Министерство образования и науки Российской Федерации
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Т.В. АВДЕЕНКО

КОМПЬЮТЕРНЫЕ МЕТОДЫ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ И ПРОГНОЗИРОВАНИЯ

Утверждено
Редакционно-издательским советом университета
в качестве учебного пособия

НОВОСИБИРСК
2008

УДК 519.246.8:004(075.8)
А 187

Инновационная образовательная программа НГТУ «Высокие технологии»

Рецензенты:

В.В. Губарев, д-р техн. наук, проф.;

Г.С. Лбов, гл. научн. сотрудник Института математики СО РАН

Работа выполнена
на кафедре экономической информатики
для студентов 5-го курса факультета бизнеса
и факультета прикладной математики

Авдеенко Т.В.

А 187 Компьютерные методы анализа временны́х рядов и прогнозирования : учеб. пособие / Т.В. Авдеенко. – Новосибирск : Изд-во НГТУ, 2008. – 272 с.

ISBN 978-5-7782-1098-1

Учебное пособие посвящено построению статистических моделей временных рядов и прогнозированию на основе моделирования. В первой части рассматриваются модели декомпозиции и сглаживания. Во второй части дается развернутое изложение методологии ARIMA. В рамках этой методологии рассматриваются не только модели стационарных, нестационарных и сезонных временных рядов, но и модели рядов с интервенциями и выбросами. Изложение математической теории сопровождается изложением методологии анализа и рассмотрением конкретных примеров построения математических моделей по реальным данным в системе Statistica.

Материалы пособия апробированы при проведении занятий по дисциплинам «Математическое обеспечение систем обработки данных» и «Компьютерные методы статистического анализа и прогнозирования» в Новосибирском государственном техническом университете.

УДК 519.246.8:004(075.8)

ISBN 978-5-7782-1098-1

© Авдеенко Т.В., 2008
© Новосибирский государственный

технический университет, 2008

ПРЕДИСЛОВИЕ

Настоящее учебное пособие посвящено современным методам анализа временны́х рядов и прогнозирования. Отличительная особенность представленного материала – сочетание строгой математической теории и описания практического применения рассмотренных методов в популярной системе анализа данных *Statistica*. С помощью этих материалов можно не только научиться «нажимать кнопки», используя тот или иной вид анализа, ту или иную модель данных, но и осознать суть методологии через математические модели.

Учебное пособие методически разделено на две части. В первой части рассмотрены методы декомпозиции и сглаживания, наиболее часто применяемые на практике. Вторая часть посвящена полному изложению методологии ARIMA, причем в рамках этой методологии рассматриваются как стационарные, так и нестационарные модели, а также сезонные модели и модели с интервенциями и выбросами. Уделено внимание оцениванию параметров и прогнозированию. Столь серьезное освещение методологии ARIMA связано с тем, что она считается наиболее прогрессивным средством анализа временных рядов, однако почти не используется в отечественной практике из-за «сложности понимания». В пособии приводится развернутый анализ конкретных данных в системе *Statistica*, и это помогает понять методологию ARIMA, даже если детально не разбираться в математических формулах.

Таким образом, учебное пособие охватывает основной спектр современных методов анализа временных рядов, реализованных в конкретной универсальной системе анализа данных.

СОДЕРЖАНИЕ

Предисловие 3

Часть I. Анализ временных рядов с использованием классической методологии декомпозиции и сглаживания 4

1. Введение 4

1.1. Основные понятия и типы временных рядов 4

1.2. Обзор возможностей системы *Statistica* для анализа временных рядов 11

2. Классическая декомпозиция и сглаживание временных рядов 15

2.1. Основные компоненты классической декомпозиции временных рядов 15

2.2. Сглаживание данных 20

2.2.1. Сглаживание методом скользящего среднего (скользящей медианы) 21

2.2.2. Метод экспоненциального сглаживания 29

2.2.2.1. Простое экспоненциальное сглаживание 29

2.2.2.2. Двухпараметрическое экспоненциальное сглаживание 32

2.2.2.3. Экспоненциальное сглаживание сезонных данных 34

2.2.2.4. Оценка точности предсказания и качества модели 36

2.2.2.5. Экспоненциальное сглаживание в системе *Statistica* 38

2.2.2.6. Экспоненциальное сглаживание временного ряда авиаперевозок 45

2.3. Сезонная декомпозиция временных рядов 50

2.3.1. Классическая сезонная декомпозиция 50

2.3.2. Сезонная декомпозиция временного ряда применительно к исходным данным, характеризующим работу терминала в сети платежной системы 56

Часть II. Анализ временных рядов с использованием методологии ARIMA 65

3. Введение в методологию ARIMA 65

3.1. Теоретические основы анализа временных рядов с использованием методологии ARIMA 65

3.2. Автоковариационная и автокорреляционная функции 68

3.2.1. Автоковариационная функция 68

3.2.2. Частная автокорреляционная функция 70

3.3. Процесс белого шума 76

3.4. Оценивание математического ожидания, автоковариаций и автокорреляций 78

3.4.1. Выборочные среднее 78

3.4.2. Выборочная автоковариационная и автокорреляционная функции 79

3.4.3. Выборочная частная автокорреляционная функция 82

3.5. Теоретические представления процессов временных рядов 83

3.5.1. Представление процессов с помощью модели скользящего среднего 83

3.5.2. Представление процессов с помощью модели авторегрессии 86

4. Модели стационарных временных рядов 87

4.1. Модели стационарных временных рядов конечного порядка 87

4.2. Модели процессов авторегрессии 88

4.2.1. Модель AR(1) – авторегрессия первого порядка 89

4.2.1.1. Теоретическая автокорреляционная функция для процесса AR(1) 89

4.2.1.2. Частная автокорреляционная функция для процесса AR(1) 90

4.2.2. Модель AR(2) – авторегрессия второго порядка 91

4.2.2.1. Теоретическая автокорреляционная функция для процесса AR(2) 94

4.2.2.2. Частная автокорреляционная функция для процесса AR(2) 95

4.2.3. Общая модель AR(p) авторегрессии порядка p 96

4.3. Модели процессов скользящего среднего 99

4.3.1. Модель MA(1) – скользящее среднее первого порядка 99

4.3.1.1. АКФ процесса MA(1) 100

4.3.1.2. ЧАКФ процесса MA(1) 101

4.3.2. Модель MA(2) – скользящее среднее второго порядка 102
4.3.2.1. АКФ процесса MA(2) 103
4.3.2.2. ЧАКФ процесса MA(2).....104
4.3.3. Общая модель скользящего среднего MA(q) 104
4.4. Авторегрессионные процессы скользящего среднего ARMA(p, q).... 106
4.4.1. АКФ для процесса ARMA(p, q) 107
4.4.2. ЧАКФ для процесса ARMA(p, q).....108
4.4.3. Процесс ARMA(1, 1).....108
4.4.3.1. АКФ для процесса ARMA(1, 1).....109
4.4.3.2. ЧАКФ для процесса ARMA(1, 1) 110

5. Модели нестационарных временных рядов.....113

5.1. Нестационарность по отношению к среднему.....113
5.1.1. Модели с детерминированным трендом 114
5.1.2. Модели стохастического тренда и оператор взятия разности 115
5.1.3. Модель авторегрессии с проинтегрированным скользящим средним 116
5.1.3.1. Модель случайного блуждания 117
5.1.3.2. Модель ARIMA (0, 1, 1), или IMA(1, 1).....120
5.2. Нестационарность по отношению к дисперсии и автоковариации 121
5.2.1. Нестационарность по дисперсии и автоковариации процесса ARIMA.....121
5.2.2. Преобразования, стабилизирующие дисперсию 123

6. Идентификация модели ARIMA 127

6.1. Этапы идентификации модели ARIMA.....127
6.1.1. Этап 1. Построение графика ряда и использование преобразования, стабилизирующего дисперсию 127
6.1.2. Этап 2. Приведение однородно нестационарного ряда к стационарному с помощью операции взятия разности 129
6.1.3. Этап 3. Идентификация стационарного временного ряда....129
6.1.4. Этап 4. Проверка наличия детерминированного тренда 131

7. Расширение методологии ARIMA для анализа сезонных временных рядов 132

7.1. Традиционные методы учета сезонности временного ряда. Регрессионный метод.....134
7.2. Сезонные ARIMA-модели 136

7.3. Вычисление автокорреляционной функции для модели авиаперевозок 139
7.4. Общая сезонная ARIMA-модель 141

8. Оценивание параметров модели ARIMA 142

8.1. Метод моментов 143
8.2. Метод максимального правдоподобия 146
8.2.1. Метод условного максимального правдоподобия 146
8.2.2. Способы определения начальных условий для метода условного максимального правдоподобия. Метод обратного прогноза 147
8.2.3. Точные функции правдоподобия 151
8.2.4. Описание процедуры нелинейного оценивания 154
8.2.5. Свойства оценок параметров 158
8.3. Оценивание обыкновенным МНК в анализе временных рядов 160
8.4. Диагностика модели 162
8.5. Критерии выбора альтернативных моделей 165

9. Анализ временных рядов с использованием методологии ARIMA 167

9.1. Анализ временного ряда Вульфа ежегодного числа солнечных пятен, снятый с 1700 по 1983 год 167
9.2. Анализ временного ряда колебаний индекса Доу-Джонса в 1980-х годах 180
9.3. Анализ сезонного временного ряда 186

10. Анализ интервенций и определение выбросов 194

10.1. Анализ интервенций 195
10.1.1. Модель интервенций 195
10.1.2. Анализ моделей с интервенцией в системе *Statistica* 200
10.1.3. Примеры анализа интервенций 208
10.1.3.1. Анализ временного ряда реализации сладких вин в Австралии 208
10.1.3.2. Анализ временного ряда анализа инвестиций Гонконга 213
10.2. Выбросы временных рядов 223
10.2.1. Аддитивные и инновационные выбросы 224
10.2.2. Оценивание эффекта выброса при известном времени выброса 225
10.2.3. Определение выбросов с использованием итерационной процедуры 227

10.3. Замечания по использованию методологии анализа интервенций и определения выбросов 229

11. Прогнозирование 230

11.1. Прогнозирование для модели ARMA 231
11.2. Прогнозирование для модели ARIMA 234
11.3. Вычисление прогнозов 237
11.4. ARIMA-прогноз как взвешенное среднее наблюдений временного ряда 240
11.5. Обновление прогнозов 244
11.6. Эвентуальные функции прогноза 245
11.6.1. Общая формула вычисления эвентуальной функции прогноза 245
11.6.2. Примеры вычисления эвентуальной функции прогноза 247

Упражнения 249

Библиографический список 266

Часть I

АНАЛИЗ ВРЕМЕННЫХ РЯДОВ С ИСПОЛЬЗОВАНИЕМ КЛАССИЧЕСКОЙ МЕТОДОЛОГИИ ДЕКОМПОЗИЦИИ И СГЛАЖИВАНИЯ

1. ВВЕДЕНИЕ

1.1. ОСНОВНЫЕ ПОНЯТИЯ И ТИПЫ ВРЕМЕННЫХ РЯДОВ

Под временным (динамическим) рядом мы будем понимать упорядоченную последовательность наблюдений некоторого параметра (случайной величины). Хотя обычно упорядочивание проводится в последовательные моменты времени (через равные промежутки времени), вообще говоря, оно может быть и по другим измерениям (например, по пространственной переменной). С временными рядами мы сталкиваемся повсюду. В сельском хозяйстве мы наблюдаем годовой объем урожая и цены на сельхозпродукцию. В бизнесе и экономике ежедневно собираются наблюдения о курсе акций на конец дня и о курсе доллара (евро), еженедельно – процентные ставки, ежемесячно – индексы цен, ежеквартально – объемы продаж, ежегодно – годовой заработок. В технических областях снимают наблюдения за звуковыми, электрическими сигналами, напряжением, причем в этом случае используются более мелкие дискретности времени, чем в экономике – минуты и даже секунды. В геофизике регистрируется турбулентность, например, волны океана или шум земли в некоторой местности. В медицинской практике для диагностики заболеваний у пациентов снимается электроэнцефалограмма, электрокардиограмма. В метеорологии ежечасно наблюдают скорость ветра, ежедневно – температуру воздуха, ежегодно – количество осадков. В социологии ежегодно регистрируются рождаемость, смертность, количество несчастных случаев, уровень преступности. Известно успешное применение анализа временных рядов при слежении за подводными лодками противника в 70–80-е годы XX века.

Различают непрерывные и дискретные временные ряды. Временной ряд, который можно регистрировать непрерывно во времени, называется непрерывным. Это, например, электрические сигналы или напряжение. Временные ряды, которые регистрируется только через некоторые временные интервалы, называются дискретными. Это, например, временные ряды, встречающиеся в экономике. В дальнейшем мы будем рассматривать только дискретные временные ряды.

Как и в общей методологии статистического моделирования, выделим следующие три основные цели исследования временных рядов. Во-первых, с помощью такого исследования мы пытаемся понять, объяснить и описать механизм, генерирующий наблюдения интересующего нас параметра. Во-вторых, на основе построенной модели и имеющихся данных хотим заглянуть в будущее и спрогнозировать последующие наблюдения. И, наконец, целью всего исследования является оптимальное управление системой, или принятие оптимальных управленческих решений. Методология подобного статистического исследования динамических данных известна под названием *анализ временных рядов*.

Анализ временных рядов существенно отличается и с точки зрения лежащей в его основе философии, и с точки зрения используемых методов от регрессионного анализа и структурного моделирования. Основная идея структурных моделей заключается в том, что изменение значений результирующей (эндогенной) переменной объясняется за счет изменений значений одной или нескольких других, объясняющих (экзогенных), переменных. Анализ временных рядов базируется на иной идее: эндогенная переменная складывается под влиянием очень большого числа факторов, многие из которых не поддаются идентификации и непосредственному наблюдению. Поэтому лучший источник информации о совокупном влиянии всех факторов – значения самой исследуемой переменной в прошлые моменты времени, а также текущие и прошлые значения случайных ошибок.

Характерной особенностью временных рядов является то, что наблюдения в них зависимы и коррелированны, поэтому важен порядок наблюдений в выборке. Следовательно, для анализа временных рядов нельзя применять статистические процедуры и методы, опирающиеся на предположения независимости наблюдений в выборке, например, методы регрессионного анализа. Далее мы опишем статистическую

методологию анализа временных рядов, которая доступна и профессиональным программистам, и обычным пользователям программного обеспечения, так как реализована в большинстве пакетов статистического анализа данных, например, в системе *Statistica*.

