,425	0,250	0,600	0,800	0,900
Ступенчатое изменение	0,050	0,050	0,050	0,050	0,003
Изменение скорости роста	0,425	0,600	0,250	0,050	0,003
Случайное импульсное отклонение	0,100	0,100	0,100	0,100	0,094

Таблица 6.5

Варианты значений параметров байесовской модели
с четырьмя состояниями,
из которых отыскивался наилучший вариант

№ варианта значений параметров	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15*
№ варианта значений относительных дисперсий (см. табл. 6.3)	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3
№ варианта априорных вероятностей состояний (см. табл. 6.4)	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5

* Вариант № 15 является рекомендацией авторов метода в [66].

Таблица 6.6

Сравнение результатов прогнозирования
по байесовской модели с четырьмя состояниями
при различных вариантах значений ее параметров
(всего 15 вариантов; см. табл. 6.5)

№ п/п	Выборка из ряда	Число точек	Число точек для определения начальных условий N_1	Основная дисперсия D_0	Вариант параметров, рекомендуемый авторами метода	Наилучший вариант параметров		Вариант параметров, дающий результаты, близкие к наилучшим	
					средний квадрат ошибки	№ варианта	средний квадрат ошибки	№ варианта	средний квадрат ошибки
1	2	3	4	5	6	7	8	9	10
1	Курс акций фирмы ИБМ, дол.	144	30	64,822	76,384	3	52,932	8	53,132
2	Курс акций фирмы ИБМ (часть 1)	48	30	64,822	55,457	4	44,723	9	45,067
3	Курс акций фирмы ИБМ (часть 2)	48	30	76,797	75,059	3	53,549	8	53,707
4	Курс акций фирмы ИБМ (часть 3)	48	30	83,738	67,170	1	52,187	6	52,346
5	Цена на золото, дол. за унцию	360	30	0,10278	0,80053	2	0,57678	1	0,57709
6	Цена на золото (часть 1)	60	30	0,10278	0,16464	3	0,08542	8	0,09595
7	Цена на золото (часть 2)	60	30	0,07807	0,25988	2	0,14159	1	0,14267
8	Цена на золото (часть 3)	60	30	0,01920	0,03223	4	0,02768	14	0,02769
9	Цена на золото (часть 4)	60	30	0,16998	0,20290	4	0,17058	3	0,17174
10	Цена на золото (часть 5)	60	30	0,07612	1,18003	2	0,93812	1	0,93817
11	Цена на золото (часть 6)	60	30	2,66693	2,27642	4	2,20804	3	2,22290
12	Поставка минеральных удобрений сельскому хозяйству, тыс. т усл. ед.	30	15	161559	19721500	2	5108376	1	5356601

№ п/п	Выборка из ряда	Число точек	Число точек для определения начальных условий N_1	Основная дисперсия D_0	Вариант параметров рекомендуемый авторами метода	Наилучший вариант параметров		Вариант параметров, дающий результаты, близкие к наилучшим	
					средний квадрат ошибки	№ варианта	средний квадрат ошибки	№ варианта	средний квадрат ошибки
1	2	3	4	5	6	7	8	9	10
13	Производство электроэнергии, млрд. кВт·ч	31	15	158,92	461,50	12	59,315	7	59,995
14	Производство гидроэлектроэнергии, млрд. кВт·ч	31	15	26,280	48,825	11	30,743	1	30,752
15	Сумма вкладов населения в сберкассы, млн. руб	26	15	186 324	$212,8 \cdot 10^5$	2	$34,307 \times 10^5$	1	$38,32 \times 10^5$
16	Численность городского населения, млн. чел.	27	15	1,0546	1,1694	12	0,4961	7	0,5008
17	Протяженность нефте- и нефтепродуктопроводов, тыс. км	21	15	0,53562	6,09359	12	1,50304	7	1,52253
18	Грузооборот нефте- и нефтепродуктопроводов, ткм	21	15	382,646	2052,27	12	529,64	7	538,76
19	Урожайность зерна, ц с га	30	15	0,63794	2,54825	5	2,49374	10	2,52021
20	Урожайность картофеля, ц с га	30	15	138,720	170,871	5	170,635	10	170,825
21	Урожайность овощей, ц с га	30	15	56,142	81,834	5	81,231	10	81,699
22	Производство яиц, млрд. шт.	30	15	0,51425	4,11174	12	2,26763	7	2,28091

ния. Прежде всего желательно сопоставить его эффективность с точностью прогнозов, получаемых другими методами. Между тем практическому использованию описанного метода посвящена, по-видимому, лишь одна работа [63]. Поэтому нами были проведены эмпирические сопоставления, результаты которых приведены в следующем параграфе. Первые значения параметров распределения коэффициентов (a_1, a_2), необходимые для начала рассмотренной в этой главе итеративной процедуры, и основная дисперсия D_0 определялись нами путем оценивания линейной регрессии на первых N_1 точках выборки. Считалось, что $Q_0^{(j)} = Pr_j$. Рассмотрим соответствующий пример.

Байесовская модель с 4 состояниями была испытана нами на 22 выборках различного объема. Было опробовано 15 вариантов задаваемых значений параметров модели (см. табл. 6.3, 6.4, 6.5), из которых вариант № 15 представлял значения, рекомендуемые авторами метода.

Результаты, приведенные в табл. 6.6, свидетельствуют о том, что рекомендация авторов относительно значений параметров не получила экспериментального подтверждения. Анализ результатов приводит к выводу, что средний квадрат ошибки особенно чувствителен к значениям априорных вероятностей состояния процесса.

§ 4. СРАВНЕНИЕ МЕТОДОВ

Теперь, познакомившись с разнообразными моделями адаптивного типа, целесообразно провести эмпирическое сравнение точности получаемых по ним прогнозов.

Результаты испытаний адаптивных моделей представлены в сводной табл. 6.7. В ней прямоугольником выделен результат, наилучший в строке, а подчеркнут — результат, лучший из тех, которые дают модели метода эволюции единственного параметра адаптации. Номера моделей в графах 8, 11, 12, 13 означают: № 1 — полиномиальная модель нулевого порядка, № 2 — полиномиальная модель первого порядка, № 3 — полиномиальная модель второго порядка.

Как видно из таблицы, превосходство осталось за адаптивной моделью авторегрессии, рассмотренной в гл. 3. Совершенно не выдерживает конкуренции намного более сложная байесовская модель с 4 состояниями (см. выборки № 11, 12, 13, 15, 18). Остальные модели в общем незначительно отстают от модели адаптивной авторегрессии.

Сравнение адаптивных методов
ретроспективных прогнозов

Выборка			Метод Байеса		Адаптивная авторегрессия	
№ п/п	из ряда	число точек	N ₁	средний квадрат ошибок	макс.-малый лаг	средний квадрат ошибок
1	2	3	4	5	6	7
1	Курс акций фирмы ИБМ, дол.	144	30	52,932	1	40,937
2	Курс акций фирмы ИБМ (часть 1)	48	30	44,723	1	33,969
3	Курс акций фирмы ИБМ (часть 2)	48	30	53,549	3	42,854
4	Курс акций фирмы ИБМ (часть 3)	48	30	52,187	1	41,898
5	Цена на золото, дол. за унцию	360	30	0,57678	1	0,50989
6	Цена на золото (часть 1)	60	30	0,09542	1	0,08964
7	Цена на золото (часть 2)	60	30	0,14159	1	0,10582
8	Цена на золото (часть 3)	60	30	0,02768	2	0,02898
9	Цена на золото (часть 4)	60	30	0,17058	1	0,14467
10	Цена на золото (часть 5)	60	30	0,93812	1	0,77920
11	Цена на золото (часть 6)	60	30	2,20804	3	0,85576
12	Поставка минеральных удобрений в сельское хозяйство, тыс. т усл. ед.	30	15	51,083 · 10 ⁵	1	21,459 · 10 ⁵

Таблица 6.7

по средним квадратам ошибок
сделанных на один шаг вперед ($\tau=1$)

Метод Брауна			Метод эволюции			
№ модели	параметр $\alpha_{\text{опт}}$	средний квадрат ошибок	средний квадрат ошибок			
			модель 1	модель 2	модель 3	комбинированная
8	9	10	11	12	13	14
1	0,9600	41,427	43,419	64,315	80,240	44,728
1	0,9450	32,956	37,565	47,913	73,710	34,863
1	0,9992	46,294	46,445	72,234	79,970	49,308
1	0,9120	43,419	53,252	58,787	68,305	49,516
1	0,8410	0,50123	0,51263	0,64924	0,76728	0,65877
1	0,7510	0,08265	0,09869	0,09877	0,14866	0,09719
1	0,9991	0,10480	0,10958	0,14941	0,27336	0,13601
2	0,2711	0,02884	0,02913	0,02902	0,03089	0,03224
1	0,0709	0,14521	0,16726	0,18879	0,30078	0,14726
1	0,9999	0,84637	0,96741	0,93089	0,98194	0,94999
1	0,5709	1,6928	1,4022	1,9341	2,6717	1,5401
2	0,8010	28,148 × 10 ⁵	201,98 × 10 ⁵	42,00 · 10 ⁵	30,90 · 10 ⁵	32,58 · 10 ⁵

Выборка			Метод Байеса		Адаптивная авторегрессия	
№ п/п	из ряда	число точек	N ₁	средний квадрат ошибок	максимальный лаг	средний квадрат ошибок
1	2	3	4	5	6	7
13	Производство электроэнергии, млрд. кВт·ч	31	15	59,315	1	19,115
14	Производство гидроэлектроэнергии, млрд. кВт·ч	31	15	30,743	1	34,933
15	Сумма вкладов населения в сберкассы, млн. руб	26	15	$34,307 \times 10^5$	1	$8,45810 \cdot 10^5$
16	Численность городского населения, млн. чел.	27	15	0,4961	1	0,37938
17	Протяженность нефте- и нефтепродуктопроводов, тыс. км	21	15	1,50304	1	1,1703
18	Грузооборот нефте- и нефтепродуктопроводов, ткм	21	15	529,64	1	99,479
19	Урожайность зерна, ц с га	30	15	2,4937	2	3,1536
20	Урожайность картофеля, ц с га	30	15	170,63	3	165,24
21	Урожайность овощей, ц с га	30	15	81,231	3	71,758
22	Производство яиц, млрд. шт.	30	15	2,2676	1	2,0802

Метод Брауна			Метод эволюции			
№ модели	параметр $\alpha_{\text{опт}}$	средний квадрат ошибок	средний квадрат ошибок			
			модель 1	модель 2	модель 3	комбинированная
8	9	10	11	12	13	14
2	0,8990	26,593	1888,7	22,829	31,063	21,722
2	0,5121	27,456	54,920	37,274	38,893	32,424
2	0,9909	$10,921 \times 10^5$	$307,6 \cdot 10^5$	$19,04 \cdot 10^5$	$9,872 \cdot 10^5$	$11,86 \cdot 10^5$
2	0,9191	0,32741	28,067	0,56278	0,81668	0,43933
2	0,6029	1,4529	13,872	1,6304	1,4723	1,5126
3	0,7298	168,73	2757,0	322,11	296,64	298,66
2	0,0191	2,2653	3,2946	3,1126	3,0164	3,2191
1	0,3060	170,72	162,68	186,43	202,33	198,12
2	0,0668	78,732	121,31	90,85	100,58	117,68
2	0,7291	1,9412	9,2810	2,2714	2,2063	1,7782

Глава 7

МОДЕЛИ АВТОРЕГРЕССИИ — СКОЛЬЗЯЩЕГО СРЕДНЕГО (МЕТОД БОКСА — ДЖЕНКИНСА)

В главе излагается метод построения смешанных моделей авторегрессии — скользящего среднего, исследуются свойства моделей, проводится их эмпирическое сравнение с адаптивной моделью авторегрессии, рассмотренной в гл. 3. Здесь рассматриваются модели для временных рядов без эффекта сезонности.

§ 1. ОБЩЕЕ ОПИСАНИЕ МОДЕЛЕЙ И ИХ СВОЙСТВ

Еще одно направление развития и обоснования адаптивных моделей дано в работе Дж. Бокса и Г. Дженкинса [5]. В их книге особое внимание уделено проблеме выбора модели и ее оцениванию, что и вызвало интерес специалистов. Однако недостаточное количество практических примеров прогнозирования с использованием этого метода вызывает определенные сомнения. Возникает необходимость проверить метод на деле, сравнить его с другими методами.

Прежде всего кратко опишем модели, лежащие в основе процедуры прогнозирования, и методы построения этих моделей, а затем проанализируем особенности их применения и конкретные результаты, полученные при проведении экспериментов.

Для описания моделей потребуются следующие операторы¹ и обозначения:

x_t — значение ряда в момент t ;

¹ О правомерности и особенностях алгебраических действий над рассматриваемыми операторами см. в [1, приложение А].

B — оператор сдвига назад, определенный следующим образом: $Bx_t = x_{t-1}$, $B^m x_t = x_{t-m}$;

∇ — оператор разности назад: $\nabla x_t = x_t - x_{t-1} = (1 - B)x_t$;

$\nabla^{-1} = S$ — оператор суммирования:

$$\nabla^{-1} x_t = Sx_t = \sum_{j=0}^{\infty} x_{t-j} = (1 + B + B^2 + \dots) x_t = (1 - B)^{-1} x_t;$$

ε_t — белый шум с дисперсией σ_ε^2 .

Остальные обозначения будут введены по ходу изложения.

Модель основывается на гипотезе, что изучаемый процесс является выходом линейного фильтра, на вход которого подан процесс белого шума, т. е. что член ряда x_t является взвешенной суммой текущего и предыдущих значений входного потока:

$$x_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots = \mu + \psi(B) \varepsilon_t,$$

где $\mu = \text{const}$ в общем случае является параметром, характеризующим процесс, и оператор $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$

Если последовательность $\psi_1, \psi_2, \dots$ конечна или бесконечна, но сходится, то фильтр называется устойчивым, а процесс x_t будет стационарным. Тогда μ — среднее значение, вокруг которого процесс варьирует. В противном случае x_t — нестационарен и μ не имеет особого смысла, кроме как некой точки отсчета уровня процесса.

Рассмотрим некоторые специфические модели, получаемые линейной фильтрацией белого шума.

Типы моделей

Авторегрессионная модель. В этой модели текущее значение процесса выражается через конечную линейную совокупность предыдущих значений процесса и возмущения ε_t .

Если обозначить через $\tilde{x}_t$ отклонение от среднего $\tilde{x}_t = x_t - \mu$, то получаем процесс

$$\tilde{x}_t = \varphi_1 \tilde{x}_{t-1} + \varphi_2 \tilde{x}_{t-2} + \dots + \varphi_p \tilde{x}_{t-p} + \varepsilon_t, \quad (7.1)$$

где $\varphi_i = \text{const}$, $i = 1, \dots, p$, называемый авторегрессионным процессом порядка p , который будем обозначать $AP(p)$. Введем авторегрессионный оператор порядка p :

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p,$$

тогда модель может быть записана экономно:

$$\varphi(B) \tilde{x}_t = \varepsilon_t. \quad (7.2)$$

В этой модели $p + 2$ неизвестных параметра $\mu, \varphi_1, \varphi_2, \dots, \dots, \varphi_p, \sigma_\varepsilon^2$, которые должны быть оценены по имеющимся данным об изучаемом процессе.

Если последовательно выразить $\tilde{x}_{t-1}, \tilde{x}_{t-2}$ и т. д. через $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots$, то получим эквивалентную запись через бесконечную взвешенную сумму реализаций белого шума:

$$\tilde{x}_t = \psi(B) \varepsilon_t. \quad (7.3)$$

Однако здесь количество неизвестных параметров модели оказывается бесконечным и форма (7.2) явно предпочтительней. Из (7.2) и (7.3) следует, что $\psi(B) = \varphi^{-1}(B)$. (7.4)

Процессы могут быть стационарными и нестационарными. Для того чтобы процесс был стационарным, коэффициенты φ должны быть такими, чтобы веса $\psi_1, \psi_2, \dots$ в (7.3) образовывали сходящийся ряд. Для решения практических задач, как правило, достаточно $p \leq 2$.

Модель скользящего среднего. Другим типом модели, имеющим практическую ценность, является модель конечного скользящего среднего, в которой $\tilde{x}_t$ линейно зависит от конечного числа предыдущих значений ε , т.е.

$$\tilde{x}_t = \varepsilon_t - Q_1 \varepsilon_{t-1} - Q_2 \varepsilon_{t-2} - \dots - Q_q \varepsilon_{t-q}. \quad (7.5)$$

Это процесс скользящего среднего порядка q или кратко СС (q). Следует отметить, что в данном случае название «скользящее среднее» вводит в заблуждение, так как веса $1, -Q_1, -Q_2, \dots, -Q_q$ не обязательно должны в сумме давать единицу и не обязательно должны быть положительными.

Если введем оператор процесса скользящего среднего порядка q

$$Q(B) = 1 - Q_1 B - Q_2 B^2 - \dots - Q_q B^q,$$

то модель СС (q) может быть записана кратко:

$$\tilde{x}_t = Q(B) \varepsilon_t. \quad (7.6)$$

Она содержит $q + 2$ неизвестных параметра $\mu, Q_1, Q_2, \dots, Q_q, \sigma_\varepsilon^2$. Обычно $q = 0, 1, 2$.

Смешанная модель АРСС. Для достижения большей гибкости при построении модели исследуемых процессов полезно включать в нее и члены скользящего среднего, и авторегрессионные члены. Это приводит к смешанной модели АРСС (p, q):

$$\varphi(B) \tilde{x}_t = Q(B) \varepsilon_t$$

с $p + q + 2$ неизвестными параметрами.

Модель может быть записана и так:

$$\begin{aligned} \tilde{x}_t &= \varphi^{-1}(B) Q(B) \varepsilon_t = \frac{Q(B)}{\varphi(B)} \varepsilon_t = \\ &= \frac{1 - Q_1 B - \dots - Q_q B^q}{1 - \varphi_1 B - \dots - \varphi_p B^p} \varepsilon_t. \end{aligned}$$

Особый практический интерес представляет случай $p = 1, q = 1$.

Нестационарные модели. В этих моделях используются идеи, разработанные в [36], о возможности трансформировать нестационарные ряды в стационарные путем перехода от исходного ряда к его разностям соответствующего порядка d . Тогда преобразованный, стационарный ряд можно описать одной из рассмотренных выше моделей.

Введем обобщенный авторегрессионный оператор $\Psi(B) = \varphi(B)(1 - B)^d$, где $\varphi(B)$ — стационарный оператор, рассмотренный выше. Тогда модель можно записать таким образом:

$$\Psi(B) x_t = \varphi(B)(1 - B)^d x_t = Q(B) \varepsilon_t,$$

или

$$\varphi(B) w_t = Q(B) \varepsilon_t; \quad (7.7)$$

$$w_t = \nabla^d x_t. \quad (7.8)$$

На практике обычно $d = 0, 1$ или, наконец, 2 . Модель (7.7, 7.8) является более общей и называется интегрированной моделью авторегрессии — скользящего среднего (АРИСС) порядка (p, d, q) . Она называется интегрированной потому, что соотношение обратное (7.8) есть $x_t = S^d w_t$, где S — оператор суммирования.

Для преобразования исходного ряда в стационарный могут использоваться и различные другие преобразования. Например, для экономических и других рядов, характери-

зующихся приблизительно экспоненциальным ростом, полезно предварительно применять операцию логарифмирования (см. § 4 этой главы).

Включение детерминированной полиномиальной тенденции роста в модель АРИСС. Иногда бывает полезна небольшая модификация модели АРИСС путем добавления постоянного члена в правую часть. Это новое слагаемое придает модели наиболее общий вид:

$$\Psi(B)x_t = \varphi(B)\nabla^d x_t = Q_0 + Q(B)\varepsilon_t. \quad (7.9)$$

Если постоянный член Q_0 опущен, то модель может отображать ряды, имеющие стохастические тренды (случайные изменения уровня, темпа роста и т.п.). В общем случае, однако, может быть желательным включение в модель детерминированной функции времени $f(t)$. В частности, можно включить полином порядка d , если положить $Q_0 \neq 0$.

Взяв математическое ожидание от обеих частей равенства (7.9), получаем

$$M(w_t) = M(\nabla^d x_t) = \mu_w = Q_0/(1 - \varphi_1 - \varphi_2 - \dots - \varphi_p).$$

Отсюда следует, что модель (7.9) эквивалентна модели

$$\varphi(B)\tilde{w}_t = Q(B)\varepsilon_t, \quad (7.10)$$

где $\tilde{w}_t = w_t - \mu_w$.

Поскольку предполагается, что временной ряд генерируется моделью из класса стационарных и обратимых, то необходимо уяснить требования, накладываемые при этом на ее параметры.

Стационарность линейного процесса

Единственное условие стационарности состоит в том, чтобы $\varphi(B)$ сходилась при $|B| \leq 1$, где B — комплексное число*.

Для процесса $AR(p)$ в общем случае $\varphi(B)$ можно рассматривать как полином порядка p относительно B , поэтому $\varphi(B)$ можно представить в виде $\varphi(B) = (1 - G_1 B) \times$

$\times (1 - G_2 B) \dots (1 - G_p B)$. Тогда, если все корни полинома различны,

$$\tilde{x}_t = \varphi^{-1}(B) \varepsilon_t = \sum_{i=1}^p \frac{K_i}{(1 - G_i B)} \varepsilon_t.$$

Так как для стационарности процесса последовательность $\varphi(B) = \varphi^{-1}(B)$ должна быть сходящейся при $|B| \leq 1$, то необходимо иметь $|G_i| < 1$, $i = 1, 2, \dots, p$. Это требование эквивалентно тому, что корни выражения $\varphi(B) = 0$ должны лежать за пределами единичного круга. Уравнение $\varphi(B) = 0$ называется характеристическим. Ясно, что для процесса АРСС условия стационарности те же.

На параметры процесса скользящего среднего из-за конечного числа членов в $\psi(B) = Q(B) = 1 - Q_1 B - Q_2 B^2 - \dots - Q_q B^q$ никакие ограничения для того, чтобы гарантировать стационарность, не накладываются.

Мы видели, что процесс АРСС является стационарным, если корни $\varphi(B) = 0$ лежат вне единичного круга. На простом примере легко убедиться, что если корни лежат внутри единичного круга, то модель проявляет нестационарность. Особое внимание авторов модели привлеч случай, когда корни лежат на единичной окружности. Так появилась нестационарная модель АРИСС.

Обратимость

Линейный процесс

$$\tilde{x}_t = \varepsilon_t + \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j}$$

может быть записан через взвешенную сумму прошлых значений $\tilde{x}$ и аддитивный скачок ε_t

$$\tilde{x}_t = \pi_1 \tilde{x}_{t-1} + \pi_2 \tilde{x}_{t-2} + \dots + \varepsilon_t = \sum_{j=1}^{\infty} \pi_j \tilde{x}_{t-j} + \varepsilon_t,$$

или

$$\pi(B) \tilde{x}_t = \varepsilon_t, \quad (7.11)$$

где

$$\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j.$$

* Необходимо отличать B — комплексную переменную от B — оператора сдвига назад. Одинаковое обозначение введено для большей формализации анализа, ибо всегда для получения характеристических уравнений (см. дальше) требуется замена оператора сдвига назад на комплексную переменную.

В общем случае линейный процесс (7.11) обратим, если веса π_j таковы, что $\pi(B)$ сходится при $|B| \leq 1$.

Получим одно полезное соотношение, применив оператор $\psi(B)$ к обеим частям (7.11):

$$\psi(B)\pi(B)\tilde{x}_t = \psi(B)\varepsilon_t = \tilde{x}_t.$$

Отсюда $\psi(B)\pi(B) = 1$, т.е. $\pi(B) = \psi^{-1}(B)$. (7.12)
Из (7.12) можно получать веса π , зная ψ , и наоборот.

Рассмотрим модель $\tilde{x}_t = (1 - Q_1 B)\varepsilon_t$, в которой

$$\psi_1 = -Q_1, \psi_j = 0, j > 1.$$

Выражая ε_t через $\tilde{x}_t$, получим $\varepsilon_t = (1 - Q_1 B)^{-1}\tilde{x}_t$.
Отсюда

$$(1 + Q_1 B + Q_1^2 B^2 + Q_1^3 B^3 + \dots)\tilde{x}_t = \varepsilon_t, \quad (7.13)$$

так что для этой модели $\pi_j = -Q_1^j$.

Из выражения (7.13) видно, что модель СС(1) выражается моделью авторегрессии с бесконечным числом членов. Справедливо и обратное: модель АР (1) эквивалентна модели СС с бесконечным числом членов. Это говорит о большой гибкости модели АРСС (1, 1).

Теперь выведем условия, которым должны удовлетворять параметры $Q_1, Q_2, \dots, Q_q$, для того чтобы гарантировать обратимость процесса СС(q). Запишем модель следующим образом:

$$\varepsilon_t = Q^{-1}(B)\tilde{x}_t.$$

Далее используем уже известный прием (считаем, что корни полинома $Q(B)$ различны):

$$Q(B) = \prod_{i=1}^q (1 - H_i B);$$

$$\pi(B) = Q^{-1}(B) = \sum_{j=1}^q \frac{M_j}{(1 - H_j B)}.$$

Таким образом, $\pi(B)$ сходится при $|B| < 1$, если $|H_j| < 1$ для $j = 1, 2, \dots, q$. Так как корни $Q(B) = 0$ есть H_j^{-1} , то, следовательно, условие обратимости для процесса СС(q) заключается в том, чтобы корни характеристического уравнения $Q(B) = 0$ лежали вне единичного круга. Для процесса АРСС условия обратимости аналогичны.

На параметры АР-процесса никакие ограничения для обеспечения обратимости не накладываются, так как $\pi(B) = \varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$ содержит конечное число членов.

Автокорреляционная функция

Автокорреляционная функция в этом методе является основным инструментом при построении модели. Изучим ее свойства сначала для АР-процессов:

$$\tilde{x}_t = \varphi_1 \tilde{x}_{t-1} + \varphi_2 \tilde{x}_{t-2} + \dots + \varphi_p \tilde{x}_{t-p} + \varepsilon_t. \quad (7.14)$$

Умножим (7.14) на $\tilde{x}_{t-k}$.

Берем математическое ожидание и получаем разностное уравнение для автоковариаций R :

$$R_k = \varphi_1 R_{k-1} + \varphi_2 R_{k-2} + \dots + \varphi_p R_{k-p}, \quad k > 0.$$

Отметим, что $M(\tilde{x}_{t-k}\varepsilon_t) = 0$, когда $k > 0$, так как $\tilde{x}_{t-k}$ может включать реализации ε , имевшие место до момента $t - k$, а они некоррелированы с ε_t . Разделив все члены на R_0 , убедимся, что автокорреляционная функция удовлетворяет разностному уравнению того же вида:

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \dots + \varphi_p \rho_{k-p}, \quad k > 0. \quad (7.15)$$

Оно аналогично разностному уравнению, которому удовлетворяет сам процесс $\tilde{x}_t$. Таким образом, (7.15) можно записать как

$$\varphi(B)\rho_k = 0. \quad (7.16)$$

Учитывая, что $\varphi(B) = \prod_{i=1}^p (1 - G_i B)$, получаем общее решение уравнения (7.16):

$$\rho_k = A_1 G_1^k + A_2 G_2^k + \dots + A_p G_p^k, \quad (7.17)$$

где $G_1^{-1}, G_2^{-1}, \dots, G_p^{-1}$ — корни характеристического уравнения $\varphi(B) = 0$ и благодаря условию стационарности $|G_i| < 1$. Если корни G_i различны, то возможны два случая.

1. Корень G_i вещественный. Тогда член $A_i G_i^k$ в (7.17) геометрически убывает к нулю с ростом k (затухающая экспонента).

Имея коэффициенты автокорреляции, можно с их помощью оценить параметры АР процесса. Для этого подставим в (7.15) $k = 1, 2, \dots, p$ и получим систему линейных уравнений для $\varphi_1, \varphi_2, \dots, \varphi_p$:

$$\left. \begin{aligned} \rho_1 &= \varphi_1 + \varphi_2 \rho_1 + \dots + \varphi_p \rho_{p-1}; \\ \rho_2 &= \varphi_1 \rho_1 + \varphi_2 + \dots + \varphi_p \rho_{p-2}; \\ &\vdots \\ \rho_p &= \varphi_1 \rho_{p-1} + \varphi_2 \rho_{p-2} + \dots + \varphi_p. \end{aligned} \right\} \quad (7.18)$$

Если перейти к матричным обозначениям:

$$\varphi = \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_p \end{bmatrix}, \quad \rho_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}, \quad P_p = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & 1 \end{bmatrix},$$

то решение системы (7.18) относительно параметров Φ можно записать в виде

$$\varphi = P_p^{-1} \rho_p. \quad (7.19)$$

Заменяя теоретические автокорреляции ρ_k на их оценки r_k , можно получить оценки Юла-Уокера для параметров модели. Отметим сразу же, что в статистике имеется несколько оценок автокорреляций, свойства которых подробно рассмотрены в [13]. Там сделан вывод о предпочтительности оценки автокорреляции ρ_k по формуле

$$r_k = \hat{R}_k / \hat{R}_0,$$

где $\hat{R}_k = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t-k} - \bar{x})$ выборочная оценка автоковариации;

$\bar{x}$ — среднее значение временного ряда по N наблюдениям.

$$R_k = M[(\varepsilon_t - Q_1 \varepsilon_{t-1} - \dots - Q_q \varepsilon_{t-q}) \times \\ \times (\varepsilon_{t-k} - Q_1 \varepsilon_{t-k-1} - \dots - Q_q \varepsilon_{t-k-q})].$$
$$R_0 = \sigma_x^2 = (1 + Q_1^2 + Q_2^2 + \dots + Q_q^2) \sigma_\varepsilon^2$$

$$R_k = \begin{cases} (-Q_k + Q_1 Q_{k+1} + Q_2 Q_{k+2} + \dots + \\ + Q_{g-k} Q_g) \sigma_\varepsilon^2 & k=1, \dots, q; \\ 0 & k > q. \end{cases} \quad (7.20)$$

Следовательно, автокорреляционная функция будет:

$$\rho_k = \begin{cases} \frac{-Q_k + Q_1 Q_{k+1} + \dots + Q_{q-k} Q_q}{1 + Q_1^2 + \dots + Q_q^2} & k = 1, 2, \dots, q; \\ 0 & k > q. \end{cases} \quad (7.21)$$

Итак, для $k > q$ автокорреляционная функция СС-процесса равна нулю. Другими словами, автокорреляционная функция процесса скользящей средней обрывается при лаге q . Это значит, что по автокорреляционной функции можно для процессов СС определить численное значение q .

Учитывая, однако, что имеются лишь грубые оценки $r_1, \dots, r_q$, необходим некий показатель того, насколько оцененное значение r_i может отличаться от соответствующего теоретического значения, в частности, для суждения, являются ли автокорреляции нулями после некоторого лага q . Для этих лагов можно подсчитать стандартные ошибки оценок r_i по упрощенной формуле Бартлетта:

$$\hat{\sigma}(r_k) \simeq \frac{1}{\sqrt{N}} \{1 + 2(r_1^2 + r_2^2 + \dots + r_q^2)\}^{1/2} \quad k > q.$$

В [22, вып. 2, с.46] показано, как определить стандартные ошибки оценок r_i в общем случае.

По оценкам $r_1, \dots, r_q$, используя q уравнений (7.21), можно получить оценки $Q_1, Q_2, \dots, Q_q$. Однако в отличие от уравнений Юла-Уокера для АР-процессов, которые являются линейными, уравнения (7.21) являются нелинейными. Поэтому, за исключением простого случая, когда $q = 1$, эти уравнения должны решаться итеративно.

Наконец, для смешанных процессов автокорреляционная функция может быть выведена аналогично тому, как она

выводилась для АР-процесса. Автоковариационная функция здесь удовлетворяет разностному уравнению:

$$R_k = \varphi_1 R_{k-1} + \dots + \varphi_p R_{k-p} + R_{xe}(k) - Q_1 R_{xe}(k-1) - \dots - Q_q R_{xe}(k-q), \quad (7.22)$$

где $R_{xe}(k) = M(\tilde{x}_t - k \varepsilon_t)$ — взаимная ковариационная функция между $\tilde{x}$ и ε

$R_{xe}(k) = 0$ для $k > 0$;

$R_{xe}(k) \neq 0$ для $k \leq 0$.

Выражение (7.22) означает, что

$$R_k = \varphi_1 R_{k-1} + \varphi_2 R_{k-2} + \dots + \varphi_p R_{k-p} \quad k \geq q+1.$$

отсюда

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \dots + \varphi_p \rho_{k-p} \quad k \geq q+1,$$

$$\text{или } \varphi(B) \rho_k = 0 \quad k \geq q+1. \quad (7.23)$$

Таким образом, для процесса АРСС (p, q) будет q автокорреляций $\rho_q, \rho_{q-1}, \dots, \rho_1$, значения которых зависят и от параметров скользящего среднего и от автокорреляционных параметров.

p величин $\rho_q, \rho_{q-1}, \dots, \rho_{q-p+1}$ составляют необходимые начальные значения для разностного уравнения (7.23), которое затем полностью определяет автокорреляции с более высокими лагами. Если $q - p < 0$, вся автокорреляционная функция $\rho_j, j = 0, 1, 2, \dots$ будет выражаться совокупностью затухающих экспонент и (или) затухающих синусоидальных волн, а ее свойства будут определяться полиномом $\varphi(B)$ и p начальными значениями автокорреляционной функции. Если, однако, $q - p \geq 0$, то будет $q - p + 1$ значений $\rho_0, \rho_1, \dots, \rho_{q-p}$, выпадающих из данной схемы. Эти моменты полезны при идентификации смешанных процессов.

Частная автокорреляционная функция

При построении модели вначале неизвестно, какого порядка авторегрессионный процесс надо ставить в соответствие фактическому ряду. Инструментом решения этого вопроса служит частная автокорреляционная функция. Здесь используется тот факт, что процесс АР (p) , имеющий автокорреляционную функцию с бесконечным числом членов, по своей природе может быть описан с помощью p ненулевых функций от автокорреляций.

Обозначим через φ_{kj} j -й коэффициент в авторегрессионном процессе порядка k , так что φ_{kk} — последний коэффициент. По (7.18) φ_{kj} удовлетворяют системе уравнений Юла-Уокера:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \varphi_{k1} \\ \varphi_{k2} \\ \vdots \\ \varphi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad (7.24)$$

или

$$\mathbf{P}_k \boldsymbol{\varphi}_k = \boldsymbol{\rho}_k.$$

Решая эту систему уравнений последовательно для $k = 1, 2, 3, \dots$, получим $\varphi_{11} = \rho_1$,

$$\varphi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2};$$

$$\varphi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}. \quad (7.25)$$

В общем случае в выражении для φ_{kk} определитель числителя имеет те же самые элементы, что и определитель знаменателя, но последний столбец заменен на $\boldsymbol{\rho}_k$. Величина φ_{kk} рассматривается как функция лага k и называется частной автокорреляционной функцией.

Для авторегрессионного процесса порядка p частная автокорреляционная функция φ_{kk} будет отлична от нуля при $k \leq p$ и равна нулю при $k > p$. Другими словами, частная автокорреляционная функция АР (p) -процесса обрывается после лага p . Если значения параметров не слишком близки к границам нестационарности, то успешно могут быть использованы оценки Юла-Уокера; при невыполнении этого условия оценки оказываются грубыми.

При гипотезе, что процесс является авторегрессионным порядка p , оценки частных автокорреляций для лагов $p+1$ и выше, как показал М. Кенуй [94], распределены приблизительно независимо с дисперсией $D(\hat{\varphi}_{kk}) \simeq \frac{1}{N} \quad k \geq p+1$.

Таким образом, стандартная ошибка оцененной частной автокорреляции $\hat{\varphi}_{kk}$ будет:

$$\sigma(\hat{\varphi}_{kk}) \simeq \frac{1}{\sqrt{N}} \quad k \geq p+1 \quad (7.26)$$

Это и служит критерием определения момента обрыва функции.

Частная автокорреляционная функция существует также и для процессов скользящего среднего. Как мы уже видели (7.13), модель СС(1) эквивалентна модели АР с бесконечным числом членов. Это означает, что частная автокорреляционная функция в данном случае будет убывать бесконечно, но обрыва иметь не будет. Это свойство имеет более общий характер и присуще моделям СС высоких порядков, а также процессам АРСС.

Рассмотренные свойства моделей сведены в табл. 7.1. Здесь наглядно видны взаимно-обратные свойства СС и АР-процессов.

Таблица 7.1

Свойства автокорреляционных функций

Процесс Функция	АР (p)	СС (q)	АРСС (p, q)
ρ_k	бесконечная убывающая	обрыв после лага q	бесконечная, после первых (q-p) лагов доминирует наложение затухающих экспонент и синусоид
$\hat{\varphi}_{kk}$	обрыв после лага p	бесконечная убывающая	бесконечная, после первых (p-q) лагов доминирует наложение затухающих экспонент и синусоид

§ 2. ИДЕНТИФИКАЦИЯ МОДЕЛЕЙ. ИНТЕРПРЕТАЦИЯ R^2 В МОДЕЛЯХ АРСС

Теперь у нас достаточно информации для того, чтобы перейти к проблемам построения, оценки и использования моделей прогнозирования.

Если имеется достаточно длинный отрезок ряда (желательно 50 или более точек) и принято решение искать адекватную модель в классе процессов АРИСС, то выбор и привязка модели этого типа к имеющимся данным лучше всего достигаются с помощью трехстадийной итеративной процедуры, включающей идентификацию, оценку и диагностическую проверку модели. После этого модель можно использовать для прогнозирования (см. рис. 7.1).

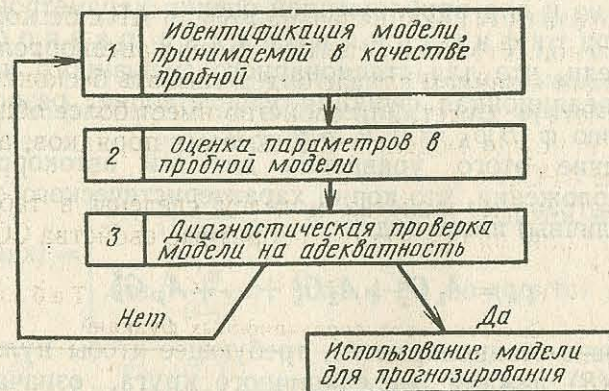


Рис. 7.1. Блок-схема построения модели

Под идентификацией имеется в виду использование наблюдаемых данных и любой другой информации для определения подкласса экономных (с точки зрения количества параметров) моделей, среди которых следует искать адекватную.

Под оценкой понимается эффективное использование данных для получения численных значений параметров модели при предположении ее адекватности процессу.

Диагностическая проверка имеет целью проверку адекватности подобранной модели и ее улучшение.

Идентификация моделей

Методы идентификации являются грубыми процедурами, применяемыми к исходным данным для определения типа модели, который стоит исследовать дальше. Специфической целью здесь является получение некоторого представления о величинах p , d , q , необходимых для конструирования ли-

нейной модели АРИСС (p, d, q) и получение предварительных оценок ее параметров. Идентификация распадается на две стадии.

1. Определение порядка разности d исходного ряда x , который обеспечивает стационарность.

2. Идентификация модели АРСС для ряда $\nabla^d x_t$.

Главными инструментами анализа на обеих стадиях являются автокорреляционная и частная автокорреляционная функции. Они используются не только для определения вида модели, но и для приближенной оценки параметров.

Идентификация порядка разности d . Мы видели, что для стационарного смешанного процесса автокорреляционная функция удовлетворяет разностному уравнению $\varphi(B)\rho_k = 0, k > q - p$.

Решение этого уравнения для k -й автокорреляции (в предположении, что корни характеристического уравнения различны) имеет вид:

$$\rho_k = A_1 G_1^k + A_2 G_2^k + \dots + A_p G_p^k. \quad (7.27)$$

Условие стационарности, требующее чтобы нули полинома $\varphi(B)$ лежали вне единичного круга, означает, что $G_1, G_2, \dots, G_p$ лежат внутри единичного круга. Из (7.27) ясно, что в случае стационарного процесса, для которого ни один из корней не лежит близко к границе единичного круга, автокорреляционная функция будет сравнительно быстро затухать с ростом k . Если же один вещественный корень, скажем G_1 , близок к 1, так что $G_1 = 1 - \delta$, где δ является некоторой малой положительной величиной, то для большого k

$$\rho_k \simeq A_1 (1 - k\delta)$$

автокорреляционная функция будет затухать медленно и почти линейно. Это свойство используется в качестве индикатора корня, близкого к 1.

Способность автокорреляционной функции сохранять постоянное значение с ростом k может означать, что нужно рассматривать исследуемый процесс как нестационарный, но возможно, что его разность при этом стационарна. Отметим, что оцененные корреляции не обязательно должны быть очень высоки.

По изложенным причинам предполагается, что порядок разности d , обеспечивающий стационарность, достигнут тог-

да, когда автокорреляционная функция процесса $w_t = \nabla^d x_t$ падает довольно быстро. На практике обычно $d = 1, 2$, и, как правило, достаточно проанализировать первые 15—20 значений автокорреляционной функции исходного ряда и его первых и вторых разностей.

Идентификация процесса $\nabla^d x_t$. Определив экспериментально d , изучаем затем общий вид оцененной автокорреляционной и частной автокорреляционной функций для разностей порядка d . На их основе выбираются p и q для АР и СС операторов. Для этого используются рассмотренные выше свойства (см. табл. 7.1). Кроме того, в работе [54] показано, что теоретически возможный максимум коэффициентов автокорреляции для процесса СС(q) определяется соотношениями:

$$\rho_k(\max) = \begin{cases} \cos\left(\frac{\pi}{l+1}\right), & \text{если } q+1 \text{ делится на } k; \\ \cos\left(\frac{\pi}{l+2}\right), & \text{если } q+1 \text{ не делится на } k, \end{cases}$$

где

$$l = \left[\frac{q+1}{k} \right] - \text{целая часть числа.}$$

$$\text{В частности, } \rho_1(\max) = \cos\left(\frac{\pi}{q+2}\right).$$

Особое практическое значение имеют АР и СС-процессы 1-го и 2-го порядков и простейший смешанный процесс (1, d , 1). Для моделей этих процессов имеют место закономерности и соотношения, сведенные в табл. 7.2.

Мы видим, что предварительные оценки параметров для моделей (0, d , 1) и (1, d , 1) находят решая квадратические уравнения, значит, формально будет получено по два решения. Однако лишь одно из каждой пары решений будет в области допустимых значений, определяемой условиями стационарности и обратимости. Из оставшихся случаев лишь (0, d , 2) представляет трудность. Для этого процесса приближенные оценки параметров находят итеративно следующим образом.

Имея оценки автоковариаций и соотношения (7.20):

$$R_0 = (1 + Q_1^2 + \dots + Q_q^2) \sigma_\varepsilon^2;$$

$$R_k = (-Q_k + Q_1 Q_{k+1} + \dots + Q_{q-k} Q_q) \sigma_\varepsilon^2, \quad k \geq 1,$$

Основные свойства и соотношения
для наиболее важных моделей АРИСС

Модель	(1, d, 0)	(0, d, 1)
Поведение ρ_k	убывает экспоненциально	только $\rho_1 \neq 0$
Поведение φ_{kk}	только $\varphi_{11} \neq 0$	доминирует экспоненциальное затухание
Предварительные оценки	$\hat{\varphi}_1 = r_1$	$r_1 = \frac{-\hat{Q}_1}{1 + \hat{Q}_1^2}$
Допустимая область	$-1 < \varphi_1 < 1$	$-1 < Q_1 < 1$
Модель	(2, d, 0)	(0, d, 2)
Поведение ρ_k	наложение затухающих экспонент и синусоид	только $\rho_1 \neq 0, \rho_2 \neq 0$
Поведение φ_{kk}	только $\varphi_{11} \neq 0, \varphi_{22} \neq 0$	доминирует наложение затухающих экспонент и синусоид
Предварительные оценки	$\hat{\varphi}_1 = \frac{r_1(1-r_2)}{1-r_1^2}$ $\hat{\varphi}_2 = \frac{r_2-r_1^2}{1-r_1^2}$	$r_1 = \frac{-\hat{Q}_1(1-\hat{Q}_2)}{1+\hat{Q}_1^2+\hat{Q}_2^2}$ $r_2 = \frac{-\hat{Q}_2}{1+\hat{Q}_1^2+\hat{Q}_2^2}$
Область допустимых значений	$-1 < \varphi_2 < 1$ $\varphi_2 + \varphi_1 < 1$ $\varphi_2 - \varphi_1 < 1$	$-1 < Q_2 < 1$ $Q_2 + Q_1 < 1$ $Q_2 - Q_1 < 1$
Модель	(1, d, 1)	
Поведение ρ_k	Экспоненциально убывает, начиная с $k=1$	

Модель	(1, d, 1)
Поведение φ_{kk}	Начиная с $k=1$, доминирует затухающая экспонента
Предварительные оценки	$r_1 = \frac{(1-\hat{Q}_1 \hat{\varphi}_1)(\hat{\varphi}_1 - \hat{Q}_1)}{1 + \hat{Q}_1^2 - 2\hat{\varphi}_1 \hat{Q}_1}; r_2 = r_1 \hat{\varphi}_1$
Область допустимых значений	$-1 < \varphi_1 < 1; -1 < Q_1 < 1$

можно подсчитать последовательно в указанном порядке оценки параметров $\sigma_\varepsilon^2, Q_q, Q_{q-1}, \dots, Q_1$, используя итерацию:

$$\sigma_\varepsilon^2 = \frac{\hat{R}_0}{1 + \hat{Q}_1^2 + \dots + \hat{Q}_q^2};$$

$$Q_j = - \left(\frac{\hat{R}_j}{\sigma_\varepsilon^2} - Q_1 Q_{j+1} - Q_2 Q_{j+2} - \dots - Q_{q-j} Q_q \right) \quad j = q, \dots, 1,$$

при условии, что $Q_0 = 0$. Для начала итерации параметры $Q_1, Q_2, \dots, Q_q$ приравняются нулю. В частности, для процесса (0, d, 2) уравнения имеют вид:

$$\sigma_\varepsilon^2 = \frac{\hat{R}_0}{1 + \hat{Q}_1^2 + \hat{Q}_2^2};$$

$$Q_2 = - \frac{\hat{R}_2}{\sigma_\varepsilon^2};$$

$$Q_1 = - \left(\frac{\hat{R}_1}{\sigma_\varepsilon^2} - Q_1 Q_2 \right).$$

При получении оценок, конечно, необходимо сверяться с областью допустимых значений.

Интерпретация R^2 в моделях АРСС

Для анализа моделей авторегрессии и скользящего среднего также можно ввести статистику R^2 , как это делается в множественной регрессии¹:

$$R^2 = 1 - \frac{\sigma_e^2}{D(x)}, \quad (7.28)$$

где σ_e^2 — дисперсия белого шума;

$D(x)$ — дисперсия временного ряда.

Величина R^2 изменяется от 0 до 1 и представляет собой относительную меру определенности прогноза, получаемого на основе прошлой информации. В работе Ч. Нельсона [88] показано, как R^2 связан с автокорреляциями временного ряда и характеристиками генерирующего процесса. Большое внимание уделено интерпретации R^2 . Эта величина может быть особенно полезна при построении модели на стадии идентификации, когда ее предварительная оценка дает исследователю представление о прогнозируемости временного ряда и помогает избежать дальнейшего бесполезного поиска лучших моделей, повышающих R^2 .

Рассмотрим временной ряд x_t , генерируемый моделью авторегрессии — скользящего среднего:

$$x_t = Q_0 + \sum_{i=1}^p \varphi_i x_{t-i} + \varepsilon_t - \sum_{j=1}^q Q_j \varepsilon_{t-j},$$

где φ и Q — фиксированные параметры, удовлетворяющие условиям стационарности и обратимости соответственно, а ε_t — белый шум.

Этот процесс может быть записан в обращенном виде как процесс скользящего среднего бесконечного порядка:

$$x_t = \varepsilon_t + \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i},$$

где ψ_i являются функциями от φ_i и Q_i и

$$D(x) = \sigma_e^2 \left(1 + \sum_{i=1}^{\infty} \psi_i^2 \right). \quad (7.29)$$

¹ Далее, в § 3, показано, что дисперсия ошибки прогноза по модели АРСС на 1 шаг вперед равна σ_e^2 .

В результате подстановки (7.29) в (7.28) получаем

$$R^2 = \frac{\sum_{i=1}^{\infty} \psi_i^2}{1 + \sum_{i=1}^{\infty} \psi_i^2}. \quad (7.30)$$

Таким образом, R^2 не зависит от σ_e^2 . Подсчет величины R^2 по формуле (7.30) на стадии идентификации затруднителен. Поэтому, учитывая, что веса ψ определяют автокорреляционную функцию временного ряда, целесообразно связать R^2 с коэффициентами автокорреляции. Проще всего это сделать для чистых авторегрессионных процессов:

$$x_t = \sum_{i=1}^p \varphi_i x_{t-i} + \varepsilon_t.$$

Обозначая автоковариации процесса по-прежнему символом R_j , получаем, что дисперсия x_t определяется выражением:

$$D(x_t) = M(x_t^2) = \varphi_1 R_1 + \dots + \varphi_p R_p + \sigma_e^2.$$

Так что, обозначая R^2 для АР-процесса порядка p через R_p^2 , имеем:

$$1 - R_p^2 = \frac{\sigma_e^2}{D(x)} = 1 - \varphi_1 \rho_1 - \dots - \varphi_p \rho_p;$$

$$R_p^2 = \rho_p' \varphi_p,$$

где ρ_p — вектор-столбец $(\rho_1, \dots, \rho_p)'$;

φ_p — вектор-столбец $(\varphi_1, \dots, \varphi_p)'$.

Далее, используя оценки параметров Юла-Уокера (см. 7.19)

$$\varphi_p = P_p^{-1} \rho_p,$$

где P_p — автокорреляционная матрица, получаем R^2 как функцию одних лишь коэффициентов автокорреляций $(\rho_1, \dots, \rho_p)$:

$$R_p^2 = \rho_p' P_p^{-1} \rho_p. \quad (7.31)$$

В частности, для процесса АР (1)

$$R_1^2 = \rho_1^2,$$

для процесса АР(2)

$$R_2^2 = \frac{\rho_1^2 + \rho_2^2 - 2\rho_1 \rho_2}{1 - \rho_1^2}.$$

Возможен еще один альтернативный способ выражения R^2 , который с точки зрения интерпретации особенно полезен.

Рассмотрим долю дисперсии x_t , которая может быть объяснена регрессией x_t на x_{t-1} , затем регрессией на x_{t-1} и x_{t-2} , потом на x_{t-1} , x_{t-2} и x_{t-3} и т. д. Если опять обозначить R^2 , характеризующий эти последовательные регрессии через R_1^2 , R_2^2 , R_3^2 и т. д., то из теории линейной регрессии следует, что

$$R_j^2 = \varphi_{jj}^2 (1 - R_{j-1}^2) + R_{j-1}^2, \quad j = 1, 2, 3, \dots \quad (7.32)$$

где φ_{jj} — коэффициент частной автокорреляции.

Это выражение означает, что дополнительный вклад в R^2 при последовательном добавлении новой лаговой объясняющей переменной определяется коэффициентом частной автокорреляции для данного лага, который показывает, какая доля необъясненной до этого дисперсии x_t теперь получает объяснение. Становится понятным, почему R_p^2 для процесса АР(p) зависит только от $(\rho_1, \dots, \rho_p)$. Это объясняется тем, что $\varphi_{jj} = 0$ для $j > p$, а также тем, что, как это следует из уравнений Юла-Уокера, частные автокорреляции до лага p включительно зависят только от $(\rho_1, \dots, \rho_p)$.

На практике можно считать, что добавление новых лаговых объясняющих переменных (увеличение порядка эмпирической модели) даст малый вклад в R^2 , если выборочные оценки частных автокорреляций для этих лагов малы.

Анализ R^2 для моделей чистого скользящего среднего СС (q)

$$x_t = Q_0 + \varepsilon_t + \sum_{j=1}^q Q_j \varepsilon_{t-j}$$

легко проводить, если известны параметры Q , так как веса ψ_i в (7.30) в этом случае равны коэффициентам Q_i :

$$R^2 = \frac{\sum_{i=1}^q Q_i^2}{1 + \sum_{i=1}^q Q_i^2} \quad (7.33)$$

Отсюда ясно, что процессы с относительно большими коэффициентами более предсказуемы.

Однако, поскольку стадия идентификации предшествует процедуре оценивания модели, желательно опять установить связь R^2 с автокорреляциями. В случае СС (1) легко получить точное соотношение:

$$R^2 = \frac{Q_1^2}{1 + Q_1^2} = \frac{1}{2} - \left\{ \frac{1}{4} - \rho_1^2 \right\}^{1/2}. \quad (7.34)$$

Но, как мы уже видели, функциональная связь между коэффициентами Q и автокорреляциями процесса СС нелинейна и именно поэтому в общем случае трудно выразить R^2 через автокорреляции. Есть два пути для выхода из положения. Первый — это учет того, что любой обратимый процесс СС можно аппроксимировать АР-процессом высокого порядка:

$$x_t \approx \pi_1 x_{t-1} + \dots + \pi_k x_{t-k} + \varepsilon_t.$$

Соответствующий этой аппроксимации R^2 подсчитывается по формуле (7.31). Увеличивая k для получения более точной аппроксимации, получаем

$$R^2 = \lim_{k \rightarrow \infty} R_k^2 = \lim_{k \rightarrow \infty} (\rho_k' P_k^{-1} \rho_k). \quad (7.35)$$

Второй способ аппроксимации R^2 для процессов СС состоит в использовании рекуррентного соотношения (7.32):

$$R^2 = \lim_{k \rightarrow \infty} R_k^2 = \lim_{k \rightarrow \infty} \{ \varphi_{kk}^2 (1 - R_{k-1}^2) + R_{k-1}^2 \}. \quad (7.36)$$

На практике аппроксимация АР-процессом даже довольно низкого порядка обычно дает удовлетворительную оценку R^2 .

Легко заметить, что какой бы ни была гипотеза относительно порядка процесса СС (q) мы получаем одну единственную АР-аппроксимацию. Отсюда, конечно, не следует, что любая модель СС (q) будет давать один и тот же R^2 . Это просто означает, что аппроксимация дает оценку теоретически достижимого R^2 , на основе которого можно делать выводы о достаточности порядка модели СС (q). Например, если R^2 для модели СС(1), подсчитанный по формуле (7.34), намного меньше R^2 , полученного путем АР-аппроксимации, то целесообразно увеличить порядок модели q .

Случай смешанных процессов АРСС столь же сложен, как и случай процессов СС, так как опять уравнения, свя-

зывающие параметры с автокорреляциями, нелинейны. Здесь также используются оценки R^2 по формулам (7.35) или (7.36).

В заключение еще раз подчеркнем, что оценки автокорреляций, лежащих в основе процедуры идентификации, могут иметь довольно большие дисперсии и быть сильно автокоррелированными. По этой причине нет строгого соответствия между теоретической и оцененной автокорреляционными функциями. Это приводит к затруднениям при выборе p, d, q и на данной стадии могут быть выбраны две или несколько различных моделей для дальнейшего исследования. На этой стадии особенно полезны графические методы, опыт самого исследователя. В сомнительных случаях может быть целесообразней использовать нестационарную модель, чем стационарную, ибо она является более гибкой. Выбор пробных моделей на этой стадии — отправной пункт для применения более формальных и эффективных методов оценивания.

§ 3. ОЦЕНИВАНИЕ МОДЕЛЕЙ И ПРОГНОЗИРОВАНИЕ

На стадии идентификации были получены грубые оценки параметров модели, теперь необходимо получить такие оценки параметров, которые приводят к наиболее точным прогнозам.

Предположим, что находимся в точке t и должны сделать прогноз $\hat{x}_\tau(t)$ величины $x_{t+\tau}$. Пусть для получения прогноза используется линейная комбинация бесконечного числа значений $e_t, e_{t-1}, \dots$ — реализаций белого шума к моменту t . Допустим далее, что наилучший прогноз может быть представлен в виде

$$\hat{x}_\tau(t) = \psi_\tau^* e_t + \psi_{\tau+1}^* e_{t-1} + \psi_{\tau+2}^* e_{t-2} + \dots,$$

где веса $\psi_\tau^*, \psi_{\tau+1}^*, \dots$ должны быть определены. Тогда, учитывая, что $x_{t+\tau} = \sum_{j=0}^{\infty} \psi_j e_{t+\tau-j}$, средний квадрат ошибки прогноза равен:

$$M[x_{t+\tau} - \hat{x}_\tau(t)]^2 = (1 + \psi_1^2 + \dots + \psi_{\tau-1}^2) \sigma_e^2 + \sum_{j=0}^{\infty} (\psi_{\tau+j} - \psi_{\tau+j}^*)^2 \sigma_e^2.$$

Он достигает минимума при $\psi_{\tau+j}^* = \psi_{\tau+j}$. Таким образом, значение прогнозируемого члена ряда можно представить в виде двух групп слагаемых:

$$x_{t+\tau} = (e_{t+\tau} + \psi_1 e_{t+\tau-1} + \dots + \psi_{\tau-1} e_{t+1}) + (\psi_\tau e_t + \psi_{\tau+1} e_{t-1} + \dots) = e_\tau(t) + \hat{x}_\tau(t),$$

где

$$e_\tau(t) = e_{t+\tau} + \psi_1 e_{t+\tau-1} + \dots + \psi_{\tau-1} e_{t+1} \quad (7.37)$$

есть ошибка прогноза $\hat{x}_\tau(t)$.

Ошибка прогноза на 1 шаг вперед будет

$$e_1(t) = x_{t+1} - \hat{x}_1(t) = e_{t+1}.$$

Как видим, ошибка прогнозирования на один шаг вперед оказывается равной значению белого шума, генерирующего процесс в следующий момент времени. И следовательно, дисперсия белого шума будет также дисперсией ошибки прогнозов на один шаг вперед.

Самый простой и наглядный способ получения модели с минимальным средним квадратом ошибки прогноза — это графическое изучение $\sum \hat{e}_i^2$ суммы квадратов расхождений между реальными членами ряда и их прогнозами, сделанными на предыдущем шаге, в зависимости от значений параметров, задаваемых в окрестности их предварительных оценок. Минимум этой суммы будет определять искомые значения параметров. При условии нормальности распределения множества значений $\hat{e}$ эти оценки будут совпадать с оценками, полученными по методу максимума правдоподобия.

Существуют и другие способы поиска оценок параметров. Например, алгоритм Марквардта для нелинейного метода наименьших квадратов. Различные модификации этого алгоритма можно найти в [74] и в [5]. Рассмотрим лишь один из способов оценивания, предложенный Боксом и Дженкинсом.

Общая модель может быть записана как $\hat{e}_t = Q^{-1}(B)\varphi(B)\tilde{w}_t$,

где $w_t = \nabla^d x_t$, $\tilde{w}_t = w_t - \mu_w$.

Необходимо минимизировать $\sum \hat{e}_i^2$.

Пусть $\beta = (\varphi, Q)$ — вектор параметров модели, а $\beta_0 = (\varphi_0, Q_0)$ — вектор их предварительных оценок. Рассматривая каждое значение $\hat{\varepsilon}_t$ как функцию от β , можно разложить $\hat{\varepsilon}_t$ в ряд Тейлора около его значения $\hat{\varepsilon}_{t,0}$, соответствующего вектору предварительных оценок параметров β_0 :

$$\hat{\varepsilon}_t = \hat{\varepsilon}_{t,0} - \sum_{j=1}^p (\varphi_j - \varphi_{j,0}) u_{j,t} - \sum_{i=1}^q (Q_i - Q_{i,0}) v_{i,t}, \quad (7.38)$$

где

$$v_{i,t} = - \left. \frac{\partial \hat{\varepsilon}_t}{\partial Q_i} \right|_{\beta=\beta_0} = v_{t-i};$$

$$u_{j,t} = - \left. \frac{\partial \hat{\varepsilon}_t}{\partial \varphi_j} \right|_{\beta=\beta_0} = u_{t-j};$$

$$u_t = Q_0^{-1}(B) \tilde{w}_t = \varphi_0^{-1}(B) \hat{\varepsilon}_{t,0};$$

$$v_t = -Q_0^{-2}(B) \varphi_0(B) \tilde{w}_t = -Q_0^{-1}(B) \hat{\varepsilon}_{t,0}.$$

Если начальные значения $\hat{\varepsilon}$, v , u положить равными нулю, то остальные могут быть подсчитаны рекурсивно следующим образом:

$$\hat{\varepsilon}_{t,0} = \tilde{w}_t - \varphi_{1,0} \tilde{w}_{t-1} - \dots - \varphi_{p,0} \tilde{w}_{t-p} +$$

$$+ Q_{1,0} \hat{\varepsilon}_{t-1,0} + \dots + Q_{q,0} \hat{\varepsilon}_{t-q,0};$$

$$u_t = \varphi_{1,0} u_{t-1} + \dots + \varphi_{p,0} u_{t-p} + \hat{\varepsilon}_{t,0};$$

$$v_t = Q_{1,0} v_{t-1} + \dots + Q_{q,0} v_{t-q} - \hat{\varepsilon}_{t,0}.$$

В соответствии с (7.38) получаем приближенное уравнение линейной регрессии

$$\hat{\varepsilon}_{t,0} = \sum_{j=1}^p (\varphi_j - \varphi_{j,0}) u_{t-j} + \sum_{i=1}^q (Q_i - Q_{i,0}) v_{t-i} + \hat{\varepsilon}_t, \quad (7.39)$$

в котором $\hat{\varepsilon}_t$ играет роль остатка. Тогда корректировками параметров, минимизирующих $\sum \hat{\varepsilon}_t^2$, оказываются коэффициенты регрессии $\hat{\varepsilon}_{t,0}$ на u_{t-j} и v_{t-i} . Добавляя корректировки к первым приближенным оценкам (φ_0, Q_0) , получаем набор вторых приближений, которые во второй итерации заменяют вектор β_0 . Итерация проводится до тех пор, пока не будет достигнута сходимость.

Поскольку итеративный цикл начинается с нулевых значений $\hat{\varepsilon}$, v , u , то несколько их первых значений будут грубыми, они отбрасываются и не используются в уравнении (7.39). Из этого уравнения легко получить частные случаи для различных p и q .