Наиболее распространенные методы анализа временных рядов можно разделить на временные (автокорреляционная и частная автокорреляционная функция, авторегрессия, модель ARIMA, экспоненциальное сглаживание) и частотные (спектральный анализ на основе спектральных функций). Для временного подхода временной ряд рассматривается как непосредственная функция времени. В случае частотного подхода временной ряд представляется как сумма периодических рядов с разными частотами. Между методами обоих классов существует взаимно однозначное соответствие. Однако бывают ситуации, когда один из подходов оказывается предпочтительнее.

Для того чтобы иметь представление, какие задачи можно решать с помощью методов, описанных в данном учебном пособии, рассмотрим примеры конкретных временных рядов, взятых из различных хозяйственных сфер. Все они проявляют основную особенность, общую для всех временных рядов, – зависимость между наблюдениями, но выявляют и другие различающие их особенности.

Временной ряд 1. График ежедневных данных о среднем числе дефектов на грузовик в конце сборочного конвейера на предприятии по производству грузовиков изображен на рис. 1.1. Как видно из графика, наблюдения осциллируют на некотором постоянном уровне. Временной ряд такого типа называется стационарным в среднем и является специальным случаем стационарных временных рядов. Такой ряд может быть описан авторегрессионной моделью скользящего среднего (ARMA), предложенной в методологии Бокса–Дженкинса.

Временной ряд 2. Данные о производстве (ежегодном) табака в США (рис. 1.2) не варьируются около постоянного значения, а, наоборот, выявляют предельный, вверх направленный тренд. Кроме того, дисперсия этого временного ряда увеличивается с увеличением времени. Временной ряд, выявляющий такую закономерность, называется нестационарным по среднему и по дисперсии и служит примером нестационарного временного ряда. Такое сочетание двух типов нестационарности нередко встречается на практике.

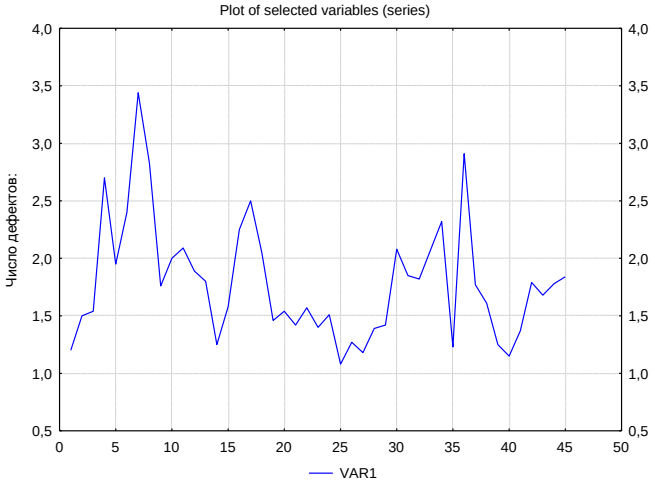


Рис. 1.1. Среднее число дефектов на грузовик в конце сборочного конвейера

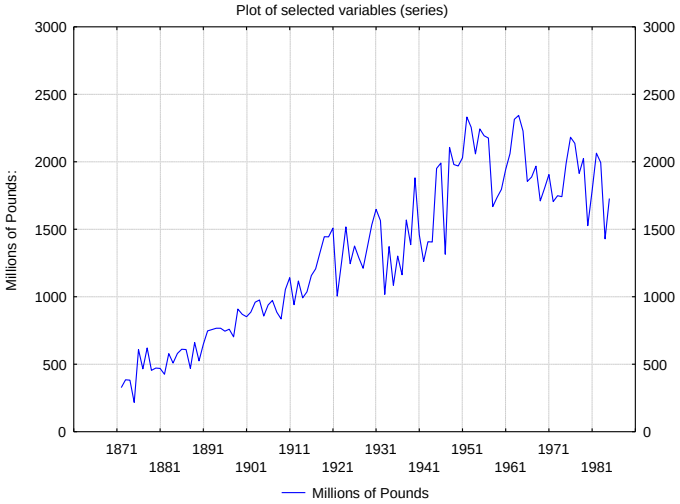


Рис. 1.2. Ежегодное производство табака в США с 1871 по 1981 год

Однако, так как осцилляции на различных уровнях выявляют одинаковую тенденцию, такая нестационарность в среднем называется однородной и легко может быть учтена с помощью оператора взятия разности. Можно также использовать классический метод декомпозиции с выделением тренда. Для учета нестационарности дисперсии здесь можно провести преобразование, стабилизирующее дисперсию (например, логарифмирование).

Временной ряд 3. Ежеквартальные данные о производстве пива в США в течение нескольких лет составляют сезонный временной ряд, проявляющий ежегодную тенденцию к повторению (график ряда изображен на рис. 1.3). Период сезонности, т.е. интервал, через который тенденция повторяется, в данном случае равен 4. Для анализа данного ряда может быть предложена модификация модели Бокса–Дженкинса. Альтернативным способом моделирования является сезонная декомпозиция.

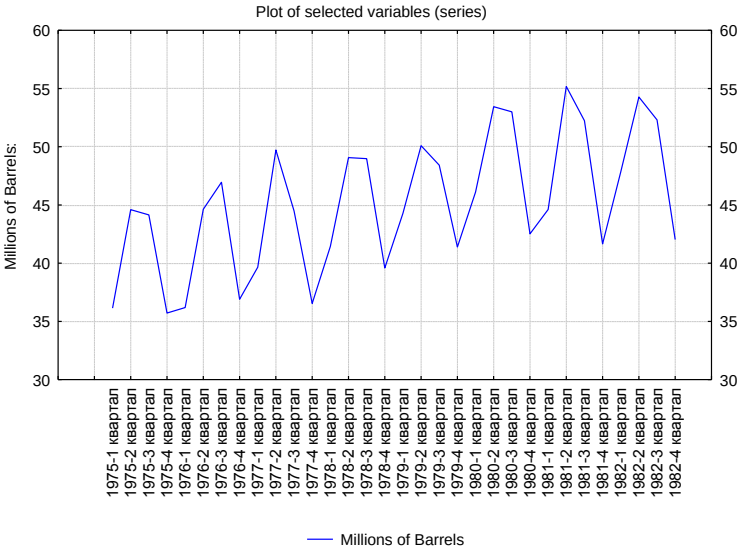


Рис. 1.3. Ежеквартальное производство пива в США

Временной ряд 4. На рис. 1.4 изображен график реализации вина сладкого сорта на территории Австралии с января 1980 по июнь 1994 года. Ряд отражает другое явление нестационарности – изменение в

структуре ряда, возникшее из-за некоторого внешнего события. Этот тип нестационарности нельзя учесть, применяя то или иное стандартное преобразование. Дополнительный пример подобных временных рядов – производственный ряд, нарушенный забастовками. Или ряд с записывающего устройства, полученный при сбое устройства. Такие внешние возмущения называются *интервенциями*, или *выбросами*. Строго говоря, интервенции и выбросы – несколько разные понятия, о чем подробнее будет сказано в разделе 10, посвященном анализу интервенций.

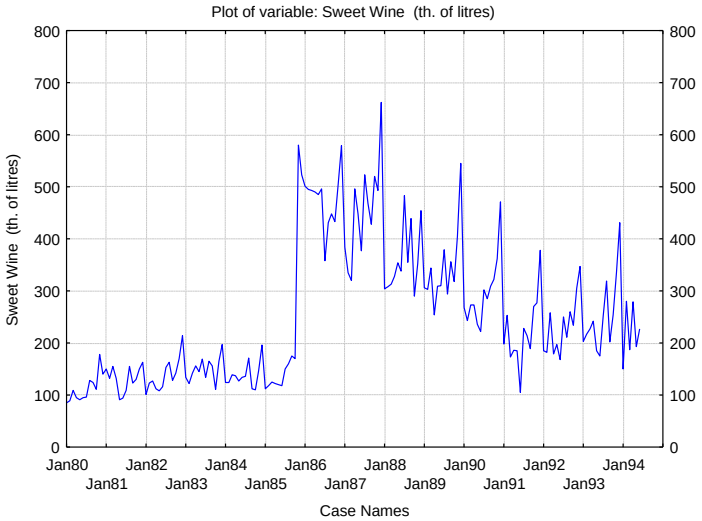


Рис. 1.4. Реализация вина сладкого сорта на территории Австралии

Рассмотренные примеры относятся к так называемым одномерным временным рядам. Однако данные могут включать одновременные наблюдения за несколькими переменными. В бизнесе, например, часто пытаются увеличить объем продаж посредством рекламы. В этих условиях в целях изучения воздействия рекламы и для улучшения прогноза об объеме продаж естественно построить динамическую модель, связывающую текущие значения объема продаж с текущими и прошлыми значениями объема рекламы, т.е. модель с одной зависимой выходной переменной и несколькими независимыми входными переменными.

Однако хотя объем средств, затраченных на рекламу, влияет на объем продаж, последний в свою очередь тоже влияет на рекламу (так как бюджет на рекламу часто определяется как процент от объема продаж). В этом случае используются более сложные векторные авторегрессионные модели скользящего среднего. В пакет *Statistica* векторный метод не входит. В нем присутствуют более простые способы исследования двух временных рядов (например, модель распределенных запаздываний).

Временной ряд 5. Временной ряд Лидии Пинкхам изображен на рис. 1.5.

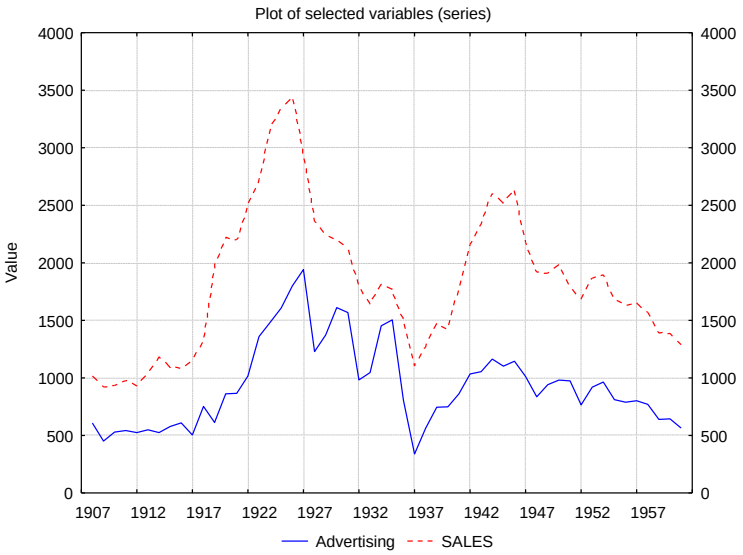


Рис. 1.5. Двумерный временной ряд Лидии Пинкхам

Здесь представлены две переменные: одна – объем средств, затраченных косметической фирмой на рекламу, вторая – объем продаж этой фирмы в те же временные интервалы. На рисунке, изображающем графики этих временных рядов, отчетлива корреляция между ними. Причем не только затраты на рекламу влияют на объем продаж, но и объем продаж в свою очередь влияет на объем средств, затраченных фирмой на рекламу.

1.2. ОБЗОР ВОЗМОЖНОСТЕЙ СИСТЕМЫ *STATISTICA* ДЛЯ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

Для работы с временными рядами в системе *Statistica* используется модуль *Time Series/Forecasting* (Анализ временных рядов/Прогнозирование) в подменю *Advanced Linear/Nonlinear Models* главного меню *Statistics* (рис. 1.6). После выбора данного пункта появляется стартовая панель модуля, изображенная на рис. 1.7.

Кнопка ***OK* (*transformations, autocorrelations, crosscorrelation, plots*)** (Да, преобразования, корреляции, кросскорреляции, графики), расположенная в верхней правой части панели, предоставляет множество предварительных исследований и преобразований исходного временного ряда, необходимых для того или иного вида анализа. Назначение предоставляемых этой кнопкой возможностей мы будем приводить при описании реализации того или иного вида анализа в системе *Statistica*.

В нижней функциональной части панели находятся функциональные кнопки, открывающие различные виды анализа временных рядов, которые будут рассмотрены нами в данном учебном пособии. Кратко опишем назначение каждой из кнопок.