При построении модели следует проанализировать нет ли параметрической избыточности, нельзя ли упростить структуру. Наличие избыточности не всегда очевидно. Например, общий множитель слева и справа в модели

$$(1 - 1,3B + 0,4B^2) \tilde{w}_t = (1 - 0,5B) \varepsilon_t$$

можно увидеть только после выделения множителей в левой части

$$(1 - 0,5B)(1 - 0,8B) \tilde{w}_t = (1 - 0,5B) \varepsilon_t,$$

т. е.

$$(1 - 0,8B) \tilde{w}_t = \varepsilon_t.$$

Модель с параметрической избыточностью создает серьезные трудности в процедуре оценивания. На практике трудности создают не только случаи точного сокращения, но и близкие к этому. Оценки при этом будут нестабильными из-за почти одинаковых множителей в обеих частях модели.

В табл. 7.3 приведены ковариационные матрицы оценок параметров для моделей, представляющих особый практический интерес.

Таблица 7.3

Ковариационные матрицы оценок параметров некоторых моделей ($n = N - d$)

Процесс	Ковариационная матрица
АР (1)	$V(\hat{\varphi}) = n^{-1}(1 - \varphi^2)$
АР (2)	$V(\hat{\varphi}_1, \hat{\varphi}_2) = n^{-1} \begin{bmatrix} 1 - \varphi_1^2 & -\varphi_1(1 + \varphi_2) \\ -\varphi_1(1 + \varphi_2) & 1 - \varphi_2^2 \end{bmatrix}$
СС (1)	$V(\hat{Q}) = n^{-1}(1 - Q^2)$
СС (2)	$V(\hat{Q}_1, \hat{Q}_2) = n^{-1} \begin{bmatrix} 1 - Q_1^2 & -Q_1(1 + Q_2) \\ -Q_1(1 + Q_2) & 1 - Q_2^2 \end{bmatrix}$
АРСС (1, 1)	$V(\hat{\varphi}, \hat{Q}) = n^{-1} \frac{1 - \varphi Q}{(\varphi - Q)^2} \begin{bmatrix} (1 - \varphi^2)(1 - \varphi Q) & (1 - \varphi^2)(1 - Q^2) \\ (1 - \varphi^2)(1 - Q^2) & (1 - Q^2)(1 - \varphi Q) \end{bmatrix}$

Из табл. 7.3 видно, что при $\varphi = Q$ дисперсии оценок $\hat{\varphi}$ и $\hat{Q}$ в модели АРСС (1,1) бесконечны. Это и есть случай сокращающихся множителей АР и СС-частей.

Диагностическая проверка модели на адекватность

Одним из способов проверки является построение модели несколько более общей, чем та, которая считается истинной, и сравнение этих моделей, выявление незначимых параметров. Данный метод предполагает, что можно определить слабое звено в модели, из-за которого она оказалась неадекватна. Однако сделать это довольно затруднительно, особенно когда выборка невелика. Более формальная проверка основана на анализе автокорреляционной функции остатков. Рассмотрим ее.

Анализ остатков. Если бы модель была адекватна исследуемому процессу, то ε_i были бы некоррелированы и оценки $r_\varepsilon(k)$ распределены приблизительно нормально около нуля с дисперсией, равной n^{-1} , или соответственно со стандартной ошибкой $n^{-1/2}$. В нашем случае $n = N - d$, где N — число наблюдений в выборке, а d — порядок разности. Это положение можно использовать для приближенной оценки статистической значимости отклонений оценок этих автокорреляций от нуля. Однако Дж. Дарбин [56] предостерегает от недооценки статистической значимости отклонений автокорреляций $r_\varepsilon(k)$ от их теоретических нулевых значений, если при малых лагах k исходить из стандартной ошибки $n^{-1/2}$. Он, например, показал, что для процесса АР (1) с параметром φ дисперсия $r_\varepsilon(1)$ есть $\varphi^2 n^{-1}$, а это может быть существенно меньше чем n^{-1} . Кроме случаев сравнительно больших лагов, $n^{-1/2}$ следует рассматривать как верхнюю границу стандартных ошибок для $r_\varepsilon(k)$.

Помимо индивидуального анализа коэффициентов $r_\varepsilon(k)$ возможен общий совокупный тест автокорреляционной функции остатков. При этом задаются целью выяснить не свидетельствуют ли первые, скажем, 20 автокорреляций остатков, взятые вместе, о неадекватности модели.

Пусть имеется K оценок автокорреляций

$$r_\varepsilon(k) \quad k = 1, 2, \dots, K.$$

Тогда можно показать [45], что если построенная модель адекватна временному ряду, то случайная величина

$$\Theta = n \sum_{k=1}^K r_\varepsilon^2(k), \quad n = N - d$$

распределена как $\chi^2(K - p - q)$. Если же модель не соответствует временному ряду, автокорреляции остатков будут существенными, а величина Θ большой. Таким образом, общую проверку гипотезы об адекватности модели можно осуществить путем сопоставления Θ с таблицей χ^2 с $(K - p - q)$ -степенями свободы.

Неадекватность модели может иметь место из-за изменений значений параметров во времени при сохранении формы модели. Для выявления этой особенности продлевается следующая операция. Исследуемый отрезок данных разбивается на два и к каждому привязывается модель.

Пометим параметры моделей для каждой из половин отрезка соответствующими индексами $Q_i^{(1)}, Q_i^{(2)}, \varphi_j^{(1)}, \varphi_j^{(2)}$. Им соответствуют стандартные ошибки оценок $\sigma_{Q_i^{(1)}}, \sigma_{Q_i^{(2)}}, \sigma_{\varphi_j^{(1)}}, \sigma_{\varphi_j^{(2)}}$. Отсюда находим, что стандартные ошибки разностей

$$Q_i^{(1)} - Q_i^{(2)} \text{ и } \varphi_j^{(1)} - \varphi_j^{(2)} \text{ будут}$$

$$\sigma_{\Delta Q_i} = (\sigma_{Q_i^{(1)}}^2 + \sigma_{Q_i^{(2)}}^2)^{1/2}; \quad \sigma_{\Delta \varphi_j} = (\sigma_{\varphi_j^{(1)}}^2 + \sigma_{\varphi_j^{(2)}}^2)^{1/2}.$$

Сопоставляя разности параметров с их стандартными ошибками, можно сделать вывод о вероятности действительного изменения параметров. Если такие изменения имели место, то берутся оценки параметров, полученные на основе более свежих данных.

Использование R^2 для проверки статистических гипотез. Опыт свидетельствует о том, что часто уже после оценивания нескольких альтернативных моделей исследователь оказывается в затруднении, не зная, какой модели отдать предпочтение. На этой стадии целесообразно получить уточненные оценки R^2 , которые в работе Ч. Нельсона [88] предложено использовать для проверки статистических гипотез. Для этого разработана асимптотическая теория, основывающаяся на анализе отношения правдоподобий.

Пусть β означает вектор $(\varphi_1, \dots, \varphi_p; Q_1, \dots, Q_q)$, а β_s — вектор подмножества элементов вектора β . Часто исследо-

вателя интересует проверка гипотезы того, что все элементы β_s равны нулю, т. е. что соответствующие s параметров не нужно включать в модель. В более общем виде можно рассмотреть гипотезу вида

$$\beta_s = \beta_s^*, \quad (7.40)$$

где β_s^* — специфицированный набор значений s параметров.

Обозначим максимальное значение функции правдоподобия для вектора β через $L(\hat{\beta})$, а для вектора β_c , на который наложены ограничения (7.40), — через $L(\hat{\beta}_c)$. Отношение $L(\hat{\beta}_c)$ к $L(\hat{\beta})$ показывает, насколько хорошо гипотеза (7.40) соответствует данным. Проверка гипотезы может быть основана на том факте, что для больших выборок статистика

$$\lambda = -2 \ln [L(\hat{\beta}_c)/L(\hat{\beta})] \quad (7.41)$$

распределена как χ^2 с s -степенями свободы. Если гипотеза неверна, то λ принимает большое значение, и в этом случае гипотеза (7.40) отвергается.

Легко показать, что в случае нормального закона распределения (7.41) сводится к

$$\lambda = \frac{\sum \hat{\epsilon}_t^2(\hat{\beta}_c) - \sum \hat{\epsilon}_t^2(\hat{\beta})}{\sigma_e^2},$$

где $\hat{\epsilon}_t(\hat{\beta})$ — остаток, подсчитанный для момента t как функция параметров $\hat{\beta}$.

Выборочная оценка R^2 для модели без наложения ограничений на параметры определяется как

$$R^2 = 1 - \frac{\sum \hat{\epsilon}_t^2(\hat{\beta})}{\sum (x_t - \bar{x})^2},$$

а для модели с наложенными ограничениями

$$R_c^2 = 1 - \frac{\sum \hat{\epsilon}_t^2(\hat{\beta}_c)}{\sum (x_t - \bar{x})^2},$$

таким образом, статистический критерий принимает вид

$$\lambda = \frac{R^2 - R_c^2}{\sigma_e^2 / \sum (x_t - \bar{x})^2},$$

все еще зависящий от неизвестного параметра σ_e^2 . Если вместо σ_e^2 подставить оценку $n^{-1} \sum \hat{\epsilon}_t^2(\hat{\beta})$, где $n = N - d$,

N — число наблюдений, то получим выражение¹, удобное для вычислений:

$$\lambda = n \left(\frac{R^2 - R_c^2}{1 - R^2} \right). \quad (7.42)$$

Статистика λ определяется относительным увеличением R^2 в результате устранения ограничений на параметры. Если увеличение большое, то гипотеза (7.40) отвергается.

Теория распределений λ является асимптотической. Между тем на практике часто имеется очень немного данных для получения выборочной оценки σ_e^2 . В связи с этим по аналогии с соответствующей ситуацией в линейной регрессии, для которой разработана теория малой выборки, предлагается также несколько видоизмененная статистика:

$$\frac{n-p-q-1}{s} \cdot \frac{R^2 - R_c^2}{1 - R^2} \approx F(s, n-p-q-1), \quad (7.43)$$

где F имеет F -распределение с s и $(n-p-q-1)$ -степенями свободы (одна степень свободы вычтена в предположении, что константа определена). Распределения F и χ^2 будут давать практически одинаковые результаты, если $(n-p-q-1)$ порядка ста или выше. Об адекватности двух статистик при выборках обычного для практики объема ничего неизвестно.

В частном случае, когда $s = 1$, F превращается в t -распределение с $(n-p-q-1)$ -степенями свободы.

На практике часто выдвигается гипотеза $\beta = 0$, т. е. что все $(p+q)$ параметров равны нулю. В этом случае $\lambda = n [R^2/(1-R^2)]$, так как $R_c^2 = 0$.

Итак, имея оценки R^2 , можно вычислить значения F или λ и, сопоставив их с табличными значениями, отвергнуть или принять проверяемую гипотезу.

При сравнении моделей возможны ошибки двух видов. Во-первых, может быть отвергнута модель с меньшим числом параметров, которая на самом деле адекватна ряду, и принята более общая модель, оценки некоторых параметров которой будут незначимыми. Другая возможная ошибка состоит в том, что может быть отвергнута более общая адекватная модель и принята модель с ограничениями, что приведет к некорректным значениям параметров. Следовательно, при анализе статистических критериев нежелательно поль-

¹ Нельсон показывает, что эта статистика приближенно равна той, которую используют А. Зельнер и Ф. Пальм в [115].

зоваться очень низкими уровнями значимости, такими, как 0,01 или 0,001. Это в какой-то мере оправдывает обычную практику принятия решения в пользу более общей модели в сомнительных случаях.

Недостаток рассмотренных статистических критериев в том, что они позволяют сравнивать две модели лишь в том случае, когда одна из них является более общей по отношению к другой, т.е. $(p_1 \leq p_2; q_1 \leq q_2)$, и не могут быть использованы в тех случаях, когда $(p_1 < p_2; q_1 > q_2)$ или $(p_1 > p_2; q_1 < q_2)$.

После того как модель выдержала диагностическую проверку и статистические критерии свидетельствуют о ее обоснованности, можно переходить к прогнозированию.

Прогнозирование

Пусть в момент t необходимо сделать прогноз величины $x_{t+\tau}$, $\tau \geq 1$.

Проще всего осуществлять прогнозирование рекурсивно непосредственно по разностному уравнению $x_{t+i} = \Psi_1 x_{t+i-1} + \dots + \Psi_{p+d} x_{t+i-p-d} - Q_1 e_{t+i-1} - \dots - Q_q e_{t+i-q}$, последовательно полагая $i = 1, 2, \dots, \tau$ и заменяя x_j для $t < j < t+i$ их прогнозами, полученными на предыдущих шагах.

Доверительные границы прогнозов. Дисперсия ошибки прогноза на τ шагов вперед есть математическое ожидание от $e_\tau^2(t) = \{x_{t+\tau} - \hat{x}_\tau(t)\}^2$, и согласно (7.37) дается выражением:

$$D_e(\tau) = \left\{ 1 + \sum_{j=1}^{\tau-1} \psi_j^2 \right\} \sigma_e^2.$$

Оценка σ_e^2 может быть получена уже на стадии оценки параметров как средний квадрат ошибок ретроспективного прогнозирования на один шаг вперед. Веса ψ можно получить исходя из следующего:

$$x_t = \psi(B)e_t; \quad (7.44)$$

$$\Psi(B)x_t = Q(B)e_t. \quad (7.45)$$

Подставляем (7.44) в (7.45):

$$\Psi(B)\psi(B)e_t = Q(B)e_t,$$

или $\Psi(B)\psi(B) = Q(B)$.

Приравнявая коэффициенты при одинаковых степенях B слева и справа, получаем:

$$\psi_0 = 1;$$

$$\psi_1 = \Psi_1 - Q_1;$$

$$\psi_2 = \Psi_1 \psi_1 + \Psi_2 - Q_2;$$

$$\psi_j = \Psi_1 \psi_{j-1} + \dots + \Psi_{p+d} \psi_{j-p-d} - Q_j,$$

где для $j < 0$ $\psi_j = 0$ и для $j > q$ $Q_j = 0$.

Таким образом, коэффициенты ψ легко подсчитываются рекурсивно. Тогда прогноз с вероятностью $1 - v$ не будет выходить за пределы

$$x_{t+\tau}(\pm) = \hat{x}_\tau(t) \pm u_{v/2} \left\{ 1 + \sum_{j=1}^{\tau-1} \psi_j^2 \right\}^{1/2} \hat{\sigma}_e,$$

где $u_{v/2}$ является квантилем уровня $1 - v/2$ стандартного нормального распределения. Это величина табличная:

$1 - v$	50%	95%
$u_{v/2}$	0,674	1,96

Адаптивные свойства модели. Рассмотрим модель $\nabla x_t = (1 - QB)e_t$. В момент $t + \tau$ она может быть записана:

$$x_{t+\tau} = x_{t+\tau-1} + e_{t+\tau} - Qe_{t+\tau-1}.$$

Берем условное математическое ожидание в момент t при известном до момента t прошлом. Учитывая, что

$$M_i(e_i) = \begin{cases} e_i & \text{для } i \leq t; \\ 0 & \text{для } i > t, \end{cases}$$

получаем

$$\hat{x}_1(t) = x_t - Qe_t; \quad (7.46)$$

$$\hat{x}_\tau(t) = \hat{x}_{\tau-1}(t) \quad \tau \geq 2.$$

Используя тот факт, что $x_t = \hat{x}_1(t-1) + \varepsilon_t$, (7.46) можно записать в виде $\hat{x}_\tau(t) = \hat{x}_\tau(t-1) + \lambda \varepsilon_t$, где $\lambda = 1 - Q$ или в виде $\hat{x}_\tau(t) = \lambda x_t + (1 - \lambda) \hat{x}_\tau(t-1)$, что является обычной моделью экспоненциальной средней (см. гл. 1), которая может быть записана и так:

$$\hat{x}_\tau(t) = \hat{a}_{1,t} = \hat{a}_{1,t-1} + \lambda \varepsilon_t,$$

где $\hat{a}_{1,t}$ — экспоненциальная средняя процесса в момент t ; $\hat{a}_{1,t}$ — адаптивный коэффициент.

Можно также показать, что предиктор процесса

$$\nabla^2 x_t = (1 - Q_1 B - Q_2 B^2) \varepsilon_t$$

приводится к модели линейного роста экспоненциального типа:

$$\begin{aligned} \hat{x}_\tau(t) &= \hat{a}_{1,t} + \tau \hat{a}_{2,t} \quad \tau > 0; \\ \hat{a}_{1,t} &= \hat{a}_{1,t-1} + \hat{a}_{2,t-1} + \lambda_1 \varepsilon_t; \\ \hat{a}_{2,t} &= \hat{a}_{2,t-1} + \lambda_2 \varepsilon_t; \\ \lambda_1 &= 1 + Q_2; \quad \lambda_2 = 1 - Q_1 - Q_2. \end{aligned}$$

Когда $\hat{x}_\tau(t)$ рассматривается как функция от τ при фиксированном t будем называть ее прогнозирующей функцией для момента t . Общий процесс $(0, d, q)$ имеет своей прогнозирующей функцией полином от τ степени $d - 1$.

$$\hat{x}_\tau(t) = \hat{a}_{1,t} + \tau \hat{a}_{2,t} + \tau^2 \hat{a}_{3,t} + \dots + \tau^{d-1} \hat{a}_{d-1,t} \quad (7.47)$$

для $\tau > q - d$.

Коэффициенты $\hat{a}_{i,t}$ должны последовательно обновляться с получением каждой новой фактической точки. Для $\tau \leq q - d$ прогноз будет зависеть от $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q+\tau}$, а после этого определяться полиномом (7.47).

Класс функций прогноза. Посмотрим, какими еще функциями можно описать будущее поведение ряда исходя из модели АРИСС.

При прогнозировании процесса

$$\Psi(B)x_t = Q(B)\varepsilon_t$$

будущие значения ε являются неизвестными и непредсказуемыми, поэтому заменим их при вычислении прогнозов нулем. Тогда класс функций прогноза для $\tau > q$ будет определяться

общим решением однородного линейного разностного уравнения:

$$\Psi(B)x_t = 0, \quad (7.48)$$

где

$$\begin{aligned} \Psi(B) &= (1 - G_1 B)(1 - G_2 B) \dots (1 - G_{p'} B), \\ p' &= p + d. \end{aligned} \quad (7.49)$$

Предположим сначала, что $G_1, G_2, \dots, G_{p'}$ различны. Тогда общее решение для момента k при условии, что ряд начинается в момент t (при прогнозировании t соответствует текущему моменту времени), будет:

$$x'_k = A_1 G_1^{k-t} + A_2 G_2^{k-t} + \dots + A_{p'} G_{p'}^{k-t},$$

где A_i — константы. В этом можно убедиться непосредственной подстановкой в исходное уравнение. Тогда при выполнении условия стационарности $|G_i| < 1$ действительный корень уравнения $\Psi(B) = 0$ даст в качестве слагаемого затухающую экспоненту G_i^{k-t} . Пара комплексных корней дает затухающую синусоидальную волну:

$$C^{k-t} \sin(2\pi f_0 + F).$$

Предположим теперь, что $\Psi(B) = 0$ имеет m равных корней G_0^{-1} , так что $\Psi(B)$ содержит множитель $(1 - G_0 B)^m$. Тогда подстановкой можно проверить, что общее решение будет:

$$\begin{aligned} x'_k &= \{A_0 + A_1(k-t) + A_2(k-t)^2 + \\ &+ \dots + A_{m-1}(k-t)^{m-1}\} G_0^{k-t}. \end{aligned}$$

В частности, при $G_0 = 1$ получаем полином от $k - t$ порядка $m - 1$.

Вообще, когда $\Psi(B)$ раскладывается на множители как $(1 - G_1 B)(1 - G_2 B) \dots (1 - G_p B)(1 - G_0 B^m)$ общее решение будет:

$$x'_k = G_0^{k-t} \sum_{j=0}^{m-1} A_j (k-t)^j + \sum_{i=1}^p C_i G_i^{k-t},$$

т. е. функция прогнозирования состоит из смеси затухающих экспонент, полиномиальных членов, затухающих синусоидальных волн и их комбинаций.

Если взять комплексные корни характеристического уравнения на границе стационарности $|G_i| = 1$, то модель

превращается в нестационарную и включает регулярные (незатухающие) периодические колебания. Например, функция прогноза, которая является синусоидальной волной с 12-точечным периодом, будет удовлетворять уравнению:

$$(1 - \sqrt{3} B + B^2) x'_t(t) = 0,$$

где оператор B относится к t .

Это означает, что модель Брауна по своей структуре и функциям прогнозирования является частным случаем модели АРИСС. Модель Брауна является более жесткой и при заданном наборе функций, включенных в нее, процесс прогнозирования полностью определяется единственным параметром β , который, кстати сказать, часто задается интуитивно. Поэтому для обработки протяженных рядов предпочтительней использовать метод Бокса-Дженкинса. Однако при небольшом количестве точек, когда этот метод бессилен, модель Брауна может оказаться полезной. Подбирать какие-либо сложные функции при малом объеме выборки нецелесообразно, поэтому для практического применения из всех моделей Брауна имеет смысл оставить лишь самые простые.

В экономических исследованиях включение в модель синусоид, даже с адаптивной фазой и амплитудой, редко приводит к удовлетворительному отображению сезонных колебаний, и в каждом случае требуется дополнительный анализ. Если включение синусоиды все же оказывается полезным, то удобнее проводить исследование по модели АРИСС для преобразованного ряда. Например, при 12-точечном периоде целесообразно перейти к анализу ряда $w_t = (1 - \sqrt{3}B + B^2)x_t$ и т. п.

Для более качественного анализа сезонных явлений Бокс и Дженкинс в [5] предлагают специальную схему. Учитывая, что в ней используются уже рассмотренные нами принципы и процедуры, не будем на этом останавливаться.

Сравнение процесса АРИСС с процессом Тейла—Вейджа

Процесс Тейла—Вейджа можно записать как

$$y_t = \nabla^2 x_t = x_t - 2x_{t-1} + x_{t-2} = u_t - 2u_{t-1} + u_{t-2} + v_t, \quad (7.50)$$

где u и v — независимые случайные переменные. Посмотрим, не является ли этот процесс частным случаем процессов класса АРИСС. Для достижения стационарности не-

обходимо взять вторую разность от исходного ряда x , т. е. $d = 2$. Автокорреляционная функция для вторых разностей в соответствии с (1.5) обрывается при лаге $k = 3$. Поэтому среди процессов АРИСС адекватными процессу Тейла—Вейджа могут оказаться только те, которые имеют порядок $(0, 2, 2)$, т. е.

$$y_t = x_t - 2x_{t-1} + x_{t-2} = \varepsilon_t - Q_1 \varepsilon_{t-1} - Q_2 \varepsilon_{t-2}. \quad (7.51)$$

Предположим, что (7.50) может быть эквивалентно представлен процессом (7.51). Тогда имеем соотношение

$$u_t - 2u_{t-1} + u_{t-2} + v_t = \varepsilon_t - Q_1 \varepsilon_{t-1} - Q_2 \varepsilon_{t-2}. \quad (7.52)$$

Рассмотрим условные математические ожидания. По (7.50) $M(y_t/x_{t-1}, x_{t-2}, \dots) = M(y_t/u_{t-1}, u_{t-2}, \dots; v_{t-1}, v_{t-2}, \dots) = -2u_{t-1} + u_{t-2}$, а по (7.51) $M(y_t/x_{t-1}, x_{t-2}, \dots) = M(y_t/\varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = -Q_1 \varepsilon_{t-1} - Q_2 \varepsilon_{t-2}$.

$$\text{Таким образом, } -2u_{t-1} + u_{t-2} = -Q_1 \varepsilon_{t-1} - Q_2 \varepsilon_{t-2}. \quad (7.53)$$

Далее, с одной стороны, $M(y_t/x_{t-2}, x_{t-3}, \dots) = M(y_t/u_{t-2}, u_{t-3}, \dots; v_{t-2}, v_{t-3}, \dots) = u_{t-2}$, а с другой — $M(y_t/x_{t-2}, x_{t-3}, \dots) = M(y_t/\varepsilon_{t-2}, \varepsilon_{t-3}, \dots) = -Q_2 \varepsilon_{t-2}$, т. е. $u_{t-2} = -Q_2 \varepsilon_{t-2}$.

$$\text{Вычтем (7.54) из (7.53) } -2u_{t-1} = -Q_1 \varepsilon_{t-1} \quad (7.55)$$

и (7.53) из (7.52)

$$v_t + u_t = \varepsilon_t. \quad (7.56)$$

Тогда связь между ε_i и u_i, v_i по (7.56) будет $\varepsilon_i = u_i + v_i$, а по (7.55) $\varepsilon_i = \frac{2}{Q_1} u_i$.

Это приводит к соотношениям:

$$u_i + v_i = \frac{2}{Q_1} u_i$$

или

$$v_i = \left(\frac{2}{Q_1} - 1 \right) u_i.$$

Таким образом, гипотеза адекватности двух рассматриваемых процессов приводит к тому, что два независимых стохастических процесса u и v выражаются один через другой. Это противоречие не устранимо ни при каких значениях Q_1 . Следовательно, процесс Тейла—Вейджа нельзя привести к процессу АРИСС с одним источником белого шума.

§ 4. ПРОГНОЗИРОВАНИЕ ПОСЛЕ ЛОГАРИФМИЧЕСКОГО ПРЕОБРАЗОВАНИЯ

В экономике часто встречаются негомогенные временные ряды, в которых рост в процентном измерении представляет, однако, гомогенный процесс. В этих случаях обычно переходят от исходного неоднородного ряда к натуральным логарифмам. Если к преобразованному ряду подобрать модель АРИСС, то прогнозы, генерированные этой моделью, будут, конечно, прогнозами логарифмов будущих наблюдений. Представляется естественным обратный переход от логарифмов к исходному представлению операций антилогарифмирования полученного прогноза. Но даст ли такая операция условное математическое ожидание будущего значения исходного ряда? Ответ будет отрицательным, так как если обозначить исходный ряд через x_t , а $\ln(x_t)$ — через y_t , предложенная операция будет давать оценку $\exp[M(y_{t+1}/X)]$, где X — совокупность прошлых значений x , а это, вообще говоря, не является математическим ожиданием.

Прежде чем продолжить рассуждения, рассмотрим логнормальное распределение [86].

Пусть y — случайная переменная с нормальным распределением $N(\mu, \sigma^2)$ и x — случайная переменная, связанная с y соотношением $x = e^y$. Тогда распределение x описывается логнормальной функцией плотности:

$$p(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(\ln x - \mu)^2\right] & x > 0; \\ 0 & x \leq 0, \end{cases}$$

где x принимает только положительные значения и при $x = 0$ $p(x) = 0$. Мода распределения будет в точке x равна $e^{\mu - \sigma^2}$, а медиана в точке x равна e^{μ} . Моменты относительно начала отсчета:

$$m_j = e^{j\mu + \frac{1}{2}j^2\sigma^2} \quad j = 1, 2, \dots$$

Отсюда математическое ожидание

$$M(x) = m_1 = e^{\mu + \frac{1}{2}\sigma^2},$$

а дисперсия

$$D(x) = m_2 - m_1^2 = e^{2\mu + 2\sigma^2}(e^{\sigma^2} - 1).$$

На рис. 7.2 представлены кривые нормального и логнормального распределений. Отметим относительное расположение моды, медианы и математического ожидания для логнормального распределения. Иллюстрируются также логнормальные кривые для различных значений μ и σ^2 .

Эти результаты непосредственно применимы к прогнозированию, когда логарифм исходного ряда представляется нормальным процессом АРИСС. Модель генерирует прогноз $y_{t+\tau}$ как условное распределение $y_{t+\tau} \sim N\{\hat{y}_\tau(t), D_e(\tau) | X\}$ при данной выборке X . Следовательно, условное математическое ожидание члена $x_{t+\tau}$ дается выражением: $\hat{x}_\tau(t) = \exp\{\hat{y}_\tau(t) + \frac{1}{2}D_e(\tau)\}$, а дисперсия $x_{t+\tau}$, или, другими словами, дисперсия ошибки прогноза $e_\tau(t) = x_{t+\tau} - \hat{x}_\tau(t)$, равна: $D(x_{t+\tau} | X) = D_e(\tau) = \exp\{2\hat{y}_\tau(t) + D_e(\tau)\} \times \{\exp[D_e(\tau)] - 1\}$.

Далее, если $\hat{y}_\tau(t) \pm k\hat{\sigma}_e(\tau)$ образует $v\%$ -ный доверительный интервал для $y_{t+\tau}$, то $\exp\{\hat{y}_\tau(t) \pm k\hat{\sigma}_e(\tau)\}$ об-

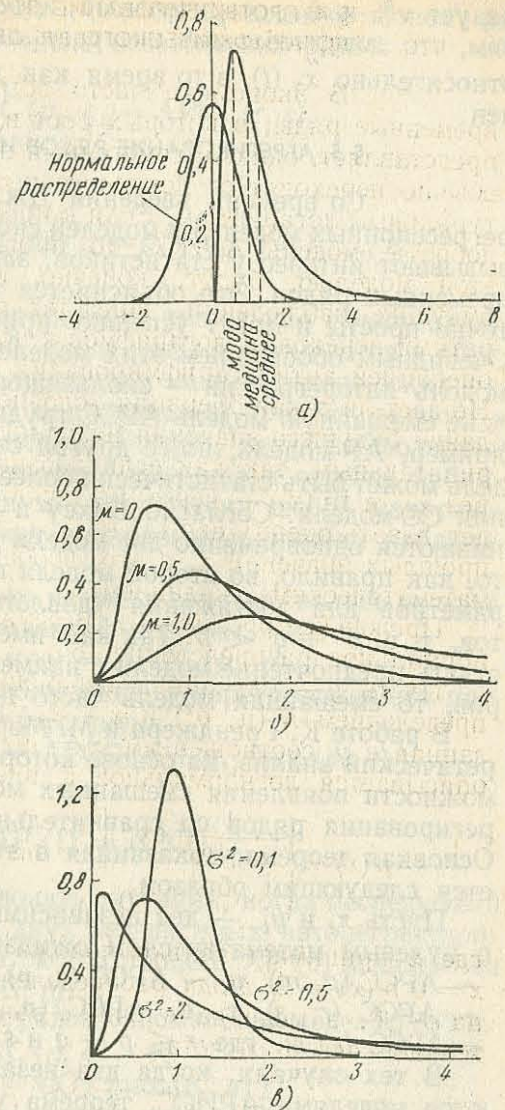


Рис. 7.2. Кривые нормального и логнормального распределений

разует $\nu\%$ -ный доверительный интервал для $x_{t+\tau}$. Заметим, что доверительный интервал для $x_{t+\tau}$ несимметричен относительно $\hat{x}_\tau(t)$, в то время как для $y_{t+\tau}$ он симметричен.

§ 5. АГРЕГИРОВАНИЕ РЯДОВ И МОДЕЛЕЙ

Со времени введения Дж. Э. Юлом (1921) авторегрессионных моделей и моделей скользящего среднего они вызывают интерес у статистиков, занимающихся анализом временных рядов. Это объясняется тем, что модели достаточно просты и могут успешно применяться на практике. Очевидным обобщением этих моделей является смешанная модель авторегрессии — скользящего среднего. На практике смешанную модель АРСС труднее оценивать, чем, например, АР-модель, но, с другой стороны, смешанная модель может быть статистически более эффективной, чем АР-или СС-модели. Согласно Боксу и Дженкинсу, если оцениваются одновременно две модели АР (p) и АРСС (p', q'), то, как правило, во второй модели потребуется меньше параметров для достижения удовлетворительных результатов, т. е. $p' + q' < p$. Так как имеются веские основания отдать предпочтение модели с наименьшим числом параметров, то смешанная модель часто предпочтительнее.

В работе К. Гренджера и М. Морриса [62] проведен теоретический анализ, на основе которого сделан вывод о возможности появления смешанных моделей в результате агрегирования рядов со сравнительно простой структурой. Основная теорема, доказанная в этой работе, формулируется следующим образом.

Пусть x_t и y_t — два независимых стационарных ряда с нулевым математическим ожиданием. Тогда если $x_t \sim \text{АРСС}(p, m)$, $y_t \sim \text{АРСС}(q, n)$ и $z_t = x_t + y_t$, то $z \sim \text{АРСС}(r, s)$, т. е. $\text{АРСС}(p, m) + \text{АРСС}(q, n) = \text{АРСС}(r, s)$, где $r \leq p + q$ и $s \leq \max(p + n, q + m)$.

В тех случаях, когда два независимых ряда описываются моделями АРИСС, теорема утверждает, что АРИСС (p_1, d_1, q_1) + АРИСС (p_2, d_2, q_2) = АРИСС (r, d', s),

где $r \leq p_1 + p_2$;

$d' = \max(d_1, d_2)$;

$$s \leq \begin{cases} \max(p_1 + q_2 + d_1 - d_2, p_2 + q_1), & \text{если } d_1 \geq d_2; \\ \max(p_1 + q_2, p_2 + q_1 + d_2 - d_1), & \text{если } d_2 \geq d_1. \end{cases}$$

Не составляет трудности обобщить основную теорему на случай суммы любого числа независимых рядов:

$$\sum_{j=1}^n \text{АРСС}(p_j, q_j) = \text{АРСС}(r, s),$$

где

$$r \leq \sum_{j=1}^n p_j \text{ и } s \leq \max(r - p_j + q_j), j = 1, \dots, n.$$

С точки зрения интерпретации два случая представляют особый интерес. Первый, когда ряды агрегируются и образуется один общий; большинство макроэкономических рядов, таких, как валовой национальный продукт, экспорт и т. д., являются агрегатными. Второй, когда наблюдаемый ряд является суммой истинного процесса и ошибки наблюдения: «сигнал плюс шум». Фактически любой макроэкономический ряд содержит существенные ошибки наблюдения.

Рассмотрим несколько частных случаев, полнее раскрывающих содержание основной теоремы.

1. АР (p) + белый шум = АРСС (p, p).

Это соответствует наблюдению авторегрессионного процесса в присутствии белого шума.

2. АР (p) + АР (q) = АРСС ($p + q, \max(p, q)$) и, в частности,

$$\text{АР}(1) + \text{АР}(1) = \text{АРСС}(2, 1).$$

Это может соответствовать ситуации, когда исследуемый ряд является агрегатом двух независимых авторегрессионных рядов, или случаю, когда процесс АР (p) искажается ошибкой наблюдения АР (q).

Отметим, что сумма k рядов, описываемых моделью АР(1), приводит к модели АРСС ($k, k - 1$).

3. СС (p) + СС (q) = СС ($\max(p, q)$)

и, в частности, СС (p) + белый шум = СС (p).

Таким образом, если истинный процесс генерируется моделью СС, то добавление белого шума ошибок наблюдения не изменит типа и порядка модели. Значения параметров при этом, конечно, изменятся.

4. АРСС (p, m) + белый шум = $\begin{cases} \text{АРСС}(p, p), & \text{если } p > m; \\ \text{АРСС}(p, m), & \text{если } p < m \end{cases}$

Это говорит о том, что добавление ошибки наблюдения может изменить порядок модели АРСС, хотя и не обязательно.

$$5. \text{AP}(p) + \text{CC}(q) = \text{АРСС}(p, p+q).$$

Этот случай также можно интерпретировать либо как агрегирование, либо как наложение ошибки наблюдения, не являющейся белым шумом.

Итак, в четырех из пяти рассмотренных случаев, кроме случая 3, после сложения независимых рядов были получены смешанные модели авторегрессии — скользящего среднего. Между тем требование независимости рядов может быть несколько ослаблено и получены те же выводы. Предположим, что каждый ряд, входящий в агрегат, состоит из двух компонент. Одна компонента, общая для всех рядов, представляет влияние общеэкономических факторов, а другая не зависит от первой и характерна только для данного ряда. Тогда для каждой компоненты могут быть приняты простые модели, а после агрегирования получена общая модель суммы рядов в соответствии с основной теоремой. Такие модели успешно использовались за рубежом при анализе цен на фондовой бирже и доходов фирм.

Наконец, рассмотрим еще один случай, когда x_t и y_t генерированы авторегрессионной схемой двух переменных с обратной связью $a(B)x_t + b(B)y_t = \varepsilon_t$, $c(B)x_t + d(B)y_t = u_t$, где ε_t , u_t — некоррелированные белые шумы;

$$b(0) = c(0) = 0.$$

Модель, описывающая только x_t , определяется элиминированием y_t и имеет вид: $[a(B)d(B) - c(B)b(B)]x_t = d(B)\varepsilon_t + b(B)u_t$.

Таким образом, вновь появляется модель АРСС (p, q) и легко показать, что вообще $p > q$.

Рассмотренные случаи говорят о том, что по-видимому, многие реальные данные описываются моделью АРСС. Фактически можно сделать вывод, что наиболее вероятной моделью на практике будет смешанная модель авторегрессии — скользящего среднего, что надо учитывать на стадии идентификации.

§ 6. ПРИМЕРЫ

Метод Бокса и Дженкинса был испытан на рядах курса акций фирмы ИБМ и цен на золото.

Ряд ИБМ также был разделен на 3 отрезка по 48 точек, а ряд цен на золото — на 6 отрезков по 60 точек. Для каж-

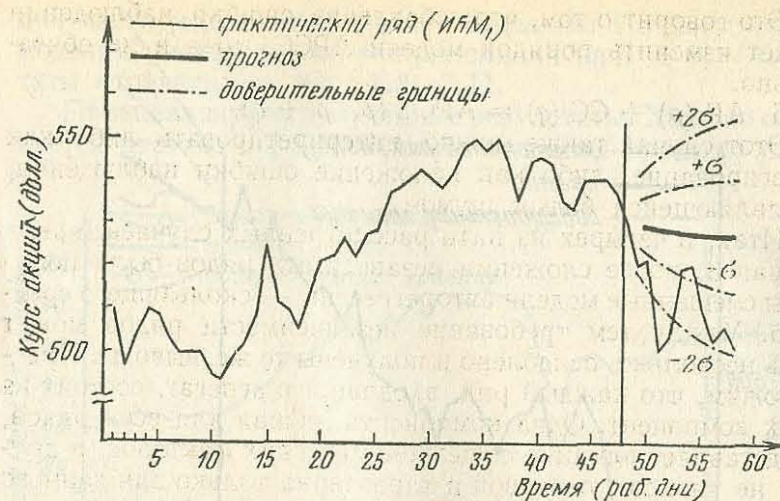


Рис. 7.3. Прогнозирование курса акций фирмы ИБМ по модели АРИСС (1, 0, 0):

$$x_t = 525 + 0,944(x_{t-1} - 525) + \varepsilon_t$$

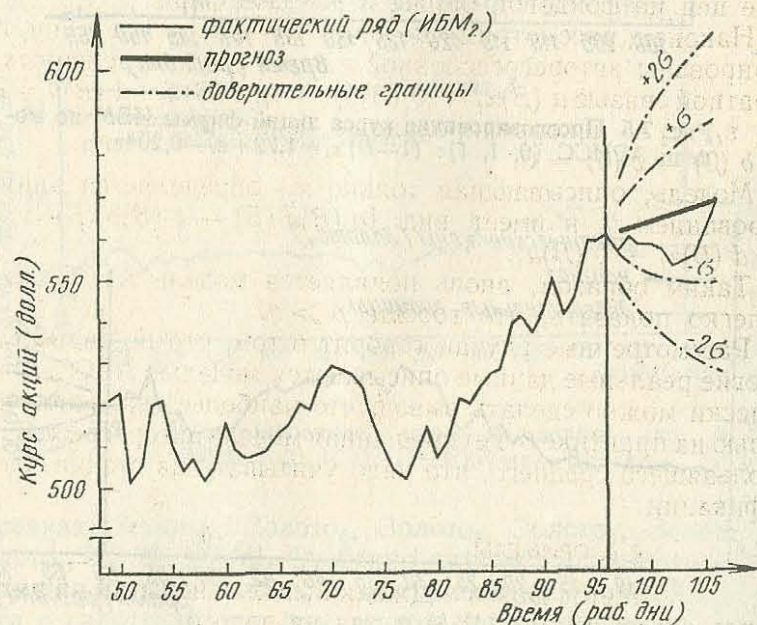


Рис. 7.4. Прогнозирование курса акций фирмы ИБМ по модели АРИСС (0, 1, 1): $(1-B)x_t = 0,830 + \varepsilon_t + 0,126\varepsilon_{t-1}$

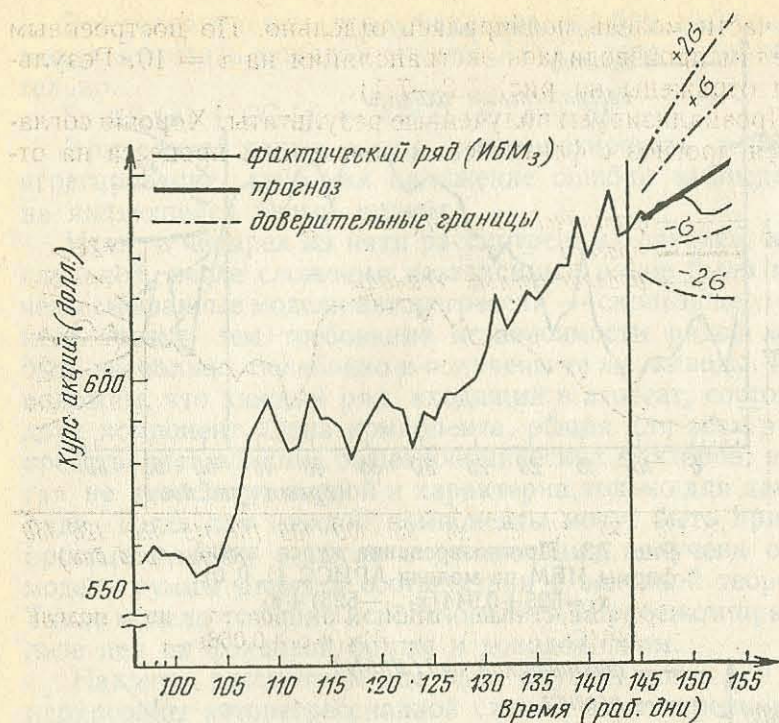


Рис. 7.5. Прогнозирование курса акций фирмы ИБМ по модели АРИСС (0, 1, 1): $(1-B)x_t = 1,72 + \varepsilon_t - 0,204\varepsilon_{t-1}$

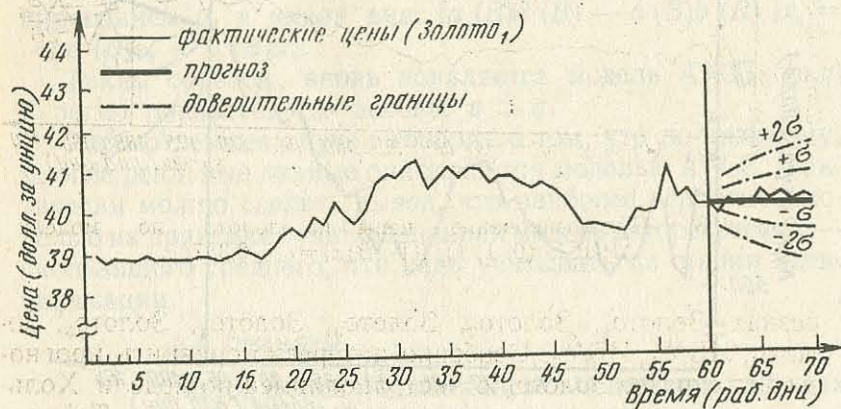


Рис. 7.6. Прогнозирование цены на золото по модели АРИСС (0, 1, 1): $(1-B)x_t = \varepsilon_t - 0,248\varepsilon_{t-1}$

дой части модель подбиралась отдельно. По построенным моделям производилась экстраполяция на $\tau = 10$. Результаты отражены на рис. 7.3—7.11.

Проанализируем полученные результаты. Хорошо согласуется прогноз с фактическим поведением процесса на от-



Рис. 7.7. Прогнозирование цены на золото по модели АРИСС (0, 1, 1): $(1-B)x_t = \varepsilon_t + 0,068\varepsilon_{t-1}$

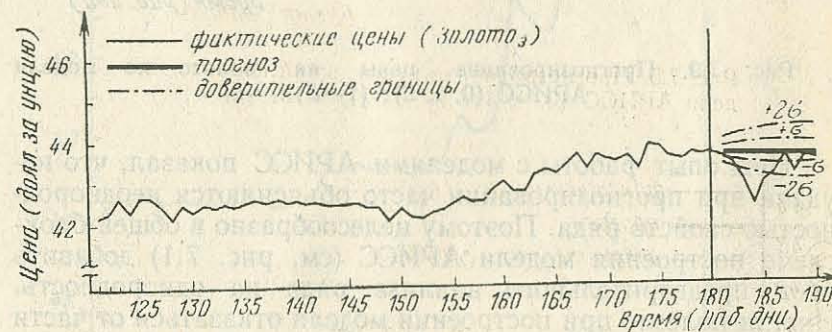


Рис. 7.8. Прогнозирование цены на золото по модели АРИСС (0, 1, 1): $(1-B)x_t = \varepsilon_t - 0,387\varepsilon_{t-1}$

резках: Золото₁, Золото₂, Золото₃, Золото₄, Золото₅, Золото₆, ИБМ₂, ИБМ₃. Особенно интересно сравнить прогнозы для отрезка Золото₆ с экстраполяцией по модели Хольта с гипотезой, что ряд генерируется процессом Тейла—Вейджа (см. рис. 1.10). Модель АРИСС дала здесь намного лучший результат.

На участке ИБМ₁ реальный процесс уходит из доверительной области. Это можно объяснить резким падением курса акций (на 40 дол.) от точки 46 до точки 51. Таких перепадов на участке 1—48 не наблюдалось. В данном случае оказалась неверна гипотеза о возможности экстраполяции динамических свойств ряда.



Рис. 7.9. Прогнозирование цены на золото по модели АРИСС (0, 1, 0): $(1-B)x_t = \varepsilon_t$

Наш опыт работы с моделями АРИСС показал, что неудачи при прогнозировании часто объясняются неоднородностью свойств ряда. Поэтому целесообразно в общей блок-схеме построения модели АРИСС (см. рис. 7.1) добавить блок предварительного анализа ряда на однородность. Иногда полезно при построении модели отказаться от части имеющейся устаревшей информации с тем, чтобы данные характеризовались большей однородностью.

В табл. 7.4 приведены результаты эмпирического сравнения моделей Бокса—Дженкинса и адаптивной модели авторегрессии, рассмотренной в гл. 3. Адаптивная модель авторегрессии хорошо себя зарекомендовала при эмпирическом сравнении с другими моделями (см. гл. 6, § 4). На основе данных табл. 7.4 можно сделать вывод о том, что модели Бокса—Дженкинса, по-видимому, приводят к более точным прогнозам, чем адаптивная модель авторегрессии.

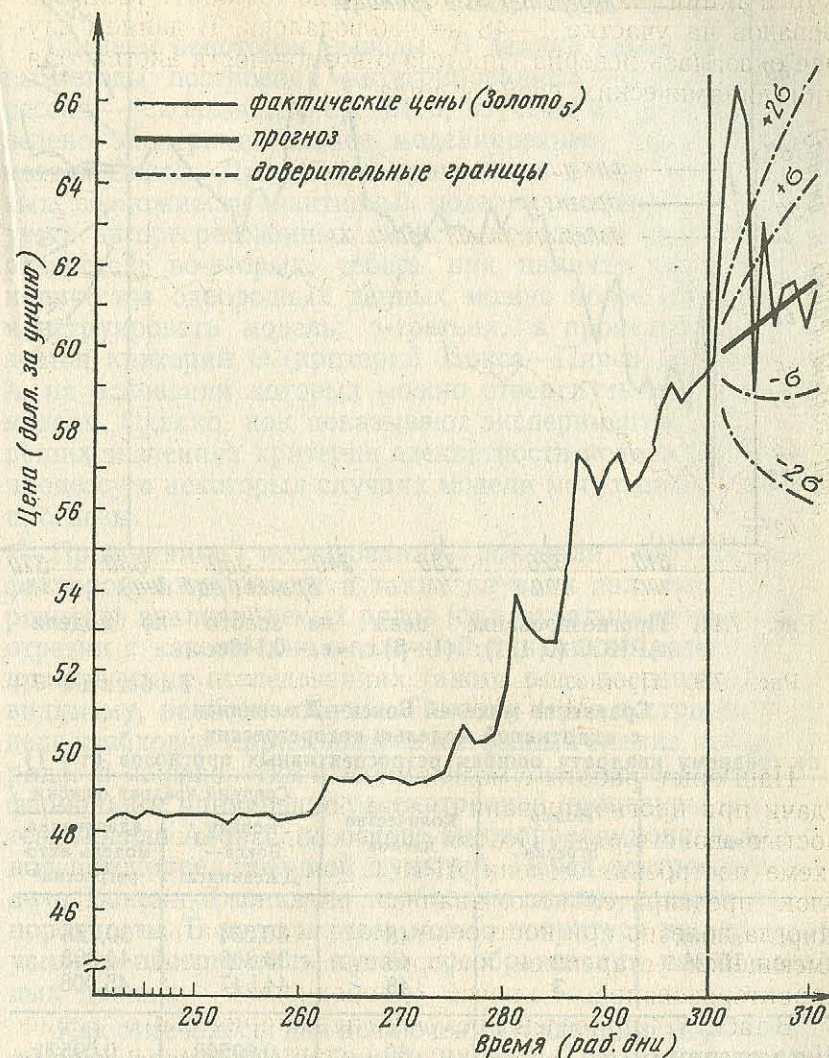


Рис. 7.10. Прогнозирование цены на золото по модели АРИСС (0, 1, 1): $(1-B)x_t = 0,196 + \varepsilon_t + 0,416\varepsilon_{t-1}$

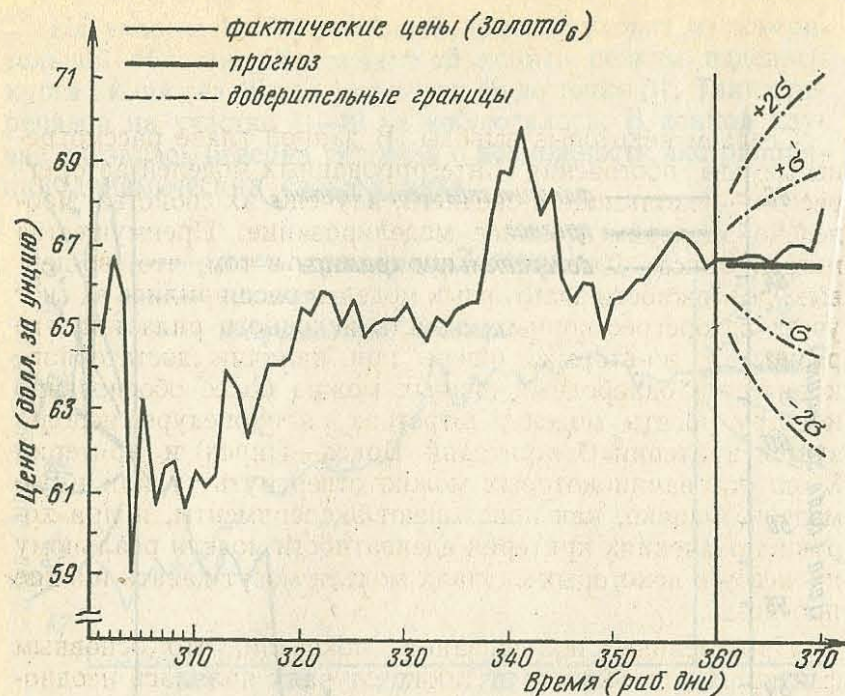


Рис. 7.11. Прогнозирование цены на золото по модели АРИСС (0, 1, 1): $(1-\beta)x_t = \varepsilon_t - 0,146\varepsilon_{t-1}$

Таблица 7.4

Сравнение моделей Бокса—Дженкинса с адаптивной моделью авторегрессии по среднему квадрату ошибки ретроспективных прогнозов ($\tau=1$)

Ряд	Номер отрезка	Количество членов	Средний квадрат ошибки	
			модель Бокса—Дженкинса	адаптивная модель авторегрессии*
Курс акций фирмы ИБМ	1	48	30,254	30,728
	2	48	33,808	44,363
	3	48	44,547	45,008
Цена на золото	1	60	0,09208	0,09571
	2	60	0,10623	0,10947
	3	60	0,02797	0,02910
	4	60	0,13410	0,13851
	5	60	0,40441	0,83498
	6	60	0,68115	0,81166

* В связи с тем, что при подсчете среднего квадрата ошибки в использованных нами программах метода Бокса—Дженкинса первые пять значений ошибки не учитываются, как связанные с начальными условиями, данные скорректированы.

Сделаем некоторые выводы. В данной главе рассмотрены методы построения интегрированных моделей авторегрессии — скользящего среднего, изучены их свойства, проведено экспериментальное моделирование. Преимущества метода Бокса—Дженкинса проявились в том, что, во-первых, возможности адаптивных моделей расширились за счет учета авторегрессионных свойств исходного ряда или его разностей; во-вторых, теперь при наличии достаточного количества однородных данных можно более обоснованно конструировать модель; в-третьих, в процедуре используется критерий Θ (критерий Бокса—Пирса) и критерий λ , на основании которых можно отвергнуть ту или иную модель. Однако, как показывают эксперименты, и при хороших значениях критерия адекватности модели реальному процессу в некоторых случаях модели могут давать плохие прогнозы.

Проведенные исследования показали, что основным фактором погрешности в таких случаях являлась неоднородность анализируемых рядов (они охватывали временные отрезки с качественно различной динамикой развития). В практических исследованиях такого рода погрешности, по-видимому, наиболее опасны. Поэтому перед построением модели необходим тщательный качественный анализ исходного ряда. В связи с этим следует считать неправомерным чисто формальное применение рассмотренной процедуры для синтеза модели АРИСС на основе автокорреляционной и частной автокорреляционной функций. Перед построением модели Бокса—Дженкинса необходим анализ данных на однородность. В ряде случаев может появиться возможность усилить однородность путем преобразования части исходных данных.

Как отмечалось, для исключения тенденции, выражаемой полиномом порядка d , Бокс и Дженкинс предложили переходить от исходного ряда к его разностям $w_t = \nabla^d x_t$ и строить модель АРСС для ряда $\tilde{w}_t = w_t - \mu_w$ (см. (7.10)), где μ_w — среднее значение w_t , т. е.

$$\mu_w = \frac{1}{N-d} \sum w_t. \quad (7.57)$$

Представим (7.57) для случая $d = 1$ в виде:

$$\mu_w = \frac{1}{n} [(x_n - x_{n-1}) + (x_{n-1} - x_{n-2}) + \dots + (x_2 - x_1)] = \frac{x_n - x_1}{n}, \quad (7.58)$$

где $n = N - 1$.

При $d = 1$ и $\mu_w \neq 0$ речь идет об исключении тенденции, выражаемой полиномом первого порядка, причем μ_w есть коэффициент линейного роста. Способ его оценки (7.58), очевидно, не является наилучшим, так как он опирается лишь на два крайних наблюдения выборки, а остальные данные в расчет не принимаются.

Точно так же можно показать, что при $d = 2$

$$\mu_w = \frac{x_n - x_{n-1} + x_2 - x_1}{n}, \quad n = N - 2 \text{ и т. д.}$$

Это свидетельствует о том, что предлагаемый метод исключения полиномиальных тенденций может быть пригоден лишь при краткосрочном прогнозировании. При увеличении же периода упреждения тенденцию лучше исключать каким-либо другим способом, а модель АРСС строить для отклонений от тенденции.

Сравнивая процессы Бокса—Дженкинса и Тейла—Вейджа, мы пришли к выводу, что это различные процессы. Между тем линейный предиктор Хольта, использованный Тейлом и Вейджем, является частным случаем предикторов Бокса—Дженкинса. Таким образом, один и тот же предиктор может быть оптимальным для прогнозирования различных процессов, при этом доверительные границы прогноза должны рассчитываться, конечно, по различным формулам.

Мы видели, что модели Брауна являются частным случаем моделей Бокса и Дженкинса. Подчеркнем теперь основное различие между этими методами. Модели Брауна полностью определяются единственным параметром β . Браун пытался прогнозировать все многообразие рядов с помощью всего лишь трех различных значений β . Таким образом, его целью было создание простых универсальных моделей, дающих минимум среднеквадратической ошибки для совокупности многих рядов в целом.

У Бокса и Дженкинса к каждому ряду подход индивидуален. Они решают проблему построения специальной, оптимальной модели отдельно в каждом случае. По-видимому, оба подхода для практики сохраняют свое значение, но различаются по целям и области применения.

Глава 8

МОДЕЛИРОВАНИЕ ВЗАИМОСВЯЗАННЫХ ВРЕМЕННЫХ РЯДОВ

Глава посвящена вопросам многомерного статистического анализа. Рассматривается возможность проверки адекватности линейных динамических эконометрических моделей реальному процессу на основе изучения и сопоставления стохастических свойств переменных, входящих в состав уравнений. Предлагается адаптивная модель множественной регрессии. На ее основе построена адаптивная производственная функция на данных, характеризующих экономику США (1947—1973 гг.).

§ 1. АНАЛИЗ ЛИНЕЙНЫХ ДИНАМИЧЕСКИХ ЭКОНОМЕТРИЧЕСКИХ МОДЕЛЕЙ

В предыдущих главах рассматривалась техника анализа изолированного временного ряда. В центре нашего внимания были методы построения прогностических моделей на основе анализа данных единственной выборки. Предполагалось, что будущая тенденция является той или иной функцией времени или предшествующих значений ряда. Вследствие того, что на коэффициенты этой функции не накладывалось требование быть неизменными во времени, а модель, по которой рассчитывались прогнозы, наделялась адаптивными свойствами, методы анализа и прогнозирования изолированного ряда были достаточно гибкими и полезными для моделирования широкого класса одномерных процессов или их отдельных сторон. В случае ограниченности исходной информации, неразработанности теории исследуемого процесса, неопределенности представления о взаимосвязях данного ряда с другими рядами эти методы могут оказаться незаменимыми и упрощение, в резуль-

тате которого совокупное воздействие всех сторонних факторов выражается в модели через время, становится необходимым.

Если же известно о воздействии на изучаемый процесс каких-то других процессов и имеется возможность получить временные ряды, описывающие их развитие, то методы анализа изолированных рядов уступают место многомерному статистическому анализу. Это позволяет включить в модель ценную дополнительную информацию, учесть структуру изучаемого объекта и получить взаимоувязанные прогнозы нескольких переменных.

В случае многомерного статистического анализа обычно строят системы линейных уравнений, модели множественной регрессии. Вопросам их построения посвящено достаточно много работ (см. например, [14]). В процедуре построения таких моделей можно выделить пять основных этапов: отбор временных рядов (факторов) для включения в модель, разделение всех отобранных факторов на экзогенные и эндогенные, принятие гипотезы о характере связи эндогенных и экзогенных переменных (т. е. структуры модели), оценивание параметров модели по заданному критерию, анализ адекватности модели. На все этапы, кроме, пожалуй, оценивания, огромное влияние оказывает субъективное мнение исследователя, его опыт, знание реальных процессов и теорий. В силу этого структура модели часто задается произвольно, интуитивно. Этим и объясняется потребность в методах анализа гипотез о структуре изучаемого объекта. Интересным в этом плане является предложение А. Зельнера и Ф. Пальма [115] сначала теоретически вывести свойства каждого включенного в модель фактора исходя из гипотетической структуры модели, а затем провести эмпирический анализ статистических данных. Если результаты эмпирического и теоретического анализа не противоречат друг другу, то предполагается, что структура модели выбрана верно. Если же обнаруживается противоречие, модель перестраивается в соответствии с новой гипотезой и вновь осуществляется проверка.

Иными словами, структура эконометрической системы уравнений проверяется на основе раздельного анализа каждого входящего в нее временного ряда методом, рассмотренным в гл. 7, и сопоставления статистических свойств эндогенных и экзогенных переменных. Таким образом, традиционный эконометрический подход соединяется с послед-

ними достижениями техники исследования временных рядов. Остановимся на этой идее более подробно.

Предположим, что исследуется линейный многомерный процесс. Наблюдаемые временные ряды многомерного процесса можно представить в виде вектора переменных $z_t = (z_{1,t}, z_{2,t}, \dots, z_{p,t})$. В предположении, что этот процесс адекватно аппроксимируется многомерным процессом АРСС, можно вслед за М. Кенуем [127] записать:

$$H(B)z_t = \Theta + F(B)\varepsilon_t, \quad t = 1, 2, \dots, T, \quad (8.1)$$

где $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \dots, \varepsilon_{p,t})$ вектор случайных ошибок; $\Theta = (\theta_1, \theta_2, \dots, \theta_p)$ — вектор констант;

$H(B)$ и $F(B)$ — матрицы операторов сдвига назад размерности $(p \times p)$, элементами которых являются полиномы от B конечного порядка, т. е.

$$H(B) = \{h_{ij}(B)\} \text{ и } F(B) = \{f_{ij}(B)\}, \text{ где}$$

$$h_{ij}(B) = \sum_{l=0}^{r_{ij}} h_{ij}^{(l)} B^l; \quad f_{ij}(B) = \sum_{l=0}^{q_{ij}} f_{ij}^{(l)} B^l;$$

r_{ij} и q_{ij} обозначают порядок $h_{ij}(B)$ и $f_{ij}(B)$ соответственно. В отношении вектора ошибок принимается, что

$$M(\varepsilon_t) = \bar{0};$$

$$M(\varepsilon_t \varepsilon_{t'}') = \delta_{tt'} I_p \text{ для всех } t, t',$$

где $\delta_{tt'}$ — дельта-функция Кронекера¹; I_p — единичная матрица $(p \times p)$.

Взаимные корреляции одновременных значений, автокорреляции, а также различные дисперсии элементов вектора ошибок могут быть введены посредством соответствующей спецификации матрицы $F(B)$. Если элементы матрицы $F(B)$ имеют нулевой порядок от B , то (8.1) чистый АР-процесс. Если элементы матрицы $H(B)$ имеют нулевой порядок от B , то (8.1) чистый СС-процесс. Если и $H(B)$ и $F(B)$ содержат полиномы, порядки которых отличны от нуля, то (8.1) представляет смешанный процесс АРСС.