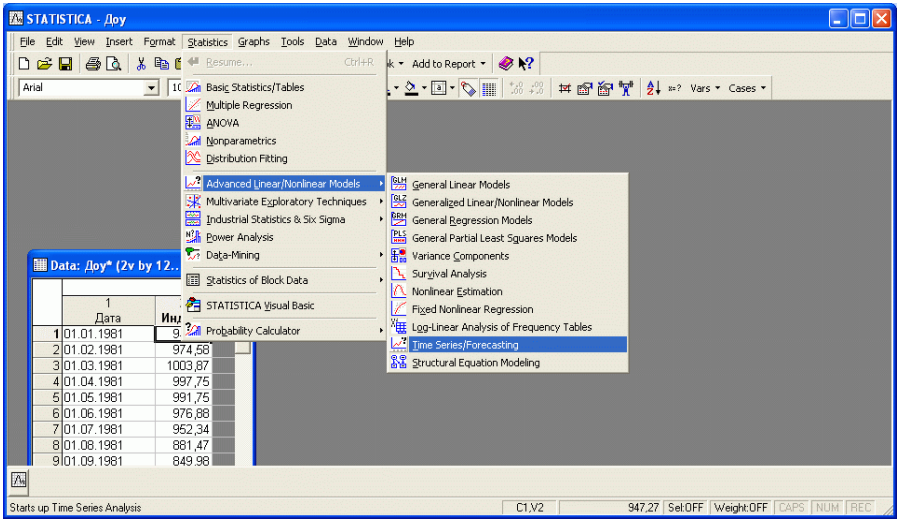


Рис. 1.6. Выбор модуля *Time Series/Forecasting*

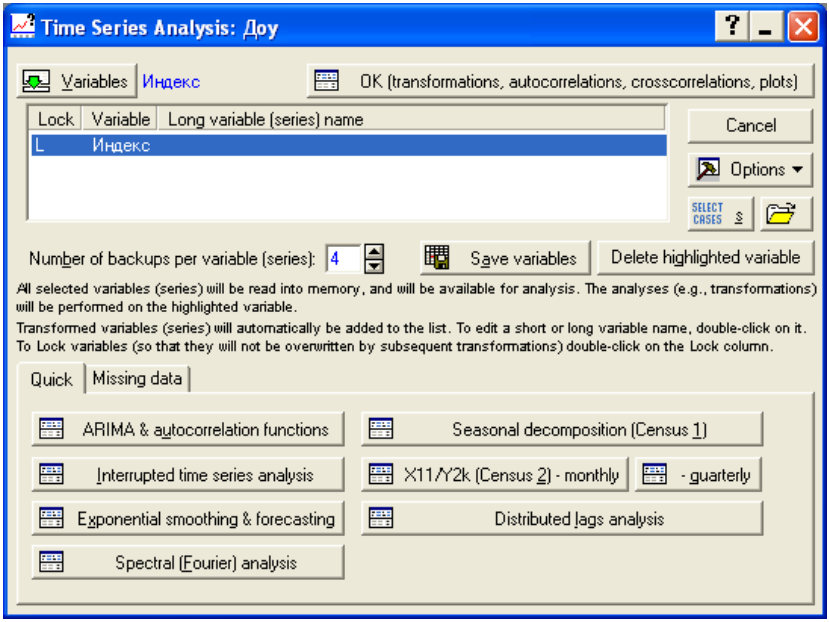


Рис. 1.7. Стартовая панель модуля *Time Series/Forecasting*

Кнопка ***ARIMA&autocorrelation functions*** (ARIMA&автокорреляционные функции) открывает работу с важным классом параметрических моделей ARIMA. В системе *Statistica* диалог ARIMA реализован в классической методологии Бокса–Дженкинса [3]. Многие ряды в экономике, маркетинге (как стационарные, так и нестационарные, и даже сезонные), могут быть успешно описаны моделью ARIMA. Эта модель дает наиболее оптимальные прогнозы, однако она редко используется в отечественной практике по двум, на наш взгляд, причинам. Во-первых, временные ряды в российской экономике не отличаются большой длиной, поэтому точная модель ARIMA не всегда может быть построена. Во-вторых, методология ARIMA-моделирования считается наиболее сложной для освоения. Однако с системой *Statistica* использование этой весьма полезной методологии сильно упрощается. Далее в учебном пособии описываются стационарные, нестационарные, сезонные модели ARIMA, их идентификация, оценивание параметров и прогнозирование на основе построенной модели. Приводятся примеры анализа конкретных временных рядов в системе *Statistica*.

Кнопка ***Interrupted time series analysis*** **(**Анализ прерванного временного ряда**)** позволяет расширить класс моделей ARIMA для анализа временных рядов с интервенцией. Необходимость такого рода анализа возникает, например, когда с некоторого момента времени по объективным причинам изменяются условия реализации наблюдаемого процесса, что вызывает изменения уровня временного ряда. Как уже отмечалось, внешнее вмешательство, повлекшее такие изменения временного ряда, называется интервенцией. Например, изменяются котировки акций из-за изменения правил торгов на бирже или каких-то политических событий, изменяется биение сердца больного после приема лекарства и т.д.

Внешнее воздействие на ряд, вызываемое разными причинами, может быть как кратким, импульсивным, так и длительным, устойчивым. В момент воздействия траектория временного ряда резко меняется, но далее вновь описывается той же моделью, что и до интервенции. Специальный раздел учебного пособия посвящен работе с моделями интервенций. Мы даем понятие переменной интервенции, отклика на интервенцию, приводим общую модель ARIMA с интервенцией как расширение общей модели ARIMA, описываем те возможности методологии, которые предоставляются системой *Statistica*, и, наконец, приводим анализ конкретных временных рядов с использованием этой полезной методологии.

Кнопка ***Exponential smoothing & forecasting*** **(**Экспоненциальное сглаживание и прогнозирование**)** открывает возможность использовать различные процедуры экспоненциального сглаживания и прогнозирования на их основе. Экспоненциальное сглаживание позволяет сгладить наблюдаемый ряд, выделить из него неслучайную составляющую – тренд, на основе его спрогнозировать будущие значения. В системе *Statistica* имеется возможность использования однопараметрического сглаживания (без тренда), двухпараметрического сглаживания (линейный и экспоненциальный тренды), а также трехпараметрического экспоненциального сглаживания – для сезонных временных рядов. В последнем случае можно выбрать мультипликативную или аддитивную модель. Предоставляется возможность использования модели сглаживания с затухающим трендом. В учебном пособии описываются математические модели одно-, двух- и трехпараметрического сглаживания и прогнозирование на их основе. Рассматриваются критерии выбора параметров сглаживания. Описываются примеры использования методов сглаживания в системе *Statistica*.

Три кнопки ***Seasonal decomposition (Census1), X11/Y2k (Census2) – monthly и X11/Y2k (Census2) – quarterly*** (Сезонная декомпозиция 1 и 2, метод Census1 и Census2, в том числе месячная и квартальная декомпозиции) позволяют проводить классическую аддитивную и мультипликативную декомпозиции сезонного временного ряда с использованием методов Census 1 и Census 2. В учебном пособии мы приводим аддитивную и мультипликативную модели, подчеркиваем различия между ними, а также описываем особенности предлагаемых системой методов выделения основных компонент модели. Рассматриваются примеры использования данного инструментария для анализа сезонных временных рядов.

Кнопка ***Distributed lag analysis*** (Анализ распределенных запаздываний) позволяет строить модель распределенных запаздываний (лагов) для двух переменных, значения каждой из которых являются временными рядами. Одна из переменных – зависимая – является функцией другой переменной – независимой, причем не только текущие значения, но и предшествующие значения независимой переменной предполагаются влияющими на зависимую переменную. В модели распределенных лагов значения зависимого ряда являются регрессией на взвешенные текущие и предшествующие значения независимого ряда**.**

Кнопка ***Spectral (Fourier) analysis*** (Спектральный (Фурье) анализ) позволяет провести полный спектральный анализ временных рядов, построить периодограмму оценки спектральной плотности с разнообразными, в том числе определяемыми пользователем, спектральными окнами. Эти методы (помимо того, что они важны сами по себе) могут быть рассмотрены в связи с другими методами обработки временных рядов. Например, они могут применяться при ARIMA-моделировании для проверки адекватности подогнанных моделей. В этом случае используется спектральный анализ остатков. Методы спектрального анализа применимы при выявлении скрытых периодичностей в данных, при определении длины периода или для спектрального анализа рядов, построенных из остатков подогнанной модели.

2. КЛАССИЧЕСКАЯ ДЕКОМПОЗИЦИЯ И СГЛАЖИВАНИЕ ВРЕМЕННЫХ РЯДОВ

2.1. ОСНОВНЫЕ КОМПОНЕНТЫ КЛАССИЧЕСКОЙ ДЕКОМПОЗИЦИИ ВРЕМЕННЫХ РЯДОВ

Одним из основных положений, на котором базируется анализ временных рядов, является то, что факторы, влияющие на отклик изучаемой системы, действовали некоторым образом в прошлом и настоящем, и ожидается, что они будут действовать аналогичным образом и в будущем. Поэтому основная цель анализа временных рядов – вычленение этих факторов для прогноза дальнейшего поведения исследуемой системы и выработки рациональных управленческих решений.

Рассмотрим временной ряд ежемесячных данных занятости пассажирских мест на международных полетах за 12 лет. Данный ряд был описан в книге [3], а файл *Series_G.sta*, содержащий данные, находится в стандартной папке *Datasets* системы *Statistica*. Чтобы начать работать с этим файлом, необходимо открыть его с помощью меню *File/Open*. График ряда можно построить, нажав кнопку ***OK(transformations, autocorrelations, crosscorrelation, plots)*** стартовой панели модуля *Time Series/Forecasting*. После нажатия этой кнопки в появившемся окне преобразований ***Transformations of Variables*** необходимо выбрать закладку ***Review&plot***, после чего появляется окно, изображенное на рис. 2.1. Для построения графика ряда необходимо выделить соответствующий файл, после чего нажать кнопку ***Plot*** справа от кнопки ***Review highlighted variable***. Предварительно в данном случае лучше выбрать ***Case names*** в качестве разметки по оси абсцисс, так как даты снятия наблюдений присутствуют в именах строк, а чтобы даты на рисунке не сливались (так как их очень много), выбрать ***Scale X axis in plots manually*** , шаг Step=6.

Полученный график данных авиаперевозок изображен на рис. 2.2. Из графика видно, что, во-первых, выявляется общая долговременная тенденция к возрастанию, так называемый тренд, причем тренд имеет вид степенной функции. Во-вторых, обнаруживается тенденция к повторению формы колебаний для каждого года, учитывающей, что большинство путешествий совершается в летние месяцы и небольшое увеличение наблюдается к концу декабря. Цель метода сезонной декомпозиции заключается в том, чтобы изолировать эти компоненты, т.е. разложить ряд на несколько составляющих с тем, чтобы по отдельности анализировать их эффект.

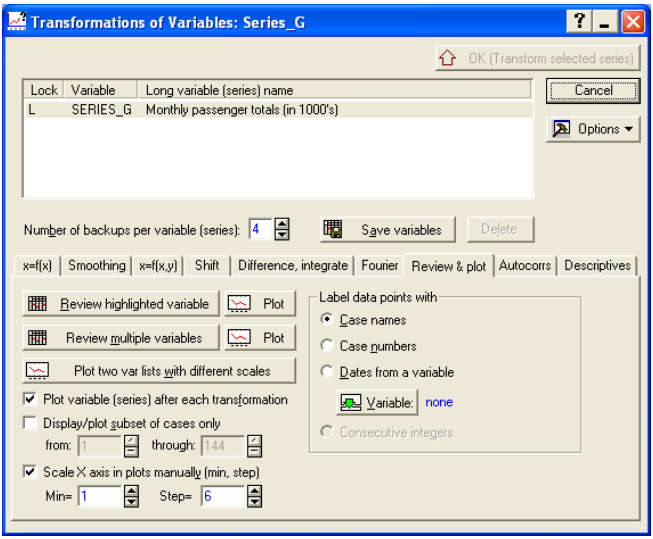


Рис. 2.1. Закладка ***Review&plot*** построения графиков в окне преобразований

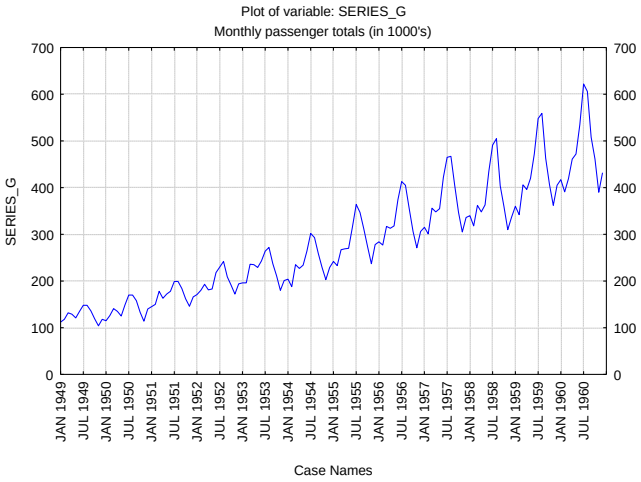


Рис. 2.2. График временного ряда данных о количестве авиаперевозок

За время развития теории временных рядов предлагалось много методов вычленения указанных факторов. Но наиболее удачной оказалась классическая мультипликативная модель временного ряда, чаще всего используемая в экономических исследованиях. Реже используется классическая аддитивная модель временного ряда. В рамках обеих моделей выделяют четыре типа факторов, под влиянием которых формируются значения временного ряда: долговременные, сезонные, циклические и случайные.

Долговременные факторы – формируют общую тенденцию изменений анализируемого показателя в длительной перспективе. Функция, с помощью которой описывается эта тенденция, называется *трендом*. Тренд определяется возрастанием или убыванием «сглаженных» значений временного ряда в течение продолжительного периода времени, составляющего обычно для экономических приложений 15–20 лет. Наличие этого компонента в поведении временного ряда обусловлено действием фундаментальных экономических или демографических факторов, таких как технический прогресс, изменение уровня жизни, изменение численности населения и т.д. Тренд может быть выделен или в результате визуального изучения ряда, или, если визуально он плохо определяется, в результате применения процедур сглаживания.

Сезонные факторы – формируют периодически повторяющиеся в определенное время года колебания анализируемого показателя. Эти факторы обусловлены главным образом погодой и обычаями. Как правило, для исследования таких сезонных изменений используется ряд ежемесячных или ежеквартальных наблюдений. Например, каждый год количество мороженого, продаваемого в летние месяцы, больше, чем в зимние месяцы. Однако сезонность не обязательно связана с годовыми изменениями. Например, при исследовании ежедневного объема корреспонденции, обрабатываемого почтовыми службами, могут наблюдаться недельные периодические изменения, когда в середине недели объем отправлений выше, чем в начале и/или в конце. Кроме того, для такого же временного ряда можно обнаружить определенные сезонные изменения в течение года, если для этого достаточно количество данных. Таким образом, один и тот же временной ряд может показать несколько сезонных составляющих с разными периодами сезонности.

Циклические (конъюнктурные) факторы – формируют изменения анализируемого показателя под влиянием действия долговременных циклов экономической, демографической и иной природы. Цик-

лические колебания представляют собой повторяющиеся движения значений ряда вверх и вниз относительно тренда с продолжительностью цикла от 2 до 15 лет. Эти циклы не являются строго периодическими. Амплитуда и продолжительность их зависят от содержания измеряемого показателя. Они отличаются в индустриальном и сельскохозяйственном рядах. Например, циклы в производстве товаров длительного пользования обычно обнаруживают более значительные относительные флуктуации, чем в производстве потребительских товаров, а конкретный временной ряд для этого направления может обладать бизнес-циклами, имеющими значительные вариации как по периодам, так и по амплитудам. Циклические компоненты не могут быть учтены и спрогнозированы в модели, если отсутствует статистика за длительный период времени. Это связано с тем, что нет вообще простых объяснений цикличности деловой активности. Поэтому в классическом анализе временных рядов применяют так называемый остаточный подход к расчету циклов: после удаления тренда и сезонного компонента остаток объясняется совокупным действием циклического и случайного факторов.

Случайные (нерегулярные) факторы – обусловливают стохастическую (случайную) природу элементов временного ряда, не поддаются учету и регистрации. Такие колебания называются иногда остаточной вариацией, так как они остаются во временном ряде после выделения тренда, сезонных и циклических составляющих. Эти изменения вызываются большим числом не поддающихся учету факторов, совместное воздействие которых хаотично.

Соотношения перечисленных компонентов в модели могут принимать разнообразные формы. В классической модели декомпозиции приняты аддитивная и мультипликативная формы зависимости. Мультипликативная форма имеет вид

$$Z_t = T_t C_t S_t I_t, \quad t = 1, \ldots, n, \tag{2.1}$$

где n – объем выборки, T_t – трендовый компонент, C_t – циклический компонент, S_t – сезонный компонент, I_t – нерегулярный компонент.

Аддитивная форма соответственно имеет вид

$$Z_t = T_t + C_t + S_t + I_t. \tag{2.2}$$

Мультипликативная форма модели более распространена по сравнению с аддитивной. Выбор между мультипликативной и аддитивной

моделями основывается на следующих соображениях. Сезонные колебания (например, увеличение объема продаж в декабре) могут быть аддитивными, если объем продаж увеличивается ежегодно в декабре в среднем, скажем, на 1 млн р. Сезонные колебания являются мультипликативными, если объем продаж увеличивается ежегодно в декабре в среднем на 30 %, т.е. в 1,3 раза. На графиках в аддитивном случае ряд показывает равномерные сезонные колебания, не зависящие от уровня ряда, а в мультипликативном случае амплитуда сезонных колебаний варьируется в зависимости от уровня ряда. Аналогичные рассуждения справедливы при выборе формы и для остальных факторов модели. Например, для ряда, изображенного на рис. 2.2, более подходит мультипликативная модель. Однако если провести предварительное логарифмирование ряда, то можно использовать и аддитивную модель. Логарифмирование ряда проводится в уже знакомом нам окне преобразований ***Transformations of Variables***, в котором надо в закладке ***x = f(x)*** выбрать преобразование ***Natural log***, как на рис. 2.3.

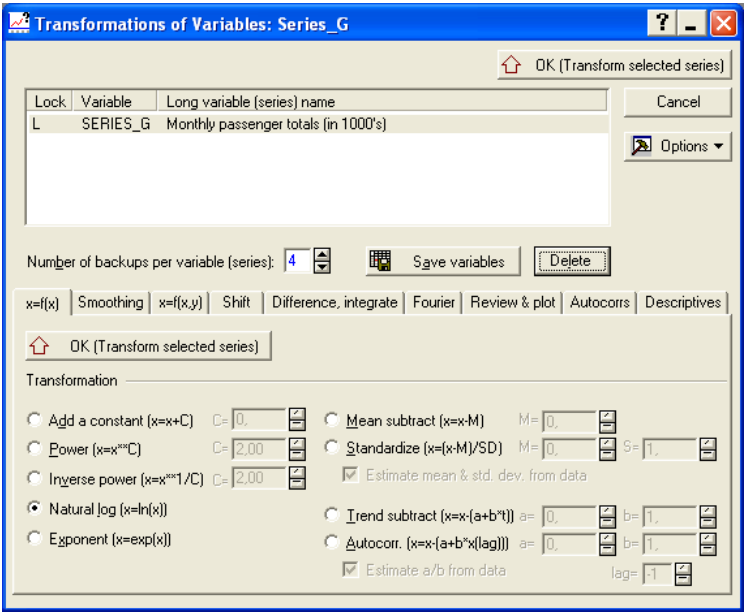


Рис. 2.3. Выбор логарифмирования в закладке ***x = f(x)*** окна преобразований

Полученный в результате логарифмирования ряд изображен на рис. 2.4. Из графика видно, что теперь ежегодные (сезонные) колебания постоянны, а тренд имеет тенденцию линейного возрастания, т.е. возрастания с постоянным приращением. Поэтому для преобразованного ряда можно использовать аддитивную модель декомпозиции.

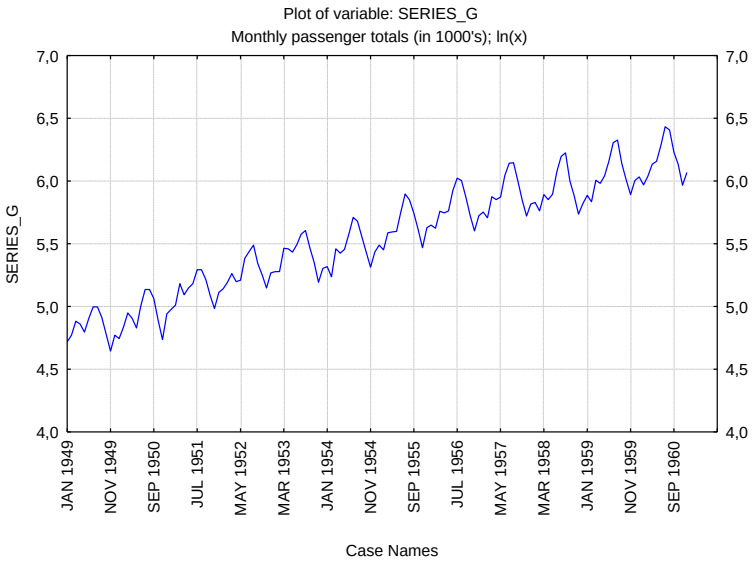


Рис. 2.4. График прологарифмированного ряда о количестве авиаперевозок

2.2. СГЛАЖИВАНИЕ ДАННЫХ

Одна из целей анализа временных рядов заключается в обнаружении каких-либо тенденций или закономерностей среди данных. Для ряда авиаперевозок долговременная тенденция – возрастающий тренд – довольно отчетливо видна на графиках 2.2, 2.4. Такого рода тренд может быть обусловлен, например, ростом численности населения, или ростом благосостояния, или научно-техническим прогрессом. В ряде случаев теория не может указать явный вид тренда как функции времени, тем не менее оказывается возможным приблизить тренд полиномом от времени достаточно низкой степени (например, линейной функцией).

Однако довольно часто тренд невозможно представить какой-либо простой функцией времени на всем рассматриваемом временном интервале. Задача усложняется, если изменчивость данных скрывает тенденцию. Для компенсации этой проблемы можно «сгладить» временной ряд за счет усреднения последовательных значений данных. Сглаженные данные менее восприимчивы к случайным флуктуациям исходного набора данных, т.е. при сглаживании в какой-то степени нивелируется влияние случайной составляющей отклика. Поэтому сглаженный временной ряд фактически оказывается суперпозицией тренда, циклического и, возможно, сезонного (это зависит от метода сглаживания) компонентов процесса. Обычно используются методы скользящего среднего или экспоненциального сглаживания, которые и будут рассмотрены в дальнейших подразделах.

2.2.1. СГЛАЖИВАНИЕ МЕТОДОМ СКОЛЬЗЯЩЕГО СРЕДНЕГО (СКОЛЬЗЯЩЕЙ МЕДИАНЫ)

Один из наиболее простых способов сглаживания данных заключается в вычислении скользящих средних. Скользящее среднее (скользящая медиана) – это среднее арифметическое значение (медиана) для N соседних наблюдений (так называемого N-точечного окна скользящего среднего).

Наиболее распространенный способ использования скользящего среднего заключается в вычислении сглаженного значения для точки, расположенной в середине окна скользящего среднего. Если N – нечетное, то формула для вычисления t-го сглаженного значения выглядит следующим образом:

$$Z_{MA(t)} = \frac{1}{N} \sum_{i=(1-N)/2}^{(N-1)/2} Z_{t+i} \,. \qquad (2.3)$$

В этом случае чаще всего применяются 3-точечные, 5-точечные и 7-точечные схемы сглаживания, когда $N = 3$, $N = 5$ и $N = 7$. Чем больше N, тем более сглаженный ряд мы получаем.

Используется также вариант сглаживания, когда значения ряда сглаживаются средним арифметическим предыдущих N наблюдений. В этом случае формула сглаживания выглядит следующим образом:

$$Z_{MA_prior(t)} = \frac{1}{N} \sum_{i=1}^{N} Z_{t-i} \,. \qquad (2.4)$$

В формулах (2.3) и (2.4) все наблюдения внутри интервала имеют одинаковый вес. Однако существует взвешенный вариант скользящего среднего, определяемый формулами

$$Z_{MA(t)} = \frac{1}{N} \sum_{i=(1-N)/2}^{(N-1)/2} \alpha_i Z_{t+i} \qquad (2.5)$$

и

$$Z_{MA_prior(t)} = \frac{1}{N} \sum_{i=1}^{N} \alpha_i Z_{t-i} \,. \qquad (2.6)$$

Отметим, что выбор весов α_i в формулах (2.5) и (2.6) крайне субъективен. В некоторых источниках рекомендуется, чтобы соблюдалось условие нормировки

$$\sum_{i=1}^{N} \alpha_i = 1 \,.$$

Центрированная формула (2.5) может быть полезной при сглаживании сезонного ряда для последующего вычисления сезонных индексов. При этом рекомендуется выбирать N равным длине периода сезонности (если данные ежемесячные, то период сезонности равен 12) плюс 1 (центральное значение).

Рассмотрим возможности для сглаживания скользящим средним, которые предоставляет система *Statistica*. В качестве примера данных, на которых производится сглаживание, рассмотрим временной ряд данных о ежемесячных средних значениях индекса Доу-Джонса в период с 1981 по 1990 год. Сначала построим и проанализируем график временного ряда, представленный на рис. 2.5. Ряд очевидно имеет тренд, т.е. видна устойчивая тенденция к возрастанию индекса Доу-Джонса. Однако одновременно с увеличением индекса возрастает и его изменчивость, количественно выражающаяся в повышении уровня стандартного отклонения.

Для владельцев акций важным фактором при принятии решения о продаже или покупке ценных бумаг выступает не сам индекс, а изменения его среднемесячного значения. Для того чтобы проанализировать помесячные изменения, получим разности исходного ряда. Операция взятия разностей – одна из самых распространенных операций в

анализе временных рядов. Предположим, что мы имеем исходный ряд Z_t, $t = 1,\ldots,n$. В результате применения операции взятия разностей получаем ряд

$$W_t = Z_t - Z_{t-1}, \qquad t = 2,\ldots,n.$$

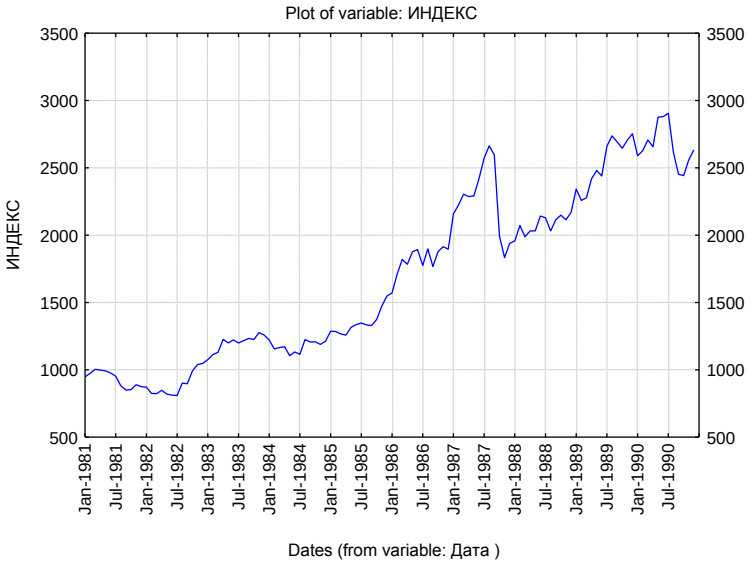


Рис. 2.5. График временного ряда колебаний индекса Доу-Джонса

В системе *Statistica* эта операция доступна в окне преобразований ***Transformations***, которое мы уже рассматривали. После выбора закладки ***Difference, integrate*** в полученном окне надо выбрать опцию ***Differencing… lag=1***, как показано на рис. 2.6.

Последующее нажатие кнопки ***OK Transform selected series*** преобразует выделенный ряд. Ряд, полученный взятием разностей, изображен на рис. 2.7. На графике прослеживается увеличение изменчивости ежемесячных значений индекса Доу-Джонса, а также отчетливо виден выброс, связанный с внезапным падением индекса в октябре 1987 года. Неравномерность изменчивости на исследуемом временном интервале легко устраняется предварительным логарифмированием, как это было показано для данных авиаперевозок в подразделе 2.1.

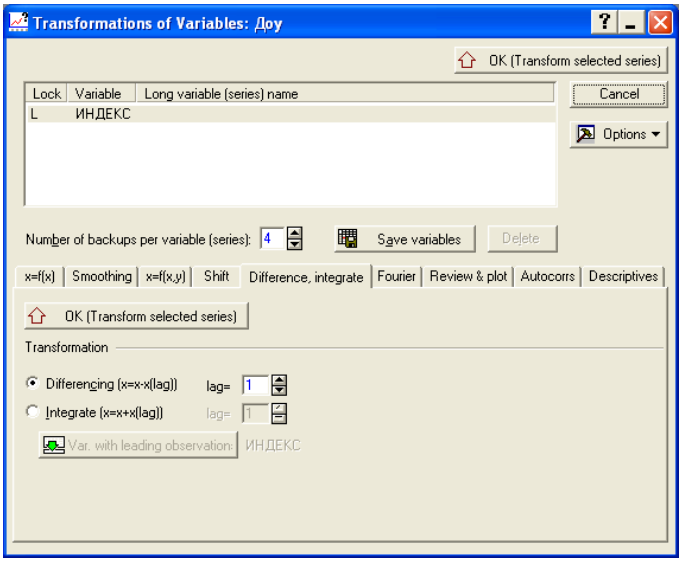


Рис. 2.6. Выполнение операции взятия разности

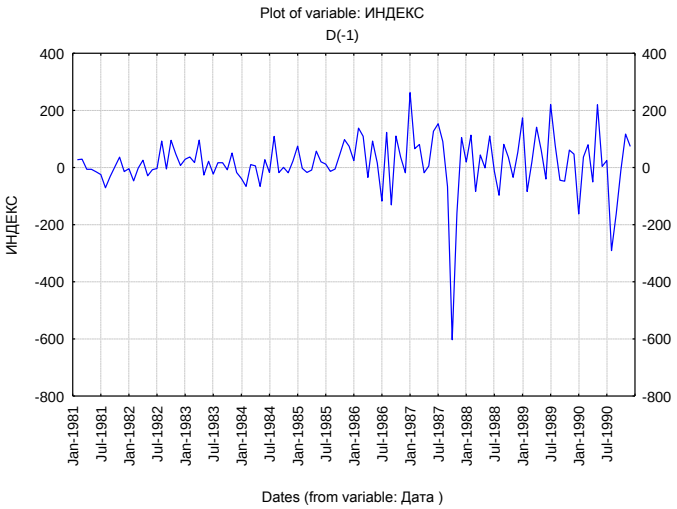


Рис. 2.7. График разностей временного ряда колебаний индекса Доу-Джонса

Ряд, полученный двумя последовательными преобразованиями – логарифмированием и последующим взятием разности, изображен на рис. 2.8. Здесь мы, во-первых, добились равномерной изменчивости; во-вторых, видим, что ряд осциллирует, т.е. случайный компонент, вызывающий колебания, скрывает основную тенденцию, которая необходима для принятия решения о покупке акций. Для выявления тренд-циклического компонента, позволяющего судить о ежемесячных изменениях индекса, сгладим временной ряд.