Если $H(B)$ обратима, т. е. имеет собственные значения вне единичного круга, то можно обе части (8.1) умножить

¹ См. стр. 29.

на матрицу $H^*(B)$, присоединенную¹ к матрице $H(B)$:

$$|H(B)|z_t = \Theta'' + H^*(B)F(B)\varepsilon_t, \quad (8.2a)$$

где $|H(B)|$ — детерминант матрицы $H(B)$, полином от B конечного порядка;

$\Theta'' = (\theta_1'', \theta_2'', \dots, \theta_p'')$ — вектор новых констант.

Таким образом, система (8.2a) состоит из уравнений вида:

$$|H(B)|z_{it} = \theta_i'' + c_i' \varepsilon_t, \quad i = 1, 2, \dots, p, \quad (8.2b)$$

где c_i' обозначает i -ю строку $H^*(B)F(B)$. Такие уравнения назовем конечными уравнениями (КУ).

Из (8.2b) видно, что отдельные элементы вектора z_t описываются АРСС-процессами. Если обе части выражения не содержат общих сокращающихся множителей, то порядок и параметры АР-частей всех уравнений в (8.2b) должны быть идентичными. Это может быть проверено анализом статистических выборок. Если эмпирический анализ каждого временного ряда в отдельности подтверждает, что порядок и параметры АР-частей КУ идентичны, то вектор может быть генерирован процессом (8.1) без наложения каких-либо дополнительных ограничений на $H(B)$. Если, однако, порядок и параметры АР-частей КУ оказались различными для различных элементов z_t , то возможно, что $H(B)$ имеет какой-либо особый вид, например, может быть треугольной или блочной диагональной матрицей.

На основе эмпирических результатов изучения КУ можно сделать выводы о возможных свойствах матрицы $H(B)$ и наложить на нее соответствующие ограничения. В частности, в больших системах, если бы на $H(B)$ не накладывались никакие ограничения, процессы (8.2a) имели бы высокий порядок АР-частей, что не согласуется с эмпирическими исследованиями, которые указывают на относительно низкий порядок АР-частей КУ для многих эко-

¹ Напомним, что присоединенной матрицей к квадратной матрице A называется матрица A^* , составленная из алгебраических дополнений к элементам матрицы A .

Известно, что произведение

$$AA^* = A^*A = \begin{pmatrix} d & 0 & \dots & 0 \\ 0 & d & \dots & 0 \\ 0 & 0 & \dots & d \end{pmatrix},$$

где $d = |A|$ — определитель матрицы A .

номических переменных. Важно отметить, что независимо от того, возможно или нет вывести свойства $H(B)$ из эмпирического анализа, определенные по статистическим данным КУ позволяют генерировать прогнозы отдельно по каждой переменной.

В эконометрических работах обычно предполагается, что не все переменные равноправны: некоторые из переменных в z_t заданы экзогенно, т. е. генерированы процессами, независимыми от процессов, генерирующих остальные переменные. Это означает, что матрицам $H(B)$ и $F(B)$ предъявляются определенные требования и они приобретают специфический вид.

Разобьем z_t следующим образом: $z_t = (y_t, x_t)$, где y_t — вектор эндогенных переменных размерностью $(p_1 \times 1)$;

x_t — вектор экзогенных переменных размерностью $(p_2 \times 1)$, $p = p_1 + p_2$.

Тогда систему (8.1) можно записать как

$$\begin{bmatrix} H_{11}(B) & H_{12}(B) \\ H_{21}(B) & H_{22}(B) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \Theta_1 \\ \Theta_2 \end{bmatrix} + \begin{bmatrix} F_{11}(B) & F_{12}(B) \\ F_{21}(B) & F_{22}(B) \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}, \quad (8.3)$$

где $H(B)$, $F(B)$, Θ , ε_t расчленены в соответствии с разбиением z_t .

Предположение, что x_t — вектор экзогенных переменных — приводит к следующим ограничениям:

$$H_{21}(B) \equiv 0; F_{21}(B) \equiv 0; F_{12}(B) \equiv 0. \quad (8.4)$$

Это означает, что на поведение x_t элементы вектора ε_{1t} не оказывают влияния, а элементы вектора ε_{2t} воздействуют на y_t только через элементы x_t .

При этих ограничениях, наложенных на (8.3), получаем (ниже матриц и векторов указана их размерность):

$$\begin{matrix} H_{11}(B) & y_t & + & H_{12}(B) & x_t & = & \Theta_1 & + & F_{11}(B) & \varepsilon_{1t} \\ (p_1 \times p_1) & (p_1 \times 1) & & (p_1 \times p_2) & (p_2 \times p_1) & (p_1 \times 1) & & (p_1 \times p_1) & (p_1 \times 1) \end{matrix} \quad (8.5a)$$

$$\text{и} \quad \begin{matrix} H_{22}(B) & x_t & = & \Theta_2 & + & F_{22}(B) & \varepsilon_{2t} \\ (p_2 \times p_2) & (p_2 \times 1) & & (p_2 \times 1) & & (p_2 \times p_2) & (p_2 \times 1) \end{matrix} \quad (8.5b)$$

Уравнения (8.5a) являются структурными, в то время как уравнения (8.5b) описывают процессы, гене-

рирующие вектор стохастических экзогенных переменных x_t .

Предположение о том, что x_t — вектор экзогенных переменных, приводя к ограничениям (8.4), сильно сказывается на форме КУ. При условии обратимости $H_{22}(B)$ КУ для элементов x_t имеют вид:

$$|H_{22}(B)| x_t = \Theta_2'' + H_{22}^*(B) F_{22}(B) \varepsilon_{2t}, \quad (8.6)$$

где $|H_{22}(B)|$ и $H_{22}^*(B)$ — детерминант и присоединенная матрица соответственно;

Θ_2'' — вектор новых констант.

Подставляя x_t из (8.6) в (8.5а) и умножая обе части на матрицу, присоединенную к $H_{11}(B)$, т. е. на $H_{11}^*(B)$, получаем КУ для элементов y_t :

$$|H_{11}(B)| |H_{22}(B)| y_t = \Theta_1'' - H_{11}^*(B) H_{12}(B) H_{22}^*(B) F_{22}(B) \varepsilon_{2t} + |H_{22}(B)| H_{11}^*(B) F_{11}(B) \varepsilon_{1t}, \quad (8.7)$$

где Θ_1'' — вектор новых констант.

При сравнении (8.6) и (8.7) видно, что КУ для элементов вектора эндогенных переменных y_t имеют АР-операторы, порядок которых равен или выше, чем порядок у АР-операторов КУ для элементов x_t . Если $|H_{11}(B)|$ полином нулевого порядка от B , то порядки АР-частей (8.6) и (8.7) совпадают. Обычно же порядок АР-оператора (8.7) будет превосходить порядок АР-оператора (8.6). Это может быть проверено эмпирически.

Далее, порядок СС-оператора в (8.7) будет выше, чем в (8.6). Это также можно проверить эмпирически. Наконец, если, как это обычно и бывает, относительно элементов $H(B)$ и $F(B)$ сделаны дополнительные предположения, то их влияние на вид КУ может быть определено теоретически и проверено эмпирически.

§ 2. АДАПТИВНАЯ МОДЕЛЬ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

Классический регрессионный анализ опирается на гипотезу о возможности аппроксимации изучаемого процесса линейным уравнением с постоянными коэффициентами. Эти коэффициенты отражают степень связи различных переменных с изучаемой величиной. В реальной жизни сила взаимодействия переменных не остается неизменной, как не остается неизменной и внешняя среда, в которой раз-

вивается исследуемый процесс. Оценки же коэффициентов, полученные на основании упомянутой гипотезы, отражают характер связи переменных лишь в среднем по выборке, поэтому трудно ожидать, чтобы они привели к хорошим краткосрочным прогнозам.

Таким образом, множественная регрессия с постоянными коэффициентами имеет ограниченное применение и весьма желательно было бы найти способ корректировки, обновления ее коэффициентов. Это открыло бы возможность исследовать направление и характер эволюции взаимосвязей переменных и получать прогнозы по модели, лучше отражающей текущее состояние процесса. Здесь мы и предлагаем способ адаптации коэффициентов множественной регрессии.

Предположим, что исследуется связь ряда y с рядами $x_1, \dots, x_M$ и что оценку значения $y_{t+\tau}$ можно получить как взвешенную сумму вида:

$$\hat{y}_\tau(t) = \omega_{1,t} x_{1,t} + \dots + \omega_{M,t} x_{M,t} = \sum_{i=1}^M \omega_{i,t} x_{i,t} \quad \tau \geq 0.$$

Это уравнение множественной регрессии.

В случае, когда $\tau = 0$, будем решать задачу чистого анализа эволюции коэффициентов связи $\omega_{i,t}$. При $\tau > 0$ — задачу анализа эволюции коэффициентов множественной регрессии и прогнозирования на τ шагов вперед на основе текущей информации.

Сравнивая оценку $\hat{y}_\tau(t)$ с фактической точкой ряда $y_{t+\tau}$, можем вычислить ошибку:

$$e_{t+\tau} = y_{t+\tau} - \hat{y}_\tau(t) = y_{t+\tau} - \sum_{i=1}^M \omega_{i,t} x_{i,t},$$

и на основе полученного результата произвести корректировку коэффициентов $\omega_{i,t}$.

Для адаптации коэффициентов $\omega_{i,t}$, как и в гл. 3, воспользуемся методом наискорейшего спуска, т. е. обновление весов будем осуществлять по следующему правилу:

$$W_n = W_c - k \text{ grad}(e_{t+\tau}^2),$$

где W_c — вектор старых коэффициентов;

W_n — вектор новых коэффициентов;

k — коэффициент ($k > 0$);

$\text{grad}(e_{t+\tau}^2)$ — вектор, градиент $e_{t+\tau}^2$.

Используя выражение для $e_{t+\tau}$, находим элементы градиента:

$$\frac{\partial e_{t+\tau}^2}{\partial \omega_{i,t}} = 2e_{t+\tau} \frac{\partial e_{t+\tau}}{\partial \omega_{i,t}} = -2e_{t+\tau} x_{i,t}.$$

В целом градиент будет равен:

$$\text{grad}(e_{t+\tau}^2) = -2e_{t+\tau} X_t,$$

где X_t — вектор $[x_{1,t}, \dots, x_{M,t}]'$.

Таким образом, корректировка коэффициентов должна осуществляться по правилу: $W_n = W_0 + 2ke_{t+\tau}X_t$.

Неизвестным в этом выражении остается лишь значение коэффициента k , определяющего скорость движения в направлении обратном градиенту. Для того чтобы выявить его роль в процессе адаптации, поступим следующим образом. Вернемся в точку t и вновь сделаем прогноз, но уже с новыми весами W_n . Получим новое значение ошибки:

$$\begin{aligned} (e_n)_{t+\tau} &= y_{t+\tau} - \sum_{i=1}^M [(\omega_c)_{i,t} + 2k(e_c)_{t+\tau} x_{i,t}] x_{i,t} = \\ &= \left[y_{t+\tau} - \sum_{i=1}^M (\omega_c)_{i,t} x_{i,t} \right] - \sum_{i=1}^M 2k(e_c)_{t+\tau} x_{i,t}^2 = \\ &= (e_c)_{t+\tau} - 2k(e_c)_{t+\tau} \sum_{i=1}^M x_{i,t}^2 = (e_c)_{t+\tau} \left(1 - 2k \sum_{i=1}^M x_{i,t}^2 \right), \end{aligned}$$

где $(e_c)_{t+\tau}$ — ошибка, полученная при старых весах $(\omega_c)_{i,t}$.

Теперь если положить $k = \frac{\alpha}{2 \sum_{i=1}^M x_{i,t}^2}$, то

$$(e_n)_{t+\tau} = (e_c)_{t+\tau} (1 - \alpha)$$

и при $0 < \alpha < 2$ $|(e_n)_{t+\tau}| < |(e_c)_{t+\tau}|$.

Следовательно, α определяет реакцию модели на полученную текущую ошибку и корректирует коэффициенты множественной регрессии так, чтобы уменьшить ошибку на $(1 - |1 - \alpha|) \cdot 100\%$. Будем называть α параметром адаптации и считать его постоянным для данной модели. Коэффициент k будет при этом меняться во времени. Оптимальное значение α можно определить методом проб, т. е. в процессе «обучения» модели.

Очевидно, что, для того чтобы включить в уравнение множественной регрессии свободный член, достаточно один экзогенный ряд, скажем x_1 , задать в виде единиц. Лаговые, запаздывающие переменные можно ввести в число экзогенных, полагая, например,

$$x_{i,t} = x_{j,t-l}, \quad t = 1, 2, \dots, T.$$

Следует сказать несколько слов о проблеме мультиколлинеарности. Мультиколлинеарность, т. е. корреляция между независимыми переменными уравнения, имеет место тогда, когда существуют линейные соотношения между экзогенными переменными. Сильная мультиколлинеарность часто возникает при введении в уравнение лаговых переменных. Проблема мультиколлинеарности, неразрывно связанная с проблемой оценивания параметров уравнения, рассмотрена практически во всех серьезных работах, посвященных регрессионному анализу. Остается отметить, что в нашем случае мультиколлинеарность проявляется в ухудшении процесса адаптации. Это приводит к тому, что оценки параметров могут значительно искажать представление о реальной структуре объекта в текущий момент времени.

Проведем эксперимент. Возьмем ряд натуральных чисел от 1 до 20. Легко убедиться в том, что этот ряд точно воспроизводится авторегрессионной схемой второго порядка:

$$x_t = x_{t-1} + (x_{t-1} - x_{t-2}) = 2x_{t-1} - x_{t-2},$$

т. е. $\omega_1 = 2$ и $\omega_2 = -1$. Значения этих коэффициентов оценим с помощью адаптивной модели множественной регрессии. В качестве начальных оценок весов произвольно положим $\omega_{1,0} = 0,5$ и $\omega_{2,0} = 0,5$. Процесс адаптации проведем многократным прогоном модели от значения $x = 3$ до $x = 20$, считая в каждом цикле (по 18 итераций в каждом) начальными те значения параметров, которые были получены в конце предыдущего цикла. Результаты, приведенные в табл. 8.1, свидетельствуют о медленной сходимости оценок к истинным значениям параметров при наличии сильной корреляции экзогенных переменных.

Перед построением модели адаптивной множественной регрессии рекомендуем строить обычную множественную регрессию методом наименьших квадратов. Это помогает на начальном этапе моделирования определить структуру уравнения множественной регрессии, отобрать переменные

Таблица 8.1

Адаптация коэффициентов множественной регрессии
при сильной мультиколлинеарности экзогенных переменных

α	Количество циклов	Общее количество итераций	Оценки параметров	
			ω_1	ω_2
1,8	0	0	0,5	0,5
	5	90	0,99674	0,04146
	50	900	1,96574	-0,96443
	100	1 800	1,99914	-0,99911
	150	2 400	1,99992	-0,99992
1,9	0	0	0,5	0,5
	5	90	1,22087	-0,29308
	50	900	1,99271	-0,99339
	100	1 800	1,99991	-0,99992
	150	2 400	1,99995	-0,99995

для их включения в уравнение. Полученные на этой стадии результаты можно рассматривать как исходные для адаптивного моделирования, их полезно также использовать для сопоставлений.

§ 3. АДАПТИВНАЯ МОДЕЛЬ ПРОИЗВОДСТВЕННОЙ ФУНКЦИИ

Построение и оценка производственных функций, отражающих связь между факторами и результатами производства, являются классической проблемой эконометрии (см. [6], [27], [31]).

Производственные функции имеют достаточно общую форму для экономического анализа. Они применяются в исследованиях различных уровней экономики в зависимости от характера исходных данных. В одних случаях производственные функции описывают отдельные технологические процессы, в других — отражают деятельность предприятия, отрасли или экономики страны в целом. Производственная функция отражает устойчивую количественную связь, существующую между затратами и выпуском продукции, но безотносительно к содержанию происходящих при этом реальных производственных процессов. Сочетание производственных факторов при этом условно называется абстрактной технологией.

Простейшая производственная функция, функция Кобба—Дугласа имеет вид: $Y = AK^\alpha L^\beta$; $\alpha > 0$; $\beta > 0$, где Y — выпуск продукции, а L и K — используемые объемы живого и овеществленного труда соответственно.

Смысл параметров α и β , статистическая оценка которых производится на основе рядов Y , K и L , легко выявляется с помощью логарифмического дифференцирования предыдущего уравнения:

$$\frac{dY}{Y} = \alpha \frac{dK}{K} + \beta \frac{dL}{L}.$$

Это выражение означает, что рост в исследуемый период объема K на 1% определяет рост выпуска продукции на $\alpha\%$. Параметр β интерпретируется аналогично для затрат живого труда. Таким образом, α и β являются эластичностями выпуска по факторам производства.

В экономической системе, описываемой производственной функцией Кобба—Дугласа, при $\alpha + \beta > 1$ имеет место экономия на масштабах производства, а при $\alpha + \beta < 1$ — рост удельных издержек. Соотношение $\alpha + \beta = 1$ означает, что совокупная эластичность факторов равна единице.

Производственная функция предполагает, что существует возможность для взаимного замещения факторов, т. е. что производственные факторы могут сочетаться в различных пропорциях. Однако при любой данной комбинации факторов можно достичь различных объемов выпуска продукции в зависимости от эффективности технологии и организации производственного процесса. Эффективность — характеристика абстрактной технологии, относящаяся только к зависимости между затратами и выпуском. Она не затрагивает отношений между затратами. Можно сказать, что эффективность абстрактной технологии выступает как масштаб преобразования затрат в выпуск. Эффективность в производственной функции отражает параметр A . Если при прочих равных условиях параметр A в производственной функции для одной экономической системы выше, чем для другой, то первая может считаться более эффективной, так как равные затраты ресурсов обеспечивают в ней больший выпуск продукции.

В связи с тем, что эффективность экономической системы, как правило, не остается постоянной на отрезке времени, для которого строится производственная функция, логично отражать эффективность в производственной функ-

ции переменным множителем $A(t)$. В качестве простейшего способа учета возрастания экономической эффективности Я. Тинберген предложил функцию $A(t) = Ae^{\kappa t}$. При этом на параметры производственной функции обычно накладывают ограничение $\alpha + \beta = 1$.

Таким образом, получаем динамическую производственную функцию:

$$Y = AK^\alpha L^\beta e^{\kappa t}, \quad \beta = 1 - \alpha, \quad (8.8)$$

где оцениваемыми параметрами являются A , α , κ , причем A теперь характеризует эффективность технологии в момент $t = 0$.

Если рост эффективности происходит при неизменном отношении α/β , то говорят, что наблюдается нейтральный технический прогресс. Если же это соотношение меняется во времени ($\alpha = \alpha(t)$ и $\beta = \beta(t)$), то в случае увеличения α/β прогресс называется капиталорасходующим (трудосберегающим), а при уменьшении α/β — капиталосберегающим (трудорасходующим).

Перейдем к проблеме оценивания. Уравнение производственной функции относится к классу нелинейных. Между тем наиболее распространенным методом оценки параметров экономико-математических моделей является метод наименьших квадратов (МНК), разработанный преимущественно для оценки линейных моделей. В тех случаях, когда использование линейной модели не отвечает поставленной задаче, нелинейные модели обычно сводят к линейным, а затем эти производные модели оценивают с помощью МНК.

Статистическая связь между тремя переменными в производственной функции выражается в форме:

$$Y_t = AK_t^\alpha L_t^{1-\alpha} e^{\kappa t} + \varepsilon_t, \quad \alpha > 0, \quad (8.9)$$

где ε_t — случайная переменная.

На практике же оценивание параметров A , α и κ часто осуществляют по регрессионному уравнению, получаемому из (8.8) путем почленного деления на L_t и логарифмирования обеих частей:

$$\ln \left(\frac{Y_t}{L_t} \right) = \ln A + \alpha \ln \left(\frac{K_t}{L_t} \right) + \kappa t + u_t, \quad (8.10)$$

где u_t — случайная переменная.

Однако применение МНК непосредственно к этому уравнению приводит к тому, что оценки параметров оказыва-

ются смещенными, так как минимум $\sum u_t^2$ в (8.10) не соответствует минимуму $\sum \varepsilon_t^2$ в (8.9).

В [12] (см. также приложение 3) делается вывод, что для компенсации смещения в случае производственной функции рассматриваемого нами вида нужно обе части уравнения (8.10) предварительно умножить на переменную, стоящую в левой части уравнения (8.9). Таким образом, приходим к задаче оценивания уравнения:

$$Y_t \ln \left(\frac{Y_t}{L_t} \right) = Y_t \ln A + \alpha Y_t \ln \left(\frac{K_t}{L_t} \right) + Y_t \kappa t + v_t,$$

где v_t — случайная переменная; $v_t = Y_t u_t \approx \varepsilon_t$.

Если теперь для оценивания этого уравнения вместо МНК применить метод адаптивной множественной регрессии, то получим адаптивную модель производственной функции.

Пример

Построим адаптивную модель производственной функции экономики США по данным за 1947—1973 гг., приведенным в табл. 8.2. В качестве переменной Y возьмем валовой национальный продукт в ценах 1958 г., в качестве K — величину основного капитала по полной стоимости в ценах 1958 г., умноженной на индекс загрузки оборудования, и в качестве L — численность занятых в народном хозяйстве.

После простого преобразования получим возможность рассчитывать производственную функцию на основе индексов, как это обычно и делается:

$$\frac{Y_t}{Y_1} = \left(\frac{K_t}{K_1} \right)^\alpha \left(\frac{L_t}{L_1} \right)^{1-\alpha} e^{\alpha + \kappa(t-1)}. \quad (8.11)$$

Делим обе части исходного уравнения (8.11) на L_t/L_1 и логарифмируем:

$$\ln \left(\frac{Y_t}{Y_1} \frac{L_1}{L_t} \right) = \alpha \ln \left(\frac{K_t}{K_1} \frac{L_1}{L_t} \right) + \kappa(t-1) + \alpha.$$

Проводим преобразование уравнения, необходимое для корректировки смещения при оценивании параметров:

$$\begin{aligned} \frac{Y_t}{Y_1} \ln \left(\frac{Y_t}{Y_1} \frac{L_1}{L_t} \right) &= \alpha \frac{Y_t}{Y_1} \ln \left(\frac{K_t}{K_1} \frac{L_1}{L_t} \right) + \kappa \frac{Y_t}{Y_1} (t-1) + \\ &+ \alpha \frac{Y_t}{Y_1}. \end{aligned} \quad (8.12)$$

Оценивание проводим методом адаптивной множественной регрессии и получаем картину эволюции параметров (см. табл. 8.2).

Таблица 8.2

Адаптивная производственная функция экономики США
(1947—1973)

Год	Исходные данные				Результаты			
	ВНП* в це- нах 1958 г., млрд. дол.	Капитал** по полной стоимости в ценах 1958 г., млн. дол.	Индекс за- ружки обо- рудования***	Занятые, тыс. чел. (16 лет и старше)*	α	κ	a	$e^{a+\kappa x} \times (t-1)$
1	2	3	4	5	6	7	8	9
1947	309,9	428 705	95,2	57 039	0,27823	0,02769	-0,00797	0,9921
1948	323,7	447 216	92,4	58 344	0,27820	0,03102	-0,00464	1,0267
1949	324,1	461 761	81,3	57 649	0,27811	0,03296	-0,00367	1,0642
1950	355,3	480 881	89,5	58 920	0,27812	0,03459	-0,00313	1,1059
1951	383,4	500 640	90,6	59 962	0,27818	0,03928	-0,00195	1,1679
1952	395,1	518 053	88,5	60 254	0,27809	0,03186	-0,00344	1,1687
1953	412,8	536 325	92,3	61 181	0,27806	0,03045	-0,00367	1,1961
1954	407,0	553 462	82,9	60 110	0,27805	0,02876	-0,00391	1,2182
1955	438,0	572 708	90,8	62 171	0,27801	0,02678	-0,00416	1,2338
1956	446,1	595 509	89,8	63 802	0,27792	0,02162	-0,00473	1,2091
1957	452,5	615 884	86,5	64 071	0,27794	0,02309	-0,00459	1,2540
1958	447,3	630 255	75,9	63 036	0,27794	0,02327	-0,00457	1,2858
1959	475,9	645 172	82,2	64 630	0,27793	0,02208	-0,00467	1,2973
1960	487,7	663 545	80,5	65 778	0,27792	0,02125	-0,00473	1,3120
1961	497,2	680 050	77,7	65 746	0,27792	0,02191	-0,00469	1,3526
1962	529,8	699 872	81,2	66 702	0,27793	0,02270	-0,00463	1,3920
1963	551,0	712 014	82,8	67 762	0,27792	0,02171	-0,00469	1,4087
1964	581,1	745 146	85,7	69 305	0,27792	0,02164	-0,00470	1,4379
1965	617,8	778 123	90,6	71 088	0,27791	0,02111	-0,00473	1,4554
1966	658,1	817 822	95,8	72 895	0,27791	0,02097	-0,00474	1,4824
1967	675,2	855 304	92,7	74 372	0,27789	0,01998	-0,00478	1,4841
1968	706,6	894 529	94,5	75 920	0,27789	0,02002	-0,00478	1,5153
1969	725,6	936 600	95,6	77 902	0,27784	0,01791	-0,00488	1,4757
1970	722,5	973 400	87,9	78 627	0,27785	0,01805	-0,00487	1,5072
1971	746,3	1006 800	85,3	79 120	0,27785	0,01856	-0,00485	1,5536
1972	792,5	1045 100	89,6	81 702	0,27784	0,01787	-0,00488	1,5556
1973	839,2	1090 200	95,8	84 409	0,27783	0,01739	-0,00490	1,5640

* Economic report of the president, transmitted to the congress, Feb. 1975.

** Fixed nonresidential business, Capital in the U. S., 1925—1970, p. 1; Survey of current business, 1974, n. 3.

*** Economic report of the president, 1976, Wharton series.

Таблица 8.3

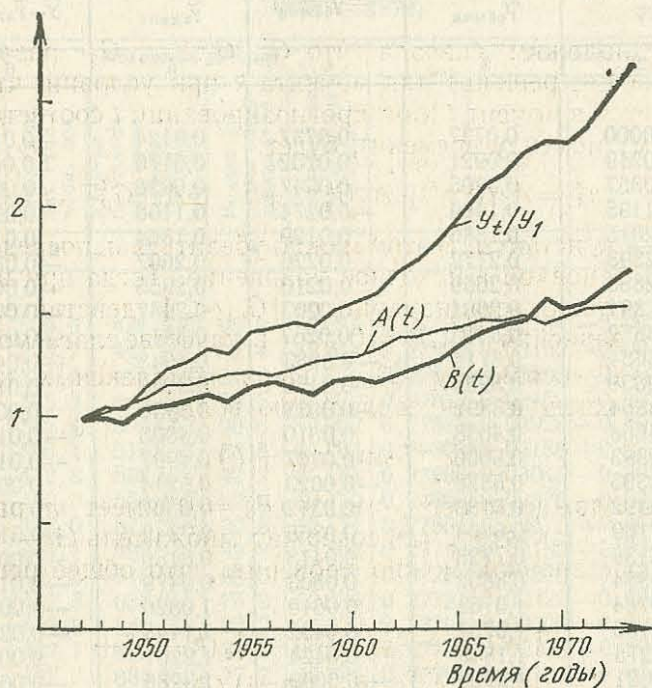
Сравнение обычной и адаптивной множественных регрессий

№ точки	Фактическое значение эндогенной переменной $\hat{Y}$	Обычная регрессия		Адаптивная регрессия	
		Модельное значение $\hat{Y}_{обыч}$	Ошибка $\hat{Y} - \hat{Y}_{обыч}$	Модельное значение $\hat{Y}_{адапт}$	Ошибка $\hat{Y} - \hat{Y}_{адапт}$
1	2	3	4	5	6
1	0,0000	0,0737	-0,0737	0,0124	-0,0124
2	0,0219	0,0921	-0,0702	0,0176	0,0042
3	0,0357	0,0905	-0,0547	0,0326	0,0031
4	0,1195	0,1469	-0,0274	0,1158	0,0038
5	0,2015	0,1886	0,0129	0,1864	0,0150
6	0,2398	0,2172	0,0226	0,2697	-0,0300
7	0,2885	0,2669	0,0216	0,2956	-0,0070
8	0,2891	0,2684	0,0207	0,2987	-0,0096
9	0,3672	0,3405	0,0267	0,3811	-0,0138
10	0,3631	0,3714	-0,0083	0,4043	-0,0412
11	0,3830	0,3985	-0,0156	0,3698	0,0132
12	0,3854	0,3892	-0,0038	0,3836	0,0018
13	0,4668	0,4658	0,0010	0,4803	-0,0135
14	0,4893	0,5000	-0,0107	0,4997	-0,0105
15	0,5305	0,5328	-0,0023	0,5214	0,0091
16	0,6492	0,6180	0,0312	0,6367	0,0125
17	0,7169	0,6804	0,0365	0,7342	-0,0173
18	0,8136	0,7723	0,0413	0,8150	-0,0014
19	0,9364	0,8862	0,0503	0,9479	-0,0115
20	1,0784	1,0165	0,0619	1,0820	-0,0036
21	1,1186	1,0754	0,0432	1,1448	-0,0262
22	1,2273	1,1849	0,0424	1,2261	0,0012
23	1,2621	1,2720	-0,0099	1,3286	-0,0665
24	1,2251	1,2776	-0,0525	1,2205	0,0046
25	1,3285	1,3584	-0,0300	1,3103	0,0182
26	1,4822	1,5154	-0,0332	1,5092	-0,0270
27	1,6363	1,6950	-0,0587	1,6572	-0,0208

Средние квадраты отклонений

В обычной регрессии, оцененной МНК	В адаптивной регрессии
0,00163	0,00043

Начальные значения параметров для процесса адаптации были получены как оценки МНК обычной регрессии, построенной на первых N_1 точках выборок. Наилучшее N_1 , найденное методом перебора, оказалось равным 9, а наилучшее значение параметра адаптации равным 1,642.



y_t — ВВП США

$A(t)$ — показатель экономической эффективности

$B(t) = (y_t/y_1) / A(t)$

Рис. 8.1. Иллюстрация результатов, полученных по адаптивной модели производственной функции экономики США по данным за 1947--1973 гг.

Динамика показателя экономической эффективности представлена в столбце 9 табл. 8.2. Следует отметить, что явная тенденция показателя к снижению свидетельствует о невыполнении обычно принимаемой гипотезы об экспоненциальном росте экономической эффективности. В действительности, как это следует из адаптивной модели, про-

исходит последовательный переход с более быстрой экспоненты на более медленную. Технический прогресс при этом носит нейтральный характер ($\frac{\alpha}{1-\alpha} \approx \text{const}$).

Для сравнения уравнение (8.12) было оценено обычным МНК по тем же данным (см. табл. 8.3).

Сопоставление средних квадратов отклонений приводит к выводу о значительно более гибком отражении функционирования экономики адаптивной моделью.

На рис. 8.1 на графике по оси ординат отложены индексы валового национального продукта (ВВП) и значения показателя экономической эффективности $A(t)$. Показатель экономической эффективности отражает роль интенсивных факторов роста, а роль экстенсивных факторов можно определить разделив индекс ВВП на соответствующий показатель экономической эффективности. Этот показатель, обозначенный как $B(t)$, также представлен на графике.

Легко заметить совпадение тенденций индекса ВВП и показателя $B(t)$ и колеблемость эффективности $A(t)$ как бы в противофазе с $B(t)$ и индексом ВВП. Первое говорит о прямом влиянии экстенсивных факторов на рост ВВП, а второе — о падении эффективности в период роста производства и ее увеличении во время кризисных явлений. Это объясняется тем, что в период экономических спадов из процесса производства в первую очередь выбывают устаревшие производственные фонды и работники с низкой квалификацией, а в период подъема начинают функционировать и отсталые предприятия, которые тянут показатель экономической эффективности вниз.

В заключение отметим, что результаты, получаемые методом адаптивного моделирования, могут давать ценную информацию о последовательном развитии экономических систем и полезны как для краткосрочного прогнозирования, так и для анализа долгосрочных тенденций.

Глава 9

НЕТРАДИЦИОННЫЙ КОРРЕЛЯЦИОННЫЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

В этой главе дана критика применения классической теории корреляции для анализа связи временных (и, в частности, экономических) рядов, когда не выполняются гипотезы об их стационарности и эргодичности. Предлагается ряд новых измерителей.

§ 1. УСЛОВИЯ ПРИМЕНИМОСТИ ТРАДИЦИОННОГО КОРРЕЛЯЦИОННОГО АНАЛИЗА

Корреляционный анализ часто используется в экономических исследованиях для изучения статистической связи между переменными, представленными временными рядами [3]. Несмотря на то что он не позволяет отделить прямые причинно-следственные связи от косвенных или даже просто от случайных совпадений колебаний переменных, этот анализ ценен тем, что предоставляет возможность придать количественную определенность степени сопряженности колебаний этих рядов, подтвердить или поставить под сомнение те или иные гипотезы. Корреляционный анализ в соединении с содержательным исследованием оказывается полезным и достаточно простым аналитическим средством. Этим и объясняется его популярность у экономистов.

Корреляционная теория была разработана для случайных стационарных процессов, допускающих получение любого числа реализаций (наблюдений, измерений). Под *стационарностью* понимается неизменность среднего уровня

случайного процесса, дисперсии отклонений и постоянство автокорреляционной функции. Эта теория предназначалась прежде всего для технических приложений, где гипотеза о стационарности считается приемлемой, а принципиальных ограничений на число реализаций нет.

На самом деле выдвигаются (или молчаливо используются) даже еще более жесткие требования к случайному процессу, заключающиеся в том, что предполагается его эргодичность. *Эргодическими случайными процессами* называется подкласс стационарных случайных процессов, для которых усреднение по множеству реализаций эквивалентно усреднению по времени. Тогда корреляционный анализ особенно упрощается, и это одна из причин его распространенности.

Именно эта теория, разработанная для стационарных и эргодических случайных процессов, некритически воспринята экономистами и используется ими для анализа процессов заведомо нестационарных (например, имеющих тренд) и даже неслучайных (случайный процесс должен описываться тем или иным законом распределения вероятностей), которые обычно представлены единственной реализацией. Большое число реализаций в экономике при одних и тех же условиях часто принципиально невозможно.

В экономических исследованиях требуется особая теория корреляционного анализа, специально предназначенная для изучения сопряженности (или связности) колебаний нестационарных процессов, представленных единственной реализацией. Такая теория, предложенная нами в [122] и [125], и будет изложена в этой главе. Вводятся новые виды корреляционных коэффициентов, позволяющие измерять изменение силы связи двух переменных во времени, а также в зависимости от амплитуды их отклонения от некоторой критической величины. Рассматривается корреляционная связь с вероятностной точки зрения.

§ 2. ПОСТАНОВКА ПРОБЛЕМЫ

Постулирование стационарности и эргодичности процессов x_{1t} и x_{2t} — вынужденная мера, к которой экономисты прибегают, чтобы обеспечить себе хоть какую-то возможность провести корреляционный анализ связи двух перемен-

ных. Арифметические средние величины m_1 и m_2 трактуются как неизменные средние уровни процессов x_{1t} и x_{2t} соответственно, а $\hat{\sigma}_1$ и $\hat{\sigma}_2$ — как постоянные среднеквадратические отклонения от среднего. И это, с нашей точки зрения, вообще говоря, неправомерно.

На самом деле переменные x_{1t} и x_{2t} , как правило, не имеют фиксированных средних уровней и каких-либо определенных среднеквадратических отклонений от них. Величины m_1 и m_2 являются некоторыми условными уровнями, относительно которых вычисляются отклонения исследуемых рядов. В этой ситуации обычный показатель корреляции r выражает скорее силу связи между такими отклонениями рядов x_{1t} и x_{2t} от m_1 и m_2 , чем между собственно рядами x_{1t} и x_{2t} . Неопределенный характер m_1 и m_2 объясняется тем, что они не отражают каких-либо действительных, устойчивых характеристик временных рядов. Дело в том, что как только мы сдвинем границы выборочного периода, т.е. расширим, сузим или как-то изменим его диапазон, все оценки m_1 и m_2 , $\hat{\sigma}_1$ и $\hat{\sigma}_2$ станут другими, особенно (как это обычно и бывает в реальных данных) если исследуемые ряды имеют временные тренды. Это приводит к тому, что одному и тому же моменту времени при одних и тех же статистических наблюдениях будут соответствовать отклонения ряда, различающиеся не только по величине, но, возможно, даже и по знаку в зависимости от таких внешних по отношению к данной точке обстоятельств, как расположение границ выборочного периода на оси времени. Тем более что, как известно, вопрос о границах выборочного периода (тесно связанный с m) решается исследователем субъективно. Таким образом, единое движение ряда от момента $t-1$ к моменту t может оказаться искусственно и произвольно разделенным величиной m на положительную и отрицательную части. В результате слагаемое в числителе формулы для коэффициента корреляции, соответствующее моменту t , также будет переменным, зависящим от m_1 и m_2 . А ведь слагаемое — фактически локальное, частное свидетельство о характере корреляционной связи рядов, поступившее в момент t . Зыбкость, изменчивость этого свидетельства, его зависимость от внешних обстоятельств представляется серьезным недостатком изложенной методики, проистекающим из ложной посылки, что классическую теорию корреляционного анализа можно формально применять для изучения нестационар-

ных экономических временных рядов. С нашей точки зрения, переход от наблюдаемых рядов x_{1t} и x_{2t} к рядам $(x_{1t} - m_1)$ и $(x_{2t} - m_2)$ в общем случае «портит» первичную информацию.

Выход из этого затруднительного положения обычно находят в том, чтобы исходный нестационарный ряд преобразовать к приблизительно стационарному виду. С этой целью предпринимается попытка предварительно исключить из рядов временные тренды. Но, во-первых, эта операция также зависит от границ выборочного периода и таких субъективных факторов, как выбор типа тренда, метода его оценивания, и, во-вторых, в результате элиминирования тренда из временного ряда удаляется весьма существенная информация. Остатки, после исключения тренда, как правило, оказываются слабо коррелированными, а исходные ряды могут обнаруживать корреляцию, близкую к идеальной из-за почти одинакового характера тренда. Получив такие результаты, практик остается в растерянности. Можно констатировать, что в случае предварительного исключения тренда искажение первичной информации лишь усугубляется.

§ 3. МОДИФИЦИРОВАННЫЙ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

Возможен, однако, другой выход: судить о наличии положительной или отрицательной корреляции в момент t по совпадению или несовпадению знаков прироста интересующих переменных. Для получения общего представления об усредненных корреляционных свойствах двух нестационарных рядов целесообразно разработать специальный статистический показатель корреляции, основывающийся на использовании приростов переменных. Для этого модифицируем коэффициент корреляции следующим образом (для упрощения будем считать лаг $\tau = 0$):

$$r_{\text{mod}} = \frac{\frac{1}{T} \sum_{t=2}^T \Delta x_{1t} \Delta x_{2t}}{\frac{1}{T-1} \sum_{t=2}^T |\Delta x_{1t} \Delta x_{2t}|} = \frac{\sum_{t=2}^T \Delta x_{1t} \Delta x_{2t}}{\sum_{t=2}^T |\Delta x_{1t} \Delta x_{2t}|}, \quad (9.1)$$

где $\Delta x_{it} = x_{it} - x_{i,t-1}$, $i = 1, 2$.

Знаменатель формулы (9.1) играет роль нормирующего коэффициента. Благодаря ему r_{mod} не может выйти за пределы $-1 \leq r_{\text{mod}} \leq 1$. При таком измерителе степени коррелированности двух рядов отмеченный выше недостаток устраняется и никакой деформации исходных данных не происходит. Каждое слагаемое в числителе и знаменателе неизменно, не зависит от каких-либо субъективных процедур или желаний исследователя.

Отметим, что усреднение по времени вместо усреднения по реализациям здесь остается, поскольку мы считаем, что имеется лишь одна реализация для каждого момента и другого пути нет. По существу, такое усреднение основывается на гипотезе о постоянстве корреляционной связи двух рядов на всем выборочном периоде. Это предположение, очевидно, также слишком сильное для экономических нестационарных процессов. Мы к нему вернемся несколько позже. А сейчас продолжим анализ формулы (9.1).

Вычисление этого коэффициента возможно для неслучайных функций, так как его определение не связано с какими-либо гипотезами вероятностного характера. В частности, легко заметить особенность предложенного коэффициента корреляции, заключающуюся в том, что модифицированный коэффициент корреляции любых двух монотонно возрастающих или монотонно убывающих функций всегда будет равен единице. В этом случае целесообразно перейти к анализу разностей второго (или даже более высокого) порядка. Получаемый коэффициент можно назвать *модифицированным коэффициентом корреляции второго (или более высокого) порядка*.

§ 4. АДАПТИВНЫЙ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

Напомним, что речь идет прежде всего о нестационарных временных рядах: корреляционная связь между ними может становиться то слабее, то сильнее, и если найти способ выявить динамику этой связи, то можно получить дополнительную интересную информацию о взаимосвязях переменных, полезную в содержательном анализе исследуемых явлений. В таком случае гипотезу о постоянстве корреляционной связи, использовавшейся в (9.1), придется при-

знать несостоятельной. Арифметическое усреднение по времени дает лишь весьма грубый результат, ничего не говорящий о движении коэффициента корреляции во времени. Поэтому в [122] и [125] предложен еще один, рекуррентный (или адаптивный), коэффициент корреляции

$$r_t(\alpha) = \frac{s_t}{d_t}, \quad (9.2)$$

$$\begin{aligned} s_t &= (1 - \alpha) s_{t-1} + \alpha (\Delta x_{1t} \Delta x_{2t}), \quad t = 1, 2, \dots, T, \\ d_t &= (1 - \alpha) d_{t-1} + \alpha |\Delta x_{1t} \Delta x_{2t}|, \quad 0 < \alpha < 1. \end{aligned}$$

Как видим, s_t и d_t являются экспоненциально-взвешенными скользящими средними произведений приростов и абсолютных произведений приростов двух рядов. Параметр α — постоянная сглаживания или параметр адаптации (гл. 1 §2). Ясно, что и здесь $-1 \leq r_t(\alpha) \leq 1$. График $r_t(\alpha)$ относительно времени показывает, как изменяется сила корреляционной связи на выборочном периоде. Открытым остается лишь вопрос о начальных значениях s_0 и d_0 , необходимых для начала рекуррентной процедуры, и величине постоянной сглаживания α .

Начальные значения s_0 и d_0 предлагаем определять как простые арифметические средние произведений и абсолютных значений произведений приростов, вычисляемые на основе T_0 наиболее ранних наблюдений выборки, $0 < T_0 \leq T-1$. Тогда $r_0 = s_0 / d_0$ будет начальным коэффициентом корреляции, совпадающим с модифицированным коэффициентом корреляции, рассчитанным по (9.1) при $T = T_0$. Величина α может выбираться из указанного диапазона $[0, 1]$, исходя из конкретного применения вычисляемого коэффициента корреляции.

Если вычисление такого коэффициента ставится как самостоятельная задача, то можно предложить следующий способ определения наилучшего α . Будем рассматривать $r_{t-1}(\alpha)$ как прогноз совпадения колеблемости рядов x_{1t} и x_{2t} в момент t , а

$$e_t = \frac{\Delta x_{1t} \Delta x_{2t}}{|\Delta x_{1t} \Delta x_{2t}|} - r_{t-1}(\alpha) \quad (9.3)$$

как ошибку этого прогноза на тот же момент. Тогда можно ввести функцию потерь $Q_v(\alpha)$

$$Q_v(\alpha) = \sum_{t=2}^T |e_t|^v, \quad (9.4)$$

где $v=1$ или 2 , и определять оптимальное значение α путем минимизации функции $Q_v(\alpha)$ по α . Минимум легко найти методом проб различных α из интервала $[0, 1]$. При необходимости точно так же методом проб устанавливается и наилучшее значение T_0 . В этом случае функция потерь рассматривается как функция двух переменных $Q_v(T_0, \alpha)$.

Возможен и другой вариант адаптивного коэффициента корреляции

$$r_t = (1-\alpha)r_{t-1} + \alpha \frac{\Delta x_{1t} \Delta x_{2t}}{|\Delta x_{1t} \Delta x_{2t}|}, \quad (9.5)$$

здесь в качестве r_0 можно взять отношение $r_0 = c_0 / q_0$,

$$\text{где } c_0 = \frac{1}{T_0} \sum_{j=1}^{T_0} \Delta x_{1j} \Delta x_{2j}, \quad q_0 = \frac{1}{T_0} \sum_{j=1}^{T_0} |\Delta x_{1j} \Delta x_{2j}|.$$

Такой коэффициент назовем *адаптивным коэффициентом знаковой корреляции приростов*. Далее, если действовать так же, как и в предыдущем случае, нетрудно вычислить ошибки e_t и построить критерий $Q_v(T_0, \alpha)$, используя (9.4) и (9.5). Оптимальные значения T_0 и α будут соответствовать минимуму Q .

§ 5. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ ОТКЛОНЕНИЙ ОТ ЗАДАНЫХ УРОВНЕЙ

В классическом корреляционном анализе отсчет отклонений переменной ведется от математического ожидания. Однако на практике такое отклонение представляет интерес лишь в частных случаях. В экономических исследо-

ваниях, например, гораздо чаще важен знак отклонения от предыдущего значения ряда (т.е. знак приращения) или от нулевого уровня. Тогда и отклонения целесообразно рассматривать от соответствующих величин.

Вообще говоря, в зависимости от задачи анализу могут подвергаться отклонения от моды, медианы, от какого-либо заданного уровня. Назовем такой уровень критическим, подразумевая под этим, что отклонения от него в разные стороны имеют существенные качественные различия. Обозначим такой уровень для ряда x_t через E_x . Пусть в этом случае требуется установить, существует ли взаимосвязь между отклонениями одной переменной, например $(x_t - E_x)$, с отклонениями другой $(z_t - E_z)$, и если существует, то насколько она сильна.

Возьмем в качестве измерителя корреляционной связи в этом случае коэффициент

$$r_E = \frac{E[(x_t - E_x)(z_t - E_z)]}{E|(x_t - E_x)(z_t - E_z)|} \quad (9.6)$$

и рассмотрим его свойства. Совершенно ясно, что он находится в обычных пределах $-1 \leq r_E \leq 1$. Его оценкой может служить корреляционный коэффициент:

$$\hat{r}_E = \frac{\sum (x_t - E_x)(z_t - E_z)}{\sum |(x_t - E_x)(z_t - E_z)|} \quad (9.7)$$

Если $E_x = 0$ и $E_z = 0$, то коэффициент r_E характеризует силу взаимосвязи между отклонениями соответствующих переменных от нуля. При $E_x = x_{t-1}$ и $E_z = z_{t-1}$ получаем рассмотренный выше модифицированный коэффициент корреляции, который также можно назвать *коэффициентом корреляции приростов*. Отметим, однако, что при $E_x = M(x)$ и $E_z = M(z)$, где M — символ математического ожидания, коэффициент r_E не идентичен классическому коэффициенту корреляции из-за различных нормирующих множителей (знаменателей).

Никаких ограничений для использования корреляционного коэффициента (9.7) нет. Он применим для переменных, имеющих смещенное (когда $E_x \neq M(x)$) и несимметричное

распределение относительно критического уровня E_x . Однако стационарность и эргодичность переменных предполагаются. Если же эти предположения слишком сильные, можно построить адаптивный коэффициент корреляции по аналогии с (9.2), где приращения следует заменить на отклонения от критических значений.

§ 6. УСЛОВНЫЙ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

Рассмотренные коэффициенты корреляции отражают силу взаимосвязи колебаний двух рядов в целом. Такая характеристика может оказаться слишком общей и не всегда удовлетворит исследователя. Поставим вопрос конкретнее: насколько сильна связь, например, положительных отклонений x_i с отклонениями z_i ? Для ее измерения введем понятие *условного коэффициента корреляции*. Определим его как

$$r_c(x_i, z_i | x_i \geq E_x) = \frac{\sum (x_i - E_x)(z_i - E_z)}{\sum |(x_i - E_x)(z_i - E_z)|} | x_i \geq E_x, \quad (9.8)$$

где суммируются только те произведения, в которых $x_i \geq E_x$. Точно так же можно вычислить коэффициент взаимосвязи отрицательных отклонений ряда x_i с отклонением z_i . Естественно, что в обоих случаях коэффициенты корреляции могут принимать неодинаковые значения. Таким образом, особенности взаимосвязи двух рядов теперь не скрываются за одним-единственным значением коэффициента корреляции, как это предполагается в классической теории, а представляются более подробно двумя коэффициентами.

Преимущества такого подхода наиболее ярко проявляются тогда, когда одна переменная (например, x_i) является контролируемой, а другая (z_i) — непосредственно недоступной, но коррелированной с первой. Тогда можно «нащупать» такое критическое значение E_x , при котором удерживание x_i в области $E_x \leq x_i$ (или $x_i \leq E_x$) имело бы наиболее благоприятные последствия (т.е. наибольшее значение коэффициента корреляции) для удержания z_i в пределах $z_i \leq E_z$ (или, наоборот, $z_i \geq E_z$).

В общем виде условия при вычислении коэффициента корреляции могут, конечно, иметь произвольный вид

$$r_c(x_i, z_i | A \leq x_i - E_x \leq B) = \frac{\sum (x_i - E_x)(z_i - E_z)}{\sum |(x_i - E_x)(z_i - E_z)|} | A \leq x_i - E_x \leq B \quad (9.9)$$

лишь бы хватило статистических данных для получения устойчивых оценок коэффициентов.

Доводом в пользу применения условного коэффициента корреляции (9.9) могут служить следующие рассуждения. При низких значениях амплитуды отклонений корреляционная связь может не просматриваться из-за влияния случайных помех, т.е. быть несущественной. А при больших отклонениях определенность этой связи возрастает почти до детерминированности. Тогда при большой амплитуде отклонений одной переменной резко возрастает степень предопределенности области нахождения (например, знак прироста) второй переменной. Легче определить и лаг (запаздывание во времени) одной переменной относительно другой. Можно считать, что в этом случае возрастает мощность критерия.

Задавая значения A и B , можно выделить любую интересующую нас область амплитуд отклонений первой переменной ($x_i - E_x$) и вычислить коэффициент корреляции с переменной z_i для этой области. Такой подход открывает путь к доскональному изучению взаимосвязи рядов, выявлению специфических локальных корреляционных свойств. Можно сказать, что корреляционные свойства предстают в «расщепленном» виде — их выражает множество коэффициентов.

§ 7. ВЕРОЯТНОСТНЫЙ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

Рассмотрим еще один метод исследования силы сопряженности двух переменных на основе оценки условной вероятности их связи [125]. Определим коэффициент корреляции как

$$r_p(x_t, z_t | A \leq x_t - E_x \leq B, C \leq z_t - E_z \leq D) = \quad (9.10)$$

$$= p(C \leq z_t - E_z \leq D | A \leq x_t - E_x \leq B),$$

где p — вероятность того, что $C \leq z_t - E_z \leq D$ при условии $A \leq x_t - E_x \leq B$. Отсюда следует, что областью существования коэффициента корреляции будет отрезок $[0, 1]$. Она отличается от области существования рассмотренных выше коэффициентов корреляции, которая охватывает отрезок $[-1, 1]$.

Определение (9.10) теоретическое. Практическую оценку на основе статистических данных можно получить как отношение

$$\hat{r}_p(x_t, z_t | A \leq x_t - E_x \leq B, C \leq z_t - E_z \leq D) = \frac{n_{zx}}{N_x}, \quad (9.11)$$

где n_{zx} — число наблюдений, когда $z_t - E_z$ оказывалось в пределах $C \leq z_t - E_z \leq D$ при $A \leq x_t - E_x \leq B$; N_x — общее число наблюдений в выборке, удовлетворяющих условию $A \leq x_t - E_x \leq B$.

Обратное направление связи характеризуется величиной

$$\hat{r}_p(z_t, x_t | C \leq z_t - E_z \leq D, A \leq x_t - E_x \leq B) = \frac{n_{xz}}{N_z}. \quad (9.12)$$

Здесь обозначения аналогичны использованным в (9.11). Отсюда очевидно, что в общем случае

$$\begin{aligned} \hat{r}_p(x_t, z_t | A \leq x_t - E_x \leq B, C \leq z_t - E_z \leq D) &\neq \\ \neq \hat{r}_p(z_t, x_t | C \leq z_t - E_z \leq D, A \leq x_t - E_x \leq B). \end{aligned} \quad (9.13)$$

Этот вид корреляционного коэффициента позволяет исследовать взаимосвязь переменных с вероятностной точки зрения и найти диапазоны отклонений переменных, взаимосвязь между которыми наблюдается с наибольшей вероятностью.

§ 8. ПРИМЕР

На рис. 9.1 показано, как меняется адаптивный коэффициент корреляции процентных ставок по долгосрочным (первый ряд) и краткосрочным (второй ряд) операциям, рассчитанный на основе годовых данных США с 1857 по 1985 г., $v = 1$. Эта связь достаточно высокая, $r_p(\alpha)$ — повсюду выше 0,5. Можно, однако, отметить, что в XX в., особенно после первой мировой войны, взаимосвязь этих двух величин стала теснее, чем во второй половине XIX столетия. Можно также видеть временное снижение корреляции двух показателей во время первой мировой войны и в период мирового энергетического кризиса после 1974 г. Найдены оптимальные значения параметров: $T_0 = 18$, $\alpha_{opt} = 0,0729$ при $v = 1$.

Для сравнения интересно отметить, что обычный коэффициент корреляции равен 0,859, а модифицированный — 0,836. Условный коэффициент корреляции рассчитывался для первых разностей двух рядов. При этом критическим уровнем был выбран нулевой. Получены следующие результаты. Условный коэффициент корреляции положительных приростов процентных ставок по долгосрочным операциям с приростами ставок по краткосрочным операциям составил 0,85, а отрицательных приростов первого ряда с приростами второго — 0,82.

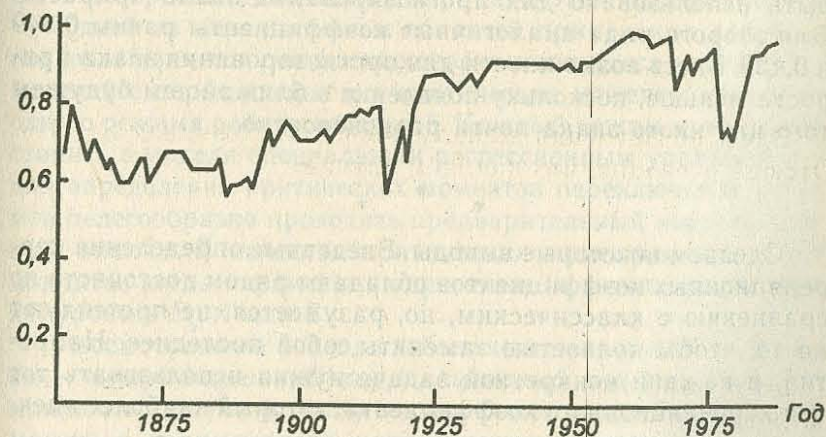


Рис. 9.1. Адаптивный коэффициент корреляции процентных ставок по краткосрочным и долгосрочным кредитам

Это свидетельствует о приблизительно одинаковой силе связи падений и подъемов первого ряда с колебаниями второго.

Вероятностный коэффициент корреляции положительных приростов первого ряда с положительными приростами второго равен 0,79, а отрицательных с отрицательными — только 0,64. Как видим, вероятность совпадения подъемов несколько выше вероятности совместного снижения рассматриваемых процентных ставок.

Вычисление условных коэффициентов автокорреляции соседних приростов ($\tau = 1$) дало следующие результаты. У первого ряда корреляционная связь положительного текущего прироста с будущим приростом характеризуется коэффициентом 0,52, а текущего отрицательного прироста с будущим приростом — 0,41. Таким образом, положительные приросты первого ряда позволяют строить более точные прогнозы на будущее, чем отрицательные. Аналогичные коэффициенты для второго ряда соответственно равны $-0,21$ и $0,08$, т.е. связь последовательных приростов у процентных ставок по краткосрочным операциям совершенно незначительна.

Вероятностные коэффициенты автокорреляции приростов позволяют сделать такие выводы. Положительные текущие приросты первого ряда влекут за собой положительные приросты этого же ряда с вероятностью 52,6% ($r = 0,526$), а отрицательные с вероятностью 60,6% ($r_p = 0,606$) сопровождаются также отрицательными приростами. Это может быть использовано для прогнозирования знака прироста. Для второго ряда аналогичные коэффициенты равны 0,549 и 0,429. Здесь возможности для прогнозирования знака прироста меньше, поскольку появление в ближайшем будущем того или иного знака почти равновероятно.

* * *

Сделаем некоторые выводы. Введенные определения корреляционных коэффициентов обладают рядом достоинств по сравнению с классическим, но, разумеется, не претендуют на то, чтобы полностью заменить собой последнее. Напротив, в каждой конкретной задаче нужно использовать тот тип корреляционного коэффициента, который наиболее адекватен поставленной цели.

В предложенном модифицированном коэффициенте корреляции учитываются приращения переменных вместе с их

знаком. Это равносильно тому, что вместо отклонений от математического ожидания здесь рассматриваются отклонения от ранее достигнутых значений. Идея отказа от математического ожидания как от уровня отсчета отклонений и замены его на некоторый критический (в сущности, произвольный, заданный исследователем) уровень нашла себе применение и в других рассмотренных нами определениях. Адаптивный коэффициент корреляции позволяет проследивать динамику корреляционной связи во времени. Для исследования корреляции знаков приращений переменных построен адаптивный коэффициент знаковой корреляции. Практическое использование этого подхода для краткосрочного прогнозирования валютных курсов освещено в [126]. Этот метод может, конечно, применяться и на рынке ценных бумаг.

Условный коэффициент корреляции, как и вероятностный, можно назвать *парциальными коэффициентами корреляции*, так как они способны отражать частичные свойства двух исследуемых временных рядов, наблюдаемые при определенных условиях. Два последних подхода позволяют проводить дифференцированный анализ взаимосвязи переменных, выявлять, переходя от одного диапазона амплитуд отклонений к другому, неоднородность связи в разных поддиапазонах и давать более богатую информацию исследователю для размышления и использования в аналитической и прогнозной деятельности. При необходимости могут быть построены адаптивные варианты этих коэффициентов.

Одной из перспективных областей применения подобных подходов является построение *переключающихся моделей*, т.е. моделей с переключением в критические моменты времени с одного режима работы на другой. Каждый режим можно представить в модели специальным регрессионным уравнением, а для определения критических моментов переключения режимов целесообразно проводить предварительный корреляционный анализ. В качестве критических моментов выступают моменты, когда переменная пересекает (в наших обозначениях) уровень E_c , A или B .

Рассмотренные подходы, объединенные общим названием — *нетрадиционный корреляционный анализ*, дают возможность получать значительно более подробную информацию о взаимосвязях переменных, чем это предлагает классическая теория, и открывают новые перспективы в познании сложных динамических систем.

Глава 10

ФАЗОВЫЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

Рассматривается подход к анализу нерегулярных циклических колебаний временных рядов, теоретические и методологические основы для изучения этих колебаний по фазам циклов, предложенные в [129]. Описан анализ циклических колебаний на примере инвестиционных процессов в США и странах Западной Европы.

§ 1. ПОСТАНОВКА ПРОБЛЕМЫ

В экономических исследованиях часто встречаются показатели, движение которых во времени представляет собой нерегулярные циклические колебания с переменной амплитудой и неустойчивым периодом. Внешне стохастический характер этих явлений отражает циклическое развитие экономики в условиях влияния на нее многих случайных, рыночных и волевых, управляющих воздействий.

Исследование фаз движения упомянутых показателей традиционными методами, такими, как спектральный и регрессионный анализ, по ряду причин не подходит. Так, спектральный метод аппроксимирует движение временного ряда, отражающего движение того или иного показателя, суммой регулярных синусоид. Между тем строгой периодичности и постоянства амплитуды от экономических показателей трудно ожидать, поэтому для хорошей аппроксимации требуется большое число синусоид различной частоты. Возникают трудности и с интерпретацией результатов [120]. Регрессионный метод аппроксимирует всю выборку в целом, и локальные свойства ряда при этом играют второстепенную

роль. Между тем в экономике каждый цикл обычно имеет свои особенности и важно не потерять их во время статистической обработки. Очень часто наблюдается асимметрия колебаний относительно уровня отсчета. Например, у положительных полуволи может быть одна амплитуда и длительность, а у отрицательных — другая.

Для исследования локальных колебаний временного ряда существенным является определение фазы.

§ 2. РАЗБИЕНИЕ ВРЕМЕННОГО РЯДА НА ФАЗЫ

При исследовании циклических колебаний экономисты используют нечеткое понятие цикла. Представление о цикле изменяется в зависимости от того, какие свойства ряда — краткосрочные или долгосрочные — интересуют аналитика. Положительная или отрицательная фаза цикла определяется знаком остатка, полученного вычитанием непосредственно из исходного ряда того или иного временного тренда. Но временной тренд, как известно, устанавливается весьма субъективно. Иногда исследователь и вовсе экспертно делит весь ряд на фазы, а затем анализирует полученное [147, 152]. Поэтому в [129] предпринята попытка разработать более объективные способы вычленения фаз во временном ряде и специальные статистические методы их изучения.

Для изложения предложенного подхода сначала введем некоторые определения.

Пусть x_t — исходный временной ряд, наблюдаемый в моменты $t = 1, 2, \dots, N$.

Флуктуация — величина отклонения x_t от некоторого установленного исследователем уровня. Это может быть от-

клонение от среднего уровня $m_x = \frac{1}{N} \sum_{t=1}^N x_t$, предшествующего значения x_{t-1} , нуля, тренда или какой-либо другой величины. Важно отметить, что уровень отсчета флуктуаций в общем случае может меняться: его динамику обычно отражает тренд. Понятие тренда достаточно неопределенно. Его вид и параметры зависят от объема выборки, опыта и желаний исследователя, от метода оценивания. Все это позволяет сделать вывод, что элиминирование (исключение) трен-

да, вообще говоря, существенно трансформирует исходный ряд. Переход от первичных наблюдений к исчисленным отклонениям от искусственно построенной линии тренда может повлечь за собой ухудшение качества исходной для анализа информации. Поэтому с операциями исключения тренда нужно быть очень осторожным. Предпочтительным способом исключения тренда, с нашей точки зрения, является переход к логарифмам, разностям, цепным индексам, относительным величинам (когда тренд элиминируется делением одной величины на другую). Но во всех случаях исследователь должен определить, какие именно отклонения он рассматривает в качестве флуктуаций ряда. Обозначим величину флуктуации в момент t через Dx_t .