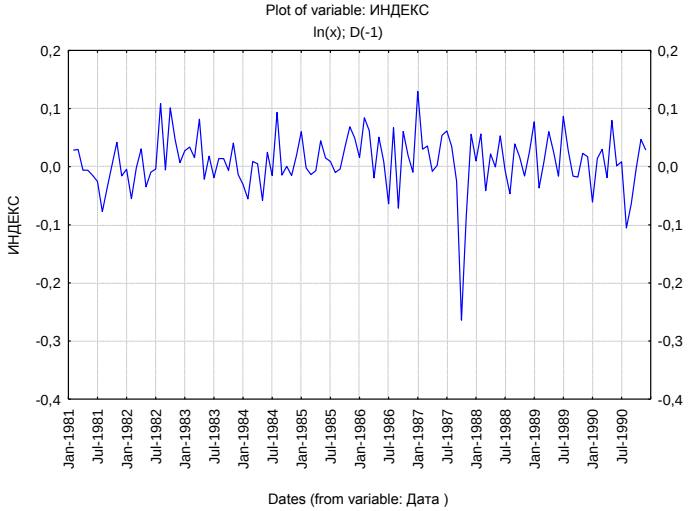


Рис. 2.8. Индекс Доу-Джонса после логарифмирования и взятия разности

Простейшие операции сглаживания становятся доступными в известном нам окне преобразований ***Transformations***. В этом окне необходимо выбрать закладку ***Smoothing***, в которой предоставляется выбор одного из четырех методов сглаживания, как показано на рис. 2.9.

Первый метод – метод сглаживания N-точечным скользящим средним. Здесь можно либо использовать центрированную формулу (2.3), либо выбрать опцию ***Prior*** – вычисление скользящего среднего по предыдущим значениям по формуле (2.4). При желании можно воспользоваться взвешенным методом вычисления по формулам (2.5) и (2.6), задав свой вес каждой точке с помощью опции ***Weighted… Specified weights***.

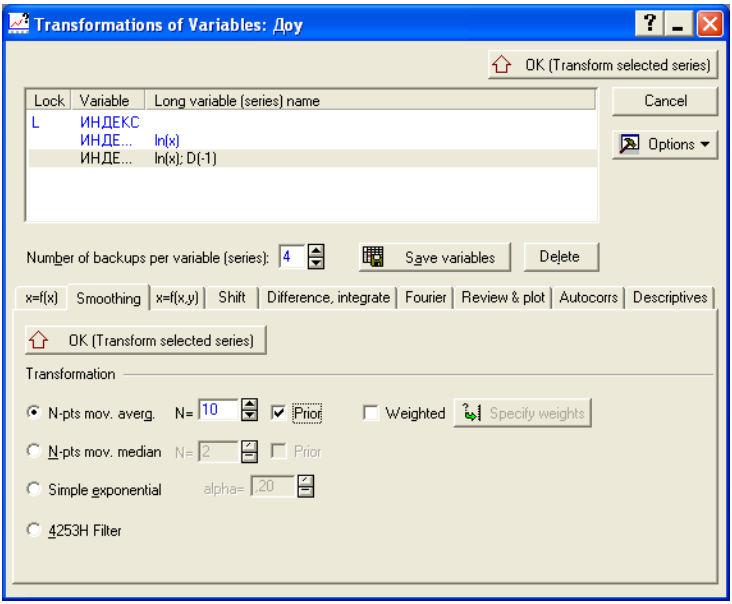


Рис. 2.9. Выбор операции сглаживания

Второй метод – метод сглаживания N-точечной скользящей медианой. В каждой точке сглаженного ряда вычисляется медиана N соседних точек. Здесь также может быть центрированный вариант или вариант, когда все точки предшествуют точке, в которой вычисляется сглаженное значение (вариант ***Prior***). Этот метод менее чувствителен к выбросам, чем метод скользящего среднего.

Третий метод – простейшая (однопараметрическая) версия метода экспоненциального сглаживания, задаваемая параметром α (метод экспоненциального сглаживания будет подробно описан в следующем подразделе).

Четвертый метод – 4253Н Фильтр – мощное средство сглаживания, состоящее из последовательной серии сглаживаний скользящим средним и скользящей медианой. На практике этот метод часто делает более заметными важные характеристики ряда.

Выберем сглаживание скользящим средним по формуле (2.4) с $N = 10$, как показано на рис. 2.9, и выполним сглаживание. В результате получаем сглаженный ряд, изображенный на рис. 2.10. Значение

$N = 10$ выбрано потому, что инвесторы часто используют стратегию покупки акций, если текущая цена превышает среднее для 10 предыдущих моментов времени. А если текущая цена ниже среднего для 10 предыдущих моментов времени, то инвесторы используют стратегию продажи акций.

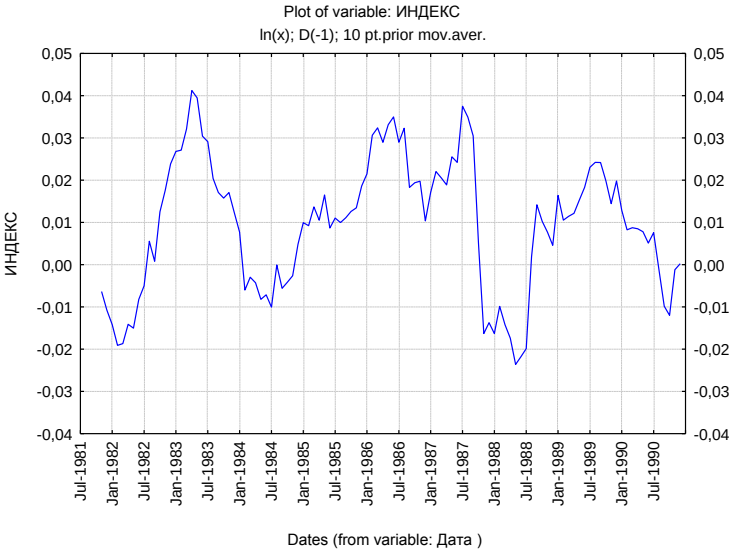


Рис. 2.10. Ряд, сглаженный скользящим средним по 10 предыдущим точкам

Анализ сглаженного ряда показывает, что скользящее среднее с периодом 10 мес было выше нуля. На графике (рис. 2.10) можно заметить три периода устойчивого роста, превышающего среднюю скорость роста: с июля 1982 года по январь 1984, с января 1985 по сентябрь 1987, и с июля 1988 по июль 1990 года. Единственный период устойчивого спада наблюдался с октября 1987 года по октябрь 1988 года, что может быть объяснено выбросом в октябре 1987 года. Если применить метод скользящей медианы, менее чувствительный к выбросам, то этого периода на сглаженном ряде не наблюдается, что видно на рис. 2.11. В целом по приведенному анализу следует сделать вывод о благоприятной ситуации на рынке акций.

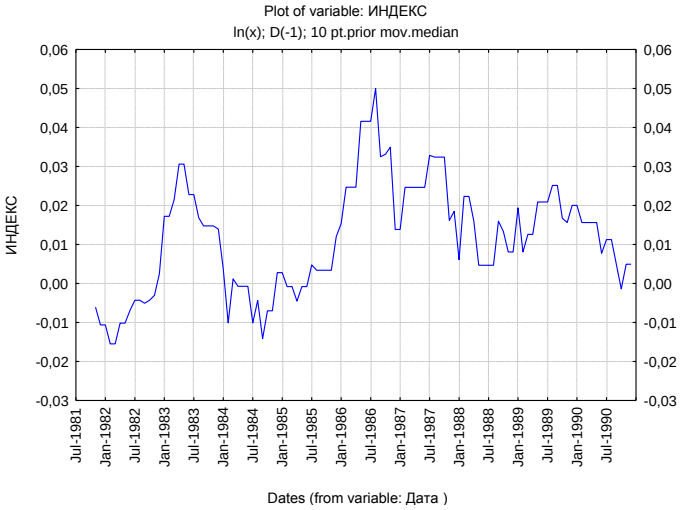


Рис. 2.11. Ряд, сглаженный скользящей медианой по 10 предыдущим точкам

Метод скользящего среднего может быть использован и для сглаживания данных, которым свойственна сезонность, т.е. повторение тенденции через определенный одинаковый интервал времени, называемый периодом сезонности. При этом рекомендуется использование симметричного оператора скользящего среднего [14]

$$\tilde{Z}_t = \sum_{i=-m}^{m} \alpha_i Z_{t-i}\,, \tag{2.7}$$

где m – положительное целое; α_i – константы, такие что $\alpha_i = \alpha_{-i}$, $\sum_{i=-m}^{m} \alpha_i = 1$.

Метод основан на предположении, что ежегодная сумма сезонного временного ряда обладает маленькими сезонными вариациями. При этом целесообразно выбирать m равным половине длины периода сезонности. Оценка сезонного компонента получается вычитанием сглаженного ряда $\tilde{Z}_t$ из исходного ряда, т.е.

$$\hat{S}_t = Z_t - \tilde{Z}_t\,.$$

Вышеприведенные оценки (2.7) могут быть получены итерационно повторным использованием операторов скользящего среднего. Удачным примером использования этой процедуры служит метод Census X-11, широко распространенный в правительственных исследованиях и в промышленности США. Ряд $Z_t - \tilde{Z}_t$, полученный после удаления сезонных колебаний, называется сезонно скорректированным, а вышеприведенная процедура декомпозиции называется сезонной корректировкой.

2.2.2. МЕТОД ЭКСПОНЕНЦИАЛЬНОГО СГЛАЖИВАНИЯ

2.2.2.1. ПРОСТОЕ ЭКСПОНЕНЦИАЛЬНОЕ СГЛАЖИВАНИЕ

Скользящее среднее предполагает равную значимость для всех наблюдений периода. Например, при анализе ежемесячных индексов Доу-Джонса величина индекса 10 мес назад имеет тот же вес, что и величина в текущем месяце. Однако специалисты часто предпочитают использовать взвешенный вариант скользящих средних, когда более свежие наблюдения имеют больший вес, а более давние наблюдения – меньший вес. Причем для более старых наблюдений используются веса, которые уменьшаются по экспоненциальному закону. В этом случае самое свежее наблюдение имеет вес α, следующее за ним наиболее свежее наблюдение имеет вес $\alpha(1-\alpha)$, следующее – вес $\alpha(1-\alpha)^2$ и т.д. В этом случае для вычисления скользящего среднего используются все предыдущие значения, а не только ограниченный набор значений. При этом формула для экспоненциально взвешенного скользящего среднего имеет вид

$$Z_{MA_prior(t)} = \alpha Z_{t-1} + \alpha(1-\alpha)Z_{t-2} + \alpha(1-\alpha)^2 Z_{t-3} + \ldots, \quad (2.8)$$

где вес α называется параметром сглаживания, $0 < \alpha < 1$. Такой способ вычисления называется экспоненциальным сглаживанием.

Экспоненциальное сглаживание выгодно отличается от сглаживания скользящим средним тем, что оно может использоваться для предсказания значения следующего наблюдения Z_{n+1} на основе текущего наблюдения Z_n и предыдущих наблюдений Z_{n-1}, $Z_{n-2}, \ldots, Z_1$. Формула для предсказания аналогична формуле (2.8):

$$\hat{Z}_{n+1} = S_n = \alpha Z_n + \alpha(1-\alpha)Z_{n-1} + \alpha(1-\alpha)^2 Z_{n-2} +$$

$$+\ldots\alpha(1-\alpha)^{n-1}Z_1 + (1-\alpha)^n S_0 . \qquad (2.9)$$

Формулу (2.9) можно представить в виде рекуррентного соотношения

$$S_n = \alpha Z_n + (1-\alpha)S_{n-1} . \qquad (2.10)$$

Таким образом, предсказанное значение S_n равно сумме взвешенных значений текущего наблюдения и предыдущего предсказания. Для того чтобы начать вычисления по формуле (2.10), необходимо задать начальное значение S_0. Можно использовать в качестве начального значения первое наблюдение Z_1 временного ряда или среднее арифметическое значение ряда. После задания S_0 расчет выполняется в такой последовательности:

$$S_1 = \alpha Z_1 + (1-\alpha)S_0,$$

$$S_2 = \alpha Z_2 + (1-\alpha)S_1,$$

$$\vdots$$

$$S_n = \alpha Z_n + (1-\alpha)S_{n-1}.$$

Хотя в принципе α может принимать любые значения из диапазона $0 < \alpha < 1$, обычно ограничиваются интервалом $0,2 < \alpha < 0,5$. При высоких значениях α в большей степени учитываются мгновенные текущие наблюдения отклика, и наоборот, при низких его значениях сглаженная величина определяется в большей степени прошлой тенденцией развития, чем текущим состоянием отклика системы. В табл. 2.1 подсчитаны веса для пяти ближайших точек при трех значениях параметра сглаживания $\alpha = 0,01$; 0,5 и 0,8. Заметим, что для $\alpha = 0,01$ все пять предыдущих наблюдений получают примерно одинаковые веса. В этом случае говорят, что сглаженное среднее имеет более длинную память. Если же параметр равен $\alpha = 0,7$, то наибольший вес получает наблюдение, ближайшее к текущему, остальные наблюдения получают очень низкие значения весов. Уже для $\alpha = 0,3$

веса точек довольно быстро убывают по мере удаления от текущего значения.

Таблица 2.1

Расчет весов экспоненциального сглаживания

α	$\alpha(1-\alpha)$	$\alpha(1-\alpha)^2$	$\alpha(1-\alpha)^3$	$\alpha(1-\alpha)^4$
0,01	0,0099	0,0098	0,0097	0,0096
0,3	0,21	0,147	0,1029	0,0720
0,7	0,21	0,063	0,0189	0,0057

Аналитики большинства фирм при обработке рядов используют свои традиционные значения параметра сглаживания α. Для динамично развивающихся фирм и рынков характерны значения α выше, чем для более консервативных компаний и стабильных рынков. Для прогнозов рекомендуется использовать более высокие величины этого параметра, чем для анализа предшествующих тенденций. Тем не менее выбор параметра сглаживания остается очень субъективной процедурой, зависящей от опыта и интуиции исследователя.

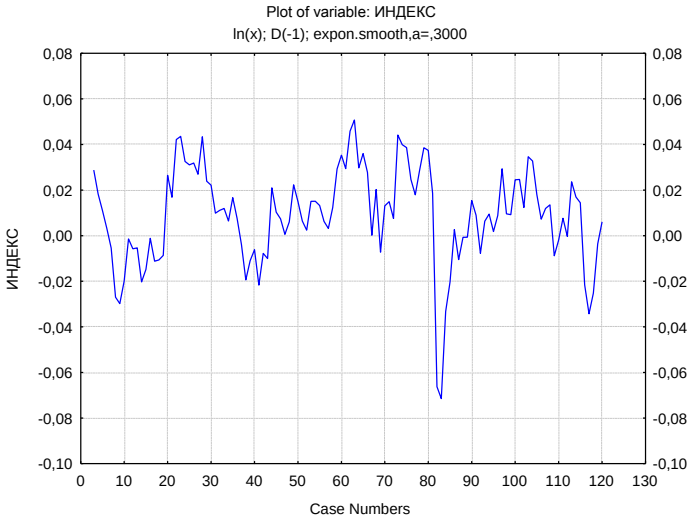


Рис. 2.12. Ряд, преобразованный методом экспоненциального сглаживания ($\alpha = 0,3$)

В системе *Statistica* простое экспоненциальное сглаживание доступно в окне преобразований по закладке ***Smoothing***, как и сглаживание скользящим средним. Произведем экспоненциальное сглаживание ряда индекса Доу-Джонса после логарифмирования и последующего взятия разностей. Выбрав данный метод сглаживания ***Simple exponential***, необходимо задать значение параметра α и нажать ***OK (Transform selected series)***. Сглаженный ряд при выбранном значении параметра сглаживания $\alpha = 0{,}3$ показан на рис. 2.12. Заметим, что выводы по данному графику те же, что мы делали после использования метода скользящего среднего.

Данный простейший вид экспоненциального сглаживания доступен в системе *Statistica* также и в окне ***Seasonal and Non-Seasonal Exponential Smoothing*** по кнопке ***Exponential smoothing&forecasting*** вместе с более сложными разновидностями, описанными дальше.

2.2.2.2. ДВУХПАРАМЕТРИЧЕСКОЕ ЭКСПОНЕНЦИАЛЬНОЕ СГЛАЖИВАНИЕ

Метод экспоненциального сглаживания, описанный выше, называется однопараметрическим экспоненциальным сглаживанием, потому что для сглаживания данных используется только один параметр. Однопараметрическое экспоненциальное сглаживание обеспечивает наглядное представление о тренде и позволяет делать краткосрочные прогнозы, а при попытке распространить прогноз на больший период создается впечатление, что развитие процесса в сторону возрастания или убывания совершенно прекратилось – на любой период будущего прогнозируются одни и те же значения отклика.

Если временной ряд демонстрирует долговременную тенденцию, например, постепенное увеличение уровня ряда со временем, то в таком случае эффективнее использовать двухпараметрическое экспоненциальное сглаживание, так называемый метод Хольта. Этот метод успешно справляется и со среднесрочными, и с долгосрочными прогнозами, поскольку он способен обнаруживать микротренды (тренды, относящиеся к коротким периодам) в моменты времени, непосредственно предшествующие прогнозным, и экстраполировать эти тренды на

будущее. И хотя в рамках метода Хольта возможна только линейная экстраполяция в будущее, в большинстве реальных ситуаций ее оказывается достаточно.

Формулы двухпараметрического сглаживания подобны формулам однопараметрического сглаживания, однако в первом случае на каждом шаге вычисляется путем сглаживания не только параметр расположения, но и параметр тренда. При использовании метода Хольта необходимо последовательно вычислять сглаженные значения ряда S_n и значение тренда (наклона) T_n, накопленное в любой точке ряда:

$$S_n = \alpha Z_n + (1-\alpha)(S_{n-1} + T_{n-1}), \tag{2.11}$$

$$T_n = \gamma(S_n - S_{n-1}) + (1-\gamma)T_{n-1}, \tag{2.12}$$

где α и γ – константы сглаживания, $0 < \alpha < 1$, $0 < \gamma < 1$.

Выбор констант сглаживания, как и в однопараметрическом методе, очень субъективен. Чаще всего они выбираются из диапазона от 0,25 до 0,5, причем значения констант не обязательно совпадают. Обычно рекомендуется начинать со значений $\alpha = \gamma = 0{,}3$, а затем варьировать их по мере необходимости. Для того чтобы начать вычисления по формулам (2.11) и (2.12), необходимо задать начальные значения S_0 и T_0. Один из методов вычисления начальных значений – построение линии регрессии по выборке. Пересечение регрессии с осью ординат выбирается в качестве значения S_0, а наклон уравнения регрессии – в качестве значения T_0.

Уравнения (2.11) и (2.12) имеют простую интерпретацию. Параметр расположения S_n равен сумме взвешенных значений текущего наблюдения и предыдущих оценок параметров расположения и тренда. А значение T_n равно взвешенной сумме предыдущей оценки параметра тренда и разности между текущей и предыдущей оценками параметра расположения.

Эта модель подходит, например, для прогнозирования количества запасных деталей к станкам. Потребность в специальных деталях может медленно увеличиваться или уменьшаться с течением времени (тренд), и тренд может медленно изменяться по мере изнашивания станков или их устаревания.

При расчете прогноза по методу Хольта предполагается, что сглаженное значение в последней точке является опорным, а определенный для нее линейный микротренд сохранит свое значение и в будущем. Функция прогноза на l шагов вперед вычисляется по формуле

$$\hat{Z}_{n+l} = S_n + lT_n,$$

где $\hat{Z}_{n+l}$ – прогноз значения Z_{n+l}. Данный прогноз считается более или менее достоверным на 4–5 периодов, если тренд достаточно близок к линейному.

2.2.2.3. ЭКСПОНЕНЦИАЛЬНОЕ СГЛАЖИВАНИЕ СЕЗОННЫХ ДАННЫХ

Если экономические показатели измеряются один раз в год, то в качестве шага по времени используется именно этот отрезок. В таком случае применимы однопараметрическое и двухпараметрическое сглаживание, о чем шла речь в предыдущих подразделах. Если же данные получены с меньшими интервалами, ежеквартально или ежемесячно, то величина отклика может периодически закономерно меняться, выявляя тенденцию увеличения или уменьшения в определенные месяцы года, как происходит, например, в строительстве или туристическом бизнесе. В этом случае существенным оказывается сезонный компонент отклика и в процедуру сглаживания вносятся соответствующие изменения, которые мы рассмотрим в этом подразделе. Если же влияние сезонных изменений мало или же вовсе отсутствует, то применяются описанные ранее более простые виды сглаживания.

Экспоненциальное сглаживание сезонных данных с использованием трех констант сглаживания называется трехпараметрическим экспоненциальным сглаживанием, или методом Винтера. Константы сглаживания в методе Винтера относятся к размещению, тренду и сезонности. Метод Винтера можно использовать для мультипликативной и аддитивной сезонности. Разница заключается в том, что в случае мультипликативной модели производится сезонная корректировка, когда компонент расположения умножается на сезонный компонент (сезонный индекс) при прогнозировании, а для аддитивной модели сезонный компонент прибавляется к компоненту расположения. Ниже приведены формулы трехпараметрического экспоненциального сглаживания для мультипликативной сезонности. Первые две формулы аналогичны формулам однопараметрического и двухпараметрического экспоненциального сглаживания:

$$S_n = \alpha \frac{Z_n}{I_{n-p}} + (1-\alpha)(S_{n-1} + T_{n-1}), \tag{2.13}$$

$$T_n = \gamma(S_n - S_{n-1}) + (1-\gamma)T_{n-1}. \tag{2.14}$$

Разница заключается в корректировке текущего наблюдения Z_n с помощью деления на сезонный индекс I_{n-p}, взятый в предыдущий момент времени $n-p$; p – период сезонности.

В данном методе к формулам (2.13) и (2.14) добавляется третья формула для сглаживания сезонных индексов:

$$I_n = \delta \frac{Z_n}{S_n} + (1-\delta)I_{n-p}, \tag{2.15}$$

где δ – константа сглаживания, $0 < \delta < 1$.

Согласно (2.15), значение сглаженного сезонного индекса является взвешенным средним значением текущего сезонного индекса и сезонного индекса из предыдущего периода времени. Для вычисления сглаженных значений по формулам (2.13) – (2.15) необходимо задание начальных значений S_0, T_0, а также набор, состоящий из p начальных индексов сезонности I_0, I_{-1}, … , I_{1-p}. Начальные значения сезонного компонента могут быть вычислены, например, как в методе сезонной декомпозиции. Для аддитивной модели сезонный индекс для конкретного месяца или квартала вычисляется как среднее значение по всем таким же месяцам или кварталам для других лет. Для мультипликативной модели такие индексы вычисляются как средние значения по конкретному месяцу или кварталу всех лет после удаления минимального и максимального значений.

Прогноз будущих значений для данного метода вычисляется в случае мультипликативной сезонности по формуле

$$\hat{Z}_{n+l} = S_{n+l} I_{n+l-p},$$

а в случае аддитивной сезонности – по формуле

$$\hat{Z}_{n+l} = S_{n+l} + I_{n+l-p}.$$

2.2.2.4. ОЦЕНКА ТОЧНОСТИ ПРЕДСКАЗАНИЯ И КАЧЕСТВА МОДЕЛИ

Таким образом, мы рассмотрели основные классические модели экспоненциального сглаживания. Как было отмечено ранее, все модели отличаются большой субъективностью в выборе констант сглаживания, а также начальных значений параметров расположения, тренда и сезонных индексов. При выборе различных значений получаем по сути разные модели сглаживания и соответственно разное качество сглаживания. Возникает вопрос, существуют ли более или менее объективные критерии, позволяющие судить о том, насколько хороша та или иная модель сглаживания.

Так как одна из важных функций экспоненциального сглаживания – прогнозирование по временному ряду, то лучшей моделью естественно считать ту, которая обеспечивает наибольшую точность прогноза. Использовать этот критерий в чистом виде проблематично, поскольку проверить качество прогноза можно будет только в будущем, сравнив предсказанное значение с реальностью. Однако не без основания считается, что модель, наиболее точно описывающая существующие данные, будет также хорошо прогнозировать. Таким образом, одним из критериев качества модели может быть некоторая мера расхождения реальных данных и рассчитанных по модели значений.

Рассмотрим три таких критерия качества. Первый из них – среднеквадратическая ошибка (*mean squared error – MSE*), основан на вычислении суммы квадратов разностей между предсказываемым и наблюдаемым значениями:

$$\mathrm{MSE} = \frac{\sum_{t=1}^{n}(Z_t - \hat{Z}_t)^2}{n}. \qquad (2.16)$$

Критерий MSE чаще всего используется для оценки точности прогноза в большинстве статистических процедур подгонки. Квадратный корень из MSE называется стандартной ошибкой (*standard error*), которую можно считать типичной ошибкой предсказания.

Еще один способ оценки качества прогноза – вычисление абсолютных значений разностей между наблюдаемым и предсказываемым значениями, критерий *MAE* (*mean absolute error*):

$$\mathrm{MAE} = \frac{\sum_{t=1}^{n} \left| Z_t - \hat{Z}_t \right|}{n}. \qquad (2.17)$$

Критерий MAE применяется в том случае, когда хотят сгладить влияние выбросов, к которым очень чувствителен критерий MSE (из-за возведения в квадрат).

Третий критерий – средняя абсолютная процентная ошибка (*Mean absolute percentage error – MAPE*), выражает точность в виде процентной доли наблюдаемого значения:

$$\mathrm{MAPE} = \frac{\sum_{t=1}^{n} \left| (Z_t - \hat{Z}_t) / Z_t \right|}{n} \cdot 100. \qquad (2.18)$$

Два предыдущих критерия (2.16) и (2.17) выражали фактическое значение ошибки. Однако может оказаться более обоснованным (особенно если наблюдения в большой мере осциллируют с течением времени) вычислять отклонение одношаговых прогнозов от наблюдаемых значений относительно величин наблюдений. Кроме того, при прогнозировании ежемесячных продаж, возможно, более удобно рассуждать о качестве предсказания, например, с ±10%-й точностью, а не со среднеквадратичной ошибкой, равной, например, 406,88.

Критерии (2.16) – (2.18) могут быть использованы для оценки качества предсказания модели, а также при выборе констант сглаживания. Однако минимизация критерия точности – не единственный фактор, который нужно учитывать при выборе той или иной модели. Вообще, желательно, чтобы ошибка предсказания была низкой. Но, например, значение константы однопараметрического сглаживания α, которое дает самую низкую ошибку, может оказаться очень высоким, например, 0,9. Если задача сглаживания состоит в том, чтобы упростить поведение временного ряда и отобразить общие тенденции, то предсказание с такой высокой константой сглаживания не позволяет достичь поставленной цели, даже если она минимизирует ошибку прогноза. Аналитики предпочитают использовать константу сглаживания α в диапазоне от 0,01 до 0,3. Выбор оптимальной константы сглаживания часто основан на интуиции и опыте.