Мощность флуктуации — абсолютная величина флуктуации $|Dx_t|$. **Нерегулярные циклические колебания** — наличие у ряда x_t разнонаправленных отклонений Dx_t с различной амплитудой без их детерминированной повторяемости. Принципиально важно для фазового анализа, чтобы в выборке присутствовали отклонения обоого знака. В противном случае выборка представляет собой одну фазу, и тогда задачи разбиения на фазы не существует.

Фаза — период положительных или отрицательных флуктуаций ряда Dx_t . **Положительная и отрицательная фазы** — периоды соответственно положительных и отрицательных флуктуаций ряда. **Длительность фазы** — временной отрезок положительных или отрицательных флуктуаций ряда. **Точка смены фаз** — момент смены знака флуктуаций. **Мощность флуктуаций ряда**, или просто **мощность ряда**, — сумма абсолютных флуктуаций всего ряда, т.е.

$$P = \sum_t |Dx_t|. \quad (10.1)$$

Мощность фазы — сумма абсолютных флуктуаций ряда внутри фазы, т.е. мощность фазы i

$$p_i = \sum_{t=t_i}^{t_{i+1}} |Dx_t|, \quad (10.2)$$

где t_i — момент начала фазы i ;
 t_{i+1} — момент окончания фазы i (поворотные точки).

Из (10.1) и (10.2) следует, что мощность ряда равна сумме мощностей составляющих его фаз, т.е.

$$P = \sum_i p_i. \quad (10.3)$$

Уровень агрегирования фаз — номер итерации в процессе последовательного объединения фаз путем фильтрации маломощных флуктуаций. **Уровень теряемой мощности** — процент мощности ряда, теряемой при элиминировании маломощных флуктуаций. **Фазовая диаграмма** — представление временного ряда в виде сменяющих друг друга фаз.

Основная идея метода заключается в том, чтобы постепенно (итеративно) очищать ряд от маломощных колебаний, отождествляемых со случайными или второстепенными, конъюнктурными флуктуациями. В ходе такого процесса фильтрации производится сглаживание наименее мощных фаз, в результате чего соседние фазы объединяются в одну более крупную. Поясним это следующим образом.

Начнем с исходного ряда x_t . Для определенности будем считать его флуктуацией величину $Dx_t = x_t - x_{t-1}$ — прирост ряда. Найдем моменты t , в которых выполняется одно из двух условий:

$$x_{t-1} \leq x_t > x_{t+1} \quad (10.4a)$$

либо

$$x_{t-1} \geq x_t < x_{t+1}. \quad (10.4b)$$

Первое условие означает, что в точке t — локальный максимум, и ее можно рассматривать в качестве поворотной, так как за ней следует снижение ряда. Второе условие свидетельствует о том, что в точке t — локальный минимум, и ее тоже можно рассматривать как поворотную, так как в ней заканчивается снижение, а после нее начинается возрастание ряда. Найденные поворотные точки обозначим $t_1, \dots, t_{m-1}$. Отметим, что между соседними поворотными точками флуктуации ряда Dx_t имеют одинаковый знак и, следовательно, образуют одну фазу. Таким образом, в данном слу-

чае поворотные точки ряда x_i разделяют исходный ряд на m фаз. Каждая фаза характеризуется своей мощностью в соответствии с данным выше определением.

Теперь вспомним, что экономист, как правило, имеет дело с временным рядом, содержащим случайные флуктуации. В исходном ряде каждая такая флуктуация или несколько соседних могут образовать фазу, не имеющую, однако, какого-либо существенного содержания. Требуется очистить ряд от случайных флуктуаций и соответствующих им фаз и получить интерпретируемые движения ряда.

Примем, что фазы, образованные случайными флуктуациями, маломощны. При независимости случайных флуктуаций это предположение вполне приемлемо. Исключить маломощные фазы предлагается с помощью следующей итеративной процедуры.

Этап 1. Задаем критерием остановки итерационного процесса. В качестве такового могут выступать:

а) уровень теряемой мощности v — процент совокупной мощности ряда, которым мы решаемся пожертвовать в процессе агрегирования фаз путем подавления случайных флуктуаций;

б) наперед заданное число фаз, на которое желательно разбить весь выборочный период. Процесс объединения (агрегирования) фаз останавливается при достижении этого числа;

в) средняя длительность фаз. Если, например, ставится задача изучения приблизительно 4-летних циклов, то процесс целесообразно остановить, когда средняя длительность фаз будет близка 2;

г) экспертное решение, принимаемое на основе анализа графиков и числовых таблиц.

Нетрудно заметить, что критерии б и в тесно связаны между собой, так как средняя длительность фаз равна объему выборки N , деленному на число образованных фаз.

Обычно трудно априори выбрать критерий остановки и задать такое его значение, которое приведет к наилучшему исключению относительно коротких, маломощных фаз, к удовлетворительному выделению динамики процесса, подлежащей исследованию. В связи с этим целесообразно выполнять указанную итерационную процедуру до образования всего лишь нескольких фаз (заведомо пройдя необходимый

уровень деления на фазы). После этого для каждой итерации слияния фаз следует вычислить значение критериев а, б, в и сравнить их изменения при переходе от одной итерации к другой. Полезно процесс образования фаз контролировать на графиках.

В конечном счете решение о необходимом уровне агрегирования фаз остается за исследователем, а рассмотренные критерии представляют собой вид информации, помогающей квалифицированно принять это решение.

Важно отметить, что роль числовых критериев остановки возрастает при сопоставлении нерегулярных колебаний в двух или нескольких временных рядах, так как тогда требуется идентичный критерий для всех сравниваемых рядов.

Выбор типа критерия, очевидно, зависит и от конкретно поставленной цели исследования, и от стохастических свойств отобранных временных рядов. Одни ряды могут содержать более мощные «шумовые» колебания, у других — их удельный вес невелик.

Этап 2. Находим наименее мощную фазу, например j . Возможны два варианта ее расположения в выборке: а — в начале или в конце; б — внутри. Рассмотрим второй вариант. Принимаем решение об устранении соответствующих ей флуктуаций. Для этого заменяем фазы $j-1$, j , $j+1$ одной, мощность которой равна $p_{j-1} - p_j + p_{j+1}$. В ходе этой операции число фаз уменьшается на 2, а совокупная мощность временного ряда сокращается на $2p_j$. Последнее легко увидеть, так как в сумму (10.3) вместо $p_{j-1} + p_j + p_{j+1}$ теперь будет входить $p_{j-1} - p_j + p_{j+1}$. Коэффициент 2 можно объяснить также тем, что при элиминировании колебания устраняется его движение как в положительном, так и в отрицательном направлении. Таким образом, $2p_j$ — мощность, теряемая в процессе агрегирования фаз в случае б. Случай а отличается от б лишь тем, что число фаз в результате элиминирования наименее мощной краевой фазы уменьшается на одну, а не на две, так как объединению подлежат не три, а две соседние фазы (j и $j+1$, если j — первая фаза; $j-1$ и j , если j — последняя фаза).

Этап 3. Этап 2 будем повторять до тех пор, пока не достигнем заданного критерия остановки (см. этап 1). Тогда процесс агрегирования прекращается и полученные фазы подлежат содержательному или статистическому анализу.

Этап 4. В результате применения изложенной процедуры весь интервал выборки ряда будет разбит точками смены фаз на отрезки, внутри которых можно представить движение ряда упрощенно линейным образом. Для этого достаточно разделить мощность фазы на ее длительность — получим среднюю мощность флуктуаций внутри данной фазы, а ее знак определит знак флуктуаций. Интерес может представлять наложение графика агрегированных фаз (фазовой диаграммы) на график исходного ряда.

Возможна аппроксимация каждой фазы синусоидальной полуволной равной мощности и соответствующего знака. Нетрудно установить, что мощность p_j синусоидальной полуволны связана с ее амплитудой A_j и периодом колебаний T_j следующим соотношением (для простоты будем вести отсчет времени внутри фазы j от нуля до n_j)

$$p_j = \left| \int_0^{n_j} A_j \sin(\omega_j t + \varphi_j) dt \right| = \frac{2A_j n_j}{\pi} = \frac{A_j T_j}{\pi}, \quad (10.5)$$

где n_j — длительность фазы j ;

T_j — период колебаний синусоиды;

$T_j = 2n_j$;

ω_j — круговая частота;

$\omega_j = 2\pi/T_j$;

$$\varphi_j = \begin{cases} 0, & \text{если фаза положительна,} \\ \pi & \text{— в противном случае.} \end{cases}$$

Поскольку речь идет об аппроксимации дискретного временного ряда, синусоидальную полуволну необходимо заменить на n_j равноотстоящих дискретных значений (флуктуаций), суммарная мощность которых равна мощности фазы. В соответствии с этим представим (10.5) в виде суммы

$$\left| \int_0^{n_j} A_j \sin(\omega_j t + \varphi_j) dt \right| = \left| \int_0^1 + \int_1^2 + \dots + \int_{n_j-1}^{n_j} \right|. \quad (10.6)$$

Интегралы вычисляются по формуле

$$\int_a^b A_j \sin(\omega_j t + \varphi_j) dt = \frac{A_j n_j}{\pi} [-\cos(\omega_j b + \varphi_j) + \cos(\omega_j a + \varphi_j)] \quad (10.7)$$

и каждый из них характеризует величину и знак колебания в момент b .

Таким образом, на графике каждая фаза будет представлена синусоидальной полуволной со своей амплитудой и периодом колебаний. Исключение составляют лишь первая и последняя фазы. Их аппроксимировать синусоидой неправомерно, так как неизвестно, когда началась первая и когда закончилась последняя фаза. Поэтому их целесообразно оставить такими, какие они есть, вычислив лишь уровень их средней мощности.

Отметим, что для аппроксимации фазы гармоникой длительность этой фазы, естественно, должна быть хотя бы не меньше трех, в противном случае гармоника вырождается.

§ 3. ФАЗОВЫЙ АНАЛИЗ ИНВЕСТИЦИОННЫХ ЦИКЛОВ В США И ЗАПАДНОЙ ЕВРОПЕ

Рассмотрим ряд валовых капитальных вложений в экономику США в постоянных ценах (в долларах 1985 г.) с 1960 по 1988 г. (см. Приложение № 4, табл. V). Обозначим этот ряд k_t . Предварительный анализ обнаруживает экспоненциальный тренд, поэтому прежде чем приступить к фазовому анализу, выполним несколько несложных преобразований исходного ряда, чтобы элиминировать тренд и выделить скрытый в нем колебательный процесс в явном виде. Для этого перейдем от ряда k_t к его логарифмам $\ln k_t$. Логарифм k_t имеет четко выраженную тенденцию линейного роста. Перейдем к первым разностям логарифмов

$$x_t = \Delta \ln k_t = \ln k_t - \ln k_{t-1} = \ln(k_t / k_{t-1}). \quad (10.8)$$

Смысл x_t нетрудно установить. Из (10.8) следует, что

$$k_t = k_{t-1} \exp(x_t) \approx k_{t-1}(1 + x_t). \quad (10.9)$$

Таким образом, x_t — это доля, на которую возрастают валовые капитальные вложения в году t , т.е. x_t — темп прироста. На графике (здесь не приводится) видно, что x_t имеет нерегулярные циклические колебания. Динамику x_t можно исследовать, изучая либо отклонения x_t от среднего

$m_x = \frac{1}{N} \sum x_t$, либо первые разности $x_t - x_{t-1}$. Среднее m_x здесь выступает в роли некоторого стандарта, от которого измеряются отклонения вверх и вниз, и представляет собой средний темп прироста. А величина $x_t - x_{t-1}$ характеризует увеличение темпов прироста, если она положительна, или их снижение — в противном случае. Правомерны постановка и решение задачи в обоих вариантах. Остановимся на первом.

Итак, считаем флуктуацией

$$Dx_t = x_t - m_x. \quad (10.10)$$

Далее последовательно проводим фазовый анализ по изложенной схеме. За критерий остановки возьмем число фаз, равное 7, но попутно будем вести расчет и других критериев остановки (табл. 10.1). На итерации V процесс заканчивается. На рис. 10.1 показаны ряд $x_t - m_x$ и его фазовая диаграмма, полученная путем усреднения мощности флуктуаций внутри каждой найденной фазы.

Результаты аналогичных исследований, выполненных для Западной Европы, представлены в табл. 10.2 и на рис. 10.2. Обобщающие характеристики обоих примеров — в табл. 10.3.

Для сравнения колебательных процессов в валовых капитальных вложениях США и Западной Европы фазовые диаграммы (рис. 10.1 и 10.2) наложены одна на другую (рис. 10.3). Нетрудно заметить, что динамика валовых инвестиций в Западной Европе повторяет динамику валовых капитальных вложений в США с некоторым опозданием, причем величина лага переменная и постепенно сокращается с четырех лет в начале выборочного периода до года в его конце. Явно повторяется даже величина средней мощности фаз (амплитуда фазовой диаграммы).

Таблица 10.1
Агрегирование фаз — фильтрация маломощных флуктуаций на данных США

Флуктуации		Номер фазы	Знаки и мощности фаз					
Год, t	$x_t - m_x$		перво-начальных	на итерации				
				I	II	III	IV	V
1961	-0,0145	1	-0,0145(1)	-0,0145(1)	0,0780(6)	0,0780(6)	0,0780(6)	0,0780(6)
1962	0,0300	2	0,0925(5)	0,0925(5)	-0,0685(4)	-0,0685(4)	-0,0685(4)	-0,0685(4)
1963	0,0112	3	-0,0211(1)	-0,0685(4)	0,0666(3)	0,0666(3)	0,0666(3)	0,0666(3)
1964	0,0085	4	0,0117(1)	0,0666(3)	-0,1382(2)	-0,1382(2)	-0,1382(2)	-0,1382(2)
1965	0,0344	5	-0,0591(2)	-0,1382(2)	0,1101(3)	0,1101(3)	0,1101(3)	0,1101(3)
1966	0,0084	6	0,0666(3)	0,1101(3)	-0,0831(2)	-0,0831(2)	-0,0831(2)	-0,1546(4)
1967	-0,0211	7	-0,1382(2)	-0,0831(2)	0,0187(1)	0,0187(1)	0,0187(1)	0,1067(6)
1968	0,0117	8	0,1101(3)	0,0187(1)	-0,0902(1)	-0,0902(1)	-0,0902(1)	
1969	-0,0050	9	-0,0831(2)	-0,0902(1)	0,1222(2)	0,1222(2)	0,1222(2)	
1970	-0,0541	10	0,0187(1)	0,1222(2)	-0,0155(4)	-0,0155(4)		
1971	0,0269	11	-0,0902(1)	-0,0155(4)				
1972	0,0166	12	0,1222(2)					
1973	0,0231	13	0,0258(2)					
1974	-0,0445	14	0,0102(2)					

Флуктуации		Номер фазы	Знаки и мощности фаз				
Год, t	$x_t - m_x$		первоначаль- ных	на итерации			
			I	II	III	IV	V
1975	-0,0937						
1976	0,0458						
1977	0,0398						
1978	0,0246						
1979	-0,0226						
1980	-0,0605						
1981	0,0187						
1982	-0,0902						
1983	0,0216						
1984	0,1006						
1985	-0,0127						
1986	-0,0130						
1987	0,0088						
1988	0,0015						

Примечание. В скобках указана длительность фазы; подчеркнуты фазы с минимальной мощностью.

Агрегирование фаз — фильтрация маломощных флуктуаций на данных Западной Европы

Флуктуации		Номер фазы	Знаки и мощности фаз				
Год, t	$x_t - m_x$		первоначаль- ных	на итерации			
			I	II	III	IV	V
1961	0,0098	1	0,0113(2)	0,0113(2)	-0,0003(3)	0,1313(10)	0,1313(10)
1962	0,0015	2	-0,0115(1)	-0,0115(1)	0,1316(7)	-0,0445(2)	-0,0445(2)
1963	-0,0115	3	0,0590(3)	0,1316(7)	-0,0445(2)	0,0443(2)	0,0443(2)
1964	0,0514	4	-0,0049(1)	-0,0445(2)	0,0443(2)	-0,0949(1)	-0,1394(4)
1965	0,0060	5	0,0775(3)	0,0443(2)	-0,0949(1)	0,0384(1)	0,0692(2)
1966	0,0016	6	-0,0445(2)	-0,0949(1)	0,0384(1)	-0,0830(2)	-0,1314(3)
1967	-0,0049	7	0,0443(2)	0,0384(1)	-0,0830(2)	0,0692(2)	0,0705(5)
1968	0,0072	8	-0,0949(1)	-0,0830(2)	0,0692(2)	-0,1130(1)	-0,1314(3)
1969	0,0384	9	0,0384(1)	0,0692(2)	-0,1130(1)	0,0225(1)	0,0705(5)
1970	0,0319	10	-0,0830(2)	-0,1130(1)	0,0225(1)	-0,0408(1)	
1971	-0,0302	11	0,0692(2)	0,0225(1)	-0,0408(1)	0,0705(5)	
1972	-0,0142	12	-0,1130(1)	-0,0408(1)	0,0705(5)		
1973	0,0316	13	0,0225(1)	0,0705(5)			
1974	0,0128	14	-0,0408(1)				

Флуктуации		Номер фазы	Знаки и мощности фаз					
Год, t	$x'_t - m_x$		первоначальных	на итерации				
			I	II	III	IV	V	VI
1975	-0,0949	15	0,0705(5)					
1976	0,0384	16	0,0167(1)					
1977	-0,0386	17	<u>-0,0001(1)</u>					
1978	-0,0444		0,0539(3)					
1979	0,0544							
1980	0,0148							
1981	-0,1130							
1982	0,0225							
1983	-0,0408							
1984	0,0167							
1985	-0,0001							
1986	0,0192							
1987	0,0114							
1988	0,0233							

Примечание. В скобках указана длительность фазы; подчеркнуты фазы с минимальной мощностью.

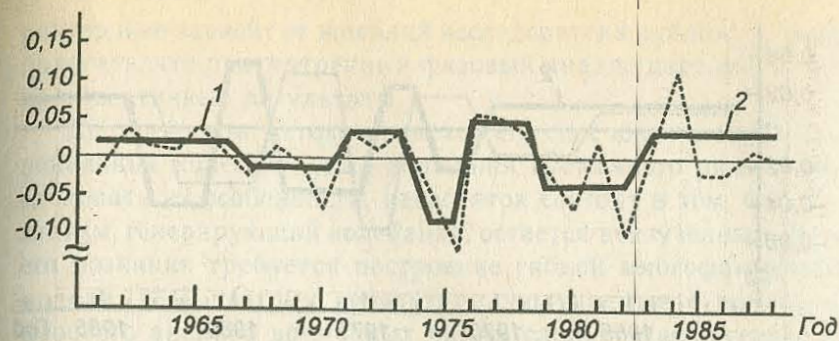


Рис. 10.1. Флуктуации (1) и фазовая диаграмма (2) темпов прироста валовых капитальных вложений в США с 1961 по 1988 г.

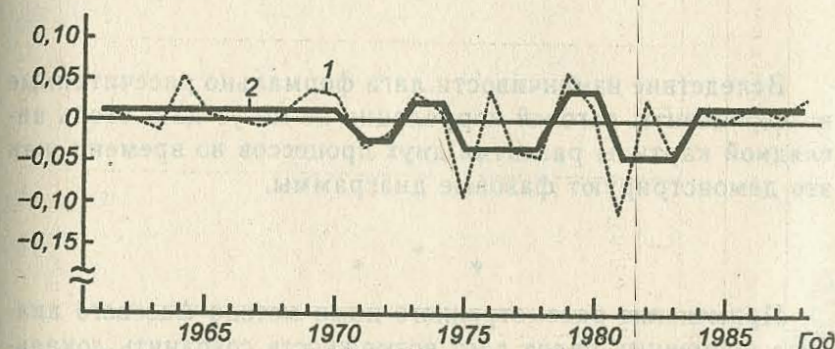


Рис. 10.2. Флуктуации (1) и фазовая диаграмма (2) темпов прироста валовых капитальных вложений в Западной Европе с 1961 по 1988 г.

Таблица 10.3

Обобщающие характеристики разбиения на фазы

Регион	Мощность ряда $x_t - m_x$	Число фаз	Средняя длительность фазы	Потери мощности, %	Число итераций
США	0,8640	7	4	16,35	5
Западная Европа	0,7854	7	4	9,93	6

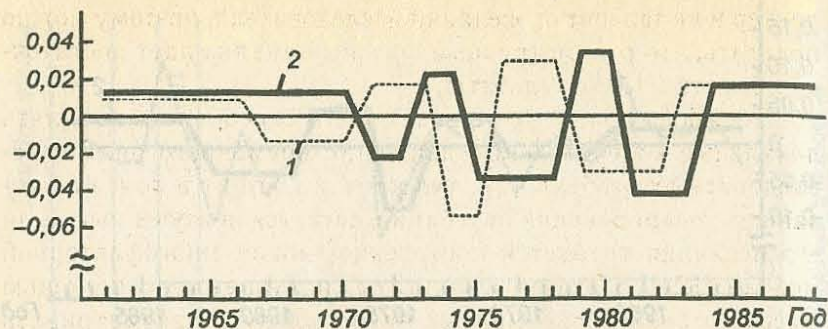


Рис. 10.3. Фазовые диаграммы темпов прироста валовых капитальных вложений в США (1) и в Западной Европе (2) с 1961 по 1988 г.

Вследствие изменчивости лага формально рассчитанные коэффициенты лаговой корреляции не могут дать столь наглядной картины развития двух процессов во времени, как это демонстрируют фазовые диаграммы.

* * *

Применение рассмотренного нами метода фазового анализа временных рядов дает возможность сохранить локальную индивидуальность нерегулярных циклических колебаний и последовательно переходить к анализу все более длинных и мощных волн, абстрагируясь от колебаний меньшего масштаба. Метод отличается простотой и ясностью концепции.

Анализ проводится во времени, что облегчает интерпретацию получаемых результатов, позволяет сопоставить их с другими событиями окружающего мира. Обработка ряда возможна с помощью универсальных электронных табличных процессоров типа Excel, Lotus 1-2-3 и т.п.

От исследователя требуется подготовка задачи к фазовому анализу, включая предварительное преобразование исходного ряда, определение уровня отсчета флуктуаций, задание критерия останова итеративного процесса агрегирования фаз. В остальном алгоритм носит жесткий, однозначный ха-

рактер и не зависит от желаний исследователя, поэтому можно полагать, что рассмотренный фазовый анализ дает достаточно объективные результаты.

Достоинством метода является его способность выделять локальные колебательные движения временного ряда и исследовать их особенности, недостаток состоит в том, что механизм, генерирующий колебания, остается неизученным. Для его познания требуется построение гибкой многофакторной модели [124], поэтому результаты, получаемые с помощью фазового анализа временных рядов, следует рассматривать как исходные для более глубокого изучения колебательных процессов.

Глава 11

АДАПТИВНАЯ ГИСТОГРАММА, ПРОБЛЕМА ОПТИМИЗАЦИИ

Рассматриваются критерий и метод оптимизации адаптивной процедуры последовательного обновления гистограммы временного ряда, изложенной в гл. 2 §1. Приводится пример использования адаптивной гистограммы для анализа динамики распределения кассовых остатков в банковском учреждении.

§ 1. ПОСТАНОВКА ПРОБЛЕМЫ

В анализе кассовых остатков банковского учреждения, как и во многих других экономических задачах, часто требуется определить форму закона распределения вероятностей случайной переменной. На практике для этих целей используется построенная на выборочных данных достаточного объема гистограмма, показывающая эмпирический закон распределения.

Так, если процесс, представляемый переменной, стационарен, то такое решение задачи следует признать вполне удовлетворительным. Однако гораздо чаще стохастический процесс, порождающий переменную, является нестационарным — изменяются его средний уровень, размах отклонений от среднего, а также форма кривой распределения амплитуд колебаний. Если эти изменения носят эволюционный характер, то, как показал Р.Г. Браун [47], может быть построена адаптивная процедура обновления гистограммы, основывающаяся на экспоненциальном сглаживании нуля или единицы. Эта процедура рассмотрена нами в гл. 2 § 1. Обновлен-

ные значения частот на момент t для разных диапазонов значений исследуемой переменной и определяют гистограмму, приближенно отражающую закон распределения.

Нерешенным, однако, остался вопрос о критерии качества этой процедуры и об оптимальном значении параметра адаптации. Решение данного вопроса изложено в [123].

§ 2. АДАПТИВНАЯ ПРОЦЕДУРА ОБНОВЛЕНИЯ ГИСТОГРАММЫ

Напомним, что в общем случае рассматривается полная система n несовместных событий, определенных на числовой оси с помощью $n+1$ границы

$$X_0 < X_1 < X_2 < \dots < X_n. \quad (11.1)$$

Событию, связанному с наблюдением x_t в момент t , соответствует номер интервала k — такой, что $X_{k-1} < x_t \leq X_k$, в который попадает наблюдаемое значение. Предполагаем, что моменты наблюдения разделены равными промежутками времени и $t = 1, 2, \dots, T$. Оценку на момент t вероятности наступления события k будем обозначать как

$$p_k(t) = P(X_{k-1} < x_t \leq X_k). \quad (11.2)$$

Разбиение оси x на интервалы произведем, разделив весь диапазон $(X_{\max} - X_{\min})$ между верхним и нижним значением x на n равных частей. В общем случае границы могут быть заданы произвольно в зависимости от характера задачи и целей исследования.

Грубые начальные оценки $\hat{p}_k(0)$, $k = 1, 2, \dots, n$, можно получить, например, используя прошлые данные или начальную часть выборки объемом $T_1 \leq T$ как частоту

$$\hat{p}_k(0) = \frac{\omega_k}{T_1}, \quad (11.3)$$

где ω_k — число точек из подвыборки $\{x_t\}$, $t=1, 2, \dots, T_1$, попавших в интервал $X_{k-1} < x_t \leq X_k$.

Далее построим n -мерный вектор-столбец $u(t)$, который имеет $(n-1)$ -нулевую компоненту, а k -я компонента равна единице. Предыдущие оценки n вероятностей можно рассматривать как n -компонентный вектор-столбец $\hat{p}(t-1)$. Процесс пересмотра этих оценок с учетом текущей информации является экспоненциальным сглаживанием вектора $u(t)$ по формуле

$$\hat{p}(t) = \alpha u(t) + (1-\alpha)\hat{p}(t-1), \quad (11.4)$$

где α — константа из интервала $0 < \alpha < 1$, параметр адаптации, определяющий скорость обновления оценок. Каждая компонента вектора модифицируется экспоненциальным сглаживанием нуля или единицы.

Так как $\hat{p}(t-1)$ — вектор вероятностей, то все его компоненты неотрицательны и их сумма должна быть точно равна единице. Экспоненциальное сглаживание не может сделать компоненту отрицательной, и сумма полученных новых оценок компонент равна предыдущей сумме, т.е. единице:

$$\sum_{k=1}^n \hat{p}_k(t) = \alpha + (1-\alpha) \sum_{k=1}^n \hat{p}_k(t-1) = 1, \quad (11.5)$$

$$\text{так как } \sum_{k=1}^n \hat{p}_k(t-1) = 1.$$

Таким образом, экспоненциальное сглаживание вектора вероятностей дает обновленный вектор вероятностей.

§ 3. ПРОБЛЕМЫ ОПТИМАЛЬНОСТИ АДАПТИВНОЙ ПРОЦЕДУРЫ ОБНОВЛЕНИЯ ГИСТОГРАММЫ

Рассмотрим событие i . Если закон распределения наблюдаемых значений x_i не меняется, то математическое ожидание значения компоненты i вектора $u(t)$, подлежащего сглаживанию, точно равно действительной вероятности p_i наступления события i и математическое ожидание оценки равно действительной вероятности:

$$\begin{aligned} M[\hat{p}_i(t)] &= \sum_{j=t}^{\infty} \alpha(1-\alpha)^{j-t} M[u_i(j)] = \\ &= p_i \sum_{j=t}^{\infty} \alpha(1-\alpha)^{j-t} = p_i. \end{aligned} \quad (11.6)$$

Вероятность того, что придется сглаживать единицу, равна p_i , а того, что сглаживать нуль, составляет $1-p_i$. Легко подсчитать, что компонента i вектора $u(t)$ характеризуется дисперсией

$$D[u_i(t)] = (1-p_i)^2 p_i + (0-p_i)^2 (1-p_i) = p_i(1-p_i). \quad (11.7)$$

Если к компоненте u_i применить процедуру экспоненциального сглаживания, то дисперсия получаемой на выходе экспоненциальной средней выражается через дисперсию входного потока по известной формуле

$$\sigma_i^2 = \frac{\alpha}{2-\alpha} D[u_i(t)], \quad (11.8)$$

где α — постоянная сглаживания.

После подстановки (11.7) получим дисперсию оценки вероятности наступления события i :

$$\sigma_i^2 = \frac{\alpha}{2-\alpha} p_i(1-p_i). \quad (11.9)$$

Из (11.9) Р.Г. Браун делает вывод, что имеются два наиболее предпочтительных способа конструирования системы границ $X_0, X_1, X_2, \dots, X_n$, при которой дисперсия оценки сравнительно невелика. Он предлагает устанавливать такие границы событий, чтобы p_i была либо очень большой (близкой к единице), либо очень маленькой (почти равной нулю). Это обеспечит малое значение произведения $p_i(1-p_i)$ и, следовательно, низкую дисперсию оценок компонент вектора вероятностей. Максимум дисперсии достигается при $p_i = 0,5$.

Однако подобные рассуждения слишком формальны, а полученные рекомендации не всегда реализуемы. Если, например, положить, что $p_i \approx 0,9$, то на все остальные события остается лишь 10% вероятности. Такое разбиение на события, конечно, возможно, если речь идет, например, о разделении стандартных и нестандартных ситуаций. Но при адаптивном анализе распределения вероятностей потребуется построить границы, разделяющие числовую ось на 5–9 или более событий. Тогда условие $p_i \approx 1$ становится неприемлемым по очевидным причинам. В этом случае целесообразно потребовать, чтобы все оценки вероятностей имели равные дисперсии $\sigma_i^2 = \text{const}$ или одинаковое отношение стандартной ошибки к оценке вероятности $\sigma_i / p_i = \text{const}$.

В первом случае выдвинутое требование означает, что

$$\sigma_i^2 = \frac{\alpha}{2-\alpha} p_i(1-p_i) = \text{const}, \quad i=1,2,\dots,n \quad (11.10)$$

или

$$p_i(1-p_i) = \text{const}, \quad i=1,2,\dots,n. \quad (11.11)$$

Несмотря на то что здесь формально можно получить два решения, годится лишь одно — то, которое меньше 0,5, поскольку таких p_i должно быть 5–9 или более.

Во втором случае

$$\frac{\alpha}{2-\alpha} p_i(1-p_i) / p_i^2 = \text{const}, \quad i=1,2,\dots,n \quad (11.12)$$

или

$$(1-p_i) / p_i = \text{const}, \quad i=1,2,\dots,n, \quad (11.13)$$

откуда

$$p_i = 1/(1+\text{const}) = \text{const}, \quad i=1,2,\dots,n. \quad (11.14)$$

Таким образом, в обоих случаях приходим к одному и тому же выводу, что для равенства дисперсий оценок σ_i^2 или коэффициентов вариации σ_i / p_i границы событий надо выбирать так, чтобы вероятности наступления этих событий были одинаковыми.

Весьма важный вопрос, оставленный Р.Г. Брауном без ответа, касается выбора оптимального значения параметра адаптации α . Формула (11.10) лишь свидетельствует о том, что чем ближе α к нулю, тем меньше дисперсия оценки вероятности. Но нас в конечном счете интересует не столько ее колеблемость, сколько степень соответствия последовательно формируемого и корректируемого распределения (точнее, гистограммы) характеру реального разброса значений исследуемой величины, по предположению, изменяющемуся во времени. Совершенно очевидно, что вопрос об оптимальном значении параметра α может быть решен только при наличии критерия качества рассматриваемой процедуры.

За основу такого критерия в [123] предлагается взять максимальный уровень значимости, при котором полученную в момент t фактическую точку можно отнести к построенному на этот же момент распределению. Сумма (или нормированная сумма) таких максимальных уровней значимости по моментам времени $t = 1, 2, \dots, T$ даст значение итогового критерия Q , который будет функцией параметра адаптации α и начальных оценок вероятностей событий на момент $t = 0$. Максимальная величина Q определяет оптимальное значение α и начальных оценок вероятностей.

Поясним этот критерий на графике (рис. 11.1). Пусть на момент t адаптивно построена функция плотности распределения вероятностей $\varphi(x, \alpha, t)$, которая, как ожидается, описывает стохастический процесс x в момент t . Полагаем, что

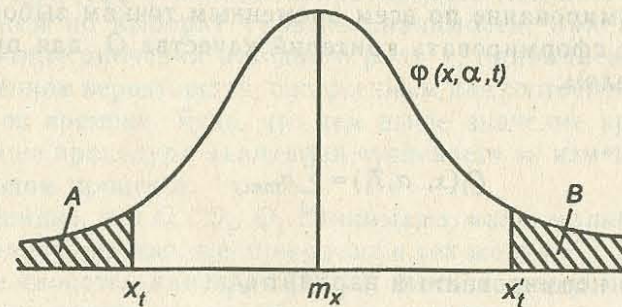


Рис. 11.1. Графическая интерпретация критерия качества адаптивной гистограммы $\varphi(x, \alpha, t)$ — оценка плотности распределения x в момент t ; α — параметр адаптации

в момент t мы получаем значение x_t (реализацию) исследуемой величины x . Тогда заштрихованная область A под кривой распределения равна вероятности того, что x примет значение меньше x_t , и представляет собой максимальный уровень значимости $q_{\max,t}$, при котором x_t еще можно считать принадлежащей генеральной совокупности с плотностью распределения $\varphi(x, \alpha, t)$.

Если бы наблюдаемое значение превосходило медиану m_x и равнялось x'_t , то максимальный уровень значимости определялся бы заштрихованной областью B .

Изложенное можно записать так:

$$q_{\max,t} = \begin{cases} \Pr(x < x_t) = \Phi(x, \alpha, t), & x_t \leq m_x, \\ \Pr(x > x_t) = 1 - \Phi(x, \alpha, t), & x_t > m_x, \end{cases} \quad (11.15)$$

где Pr — вероятность;

m_x — медиана распределения x ;

$\Phi(x, \alpha, t)$ — интегральная функция распределения x ,

$$\Phi(x, \alpha, t) = \int_{-\infty}^x \varphi(x, \alpha, t) dx. \quad (11.16)$$

Сама динамика $q_{\max,t}$ может показать, в какие периоды соответствие гистограмм и реальных наблюдений было «лучше», а в какие — «хуже».

Суммирование по всем временным точкам выборки позволяет сформировать критерий качества Q_1 для процедуры в целом:

$$Q_1(x, \alpha, T_1) = \sum_{t=1}^T q_{\max,t} \quad (11.17)$$

или его нормированный вариант Q_2 :

$$Q_2(x, \alpha, T_1) = \frac{1}{T} \sum_{t=1}^T q_{\max,t}. \quad (11.18)$$

Критерий Q_1 является агрегатом, характеризующим степень соответствия наблюдений временного ряда и ряда построенных гистограмм на всем выборочном интервале в целом. Критерий Q_2 интерпретируется как усредненное по выборке максимальное значение уровня значимости, при котором наблюдаемые значения ряда можно считать принадлежащими распределениям, представляемым соответствующими гистограммами.

В формулах (11.17) и (11.18) показано, что в обоих случаях критерий качества процедуры адаптации является функцией не только данных $\{x_t\}$ и параметра адаптации α , но и объема выборочных данных T , используемых для определения начальных значений вероятностей событий, необходимых для запуска рекуррентной процедуры.

Отметим также, что Q_1 и Q_2 сформированы как односторонние критерии, предполагающие, что сопоставление реализации процесса x_t с построенным распределением производится на одном его «хвосте». Однако на самом деле фактические точки случайным образом ложатся по обе стороны от медианы, и поэтому целесообразно перейти к двустороннему критерию, для которого все значения q нужно лишь удвоить. Тогда получим критерий Q_3 :

$$Q_3(x, \alpha, T_1) = \frac{2}{T} \sum_{t=1}^T q_{\max,t}. \quad (11.19)$$

Критерий Q_3 можно интерпретировать как достигнутый в среднем по выборке уровень значимости, при котором выборочные значения исходного ряда x_t принадлежат распределениям вероятностей, построенным для соответствующих моментов времени. Ясно, что чем выше значение критерия, тем лучше процедура адаптации «успевает» за изменениями в реальном процессе.

Очевидно, что Q_1 , Q_2 , Q_3 принимают максимальное значение одновременно, т.е. при одних и тех же α и T . Замечательное свойство критерия (11.19) состоит в том, что если, начиная с какого-то момента времени, переменная x_t принимает постоянное значение, равное медиане, то все последующие $q_{\max,t}$ равны 0,5, $Q_3 \rightarrow 1$, а распределение постепенно стягивается в точку. В общем случае значение критерия

находится в пределах $0 < Q_3 < 1$, и это делает его удобным для различных сопоставлений.

Следует отметить, что при использовании построенной гистограммы для выработки вероятностных суждений могут применяться как односторонний, так и двусторонний критерии в зависимости от задач, и результаты, конечно, будут различные.

Имея критерий качества процедуры, нетрудно определить и оптимальное значение параметра адаптации α . Для этого надо на интервале $0 < \alpha < 1$, например, методом проб найти такое его значение, которое максимизирует Q . Точно так же можно найти и оптимальное значение T_1 — начальную часть выборки, используемую для вычисления грубых оценок вероятностей $\hat{p}_k(0)$, необходимых для первого применения рекуррентных формул экспоненциального сглаживания.

Конечно, оптимизация критерия Q по двум параметрам α и T_1 несколько усложняет вычислительный процесс. Но, во-первых, область определения обоих параметров ограничена: $0 < \alpha < 1$, $1 \leq T_1 \leq T$. Во-вторых, применение сначала грубой сетки значений для локализации области поиска, а затем более тонкой помогают резко снизить объем вычислений. В-третьих, быстроедействие современных компьютеров делает проблему легко преодолимой.

§ 4. АДАПТИВНЫЙ АНАЛИЗ РАСПРЕДЕЛЕНИЯ КАССОВЫХ ОСТАТКОВ

Необходимость анализа нестационарных распределений часто возникает в экономических исследованиях. Например, в банковской деятельности это может быть задача изучения распределения сумм на корреспондентских счетах или в более общей постановке — анализ и использование вероятностных закономерностей формирования кассовых остатков.

В качестве примера рассмотрим задачу о кассовых остатках. Предположим, что имеются ежедневные данные об остатках наличных средств, являющихся результатом движения денег за день. Необходимо определить, какую сумму из этого остатка можно пустить в оборот, обеспечив при этом заданную вероятность незамедлительного удовлетворения запросов клиентов с учетом ожидаемых новых поступлений.

Для решения этой задачи нами была создана специальная программа, в основу которой положен описанный выше алгоритм. Предусмотрена оптимизация адаптивной процедуры как по параметру α , так и по начальному объему выборки T_1 . Результаты счета представляются на экране в табличном виде (рис. 11.2) и графически (рис. 11.3). На графике

АДАПТИВНЫЙ АНАЛИЗ КАССОВЫХ ОСТАТКОВ

Имя файла с данными? Cashres.dat, имя ряда: LOROinp, число точек 132

Таблица 1. Суммы, изымаемые из остатка, и вероятности обслуживания клиентов

Вероятность обслуживания, %	60 дней назад	40 дней назад	20 дней назад	10 дней назад	Прогноз
100	34,00	34,00	34,00	34,00	34,00
95	39,16	48,95	53,76	46,90	51,18
90	44,31	51,37	60,65	51,71	57,98
85	49,00	53,80	66,54	55,37	64,47
80	51,64	56,22	71,91	59,03	70,06
75	54,28	58,65	77,27	62,70	75,66
70	56,92	61,07	79,39	66,76	80,26
65	59,56	65,11	81,37	70,86	84,37
60	62,20	75,34	83,35	74,96	88,49
55	67,02	82,08	85,33	78,35	93,73
50	72,78	87,88	87,31	80,56	105,56

Таблица 2. Минимальное и максимальное значения остатков за последние дни

Значение остатков	За все дни	За 60 дней	За 40 дней	За 20 дней	За 10 дней
Минимальное	34,00	44,00	44,00	44,00	67,00
Максимальное	179,00	179,00	179,00	179,00	179,00

Таблица 3. Оптимальные значения параметров

Начальная выборка $T_1=79$; параметр адаптации $ALF=0,07$; критерий качества $Q_{MAX}=0,53898$

Рис. 11.2. Макет выдачи на экран результатов счета в виде трех таблиц

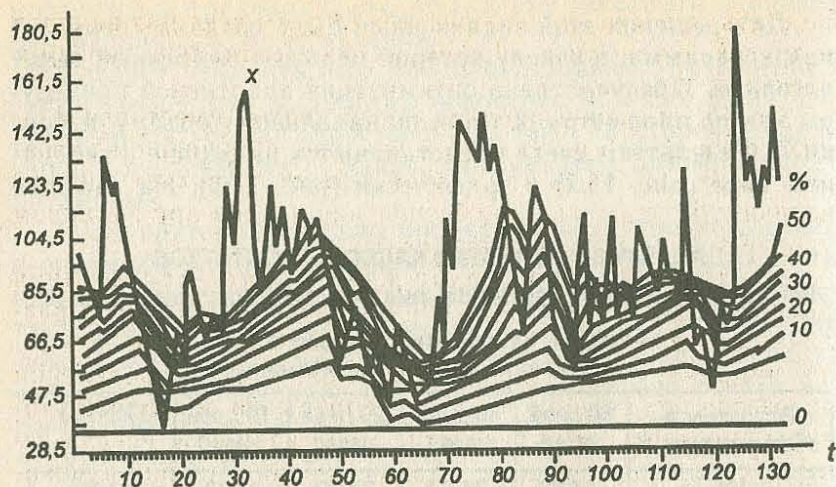


Рис. 11.3. Анализ кассовых остатков и вероятностей обслуживания клиентов: результаты счета в графическом виде (исходный ряд x и изоквантили гистограмм), t — рабочие дни

представлен исходный ряд со сложной неустойчивой динамикой и 11 кривых, соответствующих различным квантилям распределения (от 0 до 50% с шагом 5%).

Из полученных результатов следует, что при заданной вероятности обслуживания запросов клиентов без задержки, равной, например, 90%, можно часть кассовых остатков в размере 57,98 пустить в оборот. Отметим, что здесь для выработки вероятностного суждения применялся односторонний критерий.

* * *

Адаптивная гистограмма позволяет получать не только точечный и интервальный прогноз будущих значений временного ряда, но и прогноз в виде функции плотности распределений.

Вероятностная модель может предназначаться для изучения, например, распределения доходов в различных слоях общества, спроса на отдельные виды товаров по возрастным группам, эволюции распределения семейных бюджетов

по статьям расходов, распределения сумм текущих вкладов и других финансовых, экономических и социально-экономических процессов.

Особое значение предложенный подход имеет для анализа рисков, и, в частности, в анализе, получившем название Value at Risk (VAR). Такой анализ позволяет оценить величину максимального обесценения актива при заданном уровне значимости (вероятности).

Предложенная процедура адаптивного исследования распределений применима при определении размеров различных резервных и страховых фондов. С ее помощью можно, например, анализировать изменения в распределении депозитов по величине и сроку, непрерывно обновлять таблицы смертности, рождаемости, заболеваемости, используемые в страховой медицине, а также таблицы аварий и пожаров, используемые в страховании автомобилей, недвижимости и т.д.

Подобные задачи возможны и в «технологических» звеньях в связи с исследованием надежности и качества с учетом непрерывной модернизации технологий и модификации изделий и т.п.

Имеются резервы и для совершенствования рассмотренной методики. Так, в данном изложении мы полагали, что параметр адаптации постоянен. Между тем скорость деформации распределения со временем может существенно меняться. Тогда возникает задача автоматического управления параметром адаптации (см. гл. 4). Заметное влияние на эффективность процедуры могут оказывать также сильные тренды временного ряда. В этом случае процедура обновления гистограммы отстает от реальных изменений, что проявляется в снижении показателя качества Q . Такие вопросы требуют специального изучения в каждом отдельном случае.

Глава 12

КРИТЕРИИ ДИККИ-ФУЛЛЕРА ДЛЯ ИДЕНТИФИКАЦИИ ХАРАКТЕРА ТРЕНДА (ОБНАРУЖЕНИЕ ЕДИНИЧНЫХ КОРНЕЙ)

Экономические ряды очень часто характеризуются наличием временного тренда. Обнаружение трендов и их правильный учет в структуре модели представляют собой важную задачу в анализе временных рядов. Для решения этой задачи Дикки и Фуллер разработали критерии [137, 138], рассмотрению которых и посвящена данная глава.

§ 1. ДЕТЕРМИНИРОВАННЫЙ И СЛУЧАЙНЫЙ ХАРАКТЕР ТРЕНДА

В экономических исследованиях часто приходится иметь дело с нестационарными временными рядами, обнаруживающими тот или иной тренд. При построении регрессионной модели необходимо учитывать наличие этого тренда и исключать его. Известны две процедуры исключения тренда:

- 1) оценивание регрессии на время;
- 2) переход к последовательным разностям (приростам) переменной.

В регрессионном подходе предполагается, что ряд y_t генерируется моделью

$$y_t = f(t) + u_t,$$

где $f(t)$ — тренд;

u_t — стационарная случайная компонента с нулевым математическим ожиданием и дисперсией σ_u^2 .

Если предположить, что $f(t)$ является линейной функцией от времени, то

$$y_t = \alpha + \beta t + u_t. \quad (12.1)$$

После элиминирования тренда остается ряд ошибок метода наименьших квадратов (МНК) e_t со свойствами $\sum e_t = 0$ и $\sum t e_t = 0$.

Если для элиминирования тренда перейти к разностям, то получим

$$\Delta y_t = y_t - y_{t-1} = \beta + u_t - u_{t-1}.$$

Для элиминирования β нужно еще раз взять разности первого порядка

$$\Delta^2 y_t = u_t - 2u_{t-1} + u_{t-2} = \Delta^2 u_t.$$

Полученный ряд не содержит тренда. Однако, если предположить, что y_t генерирован моделью

$$y_t - y_{t-1} = \beta + \varepsilon_t, \quad (12.2)$$

где ε_t — стационарный ряд с нулевым математическим ожиданием и дисперсией σ^2 ; в этом случае первые разности оказываются стационарными с математическим ожиданием β . Эта модель называется *моделью случайного блуждания с дрейфом*.

Переходя от начального значения y_0 к y_t путем суммирования приращений, получим выражение

$$y_t = y_0 + \beta t + \sum_{i=1}^t \varepsilon_i, \quad (12.3)$$

имеющее тот же общий вид, что и уравнение (12.1), за исключением того, что член возмущения не является стационарным, так как его дисперсия $t\sigma^2$ растет во времени. К. Нельсон и К. Пlossер [153] назвали выражение (12.1) *моделью с детерминированным трендом* (trend-stationary processes — TSP), а уравнение (12.2) — *моделью стационарных приращений* (difference-stationary processes — DSP). Обе модели отражают линейный тренд, но различаются способом элиминирования тренда из исходных данных.

Для проверки гипотезы о том, что ряд генерирован моделью TSP против альтернативной, которая означает, что ряд описывается моделью DSP, К. Нельсон и К. Плоссер предложили применять критерий, разработанный Д. Дикки и В. Фуллером [137, 138]. Критерий предполагает оценку модели:

$$y_t = \alpha + \rho y_{t-1} + \beta t + \varepsilon_t. \quad (12.4)$$

Ряд y_t принадлежит к классу DSP, если $\rho = 1$, $\beta = 0$, и к классу TSP — если $\rho = 0$ или в более широком смысле, если $|\rho| < 1$.

Таким образом, нужно проверить нулевую гипотезу H_0 : $\rho = 1$, $\beta = 0$ против альтернативной H_1 : $|\rho| < 1$. Проблема, возникающая здесь, состоит в том, что нельзя использовать обычную теорию распределения оценок МНК, когда $\rho = 1$. Дикки и Фуллер показали, что оценка МНК параметра ρ при гипотезе о модели DSP (т.е., что $\rho = 1$) имеет математическое ожидание не 1, а значение, меньшее 1. Однако отрицательное смещение уменьшается с увеличением числа наблюдений. Авторы критерия получили критические значения для проверки гипотезы $\rho = 1$ против $|\rho| < 1$.

Процедура проверки гипотез относительно характера тренда является частным случаем более общей процедуры идентификации авторегрессионной зависимости, которую мы рассмотрим чуть позже, а сейчас обсудим последствия неверной спецификации модели.

§ 2. ЛОЖНЫЕ ТРЕНДЫ

Если в уравнении (12.2), представляющем модель DSP, $\beta = 0$

$$y_t - y_{t-1} = \beta + \varepsilon_t,$$

то получаем модель случайного блуждания без тренда (ее также называют *случайным блужданием с нулевым дрейфом*). Однако из представления этой модели в виде выражения (12.3)

$$y_t = y_0 + \beta t + \sum_{i=1}^t \varepsilon_i$$

видно, что даже если у среднего нет тренда, то он имеется у дисперсии.

Теперь предположим, что истинной моделью является DSP, т.е. уравнение (12.2) с $\beta = 0$, а оценивается ошибочно модель типа TSP (12.1):

$$y_t = \alpha + \beta t + u_t,$$

т.е. модель с линейным трендом среднего, но без тренда у дисперсии. Априори ясно, что возрастающая дисперсия случайной компоненты означает широкие движения u_t , которые будут приписаны движению среднего, и коэффициент β при переменной времени t окажется значимым, даже если в действительности никакого тренда у среднего нет. Каковы будут последствия ошибочной спецификации? К. Нельсон и Г. Канг [155] анализировали эту проблему и сделали следующие выводы:

1. Регрессия случайного блуждания на время, оцененная МНК, дает R^2 в районе 0,44 независимо от объема выборки, когда в действительности среднее значение переменной не связано с временем каким-либо образом.

2. В случае когда случайное блуждание имеет дрейф, т.е. $\beta \neq 0$, R^2 будет выше и будет увеличиваться с ростом объема выборки, достигая в пределе 1, независимо от значения β .

3. Остатки от регрессии на время, рассматриваемые как ряд с исключенным трендом, имеют в среднем только около 14% истинной дисперсии стохастической компоненты исходного ряда.

4. Остатки от регрессии на время автокоррелированы и коэффициент автокорреляции для лага 1 $r_1 \approx 1 - 10/n$, где n — объем выборки.

5. Обычные t -критерии для проверки значимости отдельных регрессоров более не работают. Имеется тенденция отвергать на их основе нулевую гипотезу о независимости от времени, когда она верна.

6. Регрессия одного случайного блуждания на другое с включением времени в правую часть уравнения с высокой вероятностью приводит к ложным зависимостям, т.е. обычная t -статистика часто указывает на связь между переменными, когда ее нет.

Основной вывод состоит в том, что использование регрессии на время имеет серьезные последствия, когда в действительности временной ряд относится к классу DSP и, следовательно, адекватной процедурой для исключения тренда является переход к разностям.

Теперь предположим, что истинной является модель линейного тренда, а мы перешли к разностям. Тогда ошибка имеет характер скользящей средней, и в худшем случае игнорирование этого приведет к неэффективным оценкам.

Например, пусть истинная модель

$$y_t = \alpha + \lambda x_t + \beta t + u_t,$$

где u_t — независимы, с нулевым средним и постоянной дисперсией. Если перейти к первым разностям, то получим

$$\Delta y_t = \lambda \Delta x_t + \beta + v_t,$$

где ошибка $v_t = \Delta u_t = u_t - u_{t-1}$ является скользящей средней и, следовательно, автокоррелированной. В этом случае оценивание МНК уравнения для первых разностей все еще дает состоятельные оценки. Таким образом, последствия перехода к разностям, когда этого не требуется, намного менее серьезны, чем те, что связаны с непереходом к разностям, когда это обусловлено свойствами ряда, т.е. когда истинная модель относится к классу DSP.

Нельсон и Пlossер применили критерий Дикки — Фуллера для проверки широкого круга исторических временных рядов, описывающих экономику США, и пришли к выводу, что во всех случаях следует принять гипотезу DSP, за исключением нормы безработицы. Они делают вывод, что для большинства экономических временных рядов модель DSP более адекватна, а модель TSP была бы подходящей только, если предположить, что ошибки u_t в (12.1) сильно автокоррелированы.

Таким образом, для учета тренда на практике лучше использовать переход к разностям и строить регрессии на их основе, чем регрессии для уровней с временем в качестве дополнительной объясняющей переменной.

А теперь рассмотрим подробнее процедуру Дикки — Фуллера.

§ 3. КРИТЕРИЙ ДИККИ — ФУЛЛЕРА ДЛЯ ОБНАРУЖЕНИЯ ЕДИНИЧНЫХ КОРНЕЙ

Задача проверки гипотезы $\rho = 1$ в авторегрессионном уравнении первого порядка

$$y_t = \alpha + \rho y_{t-1} + u_t \quad (12.5)$$

называется *проверкой наличия единичных корней*. Процесс y_t является стационарным, если $-1 < \rho < 1$. Если $\rho = 1$, то y_t — нестационарный и модель (12.5) представляет случайное блуждание с дрейфом. Если $\rho > 1$, ряд имеет нестационарный, взрывной характер. Поэтому гипотеза о стационарности ряда может быть проверена путем сравнения ρ с 1. Однако обычный критерий МНК — t -статистика здесь неприменим и требует корректировки. Рассмотрим этот вопрос подробнее.

Найдем дисперсию оценки МНК $\hat{\rho}$ для больших выборок в модели парной регрессии (12.5), воспользовавшись формулой

$$\text{var}(\hat{\rho}) = \frac{\sigma^2}{\sum x_i^2}, \quad (12.6)$$

где в качестве объясняющей переменной x_t выступает лаговое значение y_{t-1} . Последовательно раскрывая y_t через прошлые (лаговые) значения y , получаем

$$\begin{aligned} y_t &= \alpha + \rho y_{t-1} + u_t = \alpha + \rho(\alpha + \rho y_{t-2} + u_{t-1}) + u_t = \dots \\ &= \alpha + \alpha\rho + \alpha\rho^2 + \alpha\rho^3 + \dots + u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \dots \end{aligned}$$

Математическое ожидание:

$$M(y_t) = \alpha + \alpha\rho + \alpha\rho^2 + \alpha\rho^3 + \dots = \frac{\alpha}{1-\rho}.$$

Дисперсия:

$$\begin{aligned} \text{var}(y) &= M(y_t - M(y_t))^2 = M(u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \dots)^2 = \\ &= M(u_t^2 + \rho^2 u_{t-1}^2 + \rho^4 u_{t-2}^2 + \dots) = \frac{\sigma^2}{1-\rho^2}. \end{aligned}$$

Отсюда, после подстановки в (12.6), получаем

$$\text{var}(\hat{\rho}) = \frac{\sigma^2}{T \text{var}(y)} = \frac{\sigma^2(1-\rho^2)}{T\sigma^2} = \frac{1-\rho^2}{T},$$

где T — объем выборки.

Таким образом, $\text{var}(\hat{\rho}) = 0$ при нулевой гипотезе $H_0: \rho = 1$. Следовательно, для того чтобы можно было применять этот критерий, нужно получить предельное распределение $\hat{\rho}$ при нулевой гипотезе $H_0: \rho = 1$.

Одним из наиболее широко используемых критериев наличия единичных корней является критерий Дикки — Фуллера. Они начали с построения процедуры проверки гипотезы $H_0: \rho = 1$ в модели (12.5):

$$y_t = \alpha + \rho y_{t-1} + u_t.$$

Этот тест выполняется путем оценивания уравнения, получаемого после перехода к разностям, когда модель записывается как

$$\Delta y_t = \gamma y_{t-1} + u_t,$$

где $\gamma = \rho - 1$.

Проверка гипотезы $\rho = 1$ эквивалентна проверке гипотезы $\gamma = 0$. Затем авторы рассмотрели три различных регрессионных уравнения, в которых проверяется наличие единичного корня:

$$\Delta y_t = \gamma y_{t-1} + u_t. \quad (12.7)$$

$$\Delta y_t = \alpha + \gamma y_{t-1} + u_t. \quad (12.8)$$

$$\Delta y_t = \alpha + \gamma y_{t-1} + \beta t + u_t. \quad (12.9)$$

Отметим, что рассмотренная ранее модель (12.4) может быть приведена к виду (12.9) при $\gamma = \rho - 1$.

Различие в моделях (12.7), (12.8) и (12.9) состоит в наличии детерминированных элементов α и βt . Первая модель — это модель чисто случайного блуждания, во вторую модель

добавляется свободный член, который является параметром дрейфа, а в третью — включены и дрейф, и линейный временной тренд.

Во всех трех регрессиях интерес представляет параметр γ , если $\gamma = 0$, то ряд y_t характеризуется единичным корнем $\rho = 1$. Процедура проверки предполагает оценивание МНК одного или нескольких записанных выше уравнений с целью получения оценки γ и ее стандартной ошибки. Делением оценки на ее стандартную ошибку получают t -статистику, которая, однако, при нулевой гипотезе о наличии единичного корня имеет распределение, отличное от распределения Стьюдента. Авторы получили критические значения для t путем модельных статистических испытаний. Сравнение полученных t -статистик с критическим значением, приведенным в таблицах Дикки — Фуллера, позволяет исследователю принять или отвергнуть нулевую гипотезу $H_0: \gamma = 0$.

Пусть, например, оценивание модели

$$y_t = \rho y_{t-1} + u_t$$

дало следующие результаты: $\rho = 0,9546$ с $\sigma = 0,030$. Тогда оценка МНК уравнения в разностном виде

$$\Delta y_t = \gamma y_{t-1} + u_t$$

дает $\gamma = -0,0454$ с той же самой стандартной ошибкой 0,030. Следовательно, соответствующая t -статистика для гипотезы $\gamma = 0$ равна:

$$t = -0,0454 / 0,030 = -1,5133.$$

Для каждого из трех видов регрессионных уравнений (12.7), (12.8) и (12.9) методология остается одинаковой. Однако нужно иметь в виду, что критические значения t -статистик зависят от вида уравнения, от включения свободного члена или временного тренда. В своем исследовании методом Монте-Карло Дикки и Фуллер нашли критические значения для гипотезы $\gamma = 0$ в зависимости от вида регрессии и объема выборки. Эти статистики для трех видов регрессии обозначены как τ , τ_μ и τ_τ соответственно. В табл. 12.1

Критические значения статистик Дикки – Фуллера
для 95% и 99% доверительных интервалов

Модель	Гипотеза	Статистика критерия	Критические значения	
			95%	99%
$\Delta y_t = \alpha + \gamma y_{t-1} + \beta t + u_t$	$\gamma = 0$	τ_τ	-3,45	-4,04
	$\alpha = 0$ при $\gamma = 0$	$\tau_{\alpha\tau}$	3,11	3,78
	$\beta = 0$ при $\gamma = 0$	$\tau_{\beta\tau}$	2,79	3,53
	$\gamma = \beta = 0$	ϕ_3	6,49	8,73
	$\alpha = \gamma = \beta = 0$	ϕ_2	4,88	6,50
$\Delta y_t = \alpha + \gamma y_{t-1} + u_t$	$\gamma = 0$	τ_μ	-2,89	-3,51
	$\alpha = 0$ при $\gamma = 0$	$\tau_{\alpha\mu}$	2,54	3,22
	$\alpha = \gamma = 0$	ϕ_1	4,71	6,70
$\Delta y_t = \gamma y_{t-1} + u_t$	$\gamma = 0$	τ	-1,95	-2,60

представлены критические значения для проверки гипотезы $\gamma = 0$ при объеме выборки в 100 наблюдений на уровне значимости 5% и 1% для различных видов уравнения.

В частности, для уравнения без свободного члена и временного тренда ($\alpha = 0$ и $\beta = 0$) $\tau = -1,95$ для 5% и $\tau = -2,6$ для 1% уровня значимости. Таким образом, в числовом примере с $\gamma = -0,0454$, $\sigma = 0,030$, $t = -0,0454/0,030 = -1,5133$ находим, что $t < \tau$, и это означает, что нельзя отвергнуть нулевую гипотезу о единичном корне при заданных уровнях значимости.

Поскольку критические значения зависят от объема выборки, отметим, что при заданном уровне значимости критические значения статистик уменьшаются с ростом числа наблюдений. В 1991 г. Мак-Киннон провел более масштабные испытания, чем Дикки и Фуллер, и получил критические значения для выборок любого объема и любого числа переменных в правой части.

По-английски этот критерий называется *Augmented Dickey-Fuller Test* (ADF). Если обычный критерий Дикки – Фуллера позволяет проверять гипотезу о наличии авторегрессии первого порядка, то расширенный предназначен для идентификации авторегрессии более высоких порядков.

Итак, не все временные ряды могут быть хорошо представлены авторегрессионным процессом первого порядка. Рассмотрим авторегрессионный процесс p -го порядка

$$\Delta y_t = \alpha + \gamma y_{t-1} + \beta t + \sum_{i=2}^p \phi_i \Delta y_{t-i+1} + u_t. \quad (12.10)$$

Прибавим и вычтем $a_p y_{t-p+1}$, получим

$$y_t = \alpha + a_1 y_{t-1} + \dots + a_{p-2} y_{t-p+2} + (a_{p-1} + a_p) y_{t-p+1} - a_p \Delta y_{t-p+1} + u_t.$$

Вновь прибавим и вычтем $(a_{p-1} + a_p) y_{t-p+2}$, получим

$$y_t = \alpha + a_1 y_{t-1} + \dots - (a_{p-1} + a_p) \Delta y_{t-p+2} - a_p \Delta y_{t-p+1} + u_t.$$

Продолжая этот процесс и вычитая y_{t-1} из левой и правой части, приходим к следующему выражению:

$$y_t - y_{t-1} = \Delta y_t = \alpha + \gamma y_{t-1} + \sum_{i=2}^p \phi_i \Delta y_{t-i+1} + u_t, \quad (12.11)$$

$$\text{где } \gamma = \sum_{i=1}^p a_i - 1, \quad (12.12)$$

$$\phi_i = \sum_{j=i}^p a_j, \quad i = 2, 3, \dots, p, \quad (12.13)$$

$$\text{т.е. } \phi_p = a_p, \phi_{p-1} = \alpha_{p-1} + a_p, \phi_{p-2} = a_{p-2} + a_{p-1} + a_p, \dots$$

Интерес представляет коэффициент γ . Если он равен нулю, уравнение (12.11) оказывается полностью записанным через первые разности, и таким образом имеется единственный корень. Проверка наличия единичного корня может быть осуществлена с помощью тех же самых статистик критерия Дикки — Фуллера, распределения которых зависят от включения в регрессионное уравнение детерминированных компонент.

Другими словами, расширенный критерий Дикки — Фуллера используется для проверки наличия единичного корня в авторегрессионных уравнениях более высокого порядка, имеющих следующий вид:

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=2}^p \phi_i \Delta y_{t-i+1} + u_t; \quad (12.14)$$

$$\Delta y_t = \alpha + \gamma y_{t-1} + \sum_{i=2}^p \phi_i \Delta y_{t-i+1} + u_t; \quad (12.15)$$

$$\Delta y_t = \alpha + \gamma y_{t-1} + \beta t + \sum_{i=2}^p \phi_i \Delta y_{t-i+1} + u_t. \quad (12.16)$$

Если нет свободного члена и тренда, используется статистика τ ; если включен только свободный член, используется статистика τ_μ , а если присутствует и свободный член, и тренд, то — статистика τ_τ .

Отметим, что из (12.12) следует, что если сумма коэффициентов исходного уравнения равна 1, т.е. $\sum_{i=1}^p a_i = 1$, то $\gamma = 0$, а это значит, что имеется единичный корень. Для проверки совместных гипотез о значениях коэффициентов Дикки и Фуллера [138] предложили еще три статистики, названные ими ϕ_1 , ϕ_2 и ϕ_3 . Для уравнений (12.8) или (12.15) нулевая гипотеза $H_0: \gamma = \alpha = 0$ проверяется по статистике ϕ_1 . Если в регрессию включается временной тренд и оцениваются уравнения (12.9) или (12.16), то совместная нулевая гипотеза $H_0: \alpha = \beta = \gamma = 0$ проверяется по статистике ϕ_2 , а гипотеза $\alpha = \beta = 0$ — по статистике ϕ_3 .

Статистики ϕ_1 , ϕ_2 и ϕ_3 приведены нами в той же таблице. Они построены точно так же, как и обычные F -статистики для сравнения общей модели с частной [128]:

$$\phi_i = \frac{(R_{UR}^2 - R_R^2)/m}{(1 - R_{UR}^2)/(n - p - k)}, \quad i=1, 2, 3, \quad (12.17)$$

где R_{UR}^2 — коэффициент детерминации общей модели (без ограничений UnRestricted);
 R_R^2 — коэффициент детерминации частной модели (с ограничениями — Restricted);
 m — число ограничений (число параметров, приравняемых нулю);
 n — объем исходной выборки;
 p — порядок лага;
 k — число параметров, оцениваемых в общей модели,
 $(n - p - k)$ — число степеней свободы модели без ограничений.

Формулу (12.17) можно эквивалентным образом записать через суммы квадратов ошибок $СКО_{UR}$ и $СКО_R$ для соответствующих моделей. Так как $R^2 = 1 - \frac{СКО}{\sum (y_i - \bar{y})^2}$, то

$$\phi_i = \frac{(СКО_R - СКО_{UR})/m}{СКО_{UR}/(n - p - k)}. \quad (12.18)$$

Отметим, что критические значения статистик ϕ Дикки и Фуллера намного выше, чем в обычных таблицах F -распределения. Для сравнения приведем их для 5%-ного уровня значимости (табл. 12.2), числе степеней свободы числителя 2 и числе степеней свободы знаменателя $(n - p - 3)$.

В качестве иллюстрации рассмотрим пример, описанный Дикки и Фуллером. Они предположили, что логарифмы квартальных данных об индексе производства, рассчитываемым Федеральной резервной палатой, можно адекватно представить за период с I кв. 1950 г. по IV кв. 1977 г. моделью:

$$y_t = \alpha + \beta t + \alpha_1 y_{t-1} + \phi(y_{t-1} - y_{t-2}) + u_t,$$

которую адекватно можно записать как

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \phi \Delta y_{t-1} + u_t,$$

где $u_t \sim IN(0, \sigma^2)$ — случайная переменная, $\gamma = \alpha_1 - 1$.

Таблица 12.2

Сравнение критических значений
F-статистики и статистики $\phi(2, n - p - 3)$

Объем выборки (n - p)	Статистика Дикки - Фуллера ϕ	Стандартная статистика F
25	7,24	3,42
50	6,73	3,20
100	6,49	3,10
∞	6,25	3,00

Для проверки обоснованности структуры уравнения оценивались МНК три альтернативные модели, что дало следующие результаты.

Общая модель (без ограничений):

$$\Delta y_t = 0,52 + 0,00120t - 0,119y_{t-1} + 0,498\Delta y_{t-1} + e_t \quad (12.19)$$

$$\sigma = 0,15 \quad 0,00034 \quad 0,033 \quad 0,081$$

$$СКО_{UR} = 0,056448.$$

Частная модель 1 (с ограничениями $H_0: \beta = 0, \gamma = 0$):

$$\Delta y_t = 0,0054 + 0,447\Delta y_{t-1} + e_t \quad (12.20)$$

$$\sigma = 0,0025 \quad 0,083$$

$$СКО_R = 0,063211.$$

Частная модель 2 (с ограничениями $H_0: \alpha = 0, \beta = 0, \gamma = 0$):

$$\Delta y_t = 0,511\Delta y_{t-1} + e_t \quad (12.21)$$

$$\sigma = 0,079$$

$$СКО_R = 0,065966.$$

Для проверки значимости в модели (12.19) только одного коэффициента γ , т.е. гипотезы $H_0: \gamma = 0$, вычислим статистику

$$t_\gamma = -0,119 / 0,033 = -3,61.$$

Ее критическое значение при 5%-ном уровне значимости равно $-3,45$, т.е. наблюдаемое значение по модулю больше критического. Следовательно, нулевую гипотезу следует отвергнуть.

Выполним более сложные проверки, в которых выдвигаются совместные гипотезы относительно нескольких параметров модели.

Для проверки нулевой гипотезы, что данные генерированы моделью (12.21) против альтернативной, что «истинной» моделью является (12.19), т.е. $H_0: \alpha = 0, \beta = 0, \gamma = 0$, используется статистика ϕ_2 . Наблюдаемое значение статистики при проверке указанной гипотезы H_0 равно:

$$\phi_2 = \frac{(СКО_R - СКО_{UR})/m}{СКО_{UR}/(n - p - k)} = \frac{(0,065966 - 0,056448)/3}{0,056448/(112 - 2 - 4)} = 5,95.$$

Число степеней свободы знаменателя получено следующим образом. Всего квартальных наблюдений за полных 28 лет равно 112, однако в уравнении предполагается лаг зависимой переменной второго порядка, так что для оценивания остается только 110 наблюдений, и из этого числа вычитается число оцениваемых коэффициентов в общей модели, т.е. $110 - 4 = 106$. Критическое значение ϕ_2 при 5%-ном уровне значимости равно 4,88, т.е. наблюдаемое значение больше критического, следовательно, гипотезу H_0 (модель (12.21)) отвергаем.

Для проверки нулевой гипотезы, что данные генерированы моделью (12.20) против альтернативной, что «истинной» моделью является модель (12.19), т.е. $H_0: \beta = 0, \gamma = 0$, используется статистика ϕ_3 . Наблюдаемое значение статистики при проверке указанной гипотезы H_0 равно:

$$\phi_3 = \frac{(\text{СКО}_R - \text{СКО}_{UR})/m}{\text{СКО}_{UR}/(n-p-k)} = \frac{(0,063211 - 0,056448)/2}{0,056448/(112 - 2 - 4)} = 6,34.$$

При объеме данных около 100 наблюдений критическое значение $\phi_3 = 6,49$ при 5%-ном уровне значимости. Таким образом, наблюдаемое значение ϕ_3 меньше критического, гипотеза H_0 , которая означает наличие единичного корня и (или) детерминированного временного тренда, не отвергается, т.е. принимается модель (12.20).

Критерий Дикки — Фуллера предполагает, что ошибки независимы и имеют постоянную дисперсию. Здесь возникают четыре проблемы:

1. Процесс, порождающий временной ряд, может содержать как авторегрессию, так и скользящую среднюю. Необходимо понять, как нужно проводить тест, если порядок скользящей средней неизвестен.

2. Нельзя получить хорошую оценку γ и ее стандартной ошибки, если не все авторегрессионные члены включены в уравнение. Другими словами, возникает проблема выбора порядка авторегрессии (максимального лага).

3. Критерий Дикки — Фуллера предназначен для обнаружения единственного единичного корня. Однако авторегрессия порядка p имеет p характеристических корней. Если имеется $d \leq p$ единичных корней, то для достижения стационарности нужно перейти к разностям временного ряда порядка d .

4. Обычно заранее неизвестно, включать свободный член и временной тренд в уравнение (12.10).

Рассмотрим эти проблемы. Поскольку обратимая модель скользящего среднего (СС) может быть преобразована в авторегрессионную модель, то рассмотренная выше процедура Дикки — Фуллера применима и при наличии в модели компонент скользящего среднего. Пусть последовательность

значений $\{y_t\}$ генерирована смешанной моделью авторегрессии-скользящего среднего АРССС(p, q)

$$\phi(B)y_t = Q(B)u_t,$$

где $\phi(B)$ и $Q(B)$ — полиномы от оператора сдвига назад B соответственно порядка p и q . Оператор сдвига назад B означает, что $B y_t = y_{t-1}$, $B^2 y_t = y_{t-2}$ и т.д.

Если корни $Q(B)$ лежат вне единичного круга, то последовательность $\{y_t\}$ можно представить авторегрессионным процессом

$$\phi(B)y_t/Q(B) = u_t,$$

или, введя обозначение $D(B) = \phi(B)/Q(B)$, получаем

$$D(B)y_t = u_t.$$

Даже если оператор $D(B)$ будет полиномом бесконечного порядка, используя тот же прием, что и в (12.11), можно прийти к авторегрессионной модели бесконечного порядка вида

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=2}^{\infty} \beta_i \Delta y_{t-i+1} + u_t. \quad (12.22)$$

Бесконечную авторегрессию (12.22) нельзя оценить на конечной выборке. Однако Саид и Дикки в 1984 г. показали, что процесс интегрированной модели авторегрессии-скользящего среднего АРИСС($p, 1, q$), т.е. смешанный процесс авторегрессии-скользящего среднего для первых разностей, может быть хорошо аппроксимирован процессом АРИСС($p, 1, 0$), где порядок авторегрессии p не больше, чем $n^{1/3}$. Таким образом, первую проблему можно решить, используя авторегрессию конечного порядка для аппроксимации (12.22). Проверка нулевой гипотезы $H_0: \gamma = 0$ может быть выполнена с помощью статистик τ , τ_μ или τ_τ рассмотренного выше критерия Дикки — Фуллера.

Теперь возникает проблема определения подходящего лага. Завышенный лаг снижает мощность критерия по про-

верке нулевой гипотезы о единичном корне, так как в этом случае требуется оценить большее число параметров, а, кроме того, уменьшается число наблюдений, используемых для оценивания. И то, и другое снижает число степеней свободы. С другой стороны, заниженный лаг недостаточно хорошо будет отражать действительные свойства ошибок, так что γ и ее стандартная ошибка будут оценены плохо. Как же выбрать хорошее значение лага в таком случае?

Один подход состоит в том, чтобы начать с достаточно большого лага и постепенно сокращать его в модели на основе обычных t - и F -критериев. Повторять процесс до тех пор, пока коэффициент при лаговом значении не станет существенно отличным от нуля. В чисто авторегрессионном процессе это приведет к определению истинного значения лага. При наличии сезонных явлений процесс будет несколько иным. Например, используя квартальные данные, можно начать с трехлетнего лага, т.е. $m = 12$. Если t -статистика у лага 12 незначительна, а F -критерий указывает на то, что лаги с 9 по 12 также незначительны, переходите к лагам с 1 по 8. Повторите этот процесс для лага 8 и лагов с 5 по 8 и так далее, пока не будет достигнута приемлемая величина лага.

После того как лаг выбран, требуется выполнить диагностическую проверку. Прежде всего следует изучить график остатков: в них не должно быть признаков структурных сдвигов или автокорреляции. Коррелограмма остатков должна быть такой, как у белого шума. Статистика Бокса — Пирса Q (см. гл. 7 § 3) или ее модификация — статистика Льянга — Бокса

$$Q^* = n(n+2) \sum_{k=1}^m r_k^2 / (n-k) \sim \chi^2(m-p-q)$$

не должна показывать существенную автокорреляцию остатков. Не рекомендуется альтернативная процедура, начинающая с низкого значения лага, а затем увеличивающая его до тех пор, пока не будет достигнут незначимый лаг. Исследования методом Монте-Карло показали, что для этой процедуры характерно смещение в сторону занижения лага.

Множественность корней

Дикки и Пентала предложили простое расширение критерия Дикки — Фуллера на случай, когда корней больше, чем один. По существу, речь идет о применении процедуры Дикки — Фуллера к последовательным разностям y_t . Если подозревается один корень, то оценивается уравнение вида

$$\Delta y_t = \alpha + \gamma y_{t-1} + u_t$$

Если предполагаются два корня, то оценивается уравнение

$$\Delta^2 y_t = \alpha + \gamma_1 \Delta y_{t-1} + u_t \quad (12.23)$$

Для проверки существенного отличия γ_1 от нуля используются те же самые статистики τ , τ_μ или τ_c (в зависимости от детерминированных элементов, включенных в модель). Если нельзя отвергнуть гипотезу, что $\gamma_1 = 0$, то гипотеза о втором порядке разностей принимается и делается вывод, что ряд y_t является интегрированным порядка 2, т.е. $I(2)$. Если же γ_1 отличен от нуля, следует проверить гипотезу о том, что имеется один корень, путем оценивания уравнения

$$\Delta^2 y_t = \alpha + \gamma_1 \Delta y_{t-1} + \gamma_2 y_{t-2} + u_t \quad (12.24)$$

Наличие одного корня означает, что либо γ_1 , либо γ_2 , либо оба коэффициента отличны от нуля. При нулевой гипотезе о наличии единственного корня $\gamma_1 < 0$ и $\gamma_2 = 0$. При альтернативной гипотезе y_t является стационарным, так что γ_1 и γ_2 отрицательны. Таким образом, нужно оценить модель (12.24) и использовать критические значения статистик Дикки — Фуллера для проверки гипотезы $\gamma_1 = 0$. Если нулевая гипотеза отвергается, делается вывод о том, что y_t стационарен.

Отметим, что экономические ряды не требуют перехода к разностям выше второго порядка.

Сезонность и единичные корни

Для простоты изложения предположим, что квартальные наблюдения y_t генерируются моделью

$$y_t = y_{t-4} + u_t$$

т.е. разности лага 4 стационарны:

$$\Delta_4 y_t = y_t - y_{t-4} = u_t.$$

Если начальные значения положить равными нулю: $y_0 = y_{-1} = \dots = 0$, то

$$y_t = u_t + u_{t-4} + u_{t-8} + \dots,$$

так что

$$y_t - y_{t-1} = \sum_{i=0}^{t/4} u_{4i} - \sum_{i=0}^{t/4} u_{4i-1}.$$

Следовательно, первые разности Δy_t являются разностями между двумя стохастическими трендами. Поскольку дисперсия Δy_t неограниченно возрастает с ростом t , последовательность Δy_t нестационарна. Однако разности лага, равного периоду сезонности, могут быть стационарными. Например, если y_t генерируется моделью $y_t = y_{t-1} + u_t$, то разности лага 4 $\Delta_4 y_t = y_t - y_{t-4} = u_t + u_{t-1} + u_{t-2} + u_{t-3}$ стационарны, однако их дисперсия больше дисперсии первых разностей.

Процедура Дикки — Фуллера может быть модифицирована для обнаружения единичных корней, связанных с сезонностью, и для различения сезонных и несезонных единичных корней.

Возможно несколько альтернативных способов обработки сезонности в нестационарных рядах. Наиболее прямой метод применим, когда сезонность является чисто детерминированным явлением. Например, пусть D_1, D_2, D_3 представляют квартальные сезонные фиктивные переменные, так что $D_i = 1$ в квартале i и 0 — в других случаях. Оценим регрессионное уравнение

$$y_t = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \hat{y}_t,$$

где $\hat{y}_t$ — регрессионный остаток, который можно рассматривать как y_t , очищенный от сезонных колебаний.

Затем используем регрессионные остатки для оценивания регрессии:

$$\Delta \hat{y}_t = \gamma \hat{y}_{t-1} + \sum_{i=2}^p \beta_i \Delta \hat{y}_{t-i+1} + u_t. \quad (12.25)$$

Нулевая гипотеза о наличии единичного корня (т.е. $\gamma = 0$) может быть проверена с помощью статистики τ_μ Дикки — Фуллера. Отклонение нулевой гипотезы эквивалентно принятию альтернативной, что последовательность y_t является стационарной. Этот критерий применим, так как Д. Дикки, В. Белл и Р. Миллер [139] показали, что на предельное распределение для γ не влияет удаление детерминированных сезонных компонент. Если желательно включить временной тренд в уравнение (12.25), то следует использовать статистику τ_τ .

§ 5. СОВРЕМЕННЫЕ МЕТОДЫ ПОСТРОЕНИЯ МОДЕЛИ АРИСС

В главе 7 была рассмотрена модель, предложенная Дж. Боксом и Г. Дженкинсом [5], которая объединяет авторегрессию исходного временного ряда и скользящую среднюю для ошибок.

Нестационарные модели авторегрессионного типа разработаны для тех случаев, когда нестационарные временные ряды могут быть преобразованы в стационарные путем перехода от исходного ряда к его разностям соответствующего порядка d . Такая модель является более общей и называется *интегрированной моделью авторегрессии-скользящего среднего* порядка p, d, q и обозначается АРИСС(p, d, q), где p — порядок авторегрессии, q — порядок скользящей средней. Обычно порядок разностей $d \leq 2$.

Первый шаг в построении модели АРИСС(p, d, q) для одномерного временного ряда состоит в выборе порядков p, d, q , который осуществляется в два этапа. Сначала определяют порядок разности d , при котором достигается стационарность процесса. Этот порядок называется *порядком интеграции* и обозначается $I(d)$. Он определяется с помощью расширенного критерия Дикки — Фуллера (ADF test). За-

тем определяются порядки модели АРСС p и q для разностей порядка d . Дж. Бокс и Г. Дженкинс делали это на основе изучения коррелограммы временного ряда и сравнения ее с теоретической коррелограммой конкретной теоретической модели. Однако сейчас выбор порядков p и q осуществляется на основе информационного критерия Акайка или байесовского критерия Шварца.

Оценивание параметров моделей АР, СС и АРСС производится путем минимизации суммы квадратов ошибок $\sum u_i^2$; для этого применяются итеративные процедуры или процедуры поиска наилучших оценок на сетке значений. Значения ошибок u_i интерпретируются как ошибки прогнозов на один шаг (на одну единицу времени) вперед. Одной из целей при построении модели является поиск модели с наименьшим числом параметров — *принцип экономности*.

Общим критерием качества модели АРСС является критерий Акайка и байесовский критерий Шварца. Они аналогичны критерию максимума скорректированного коэффициента детерминации $\bar{R}^2$ или минимума дисперсии ошибки $\hat{\sigma}^2$. Эти критерии позволяют найти наилучшие значения порядков p и q модели. Однако наилучшая модель из заданного класса моделей еще не гарантирует того, что она является хорошей. Необходимы дополнительные исследования ее качества. В частности, следует проверить автокорреляцию остатков. Для авторегрессионных моделей нельзя использовать статистику Дарбина — Уотсона (DW). В этом случае следует применять критерий h Дарбина или критерий множителей Лагранжа (LM). Как уже упоминалось, Дж. Бокс и Дж. Пирс [45] предложили изучать автокорреляцию не только первого порядка, а всех порядков. Статистика Q Бокса — Пирса (Box — Pierce test) имеет вид

$$Q = n \sum_{k=1}^m r_k^2,$$

где r_k — коэффициент автокорреляции лага k ;
 n — число наблюдений в ряде.

Если модель адекватна, то Q имеет асимптотическое распределение χ^2 с числом степеней свободы $m - p - q$, где

p и q — порядки АР и СС частей модели. Льянгом и Боксом предложена модификация этого теста для выборок умеренного объема:

$$Q^* = n(n+2) \sum_{k=1}^m r_k^2 / (n-k) \sim \chi^2(m-p-q). \quad (12.26)$$

Однако имеются свидетельства того, что и критерий Q , и критерий Q^* имеют малую мощность, а критерий LM является более предпочтительным, и он уже введен в современные программы. Однако в настоящее время в программах и публикациях чаще встречается критерий Q или его модификация.

Информационный критерий Акайка

Пусть $l_n(\tilde{\theta})$ будет максимальное значение логарифмической функции правдоподобия эконометрической модели, где $\tilde{\theta}$ — оценка максимального правдоподобия (ML) вектора параметров θ , полученная на выборке объема n . Информационный критерий Акайка (Akaike information criterion — AIC) [132, 133], предназначенный для выбора наилучшей модели для временного ряда y_t из некоторого их множества, имеет вид

$$AIC_i = l_n(\tilde{\theta}) - p, \quad (12.27)$$

где p — размерность вектора θ , т.е. число оцениваемых коэффициентов регрессионной модели.

В случае линейной или нелинейной регрессионной модели, состоящей из одного уравнения, этот критерий эквивалентным образом может быть переписан как

$$AIC_{\sigma} = \ln(\tilde{\sigma}^2) + \frac{2p}{n}, \quad (12.28)$$

где $\tilde{\sigma}^2$ — оценка ML-дисперсии регрессионных остатков e_i ;

$\tilde{\sigma}^2 = \frac{e'e}{n}$, e — вектор ошибок.

Оба варианта критерия Акайка дают одинаковые результаты, но в первом варианте выбирается модель с наибольшим значением AIC_1 , а во втором — с наименьшим AIC_0 . Таким образом, и в том, и в другом случае в этом критерии вводится штраф за добавление в правую часть уравнения каждой новой объясняющей переменной.

В случае линейных регрессионных моделей эквивалентность двух вариантов критерия вытекает из следующих преобразований. Подставим максимальное значение логарифмической функции правдоподобия

$$l_n(\tilde{\theta}) = -\frac{n}{2} \{1 + \ln(2\pi\tilde{\sigma}^2)\} \quad (12.29)$$

в (12.27):

$$AIC_1 = -\frac{n}{2}(1 + \ln 2\pi) - \frac{n}{2} \ln \tilde{\sigma}^2 - p \quad (12.30)$$

и, используя (12.28), получаем

$$AIC_1 = -\frac{n}{2}(1 + \ln 2\pi) - \frac{n}{2} AIC_0. \quad (12.31)$$

Так что максимум AIC_1 будет достигаться при минимуме AIC_0 .

Байесовский критерий Шварца

Байесовский критерий Шварца (Schwarz Bayesian criterion — SBC) также предназначен для выбора наилучшей модели, как и информационный критерий Акайка. Он представляет собой аппроксимацию на больших выборках апостериорного отношения вероятностей сравниваемых моделей. Он определяется как

$$SBC_1 = l_n(\tilde{\theta}) - \frac{1}{2} p \ln n. \quad (12.32)$$

Из ряда сопоставляемых моделей выбирается та, у которой наибольшее значений SBC .

Альтернативный вариант критерия для регрессионных моделей выражается через $\tilde{\sigma}^2$

$$SBC_0 = \ln(\tilde{\sigma}^2) + \left(\frac{\ln n}{n}\right)p. \quad (12.33)$$

В соответствии с этим вариантом критерия выбирается модель с наименьшим значением SBC_0 . Оба варианта критерия приводят к одинаковому выбору модели. Однако результаты критериев Акайка и Шварца могут немного отличаться. Как правило, при выборе порядка лага в авторегрессионной модели критерий Акайка указывает на больший порядок, что означает большую величину «штрафа» за добавление переменной в правую часть уравнения у критерия Шварца.

Общий критерий множителей Лагранжа для обнаружения автокорреляции остатков

Обычно для обнаружения автокорреляции в регрессионных остатках применяется критерий Дарбина — Уотсона, который обозначается как DW или как d . Однако он неприменим в случае, когда среди регрессоров имеются лаговые значения эндогенной переменной. На это указали сами авторы критерия, и одним из них, Дарбиным, был предложен альтернативный h -критерий, который, однако, можно вычислить не всегда. И, кроме того, оба критерия и DW , и h являются критериями автокорреляции только первого порядка. Breusch (1978), и Godfrey (1978) рассматривают несколько общих тестов, которые легко применять и сохраняют свою силу при очень общих гипотезах относительно автокорреляции в регрессионных остатках. Эти критерии получены на основе общего принципа, называющегося принципом множителей Лагранжа. Один из них, критерий множителей Лагранжа (LM-test), является более общим и позволяет обнаруживать в регрессионных остатках автокорреляцию и более высоких порядков. Как и h -тест, он применим только для больших выборок.

Рассмотрим регрессионную модель

$$y_t = \sum_{i=1}^k x_{ti} \beta_i + u_t, \quad t = 1, 2, \dots, n, \quad (12.34)$$

в которой

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \varepsilon_t, \quad (12.35)$$

$$\varepsilon_t \sim IN(0, \sigma^2).$$

Необходимо проверить нулевую гипотезу H_0 :

$$\rho_1 = \rho_2 = \dots = \rho_p = 0.$$

Среди объясняющих переменных могут быть лаговые значения зависимой переменной. LM-тест состоит из следующих шагов:

1) оценить регрессию (12.34) методом наименьших квадратов, получить остатки $\hat{u}_t$;

2) оценить регрессию

$$\hat{u}_t = \sum_{i=1}^k x_{it} a_i + \sum_{i=1}^p \hat{u}_{t-i} \rho_i + v_t \quad (12.36)$$

и проверить, не являются ли все коэффициенты ρ_i при лаговых значениях остатков нулевыми. Для этого вычисляется F -статистика и рассматривается pF как $\chi^2(p)$. Здесь используется χ^2 -критерий, а не F , так как LM-тест предназначен для больших выборок.

Этот критерий может быть использован для различных спецификаций процесса, описывающего ошибки регрессии. Например, при проверке автокорреляции четвертого порядка (связанной с возможной сезонностью в квартальных данных)

$$u_t = \rho_4 u_{t-4} + \varepsilon_t \quad (12.37)$$

нужно оценить регрессию

$$\hat{u}_t = \sum_{i=1}^k x_{it} a_i + \rho_4 \hat{u}_{t-4} + v_t \quad (12.38)$$

и проверить нулевую гипотезу $H_0: \rho_4 = 0$.

Процедура одна и та же для авторегрессионной или скользящей средней модели, генерирующей ошибки. Например, если ошибка описывается скользящей средней

$$u_t = \varepsilon_t + \rho_4 \varepsilon_{t-4},$$

то все равно следует оценить уравнение (12.38) и проверить $\rho_4 = 0$.

Пример

Выбор порядков p и q в модели АРСС. Сравнивая значения AIC или SBC для различных порядков p и q , в пакете MicroFit [156] выбирают спецификацию модели с наибольшим значением критерия (в некоторых программах, напротив, с наименьшим значением — это зависит от вида критерия, выбранного программистом). Данные удобно представлять в табличном виде (табл. 12.3).

Таблица 12.3

Таблица значений AIC

p	q			
	0	1	2	3
0	42,59	46,47	48,87	48,02
1	48,35	47,92	48,08	47,10
2	48,76	48,21	47,22	46,25
3	48,13	47,29	46,30	45,35

По этим данным выбирается модель АРСС(0,2), т.е. СС(2) — скользящая средняя второго порядка (в табл. 12.3 она выделена полужирным шрифтом).

В таблице 12.4 приведены для тех же данных значения SBC .

Таблица 12.4

p	q			
	0	1	2	3
0	41,15	43,59	44,55	42,25
1	45,46	43,60	42,32	39,89
2	44,43	42,44	40,02	37,60
3	42,37	40,08	37,65	35,26

По данным табл. 12.4 выбирается модель АРСС(1,0), т.е. АР(1) — авторегрессия первого порядка.

Как видим, два критерия могут давать различные результаты.

Определение типа тренда, порядка разностей d . Тип тренда во временном ряде может быть детерминированным либо стохастическим. Тип тренда устанавливается с помощью критерия ADF . Этот же критерий позволяет установить порядок разностей, обеспечивающий стационарность.

Рассмотрим, например, процесс $AR(1)$

$$y_t = \alpha + (1 - \phi)\delta t + \phi y_{t-1} + u_t,$$

где u_t распределены идентично и независимо, т.е. $u_t \sim iid(0, \sigma^2)$. Если $|\phi| < 1$, то тренд детерминированный, а если $\phi = 1$, то имеется единичный корень и первые разности стационарны, т.е. тренд стохастический с ненулевым дрейфом α и $d = 1$.

Попытка определить тип тренда в общем виде принимает форму проверки односторонней гипотезы о наличии единичного корня против альтернативной гипотезы о стационарности стохастического тренда:

$$H_0: \phi = 1 \text{ против } H_1: \phi < 1.$$

Важно отметить, что при использовании t -статистики для проверки, что $\phi = 1$, критические значения нужно брать не из таблиц стандартного t -распределения, а из нестандартного распределения Дикки – Фуллера.

В более общем случае, когда возмущения u_t , $t = 1, 2, \dots, n$, автокоррелированы, следует применять расширенный критерий Дикки – Фуллера для проверки наличия единичных корней, который предполагает учесть автокорреляцию ошибок путем добавления лаговых разностей y_t :

$$y_t = \alpha + (1 - \phi)\delta t + \phi y_{t-1} + \sum_{i=1}^{p-1} \phi_i y_{t-i} + u_t, \quad t = 1, \dots, n,$$

что может быть также записано как

$$\Delta y_t = \alpha + \rho \delta t - \rho y_{t-1} + \sum_{i=1}^p \gamma_i \Delta y_{t-i} + u_t,$$

где нулевая гипотеза теперь имеет вид $H_0: \rho = 1 - \phi = 0$.

В ходе проверки на наличие единичных корней необходимо выбрать оптимальный порядок p регрессии ADF . Это делается методом проб на основе критерия AIC или SBC .

Отметим, что к полученным первым разностям следует повторно применить критерий ADF с целью проверки, не следует ли перейти ко вторым разностям для обеспечения стационарности и т.д., пока не будет определен порядок интегрированности d . Напомним также, что критические значения статистик Дикки – Фуллера зависят от включения в уравнение свободного члена и временного тренда. Отметим, что, например, в пакете *MicroFit* имеется специальная, позволяющая проводить анализ наличия единичных корней процедура, которая автоматически выводит и критические значения статистик Дикки – Фуллера.

* * *

Данная глава посвящена проблемам тренда во временном ряде. Еще в гл. 1, § 1 были введены понятия детерминированного и стохастического трендов. Теперь же рассмотрены критерии обнаружения трендов и установления их типа. Материал этой главы весьма важен для построения моделей Бокса – Дженкинса $ARISCC(p, d, q)$, рассмотренных нами в гл. 7, в частности, для определения необходимого порядка разностей d . Это первый, но, очевидно, важнейший шаг в построении моделей такого класса. Определение порядка авторегрессии p и порядка скользящей средней q осуществляется с помощью информационного критерия Акайка (AIC) или байесовского критерия Шварца (SBC), которые имеются в стандартных эконометрических пакетах программ, например в пакете *MicroFit*.

Глава 13

ИНТЕГРИРОВАННОСТЬ И КОИНТЕГРИРОВАННОСТЬ ПЕРЕМЕННЫХ

В эконометрическом моделировании часто требуется предварительное элиминирование временных трендов из исходных данных. Кроме того, необходима сбалансированность динамических свойств переменных, стоящих в левой и правой частях регрессионного уравнения. Каким образом осуществляется контроль в этом случае, мы и рассмотрим в этой главе.

§ 1. ИНТЕГРИРОВАННОСТЬ И КОИНТЕГРИРОВАННОСТЬ ДВУХ ПЕРЕМЕННЫХ

Большинство экономических рядов нестационарно, поскольку обычно они имеют линейный, параболический или экспоненциальный временной тренд. Однако часто временной ряд можно привести к стационарному виду, если перейти к его разностям (приростам) вида $y_t - y_{t-1}$. Если первые разности ряда стационарны, то говорят, что этот ряд является *интегрированным первого порядка*. Переход к разностям приводит к потере ценной долгосрочной информации в данных. Сравнительно недавно для решения этой проблемы были предложены понятия интегрированности и коинтегрированности временных рядов [142]. Несмотря на то что предположение о нормальности и независимости не является необходимым, дадим определения при этом предположении.

Если стационарный временной ряд $u_t \sim \text{IN}(0, \sigma^2)$, говорят, что u_t является интегрированным процессом порядка ноль и записывают это символически как $I(0)$.

Если y_t генерирован моделью случайного блуждания

$$y_t = y_{t-1} + u_t$$

то, приняв, что $y_0 = 0$, путем последовательных подстановок получаем

$$y_t = \sum_{j=0}^{t-1} u_{t-j}.$$

Таким образом, y_t является суммой u_t и

$$\Delta y_t = u_t$$

является процессом $I(0)$. В этом случае говорят, что y_t является интегрированным процессом порядка один, т.е. $y_t \sim I(1)$.

Если y_t есть $I(1)$ и мы прибавляем к нему z_t , который является $I(0)$, то сумма $(y_t + z_t)$ будет $I(1)$.

Когда специфицируют регрессионную модель для временных рядов, то обращают внимание на то, чтобы различные переменные уравнения имели одинаковый порядок интегрированности, т.е. чтобы при одном порядке разности они достигали стационарного вида. В противном случае уравнение не имеет смысла. Например, если специфицируется регрессионная модель

$$y_t = \beta x_t + u_t \quad (13.1)$$

и предполагается, что $u_t \sim \text{IN}(0, \sigma^2)$, т.е. что u_t является $I(0)$, то нужно удостовериться, что y_t и x_t являются интегрированными процессами одного порядка. Например, если y_t есть $I(1)$, а x_t — процесс $I(0)$, то ни при каком β соотношение (13.1) не будет удовлетворено.

Пусть y_t есть $I(1)$ и x_t тоже $I(1)$, тогда, если существует такое ненулевое значение β , что $y_t - \beta x_t$ есть процесс $I(0)$, то говорят, что y_t и x_t *коинтегрированы*.

Предположим, что y_t и x_t являются процессами случайного блуждания, так что оба они $I(1)$. Тогда уравнение в первых разностях вида

$$\Delta y_t = \alpha \Delta x_t + \lambda(y_t - \beta x_t) + u_t \quad (13.2)$$

является правомерным, так как Δy_t , Δx_t , $(y_t - \beta x_t)$ и все u_t являются процессами $I(0)$.

Уравнение (13.1) рассматривается как долгосрочное соотношение между y_t и x_t , а уравнение (13.2) описывает краткосрочную динамику. Энгл и Гренжер предлагают оценивать выражение (13.1) методом наименьших квадратов, получать оценку $\hat{\beta}$ и подставлять ее в уравнение (13.2) с тем, чтобы затем получить оценки параметров α и λ . Однако эта двухшаговая процедура основывается на предположении, что y_t и x_t коинтегрированы. Поэтому предварительно нужно проверить наличие коинтегрированности. Энгл и Гренжер предлагают оценивать модель (13.1) МНК, получать остатки $\hat{u}_t$, а затем применять критерий Дикки — Фуллера для анализа этих остатков. Целью данного теста является проверка гипотезы, что $\rho = 1$ в уравнении

$$u_t = \rho u_{t-1} + e_t$$

т.е. проверка гипотезы $H_0: u_t \sim I(1)$. По сути дела, проверяется гипотеза, что y_t и x_t не коинтегрированы. Поскольку $y_t \sim I(1)$ и $x_t \sim I(1)$, мы хотим убедиться, что u_t не является процессом $I(1)$.

Р. Бьюли, М. Уиккенс и Т. Бройш [134, 162] показали, что двухшаговая процедура оценивания, предложенная Р. Энглом и К. Гренжером, предполагающая сначала оценивание параметра долгосрочной связи β в уравнении (13.1), а затем оценивание параметров краткосрочной связи α и λ в уравнении (13.2), не является необходимой. Они предлагают одновременное оценивание долгосрочных и краткосрочных параметров, что позволяет получить более эффективные оценки долгосрочного параметра β . Разделив уравнение (13.2) на λ и немного перестроив уравнение, получаем

$$y_t = \beta x_t + \frac{1}{\lambda} \Delta y_t - \frac{\alpha}{\lambda} \Delta x_t - \frac{u_t}{\lambda} \quad (13.3)$$

Поскольку Δy_t будет коррелирована с ошибкой u_t , уравнение (13.3) следует оценивать методом инструментальной переменной. Коэффициенты при Δy_t и Δx_t описывают краткосрочную динамику. Отметим, что если y_t и x_t являются $I(1)$, то Δy_t и Δx_t , как и u_t , есть $I(0)$. М. Уиккенс и Т. Бройш показали, что ошибки в спецификации краткосрочной динамики не оказывают большого влияния на оценивание долгосрочных параметров. Например, даже если в уравнении (13.3) опущена Δx_t , оценка параметра β будет состоятельной.

§ 2. КОИНТЕГРАЦИЯ МНОГИХ ПЕРЕМЕННЫХ

Итак, в простейших случаях стохастический тренд в переменной устраняется путем перехода к разностям. Полученный стационарный ряд может быть представлен моделью АРИСС, оцененной на статистических данных. Эта идея может быть обобщена на случай множественного регрессионного анализа, когда все нестационарные ряды приведены к стационарному виду путем перехода к разностям соответствующего порядка. Однако в многомерном случае работа с нестационарными рядами не столь проста.

Определение коинтеграции. Переменные, входящие в регрессионную модель, должны находиться в некотором равновесии, что означает наличие связи между их стохастическими трендами. Формальный анализ начнем с рассмотрения множества экономических переменных в долгосрочном равновесии, когда

$$\beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} = 0.$$

Если ввести вектора $x_t = (x_{1t}, x_{2t}, \dots, x_{kt})$ и $\beta = (\beta_1, \beta_2, \dots, \beta_k)$, то в векторном виде долгосрочное равновесие можно записать как

$$x_t \beta = 0,$$

где $t = 1, 2, \dots, n$; n — объем выборки.

Отклонение от долгосрочного равновесия $e_t = x_t \beta$ назовем *ошибкой равновесия*. Если равновесие в действитель-

ности существует, то ошибка равновесия должна быть стационарным процессом. Р. Энгл и К. Гренжер в общем случае дают следующее определение коинтеграции.

Говорят, что компоненты вектора $x_t = (x_{1t}, x_{2t}, \dots, x_{kt})$ коинтегрированы порядка (d, b) , что обозначается как $x_t \sim CI(d, b)$, если:

1) все компоненты вектора x_t — интегрированные процессы порядка d ;

2) существует вектор $\beta = (\beta_1, \beta_2, \dots, \beta_k)'$ такой, что линейная комбинация

$$x_t \beta = \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt}$$

представляет собой интегрированный процесс порядка $(d - b)$, где $b > 0$; вектор β называется *вектором коинтеграции*.

Следует сделать четыре важных замечания относительно этого определения.

1. Понятие коинтеграции относится к линейной комбинации нестационарных переменных. Теоретически вполне возможно, что существуют нелинейные долгосрочные соотношения между множеством интегрированных переменных. Однако современная эконометрика не располагает критериями для проверки нелинейных коинтеграционных связей. Отметим также, что вектор коинтеграции не единствен. Если $(\beta_1, \beta_2, \dots, \beta_k)'$ является таким вектором, то для любого отличного от нуля λ вектор $(\lambda\beta_1, \lambda\beta_2, \dots, \lambda\beta_k)'$ также является вектором коинтеграции. Поэтому обычно одна из переменных используется для нормализации вектора коинтеграции путем приравнивания ее коэффициента единице. Для нормализации вектора коинтеграции относительно x_1 просто следует взять $\lambda = 1/\beta_1$.

2. Все коинтегрированные переменные должны иметь одинаковый порядок интеграции (необходимое условие). Это, конечно, не означает, что все переменные одного порядка интеграции являются коинтегрированными. Обычно как раз наоборот. Отсутствие коинтеграции означает отсутствие долгосрочного равновесия в переменных, так что они могут уходить сколь угодно далеко одна от другой. Однако, если переменные имеют разные порядки интеграции, то они не могут быть коинтегрированы. Например, если $x_{1t} \sim I(d_1)$, а $x_{2t} \sim I(d_2)$ имеют разные порядки интеграции, т.е. $d_1 \neq d_2$, то

из равенства $\beta_1 x_{1t} + \beta_2 x_{2t} = 0$ следует $x_{1t} = -\frac{\beta_2}{\beta_1} x_{2t}$, что является недопустимым, так как $d_1 \neq d_2$.

3. Если вектор x_t имеет k компонент, то может быть $k - 1$ линейно независимых векторов коинтеграции. Ясно, что если x_t содержит только две переменные, то самое большее может быть один независимый вектор коинтеграции. Число векторов коинтеграции называется *рангом коинтеграции вектора x_t* .

4. В большинстве работ, посвященных проблемам коинтеграции, рассматривается случай, когда каждая переменная имеет единственный единичный корень. Причина этого состоит в том, что традиционный регрессионный анализ временных рядов имеет дело с переменными $I(0)$, $I(1)$ и очень редко с экономическими показателями, имеющими порядок интеграции выше 1. Поэтому часто термин коинтеграция употребляется в узком смысле для обозначения процессов $CI(1,1)$, хотя можно себе представить множество переменных $I(2)$, имеющих порядок коинтеграции $CI(2,1)$, так что существует их линейная комбинация $I(1)$.

Коинтеграция и тренды. Рассмотрим простейший случай, когда вектор x_t содержит только две переменные y_t и z_t , очищенные от циклических и сезонных колебаний, т.е. $x_t = (y_t, z_t)$, и каждую переменную можно представить моделью случайного блуждания с аддитивной случайной компонентой:

$$y_t = \mu_{y,t} + u_{y,t};$$

$$z_t = \mu_{z,t} + u_{z,t},$$

где $\mu_{y,t}$ и $\mu_{z,t}$ представляют процессы случайного блуждания трендов переменных y_t и z_t соответственно в момент t , а $u_{y,t}$ и $u_{z,t}$ — стационарные случайные отклонения.

Если $u_{y,t}$ и $u_{z,t}$ коинтегрированы $CI(1,1)$, то должны существовать ненулевые значения β_1 и β_2 , для которых линейная комбинация

$$\begin{aligned} \beta_1 y_t + \beta_2 z_t &= \beta_1 (\mu_{y,t} + u_{y,t}) + \beta_2 (\mu_{z,t} + u_{z,t}) = \\ &= (\beta_1 \mu_{y,t} + \beta_2 \mu_{z,t}) + (\beta_1 u_{y,t} + \beta_2 u_{z,t}) \end{aligned}$$

является стационарной. Но для стационарности этой суммы член $(\beta_1 \mu_{y_t} + \beta_2 \mu_{z_t})$ должен быть равен нулю. Если он не равен нулю, то $\beta_1 y_t + \beta_2 z_t$ также будет иметь тренд. Поскольку второй член в скобках стационарен, необходимым и достаточным условием того, чтобы y_t и z_t были коинтегрированными процессами $CI(1,1)$, является

$$\beta_1 \mu_{y_t} + \beta_2 \mu_{z_t} = 0.$$

Ясно, что μ_{y_t} и μ_{z_t} непрерывно изменяются во времени. Поскольку $\beta_1 \neq 0$ и $\beta_2 \neq 0$, то

$$\mu_{y_t} = -\frac{\beta_2}{\beta_1} \mu_{z_t},$$

что означает идентичность стохастических трендов с точностью до постоянного (масштабного) множителя. Таким образом, можно сделать вывод: с точностью до постоянного множителя $-\beta_2/\beta_1$ два стохастических процесса с порядком интеграции один $I(1)$ должны иметь одинаковый стохастический тренд, если они коинтегрированы с порядком $CI(1,1)$.

Пример. Рассмотрим три процесса

$$y_t = \mu_{y_t} + u_{y_t};$$

$$z_t = \mu_{z_t} + u_{z_t};$$

$$w_t = \mu_{w_t} + u_{w_t} = \mu_{y_t} + \mu_{z_t} + u_{w_t},$$

где $\mu_{y_t} = \mu_{z_t} = \mu_t = \mu_{t-1} + \varepsilon_t$

и все случайные компоненты u_{y_t} , u_{z_t} , u_{w_t} , ε_t — независимо распределенные отклонения (белые шумы).

Каждый процесс является интегрированным первого порядка $I(1)$ и критерий Дикки — Фуллера не отверг бы гипотезу о наличии в каждом из них единичного корня. Это означает, что процессы расходятся во времени. Однако тренд у показателя w_t является суммой трендов в y_t и z_t

$$\mu_{w_t} = \mu_{y_t} + \mu_{z_t}.$$

Следовательно, вектор $x_t = (y_t, z_t, w_t)$ имеет вектор коинтеграции $(1, 1, -1)$, так как линейная комбинация

$$\begin{aligned} y_t + z_t - w_t &= (\mu_{y_t} + u_{y_t}) + (\mu_{z_t} + u_{z_t}) - (\mu_{w_t} + u_{w_t}) = \\ &= u_{y_t} + u_{z_t} - u_{w_t} \end{aligned}$$

является стационарной.

Этот пример иллюстрирует тот факт, что коинтеграция имеет место тогда, когда тренд одной переменной может быть выражен через линейную комбинацию трендов других переменных. Элементы вектора коинтеграции должны быть таковы, чтобы тренд был исключен из линейной комбинации переменных. Такой результат может быть обобщен на случай k переменных. Рассмотрим векторное уравнение

$$x_t = \mu_t + u_t, \quad (13.4)$$

где $x_t = (x_{1t}, x_{2t}, \dots, x_{kt})$ — вектор значений переменных в момент t ;

$\mu_t = (\mu_{1t}, \mu_{2t}, \dots, \mu_{kt})$ — вектор стохастических трендов;

$u_t = (u_{1t}, u_{2t}, \dots, u_{kt})$ — вектор случайных компонент.

Если тренд одной переменной может быть выражен в виде линейной комбинации трендов других переменных системы, то это означает, что существует вектор-столбец $\beta = (\beta_1, \beta_2, \dots, \beta_k)'$, такой, что:

$$\mu_t \beta = \beta_1 \mu_{1t} + \beta_2 \mu_{2t} + \dots + \beta_k \mu_{kt} = 0.$$

Умножив справа уравнение (13.4) на β , получим

$$x_t \beta = \mu_t \beta + u_t \beta.$$

Так как $\mu_t \beta = 0$, то $x_t \beta = u_t \beta$. Следовательно, линейная комбинация $x_t \beta$ стационарна.

Легко сделать обобщение и на случай, когда имеется множество линейных соотношений между трендами. Если ранг коинтеграции r , то имеется $r < k$ линейных соотношений между трендами и можно написать:

$$\mu_t \beta = 0,$$

где β — матрица $(r \times k)$ с элементами β_{ij} .

Например, если имеются два независимых вектора коинтеграции для k переменных, то они могут быть представлены как

$$\begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1k} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2k} \end{bmatrix}.$$

Заметим, что если умножить вторую строку на β_{11}/β_{21} и вычесть ее из первой, то получим другую линейную комбинацию элементов вектора x_t , которая будет стационарной. Однако в этой комбинации будет только $k - 1$ ненулевых коэффициентов при x .

В более общем случае, когда имеется r векторов коинтеграции между k переменными, существует вектор коинтеграции для каждого подмножества из $(k-r)$ переменных.

§ 3. КОИНТЕГРАЦИЯ И МОДЕЛИ КОРРЕКТИРОВКИ ОШИБОК

В последние годы получил распространение класс моделей с распределенными лагами, названных моделями коррективы ошибок (Error correction models — ECM). Эти модели основываются на том факте, что часто между двумя переменными существует долговременное равновесное соотношение. Например, между потреблением и доходом, зарплатой и ценами и т.п. Однако в краткосрочном плане равновесие может нарушаться. Отклонение от равновесия в одном периоде корректируется в следующем. Например, изменение цены в одном периоде может зависеть от избыточного спроса в предыдущем. Процесс коррективы ошибок таким образом является средством согласования краткосрочной и долгосрочной динамики изучаемого процесса. Вклад в разработку этой темы сделали Д. Сарган [159], Д. Дэвидсон, Д. Хендри, Ф. Срба, С. Йо [136], Р. Энгл и К. Гренжер [142].

Предположим, что долгосрочное соотношение между Y_t и X_t имеет вид

$$Y_t = KX_t, \quad (13.5)$$

где K — константа.

В качестве примера Фридман сформулировал гипотезу о постоянном доходе, в которой утверждается, что потребление Y_t пропорционально постоянному доходу X_t . Он аппроксимировал постоянный доход процессом с распределенными лагами. Другой пример — это гипотеза о жизненном цикле, в которой утверждается, что в долгосрочном плане потребление есть постоянная доля богатства. Еще один пример: зарплаты и цены должны иметь примерно равные долгосрочные темпы роста, хотя в краткосрочном плане их темпы могут различаться.

Взяв логарифмы от обеих частей уравнения (13.5), получаем

$$\ln Y_t = \ln K + \ln X_t,$$

или

$$y_t = k + x_t, \quad (13.6)$$

где малые буквы обозначают логарифмы.

Поскольку $y_{t-1} = k + x_{t-1}$, то, вычитая это равенство из выражения (13.6), получаем

$$\Delta y_t = \Delta x_t. \quad (13.7)$$

В общем виде краткосрочная модель с лаговой коррективкой выглядит так:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \alpha_1 y_{t-1} + u_t. \quad (13.8)$$

Теперь установим, при каких условиях краткосрочная модель будет совпадать с долгосрочной. Для этого положим, что в установившемся состоянии $y_t = y^*$ и $x_t = x^*$ для всех t . Тогда уравнение (13.8), полагая, что в долгосрочном плане $u_t = 0$, принимает следующий вид:

$$y^*(1 - \alpha_1) = \beta_0 + (\beta_1 + \beta_2)x^*.$$

Для того чтобы это совпадало с уравнением (13.6), должно выполняться условие

$$1 - \alpha_1 = \beta_1 + \beta_2,$$

которое обеспечивает равенство

$$y^* = k^* + x^*,$$

где $k^* = \beta_0 / (1 - \alpha_1)$.

Обозначим $1 - \alpha_1 = \beta_1 + \beta_2 = \gamma$. Тогда $\alpha_1 = 1 - \gamma$ и $\beta_2 = \gamma - \beta_1$. Делая эти подстановки в (13.8), получаем

$$y_t = \beta_0 + \beta_1 x_t + (\gamma - \beta_1) x_{t-1} + (1 - \gamma) y_{t-1} + u_t$$

или

$$y_t - y_{t-1} = \beta_0 + \beta_1 (x_t - x_{t-1}) + \gamma (x_{t-1} - y_{t-1}) + u_t.$$

т.е.

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \gamma (x_{t-1} - y_{t-1}) + u_t. \quad (13.9)$$

Полученное уравнение представляет собой простейшую модель с корректировкой ошибки. Оно связывает изменение в одной переменной с изменениями в другой и с разрывом между двумя переменными в предыдущем периоде. Важно отметить, что уравнение содержит краткосрочную корректировку и в то же самое время основывается на долгосрочной теории. Член $x_{t-1} - y_{t-1}$ обеспечивает корректировку краткосрочного отклонения от равновесия, поэтому тест относительно γ есть тест для компоненты, отражающей нарушение равновесия.

В общем виде спецификация модели с корректировкой ошибки такая:

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \gamma_1 x_{t-1} + \gamma_2 y_{t-1} + u_t. \quad (13.10)$$

Однако эта общая формулировка не предполагает, что $\gamma_2 = -\gamma_1$. На практике мы должны это проверить, поэтому тест относительно γ_1 и γ_2 является тестом для члена корректировки отклонения от равновесия.

Наиболее общий вариант модели с корректировкой ошибки имеет еще одно, второе уравнение, связывающее изменения в x_t с изменениями в y_t и лаговыми значениями. Таким образом, имеем

$$\Delta x_t = \varphi_0 + \varphi_1 \Delta y_t + \theta_1 x_{t-1} + \theta_2 y_{t-1} + v_t$$

Построение модели корректировки ошибок тесно связано с изучением коинтеграции переменных.

Пример. Несмотря на то что модель с корректировкой ошибки ввел Филлипс, первым применил ее Дж. Сарган [159]. Он исследовал связь между зарплатой и ценами в Великобритании. Им сформулировано большое число моделей, некоторые из которых имели структуру модели с корректировкой ошибки.

В частности, на квартальных данных он оценил следующую модель (ниже даны t -статистики):

$$\begin{aligned} w_t - w_{t-1} = & -0,005(p_{t-1} - p_{t-4}) - 0,0143u_{t-1} - 0,395(w_{t-1} - p_{t-1}) + \\ & (-0,058) \quad (-2,234) \quad (-2,904) \\ & + 0,00085D_t + 0,00119t, \\ & (1,149) \quad (3,216) \end{aligned}$$

где $w_t = \ln(J_{w,t})$ — логарифм индекса зарплаты;

$p_t = \ln(J_{p,t})$ — логарифм индекса цен;

$u_t = \ln(R_{u,t})$ — логарифм доли безработных;

D_t — фиктивная переменная, равная 0 до конца 1954 г. и 1 — в дальнейшем (введена для учета замораживания зарплаты в раннем периоде).

Зависимая переменная представляет темп прироста уровня зарплаты, так как

$$w_t - w_{t-1} = \ln \left(\frac{J_{w,t}}{J_{w,t-1}} \right) \approx \tau_{w,t} - \text{темп прироста зарплаты за квартал.}$$

В качестве объясняющих переменных Дж. Сарган использовал темп прироста цен за целый год, а не за один квартал; переменная безработицы введена для отражения эффекта экономического цикла, в предположении, что в периоды высокой безработицы рост зарплаты не будет большим; член $(w_{t-1} - p_{t-1})$ — корректировка отклонения от равновесия; временной тренд включен для отражения роста зарплаты вследствие технического прогресса.

По оцененному уравнению видно, что темп прироста цен и фактор замораживания зарплаты незначимы, но член кор-

ректировки ошибки значим. Когда незначимые члены были опущены, оцененная модель приняла вид:

$$w_t - w_{t-1} = -0,0120u_{t-1} - 0,271(w_{t-1} - p_{t-1}) + 0,00133t \\ (-2,069) \quad (-3,712) \quad (3,694)$$

Значимость члена корректировки возросла. Отрицательный знак его коэффициента означает, что рост реальной зар-

платы $w_{t-1} - p_{t-1} = \ln \left(\frac{J_{w,t-1}}{J_{p,t-1}} \right)$ уменьшает темп прироста зар-
платы.

§ 4. КРИТЕРИИ КОИНТЕГРАЦИИ

Поскольку тип модели зависит от ответа на вопрос, является ли переменная (кандидат на фигурирование в качестве регрессанда) коинтегрированной с независимой переменной, необходима проверка коинтегрированности двух или большего числа переменных. Р. Энгл и К. Гренжер рассмотрели множество тестов на коинтегрированность. Мы рассмотрим два, рекомендованных ими.

Первый тест основывается на статистике Дарбина — Уотсона DW. Процедура предполагает сначала оценку уравнения

$$y_t = \alpha + \beta x_t + u_t$$

и вычисление обычной статистики Дарбина — Уотсона DW. Если y_t и x_t являются $I(1)$, то можно ожидать, что и u_t может оказаться $I(1)$. Если это так, то статистика DW будет близка к нулю и ряды y_t и x_t не коинтегрированы. Если же DW значимо больше нуля, то это может означать коинтегрированность двух переменных и $u_t \sim I(0)$. Однако стандартные таблицы для теста DW здесь не применимы, поскольку они построены для проверки нулевой гипотезы, что $DW = 2$, а не $DW = 0$. Энгл и Гренжер с помощью имитационного моделирования получили критические значения, представленные в табл. 13.1 для выборки в 100 наблюдений.

Таблица 13.1

Критические значения для проверки коинтегрированности,
 $n = 100$

Уровень значимости	Статистика DW	t-статистика расширенного критерия Дикки — Фуллера
0,01	0,511	3,77
0,05	0,386	3,17
0,10	0,322	2,84

Второй тест использует расширенный критерий Дикки — Фуллера. Он также предполагает сначала оценку уравнения $y_t = \alpha + \beta x_t + u_t$, получение ошибок

$$\hat{u}_t = y_t - \hat{\alpha} - \hat{\beta}x_t,$$

а затем оценивание регрессии Дикки — Фуллера

$$\Delta \hat{u}_t = \gamma \hat{u}_{t-1} + \sum_{i=1}^p \beta_i \Delta \hat{u}_{t-i} + \varepsilon_t,$$

где p — предварительно выбранный порядок лага для остатков.

Статистикой критерия является t-статистика для γ , но t-распределение не подходит. Для этого случая Р. Энгл и К. Гренжер получили критические значения с помощью имитационной процедуры. Для выборки в 100 наблюдений критические значения приведены в той же табл. 13.1.

* * *

Вопросы, рассмотренные в этой главе, имеют большое значение для правильной спецификации регрессионной модели, содержащей несколько объясняющих переменных. Если с содержательной точки зрения экономиста могут удовлетворить модели с различной структурой уравнения, то эконометрик должен уделять внимание сбалансированности динамических свойств статистических рядов, используемых в обеих частях равенства, и их связанности (коинтегрированности). В случае несбалансированности требуется то или иное предварительное преобразование данных, поиск новых переменных.

Глава 14

РЕКУРРЕНТНЫЕ АЛГОРИТМЫ ОЦЕНКИ ТРАЕКТОРИЙ ПАРАМЕТРОВ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

Линейная регрессионная модель, или просто регрессия, отражает связь между одной переменной y и несколькими другими переменными $x_2, \dots, x_p$ в виде линейного уравнения

$$y_t = \beta_1 + \beta_2 x_{2t} + \dots + \beta_p x_{pt} + u_t, \quad (14.1)$$

где t — номер наблюдения, $t = 1, 2, \dots, T$, T — число наблюдений или объем выборки;

β_i — параметры уравнения, регрессионные коэффициенты, $i = 1, 2, \dots, p$;

u_t — случайное возмущение, ошибка уравнения, остаток.

В матричных обозначениях эта модель записывается как

$$y = X\beta + u, \quad (14.2)$$

где

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_T \end{bmatrix}, X = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{p1} \\ x_{12} & x_{22} & \dots & x_{p2} \\ \dots & \dots & \dots & \dots \\ x_{1T} & x_{2T} & \dots & x_{pT} \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_p \end{bmatrix}, u = \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_T \end{bmatrix},$$

$$x_{1t} = 1 \text{ для всех } t.$$

В классическом регрессионном анализе параметры регрессионного уравнения считаются постоянными и их оценки

получают прямым вычислением по формуле метода наименьших квадратов

$$\hat{\beta} = (X'X)^{-1} X'y. \quad (14.3)$$

Ковариационная матрица вектора оценок вычисляется также по формуле

$$V(\hat{\beta}) = s^2 (X'X)^{-1}, \quad (14.4)$$

где $s^2 = \frac{e'e}{T-p}$ — оценка дисперсии ошибки уравнения;

$e = y - X\hat{\beta}$ — вектор ошибок.

При оценивании используются все выборочные наблюдения сразу. Однако предположение о постоянстве параметров уравнения в экономических исследованиях не всегда реалистично. В гл. 8 § 2 был предложен градиентный алгоритм адаптации регрессионных коэффициентов. Здесь мы рассмотрим рекуррентные методы оценивания параметров множественной регрессии.

§ 1. РЕКУРРЕНТНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ РЕГРЕССИИ

Для рекуррентного подхода характерно то, что информация в виде выборочных, упорядоченных во времени значений (временных рядов) обрабатывается последовательно. На каждом шаге рекуррентных вычислений полученные на предыдущем шаге оценки обновляются с учетом новой порции данных, поступивших за очередную единицу времени. Это означает, что при рекуррентных вычислениях статистическая база оценивания расширяется вместе с продвижением модели вперед по оси времени. При одинаковой статистической базе рекуррентные оценки эквиваленты оценке «по формуле». Достоинством рекуррентных вычислений является то, что в ходе их выполнения будут получены оценки параметров для промежуточных моментов. Эти оценки, отложенные на графике вместе со своими доверительными интервалами и дисперсией ос-

татков, дают дополнительную полезную информацию для размышления относительно устойчивости (изменчивости) оценок и наилучшего объема выборки.

Рассмотрим теоретические основания рекуррентной процедуры оценивания регрессии [160, 124]. Запишем множественную линейную регрессию в векторной форме для момента t :

$$y_t = x_t' \beta + u_t, \quad t = 1, 2, \dots, T \quad (14.5)$$

со стандартными предположениями, что

- β — вектор постоянных коэффициентов;
- $x_t = [x_{t1}, \dots, x_{tp}]'$;
- u_t — независимые ошибки с нулевым математическим ожиданием и постоянной дисперсией σ^2 .

Оценку МНК (14.3) можно переписать в векторной форме как

$$\hat{\beta} = \left[\sum_{t=1}^T x_t x_t' \right]^{-1} \left(\sum_{t=1}^T x_t y_t \right). \quad (14.6)$$

Нам необходимо построить рекуррентную процедуру оценивания. В рекуррентном алгоритме целью является получение оценки $\hat{\beta}_t$ на момент t как функции от предшествующей оценки $\hat{\beta}_{t-1}$ и значений переменных, наблюдаемых в момент t . Рекуррентную оценку в общем случае будем представлять в виде произведения двух рекуррентных величин, одна из которых содержит операцию обращения:

$$\hat{\beta}_t = F_t g_t, \quad t = p, \dots, T, \quad (14.7)$$

где

$$F_t = \left[\sum_{i=1}^t x_i x_i' \right]^{-1}; \quad (14.8)$$

$$g_t = \sum_{i=1}^t x_i y_i. \quad (14.9)$$

Выражения (14.8) и (14.9) означают, что

$$F_t^{-1} = F_{t-1}^{-1} + x_t x_t', \quad (14.10)$$

$$g_t = g_{t-1} + x_t y_t. \quad (14.11)$$

Как видим, рекуррентное обновление g_t не вызывает затруднений. Обновление же F_t требует на каждом шаге обращения матрицы размерности $p \times p$. Существует более экономный с вычислительной точки зрения алгоритм, в котором удастся избежать прямого обращения. Рассмотрим этот алгоритм подробнее.