Важная процедура проверки качества любой статистической модели – анализ остатков. Под остатками в модели экспоненциального сглаживания понимают ряд ошибок, полученный взятием разности между наблюдаемыми и предсказанными значениями:

$$\hat{e}_t = Z_t - \hat{Z}_t\,,\ t = 1,\ldots,n\,.$$

Для качественного анализа моделей строят график зависимости $\hat{e}_t$ от t по полному периоду наблюдений. Так как сглаживание должно выделять тренд-циклический и сезонный компоненты (т.е. все регулярные компоненты модели), то остатки представляют собой случайный (нерегулярный) компонент. Если модель сглаживания выбрана верно, то остатки должны быть случайно разбросаны относительно нуля, не демонстрируя никакого систематического шаблона. Если же в модели есть дефект, то случайность распределения остатков уступает место некоторой системе, анализ которой иногда позволяет обнаружить пути улучшения модели.

В самом лучшем случае остатки представляют собой некоррелированные случайные величины с нулевым математическим ожиданием. Случайный процесс, генерирующий такие остатки, называется белым шумом. Проверить гипотезу о том, что процесс остатков – это белый шум, можно, проведя исследование автокорреляционной и частной автокорреляционной функций. Подробно эти функции будут рассмотрены далее, так как они являются одним из основных инструментов идентификации модели ARIMA. Сейчас отметим лишь, что для процесса белого шума оценки автокорреляционной и частной автокорреляционной функций должны быть близки к нулю и не выявлять каких-либо регулярных зависимостей. Если после выполнения экспоненциального сглаживания ряд остатков не выявляет регулярных зависимостей, то сглаживание выполнено качественно – значит, удалось выделить в сглаженном ряде все регулярные составляющие, оставив случайную часть в ряде остатков.

2.2.2.5. ЭКСПОНЕНЦИАЛЬНОЕ СГЛАЖИВАНИЕ В СИСТЕМЕ *STATISTICA*

В системе *Statistica* набор методов экспоненциального сглаживания доступен в окне ***Seasonal and Non-Seasonal Exponential Smoothing***, которое становится активным при нажатии кнопки ***Exponential Smoothing&Forecasting*** на стартовой панели модуля анализа временных рядов. При выборе закладки ***Advanced*** появляется группа возможных моделей, предложенных в [12].

Временной ряд может либо не содержать сезонного компонента, либо содержать аддитивный или мультипликативный сезонный компонент. Кроме того, можно выбрать модель без тренда либо выбрать модель с линейным, экспоненциальным или затухающим трендом. Комбинация этих характеристик дает 12 различных моделей экспоненциального сглаживания (рис. 2.13) и включает классические модели Хольта и Винтера, а также модель простого экспоненциального сглаживания, доступную в окне ***Transformations*** по закладке ***Smoothing***.

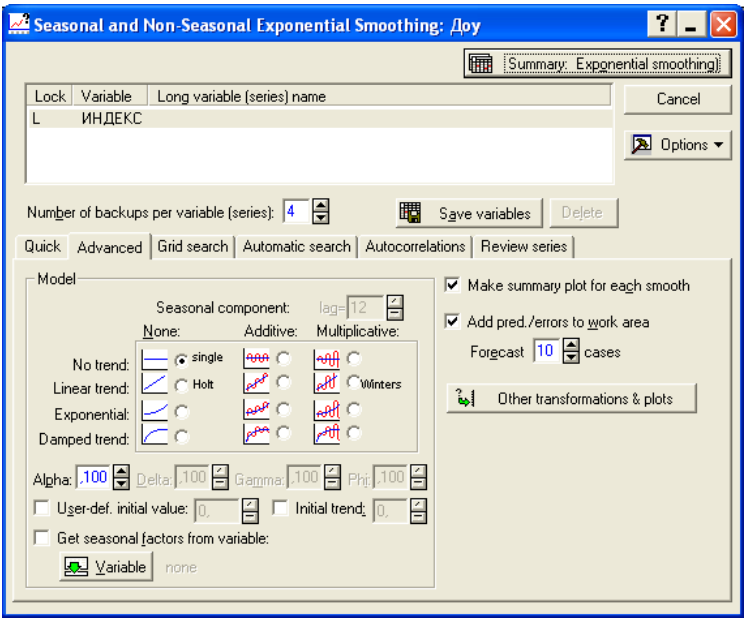


Рис. 2.13. Выбор модели экспоненциального сглаживания

Как подчеркивалось при описании основных моделей экспоненциального сглаживания, исследователю нужно самому задать значения констант сглаживания, вводя их в соответствующие окошки ***Alpha***, ***Delta***, ***Gamma*** и ***Phi*** в закладке ***Advanced***.

Еще ниже в закладке ***Advanced*** можно задать определенные пользователем начальные значения размещения ***User-def. Initial value*** S_0 и тренда ***Initial trend*** T_0, а также массив сезонных индексов ***Get seasonal factors from variable.*** Если пользователь не ввел эти значения, то они

определяются системой по умолчанию. Справа в этой же закладке можно задать количество шагов, на которые необходимо вычислить прогноз. Заметим, что для модели без тренда и сезонности прогноз вычисляется на один шаг, для модели с трендом – на 4–5 шагов. Для сезонной модели прогноз может вычисляться на целый период и даже больше, в зависимости от числа целых периодов в выборке.

В табл. 2.2 приводятся все 12 моделей экспоненциального сглаживания. Для каждой из них в соответствующей клетке записываются константы сглаживания, приводится пример задачи прогнозирования, для которой эта модель может быть использована, а также даны формулы для вычисления начальных значений размещения S_0 и тренда T_0. Приведенные формулы используются системой, если пользователь не ввел значения в соответствующие клетки. Обозначения в формулах следующие: n – общее количество наблюдений; M_1 – среднее для первого сезонного цикла; M_2 – среднее для второго сезонного цикла; k – число полных сезонных циклов; M_k – среднее для последнего сезонного цикла; p – длина сезонного цикла; Z_1 – первое значение временного ряда; Z_2 – второе значение временного ряда; Z_n – последнее значение временного ряда.

Отметим, что во всех случаях сезонного сглаживания, если пользователь не ввел в соответствующем окне набор сезонных индексов, то они определяются системой по данным, содержащимся в выборке, как в методе сезонной декомпозиции, что будет рассмотрено нами далее.

Как было сказано ранее, выбор констант сглаживания очень субъективный. В системе *Statistica* реализованы две процедуры, помогающие организовать поиск подходящих значений констант. Это поиск на сетке и автоматический поиск. Поиск на сетке (***Grid search***) реализуется при выборе закладки с одноименным названием, как показано на рис. 2.14. После выбора начального значения, шага сетки и конечного значения для всех констант сглаживания выбранной модели и последующего нажатия кнопки ***Perform grid search*** на экран выводятся 10 комбинаций значений констант, дающих минимальные значения критерия минимальной среднеквадратичной ошибки (2.16). Следует обратить внимание, что, кроме значения критерия минимальной среднеквадратичной ошибки для этих 10 моделей, на экран выводятся значения пяти других критериев, в частности, средней абсолютной ошибки (2.17) и средней абсолютной процентной ошибки (2.18), как показано на рис. 2.15.

Таблица 2.2

Модели экспоненциального сглаживания

Модели	Несезонные	Аддитивная сезонность	Мультипликативная сезонность
Без тренда	α; простое экспоненциальное сглаживание; S_0 равно среднему по всей выборке	α, δ; предсказание месячного количества осадков (общее количество слабо меняется от года к году, но в разные месяцы одного года сильно различается – сезоны дождей); S_0 равно среднему по всем сезонным периодам.	α, δ; предсказание объема продаж редких игрушек (слабо меняется от года к году, но к декабрю возрастает в зависимости от общего объема продаж; S_0 равно среднему по всем сезонным периодам
Л Без тренда тренд	α, γ; двухпараметрический метод Хольта; прогноз необходимого количества запасных деталей станков; $T_0 = (Z_n - Z_1)/n$; $S_0 = Z_1 - T_0/2$	α, δ, γ; предсказание месячного бюджета для уборки снега в некоторой области, характеризующегося возрастающим трендом вследствие ежегодного роста затрат и аддитивной сезонностью, если затраты постоянны от года к году; $T_0 = (M_k - M_1)/(p(k-1))$; $S_0 = M_1 - T_0/2$	α, δ, γ; трехпараметрический метод Винтера; предсказание месячного бюджета уборки снега в некоторой области, характеризующегося возрастающим трендом, а также мультипликативной сезонностью, если затраты определяются процентной частью бюджета области; $T_0 = (M_k - M_1)/(p(k-1))$; $S_0 = M_1 - pT_0/2$

41

Окончание табл. 2.2

Модели	Несезонные	Аддитивная сезонность	Мультипликативная сезонность
Экспоненциальный тренд	α, γ; предсказание месячной стоимости ремонта производственного оборудования, если ежегодно затраты увеличиваются на некоторый процент; $T_0 = Z_2 / Z_1$; $S_0 = Z_1 / \sqrt{T_0}$	α, δ, γ; предсказание ежемесячного дохода от курортной деятельности (ежегодно доход может увеличиваться в процентном соотношении, но к декабрю каждого года увеличение дохода фиксированное); $T_0 = e^{(\ln(M_k) - \ln(M_1))/p}$; $S_0 = e^{\ln(M_1) - p\ln(T_0)/2}$	α, δ, γ; предсказание ежемесячного дохода от курортной деятельности (ежегодно доход может увеличиваться в процентном соотношении, и к декабрю каждого года доход увеличивается в процентах от среднего уровня); $T_0 = e^{(\ln(M_2) - \ln(M_1))/p}$; $S_0 = e^{\ln(M_1) - p\ln(T_0)/2}$

Экспоненциальный тренд

42

Затухающий тренд	α, γ; ϕ для эффекта затухания; исследование процентного соотношения домохозяйств, владеющих тем или иным электронным устройством, наблюдается ежегодное увеличение, но тренд будет затухающим по мере насыщения рынка; $T_0 = (1/\phi)(Z_n - Z_1)/(n-1)$; $S_0 = Z_1 - T_0/2$	α, δ; ϕ для эффекта затухания; исследование процентного соотношения домохозяйств, владеющих каким-либо электронным устройством, наблюдается ежегодное увеличение, но тренд будет затухающим по мере насыщения рынка; кроме того, наблюдается фиксированное снижение в летние месяцы и увеличение к декабрю; $T_0 = (1/\phi)(M_k - M_1)/((k-1)p)$; $S_0 = M_1 - pT_0/2$	α, δ; ϕ для эффекта затухания; исследование процентного соотношения домохозяйств, владеющих каким-либо электронным устройством, наблюдается ежегодное увеличение, но тренд будет затухающим по мере насыщения рынка; кроме того наблюдается снижение в летние месяцы и увеличение к декабрю на определенный процент от общего числа; $T_0 = (1/\phi)(M_k - M_1)/((k-1)p)$; $S_0 = M_1 - pT_0/2$

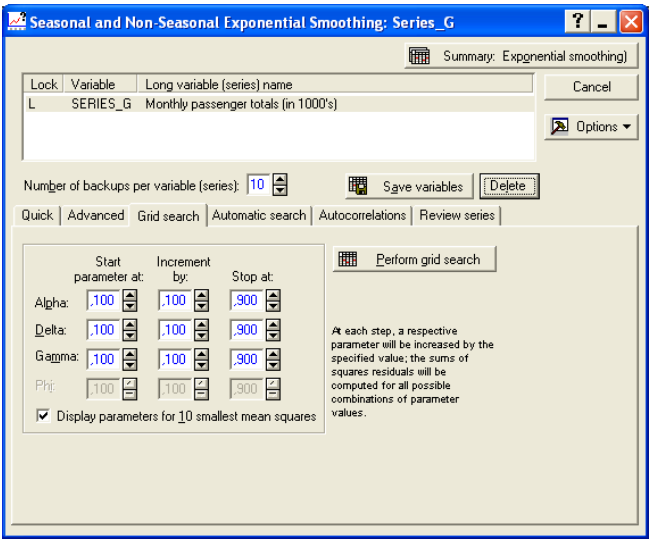
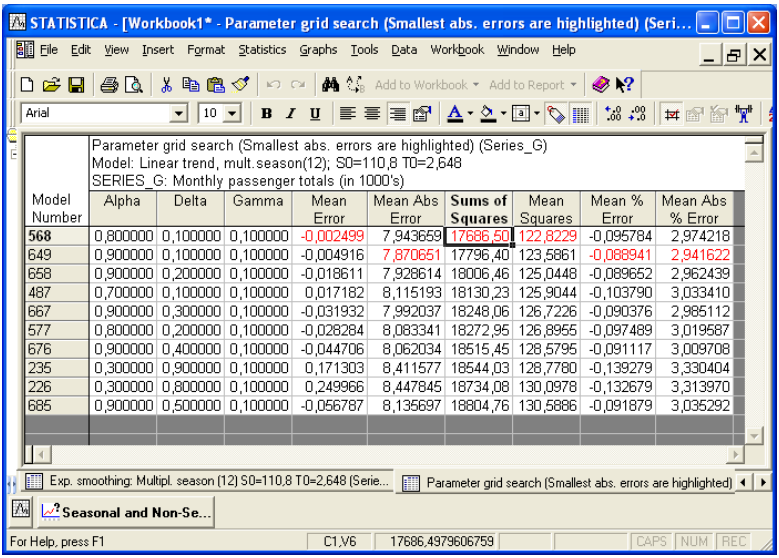


Рис. 2.14. Закладка ***Grid search*** окна экспоненциального сглаживания



Parameter grid search (Smallest abs. errors are highlighted) (Series_G)
Model: Linear trend, mult.season(12); S0=110,8 T0=2,648
SERIES_G: Monthly passenger totals (in 1000's)

Model Number	Alpha	Delta	Gamma	Mean Error	Mean Abs Error	**Sums of Squares**	Mean Squares	Mean % Error	Mean Abs % Error
568	0,800000	0,100000	0,100000	-0,002499	7,943659	17686,50	122,8229	-0,095784	2,974218
649	0,900000	0,100000	0,100000	-0,004916	7,870651	17796,40	123,5861	-0,088941	2,941622
658	0,900000	0,200000	0,100000	-0,018611	7,928614	18006,46	125,0448	-0,089652	2,962439
487	0,700000	0,100000	0,100000	0,017182	8,115193	18130,23	125,9044	-0,103790	3,033410
667	0,900000	0,300000	0,100000	-0,031932	7,992037	18248,06	126,7226	-0,090376	2,985112
577	0,800000	0,200000	0,100000	-0,028284	8,083341	18272,95	126,8955	-0,097489	3,019587
676	0,900000	0,400000	0,100000	-0,044706	8,062034	18515,45	128,5795	-0,091117	3,009708
235	0,300000	0,900000	0,100000	0,171303	8,411577	18544,03	128,7780	-0,139279	3,330404
226	0,300000	0,800000	0,100000	0,249966	8,447845	18734,08	130,0978	-0,132679	3,313970
685	0,900000	0,500000	0,100000	-0,056787	8,135697	18804,76	130,5886	-0,091879	3,035292

Рис. 2.15. Первые 10 оптимальных моделей сглаживания

В системе реализована и более эффективная автоматическая процедура поиска оптимальных констант сглаживания, рекомендованная в [12]. Ее преимущество больше всего проявляется в случае нескольких констант сглаживания.