Умножив выражение (14.10) слева на F_t и справа на F_{t-1} , получаем

$$F_t = F_{t-1} - F_t x_t x_t' F_{t-1}. \quad (14.12)$$

Умножим выражение (14.12) справа на x_t и соберем члены с $F_t x_t$ вместе. Затем после умножения справа на $(1 + x_t' F_{t-1} x_t)^{-1} x_t' F_{t-1}$ и подстановки $F_t x_t x_t' F_{t-1} = F_{t-1} - F_t$ из уравнения (14.12), приходим к равенству, часто называемому *леммой об обращении матрицы*

$$F_t = F_{t-1} - F_{t-1} x_t (1 + x_t' F_{t-1} x_t)^{-1} x_t' F_{t-1}. \quad (14.13)$$

Поскольку $1 + x_t' F_{t-1} x_t$ — скаляр, операция обращения не представляет никаких сложностей, и рекуррентная процедура с вычислительной точки зрения будет иметь преимущество по сравнению с решением по формуле, где требуется матричное обращение на каждом шаге продвижения во времени.

Запишем алгоритм в собранном виде. Введем вектор

$$k_t = F_{t-1} x_t (1 + x_t' F_{t-1} x_t)^{-1}. \quad (14.14)$$

Тогда формула (14.13) переписывается как

$$F_t = F_{t-1} - k_t x_t' F_{t-1}. \quad (14.15)$$

Сравнивая равенство (14.12) с выражением (14.15), получаем

$$k_t = F_t x_t, \quad (14.16)$$

а из равенства (14.7) и (14.11)

$$\beta_t = F_t g_t = F_t (g_{t-1} + x_t y_t) = F_t g_{t-1} + F_t x_t y_t \quad (14.17)$$

Используя выражения (14.7), (14.15) и подставляя формулу (14.16) в (14.17), приходим к выводу, что оценка равна:

$$\beta_t = \beta_{t-1} + k_t (y_t - x_t \beta_{t-1}). \quad (14.18)$$

Эта формула рекуррентного оценивания имеет ясную интерпретацию. Она означает, что оценка вектора параметров в момент t равна предыдущей оценке β_{t-1} плюс корректирующий член, являющийся произведением множителя k_t на ошибку уравнения. Рекуррентная оценка фактически выражает существо байесовского подхода в том смысле, что при известной априорной оценке β_{t-1} дает возможность получить апостериорную оценку β_t .

Выражения (14.13), (14.16) и (14.18) образуют рекуррентный алгоритм МНК для оценивания коэффициентов линейной множественной регрессии общего вида.

Нетрудно вычислить и дисперсии ошибок рекуррентных оценок. При прямом решении по формуле при принятых допущениях оценке МНК соответствует ковариационная матрица (14.4), которая в векторном виде может быть записана как

$$V(\hat{\beta}) = s^2 \left[\sum_{t=1}^T x_t x_t' \right]^{-1}.$$

Таким образом, при выборке объема t

$$V_t = \text{cov}(\beta_t) = s_t^2 F_t. \quad (14.19)$$

В качестве оценки s_t^2 можно использовать величину

$$s_t^2 = \frac{1}{t-p} \sum_{i=1}^t (x_i \beta_i - y_i)^2. \quad (14.20)$$

При условии гауссовского распределения остатков полученные оценки могут быть интерпретированы как оценки максимального правдоподобия или как оценки байесовского типа.

Легко также построить рекуррентный алгоритм для получения оценок обобщенным методом наименьших квадратов (ОМНК). Как известно, этот метод применяется в случае гетероскедастичности остатков, т. е. тогда, когда дисперсия остатков не постоянна. Тогда оценка ОМНК может быть получена с помощью обычного МНК, примененного к преобразованным данным. Следовательно, и для рекуррентного оценивания требуются лишь переход к новым переменным и использование только что изложенной процедуры.

§ 2. СКОЛЬЗЯЩАЯ РЕГРЕССИЯ

Одним из способов проверки постоянства параметров регрессии состоит в оценке регрессии на последовательно сдвигаемом во времени интервале постоянной длины и построении траекторий оценок коэффициентов вместе с их доверительными интервалами. Такую процедуру можно назвать *скользящей регрессией* по аналогии со *скользящей средней* (в англоязычной литературе ее называют *Rolling regression*).

Пусть, например, у нас имеется выборка достаточно большого объема T и определена спецификация уравнения регрессии y на $x_1, x_2, \dots, x_p$. Тогда процедура оценивания скользящей регрессии предполагает сначала оценку регрессии на выборке от наблюдения 1 до m , где $m \ll T$. Величину m можно назвать *окном* или *скользящим интервалом оценивания*. Далее сдвигаем «окно» на одну единицу времени вперед и вновь оцениваем регрессию на выборочном интервале от наблюдения 2 до $m+1$ и т.д. Полученные оценки вместе с их доверительными границами можно отложить на графике и изучать особенности движения параметров регрессии и ширину доверительных интервалов.

Такой подход позволяет вскрыть характер движения параметров, принять реалистичные гипотезы относительно моментов сдвига и типа эволюции регрессионных коэффициентов, произвести периодизацию изучаемого явления, а в

случае необходимости перейти к кусочно-линейным регрессиям или сплайнам. В этой процедуре на каждой итерации параметры регрессии могут оцениваться как константы обычным МНК. Однако такой прием не очень эффективен с вычислительной точки зрения, поскольку на каждой итерации требуется обращение матрицы. Гораздо более эффективен рекуррентный подход [117, 124, 161]. Поскольку скользящую регрессию можно рассматривать как частный случай взвешенной регрессии, то сейчас мы перейдем к взвешенной регрессии, а затем в § 4 будет изложен рекуррентный метод оценивания параметров взвешенной регрессии.

§ 3. ВЗВЕШЕННАЯ РЕГРЕССИЯ

В быстро изменяющейся экономике уравнения и параметры взаимосвязи различных факторов, характерные для прошлого периода, перестают быть адекватными в новых условиях. На смену старым технологиям приходят новые, изменяются как материальные, так и финансовые потоки ресурсов и результатов производства, происходит структурная перестройка экономики и финансовых рынков, меняются законы, уровень инфляции, степень риска, налогообложение и т.п. Такие изменения, конечно же, желательно учесть в модели в явном виде, особенно, когда речь идет о прогнозировании. Однако это не всегда или не сразу возможно, поскольку новое постепенно вырастает из старого.

Перед эконометриком встает сложная и противоречивая задача отразить в регрессионной модели новейшие тенденции изучаемого явления, но при этом опираться на достаточный объем статистических данных, с тем чтобы получить статистически значимые оценки регрессионных коэффициентов. Выходом из положения является компромисс, в соответствии с которым сохраняется старая форма регрессионного уравнения с постоянными коэффициентами, но ее несовершенство отчасти компенсируется неравноценным отношением к ошибкам уравнения в начале и в конце выборочного периода: старые ошибки имеют меньший вес, а ошибки последних моментов — больший. В этом случае оценки регрессионных коэффициентов в модели

$$y_t = \sum_i \beta_i x_{it} + v_t \quad (14.21)$$

получают минимизацией взвешенной суммы квадратов остатков

$$Q = \sum_{t=1}^T w_t e_t^2, \quad (14.22)$$

где w_t = веса;

$$e_t = y_t - \hat{y}_t;$$

$$\hat{y}_t = \sum_i \hat{\beta}_i x_{it}.$$

Такой метод оценивания называется *взвешенным*, или *дисконтированным МНК*. Введение критерия Q означает, что статистическим данным, относящимся к разным моментам времени, приписывается различная информационная ценность. Веса являются функцией времени. Эта функция характеризует «память» модели. Если предполагается, что изучаемый процесс претерпевает быстрые изменения, то и весовая функция должна быстро убывать (затухать) при движении от текущего момента t в прошлое.

Для того чтобы получить оценки $\hat{\beta}_i$, минимизирующие взвешенную сумму квадратов остатков, нужно последовательно продифференцировать (14.22) по параметрам и частные производные приравнять нулю. Сделаем это, используя для наглядности сначала скалярные обозначения:

$$\frac{\partial Q}{\partial \beta_i} = -2 \sum_{t=1}^T w_t (y_t - \sum_i \beta_i x_{it}) = 0, \quad i = 1, 2, \dots, p.$$

Отсюда получим систему линейных уравнений относительно коэффициентов β_i :

$$\left. \begin{aligned} \beta_1 \sum_{t=1}^T w_t x_{1t}^2 + \dots + \beta_p \sum_{t=1}^T w_t x_{1t} x_{pt} &= \sum_{t=1}^T w_t y_t x_{1t} \\ \beta_1 \sum_{t=1}^T w_t x_{1t} x_{2t} + \dots + \beta_p \sum_{t=1}^T w_t x_{2t} x_{pt} &= \sum_{t=1}^T w_t y_t x_{2t} \\ &\vdots \\ \beta_1 \sum_{t=1}^T w_t x_{1t} x_{pt} + \dots + \beta_p \sum_{t=1}^T w_t x_{pt}^2 &= \sum_{t=1}^T w_t y_t x_{pt} \end{aligned} \right\} \quad (14.23)$$

В матричных обозначениях получаем

$$X'WX\beta = X'Wy,$$

где

$$W = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_T \end{bmatrix} - \text{матрица весов.}$$

Отсюда находим оценку вектора коэффициентов

$$\hat{\beta}_w (X'WX)^{-1} X'Wy. \quad (14.24)$$

Легко заметить, что, если положить $B = W^{-1}$, то эту оценку можно рассматривать как оценку обобщенным МНК (ОМНК) регрессии

$$y_t = \sum_i \beta_i x_{it} + u_t, \quad (14.25)$$

где

$$M(uu') = \sigma^2 B. \quad (14.26)$$

Таким образом, оценивание линейной регрессии (14.21) по критерию (14.22) эквивалентно оцениванию регрессии с преобразованными переменными

$$\sqrt{w_t} y_t = \sum_{i=1}^p \beta_i (\sqrt{w_t} x_{it}) + v_t \quad (14.27)$$

обычным МНК.

Проанализируем характер остатков u_t взвешенной регрессии (14.25). Если веса в матрице W уменьшаются от момента $t = T$ к $t = 1$, то дисперсия $\sigma^2 B$ ошибки u_t в уравнении (14.25) будет возрастать в этом направлении таким образом, что в начальные моменты допускаются большие ошибки.

Вообще говоря, весовая функция может быть любой функцией времени. Но в целях упрощения работы в качестве весовой функции выбирают какую-либо известную аналитическую функцию с одним параметром, удовлетворяющую качественным представлениям исследователя о темпах и характере перемен.

Двумя такими функциями могут быть арифметическая прогрессия, генерирующая веса, лежащие на графике на прямой линии, и убывающая в прошлое геометрическая прогрессия, обеспечивающая экспоненциальное падение величины весового коэффициента. Могут быть, конечно, и другие функции, например ступенчатая или логистическая.

Отметим, что особой популярностью пользуются веса, геометрически убывающие в прошлое [161]. В этом случае весовой коэффициент имеет вид

$$w_t = q^{T-t}, \quad (14.28)$$

где $0 < q < 1$,

а минимизируется взвешенная сумма квадратов остатков:

$$Q = \sum_{t=1}^T q^{T-t} (y_t - \sum_{i=1}^p \hat{\beta}_{wi} x_{it})^2. \quad (14.29)$$

Суммы смешанных произведений переменных в системе линейных уравнений (14.23) тоже оказываются взвешенными экспоненциально:

$$\sum_{t=1}^T w_t x_{it} x_{jt} = \sum_{t=1}^T q^{T-t} x_{it} x_{jt}. \quad (14.30)$$

К этому интересному свойству мы еще вернемся в § 5. Что касается выбора величины q_t и вообще w_t , то они считаются здесь заданными или известными. Вопрос о выборе оптимальных весов будет рассмотрен в § 4.

§ 4. РЕКУРРЕНТНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ ВЗВЕШЕННОЙ РЕГРЕССИИ

Альтернативой оценивания коэффициентов взвешенной регрессии по формуле (14.24) является рекуррентное оценивание. Рекуррентное оценивание параметров взвешенной регрессии при известных весовых функциях сводится к задаче рекуррентного оценивания обобщенным МНК или применению рекуррентного алгоритма обычного МНК к преобразованному данным. Более подробно мы остановимся на рекуррентном оценивании взвешенной регрессии лишь для двух часто используемых весовых функций.

В общем виде взвешенная регрессия имеет тот же вид, что и обычная линейная регрессия

$$y_i = x_i' \beta + u_i, \quad (14.31)$$

но предполагается, что коэффициенты β_i , $i = 1, 2, \dots, p$, могут иметь некоторое движение, например, с целью компенсации определенного упрощения взаимосвязей, приближенно выражаемых уравнением (14.31). Ожидается, что это движение параметров можно с достаточной точностью уловить путем минимизации взвешенной суммы квадратов регрессионных остатков

$$Q_t = \sum_{i=1}^t w_i (y_i - x_i' \hat{\beta}_t)^2, \quad t = p, \dots, T. \quad (14.32)$$

Рекуррентное оценивание скользящей регрессии. На практике часто используется регрессия, в которой для обновления оценок параметров в расчет всегда принимаются только m последних точек:

$$Q_{m,t} = \sum_{i=t-m+1}^t (y_i - x_i' \hat{\beta}_t)^2. \quad (14.33)$$

Этот вид регрессии, которую можно назвать скользящей регрессией, был рассмотрен в § 2. Ее можно представить в

виде взвешенной регрессии и критерий Q записать как выражение (14.32) с весами

$$w_i = \begin{cases} 1, & \text{если } i \in [t-m+1, t] \\ 0, & \text{в остальных случаях} \end{cases} \quad (14.34)$$

Таким образом, для вычисления функции потерь на каждом шаге используется «окно» шириной m , передвигаемое по оси времени, и только данные, попавшие в это «окно», используются для вычисления критерия Q_t . Остальные данные считаются невидимыми или несущественными и в расчете оценок не участвуют.

Получить рекуррентный алгоритм для этой весовой функции можно в два этапа, представив продвижение «окна» на один шаг вперед в виде двух событий. Сначала добавим новую выборочную точку, т.е. расширим «окно» вправо, и с помощью формул (14.11), (14.13) и (14.17) получим промежуточные величины F_t^* и $\hat{\beta}_t^*$, а затем с помощью тех же самых уравнений устраним устаревшую, ставшую несущественной точку $t-m$. При этом мы лишь меняем знак «плюс» на «минус», а x_{t-m} , y_{t-m} — на x_{t-m}^* , y_{t-m}^* и находим F_t и $\hat{\beta}_t$.

На основе изложенной процедуры полезно проводить анализ ошибок прогнозов, получаемых с помощью скользящей регрессии с различной длиной скользящего отрезка (окна). Для этого можно вычислить средний квадрат ошибок прогнозирования на один шаг вперед Q_1 для различных m

$$Q_1(m) = \frac{1}{T-m} \sum_{r=m+1}^T [y_r - x_r' \hat{\beta}(r-m, r-1)]^2, \quad (14.35)$$

где $\hat{\beta}(r-m, r-1)$ — оценка вектора регрессионных коэффициентов на временном отрезке $[r-m, r-1]$. Величина m , при которой Q_1 достигает минимального значения, указывает на оптимальный объем выборки для целей прогнозирования с помощью регрессионной модели данной спецификации.

Целесообразно также вычислять величину

$$Q_2(m) = \frac{1}{T-n} \sum_{r=n+1}^T [y_r - x_r' \hat{\beta}(r-m, r-1)]^2, \quad (14.36)$$

где n — максимальная длина рассматриваемых скользящих отрезков (ширина «окна»). Величина Q_2 отличается от Q_1 лишь тем, что при ее вычислении принимаются во внимание ошибки, полученные для одного и того же периода от $r = n$ до T .

Рекуррентное оценивание регрессии с геометрически убывающими весами. Другой часто используемой весовой функцией является

$$w_i = \alpha(1 - \alpha)^{t-i}, \quad (14.37)$$

где $0 \leq \alpha \leq 1, i = 1, 2, \dots, t$.

Из выражения (14.32) следует, что умножение всех весовых коэффициентов на одну и ту же константу не меняет значений оценок параметров, при которых Q_1 достигает минимума. Смысл множителя α в выражении (14.37) состоит в том, чтобы нормализовать веса таким образом, чтобы в сумме они давали 1 при $t \rightarrow \infty$.

Решение по формуле для выборки объема t находится из уравнения

$$\left[\sum_{i=1}^t w_i x_i x_i' \right] \hat{\beta}_t - \sum_{i=1}^t w_i x_i y_i = 0. \quad (14.38)$$

Для построения рекуррентного алгоритма запишем оценку $\hat{\beta}_t$ в виде

$$\hat{\beta}_t = F_t g_t, \quad (14.39)$$

где

$$F_t^{-1} = (1 - \alpha) F_{t-1}^{-1} + \alpha x_t x_t', \quad (14.40)$$

и

$$g_t = (1 - \alpha) g_{t-1} + \alpha x_t y_t. \quad (14.41)$$

Справедливость такого представления оценки нетрудно показать с помощью повторных подстановок. Отметим, что при $\alpha = 1$ прошлые данные полностью игнорируются («забываются») и имеют значение только текущие значения переменных, а при $\alpha = 0$ начальная оценка остается неизменной, все последующие наблюдения не принимаются во внимание.

Та же последовательность переходов, которую мы применяли раньше (в § 1) при выводе рекуррентного алгоритма МНК, приводит в данном случае к рекуррентному алгоритму, состоящему из аналогичных уравнений:

$$F_t = \frac{1}{1 - \alpha} F_{t-1} - \frac{\alpha}{(1 - \alpha)^2} F_{t-1} x_t x_t' \times \left(1 + \frac{\alpha}{1 - \alpha} x_t' F_{t-1} x_t \right)^{-1} x_t' F_{t-1}; \quad (14.42)$$

$$k_t = \frac{1}{1 - \alpha} F_{t-1} x_t \left(1 + \frac{\alpha}{1 - \alpha} x_t' F_{t-1} x_t \right)^{-1}; \quad (14.43)$$

$$k_t = F_t x_t; \quad (14.44)$$

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \alpha k_t (y_t - x_t' \hat{\beta}_{t-1}). \quad (14.45)$$

Начальные оценки $\hat{\beta}_p, F_p$ и g_p при $t = p$ находят по формулам (14.7), (14.8) и (14.9) для МНК без учета весов, а далее продолжают рекуррентные вычисления с учетом весов. Выбор оптимального α может быть сделан, например, методом проб различных значений, взятых из интервала $[0, 1]$, и сравнения соответствующих сумм квадратов ошибок прогнозирования на один шаг вперед. Наилучшему α соответствует наименьшая из этих сумм.

§ 5. ОЦЕНКА ТРАЕКТОРИЙ ПАРАМЕТРОВ РЕГРЕССИИ МЕТОДОМ АДАПТИВНЫХ КОВАРИАЦИЙ

Рассмотрим еще один метод последовательной корректировки параметров линейной множественной регрессии с помощью экспоненциально-взвешенных скользящих средних, предложенный нами в [149, 150, 124]. Будем считать, что коэффициенты регрессии являются переменными во времени, т.е.

$$y_t = \sum_{i=1}^p \beta_{it} x_{it} + u_t. \quad (14.46)$$

Однако сначала рассмотрим оценку МНК вектора параметров регрессии с постоянными коэффициентами и запишем ее в несколько преобразованном виде:

$$\hat{\beta} = (X'X)^{-1}X'y = \left[\frac{1}{T}(X'X) \right]^{-1} \left[\frac{1}{T}(X'y) \right] =$$

$$= \begin{bmatrix} g_{11} & g_{12} & \dots & g_{1p} \\ g_{21} & g_{22} & \dots & g_{2p} \\ \vdots & \vdots & \dots & \vdots \\ g_{p1} & g_{p2} & \dots & g_{pp} \end{bmatrix}^{-1} \begin{bmatrix} g_{1,p+1} \\ g_{2,p+1} \\ \vdots \\ g_{p,p+1} \end{bmatrix} = G^{-1}g, \quad (14.47)$$

где

$$g_{ij} = \frac{1}{T} \sum_{t=1}^T x_{it} x_{jt}, \quad i = 1, 2, \dots, p; \quad j = 1, 2, \dots, p+1.$$

Здесь для общности обозначений индекс $p+1$ присвоен эндогенной переменной y , т.е. $x_{p+1,t} \equiv y_t$. Обратим внимание на величину g_{ij} . Ключ к практическому решению проблемы построения адаптивной регрессии (14.46) состоит, очевидно, в том, чтобы найти хороший способ обновления оценок сред-

них значений g_{ij} . Другими словами, мы предлагаем выборочные средние g_{ij} заменить на $g_{ij,t}$ — текущие оценки локальных средних значений. Получить $g_{ij,t}$ можно различными способами. Например, в скользящей регрессии, т.е. в регрессии, оцениваемой МНК на последовательно сдвигаемой отрезке времени, $g_{ij,t}$ фактически оцениваются как скользящие средние. Можно также отыскивать тренды смешанных произведений и аппроксимировать их известными функциями времени или какими-либо моделями. Однако, по-видимому, лучше заранее исключить тренд из переменных (регрессанда и регрессоров), переходя к разностям, цепным индексам, темпам приростов и т.п. с учетом порядков интегрированности переменных.

Таким образом, в рассматриваемом методе в общем виде задача сводится к изучению динамики средних значений попарных произведений наблюдаемых переменных. Эти произведения образуют временные ряды, которые можно откладывать на графике и анализировать визуально или с помощью математических средств. Из уравнения (14.46) видно, что ступенчатый скачок параметра β , непременно найдет отражение в уровне эндогенной переменной y , а следовательно, и во всех $g_{i,p+1}$, $i = 1, 2, \dots, p$, в которых она участвует. Совокупность таких графиков позволяет раскрыть структуру и динамику процесса, локализовать точки, в которых подозревается сдвиг. В целом уже на стадии анализа многомерная задача построения множественной регрессии разлагается на $p(p+3)/2$ одномерных. Отметим, однако, что указанный одномерный анализ нельзя проводить изолированно от аналогичных параллельных задач. Все эти одномерные задачи должны быть согласованы, подчинены одной общей конечной цели, одному критерию. Промежуточные, частные или косвенные критерии могут уводить от цели вследствие рассогласованности отдельных частей модели, приводить к «раскачиванию» ее как целого.

Перейдем теперь к практическому построению адаптивной регрессии. Будем полагать, что все переменные, участвующие в модели, не имеют ярко выраженного тренда, т.е. их средний уровень подвержен лишь слабой эволюции. Тогда предлагается в формуле (14.47) заменить арифметические средние g_{ij} на экспоненциально-взвешенные скользящие средние $g_{ij,t}$. В этом случае обновление g_{ij} будет последовательно производиться по формуле

$$g_{ij,t} = (1-\alpha)g_{ij,t-1} + \alpha(x_{it}x_{jt}), \quad i=1, 2, \dots, p; \quad j=1, 2, \dots, p+1, \quad (14.48)$$

где α — постоянная сглаживания, величина которой находится в интервале $0 < \alpha < 1$.

Далее по формуле (14.47) с $g_{ij,t}$ вместо g_{ij} можно определить текущие оценки коэффициентов β_i в момент t . Для начала вычислений по формуле рекуррентного типа (14.48) необходимо задать некое начальное значение $g_{ij,0}$ для момента $t=0$. Его можно определить, например, как простое арифметическое среднее на основе первых T_0 членов выборки, т.е.

$$g_{ij,0} = \frac{1}{T_0} \sum_{k=1}^{T_0} x_{ik}x_{jk}. \quad (14.49)$$

Для упрощения дела для всех пар i и j будем задавать T_0 и параметр сглаживания α одинаковыми. Но для того чтобы найти их оптимальные значения, необходимо задать критерий качества модели.

Будем считать ту модель лучшей, на основе которой на выборочном периоде получают более точные прогнозы на один шаг вперед. Ошибка прогноза равна:

$$e_{t+1} = y_{t+1} - \hat{y}_{t+1}, \quad (14.50)$$

где $\hat{y}_{t+1} = x_{t+1} \hat{\beta}_t$

$\hat{\beta}_t$ — оценка вектора параметров в момент t .

В задаче поиска оптимальных значений параметров совокупным показателем погрешности модели будем считать сумму квадратов ошибок прогнозов:

$$Q(T_0, \alpha) = \sum_{t=1}^T e_t^2, \quad (14.51)$$

зависящую от параметров T_0 и α . Поиск минимума суммы квадратов ошибок можно производить итеративно. Параметр T_0 может принимать только дискретные значения в

ограниченном диапазоне $0 < T_0 \leq T$. Поэтому его оптимальное значение легко найти перебором. При этом для каждого значения T_0 следует определить оптимальное значение α путем разбиения отрезка $[0, 1]$ на сетку значений и поиска минимума $Q(T_0, \alpha)$ на этом множестве. Постепенно переходя от грубой сетки к более частой, оценку α можно получить с любой желательной точностью. Этим построение адаптивной процедуры обновления регрессионных коэффициентов завершается.

Весьма важное свойство изложенной процедуры следует из сопоставления ее с взвешенной регрессией, рассмотренной нами в § 3 данной главы. Там установлено, что если при оценивании регрессии минимизировать взвешенную сумму квадратов ошибок с весами, геометрически убывающими в прошлое, то коэффициентами при β_i в системе нормальных уравнений (14.23) будут суммы экспоненциально-взвешенных произведений регрессионных переменных, взятых попарно, см. (14.30). Но из выражения (14.48) вытекает, что

$$g_{ij,t} = \alpha \sum_{k=1}^t (1-\alpha)^{t-k} x_{ik}x_{jk}. \quad (14.52)$$

Это означает, что, положив в (14.23) веса

$$w_k = \alpha(1-\alpha)^{t-k}, \quad k = 1, 2, \dots, t, \quad (14.53)$$

где t — текущий момент времени, на который определяется адаптивная оценка, можно прийти к выводу, что в каждый момент t оценка $\hat{\beta}_t$, полученная с помощью адаптивной регрессии, построенной на основе экспоненциально-сглаженных попарных произведений (ковариаций), минимизирует экспоненциально-взвешенную сумму квадратов ошибок:

$$Q = \sum_{k=1}^t (y_k - \sum_{i=1}^p \hat{\beta}_{ik}x_{ik})^2 (1-\alpha)^{t-k}. \quad (14.54)$$

Таким образом, поиск оптимального значения α означает выбор геометрически убывающих весов, наилучшим образом отражающих обесценение информации с течением времени, весов, доставляющих минимум функции потерь (14.51).

Проверку метода на примерах можно найти в [124, с. 226–229].

§ 6. СВЯЗЬ АДАПТИВНОЙ РЕГРЕССИИ С АДАПТИВНЫМ КОРРЕЛЯЦИОННЫМ АНАЛИЗОМ

Связь коэффициентов адаптивной регрессии с адаптивными коэффициентами корреляции будем исследовать в предположении, что уравнение регрессии записано для первых разностей переменных

$$\Delta y_t = \sum_{j=1}^p \beta_{jt} \Delta x_{jt} + u_t, \quad t = 2, 3, \dots, T. \quad (14.55)$$

В соответствии с алгоритмом адаптивной регрессии текущая оценка вектора параметров равна:

$$\hat{\beta}_t = G_t^{-1} g_t, \quad (14.56)$$

где

$$G_t = \begin{bmatrix} g_{11,t} & g_{12,t} & \dots & g_{1p,t} \\ g_{21,t} & g_{22,t} & \dots & g_{2p,t} \\ \dots & \dots & \dots & \dots \\ g_{p1,t} & g_{p2,t} & \dots & g_{pp,t} \end{bmatrix},$$

$$g_t = (g_{1,p+1,t}, \dots, g_{p,p+1,t})',$$

$$g_{ij,t} = (1 - \alpha)g_{ij,t-1} + \alpha(\Delta x_{ij} \Delta x_{jt}), \quad i = 1, 2, \dots, p; \\ j = 1, 2, \dots, p + 1,$$

$$\Delta x_{p+1,t} \equiv \Delta y_t,$$

α — постоянная сглаживания, $0 < \alpha < 1$.

Формула (14.56) означает, что оценка $\hat{\beta}_t$ является решением системы линейных уравнений

$$\left. \begin{aligned} \hat{\beta}_{1t} g_{11,t} + \hat{\beta}_{2t} g_{12,t} + \dots + \hat{\beta}_{pt} g_{1p,t} &= g_{1,p+1,t} \\ \hat{\beta}_{1t} g_{21,t} + \hat{\beta}_{2t} g_{22,t} + \dots + \hat{\beta}_{pt} g_{2p,t} &= g_{2,p+1,t} \\ \dots &\dots \\ \hat{\beta}_{1t} g_{p1,t} + \hat{\beta}_{2t} g_{p2,t} + \dots + \hat{\beta}_{pt} g_{pp,t} &= g_{p,p+1,t} \end{aligned} \right\} \quad (14.57)$$

Для того чтобы установить связь оценок регрессионных параметров $\hat{\beta}_{jt}$ с адаптивными коэффициентами корреляции (см. гл. 9), перепишем систему (14.57) в преобразованном виде

$$\left. \begin{aligned} \hat{\beta}_{1t} \frac{g_{11,t}}{d_{11,t}} + \dots + \hat{\beta}_{pt} \frac{g_{1p,t}}{d_{1p,t}} &= \frac{g_{1,p+1,t}}{d_{1,p+1,t}} \\ \hat{\beta}_{1t} \frac{g_{21,t}}{d_{21,t}} + \dots + \hat{\beta}_{pt} \frac{g_{2p,t}}{d_{2p,t}} &= \frac{g_{2,p+1,t}}{d_{2,p+1,t}} \\ \dots &\dots \\ \hat{\beta}_{1t} \frac{g_{p1,t}}{d_{p1,t}} + \dots + \hat{\beta}_{pt} \frac{g_{pp,t}}{d_{pp,t}} &= \frac{g_{p,p+1,t}}{d_{p,p+1,t}} \end{aligned} \right\} \quad (14.58)$$

где $d_{ij,t} = (1 - \alpha)d_{ij,t-1} + \alpha|\Delta x_{ij} \Delta x_{jt}|$, $i = 1, 2, \dots, p$; $j = 1, 2, \dots, p + 1$.

Величины $\frac{g_{ij,t}}{d_{ij,t}}$ являются адаптивными коэффициентами корреляции $r_{ij,t}(\alpha)$, которые мы будем обозначать просто как $r_{ij,t}$. Поэтому систему (14.58) можно переписать как

$$\left. \begin{aligned} \hat{\beta}_{1t} r_{11,t} d_{11,t} + \dots + \hat{\beta}_{pt} r_{1p,t} d_{1p,t} &= r_{1,p+1,t} d_{1,p+1,t} \\ \dots &\dots \\ \hat{\beta}_{1t} r_{p1,t} d_{p1,t} + \dots + \hat{\beta}_{pt} r_{pp,t} d_{pp,t} &= r_{p,p+1,t} d_{p,p+1,t} \end{aligned} \right\} \quad (14.59)$$

Система (14.59) выражает связь адаптивных регрессионных коэффициентов с адаптивными корреляционными коэффициентами.

Отметим, что в системе (14.59) подразумевается, что и в регрессионном, и в корреляционном анализе используется одно и то же значение α , выбираемое таким образом, чтобы наилучшим образом согласовать оценивание адаптивных коэффициентов корреляции для различных пар переменных с главной целью адаптивных вычислений — достижения наивысшей точности ретроспективных прогнозов, измеряемой критерием $Q(T_0, \alpha)$, представленным формулой (14.51).

§ 7. МОДЕЛИ С АВТОРЕГРЕССИОННОЙ УСЛОВНОЙ ГЕТЕРОСКЕДАСТИЧНОСТЬЮ

В эконометрическом моделировании неопределенность взаимосвязей переменных отражается дисперсией остаточного члена регрессии u_t . Признавая важность этого показателя, его влияние на те или иные экономические переменные, эконометриками предлагается исследовать его движение во времени с помощью авторегрессионных моделей.

Модель ARCH

Простейшей моделью является модель с авторегрессионной условной гетероскедастичностью (AutoRegressive Conditional Heteroscedasticity — ARCH), предложенная Р. Энглom [141]. В этой модели безусловная дисперсия $M(u_t^2) = \text{var}(u_t)$ является постоянной, а условная дисперсия $M(u_t^2 | I_{t-1})$, где I_{t-1} — совокупностью информации, известной на момент $t-1$, включающей значения x_t и лаговые значения y_t и x_t , т.е. $I_{t-1} = (x_t, x_{t-1}, \dots, y_{t-1}, y_{t-2}, \dots)$, изменяется во времени.

Обозначая условную дисперсию через h_t^2 , в простейшем случае модель ARCH можно записать как

$$\begin{aligned} y_t &= \beta_1 + \beta_2 x_{2t} + u_t; \\ h_t^2 &= \text{var}(u_t) = \alpha_0 + \alpha_1 u_{t-1}^2, \end{aligned}$$

где $\alpha_1 > 0$.

Последнее уравнение означает следующее: если ошибка в предыдущий момент была большой, то и дисперсия текущей ошибки будет увеличенной. Если же $\alpha_1 = 0$, то это означает, что эффекта ARCH не наблюдается, дисперсия постоянна, т.е. имеет место гомоскедастичность и можно применять обычные методы оценивания (МНК). При $\alpha_1 \neq 0$ для оценивания модели используются более сложные методы максимального правдоподобия.

Проверка нулевой гипотезы $H_0: \alpha_1 = 0$ выполняется с помощью оценивания МНК первого уравнения, получения остатков $\hat{u}_t^2$, оценивания регрессии квадратов остатков $\hat{u}_t^2$ на $\hat{u}_{t-1}^2$ (с постоянным членом) и проверки значимости коэффициента при $\hat{u}_{t-1}^2$.

Если гипотеза о гетероскедастичности принимается, то полученные оценки h_t применяются для новой оценки первого уравнения модели уже с учетом характера гетероскедастичности. Далее этот процесс может быть повторен итеративным образом до достижения сходимости. Проблемы могут возникнуть, если оценка α_1 меньше нуля или больше единицы.

Отметим, что эффекты ARCH обнаружены у курсов акций и у других активов, с которыми производят спекулятивные операции. Читатель легко может вспомнить, что после резких сдвигов в курсе валюты разница между курсом продажи и курсом покупки сразу увеличивается, так как возрастает неопределенность относительно дальнейшего его движения: то ли он вернется к прежнему уровню, то ли продолжит движение в том же направлении. Это говорит о воздействии дисперсии колебаний случайной компоненты на уровень курса. Эффекты ARCH признаны полезными в моделях инфляции, когда последовательности больших и малых ошибок прогнозирования образуют кластеры. Все это объясняется краткосрочным ажиотажным спросом, возникающим при очередном повышении цен и ожиданиях их дальнейшего роста.

Уточненная дисперсия остаточного члена позволяет гибко строить доверительные интервалы для будущих значений эндогенной переменной, т.е. для y_{t+1} .

Модель Энгла была расширена до включения большего числа лагов остаточного члена. В этом более общем случае спецификация модели ARCH имеет вид

$$y_t = x_t \beta + u_t; \quad (14.60)$$

$$\text{var}(u_t | I_{t-1}) = h_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^2, \quad (14.61)$$

где $x_t = (x_{t1}, x_{t2}, \dots, x_{tk})$ — вектор-строка значений экзогенных переменных в момент t ;

$\beta = (\beta_1, \beta_2, \dots, \beta_k)'$ — вектор-столбец коэффициентов при экзогенных переменных;

I_{t-1} — совокупность информации, известной на момент $t-1$, включающая значения x_t и лаговые значения y_t и x_t , т.е.

$$I_{t-1} = (x_t, x_{t-1}, \dots, y_{t-1}, y_{t-2}, \dots),$$

где h_t^2 — дисперсия u_t , условная по I_{t-1} .

Безусловная дисперсия u_t постоянна и равна:

$$\text{var}(u_t) = \frac{\alpha_0}{1 - \sum_{i=1}^q \alpha_i} > 0.$$

Модель GARCH

Модель GARCH(p, q) — обобщенная модель ARCH (generalized ARCH), введенная Т. Болерслеевым [135] является особым случаем модели ARCH(p, q), в уравнении которой есть авторегрессия и скользящая средняя

$$h_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^2 + \sum_{i=1}^p \phi_i h_{t-i}^2; \quad (14.62)$$

$$\text{var}(u_t) = \frac{\alpha_0}{1 - \sum_{i=1}^q \alpha_i - \sum_{i=1}^p \phi_i}. \quad (14.63)$$

Модель GARCH-M

Обобщенная модель ARCH в среднем значении (по-английски называется GARCH-in-mean, кратко она обозначается как GARCH(p, q) — M) имеет следующую спецификацию:

$$y_t = x_t \beta + \gamma h_t^2 + u_t, \quad (14.64)$$

где $h_t^2 = \text{var}(u_t | I_{t-1})$ определена в уравнении (14.62).

Таким образом, среднее значение зависимой переменной y_t в момент t зависит от условной дисперсии остаточного члена u_t , представляемой, в свою очередь, моделью ARCH.

Необходимое условие ковариационной стационарности (14.64) состоит в том, чтобы

$$\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \phi_i < 1. \quad (14.65)$$

В дополнение к ограничениям (14.63) и (14.65) Т. Болерслев предполагает также, что $\alpha_i \geq 0, i = 1, 2, \dots, q$ и $\phi_i \geq 0, i = 1, 2, \dots, p$. Этих дополнительных ограничений достаточно для того, чтобы условная дисперсия была положительной, но они не являются необходимыми.

Модель AGARCH

Модель GARCH(p, q), в которой условная стандартная ошибка u_t в (14.60) определяется как

$$h_t = \sqrt{\text{var}(u_t | I_{t-1})} = \alpha_0 + \sum_{i=1}^q \alpha_i |u_{t-i}| + \sum_{i=1}^p \phi_i h_{t-i} \quad (14.66)$$

называется *абсолютной GARCH* и обозначается AGARCH(p, q). Она также содержит авторегрессионную часть порядка p и скользящую среднюю порядка q .

Модель AGARCH-M

AGARCH-M — это модель AGARCH в среднем значении (AGARCH-in-mean) — специфицируется уравнениями (14.64) и (14.66).

Модель EGARCH

EGARCH — это экспоненциальная модель GARCH(p, q), в которой логарифм условной дисперсии ошибки уравнения (14.60) имеет следующую спецификацию:

$$\ln h_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \left(\frac{u_{t-i}}{h_{t-i}} \right) + \sum_{i=1}^q \alpha_i^* \left(\left| \frac{u_{t-i}}{h_{t-i}} \right| - \mu \right) + \sum_{i=1}^p \phi_i \ln h_{t-i}^2, \quad (14.67)$$

где $\mu = M\left(\left|\frac{u_t}{h_t}\right|\right)$.

Значение μ зависит от функции плотности вероятностей, которая, по предположению, описывает стандартизованные возмущения $\varepsilon_t = \frac{u_t}{h_t}$. Эта модель, разработанная Д. Нельсоном [154], допускает асимметричные воздействия прошлых ошибок на условные дисперсии ошибок.

Модель EGARCH-M

Это модель EGARCH(p, q) в среднем значении (EGARCH(p, q)-in-mean), специфицируемая уравнениями (14.64) и (14.67).

Оценивание моделей семейства ARCH

Параметры моделей семейства ARCH получают методом максимального правдоподобия, применяя приближенные итеративные методы, например алгоритм Ньютона — Рафсона, использующего численные оценки производных.

Сходимость обычно зависит от характера условной гетероскедастичности в данных и от выбора начальных оценок параметров.

Тесты на наличие эффектов ARCH или GARCH

Самый простой способ обнаружения эффектов ARCH(p) состоит в использовании процедуры множителей Лагранжа, предложенной Р. Энглom [141]. Процедура предполагает два шага. На первом шаге получают остатки $\hat{u}_t$ после применения обычного МНК для оценивания регрессии y_t на x_t , а на втором — оценивается регрессия квадратов остатков $\hat{u}_t^2$ на константу и p лаговых значений квадратов остатков

$$\hat{u}_t^2 = \alpha_0 + \alpha_1 \hat{u}_{t-1}^2 + \dots + \alpha_p \hat{u}_{t-p}^2 + \varepsilon_t$$

для $t = p + 1, p + 2, \dots, n$. Далее проверяется статистическая значимость коэффициентов, т.е. нулевая гипотеза $H_0: \alpha_1 = \dots = \alpha_p = 0$.

* * *

В этой главе представлены методы рекуррентного оценивания параметров линейной множественной регрессии. Данный подход позволяет получить важную дополнительную информацию об устойчивости параметров, оценить траектории движения коэффициентов. На основе такого рода информации могут быть выдвинуты более адекватные гипотезы о структуре регрессионного уравнения или добавлены уравнения, отражающие динамику коэффициентов, а также определены моменты скачков в значениях коэффициентов, что позволит весь выборочный период разделить на отдельные подпериоды относительно однородного поведения исследуемого временного ряда.

Критерий качества, использовавшийся здесь, предполагает минимизацию суммы квадратов ошибок прогнозирования на один шаг вперед. И этот критерий, по нашему мнению, обладает преимуществом по сравнению с коэффициентом детерминации R^2 , который не является операциональным в

том смысле, что он не имеет максимума и может только возрастать при увеличении числа регрессоров. Так, в одном исследовании [151] в результате статистического анализа нами были получены три модели практически одинакового качества, если судить по обычным статистическим критериям. И только с применением критерия, построенного на анализе ошибок ретроспективных прогнозов на один шаг вперед, полученных рекуррентным способом с использованием адаптивной регрессии, сразу удалось выявить достоинства одной из моделей.

Авторегрессионные модели семейства ARCH, предполагающие условную (локальную) изменчивость дисперсии остаточного члена регрессии, позволяют на каждом шаге обновлять оценку этой дисперсии и более гибко строить точечные и интервальные оценки прогнозов.

Глава 15

КРАТКОСРОЧНОЕ ПРОГНОЗИРОВАНИЕ КУРСОВ ВАЛЮТ С ПОМОЩЬЮ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ

Повышение эффективности краткосрочных операций с валютой — одна из важных задач в деятельности банков и других инвесторов. Ежедневно (а иногда и несколько раз в день) банки продают и покупают различные валюты в значительных объемах, стремясь придать движению имеющимся в наличии свободным валютным резервам с целью избежать потерь от конъюнктурных колебаний курсов и получить дополнительную прибыль. Причем валютные операции осуществляются с большой скоростью путем договоренности представителей банков с помощью современных средств связи, и очень важно выйти на валютный рынок с предложениями раньше конкурентов. Все это, по сути дела, составная часть непрерывного процесса формирования оптимальной структуры валютных резервов.

Эффективность валютных операций существенным образом зависит от надежности прогнозов колебания курсов валют. Именно поэтому краткосрочное прогнозирование курсов имеет большое практическое значение для оперативной деятельности банков и прочих инвесторов, а вопрос о возможности применения статистических методов для этой цели представляется актуальным и естественным. Проблема краткосрочного прогнозирования курсов валют с применением статистических моделей будет рассмотрена здесь исходя из того, что для успешного ведения валютных операций требуется получение прогнозов на одни сутки вперед.

Информация о динамике курсов национальных валют создает впечатление хаотического движения: падение и рост курсов сменяют друг друга в каком-то случайном порядке. Даже если за большой интервал времени отмечается тенденция, например, к росту, то на графике легко можно увидеть, что эта тенденция прокладывает себе путь через сложные движения временного ряда курса валюты. Направление ряда все время меняется под воздействием нерегулярных и часто неизвестных сил. Исследуемый объект в полной мере подвержен воздействию стихии мирового рынка, и точной информации о будущем движении курса нет. Необходимо сделать прогноз. При этом совершенно очевидно, что прогнозировать даже знак прироста курса очень сложно. Делать это обычно поручают экспертам, которые анализируют текущую конъюнктуру, а также пытаются выделить факторы, регулярным образом связанные с движением курса (фундаментальный анализ). При построении формальных моделей также пытаются выделить круг существенных факторов и на их основе сконструировать какой-либо индикатор или предиктор.

Насколько нам известно, ни эксперты-практики, ни формальные методы не дают пока устойчивых хороших результатов. Полагаем, объясняется это прежде всего тем, что если и есть действительно какой-либо круг факторов, влияющих стабильным образом на курс, то их воздействие надежно скрыто наложенной случайной составляющей, да и управляющими воздействиями центральных банков. В результате эти факторы и их влияние выделить довольно трудно.

Мы склоняемся к тому, чтобы считать краткосрочное прогнозирование курса по существу задачей прогнозирования последовательного движения изолированного временного ряда, причиной которого является главным образом массовое поведение на валютном рынке мелких дельцов и крупных финансовых воротил, совершающих основной объем финансовых операций с валютой. Такой подход можно отнести к так называемому техническому анализу.

Конечно, отдельно взятый участник валютной игры волен совершенно произвольно менять свою стратегию. И все же можно предположить, что поведение всей массы участ-

ников через соотношение спроса и предложения, влияющее на курс валюты, обладает в текущий период времени какой-то определенной доминирующей логикой, обнаруживающейся через закон больших чисел. Например, при падении курса валюты ее могут скупать, ожидая в дальнейшем повышения курса. И такой массовый спрос валюты действительно ведет к росту ее курса. Или наоборот, если после падения курса валюты доверие к ней падает и ожидается ее дальнейшее обесценение, то преобладает массовое предложение и курс падает еще ниже.

Заметим, что при таком упрощенном подходе самой динамике временного ряда можно прочесть как хронологическую запись о массовом поведении участников валютного рынка. Это дает возможность при построении модели исходить из самого ряда, не привлекая дополнительной информации, а все рассуждения о массовом поведении участников рынка использовать лишь для качественной интерпретации.

Если бы удалось найти в динамике ряда хотя бы краткосрочные закономерности, реализующиеся с вероятностью более 50%, то это дало бы основания рассчитывать на успех. Тогда для прогнозирования курсов стало бы возможным применение статистических методов, улавливающих более или менее устойчивые отношения последовательных событий временного ряда.

В данном случае мы ставим перед собой следующую задачу. Во-первых, выяснить применимость для краткосрочного прогнозирования валютных курсов каких-либо статистических методов, назначение которых — описывать повторяющиеся события или ситуации, характеризующиеся относительно устойчивыми связями. Во-вторых, если статистические методы применимы для решения поставленной задачи, то установить их наиболее перспективный класс, указать характерные особенности этих методов, особое внимание уделить простейшим из них. В-третьих, показать на примере практические результаты хотя бы некоторых попыток.

Отметим, что вопросам прогнозирования курсов валют всегда уделялось большое внимание. Из публикаций на близкую тему укажем, например, работу К. Гренжера и О. Моргенштерна (Granger Clive W.J., Morgenstern Oscar. Predictability of stock market prices. Massachusetts, 1970), в

которой исследуется динамика курсов акций и приведена обширная библиография. В этой монографии фактически сделан вывод о том, что если и есть какая-либо корреляция в рядах подобного рода, то наиболее вероятно, что она имеется между смежными приростами курсов. В дальнейшем появились модели класса ARCH, GARCH и др. (см. гл. 14, § 7), позволяющие учитывать волатильность.

Однако возникает вопрос, не пытаемся ли мы прогнозировать совершенно случайные колебания курсов валют. Для ответа на этот вопрос в [126] проведено специальное исследование.

§ 2. АНАЛИЗ СЛУЧАЙНОСТИ ДВИЖЕНИЯ КУРСОВ ВАЛЮТ

Объектом исследования являлся временной ряд последовательных ежедневных данных (наблюдений) о курсе той или иной валюты по отношению к американскому доллару, который будем записывать как $x_1, x_2, \dots, x_N$, где N — длина ряда, точнее длина исследуемого отрезка ряда. Задача состояла в том, чтобы выявить наличие более или менее устойчивой зависимости i -го наблюдения от предшествующих и на этом основании сделать прогноз на $(N+1)$ -й момент. В качестве статистической базы исследования взяты данные о курсах пяти валют: британского фунта стерлингов, западногерманской марки, японской иены, французского и швейцарского франков. Длина каждого временного ряда была произвольно принята равной 150 ежедневным наблюдениям за период с 31 декабря 1982 г. по 29 июля 1983 г. Таким образом, ряды взяты достаточно протяженными для того, чтобы можно было сделать более или менее основательные выводы.

Прежде чем приступить к изучению механизма связи между последовательными значениями курса валюты, была предпринята попытка выяснить, не являются ли исходные ряды абсолютно случайными, представляющими так называемый белый шум, в котором отсутствует зависимость между значениями ряда, относящимися к различным моментам, и направление движения которого в будущем равновероятно, т.е. непредсказуемо. Для этого отобранные ряды испы-

таны с помощью критериев случайности и в результате получено первое представление об исходном материале.

В качестве критериев случайности использовались: критерий поворотных точек, критерий распределения длины фазы, критерий, основанный на знаках разностей, и критерий, основанный на ранговой корреляции. Подробнее об этих критериях см., например, [120].

Критерий поворотных точек состоит в подсчете числа пиков и впадин. *Пиком* называется наблюдение, величиной более двух соседних; *впадина*, наоборот, — значение, которое меньше двух соседних. Оба эти значения называются *поворотными точками*. Число поворотных точек в исследуемом ряде сравнивается с их числом в совершенно случайном ряде и на этом основании делается вывод, случаен ряд или нет. Проверка показала, что для всех рассматриваемых рядов наблюдаемое число поворотных точек существенно меньше, чем это следует теоретически для случайного ряда. Другими словами, данный критерий свидетельствует о том, что эти ряды не совсем случайны: в них может быть скрыта некоторая закономерность движения вперед.

Критерий распределения длины фазы предназначен для изучения распределения интервалов между соседними поворотными точками, называемых *фазами*. Расчеты обнаружили заметные отличия теоретического распределения длины фазы у случайного ряда от тех, которыми обладают временные ряды валютных курсов. Для всех выбранных рядов характерно меньшее число коротких фаз и большее число более длинных фаз, чем для случайного ряда, а также меньшее общее число фаз, т.е. фазы движения курсов валют оказываются в среднем более продолжительными, и их смена происходит реже, чем у чисто случайного ряда. Это наводит на мысль, что, несмотря на большую колеблемость, движение валютных курсов генерируется каким-то инерционным механизмом, познание которого могло бы послужить известной базой для получения полезных прогнозов.

Критерий, основанный на знаках разностей, состоит в подсчете числа положительных разностей первого порядка, иначе говоря, числа *точек возрастания*. В основном этот критерий используется для проверки нали-

чия линейного тренда. Здесь можно отметить некоторый разброс наблюдаемого числа точек возрастания относительно их теоретического числа. Так, швейцарский франк и британский фунт имеют число точек возрастания меньше теоретического, а остальные валюты — больше. Однако весь замеченный разброс, хотя и на пределе, укладывается в 99%-ный доверительный интервал. Это свидетельствует об отсутствии четко выраженного устойчивого линейного тренда, но одновременно величина разброса дает основания предположить наличие относительно долгосрочных «кренов» в ту или иную сторону, что может быть полезным при прогнозировании.

Критерии, основанные на ранговой корреляции, усиливают подозрение о наличии линейных трендов у четырех из пяти валют за исключением фунта стерлингов. Для британского фунта коэффициент τ Кендэла близок к нулю, а для остальных валют положителен и существенно отличен от нуля.

В целом проведенное изучение рядов валютных курсов показало, что, несмотря на сильную колеблемость ежедневных данных, они не являются совершенно случайными. В них проявляется некоторая инерционность движения, и это позволяет надеяться на применимость статистических методов для прогнозирования курсов. В то же время их исследованные характеристики не очень сильно отличаются от тех, которые теоретически выведены для совершенно случайного ряда. Вследствие этого нельзя ожидать от статистических прогнозов большой степени осуществимости. Однако в совокупности за достаточно протяженный отрезок времени они, по-видимому, могут оказаться полезными.

§ 3. АДАПТИВНАЯ МОДЕЛЬ ПРОГНОЗИРОВАНИЯ ВРЕМЕННОГО РЯДА С НЕУСТОЙЧИВЫМ ХАРАКТЕРОМ КОЛЕБАНИЙ

Исходя из того, что средний уровень ряда подвержен дрейфу, а флуктуациям относительно него присуща сильная изменчивость, при построении прогнозной модели выбор был сделан в пользу класса адаптивных моделей. Модель такого типа всегда находится в движении, впитывает новую информацию и на основе ошибки прогноза на

предыдущем шаге корректирует свои параметры, подстраиваясь к механизму, генерирующему временной ряд (подробнее см. гл. 1).

Нам необходим прогноз курса валюты на следующий день. Конечно, желательно знать и направление изменения курса за день и его величину. Однако для получения положительного эффекта от валютных операций практически достаточно предвидеть знак приращения курса валюты, т.е. определить движение курса вверх или вниз. В связи с этим для упрощения задачи сделаем некоторые преобразования исходных рядов. Во-первых, от данных $x_1, x_2, \dots, x_N$ перейдем к $(N-1)$ -м первым разностям этого ряда $\Delta x_1, \dots, \Delta x_{N-1}$, где $\Delta x_i = x_{i+1} - x_i$, $i = 1, 2, \dots, N-1$. Затем сделаем замену разностей на k_i по следующему правилу:

$$k_i = \begin{cases} +1, & \Delta x_i > 0 \\ 0, & \Delta x_i = 0 \\ -1, & \Delta x_i < 0 \end{cases}$$

Ряд $k_1, k_2, \dots, k_{N-1}$, состоящий из единиц и нулей, будем для краткости называть *знаковым рядом*. Разумеется, упрощение, достигнутое переходом от исходного временного ряда к знаковому, сопряжено с потерей части информации. По знаковому ряду, например, ничего нельзя сказать об амплитуде колебаний исходного ряда. Однако несомненно, что в знаковом ряде основная информация о характере и последовательности изменений курса валюты сохраняется, причем в самом рафинированном виде, а исключается информация, имеющая для нас второстепенное значение. Одним из следствий такого преобразования исходной статистической базы является то, что довольно-таки сложный вопрос о наличии тренда и его использовании при прогнозировании в этом случае переходит в другую плоскость и сводится к изучению закономерности чередования или сохранения знаков. Знаковый ряд, очевидно, более адекватен поставленной задаче.

Итак, имеется ряд, характеризующий динамику валютного курса, состоящий из $-1, 0$ и $+1$. Требуется определить, каким будет следующее значение этого ряда. Для прогнозирования временного ряда обычно строится модель, отражающая свойства генерирующего его механизма. В основу

прогнозной модели в данном случае положим гипотезу об инерционности действий основной массы участников валютной игры. Разумеется, мы не рассматриваем периоды нарушения обычной жизни валютного рынка внезапными крупными чрезвычайными событиями.

В сущности, могут быть лишь три ситуации. Первая, когда основной массой участников валютного рынка ожидается неизменное движение курса, и их совместные действия через спрос и предложение способствуют сохранению знака прироста курса. Вторая, когда большинством игроков предполагается изменение направления движения курса и массовый характер валютных операций в предвидении этого действительно вызывают перемену знака прироста курса. Третья, когда не ожидается ни роста, ни падения, и индифферентность к валюте оставляет ее курс в неопределенном состоянии. Направление ее движения в будущем абсолютно случайно. О нем заранее ничего нельзя сказать. Модель должна уловить, какая именно из трех ситуаций превалирует в последнее время, и исходя из гипотезы, что она сохранится по крайней мере в течение еще одной единицы времени, дать прогноз на следующий момент.

Построение модели начнем с рассмотрения произведения $m_i = k_i k_{i-1}$. Легко заметить, что

$$m_i = \begin{cases} +1, & \text{когда } k_i = 1 \text{ и } k_{i-1} = 1 \text{ либо когда } k_i = -1 \text{ и } k_{i-1} = -1; \\ 0, & \text{когда } k_i = 0 \text{ или (и) } k_{i-1} = 0; \\ -1, & \text{когда } k_i = 1 \text{ и } k_{i-1} = -1 \text{ либо когда } k_i = -1 \text{ и } k_{i-1} = 1. \end{cases}$$

Первое равенство в этом выражении соответствует сохранению тенденции движения курса при переходе от одного момента к другому, третье — смене тенденций, т.е. в этом случае речь идет о поворотной точке, а второе — неопределенной ситуации.

Заметим, что m_i описывает движение курса лишь на единичном интервале от точки $i-1$ к точке i . Для того чтобы выяснить, какая же ситуация чаще встречается в последнее время, целесообразно каким-то образом усреднить ряд m_i за известный интервал. Для этого применим метод экспоненциального сглаживания.

Экспоненциальное сглаживание ряда m_i осуществляется по рекуррентной формуле

$$S_t = \alpha m_t + \beta S_{t-1}, \quad (15.1)$$

где S_t — значение экспоненциальной средней в момент t ;
 α — постоянная сглаживания (параметр адаптации), $0 < \alpha \leq 1$,
 $\beta = 1 - \alpha$.

Свойства экспоненциальной средней подробно рассмотрены в гл. 1.

Выражение (15.1) можно переписать как

$$S_t = \alpha m_t + (1 - \alpha) S_{t-1} = S_{t-1} + \alpha (m_t - S_{t-1}). \quad (15.2)$$

Здесь S_{t-1} можно рассматривать как прогноз на один шаг вперед, т.е. как прогноз величины m_t , сделанный в момент $t-1$, тогда величина $(m_t - S_{t-1})$ есть погрешность этого прогноза, а новый прогноз S_t получается в результате корректировки предыдущего прогноза с учетом его ошибки. В этом и состоит адаптация прогнозной модели. Наилучшее значение α легко устанавливается методом проб и поиска минимума потерь или максимума выигрыша.

Приняв во внимание, что

$$k_i = \frac{\Delta x_i}{|\Delta x_i|},$$

и

$$m_i = k_i k_{i-1} = \frac{\Delta x_i \Delta x_{i-1}}{|\Delta x_i \Delta x_{i-1}|},$$

нетрудно убедиться в том, что S_t есть не что иное, как адаптивный коэффициент корреляции, рассмотренный в гл. 9.

Ясно, что значение S_t , являющееся результатом усреднения единиц и нулей, будет дробным числом из интервала $[-1, 1]$, поэтому прогноз m на момент $t+1$ будем определять так:

$$\hat{m}_{t+1} = \begin{cases} +1, & S_t > 0, \\ 0, & S_t = 0, \\ -1, & S_t < 0. \end{cases}$$

Положительный знак $\hat{m}_{t+1}$ означает сохранение того знака прироста, который имел место в момент t , а отрицательный — перемену. Прогноз знака прироста курса валюты на момент $t+1$ определим как

$$\text{sign}(\Delta x_{t+1}) = \text{sign}(\hat{m}_{t+1} k_t).$$

Обратим внимание на то, что эта модель способна в различное время отражать то положительную, то отрицательную корреляцию соседних приростов. Она время от времени, приспосабливаясь к наблюдениям, как бы меняет свои свойства на противоположные. Перейдем теперь к практическим испытаниям этой модели.

§ 4. ПРОГНОЗИРОВАНИЕ КУРСОВ ВАЛЮТ

Прежде всего нужно выбрать критерии, по которым будем судить о полезности работы прогнозной модели.

Критерии оценки результатов испытаний. Нас, разумеется, будет интересовать не только точность прогнозов, но и получаемый от них удельный экономический эффект. Поэтому при расчете критериев условно будем считать, что объем операций с данной валютой (т.е. объем ее покупки или продажи) равен 1 дол. США в день. Если прогноз оказался правильным, то выигрыш равен разности $|x_{t+1} - x_t|$, а если прогноз ошибочен, то $|x_{t+1} - x_t|$ будет уже размером ущерба.

Учитывая, что исследуемому статистическому материалу присущи сильные стохастические колебания, в основе критериев оценки результатов положим средний (на базовом периоде) выигрыш (или проигрыш) от операций с валютой за единицу времени (т.е. за день) в расчете на один доллар или совокупный выигрыш за определенный период функционирования модели. Кроме того, целесообразно рассчитывать и некоторые другие абсолютные и относительные показатели. Вот перечень критериев, которые мы используем в наших экспериментах:

L — число оправдавшихся прогнозов знака прироста курса валюты, т.е. число выигрышей;

M — число ошибочных прогнозов знака прироста, т.е. число проигрышей;

$PL = [L/(L + M)]100\%$ — процент оправдавшихся прогнозов знака прироста;

$PM = [M/(L + M)]100\%$ — процент ошибочных прогнозов знака прироста;

SPR — сумма выигрышей (т.е. суммируются только выигрыши) в валютных операциях за весь исследуемый период функционирования прогнозной модели, который будем обозначать через T (измеряется в национальной валюте);

$SLOS$ — сумма потерь в валютных операциях за период T (в национальной валюте);

$SPR-SLOS$ — сальдо, сумма чистого выигрыша (в национальной валюте);

$R = SPR/SLOS$ — отношение суммы выигрышей к сумме потерь;

$REL = SPR/(SPR + SLOS)100\%$ — процент реализованных возможностей ($SPR + SLOS$ — максимально возможный выигрыш).

Эксперимент № 1. Проведем испытание изложенной выше адаптивной модели. Всего имеется 150 точек ряда. Адаптируем нашу модель, например, на первых 20 точках, которые принимаем за статистическую базу модели. Число 20 выбрано, строго говоря, произвольно; основные соображения, которыми мы руководствовались, состоят в том, что за 20 шагов влияние на прогноз начального значения практически исчезает. Сделаем прогноз 21-й точки. Затем сдвинем базу на один шаг вперед и прогоним модель от точки 2 до точки 21 и сделаем прогноз 22-й точки и т.д. Всего получим 130 прогнозов. Однако прежде чем получать прогноз, будем на каждой статистической базе из 20 точек методом перебора определять наилучшее для данного этапа значение α (из 10 значений от 0,1 до 1,0 с интервалом 0,1). Результаты приведены в табл. 15.1, из которой видим, что по четырем валютам (британский фунт, западногерманская марка, японская иена, швейцарский франк) получена прибыль, причем процент реализованных возможностей заметно превышает 50%. Операции с французским франком привели к некоторым потерям. Несколько слов нужно сказать о

показателе среднего выигрыша за день. Он, как правило, представлен очень малым числом. Однако если речь идет не об одной операции в 1 дол., а об операциях на крупную сумму в течение большого числа дней, то суммарный выигрыш достигнет значительной величины.

Сделаем теперь оценку результатов валютных операций в пересчете на годовую норму прибыли. Для этого отнесем средний выигрыш за день к среднему уровню курса данной валюты за исследуемый период (т.е. к среднему значению 130 последних точек) и получим относительную прибыль r за один день. Будем считать, что в году 260 рабочих дней (имеются в виду рабочие дни банков). Тогда за год норма прибыли составит $[(1 + r)^{260} - 1] \cdot 100\%$. Результаты расчетов по этой формуле приведены в табл. 15.2. Как видим, по всем валютам, кроме французского франка, получены вполне положительные итоги. По французскому франку понесены относительно небольшие убытки.

Хотелось бы также оценить, насколько устойчив положительный результат в отдельно взятой валютной операции, велика ли зависимость от выпадения благоприятного случая. Для измерения относительной устойчивости возьмем математическое ожидание выигрыша и стандартное отклонение от него. Исходя из предположения, что результаты подчиняются нормальному закону распределения, вычислим площадь под кривой распределения, находящуюся в области отрицательного аргумента (т.е. определим вероятность проигрыша). Эту площадь обозначим через F . Находить ее будем приближенно по известным таблицам, а выражать в процентах. Отметим, что показатель $PERDAY$ является оценкой математического ожидания выигрыша, а аргументом для определения F по таблицам нормального распределения будет отношение $PERDAY/\sigma$, где σ — стандартное отклонение выигрыша от его математического ожидания. Значения F также приведены в табл. 15.2, и они свидетельствуют о том, что в каждой конкретной валютной операции вероятность проигрыша весьма велика.

Эксперимент № 2. Теперь получим прогноз по наивным моделям. Будем различать два типа так называемых наивных моделей. Согласно модели первого типа все время предполагается, что прирост курса в следующий момент будет таким же по знаку, что и в текущий момент. В модели

Таблица 15.1

Прогнозирование курсов валют по адаптивной и наивной моделям
(результаты экспериментов № 1 и 2)

Критерий	Британский фунт		Западногерманская марка		Японская иена		Французский франк		Швейцарский франк	
	Адаптивная	Наивная	Адаптивная	Наивная	Адаптивная	Наивная	Адаптивная	Наивная	Адаптивная	Наивная
L , число выигрышей	70	63	73	71	65	65	63	76	71	69
M , число проигрышей	60	67	56	57	63	61	65	50	57	57
PL , % выигрышей	53,85	48,46	56,59	55,47	50,78	51,59	49,22	60,32	55,47	54,76
PM , % проигрышей	46,15	51,54	43,41	44,53	49,22	48,41	50,78	39,68	44,53	45,24
SPR , сумма выигрышей	0,52	0,46	0,72	0,71	72,47	77,02	2,32	0,77	0,67	2,88
$SLOS$, сумма проигрышей	0,43	0,49	0,53	0,54	52,89	48,34	2,38	0,39	0,48	1,82
$SPR-SLOS$, сальдо	0,09	-0,03	0,19	0,17	19,58	28,68	-0,06	0,38	0,19	1,06
$R=SPR/SLOS$	1,22	0,95	1,37	1,32	1,37	1,59	0,97	1,96	1,39	1,59
$PERDAY$, средний выигрыш за день на 1 дол. $\times 10^3$	0,737	-0,197	1,501	1,321	150,615	220,615	-0,498	2,888	1,464	8,197
REL , % реализованных возможностей	55,02	48,66	57,83	56,89	57,81	61,44	49,31	66,25	58,24	61,33

Оценки нормы прибыли операций с валютами
по адаптивной и наивной моделям

Валюта	Модель	Относительная прибыль за день, %	Норма прибыли за год, %	Вероятность проигрыша в отдельной валютной операции (F)
Британский фунт	адаптивная наивная	0,048 -0,013	13,28 -3,27	47,0 50,8
Западногерманская марка	адаптивная наивная	0,060 0,053	17,02 14,83	45,2 45,8
Японская иена	адаптивная наивная	0,063 0,093	17,86 27,24	45,4 43,3
Французский франк	адаптивная наивная	-0,0068 0,112	-1,745 33,64	50,3 40,6
Швейцарский франк	адаптивная наивная	0,071 0,139	20,16 43,66	45,3 44,6

второго типа логика противоположна: ожидается, что в следующий момент знак прироста изменится. Ясно, что правильное поочередное применение этих двух моделей может дать абсолютно точные прогнозы. Но проблема состоит в определении момента, когда прибегать к первой, а когда — ко второй.

Построим на основе этих двух моделей новую, в которую они входят как альтернативы. Организуем автоматический выбор той или иной модели по такому правилу: если на последних n точках (статистической базе) применение первой модели дало отрицательный результат, то для прогнозирования $(n+1)$ -й точки применяется вторая модель, и наоборот. Результаты испытания этой модели на прежних данных со смещаемой базой с $n=20$ приведены в табл. 15.1 и 15.2. По четырем валютам (на этот раз — кроме британского фунта) получен положительный итог. Особенно велика норма прибыли у швейцарского и французского франка. Риск, однако, остается.

* * *

В этой главе мы хотели показать, что простейшие статистические модели могут оказаться полезными даже в столь сложной задаче, как прогнозирование валютных курсов. Была исследована стохастическая природа валютных курсов. Сделан вывод о наличии в их движении определенной инерционности. Рассмотрены адаптивная и наивная модели, которые применены к реальным данным. Результаты в целом оказались положительными, хотя полностью риска избежать не удастся.

Предложенные подходы к краткосрочному прогнозированию валютных курсов позволяют эксперту оперативно принимать решения по текущим валютным операциям. Сам прогноз, вообще говоря, является исходной информацией для второго этапа работы — выработки оптимального решения. На этом этапе решение задачи по одной валюте целесообразно сопоставить с решениями задач по другим валютам в рамках комплексного подхода. Одномерная задача естественным образом перерастает в многомерную.

Глава 16

СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ВАЛЮТНЫХ АУКЦИОНОВ НА МОСКОВСКОЙ МЕЖБАНКОВСКОЙ ВАЛЮТНОЙ БИРЖЕ

В этой главе представлена методика построения модели валютных торгов, позволяющей прогнозировать курс доллара к рублю на основе предварительных заявок на покупку и продажу валюты с учетом возможных вариантов интервенции Центрального банка Российской Федерации (ЦБ РФ).

§ 1. ЦЕЛИ ИССЛЕДОВАНИЯ

Актуальность построения модели валютных торгов объясняется важностью курса доллара по отношению к рублю для всей экономики России, для ее внешней торговли, для юридических и физических лиц, для руководящих органов, и в первую очередь для ЦБ РФ.

В рамках рассматриваемого эконометрического исследования* валютного рынка основное внимание было сосредоточено на механизмах формирования курса рубля на торгах Московской межбанковской валютной биржи (ММВБ). В нашем распоряжении имелась статистическая информация за январь–ноябрь 1992 г. Это один из драматических периодов экономической истории нашего государства — период революции цен, гиперинфляции, структурной ломки народного хозяйства.

* Это исследование выполнено нами совместно с канд.экон. наук А.С. Лушиным [130].

Основные цели работы:

- исследовать статистическую зависимость динамики курса от спроса и предложения валюты со стороны коммерческих банков, а также интервенции ЦБ РФ на торгах ММВБ;
- рассмотреть теоретические модели торгов и разработать методику, позволяющую: отобрать модели, наиболее адекватные реальным статистическим наблюдениям и перспективные для прогнозирования и управления курсом, оценить параметры этих моделей статистическими методами и выбрать наилучшие из них для практической деятельности экспертов ММВБ и ЦБ РФ;
- использовать модели для оценивания курса доллара на предстоящих валютных торгах на основе начального спроса и предложения коммерческих банков с учетом возможных вариантов участия ЦБ РФ.

§ 2. ТЕОРЕТИЧЕСКИЕ МОДЕЛИ ВАЛЮТНЫХ ТОРГОВ

Для того чтобы выявить основные количественные соотношения между спросом, предложением и курсом доллара к рублю, проведем небольшое теоретическое исследование.

Пусть накануне торгов сформировалось некоторое начальное предложение долларов к продаже S_0 и начальный спрос D_0 при цене (курсе) P_0 , установившейся на предыдущих торгах (рис. 16.1). Разность $D_0 - S_0$ характеризует дефицит долларов на торгах, который должен быть устранен путем регулирования цены долларов, т.е. корректировки его курса. Цена, при которой наступает равенство спроса и предложения, называется *равновесной* (P_1 — рис. 16.1).

Простейшая модель торгов может быть сформулирована как

$$\Delta P = A(D_0 - S_0), \quad (16.1)$$

где

$$\Delta P = P_1 - P_0.$$

Модель (16.1) означает, что прирост курса пропорционален дефициту долларов на рынке при прежнем курсе, где A — константа, коэффициент пропорциональности.

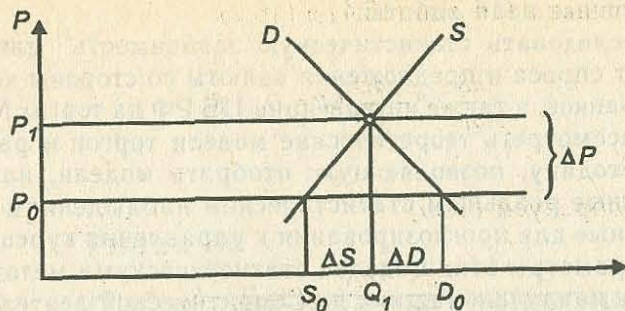


Рис. 16.1. Спрос и предложение долларов как функции курса:

S_0 — предложение долларов на начало торгов при курсе P_0 ;
 D_0 — заявки на покупку долларов на начало торгов при курсе P_0 ;
 P_0 — курс доллара в рублях, сложившийся к концу предыдущих торгов;
 P_1 — новый равновесный курс доллара;
 Q_1 — объем продаж в равновесной точке: $\Delta P = P_1 - P_0$, $\Delta D = Q_1 - D_0$,
 $\Delta S = Q_1 - S_0$

Рассмотрим, однако, этот вопрос подробнее, в частности, сосредоточим внимание на коэффициенте пропорциональности A — является ли он в действительности постоянным?

На рис. 16.1 зависимости предложения и спроса от цены упрощенно представлены линейными функциями. Переход от P_0 к новой равновесной цене P_1 описывается уравнением

$$S_0 + \Delta S = D_0 + \Delta D$$

или

$$\Delta S - \Delta D = D_0 - S_0. \quad (16.2)$$

Прирост предложения ΔS можно аппроксимировать как

$$\Delta S = \frac{dS}{dP} \Delta P, \quad (16.3)$$

а прирост спроса ΔD как

$$\Delta D = \frac{dD}{dP} \Delta P. \quad (16.4)$$

Подставим (16.3) и (16.4) в (16.2):

$$\frac{dS}{dP} \Delta P - \frac{dD}{dP} \Delta P = D_0 - S_0. \quad (16.5)$$

Преобразуем (16.5)

$$\left(\frac{dS}{dP} - \frac{dD}{dP} \right) \Delta P = D_0 - S_0,$$

откуда

$$\Delta P = \frac{1}{\left(\frac{dS}{dP} - \frac{dD}{dP} \right)} (D_0 - S_0). \quad (16.6)$$

Из сопоставления (16.1) и (16.6) видим, что множитель

$$A = \frac{1}{\left(\frac{dS}{dP} - \frac{dD}{dP} \right)}$$

характеризует связь скачка курса с дефицитом долларов на торгах. Если верна гипотеза о линейной зависимости спроса и предложения от цены, то $dS/dP = \text{const}$, $dD/dP = \text{const}$ и, как следствие, $A = \text{const}$. Тогда (16.6) можно исследовать методами регрессионного анализа и, в частности, оценить регрессию вида

$$\Delta P_t = A(D_{0t} - S_{0t}) + \varepsilon_t, \quad (16.7)$$

где t — номер или дата торгов (пусть $t = 1, \dots, T$, где T — объем выборки);

ε_t — случайная ошибка уравнения для торгов в день t .

В экономических исследованиях при изучении зависимостей между переменными часто рассматривают не сами

производные, а эластичности. Преобразуем выражение (16.6) с тем, чтобы ввести в него эластичности предложения и спроса по цене доллара. С этой целью рассмотрим сначала знаменатель (16.6)

$$\frac{dS}{dP} - \frac{dD}{dP} = \left(\frac{dS}{S} \frac{P}{dP} \right) \frac{S}{P} - \left(\frac{dD}{D} \frac{P}{dP} \right) \frac{D}{P} = \frac{a_1 S_0 - a_2 D_0}{P_0}, \quad (16.8)$$

где индекс '0' означает «на начало торгов», a_1 и a_2 эластичности предложения и спроса по курсу доллара

$$a_1 = \frac{dS}{S} \frac{P}{dP}, \quad a_2 = \frac{dD}{D} \frac{P}{dP}.$$

Тогда (16.6) можно переписать в виде

$$\Delta P = \frac{P_0}{a_1 S_0 - a_2 D_0} (D_0 - S_0). \quad (16.9)$$

Если принять, что эластичности постоянны

$$a_1 = \text{const}, \quad a_2 = \text{const}, \quad (16.10)$$

то их оценки можно получить статистически, рассматривая (16.9) как регрессионное уравнение. Для его оценки имеются два способа: 1) непосредственное оценивание регрессии нелинейной по параметрам; 2) оценивание параметров после преобразования (16.9) и приведения его к линейному виду:

$$D_0 - S_0 = a_1 \frac{\Delta P}{P_0} S_0 - a_2 \frac{\Delta P}{P_0} D_0. \quad (16.11)$$

Легко заметить, что у регрессии (16.11) оценки a_1 и a_2 будут иметь большие ошибки из-за очевидной коллинеарности регрессоров. Поэтому предпочтительнее получать оценки нелинейными методами.

Рассмотрим, что означает гипотеза о постоянстве эластичностей на примере предложения. Для спроса результат аналогичен. Имеем

$$\frac{dS}{S} \frac{P}{dP} = c_1 = \text{const}. \quad (16.12)$$

Отсюда

$$\frac{dS}{S} = c_1 \frac{dP}{P} \quad (16.13)$$

или

$$\frac{dS}{dP} = c_1 \frac{S}{P}. \quad (16.14)$$

Это означает, что производная предложения по цене не является более постоянной — она изменяется пропорционально отношению предложения к цене. Преобразуем (16.13) к виду

$$d(\ln S) = c_1 d(\ln P).$$

Отсюда интегрированием обеих частей равенства получаем

$$\ln S = c_1 \ln P + c,$$

где $c = \text{const}$,

или

$$S = c_2 P^{c_1}. \quad (16.15)$$

Полученный результат свидетельствует о том, что гипотеза о постоянстве эластичностей эквивалентна гипотезе о степенной зависимости предложения (и спроса) от цены. Изме-

нение предложения (и спроса) для перехода от первоначальной P_0 к новой равновесной цене P_1 можно найти так:

$$\Delta S = c_2(P_1^{c_1} - P_0^{c_1}). \quad (16.16)$$

Таким образом, в этом случае зависимость предложения и спроса от цены уже не является линейной. Действительно, более вероятно, что зависимости спроса и предложения от цены криволинейны и, возможно, не постоянны во времени. Поэтому целесообразно попытаться аппроксимировать эти зависимости известными функциями и выбрать из них наилучшую пару для отражения предложения и спроса и на их основе построить модель торгов. В качестве таких известных функций были взяты парабола второго порядка ($F1$), гипербола ($F2$), логарифмическая зависимость ($F3$), экспонента ($F4$) и степенная функция ($F5$). Тогда

$$F1(P) = a_0 + a_1 P + a_2 P^2; \quad (16.17)$$

$$F2(P) = a_0 + a_1 \frac{1}{P}; \quad (16.18)$$

$$F3(P) = a_0 + a_1 \ln P; \quad (16.19)$$

$$F4(P) = a_0 + a_1 \exp(a_2 P); \quad (16.20)$$

$$F5(P) = a_0 + a_1 P^{a_2}, \quad (16.21)$$

где a_0, a_1, a_2 — постоянные параметры.

Основным равенством для построения модели торгов по-прежнему служит модель (16.2), где присутствуют не сами функции, а приросты предложения и спроса, приводящие первоначальную цену P_0 к новому равновесному значению P_1 . Приросты функций (16.17)–(16.21) будем описывать с помощью следующего приближенного равенства (формула дифференциала):

$$\Delta F \approx \frac{dF}{dP} \Delta P, \quad (16.22)$$

где $\Delta P = P_1 - P_0$.

Используя соотношение (16.22), построим, например, модель торгов на основе модели (16.2) при гипотезе, что спрос отражается логарифмической функцией, а предложение — экспоненциальной:

$$\Delta S - \Delta D = D_0 - S_0,$$

$$S = b_1 \exp(b_2 P), \quad \Delta S \approx b_2 S_0 \Delta P,$$

$$D = a_0 + a_1 \ln P, \quad \Delta D \approx a_1 \frac{\Delta P}{P_0},$$

где a_1 и b_2 — постоянные параметры. После подстановки ΔS и ΔD в выражение (16.2) имеем

$$b_2 S_0 \Delta P - a_1 \frac{\Delta P}{P_0} = D_0 - S_0.$$

Отсюда

$$\Delta P = \left(\frac{1}{b_2 S_0 - a_1 / P_0} \right) (D_0 - S_0)$$

или, переобозначив параметры для упрощения записи, получим:

$$\Delta P = \left(\frac{P_0}{\beta S_0 P_0 - \alpha} \right) (D_0 - S_0).$$

Открытым остается лишь вопрос об интервенции ЦБ РФ и ее учете в модели. До сих пор мы считали, что спрос и предложение формируются свободно без вмешательства ЦБ РФ. Роль ЦБ РФ на торгах можно учесть следующим образом.

Будем считать, что интервенция ЦБ РФ (обозначим ее I) увеличивает предложение до величины $(S_0 + I)$, если Банк России продает валюту. Причем в качестве интервенции будем рассматривать фактически реализованный объем продаж или покупок ЦБ РФ, а не его первоначальную заявку.

Поскольку ЦБ РФ на торгах играет регулируемую роль, его реакция на рост или падение курса коренным образом отличается от действия обычного участника торгов. Интервенцию ЦБ РФ можно также рассматривать как погашение части или всего первоначального спроса. В этой интерпретации назовем $(D_0 - I)$ реальным или эффективным спросом, воздействующим на торгах на уровень курса доллара. Так или иначе, в модели вместо $(D_0 - S_0)$ для учета интервенции ЦБ РФ используем $(D_0 - S_0 - I)$, где интервенция $I > 0$, если ЦБ РФ продает доллары, и $I < 0$, если ЦБ РФ покупает и тем увеличивает эффективный спрос.

Окончательно модель принимает вид

$$\Delta P = \frac{P_0}{\beta S_0 P_0 - \alpha} (D_0 - S_0 - I). \quad (16.23)$$

Множество вариантов сочетания различных пар функций для аппроксимации спроса и предложения порождает множество возможных моделей торгов. Все они могут быть обобщенно представлены как

$$\Delta P = A(D_0 - S_0), \quad (16.24)$$

где A — коэффициент пропорциональности, который теперь не является в общем случае постоянным, а сам зависит от некоторых конъюнктурных переменных (см. табл. 16.2 и 16.3).

Однако все эти модели нецелесообразно строить и исследовать: необходимо выделить из них те, от которых можно надеяться получить хороший результат. Для этого было проведено вспомогательное исследование, целью которого было выявить наиболее подходящие функции для аппроксимации спроса и предложения.

§ 3. АНАЛИЗ ИСХОДНЫХ ДАННЫХ

Предварительный анализ статистических данных о торгах на ММВБ в исследуемом периоде показал, что механизм формирования курса существенно различался в за-

висимости от того, происходит ли его рост или снижение. В связи с этим было принято решение строить модели отдельно для падающего и растущего курсов, поскольку «усредненная» модель может привести к большим ошибкам при прогнозировании.

Разделение на подвыборки. На основе исходных данных о торгах ММВБ за январь–ноябрь 1992 г. были сформированы две подвыборки: 1) для растущего курса $\Delta P > 0$ и 2) для падающего курса $\Delta P < 0$. Торги, на которых курс не менялся ($\Delta P = 0$), были исключены из рассмотрения. Кроме того, из подвыборки 1 были исключены два наблюдения с резким ростом курса, относящиеся к январю–февралю 1992 г. и вызванные паническим состоянием рынка. Что касается подвыборки 2, то здесь были исключены 12 точек, для которых характерно неэластичное поведение предложения ($\Delta S = 0$) и сверхэластичный спрос. Данные торги относились в основном к первой половине года и сопровождались массивными интервенциями ЦБ РФ при незначительном снижении курса.

При формировании подвыборок следовало также учесть, что валютный рынок России функционировал в 1992 г. в двух режимах: множественных курсов (до 1 июля) и единого курса (после 1 июля). Это нашло отражение в характере торгов на ММВБ и было исследовано нами для случая растущего курса (для падающего курса после 1 июля было слишком мало наблюдений).

Приведем перечень выборок, для которых проводились статистический анализ и оценивание регрессионных моделей.

Растущий курс (выборка 1):

- 1.1 — март–ноябрь 1992 г. — 34 точки;
- 1.2 — июль–ноябрь — 28 точек;
- 1.3 — сентябрь–ноябрь — 17 точек;
- 1.4 — октябрь–ноябрь — 11 точек.

Падающий курс (выборка 2):

- 2.1 — январь–ноябрь 1992 г. — 26 точек;
- 2.2 — январь–июнь — 21 точка.

В принципе для падающего курса было бы желательно построить выборку для случая отрицательных интервенций ЦБ РФ, т.е. когда ЦБ РФ покупал валюту для уменьшения глубины падения курса. Таких наблюдений (за сентябрь–ноябрь) в нашем распоряжении было всего четыре. Этого недостаточно для статистического оценивания.

Корреляционный анализ. Самые общие представления о характере взаимосвязи курса со спросом и предложением можно получить, рассмотрев корреляционные матрицы, содержащие коэффициенты корреляции представленных в них показателей. Соответствующие матрицы для растущего и падающего курсов позволили сделать следующие предварительные выводы.

В случае растущего курса, т.е. когда $\Delta P > 0$, отмечается достаточно высокая корреляция между изменениями спроса и предложения банков в ходе торгов и изменениями курса. При этом спрос в большей степени реагировал на относительный прирост курса, а предложение — на его абсолютное изменение; связь с курсом более тесная у спроса, нежели у предложения. Кроме того, прослеживалась достаточно высокая зависимость изменений курса от дефицита валюты на торгах, что подтверждает правильность рассмотренных в § 2 этой главы теоретических моделей.

При $\Delta P < 0$ связь между изменениями спроса, предложения и курса была выражена менее отчетливо и лучше прослеживалась для предложения. Другой вывод состоял в том, что в ходе торгов банки в большей степени ориентировались на абсолютное снижение курса, нежели на его относительное падение. Зависимость величины падения курса от избытка валюты на торгах также имела место, хотя и не такая тесная, как в случае $\Delta P > 0$. В целом можно заключить, что в силу высокой степени неопределенности механизм падения курса хуже поддавался количественному описанию, чем его рост.

§ 4. МОДЕЛИ ТОРГОВ ДЛЯ РАСТУЩЕГО КУРСА

Зависимость спроса и предложения от курса. Дальнейшее исследование зависимости спроса и предложения от курса заключалось в подборе кривых, наилучшим образом описывающих данные взаимосвязи. Этот подбор заключался в поиске наилучшего уравнения для приростов, причем тип уравнения вытекал из аналитического вида рассматриваемых функций. Результаты статистического оценивания для двух выборок с растущим курсом приведены в табл. 16.1.

Наилучшие кривые отбирались по величине коэффициента детерминации R^2 , а также значимости параметров α и β , определяемой по величине t -статистики. Как следует из табл. 16.1, для спроса лучшими функциями оказались логарифмическая, параболическая и гиперболическая, а для предложения — линейная, экспоненциальная и параболическая. Этот вывод вполне согласуется с данными корреляционного анализа, поскольку в полученных функциях (за исключением параболы) изменение спроса зависит от темпа, а изменение предложения — от абсолютного прироста курса.

Возможные модели торгов. На основе выявленных лучших кривых можно сформировать наиболее перспективные модели динамики курса, общий вид которых задается соотношением (16.24), а конкретное выражение для параметра A вытекает из функций спроса и предложения. Перечень этих 13 моделей (отобранных из 36) приводится в табл. 16.2, а их аналитическое выражение — в табл. 16.3. Параметры α и β подлежат оцениванию. Во всех моделях табл. 16.3 параметр α отражает чувствительность спроса к увеличению курса, а параметр β — чувствительность предложения.

Статистическое оценивание регрессионных моделей. Результаты статистического оценивания отобранных моделей нелинейным МНК для четырех выборок с растущим курсом доллара приведены в табл. 16.4 (показаны только наиболее удачные результаты).

Как и в случае анализа функций спроса и предложения, наилучшие уравнения (модели) отбирались исходя из величины коэффициента детерминации R^2 , дисперсии ошибки уравнения s^2 и значимости параметров (t -статистики). Принимались также во внимание коэффициент Дарбина-Уотсона (DW) и коэффициент корреляции параметров $\text{cor}(\alpha, \beta)$.

Из представленных в табл. 16.4 наиболее адекватными оказались модели 4, 5 и 12, имеющие наилучшие статистические характеристики. Именно они и были выбраны для прогнозирования величины растущего курса на торгах ММВБ. Расчетные приросты курса (ΔP) для модели 4 в сопоставлении с их фактическими значениями приведены на рис. 16.2.

Статистический анализ зависимости спроса и предложения от валютного курса при $\Delta P > 0$

Зависимость от курса	Уравнение для приростов	Выборка 1.1			Выборка 1.2		
		R^2	α	t-статистика	R^2	α	t-статистика
Спрос							
Линейная	$\Delta D = \alpha \Delta P$	0,775	-0,53	10,7	0,774	-0,52	9,6
Логарифмическая	$\Delta D = \alpha \Delta P / P_0$	0,871	-123,0	14,9	0,897	-131,8	15,3
Экспоненциальная	$\Delta D = \alpha D_0 \Delta P$	0,737	-0,0074	9,6	0,764	-0,0072	9,3
Степенная	$\Delta D = \alpha D_0 \Delta P / P_0$	0,837	-1,79	13,0	0,853	-1,78	12,5
Параболическая	$\Delta D = \alpha \Delta P + \beta P_0 \Delta P$	0,870	$\alpha = -1,1$ $\beta = 0,002$	8,9	0,882	$\alpha = -1,184$ $\beta = 0,002$	8,3
Гиперболическая	$\Delta D = -\alpha \Delta P / P_0^2$	0,780	204447	4,8	0,881	26355	4,9
Предложение							
Линейная	$\Delta S = \alpha \Delta P$	0,601	0,141	7,1	0,647	0,152	7,0
Логарифмическая	$\Delta S = \alpha \Delta P / P_0$	0,442	26,4	5,1	0,555	33,4	5,8
Экспоненциальная	$\Delta S = \alpha S_0 \Delta P$	0,624	0,0062	7,4	0,625	0,0062	6,7
Степенная	$\Delta S = \alpha S_0 \Delta P / P_0$	0,567	1,51	6,6	0,568	1,52	6,0
Параболическая	$\Delta S = \alpha \Delta P + \beta P_0 \Delta P$	0,631	$\alpha = 0,045$ $\beta = 0,004$	0,7	0,655	$\alpha = 0,1$ $\beta = 0,0002$	1,2
Гиперболическая	$\Delta S = -\alpha \Delta P / P_0^2$	0,254	-3508	1,6	0,406	-5729	0,8
				3,3			4,3

Возможные модели торгов с учетом зависимости спроса и предложения от курса доллара

Спрос	Предложение			
	Линейная	Логарифмическая	Экспоненциальная	Параболическая
Линейная	Модель 1			
Логарифмическая	Модель 2	Модель 3	Модель 4	Модель 5
Экспоненциальная			Модель 6	
Степенная	Модель 7		Модель 8	Модель 9
Параболическая				Модель 10
Гиперболическая	Модель 11		Модель 12	Модель 13

Модели торгов, отображенные для статистического оценивания

Модель	Зависимость от курса доллара		Оцениваемое уравнение
	предложения	спроса	
1	Линейная	Линейная	$\Delta P = \frac{1}{\beta - \alpha} (D_0 - S_0 - I)$
2	Линейная	Логарифмическая	$\Delta P = \frac{1}{\beta - \alpha / P_0} (D_0 - S_0 - I)$
3	Логарифмическая	Логарифмическая	$\Delta P = \frac{1}{\beta - \alpha} (D_0 - S_0 - I)$
4	Экспоненциальная	Логарифмическая	$\Delta P = \frac{1}{\beta S_0 - \alpha / P_0} (D_0 - S_0 - I)$
5	Степенная	Логарифмическая	$\Delta P = \frac{P_0}{\beta S_0 - \alpha} (D_0 - S_0 - I)$
6	Экспоненциальная	Экспоненциальная	$\Delta P = \frac{1}{\beta S_0 - \alpha D_0} (D_0 - S_0 - I)$

7	Линейная	Степенная	$\Delta P = \frac{1}{\beta - \alpha D_0 / P_0} (D_0 - S_0 - I)$
8	Экспоненциальная	Степенная	$\Delta P = \frac{1}{\beta S_0 - \alpha D_0 / P_0} (D_0 - S_0 - I)$
9	Степенная	Степенная	$\Delta P = \frac{P_0}{\beta S_0 - \alpha D_0} (D_0 - S_0 - I)$
10	Параболическая	Параболическая	$\Delta P = \frac{1}{\beta P_0 - \alpha} (D_0 - S_0 - I)$
11	Линейная	Гиперболическая	$\Delta P = \frac{1}{\beta - \alpha / P_0^2} (D_0 - S_0 - I)$
12	Экспоненциальная	Гиперболическая	$\Delta P = \frac{1}{\beta S_0 - \alpha / P_0^2} (D_0 - S_0 - I)$
13	Степенная	Гиперболическая	$\Delta P = \frac{P_0}{\beta S_0 - \alpha / P_0} (D_0 - S_0 - I)$

Результаты оценивания регрессионных моделей
на различных выборках при $\Delta P > 0$

Модель	Выборка	Параметры и критерии качества						
		R^2	s^2	DW	α	t_α	β	t_β
1	1.1	0,836	53,5	1,42	-1,243	12,9		
	1.2	0,837	59,6	1,20	-1,247	11,8		
	1.3	0,900	51,0	1,22	-1,497	12,0		
	1.4	0,936	43,8	1,43	-1,701	12,1		
3	1.1	0,895	34,2	2,07	-0,005	16,8		
	1.2	0,912	32,2	2,20	-0,005	16,7		
	1.3	0,918	41,8	1,99	-0,005	13,4		
	1.4	0,924	51,8	1,66	-0,006	11,0		
4	1.1	0,912	29,7	2,02	-146,0	9,8	0,006	2,3
	1.2	0,924	29,1	2,06	-154,1	9,3	0,005	1,8
	1.3	0,938	33,6	1,95	-134,3	6,1	0,007	2,1
	1.4	0,954	34,4	1,87	-127,0	5,9	0,007	2,2

5	1.1	0,912	29,4	2,17	-133,5	7,8	2,244	2,5	0,837
	1.2	0,923	29,3	2,19	-145,1	7,2	1,768	1,9	0,862
	1.3	0,936	35,1	2,20	-126,8	5,0	2,352	2,1	0,885
	1.4	0,956	33,3	2,10	-112,1	4,3	3,110	2,3	0,886
8	1.1	0,820	60,5	2,25	-2,614	6,9	0,005	1,1	0,767
	1.2	0,827	65,6	2,19	-2,509	5,7	0,005	1,1	0,802
	1.3	0,909	49,5	2,36	-1,726	4,4	0,010	2,3	0,849
	1.4	0,943	43,2	3,36	-1,717	4,8	0,010	2,4	0,823
10	1.1	0,889	37,1	1,95	-1,241	8,6	-0,002	4,2	0,954
	1.2	0,897	39,1	1,85	-1,273	7,8	-0,002	4,0	0,959
	1.3	0,921	43,2	1,67	-1,034	5,1	-0,001	2,1	0,972
	1.4	0,939	46,2	1,57	-0,770	3,0	-0,001	0,8	0,981
12	1.1	0,907	31,3	2,23	-25962	8,4	0,013	5,4	0,691
	1.2	0,929	26,9	2,57	-28448	8,8	0,011	5,3	0,708
	1.3	0,938	34,0	2,35	-30519	6,1	0,011	3,9	0,775
	1.4	0,959	31,1	2,10	-30247	6,1	0,011	4,1	0,750

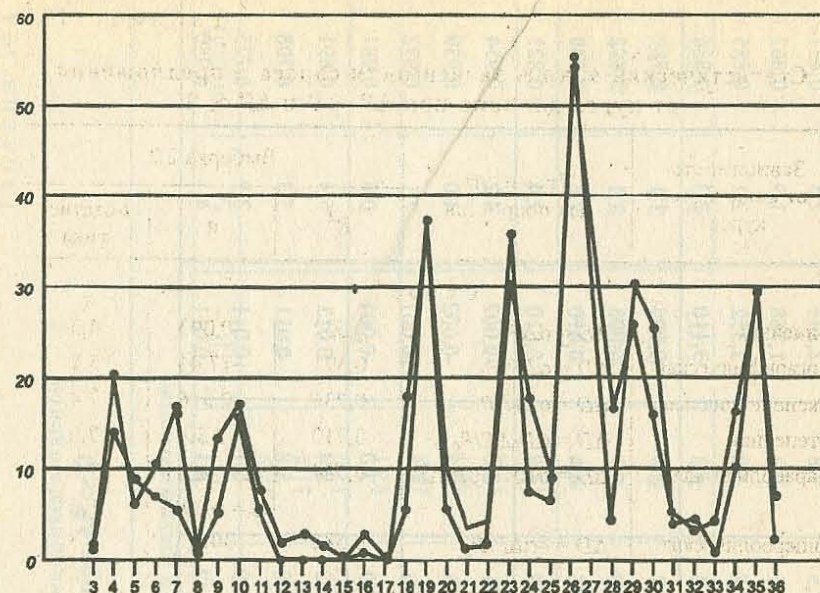


Рис. 16.2. Фактические (кривая со светлыми кружочками) и расчетные (кривая с черными кружочками) приросты курса доллара для модели 4 при $\Delta P > 0$

§ 5. МОДЕЛИ ТОРГОВ ДЛЯ ПАДАЮЩЕГО ВАЛЮТНОГО КУРСА

Зависимость спроса и предложения от курса доллара. Для случая $\Delta P < 0$ поиск наилучших кривых для аппроксимации спроса и предложения в зависимости от изменений курса валют производился теми же методами, которые описаны в § 4 этой главы. Исследование проводилось на выборке 2.2 (январь–июль), поскольку связь курса со спросом и предложением здесь прослеживается наилучшим образом.

Результаты проведенного анализа (табл. 16.5) показывают, что при $\Delta P < 0$ и спрос, и предложение лучше всего описываются экспоненциальной и степенной функциями.

Таблица 16.5

Статистический анализ зависимости спроса и предложения от курса доллара при $\Delta P < 0$ и $\Delta P > 0$

Зависимость от валютного курса	Уравнение для приростов	Выборка 2.2		
		R^2	α	t-статистика
Спрос				
Линейная	$\Delta D = \alpha \Delta P$	0,532	-0,093	4,5
Логарифмическая	$\Delta D = \alpha \Delta P / P_0$	0,571	-17,44	5,2
Экспоненциальная	$\Delta D = \alpha D_0 \Delta P$	0,734	-0,016	7,4
Степенная	$\Delta D = \alpha D_0 \Delta P / P_0$	0,710	-2,50	7,0
Параболическая	$\Delta D = \alpha \Delta P + \beta P_0 \Delta P$	0,584	$\alpha = -0,31$	2,2
			$\beta = 0,0011$	1,5
Гиперболическая	$\Delta D = -\alpha \Delta P / P_0^2$	0,582	3027	5,3
Предложение				
Линейная	$\Delta S = \alpha \Delta P$	0,668	0,21	6,4
Логарифмическая	$\Delta S = \alpha \Delta P / P_0$	0,617	36,44	5,7
Экспоненциальная	$\Delta S = \alpha S_0 \Delta P$	0,740	0,017	7,8
Степенная	$\Delta S = \alpha S_0 \Delta P / P_0$	0,722	3,17	7,2
Параболическая	$\Delta S = \alpha \Delta P + \beta P_0 \Delta P$	0,700	$\alpha = -0,11$	0,5
			$\beta = 0,002$	1,4
Гиперболическая	$\Delta S = -\alpha \Delta P / P_0^2$	0,544	-5888	4,9

Именно эти функции и были использованы при построении моделей падающего курса. Необходимо отметить, что качество аппроксимации функций спроса и предложения при падающем курсе хуже, чем при растущем. Соответственно несколько хуже оказалось и качество окончательных прогнозных моделей.

Статистическое оценивание регрессионных моделей. В табл. 16.6 представлены результаты статистического оценивания для выборок с падающим курсом доллара по моделям 6, 8 и 9, которые оказались наиболее удачными. Для прогнозирования курса при $\Delta P < 0$ окончательно отобраны модели 6 и 8. График фактических и расчетных приростов курса (ΔP) для модели 6 представлен на рис. 16.3.

Результаты оценивания регрессионных моделей
на различных выборках при $\Delta P < 0$

Модель	Выборка	Параметры и критерии качества									
		R^2	s^2	DW	α	t_α	β	t_β	$\text{cor}(\alpha, \beta)$		
6	2.1	0,860	21,9	2,14	-0,030	3,4	0,014	3,5	0,836		
	2.2	0,871	24,8	2,18	-0,030	3,2	0,014	3,3	0,834		
8	2.1	0,860	22,0	2,17	-4,605	3,3	0,161	4,7	0,783		
	2.2	0,870	25,1	2,15	-4,517	3,1	0,016	4,4	0,776		
9	2.2	0,857	27,5	2,16	-4,151	2,3	3,406	4,2	0,823		

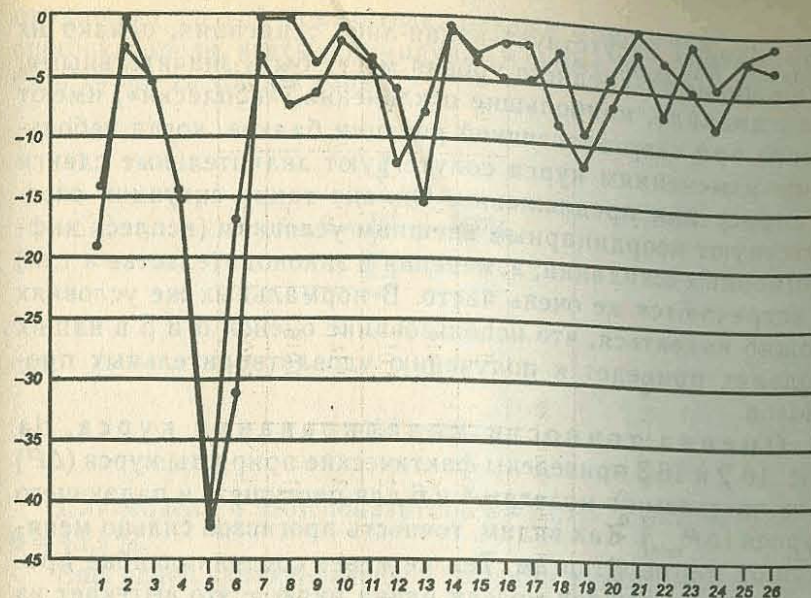


Рис. 16.3. Фактические (кривая со светлыми кружочками) и расчетные (кривая с черными кружочками) снижения курса доллара для модели 6 при $\Delta P < 0$

§ 6. ОЦЕНКА КАЧЕСТВА СТАТИСТИЧЕСКИХ МОДЕЛЕЙ

Анализ устойчивости параметров.

Проблема устойчивости параметров регрессионных моделей — одна из ключевых при эконометрическом прогнозировании. Если параметры устойчивы, то при использовании соответствующей модели можно рассчитывать на получение достаточно достоверного прогноза, в противном случае прогноз может оказаться ошибочным, если не предусмотрена корректировка и адаптация параметров.

Значения параметров α и β для моделей 4 (выборка $\Delta P > 0$) и 6 (выборка $\Delta P < 0$), рассчитанные непосредственно по формулам для приростов спроса и предложения в каждой точке, показали, что в обоих случаях в динамике

параметров отсутствует какая-либо тенденция, однако их отклонения от среднего уровня могут быть значительными. Как правило, наибольшие отклонения («всплески») имеют место при сверхэластичной реакции банков, когда небольшим изменениям курса сопутствуют значительные сдвиги в спросе или предложении. Однако такие ситуации соответствуют неординарным внешним условиям (всплеск инфляционных ожиданий, изменения в законодательстве и т.п.) и встречаются не очень часто. В нормальных же условиях можно надеяться, что использование оценок α и β в наших моделях приведет к получению удовлетворительных прогнозов.

Оценка точности моделирования курса. На рис. 16.2 и 16.3 приведены фактические приросты курса (ΔP) и их прогнозы по моделям 4 и 6 для растущего и падающего курсов (ΔP_{mod}). Как видим, точность прогнозов сильно меняется от торгов к торгам. Тем не менее средняя ошибка прогноза ΔP по каждой модели равна нулю — это вытекает из способа оценивания параметров регрессионных моделей. Поэтому требуется какой-то иной устойчивый показатель точности, характеризующий качество прогнозов в среднем по выборке. Точность прогнозирования по регрессионным моделям может оцениваться по-разному, в связи с чем рассмотрим несколько показателей.

Одним из них является *стандартная ошибка уравнения*. В наших моделях она представляется стандартной ошибкой прогноза прироста курса на торгах. В модели 4 она равна $\sigma_1 = 5,33$ руб., а в модели 6 $\sigma_2 = 4,58$ руб.

Вызывает интерес также и *относительная ошибка прогноза*. Иногда для ее измерения берут отношение абсолютной ошибки модели к реальному значению объясняемой переменной (т.е. в нашем случае отношение модуля ошибки к реальному изменению курса на данных торгах) и усредняют эту статистику по выборке. Однако здесь подобный показатель не годится, так как если на торгах t ошибка была отличной от нуля, а курс почти не изменился, то относительная ошибка будет очень большой, и это проявится при усреднении по всей выборке.

Для того чтобы получить представление об относительной точности построенных регрессионных моделей, нами

предложен другой показатель, равный отношению суммы ошибок модели, взятых по модулю, к сумме модулей приростов курса за тот же период:

$$\theta = \frac{\sum_{t=1}^T |\varepsilon_t|}{\sum_{t=1}^T |\Delta P_t|} 100\%,$$

где

$$\varepsilon_t = \Delta P_t - \Delta P_{\text{mod}t}.$$

Для модели 4 этот показатель $\theta = 22,7\%$, а для модели 6 $\theta = 45,1\%$.

Еще одной характеристикой точности модели может быть ошибка прогноза курса, равная ошибке прогноза прироста курса, так как $P_t = P_{t-1} + \Delta P_t$, а относительная погрешность здесь может быть измерена статистикой

$$\lambda = \frac{1}{T} \sum_{t=1}^T \left| \frac{\varepsilon_t}{P_t} \right| 100\%.$$

Для модели 4 $\lambda_1 = 1,85\%$, для модели 6 $\lambda_2 = 2,15\%$.

Таким образом, можно сделать следующие выводы.

1. Ожидаемая стандартная ошибка модельного прогноза равна $\sigma_1 = 5,33$, если курс растет, или $\sigma_2 = 4,58$, когда он снижается.

2. Относительный показатель ошибки моделирования прироста курса в среднем на выборке равен $\theta_1 = 22,7\%$ при растущем курсе и $\theta_2 = 45,1\%$ — при падающем. Это свидетельствует о том, что модели возрастающего курса можно доверять в большей степени, чем модели снижающегося курса, хотя и здесь ошибка достаточно велика.

3. Относительная ошибка прогноза абсолютного уровня курса в среднем равна $\lambda_1 = 1,85\%$, если он растет, и $\lambda_2 = 2,15\%$, когда он снижается.

§ 7. ВОЗМОЖНЫЕ СПОСОБЫ ИСПОЛЬЗОВАНИЯ МОДЕЛЕЙ

Прогнозирование курса накануне торгов на ММВБ. Как указывалось в § 2 этой главы, все многообразие моделей торгов описывается так:

$$\Delta P_t = A_t(D_{0t} - S_{0t} - I), \quad (16.25)$$

где коэффициент A_t в общем случае не является константой и может быть функцией от некоторых известных переменных (в частности, начального спроса и предложения банков и курса предыдущих торгов). Оценивание моделей заключается в том, чтобы на ретроспективном статистическом материале (на данных о прошлых торгах) найти оценки коэффициента A_t . Получив эти оценки на основе регрессионного анализа, можно использовать их для прогнозирования курса на предстоящих торгах в зависимости от заявок банков на продажу и покупку, а также от различных вариантов вмешательства ЦБ РФ.

Регулирование курса с помощью интервенций ЦБ РФ. Центральный банк Российской Федерации является на торгах регулирующим органом. В первой половине 1992 г. его функция выражалась в сдерживании роста курса доллара, а впоследствии — в сглаживании его колебаний. Возникла задача — оценить затраты ЦБ РФ по регулированию курса и выбору наилучшей стратегии.

Обратимся к выражению (16.25). Очевидно, что здесь коэффициент A_t , изменяющийся во времени, может быть интерпретирован как мера чувствительности курса доллара к интервенциям ЦБ РФ, показывающая, какое снижение курса доллара приходится на единицу продаж ЦБ РФ на торгах t , так как

$$\frac{\partial \Delta P_t}{\partial I_t} = -A_t. \quad (16.26)$$

В связи с этим динамика A_t представляет интерес. В результате можно сделать важное заключение: чем выше A_t , тем большего эффекта (при прочих равных условиях) достигнут продажи ЦБ РФ.

Отметим, что, если переписать (16.25) в виде

$$I_t = (D_{0t} - S_{0t}) - \frac{\Delta P_t}{A_t}, \quad (16.27)$$

где A_t определяется типом модели, то можно получать оценки величины интервенции ЦБ РФ, необходимой для удержания изменения курса в заданных пределах.

* * *

В данной главе достаточно подробно изложен порядок построения модели валютных торгов на ММВБ. Начав с простейшей модели с постоянным параметром A , постепенно, на основе анализа с использованием микроэкономических подходов, были получены альтернативные модели с переменным параметром A . Это привело к новой задаче — проблеме идентификации, т.е. к выбору наилучшей модели. Такой выбор был сделан на основе статистических показателей качества моделей. Насколько нам известно, эти модели несколько лет использовались для практических целей регулируемыми органами.

Разумеется, статистическая модель описывает наиболее устойчивые связи переменных. У эксперта информации гораздо больше, поэтому последнее решающее слово остается за экспертом, за «лицом, принимающим решение», а статистические модели можно рассматривать как инструмент получения дополнительной информации об интересующем объекте или процессе.

ЗАКЛЮЧЕНИЕ

Заканчивая рассмотрение адаптивных методов прогнозирования, отметим их некоторые особенности. Суть адаптации состоит в том, что модель следует за процессом. Это обуславливает отставание изменений в модели от новых тенденций в реальном процессе, и чем больше время упреждения, тем больше несоответствие между прогнозом и фактическим значением ряда. Следовательно, модели рассматриваемого класса можно рекомендовать для получения в основном краткосрочных прогнозов.

Многие из рассмотренных моделей характеризуют связь между исследуемой величиной и временем. Это обстоятельство само по себе является довольно серьезным ограничением. С другой стороны, время в модели выражает эволюцию всего комплекса условий протекания процесса. Через время исходный ряд неявно связан с множеством взаимосвязанных факторов, учесть влияние которых порознь затруднительно. За счет упрощенного представления исследуемой величины, связанной с одним лишь фактором времени, моделирование становится возможным даже при самой скудной информации. Положительной чертой адаптивных методов является то, что с их помощью тщательно изучается внутренняя структура временного ряда, взаимосвязь его последовательных членов, а модели, являющиеся инструментом прогноза, чутко реагируют на динамические изменения и соответственно перестраиваются тем или иным образом, учитывая обесценение устаревшей информации.

Сфера применения адаптивных моделей одномерного ряда довольно обширна. Они могут быть использованы для прогнозирования спроса и предложения, конъюнктурных колебаний финансового рынка, отдельных экономических и

технико-экономических показателей, уровня запасов в системах материально-технического снабжения, для прогнозирования структурных и технологических сдвигов, для определения траекторий некоторых глобальных показателей.

Наиболее простыми являются полиномиальные модели Брауна. Их отличает ясность концепции, достаточная гибкость, универсальность и простота расчетов. Особое значение имеют адаптивные модели для более гибкого, чем обычно, выравнивания рядов, очищения их от случайных наслоений. Однако точность получаемых по ним прогнозов не всегда удовлетворительна. Более строгим является метод, основывающийся на теории, разработанной Дж. Боксом и Г. Дженкинсом, — метод построения смешанных моделей авторегрессии-скользящего среднего. Имеются сведения [85], что, несмотря на ограниченность используемой информации, модели Бокса — Дженкинса в деле прогнозирования могут с успехом конкурировать со сложными эконометрическими моделями. Очевидно, не случайно положительно показала себя в опытах по прогнозированию экономических рядов и адаптивная модель авторегрессии (например, для прогнозирования курсов валют), и адаптивная гистограмма (для анализа кассовых остатков).

Рассмотренные в данном пособии методы анализа и построения моделей изолированного временного ряда могут найти применение и в качестве вспомогательного средства. Например, для прогнозирования экзогенных переменных при построении многофакторных моделей, а это очень важно, так как от точности прогноза экзогенных переменных зависит успех прогнозирования по многофакторной модели в целом.

Модели Бокса — Дженкинса можно также использовать как своеобразную меру динамических свойств ряда, и это их качество положено в основу анализа эконометрических систем линейных динамических уравнений. С помощью таких моделей можно также получить ряды с некоторыми заранее заданными автокорреляционными свойствами, что может найти применение в имитационных моделях.

Получил развитие адаптивный многомерный анализ: разработаны адаптивный корреляционный и адаптивный регрессионный анализ. Применение моделей АРСС для отражения колебаний дисперсии регрессионных остатков при-

вело к появлению моделей с авторегрессионной условной гетероскедастичностью. В экономике большую роль играют циклические колебания показателей. Колебания такого рода имеют несистематический характер: у них переменные амплитуда и период, деформированные положительная и отрицательная фазы. Для исследования подобных колебательных процессов нами предложен метод фазового анализа.

Этим, однако, не ограничивается область использования адаптивных моделей. Представим себе случай, когда для отображения реального процесса построена какая-либо сколь угодно сложная модель. Для лучшего ее согласования с реальными данными можно в качестве корректирующего элемента использовать *адаптивную модель*. Ее назначение состоит в том, чтобы прогнозировать ошибку, даваемую основной моделью, и вносить соответствующую корректировку. Это значительно расширяет сферу применения адаптивных моделей.

Тем не менее некоторые направления адаптивного моделирования, по-видимому, не находят широкого применения в экономических исследованиях по объективным причинам. Например, вызывает сомнение возможность плодотворного развития метода, связывающего адаптивные свойства модели с изменениями в спектральных характеристиках динамического ряда, так как оценки спектральной плотности, определяемые на нескольких десятках точек, весьма ненадежны.

Как мы видели, в адаптивном моделировании используется целый ряд подходов. Это обогащает арсенал исследователя, но одновременно создает трудности при выборе метода в конкретном случае. Многие методы сравнительно новые, и поэтому еще не накоплен достаточный практический опыт работы с ними, в связи с чем в пособии особое внимание уделено изучению и сопоставлению свойств различных методов и моделей, что должно помочь делать правильный выбор с учетом целей и ресурсов исследователя.

Полагаем, что в условиях рыночной экономики адаптивные методы экономических показателей являются наиболее адекватным средством прогнозирования.

ПРИЛОЖЕНИЯ

№ 1

ДИСПЕРСИЯ ОШИБКИ ПРОГНОЗИРОВАНИЯ СТОХАСТИЧЕСКОГО ПРОЦЕССА ТЕЙЛА — ВЕЙДЖА С ПОМОЩЬЮ ЛИНЕЙНОГО ПРЕДИКТОРА ХОЛЬТА

Г. Тейлом и С. Вейджем [103] было предложено выражение для оценки дисперсии ошибки прогноза на один шаг вперед ($\tau = 1$). Здесь мы выводим более общую формулу для любого $\tau \geq 1$. Напомним вкратце модель.

Процесс

$$\left. \begin{aligned} x_t &= a_{1,t} + \varepsilon_t, \\ a_{1,t} &= a_{1,t-1} + a_{2,t}, \\ a_{2,t} &= a_{2,t-1} + v_t; \end{aligned} \right\} \quad (1)$$

$$M(\varepsilon_t \varepsilon_{t'}) = \begin{cases} \sigma_\varepsilon^2 & t = t', \\ 0 & t \neq t'; \end{cases} \quad M(v_t v_{t'}) = \begin{cases} \sigma_v^2 & t = t', \\ 0 & t \neq t'. \end{cases} \quad (2)$$

$M(\varepsilon_t v_{t'}) = 0$ для любой пары (t, t') .

Схема прогнозирования:

$$\hat{a}_{1,t} = \hat{a}_{1,t-1} + \hat{a}_{2,t-1} + \alpha_1 \varepsilon_1(t-1); \quad (3)$$

$$\hat{a}_{2,t} = \hat{a}_{2,t-1} + \alpha_1 \alpha_2 \varepsilon_1(t-1); \quad (4)$$

$$\hat{x}_\tau(t) = \hat{a}_{1,t} + \tau \hat{a}_{2,t},$$

где

$$0 < \alpha_1, \alpha_2 < 1;$$

$\varepsilon_1(t-1) = x_t - \hat{x}_1(t-1)$ — ошибка прогноза, сделанного в момент $t-1$ на один шаг вперед.

Ошибка прогноза, сделанного в момент t на τ шагов вперед, будет:

$$e_{\tau}(t) = x_{t+\tau} - \hat{x}_{\tau}(t) = \left(a_{1,t} + \tau a_{2,t} + \sum_{i=1}^{\tau} (\tau - i + 1) \times \right. \\ \left. \times v_{t+i} + \varepsilon_{t+\tau} \right) - \hat{a}_{1,t} - \tau \hat{a}_{2,t} = A_t + \tau B_t + \\ + \varepsilon_{t+\tau} + \sum_{i=1}^{\tau} (\tau - i + 1) v_{t+i}, \quad (5)$$

где

$$A_t = a_{1,t} - \hat{a}_{1,t}; \quad B_t = a_{2,t} - \hat{a}_{2,t}.$$

Ошибка прогноза имеет три источника: ошибку оценки коэффициента a_1 в момент t , ошибку оценки коэффициента a_2 в момент t и кумулятивную сумму случайных компонент за промежуток времени $(t, t + \tau)$ с учетом их влияния на точность прогноза.

Используя уравнения адаптации (3), (4) и (5), можно найти выражения для ошибок моделирования A_t и B_t .

Запишем (3) в виде:

$$a_{1,t} - \hat{a}_{1,t} - (a_{1,t-1} - \hat{a}_{1,t-1}) - (a_{1,t} - a_{1,t-1} - a_{2,t-1}) = \\ = -\alpha_1 e_1(t-1).$$

Затем после простых преобразований для A_t может быть записано следующее соотношение:

$$A_t = (1 - \alpha_1)(A_{t-1} + B_{t-1}) - \alpha_1 \varepsilon_t + (1 - \alpha_1) v_t. \quad (6)$$

Тем же путем для B_t получаем

$$B_t = -\alpha_1 \alpha_2 A_{t-1} + (1 - \alpha_1 \alpha_2) B_{t-1} - \alpha_1 \alpha_2 \varepsilon_t + (1 - \alpha_1 \alpha_2) v_t. \quad (7)$$

Запишем (6) и (7) в векторной форме:

$$\begin{bmatrix} A_t \\ B_t \end{bmatrix} = P \begin{bmatrix} A_{t-1} \\ B_{t-1} \end{bmatrix} - Q \begin{bmatrix} \varepsilon_t \\ v_t \end{bmatrix}, \quad (8)$$

где

$$P = \begin{bmatrix} 1 - \alpha_1 & 1 - \alpha_1 \\ -\gamma & 1 - \gamma \end{bmatrix}; \quad Q = \begin{bmatrix} \alpha_1 & -(1 - \alpha_1) \\ \gamma & -(1 - \gamma) \end{bmatrix}; \\ \gamma = \alpha_1 \alpha_2.$$

Используя соотношение (8) n раз, получаем

$$\begin{bmatrix} A_t \\ B_t \end{bmatrix} = P^n \begin{bmatrix} A_{t-n} \\ B_{t-n} \end{bmatrix} - \sum_{k=0}^{n-1} P^k Q \begin{bmatrix} \varepsilon_{t-k} \\ v_{t-k} \end{bmatrix}. \quad (9)$$

Первый член справа сходится к нулю при $n \rightarrow \infty$ (см. [103]). Объединяем (9) и (5) и получаем

$$e_{\tau}(t) = -[1\tau] \sum_{k=0}^{\infty} P^k Q \begin{bmatrix} \varepsilon_{t-k} \\ v_{t-k} \end{bmatrix} + \varepsilon_{t+\tau} + \sum_{i=1}^{\tau} (\tau - i + 1) v_{t+i}.$$

Возводим обе части в квадрат и берем математическое ожидание:

$$D_e(\tau) = M[e_{\tau}^2(t)] = [1\tau] (QDQ' + PQDQ'P' + \\ + P^2QDQ'P'^2 + \dots) \begin{bmatrix} 1 \\ \tau \end{bmatrix} + \sigma_{\varepsilon}^2 + \sigma_v^2 \sum_{i=1}^{\tau} i^2, \quad (10)$$

где

$$D = \begin{bmatrix} \sigma_{\varepsilon}^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix}.$$

Выражение (10) можно упростить. Пусть

$$S = QDQ' + PQDQ'P' + P^2QDQ'P'^2 + \dots,$$

вычтем

$$PSP' = PQDQ'P' + P^2QDQ'P'^2 + \dots$$

В результате получим

$$S - PSP' = QDQ'. \quad (11)$$

Выражение (11) можно рассматривать как систему линейных уравнений с тремя элементами s_{11} , s_{12} , s_{22} матрицы S . Ясно, что S симметрична, т. е. $s_{21} = s_{12}$.

Можно (11) записать в развернутой форме:

$$\begin{bmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{bmatrix} - \begin{bmatrix} 1 - \alpha_1 & 1 - \alpha_1 \\ -\gamma & 1 - \gamma \end{bmatrix} \begin{bmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{bmatrix} \begin{bmatrix} 1 - \alpha_1 & -\gamma \\ 1 - \alpha_1 & 1 - \gamma \end{bmatrix} = \\ = \begin{bmatrix} \alpha_1 & -(1 - \alpha_1) \\ \gamma & -(1 - \gamma) \end{bmatrix} \begin{bmatrix} \sigma_{\varepsilon}^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix} \begin{bmatrix} \alpha_1 & \gamma \\ -(1 - \alpha_1) & -(1 - \gamma) \end{bmatrix}.$$

что эквивалентно следующему:

$$\begin{bmatrix} \alpha_1(2-\alpha_1) & -2(1-\alpha_1)^2 & -(1-\alpha_1)^2 \\ (1-\alpha_1)\gamma & \alpha_1 + 2\gamma - 2\alpha_1\gamma & -(1-\alpha_1)(1-\gamma) \\ -\gamma^2 & 2\gamma(1-\gamma) & \gamma(2-\gamma) \end{bmatrix} \times \\ \times \begin{bmatrix} s_{11} \\ s_{12} \\ s_{22} \end{bmatrix} = \begin{bmatrix} \alpha_1^2 \sigma_e^2 + (1-\alpha_1)^2 \sigma_v^2 \\ \alpha_1 \gamma \sigma_e^2 + (1-\alpha_1)(1-\gamma) \sigma_v^2 \\ \gamma^2 \sigma_e^2 + (1-\gamma)^2 \sigma_v^2 \end{bmatrix}.$$

Решение будет такое:

$$\begin{bmatrix} s_{11} \\ s_{12} \\ s_{22} \end{bmatrix} = \frac{1}{\alpha_1 \gamma (4-2\alpha_1-\gamma)} \times \\ \times \begin{bmatrix} (2-\alpha_1)\gamma & 2(1-\alpha_1)^2\gamma & (1-\alpha_1)^2(2-\alpha_1) \\ -(1-\alpha_1)\gamma^2 & (4\alpha_1-2\alpha_1^2-\gamma)\gamma & (1-\alpha_1)(2\alpha_1-\alpha_1^2-\gamma) \\ (2-\alpha_1)\gamma^2 & 2(\alpha_1^2-2\alpha_1+\gamma)\gamma & 2\alpha_1^2-\alpha_1^3+2\gamma-2\alpha_1\gamma \end{bmatrix} \times \\ \times \begin{bmatrix} \alpha_1^2 \sigma_e^2 + (1-\alpha_1)^2 \sigma_v^2 \\ \alpha_1 \gamma \sigma_e^2 + (1-\alpha_1)(1-\gamma) \sigma_v^2 \\ \gamma^2 \sigma_e^2 + (1-\gamma)^2 \sigma_v^2 \end{bmatrix}.$$

Возвращаясь к (10) и (11), видим, что дисперсия ошибки

$$D_e(\tau) = s_{11} + 2\tau s_{12} + \tau^2 s_{22} + \sigma_e^2 + \sigma_v^2 \sum_{i=1}^{\tau} i^2.$$

После подстановки получим искомый результат:

$$D_e(\tau) = [((2-\alpha_1)\gamma + 2\tau(\alpha_1-1)\gamma + \tau^2\gamma(2-\alpha_1))\gamma \times \\ \times (\alpha_1^3 - (1-\alpha_1)^3 g^2) + (2\gamma - 4\alpha_1 + 2\alpha_1^2 + 2\tau(4\alpha_1 - 2\alpha_1^2 - \gamma) + \\ + \tau^2(2\alpha_1^2 - 4\alpha_1 + 2\gamma))\gamma \times (\alpha_1\gamma + (1-\alpha_1)(1-\gamma)g^2) + \\ + (2 - 5\alpha_1 + 4\alpha_1^2 - \alpha_1^3 + 2\tau(2\alpha_1 - 3\alpha_1^2 + \alpha_1^3 + \alpha_1\gamma - \gamma) + \\ + \tau^2(2\alpha_1^2 - \alpha_1^3 + 2\gamma - 2\alpha_1\gamma)) \times (\gamma^2 + \\ + (1-\gamma)^2 g^2)] / (\alpha_1 \gamma (4-2\alpha_1-\gamma)) + 1 + g^2 \sum_{i=1}^{\tau} i^2 \sigma_e^2,$$

где

$$g^2 = \frac{\sigma_v^2}{\sigma_e^2}; \quad \sigma_e^2 = \frac{1-h}{1+h} D_e(1).$$

Дисперсия ошибки прогноза на один шаг вперед $D_e(1)$ находится экспериментально в результате оптимизации параметров модели.

№ 2

СОВМЕСТНОЕ АПОСТЕРИОРНОЕ РАСПРЕДЕЛЕНИЕ ПАРАМЕТРОВ ЛИНЕЙНОЙ МОДЕЛИ

В приложении дан вывод результатов, используемых в гл. 6. В разделе А рассматривается общий случай, в разделе Б — частный случай линейной модели прогнозирования с двумя параметрами.

А. Общий случай

Рассмотрим модель, генерирующую временной ряд, которая в момент t имеет вид:

$$X_t = F A_t + \varepsilon_t; \quad (A.1)$$

$$A_t = L A_{t-1} + C u_t, \quad (A.2)$$

где X_t является вектором значений, наблюдаемых в момент t размерности $(l \times 1)$. Двумерные матрицы L , C и F имеют размерности $(n \times n)$, $(n \times p)$ и $(l \times n)$ соответственно. Вектор параметров A_t имеет размерность $(n \times 1)$. Вектора ε_t и u_t размерности $(l \times 1)$ и $(p \times 1)$ соответственно являются векторами случайных независимых переменных с математическим ожиданием 0. Предполагается, что законы распределения вероятностей ε_t и u_t и матрицы F , L и C , соответствующие каждому периоду времени j ($j \leq t+1$), известны, но не оговорено как они изменяются во времени. $X(t-1)$ определяется как матрица, состоящая из векторов прошлых наблюдений X_{t-1-i} ($i = 1, \dots, \infty$), а $\hat{A}_{t-1}$ и ε_t определены как:

$$\hat{A}_{t-1} = M[A_{t-1} | X(t-1)]; \quad (A.3)$$

$$\varepsilon_t = X_t - F \hat{A}_{t-1}. \quad (A.4)$$

По формуле Байеса, записывая $X(t-1)$ как X , совместное распределение вероятностей равно:

$$p(A_t | X_t, X(t-1)) = k p(X_t | A_t, X) p(A_t | X), \quad (A.5)$$

где

$$k = \frac{1}{p(X_t | X)}. \quad (A.6)$$

Так как знание X_t и X эквивалентно знанию e_t и X :

$$p(A_t | X_t, X) = p(A_t | e_t, X) \quad (A.7)$$

и

$$p(A_t | X_t, X) = k p(e_t | A_t, X) p(A_t | X). \quad (A.8)$$

Теперь рассмотрим случай нормальных законов распределения, в котором:

$$e_t \sim N(0; W); \quad (A.9)$$

$$u_t \sim N(0; U); \quad (A.10)$$

$$(A_{t-1} | X) \sim N(\hat{A}_{t-1}; B_{t-1}). \quad (A.11)$$

Из уравнений (A.2) и (A.3) непосредственно следует, что

$$M[A_t | X] = L\hat{A}_{t-1}$$

и, так как u_t не зависит от X , ковариационная матрица $(A_t | X)$ является суммой ковариационных матриц LA_{t-1} и Cu_t .

Обозначая эту сумму через R , получим

$$R = LB_{t-1}L' + CUC'. \quad (A.12)$$

Таким образом,

$$(A_t | X) \sim N(L\hat{A}_{t-1}; R). \quad (A.13)$$

Вычитая $FL\hat{A}_{t-1}$ из обеих частей уравнения (A.1) и используя уравнение (A.4), получаем

$$e_t = F(A_t - L\hat{A}_{t-1}) + \varepsilon_t, \quad (A.14)$$

поэтому

$$M[e_t | A_t, X] = F(A_t - L\hat{A}_{t-1})$$

и ковариационные матрицы e_t и ε_t идентичны, так что

$$(e_t | A_t, X) \sim N[F(A_t - L\hat{A}_{t-1}); W]. \quad (A.15)$$

Так как

$$p(A_t, e_t | X) = p(e_t | A_t, X) p(A_t | X), \quad (A.16)$$

то, используя результаты, полученные в [2] для многомерного нормального распределения, получаем, что

$$(A_t, e_t | X) \sim N(L\hat{A}_{t-1}, 0; V), \quad (A.17)$$

где V может быть записана в расчлененном виде как:

$$V = \begin{bmatrix} R & RF' \\ FR & W + FRF' \end{bmatrix}. \quad (A.18)$$

Аналогично, используя статистические результаты, введенные для условных многомерных нормальных распределений, получаем требуемое распределение вероятностей A_t при условии, что заданы все наблюдения до X_t включительно, как условное распределение при заданных e_t и X . Таким образом,

$$(A_t | X(t)) = (A_t | e_t, X(t-1)) \sim N(L\hat{A}_{t-1} + RF'[W + FRF']^{-1}e_t; B_t), \quad (A.19)$$

где

$$B_t = R - RF'[W + FRF']^{-1}FR. \quad (A.20)$$

По (A.14) и (A.13) можно заметить, что

$$(e_t | X(t-1)) \sim (0; W + FRF'),$$

так что $W + FRF'$ является ковариационной матрицей прогноза, сделанного в момент $(t-1)$ на один шаг вперед.

Б. Частный случай

Рассмотрим процесс:

$$x_t = f_t a_{1,t} + e_t;$$

$$a_{1,t} = a_{1,t-1} + a_{2,t} + u_t;$$

$$a_{2,t} = a_{2,t-1} + v_t,$$

где e_t, u_t, v_t являются независимыми нормальными случайными переменными с нулевыми математическими ожиданиями и дисперсиями D_e, D_u, D_v .

В обозначениях рассмотренного выше общего случая:

$$X_t = x_t;$$

$$F = (f_t, 0);$$

$$A_t' = (a_{1,t}, a_{2,t});$$

$$u_t' = (u_t, v_t);$$

$$e_t = \varepsilon_t;$$

$$L = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix};$$

$$C = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix};$$

$$W = D_e;$$

$$U = \begin{bmatrix} D_u & 0 \\ 0 & D_v \end{bmatrix}.$$

Из результатов, полученных в общем случае, следует, что если $x(t)$ представляет собой все наблюдаемые значения до момента t включительно и если

$$(a_{1,t-1}, a_{2,t-1} | x(t-1)) \sim N(a_{1,t-1}, a_{2,t-1}; B_{t-1}),$$

где

$$B_{t-1} = \begin{bmatrix} m_{t-1}(a_1, a_1) & m_{t-1}(a_1, a_2) \\ m_{t-1}(a_1, a_2) & m_{t-1}(a_2, a_2) \end{bmatrix},$$

то из (19)

$$(a_{1,t}, a_{2,t} | x(t)) \sim N(a_{1,t}, a_{2,t}; B_t),$$

где

$$a_{1,t} = a_{1,t-1} + a_{2,t-1} + \alpha_1 e_t;$$

$$a_{2,t} = a_{2,t-1} + \alpha_2 e_t$$

и

$$\alpha_1 = f_t r_{11} / D_e;$$

$$\alpha_2 = f_t r_{12} / D_e,$$

а $D_e = f_t^2 r_{11} + D_e$ является дисперсией ошибки прогноза на один шаг вперед.

Величины r_{ij} определены (A.12) как

$$R = \begin{bmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{bmatrix},$$

где

$$r_{11} = m_{t-1}(a_1, a_1) + 2m_{t-1}(a_1, a_2) + m_{t-1}(a_2, a_2) + D_u + D_v;$$

$$r_{12} = m_{t-1}(a_1, a_2) + m_{t-1}(a_2, a_2) + D_v;$$

$$r_{22} = m_{t-1}(a_2, a_2) + D_v.$$

Используя уравнение (A.20):

$$B_t = R - f_t^2 \begin{bmatrix} r_{11}^2 & r_{11} r_{12} \\ r_{11} r_{12} & r_{12}^2 \end{bmatrix} / D_e$$

и обозначение:

$$B_t = \begin{bmatrix} m_t(a_1, a_1) & m_t(a_1, a_2) \\ m_t(a_1, a_2) & m_t(a_2, a_2) \end{bmatrix},$$

получаем:

$$m_t(a_1, a_1) = r_{11} - \alpha_1^2 D_e;$$

$$m_t(a_1, a_2) = r_{12} - \alpha_1 \alpha_2 D_e;$$

$$m_t(a_2, a_2) = r_{22} - \alpha_2^2 D_e.$$

№ 3

ОЦЕНКА ПАРАМЕТРОВ В НЕЛИНЕЙНОЙ РЕГРЕССИИ

Очень часто при оценивании экономико-математических моделей нелинейные модели приходится сводить к линейным и оценивать уже эти производные модели с помощью линейных методов.

Предположим, что необходимо оценить параметры производственной функции типа Кобба — Дугласа:

$$Y = A e^{\kappa t} K^\alpha L^{1-\alpha}. \quad (1)$$

Относительно неизвестных параметров A, κ, α это уравнение нелинейно и непосредственно применять метод наименьших квадратов (МНК) нельзя. Уравнение (1) преобразуют в линейное путем деления на L и логарифмирования:

$$\frac{Y}{L} = A e^{\kappa t} \left(\frac{K}{L} \right)^\alpha;$$

$$\ln\left(\frac{Y}{L}\right) = \ln A + \kappa t + \alpha \ln\left(\frac{K}{L}\right). \quad (2)$$

Последнее уравнение уже допускает использование обычного МНК для оценки параметров $\ln A, \kappa$ и α . Однако если не принять определенных мер (см. [12]¹), то оценки

¹ Так как указанная работа опубликована малым тиражом и широкому кругу читателей не известна, повторяем ход рассуждений.

параметров вследствие проведенных преобразований окажутся смещенными.

Запишем (1) в общем виде:

$$Y_i = f(X_i; \beta) + \varepsilon_i, \quad i = 1, \dots, N,$$

или

$$Y_i = \hat{Y}_i + \varepsilon_i, \quad (3)$$

где $\hat{Y}_i = f(X_i; \beta)$ — теоретическое значение Y_i ;

Y — эндогенная переменная модели;

X — вектор экзогенных переменных;

β — вектор неизвестных оцениваемых параметров;

ε — случайная ошибка.

Разделим обе части (3) на элемент некоторого ряда R_i и получим новое уравнение:

$$\frac{Y_i}{R_i} = \frac{f(X_i; \beta)}{R_i} + \frac{\varepsilon_i}{R_i} \quad (4)$$

или

$$Y_i' = \hat{Y}_i' + u_i.$$

Применение МНК к уравнению (4) означает минимизацию суммы квадратов новых отклонений u_i :

$$Q'(\beta) = \sum (Y_i' - \hat{Y}_i')^2 = \sum u_i^2 = \sum \frac{\varepsilon_i^2}{R_i^2}. \quad (5)$$

Из этого уравнения видно, что после деления обеих частей уравнения (3) на R_i мы приходим к минимизации суммы квадратов отклонений, отличной от $Q = \sum \varepsilon_i^2$, которая минимизируется при непосредственном применении МНК к уравнению (3). Поскольку $Q' \neq Q$, то и вектор оцененных параметров β' будет отличен от β .

Значения R_i^2 являются как бы весами по отношению к ошибкам ε_i^2 . Следовательно, большим значениям R_i^2 будет соответствовать меньшее слагаемое ε_i^2/R_i^2 , входящее в общую сумму Q' . В том случае, когда ряд R возрастающий, ошибки, соответствующие более ранним моментам времени, входят в общую сумму Q' с большим весом, а ошибки недалекого прошлого с меньшим весом, т. е. уравнение регрессии (4), будет лучше приближать значения Y_i , наблюдавшиеся в первые моменты времени. Таким образом, можно говорить о смещении уравнения регрессии в прошлое.

Аналогичное явление наблюдается в случае оценки уравнения регрессии линейной в логарифмах. Пусть исходное нелинейное уравнение линейно в логарифмах. Производное уравнение тогда можно представить в виде:

$$\ln Y_i = \ln \hat{Y}_i + u_i$$

или

$$Y_i' = \hat{Y}_i' + u_i.$$

Применение МНК к последнему уравнению означает минимизацию выражения

$$Q'(\beta) = \sum (Y_i' - \hat{Y}_i')^2 = \sum (\ln Y_i - \ln \hat{Y}_i')^2 = \sum \ln^2 \frac{\hat{Y}_i}{Y_i},$$

т. к. $\hat{Y}_i$ является приближением Y_i , то можно считать, что в среднем $Y_i \approx \hat{Y}_i$, поэтому

$$\ln \frac{\hat{Y}_i}{Y_i} \approx \frac{\hat{Y}_i}{Y_i} - 1 = \frac{\hat{Y}_i - Y_i}{Y_i}$$

и

$$Q'(\beta) = \sum \frac{(\hat{Y}_i - Y_i)^2}{Y_i^2} = \sum \frac{\varepsilon_i^2}{Y_i^2}.$$

Последнее уравнение аналогично уравнению (5), но с той разницей, что ошибки ε_i входят с весами Y_i^2 , а не с R_i^2 . Если предположить, что ряд Y возрастающий, то Q' дает смещение оценок параметров в прошлое и линия регрессии тоже будет смещена.

Рассмотрим подробнее вопрос о том, с какими весами входят ошибки ε_i в минимизируемую сумму квадратов отклонений для производственной функции (1).

После деления на L_i ошибки ε_i входят с весом $1/L_i$, т. е. $\varepsilon_i' = \varepsilon_i/L_i$. После логарифмирования получаем новую ошибку:

$$\varepsilon_i'' = \varepsilon_i'/(Y_i/L_i) = (\varepsilon_i/L_i)(L_i/Y_i) = \varepsilon_i/Y_i.$$

Итак, окончательно $\varepsilon_i'' = \varepsilon_i/Y_i$. Поэтому, чтобы вернуться к минимизации исходной суммы квадратов отклонений $Q = \sum \varepsilon_i^2$, нужно обе части уравнения (2) предварительно умножить на Y и оценить параметры $\ln A$, α и β .

СТАТИСТИЧЕСКИЕ ВРЕМЕННЫЕ РЯДЫ, ИСПОЛЬЗОВАННЫЕ ДЛЯ ИСПЫТАНИЯ МОДЕЛЕЙ

I. Курс акций фирмы ИБМ, дол.

№ точки	Дата	Курс	№ точки	Дата	Курс	№ точки	Дата	Курс
	1960			Сентябрь				
1	Июль		39	1	545	78	26	509
2	11	510	40	2	546	79	27	515
3	12	497	41	5	545	80	28	507
4	13	504	42	6	536	81	31	511
5	14	510	43	7	534		Ноябрь	
6	15	509	44	8	541	82	1	521
7	18	503	45	9	541	83	2	520
8	19	500	46	12	541	84	3	524
9	20	500	47	13	538	85	4	527
10	21	500	48	14	531	86	7	528
11	22	495	49	15	521	87	8	529
12	25	494	50	16	523	88	9	541
13	26	499	51	19	501	89	10	537
14	27	502	52	20	505	90	14	549
15	28	509	53	21	520	91	15	551
	29	525	54	22	521	92	16	541
16	Август		55	23	511	93	17	547
17	1	512	56	26	504	94	18	559
18	2	510	57	27	507	95	21	559
19	3	506	58	28	502	96	22	560
20	4	515	59	29	505	97	23	556
21	5	522	60	30	510	98	25	560
22	8	523		Октябрь		99	28	558
23	9	527	61	8	509	100	29	558
24	10	523	62	4	507	101	30	557
25	11	528	63	5	508		Декабрь	
26	12	529	64	6	509	102	1	553
27	15	538	65	7	518	103	2	554
28	16	539	66	10	515	104	5	555
29	17	541	67	11	520	105	6	562
30	18	543	68	12	519	106	7	569
31	19	541	69	13	526	107	8	585
32	22	539	70	14	529	108	9	590
33	23	543	71	17	528	109	12	596
34	24	548	72	18	527	110	13	589
35	25	550	73	19	524	111	14	583
36	26	548	74	20	515	112	15	584
37	29	550	75	21	509	113	16	597
38	30	544	76	24	504	114	19	591
	31	536	77	25	502	115	20	591

Продолжение

№ точки	Дата	Курс	№ точки	Дата	Курс	№ точки	Дата	Курс
116	21	589	128	10	599	143	31	636
117	22	581	129	11	601		Февраль	
118	23	588	130	12	608			
119	27	592	131	13	622			
120	28	597	132	16	613	144	1	637
121	29	594	133	17	618	145	2	641
122	30	593	134	18	622	146	3	640
	1961		135	19	620	147	6	643
	Январь		136	20	624	148	7	644
			137	23	628	149	8	650
123	3	583	138	24	628	150	9	644
124	4	592	139	25	641	151	10	641
125	5	591	140	26	630	152	13	641
126	6	597	141	27	638	153	14	642
127	9	597	142	30	647	154	15	643

Источник. Brown R. G. Smoothing forecasting and prediction of discrete time series. N. Y., 1963, Table C. 5, p. 422.

II. Цены на золото на Лондонском рынке, дол. за унцию

№ точки	Дата	Цена	№ точки	Дата	Цена	№ точки	Дата	Цена
	1971							
	Апрель		18	28	39,15	35	21	41,05
1	1	38,95	19	29	39,42	36	24	40,75
2	2	38,82	20	30	39,60	37	25	40,97
3	5	38,90		Май		38	26	40,96
4	6	38,90	21	3	39,42	39	27	40,70
5	7	38,92	22	4	39,86	40	28	40,80
6	8	38,95	23	5	39,66		Июнь	
7	13	38,90	24	6	40,20	41	1	40,85
8	14	38,90	25	7	39,87	42	2	40,65
9	15	38,89	26	10	39,65	43	3	40,52
10	16	38,93	27	11	40,00	44	4	40,60
11	19	38,95	28	12	40,70	45	7	40,50
12	20	38,95	29	13	40,85	46	8	40,50
13	21	38,90	30	14	40,80	47	9	39,95
14	22	38,97	31	17	41,10	48	10	39,60
15	23	39,08	32	18	41,20	49	11	39,65
16	26	39,00	33	19	40,50	50	14	39,60
17	27	38,91	34	20	40,80	51	15	39,49

№ точки	Дата	Цена	№ точки	Дата	Цена	№ точки	Дата	Цена
52	16	39,47	92	11	42,75	132	8	42,50
53	17	39,95	93	12	42,23	133	11	42,45
54	18	39,90	94	13	43,27	134	12	42,52
55	21	37,95	95	17	43,05	135	13	42,60
56	22	41,10	96	18	43,00	136	14	42,54
57	23	40,25	97	19	43,40	137	15	42,57
58	24	40,40	98	20	43,30	138	18	42,51
59	25	40,11	99	23	43,25	139	19	42,55
60	28	40,08	100	24	43,10	140	20	42,50
61	29	39,87	101	25	43,10	141	21	42,52
62	30	40,20	102	26	42,75	142	22	42,47
Июль			103	27	41,40	143	25	42,47
			104	31	41,00	144	26	42,50
			Сентябрь			145	27	42,51
						146	28	42,25
63	1	40,10				147	29	42,40
64	2	40,41				Ноябрь		
65	5	40,22	105	1	40,70			
66	6	40,30	106	2	41,60			
67	7	40,20	107	3	41,72			
68	8	40,27	108	6	41,80	148	1	42,12
69	9	40,24	109	7	41,70	149	2	42,40
70	12	40,35	110	8	41,32	150	3	42,20
71	13	40,35	111	9	41,70	151	4	42,15
72	14	40,40	112	10	41,57	152	5	42,25
73	15	40,44	113	13	41,59	153	8	42,35
74	16	40,60	114	14	41,85	154	9	42,47
75	19	41,01	115	15	41,89	155	10	42,62
76	20	40,97	116	16	42,22	156	11	42,55
77	21	41,20	117	17	42,10	157	12	42,55
78	22	41,59	118	20	42,05	158	15	42,85
79	23	41,80	119	21	42,30	159	16	43,00
80	26	41,50	120	22	42,20	160	17	42,81
81	27	41,90	121	23	42,25	161	18	42,80
82	28	42,10	122	24	42,65	162	19	43,15
83	29	42,30	123	27	42,35	163	22	43,25
84	30	42,47	124	28	42,72	164	23	43,20
Август			125	29	42,70	165	24	43,37
			126	30	42,47	166	25	43,64
			Октябрь			167	26	43,32
						168	29	43,65
85	2	42,47				169	30	43,55
86	3	42,15				Декабрь		
87	4	42,19	127	1	42,17			
88	5	42,80	128	4	42,50			
89	6	42,65	129	5	42,35			
90	9	42,90	130	6	42,44	170	1	43,50
91	10	43,30	131	7	42,45	171	2	43,60

№ точки	Дата	Цена	№ точки	Дата	Цена	№ точки	Дата	Цена
172	3	43,32	Февраль			252	28	48,22
173	6	43,90						
174	7	43,83						
175	8	43,60						
176	9	43,50	215	4	47,55	Апрель		
177	10	43,67	216	7	47,45			
178	13	43,65	217	8	48,20			
179	14	43,75	218	9	47,92			
180	15	43,61	219	10	48,40	255	4	48,40
181	16	43,40	220	11	48,25	256	5	48,40
182	17	43,05	221	14	48,25	257	6	48,30
183	20	42,40	222	15	47,98	258	7	48,35
184	21	42,95	223	16	48,05	259	10	48,40
185	22	43,15	224	17	48,50	260	11	48,45
186	23	43,52	225	18	48,24	261	12	48,81
187	28	43,20	226	21	48,40	262	13	49,42
188	29	43,52	227	22	48,60	263	14	49,35
189	30	43,50	228	23	48,62	264	17	49,35
190	31	43,64	229	24	48,70	265	18	49,47
1972 Январь			230	25	48,60	266	19	49,25
			231	28	48,42	267	20	49,42
			232	29	48,40	268	21	49,37
			Март			269	24	49,35
191	3	43,72				270	25	49,21
192	4	44,40				271	26	49,30
193	5	44,42				272	27	49,37
194	6	44,55	233	1	47,80	273	28	49,50
195	7	45,75	234	2	48,20	Май		
196	10	45,70	235	3	48,05			
197	11	45,34	236	6	48,00			
198	12	45,25	237	7	48,27			
199	13	45,45	238	8	48,25	274	1	50,20
200	14	45,72	239	9	48,30	275	2	50,70
201	17	45,97	240	10	48,20	276	3	50,30
202	18	46,09	241	13	48,35	277	4	50,35
203	19	45,90	242	14	48,40	278	5	50,52
204	20	46,75	243	15	48,30	279	8	51,60
205	21	45,89	244	16	48,47	280	9	54,00
206	24	45,92	245	17	48,40	281	10	53,30
207	25	46,10	246	20	48,35	282	11	53,00
208	26	47,07	247	21	48,40	283	12	52,80
209	27	46,20	248	22	48,40	284	15	52,80
210	28	47,00	249	23	48,40	285	16	54,50
211	31	47,95	250	24	48,40	286	17	57,50
			251	27	48,35	287	18	57,15
						288	19	56,50
						289	22	57,20
						290	23	57,50

Продолжение

№ точки	Дата	Цена	№ точки	Дата	Цена	№ точки	Дата	Цена
291	24	56,65	318	Июль 3	64,25	345	9	65,70
292	25	56,92	319	4	65,50	346	10	66,15
293	26	57,50	320	5	65,30	347	11	66,10
294	30	58,70	321	6	65,90	348	14	64,75
295	31	59,30	322	7	65,80	349	15	65,75
Июнь			323	10	65,05	350	16	65,80
			324	11	65,70	351	17	66,25
			325	12	64,85	352	18	66,25
			326	13	65,10	353	21	66,60
296	1	58,80	327	14	65,20	354	22	66,97
297	2	59,20	328	17	65,20	355	23	67,40
298	5	59,40	329	18	65,60	356	24	67,30
299	6	59,75	330	19	65,10	357	25	67,05
300	7	64,75	331	20	64,70	358	29	66,60
301	8	66,75	332	21	65,30	359	30	66,80
302	9	65,60	333	24	65,10	360	31	66,80
303	12	59,00	334	25	65,45	Сентябрь		
304	13	63,30	335	26	65,57			
305	14	60,70	336	27	65,65	361	1	66,80
306	15	61,60	337	28	67,45	362	4	66,35
307	16	61,80	338	31	68,90	363	5	66,85
308	19	60,60	Август			364	6	67,00
309	20	61,60				365	7	66,97
310	21	61,12				366	8	67,10
311	22	61,47				367	11	67,10
312	23	64,00	340	2	70,00	368	12	67,00
313	26	63,50	341	3	66,70	369	13	67,95
314	27	62,30	342	4	67,75	370	14	66,85
315	28	63,30	343	7	63,40			
316	29	64,10	344	8	66,25			
317	30	64,10						

Источник. Московский Народный Банк, Лондон, «Экономическая информация» за соответствующий период.

III. Среднемесячные котировки Лондонской биржи металлов на свинец, дол. за т

Месяц	1960	1961	1962	1963	1964	1965
Январь	206	173	163	150	218	349
Февраль	204	180	162	151	223	391
Март	210	182	167	154	224	396
Апрель	214	185	167	160	227	353
Май	213	184	165	167	244	303
Июнь	202	179	159	179	247	279
Июль	196	179	148	179	271	272
Август	195	179	141	185	301	272
Сентябрь	193	177	143	184	308	289
Октябрь	135	173	146	192	332	307
Ноябрь	188	167	150	191	366	300
Декабрь	179	167	153	205	385	301

Месяц	1966	1967	1968	1969	1970	1971
Январь	301	220	219	254	325	265
Февраль	292	222	228	258	335	268
Март	293	229	234	261	336	270
Апрель	281	227	235	268	321	270
Май	257	229	238	278	313	267
Июнь	259	228	239	286	308	268
Июль	264	231	247	311	301	264
Август	264	233	250	316	286	259
Сентябрь	253	230	252	302	285	239
Октябрь	234	226	248	295	285	229
Ноябрь	223	226	258	316	280	221
Декабрь	224	223	250	334	277	231

Источник. Lead and Zinc statistics, Monthly Bulletin of the International Lead and Zinc Study Group, N. Y., 1961—1971.

IV. Некоторые статистические ряды, характеризующие развитие народного хозяйства бывшего СССР

Номер выборки	Ряд	Единица измерения	Годы	Количество точек
1	Поставка минеральных удобрений сельскому хозяйству	тыс. т усл. ед.	1946–1975	30
2	Производство электроэнергии	млрд. кВт·ч	1945–1975	31
3	Производство гидроэлектроэнергии	»	1945–1975	31
4	Сумма вкладов населения в сберкассы	млн. руб.	1950–1975	26
5	Численность городского населения на 1 января	млн. чел.	1950–1976	27
6	Протяженность нефте- и нефтепродуктопроводов	тыс. км	1955–1975	21
7	Грузооборот нефте- и нефтепродуктопроводов	ткм	1955–1975	21
8	Урожайность зерна	ц/га	1946–1975	30
9	Урожайность картофеля	»	1946–1975	30
10	Урожайность овощей	»	1946–1975	30
11	Производство яиц	млрд. шт.	1946–1975	30

Источники. Статистические ежегодники «Народное хозяйство СССР» за соответствующие годы.

V. Валовые инвестиции в 1960–1988 гг. в постоянных ценах, млрд. дол.*

Год	США	Западная Европа
1960	324,16	329,77
1961	324,01	346,85
1962	358,79	357,89
1963	380,47	358,37
1964	400,98	414,83
1965	448,53	432,50

Продолжение

Год	США	Западная Европа
1966	472,57	446,36
1967	465,21	453,84
1968	493,93	474,55
1969	504,55	533,09
1970	460,37	589,93
1971	506,17	565,85
1972	543,46	563,13
1973	592,31	622,73
1974	552,49	659,49
1975	460,14	545,11
1976	528,35	612,37
1977	598,39	576,17
1978	654,37	534,92
1979	642,02	623,47
1980	577,15	663,32
1981	622,65	525,82
1982	522,77	569,41
1983	567,81	533,04
1984	739,75	569,60
1985	742,39	585,57
1986	744,54	629,35
1987	785,11	664,46
1988	814,13	720,89

* Прирост запасов + валовые капитальные вложения (Increment in stocks + Gross fixed capital formation) пересчитаны в постоянных ценах в долларах США 1985 г.

Источники. National Accounts, Main Aggregates, 1960–1988. – Paris, 1990. – V.1.

ЛИТЕРАТУРА

К главам 1 – 8

1. Аллен Р. Дж. Математическая экономия. — М.: ИЛ, 1963.
2. Андерсон Т. Введение в многомерный статистический анализ. — М.: Физматгиз, 1963.
3. Андерсон Т. Статистический анализ временных рядов. — М.: Мир, 1976.
4. Бартлетт М.С. Введение в теорию случайных процессов. — М.: ИЛ, 1958.
5. Бокс Лж., Дженкинс Г. Анализ временных рядов, прогноз и управление. — М.: Мир, 1974. — Вып. 1.
6. Браун М. Теория и измерение технического прогресса. — М.: Статистика, 1971.
7. Винер Н. Кибернетика или управление и связь в животном и машине. 2-е изд. — М.: Советское радио, 1968.
8. Гельфонд А.О. Исчисление конечных разностей. 3-е изд. — М.: Наука, 1967.
9. Геминтерн В.И., Френкель А.А. Обобщенные модели авторегрессии и скользящего среднего в анализе временных рядов // Статистический анализ экономических временных рядов и прогнозирование. — М.: Наука, 1973.
10. Гренджер К., Хатанака М. Спектральный анализ временных рядов в экономике. — М.: Статистика, 1972.
11. Грицевич И.Г. Моделирование временных рядов с помощью схем Бокса — Дженкинса // Математические методы в экономике и международных отношениях. Вып. 3. Методы эконометрического моделирования. — М.: ИМЭМО АН СССР, 1974.
12. Демиденко Е.З. Оценки параметров в нелинейной регрессии. Серия «Математические методы в экономике и международных отношениях». Вып. «Проблемы эконометрического моделирования». — М.: ИМЭМО АН СССР, 1972.
13. Дженкинс Г., Ваттс Д. Спектральный анализ и его приложения. Вып. 1 и 2. — М.: Мир, 1971, 1972.

14. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. — М.: Статистика, 1973.
15. Кейн Э. Экономическая статистика и эконометрия. Вып. 1 и 2. — М.: Статистика, 1977.
16. Кендалл М. Дж., Стьюарт А. Статистические выводы и связи. — М.: Наука, 1973.
17. Кендалл М. Дж., Стьюарт А. Многомерный статистический анализ и временные ряды. — М.: Наука, 1976.
18. Кильдишев Г.С., Френкель А.А. Анализ временных рядов и прогнозирование. — М.: Статистика, 1973.
19. Клеандров Д.И., Френкель А.А. Прогнозирование экономических показателей с помощью метода простого экспоненциального сглаживания // Статистический анализ экономических временных рядов и прогнозирование. — М.: Наука, 1973.
20. Лэнинг Дж.Х., Бэттин Р.Г. Случайные процессы в задачах автоматического управления. — М.: ИЛ, 1958.
21. Лукашин Ю.П. Анализ временных рядов по методу интегрированной авторегрессии — скользящей средней. Статистические методы анализа (алгоритмы и программы). Вып. 5. — М.: ИМЭМО АН СССР, 1975.
22. Лукашин Ю.П. Программы адаптивных методов прогнозирования на Фортране-IV. — М.: ИМЭМО АН СССР, 1978.
23. Маленко Э. Статистические методы эконометрии. Вып. 1 и 2. — М.: Статистика, 1975, 1976.
24. Моррис У. Наука об управлении, байесовский подход. — М.: Мир, 1971.
25. Налимов В.В. Теория эксперимента. — М.: Наука, 1971.
26. Налимов В.В., Чернова Н.А. Статистические методы планирования экстремальных экспериментов. — М.: Наука, 1965.
27. Прогнозирование капиталистической экономики, проблемы методологии/Отв. ред. А.И. Шапиро. — М.: Мысль, 1970.
28. Рывкин А.А., Сараджишвили В.Ю. Адаптивные методы прогнозирования экономических временных рядов. Материалы к симпозиуму «Актуальные проблемы моделирования капиталистической экономики». Вып. 1. — М. — 1968.
29. Смирнов А.Д. К проблеме оптимального экономического прогнозирования // Экономика и математические методы. — Вып. 5. — М. — 1966. — Т. 2.
30. Смирнов А.Д. Моделирование и прогнозирование социалистического воспроизводства. — М.: Экономика, 1970.

31. Тинтнер Г. Введение в эконометрию. — М.: Статистика, 1965.
32. Тутубалин В.Н. Статистическая обработка рядов наблюдений. — М.: Знание, 1973.
33. Френкель А. Математические методы анализа динамики и прогнозирования производительности труда. — М.: Экономика, 1972.
34. Четыркин Е.М. Статистические методы прогнозирования. — 2-е изд. — М.: Статистика, 1977.
35. Эшби У. Росс. Конструкция мозга, происхождение адаптивного поведения. — М.: Мир, 1964.
36. Яглом А.М. Корреляционная теория процессов со случайными стационарными n -ми приращениями. Матем. сб. — Т. 37 (79), № 1. — М.: Изд-во АН СССР, 1955.
37. Amemiya T., Wu R.Y. The effect of aggregation on prediction in the autoregressive model. — J. Amer. Statist. Ass., 1972. — Vol. 67. — PP. 628–632.
38. Bailey M.G. Prediction of an autoregressive variables subject both to disturbances and to errors of observation // J. Amer. Statist. Ass. — 1965. Vol. 60. — PP. 164–181.
39. Batty M. Monitoring an exponential smoothing forecasting system // Oper. Res. Quart. — 1969. — Vol. 20. — № 3.
40. Bates J.M., Granger C.W.J. The combination of forecasts // Oper. Res. Quart. — 1969. — Vol. 20. — № 4.
41. Bhattacharyya M.N. Forecasting the demand for telephones in Australia // Applied Statistics. — 1974. — Vol. 23. — № 1.
42. Bossons J. The effects of parameter misspecification and nonstationarity on the applicability of adaptive forecasts // Management Science. — 1966. — Vol. 12. — № 9.
43. Box G.E.P. Evolutionary operation: a method for increasing industrial productivity // Applied Statistics. — 1957. — Vol. 6. — № 2.
44. Box G.E.P., Jenkins G.M. Some statistical aspects of adaptive optimization and control // J. of the Royal Stat. Soc. — 1962. — Ser. B. — Vol. 24. — № 2.
45. Box G.E.P., Pierce D.A. Distribution of residual autocorrelation in autoregressive — integrated moving average time series models // J. Amer. Statist. Ass. — 1970. — Vol. 65. — PP. 64.
46. Brown R.G. Statistical forecasting for inventory control. — N.Y., 1959.

47. Brown R.G. Smoothing forecasting and prediction of discrete time series. — N.Y., 1963.
48. Brown R.G., Meyer R.F. The fundamental theorem of exponential smoothing // Oper. Res. — 1961. — Vol. 9. — № 5.
49. Chatfield C., Prothero D.L. Box-Jenkins seasonal forecasting problems in a case study // J. of the Royal Stat. Soc., 1973. — Ser. A. — Vol. 136. — PP. 295–352.
50. Chow W.M. Adaptive control of the exponential smoothing constant. // J. of Indust. Engineering. — 1965. — Vol. 16. — № 5.
51. Cogger K.O. The optimality of general-order exponential smoothing // Oper. Res. — 1974. — Vol. 22. — № 4.
52. Cohen G.D. A note on exponential smoothing and autocorrelated inputs // Oper. Res. — 1963. — Vol. 11. — № 3.
53. Cox D.R. Prediction by exponentially weighted moving averages and related methods // J. of the Royal Stat. Soc. — 1961. — Vol. 23. — № 2.
54. Davies N., Pate M.B., Frost M.G. Maximum autocorrelations for moving average processes // Biometrika. — 1974. — Vol. 61. — № 1.
55. Dobbie J.M. Forecasting periodic trends by exponential smoothing // Oper. Res. — 1963. — Vol. 11. — PP. 908–918.
56. Durbin J. Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables // Econometrica. — 1970. — Vol. 38. — PP. 410–421.
57. Eilon S., Elmaleh J. Adaptive limits in inventory control // Management Science. — 1970. — Vol. 16. — № 8.
58. D'Esopo D.A. A note on forecasting by the exponential smoothing operator // Oper. Res. — 1961. — Vol. 9. — № 5.
59. Fagan T.L., Wilson M.A. Exponential smoothing for prediction of reliability growth // Industrial Quality Control. — 1967. — Vol. 23. — № 7.
60. Geoffrian A.M. A summary of exponential smoothing // J. of Industr. Engineering. — 1962. — Vol. 13. — № 4.
61. Goodman M.L. A new look at higher-order exponential smoothing // Oper. Res. — 1974. — Vol. 24. — № 4.
62. Granger C.W., Morris M.J. Time series modelling and interpretation // J. of the Royal Stat. Soc. — 1976. — Ser. A. — Vol. 139. — Part. 2.
63. Green M., Harisson P.J. Fashion forecasting for a mail order company using a Bayesian approach // Oper. Res. Quart. — 1973. — Vol. 24. — № 2.

64. *Harisson P.J., Davies O.L.* The use of cumulative sum (CUSUM) techniques for the control of routine forecasts of product demand // *Oper. Res.* — 1964. — Vol. 12. — № 2.
65. *Harisson P.J.* Short-term sales forecasting. — *Applied Statistics, J. of the Royal Stat. Soc.* — 1965, — Ser. C. — Vol. 14. — № 2,3.
66. *Harisson P.J., Stevens C.F.* A Bayesian approach to short-term forecasting // *Oper. Res. Quart.* — 1971. — Vol. 22. — № 4.
67. *Harisson P.J.* Exponential smoothing and short-term sales forecasting // *Management Science.* — 1967. — Vol. 13. — № 11.
68. *Holt C.C.* Forecasting trends and seasonals by exponentially weighted moving averages // *O.N.R. Memorandum, Carnegie Inst. of Technology.* — 1957. — № 2.
69. *Kalman R.E.* New methods in Wiener filtering // *Proceedings of the first symposium on engineering applications of random function theory and probability: N.Y., L., 1964.*
70. *Kalman R.E., Bucy R.S.* New result in linear filtering and prediction theory // *J. of Basic Engineering (Trans. ASME).* — 1961. — Vol. 83 D.
71. *Kendall M.G.* Time-series. — L.: 1973.
72. *Landi D.M., Johnson B. McK.* Optimal linear inventory control and first order exponential smoothing // *Oper. Res.* — 1967. — Vol. 15. — № 2.
73. *Markland R.E.* A comparative study of demand forecasting techniques for military helicopter spare parts // *Naval Research Logistics Quarter.* — 1970. — Vol. 17. — № 1.
74. *Marquardt D.W.* An algorithm for least-squares estimation of nonlinear parameters // *J. of Soc. for Industr. and Appl. Math.* — 1963. — Vol. 11. — № 2.
75. *McClain J.O.* Dynamics of exponential smoothing with trends and seasonal terms // *Management Science.* — 1974. — Vol. 20. — PP. 1300–1304.
76. *McClain J.O., Thomas L.J.* Response-variance trade-offs in adaptive forecasting // *Oper. Res.* — 1973. — Vol. 21. — PP. 554–568.
77. *McKenzie Ed.* An analysis of general exponential smoothing // *Oper. Res.* — 1976. — Vol. 24. — № 1.
78. *McKenzie Ed.* A comparison of standard forecasting systems with the Box-Jenkins Approach // *The Statistician.* — 1974. — Vol. 23. — № 2.
79. *Montgomery D.C.* An introduction to short-term forecasting // *J. of Indust. Engineering.* — 1968. — Vol. 19. — № 10.
80. *Montgomery D.C.* Adaptive control of exponential smoothing parameters by evolutionary operation // *AIIE Transactions.* — 1970. — Vol. 2. — № 3.
81. *Montgomery D.C.* A note on forecasting with adaptive filtering // *Oper. Res. Quart.* — 1977. — Vol. 28. — № 1i.
82. *Morris R.N., Glassey C.R.* The dynamics and statistics of exponential smoothing operators // *Oper. Res.* — 1963. — Vol. 11. — PP. 561–569.
83. *Muir A.* Automatic sales forecasting // *Brit. Computer J.* — 1958. — Vol. 1. — P. 113.
84. *Muth J.F.* Optimal properties of exponentially weighted forecasts of time series with permanent and transitory components // *J. Amer. Statist. Ass.* — 1960. — Vol. 55. — P. 299.
85. *Nelson Ch.R.* The prediction performance of the FRB-MIT-PENN model of the U.S. economy // *The American Economic Review.* — 1972. — Vol. LXII. — № 5.
86. *Nelson Ch.R.* Applied time series analysis for managerial forecasting. — San Francisco, 1973.
87. *Nelson Ch.R.* The first order moving average process: identification, estimation and prediction // *J. of Econometrics.* — 1974. — Vol. 2. — PP. 121–141.
88. *Nelson Ch.R.* The interpretation of R^2 in autoregressive-moving average time series models // *The American Statistician.* — 1976. — Vol. 30. — № 4.
89. *Nerlove M., Wage S.* The optimality of adaptive forecasting // *Management Science.* — 1964. — Vol. 10. — № 2.
90. *Newbold P., Granger C.W.J.* Experience with forecasting univariate time series and the combination of forecasts // *J. of the Royal Stat. Soc.* — 1974. — Ser. A. — Vol. 137. — Part. 2.
91. *Packer A.H.* Simulation and adaptive forecasting as applied to inventory control // *Oper. Res.* — 1967. — Vol. 15. — № 4.
92. *Pandit S.M., Wu S.M.* Exponential smoothing as a special case of a linear stochastic system // *Oper. Res.* — 1974. — Vol. 24. — № 4.
93. *Regels C.C.* Exponential forecasting: some new variations // *Management Science.* — 1969. — Vol. 15. — № 5.
94. *Quenouille M.H.* Approximate tests of correlation in time series // *J. of the Royal Stat. Soc.* — 1949. Ser. B. — Vol. 11. — № 68.
95. *Quenouille M.H.* The analysis of multiple time series. — L.: 1957.

96. Rao A.G., Shapiro A. Adaptive smoothing using evolutionary spectra // Management Science. — 1970. — Vol. 17. — № 3.
- 96a. Letters. Management Science. — 1971. — Vol. 18. — № 1.
97. Roberts S.D., Reed R.Jr. The development of a self-adaptive forecasting technique // AIIE Transaction. — 1969. — Vol. 1. — № 4.
98. Roberts S.W. Control chart tests based on geometric moving averages // Technometrics. — 1959. — Vol. 1. — PP. 239–250.
99. Shone M.L. Viewpoint // Oper. Res. Quart. — 1967. — Vol. 18. — P. 318.
100. Smith T.M.F. A comparison of some models for predicting time series subject to seasonal variation // The Statistician. — 1967. — Vol. 17. — PP. 301–305.
101. Spendley W., Hext G.R., Himsworth F.R. Sequential application of simplex designs in optimization and evolutionary operation // Technometrics. — 1962. — Vol. 4. — № 4.
102. Taylor C.J. A simple graphical method of exponential smoothing with a linear trend // Oper. Res. Quart. — 1967. — Vol. 18. — № 1.
103. Theil H., Wage S. Some observations on adaptive forecasting // Management Science. — 1964. — Vol. 10. — № 2.
104. Trigg D.W. Monitoring a forecasting system // Oper. Res. Quart. — 1964. — Vol. 15. — № 3.
105. Trigg D.W., Leach A.G. Exponential smoothing with an adaptive response rate // Oper. Res. Quart. — 1967. — Vol. 18. — № 1.
106. Wade R.C. A technique for initializing exponential smoothing forecasts // Management Science. — 1967. — Vol. 13. — № 7.
107. Wagle B.V. A review of two statistical aids in forecasting // The Statistician. — 1965. — Vol. 15. — № 2.
108. Ward D.H. Comparison of different systems of exponentially weighted prediction // The Statistician. L. — 1963. — Vol. 13. — № 3.
109. Wheelwright S.C., Makridakis S. Forecasting methods for management. — N.Y.: 1973.
110. Wheelwright S.C., Makridakis S. Forecasting with adaptive filtering. — Revue Française d'Automatique d'Informatique et de Recherche Operationelle. — Paris: 1973. Ser. V. — Vol. 7. — № 1.
111. Wheelwright S.C., Makridakis S. An examination of the use of adaptive filtering in forecasting // Oper. Res. Quart. — 1973. — Vol. 24. — № 1.

112. Winer N. Extrapolation, interpolation and smoothing of stationary time series. — N.Y.: 1949.
113. Winters P.R. Forecasting sales by exponentially weighted moving averages // Management Science. — 1960. — Vol. 6. — № 3.
114. Zadeh L.A. On the definition of adaptivity // Proceedings of the IRE. — 1963. — Vol. 51. — № 3.
115. Zellner A., Palm F. Time series analysis and simultaneous equation econometric models // J. Econometrics. — 1974. — Vol. 2. — № 1.

К главам 9 – 16

116. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики: Учебник. — М.: ЮНИТИ, 1998.
117. Алберт А. Регрессия, псевдорегрессия и рекуррентное оценивание. — М.: Наука, 1977.
118. Дугерти К. Введение в эконометрику, университетский учебник / Пер. с англ. — М.: Инфра-М, 1999.
119. Канторович Г.Г. Анализ временных рядов // Экономический журнал Высшей школы экономики. — 2002. — Т. 6. — № 1. — С. 85–116. — № 2. — С. 251–273. — № 3. — С. 379–401.
120. Кендэл М. Временные ряды. — М.: Финансы и статистика, 1981.
121. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования. — М.: Статистика, 1979.
122. Лукашин Ю.П. Адаптивный корреляционный анализ экономических показателей // Вестник статистики. — 1991. — № 9. — С. 78–79.
123. Лукашин Ю.П. Анализ распределения кассовых остатков: адаптивная гистограмма, проблема оптимизации // Экономика и математические методы. — 1997. — Т. 33. — Вып. 3. — С. 90–97.
124. Лукашин Ю.П. Линейная регрессия с переменными параметрами. — М.: Финансы и статистика, 1992. — 256 с.
125. Лукашин Ю.П. Нетрадиционный корреляционный анализ временных рядов // Экономика и математические методы. — 1992. — Т. 28. — Вып. 3. — С. 406–413.
126. Лукашин Ю.П. О возможности краткосрочного прогнозирования курсов валют с помощью простейших статистических моделей // Вестник МГУ. — 1990. — Сер. 6. Экономика. — № 1. — С. 75–84.

127. Лукашин Ю.П. Оптимизация структуры портфеля ценных бумаг // Экономика и математические методы. — 1995. — Т. 31. — Вып. 1. — С. 138–150.
128. Лукашин Ю.П. Проверка гипотез в эконометрике. — М.: ИМЭМО РАН, 2002.
129. Лукашин Ю.П. Фазовый анализ временных рядов // Экономика и математические методы. — 1993. — Т. 29. — Вып. 3. — С. 503–511.
130. Лукашин Ю.П., Лушин А.С. Статистическое моделирование торгов на Московской межбанковской валютной бирже // Экономика и математические методы. — 1994. — Т. 30. — Вып. 3. — С. 84–97.
131. Эконометрика: Учебник/Под ред. И.И. Елисеевой. — М.: Финансы и статистика, 2002.
132. Akaike H. Information theory and the extension of the maximum likelihood principle. /In 'Proceeding of the Second International symposium on information theory', eds. B.N. Petrov and F. Csaki, Budapest. — 1973. — PP. 267–281.
133. Akaike H. A new look at the statistical identification model // IEEE: Trans. auto. Control. — 1974. — V. 19. — PP. 716–723.
134. Bewley R. The direct estimation of the equilibrium response in a linear dynamic model // Economics letters. — 1979. — V. 55. — PP. 251–276.
135. Bollerslev T. Generalized autoregressive conditional heteroscedasticity // Journal of econometrics. — 1986. — V. 31. — PP. 307–327.
136. Davidson J.E.H., Hendry D.F., Srba F., Yeo S. Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom // Economic journal. — 1978. — V. 88 (352). — PP. 661–692.
137. Dickey D.A., Fuller W.A. Distribution of the estimators for autoregressive time-series with a unit root // Journal of the American statistical association. — 1979. — V. 74. — PP. 427–431.
138. Dickey D.A., Fuller W.A. Likelihood ratio statistics for autoregressive time series with a unit root // Econometrica. — 1981. — V. 49. — № 4. — PP. 1057–1072.
139. Dickey D.A., Bell W.R., Miller R.B. Unit roots in time series models: Tests and applications // American statistician. — 1986. — V. 40. — PP. 12–26.
140. Enders W. Applied Econometric Time Series. — N.Y.: John Wiley & Sons, Inc. — 1995.

141. Engle R.F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation // Econometrica. — 1982. — V. 50. — PP. 987–1007.
142. Engle R.F., Granger C.W.J. Co-integration and error correction: representation, estimation and testing // Econometrica. — 1987. — V. 55. — PP. 251–276.
143. Granger C.W.J., Newbold P. Forecasting economic time series. 2nd ed. — N.Y.: Academic Press, 1986.
144. Green W. H. Econometric Analysis. — 3-d ed., International Edition.
145. Gujarati D.N. Basic Econometrics. — 2-nd ed. McGraw-Hill Book Company, International Edition, Singapore, 1988.
146. Harvey A.C. The econometric analysis of time series. — London: Philip Allan, 1981.
147. Klien P., Moore H. Monitoring grows cycles in market-oriented countries. — Cambridge, 1985.
148. Lukashin Y.P. An adaptive method of regression analysis // Statistical Analysis and Forecasting of Economic Structural Change, Peter Hackle, Ed. — IIASA, Springer-Verlag, 1989. — Ch.13. — PP. 209–216.
149. Lukashin Y.P. Analysis of data when constructing an adaptive regression model // Model-Oriented Data Analysis, Proceedings. V. Fedorov, H. Lauter, eds. — Eisenach, GDR/ Lecture Notes in Economics and Mathematical Systems. — Springer-Verlag, 1987. — PP. 189–194.
150. Lukashin Y.P. Econometric Analysis of Managers Judgement on the Determinants of the Financial Situation in Russia. — Research Memorandum № 98/8 ACE Project, University of Leicester, 1998.
151. Lukashin Y.P. Econometric Analysis of Managers' Judgements on the Determinants of the Financial Situation in Russia // Economics of Planning. Kluwer Academic Publishers. — 2000. — V.33. Special issue. — PP. 85–101.
152. Mintz J. Cyclical fluctuations in the exports of the US since 1879. — N.Y., 1967.
153. Nelson C.R., Plosser C.I. Trends and random walks in macroeconomic time series: some evidence and implications // Journal of monetary economics. — 1982. — V. 10. — PP. 139–162.
154. Nelson D.B. Conditional heteroscedasticity in asset returns: a new approach // Econometrica. — 1991. — V. 59. — PP. 347–370.

155. *Nelson C.R., Kang H.* Pitfalls in the use of time as explanatory variable in regression // *Journal of business and economic statistics*. — 1984. — V. 2. — PP. 73–82.
156. *Pesaran M. & Pesaran B.* Working with Microfit 4.0, Interactive Econometric Analysis. — Oxford: Oxford University Press, 1997.
157. *Raj B., Ullah A.* Econometrics: a varying coefficient approach. — L.: Croom Helm, 1981.
158. *Ramu Ramanathan.* Introductory Econometrics with Applications. — 4th ed., Harcourt, Inc. — 1998.
159. *Sargan J.D.* Wages and prices in the United Kingdom: A study in econometric methodology // *Econometrics analysis for national economic planning*. Eds. Hart P.E., Mills G., Whitaker J.K. — London: Butterworth, 1964.
160. *Spriet J.A., Vansteekiste G.C.* Computer-aided modeling and simulation/ L., N.Y., Paris: Academic Press, 1982.
161. *Von J. Griesse, Matt G.* Beschreibung des Verfahrens der adaptiven Einflussgrossenkombination. In *Prognoserechnung/ Herausgegeben von Peter Mertens*. — Wurzburg-Wien: Physica-Verlag, 1973. — PP. 160–166.
162. *Wickens M.R., Breusch T.S.* Dynamic specification, the long run and the estimation of transformed regression models // *Economic journal*. — 1988 (supplement). — PP. 189–205.

ОГЛАВЛЕНИЕ

Предисловие	3
Введение	6
Глава 1. ПРОСТЕЙШИЕ АДАПТИВНЫЕ МОДЕЛИ И ИХ СВОЙСТВА	14
§ 1. Временные ряды и стохастические процессы	14
§ 2. Экспоненциальное сглаживание	17
§ 3. Начальные условия экспоненциального сглаживания	22
§ 4. Выбор постоянной сглаживания	24
§ 5. Реакция модели на некоторые стандартные входные потоки данных	29
§ 6. Свойство оптимальности	34
§ 7. Модели линейного роста	35
§ 8. Стохастический процесс Тейла и Вейджа	37
§ 9. Примеры	41
Глава 2. РАЗВИТИЕ МОДЕЛЕЙ С ПОСТОЯННЫМИ ПАРАМЕТРАМИ АДАПТАЦИИ	47
§ 1. Адаптивная модель для изучения эволюционирующих законов распределения вероятностей	47
§ 2. Сезонные модели	50
§ 3. Аппроксимация полиномиальных трендов с помощью многократного сглаживания	62
§ 4. Обобщенная модель Брауна	76
§ 5. Примеры	88
Глава 3. АДАПТИВНАЯ МОДЕЛЬ ПРОГНОЗИРОВАНИЯ ВРЕМЕННОГО РЯДА, ГЕНЕРИРУЕМОГО АВТОРЕГРЕССИОННОЙ СХемой С ДРЕЙФУЮЩИМИ КОЭФФИЦИЕНТАМИ	90
§ 1. Общая схема адаптивного фильтра	90
§ 2. Адаптация коэффициентов модели авторегрессии ...	93
§ 3. Примеры	96

Глава 4. МОДЕЛИ С АДАПТИВНЫМИ ПАРАМЕТРАМИ АДАПТАЦИИ	98
§ 1. Скорость реакции как функция следящего контрольного сигнала (модель Тригга — Лича)	98
§ 2. Регулирование параметра адаптации по изменениям спектральных характеристик	105
§ 3. Адаптация параметра методом эволюции	114
Глава 5. АДАПТИВНЫЕ КОМБИНИРОВАННЫЕ МОДЕЛИ	121
§ 1. Адаптивная селективная модель	121
§ 2. Адаптивная гибридная модель	124
§ 3. Примеры	126
Глава 6. БАЙЕСОВСКИЙ ПОДХОД К КРАТКО-СРОЧНОМУ ПРОГНОЗИРОВАНИЮ	136
§ 1. Модель с множеством состояний	136
§ 2. Байесовский подход	140
§ 3. Реализация метода	147
§ 4. Сравнение методов	155
Глава 7. МОДЕЛИ АВТОРЕГРЕССИИ — СКОЛЬЗЯЩЕГО СРЕДНЕГО (метод Бокса — Дженкинса)	160
§ 1. Общее описание моделей и их свойств	160
§ 2. Идентификация моделей. Интерпретация R^2 в моделях АРСС	172
§ 3. Оценивание моделей и прогнозирование	182
§ 4. Прогнозирование после логарифмического преобразования	196
§ 5. Агрегирование рядов и моделей	198
§ 6. Примеры	200
Глава 8. МОДЕЛИРОВАНИЕ ВЗАИМОСВЯЗАННЫХ ВРЕМЕННЫХ РЯДОВ	209
§ 1. Анализ линейных динамических эконометрических моделей	209
§ 2. Адаптивная модель множественной регрессии	214
§ 3. Адаптивная модель производственной функции	218
Глава 9. НЕТРАДИЦИОННЫЙ КОРРЕЛЯЦИОННЫЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ	226
§ 1. Условия применимости традиционного корреляционного анализа	226
§ 2. Постановка проблемы	227
§ 3. Модифицированный коэффициент корреляции	229

§ 4. Адаптивный коэффициент корреляции	230
§ 5. Корреляционный анализ отклонений от заданных уровней	232
§ 6. Условный коэффициент корреляции	234
§ 7. Вероятностный коэффициент корреляции	235
§ 8. Пример	237

Глава 10. ФАЗОВЫЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ	240
§ 1. Постановка проблемы	240
§ 2. Разбиение временного ряда на фазы	241
§ 3. Фазовый анализ инвестиционных циклов в США и Западной Европе	247

Глава 11. АДАПТИВНАЯ ГИСТОГРАММА, ПРОБЛЕМА ОПТИМИЗАЦИИ	256
§ 1. Постановка проблемы	256
§ 2. Адаптивная процедура обновления гистограммы	257
§ 3. Проблемы оптимальности адаптивной процедуры обновления гистограммы	258
§ 4. Адаптивный анализ распределения кассовых остатков	264

Глава 12. КРИТЕРИИ ДИККИ — ФУЛЛЕРА ДЛЯ ИДЕНТИФИКАЦИИ ХАРАКТЕРА ТРЕНДА (обнаружение единичных корней)	268
§ 1. Детерминированный и случайный характер тренда	268
§ 2. Ложные тренды	270
§ 3. Критерий Дикки — Фуллера для обнаружения единичных корней	273
§ 4. Расширенный критерий Дикки — Фуллера	277
§ 5. Современные методы построения модели АРИСС	287

Глава 13. ИНТЕГРИРОВАННОСТЬ И КОИНТЕГРИРОВАННОСТЬ ПЕРЕМЕННЫХ	296
§ 1. Интегрированность и коинтегрированность двух переменных	296
§ 2. Коинтеграция многих переменных	299
§ 3. Коинтеграция и модели корректировки ошибок	304
§ 4. Критерии коинтеграции	308

Глава 14. РЕКУРРЕНТНЫЕ АЛГОРИТМЫ ОЦЕНКИ ТРАЕКТОРИЙ ПАРАМЕТРОВ МНОЖЕСТВЕННОЙ РЕГРЕССИИ	310
§ 1. Рекуррентное оценивание параметров регрессии	311
§ 2. Скользящая регрессия	315
§ 3. Взвешенная регрессия	316
§ 4. Рекуррентное оценивание параметров взвешенной регрессии	320
§ 5. Оценка траекторий параметров регрессии методом адаптивных ковариаций	324
§ 6. Связь адаптивной регрессии с адаптивным корреляционным анализом	328
§ 7. Модели с авторегрессионной условной гетероскедастичностью	330
Глава 15. КРАТКОСРОЧНОЕ ПРОГНОЗИРОВАНИЕ КУРСОВ ВАЛЮТ С ПОМОЩЬЮ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ	337
§ 1. Постановка проблемы	338
§ 2. Анализ случайности движения курсов валют	340
§ 3. Адаптивная модель прогнозирования временного ряда с неустойчивым характером колебаний	342
§ 4. Прогнозирование курсов валют	346
Глава 16. СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ВАЛЮТНЫХ АУКЦИОНОВ НА МОСКОВСКОЙ МЕЖБАНКОВСКОЙ ВАЛЮТНОЙ БИРЖЕ	352
§ 1. Цели исследования	352
§ 2. Теоретические модели валютных торгов	353
§ 3. Анализ исходных данных	360
§ 4. Модели торгов для растущего курса	362
§ 5. Модели торгов для падающего валютного курса ...	370
§ 6. Оценка качества статистических моделей	373
§ 7. Возможные способы использования моделей	376
Заключение	378
Приложения	381
Литература	400

Учебное издание

Лукашин Юрий Павлович

АДАПТИВНЫЕ МЕТОДЫ КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

Ответственный за выпуск *Л.А. Табакова*
Младший редактор *Н.А. Федорова*
Художественный редактор *Ю.И. Артюхов*
Технический редактор *Т.С. Маринина*
Корректор *Г.В. Хлопцева*
Компьютерная верстка *Е.Ф. Тимохиной*
Обложка художника *О.В. Толмачева*

ИБ № 4637

Подписано в печать 30.05.2003. Формат 60х88/16
Гарнитура «Таймс». Печать офсетная
Усл. п.л. 25,48. Уч. изд.-л. 21,03
Тираж 3000 экз. Заказ 1866. «С» 149

Издательство «Финансы и статистика»
101000, Москва, ул. Покровка, 7
Телефоны: (095) 925-47-08, 925-35-02, факс (095) 925-09-57
E-mail: mail@finstat.ru http://www.finstat.ru

ГУП «Великолукская городская типография»
Комитета по средствам массовой информации Псковской области,
182100, Великие Луки, ул. Полиграфистов, 78/12
Тел./факс: (811-53) 3-62-95
E-mail: VTL@MART.RU

Издательство
“ФИНАНСЫ И СТАТИСТИКА”
предлагает учебное пособие

Б.А. Лагоша

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ В ЭКОНОМИКЕ

192 с.: ил.



Исследуются методы и задачи оптимального управления в экономике, математический аппарат в форме теорем о достаточных условиях оптимальности. Из этих теорем выводятся вычислительные средства оптимального управления, доступные для практического применения. Теория сопровождается иллюстрациями экономической и другой природы.

Для обучающихся по специальности 061800 «Математические методы в экономике». Может использоваться в рамках специальности 351400 «Прикладная информатика (по областям)», а также всеми, кто интересуется применением математических методов в экономике.

**Книгу можно приобрести в киосках издательства
или заказать по почте**

Адрес: 101000, Москва, ул.Покровка, 7
(метро “Китай-город”, выход на ул.Маросейка)

Варшавка, 9

(метро “Тульская”, “Нагатинская”)

Тел.: (095)925-35-02, 923-18-68

Факс (095)925-09-57

E-mail: mail@finstat.ru <http://www.finstat.ru>