Провести автоматическую оптимизацию можно, выбрав закладку ***Automatic search*** (рис. 2.16). В окне автоматической оптимизации предлагается выбрать один из трех критериев оптимальности (2.16), (2.17) или (2.18), начальные значения констант сглаживания, а также условия сходимости итерационного алгоритма.

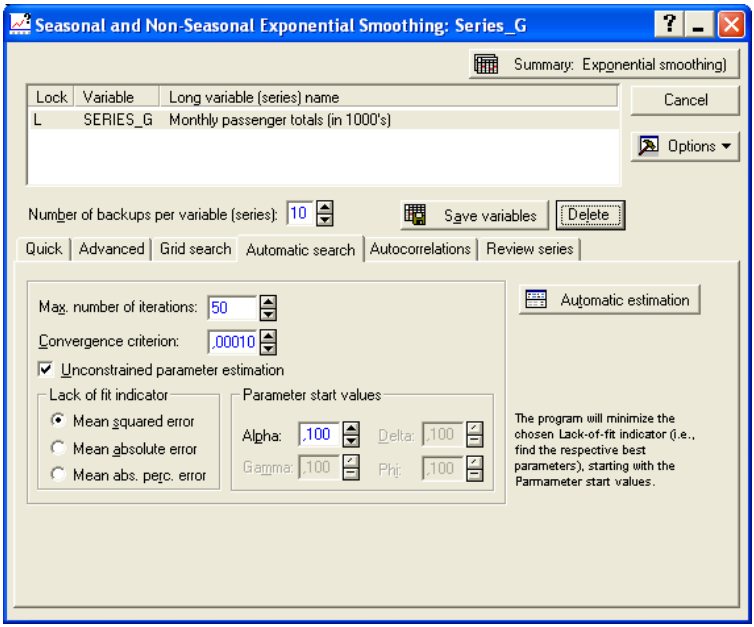


Рис. 2.16. Закладка ***Automatic search*** окна экспоненциального сглаживания

После нажатия кнопки ***Automatic estimation*** начинает работать итерационный алгоритм оптимизации методом квази-Ньютона, в результате которого получается оптимальная модель, характеризующаяся соответствующим оптимальным набором сглаживающих констант. В двух отдельных окнах в числовом и графическом виде представляются исходный ряд наблюдений, сглаженный ряд и ряд остатков, а для

сезонных моделей – ряд сезонных индексов. В третьем окне представлены значения шести критериев для полученной модели, оптимальной по выбранному критерию.

2.2.2.6. ЭКСПОНЕНЦИАЛЬНОЕ СГЛАЖИВАНИЕ ВРЕМЕННОГО РЯДА АВИАПЕРЕВОЗОК

Рассмотрим технику экспоненциального сглаживания в системе *Statistica* на примере временного ряда авиаперевозок, график которого представлен на рис. 2.1. На графике временного ряда наблюдаются явный тренд, возрастающий более чем линейно, а также сезонность с увеличивающимся ежегодным приращением. Поэтому наиболее подходящим методом сглаживания предположительно является трехпараметрический метод с мультипликативной сезонностью и экспоненциальным трендом.

Выбрав соответствующий метод в закладке ***Advanced***, применим автоматический поиск оптимальных констант сглаживания, открыв закладку ***Automatic search***. Выбрав параметры оптимизации, как показано на рис. 2.17, после завершения итерационного процесса получим оптимальные значения констант сглаживания:

$$\alpha = 0{,}260; \quad \delta = 0{,}831; \quad \gamma = 0{,}001,$$

для которых достигается минимальное значение критерия среднеквадратичной ошибки MSE = 116,08. Следует заметить, что сходимость итерационного процесса для данного временного ряда оказалась сильно зависимой от выбора начальных значений констант и условий сходимости.

Проанализируем полученные результаты. На рис. 2.18 представлены три временных ряда: исходный ряд, сглаженный ряд и ряд остатков, причем для сглаженного ряда приводится прогноз на 10 шагов вперед. Визуальный анализ позволяет сказать, что сглаженный ряд хорошо повторяет закономерности исходного ряда, а прогноз чутко улавливает основную тенденцию как тренда, так и сезонного компонента. Ряд остатков выявляет случайный разброс относительно нуля. Однако автокорреляционная и частная автокорреляционная функции, полученные для ряда остатков в закладке ***Autocorrelations*** и представленные на рис. 2.19 и 2.20, все же выявляют некоторые закономерности и выбросы, оставшиеся в ряде остатков после выделения сглаживающего компонента. Отметим, что для метода экспоненциального сглаживания этот результат довольно неплохой.

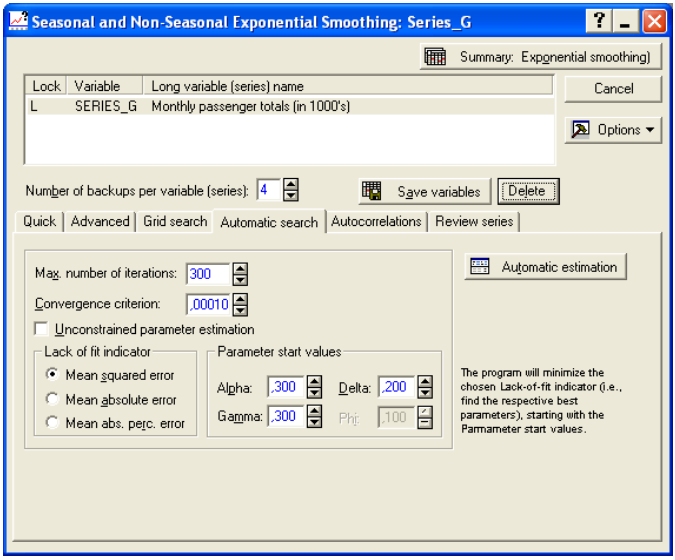


Рис. 2.17. Выбор критерия оптимальности и параметров сходимости

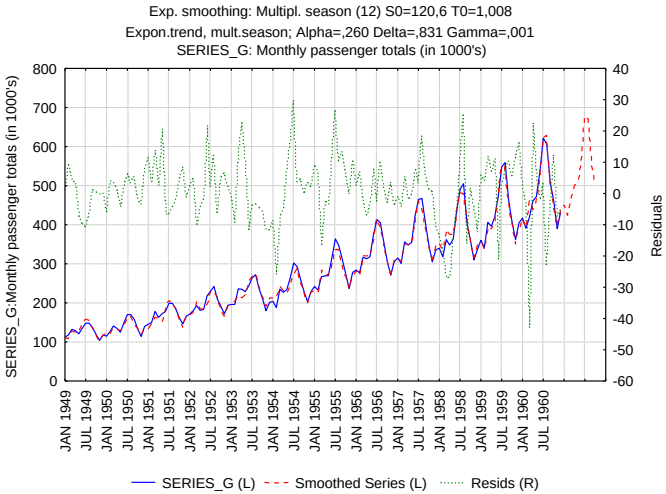


Рис. 2.18. Результаты трехпараметрического экспоненциального сглаживания

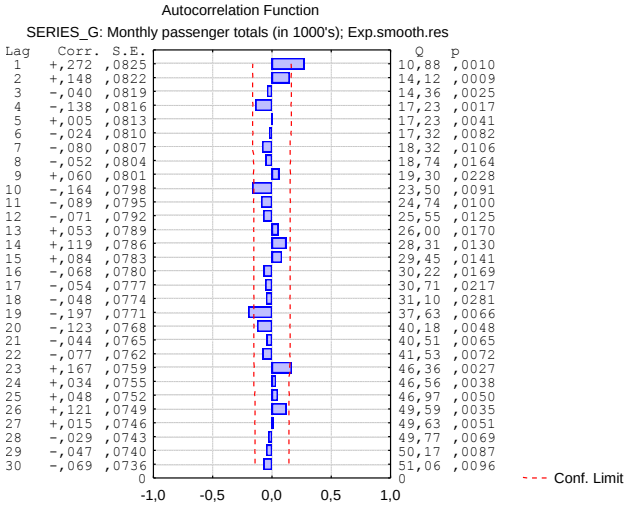


Рис. 2.19. Автокорреляционная функция остатков трехпараметрической модели

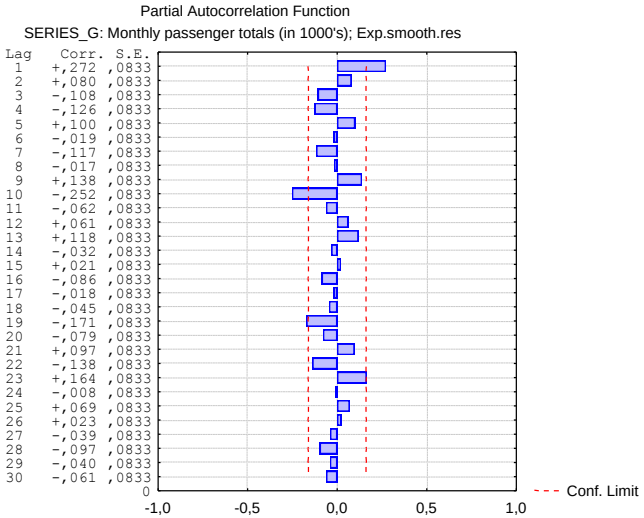


Рис. 2.20. Частная автокорреляционная функция трехпараметрической модели

Для сравнения приведем пример неадекватной сглаживающей (по тому же самому ряду авиаперевозок). Пусть мы учли только тренд, не обратив внимания на сезонность ряда, и выбрали для сглаживания двухпараметрическую модель Хольта. Результаты сглаживания приведены на рис. 2.21.

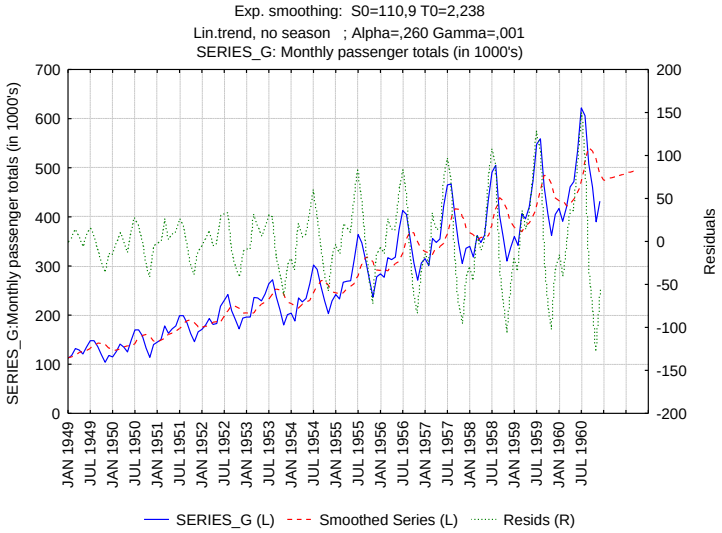


Рис. 2.21. Результаты двухпараметрического экспоненциального сглаживания

Во-первых, сглаженный ряд плохо отслеживает сезонный компонент. Это видно также по ряду остатков, в котором наблюдается оставшаяся сезонность – ежегодные повторения тенденции. Во-вторых, в ряде остатков присутствует явное увеличение изменчивости с увеличением времени, так называемая неоднородность дисперсии. В-третьих, прогноз совершенно не улавливает сезонную тенденцию, а лишь линейный возрастающий тренд. Кроме того, значение критерия MSE равно 2073,9, что более чем на порядок больше значения, полученного для трехпараметрической модели сглаживания. Рассмотрим автокорреляционную и частную автокорреляционную функции остатков (рис. 2.22 и 2.23).

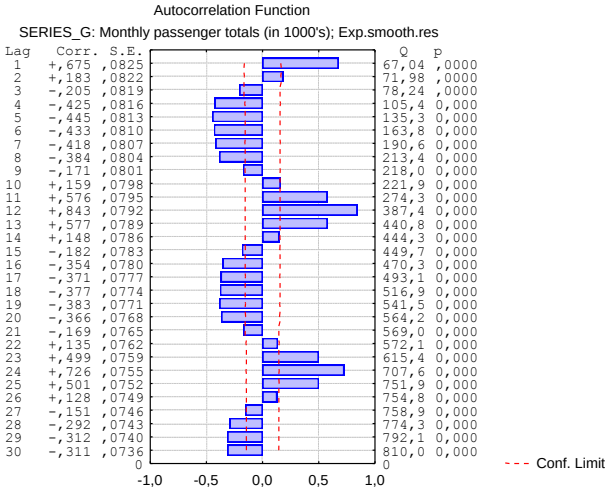


Рис. 2.22. Автокорреляционная функция остатков модели Хольта

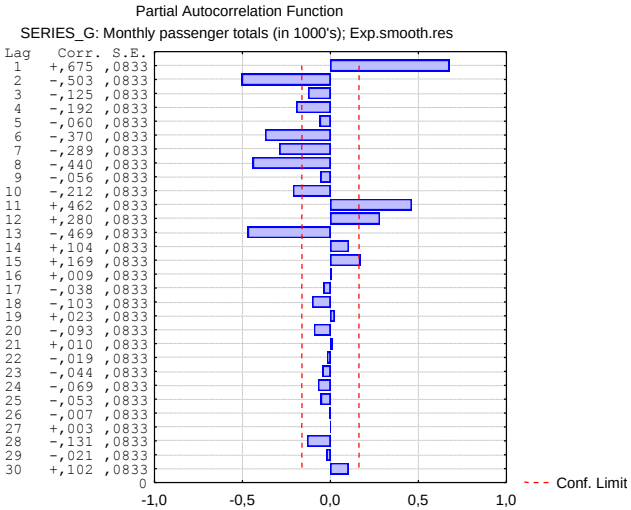


Рис. 2.23. Частная автокорреляционная функция остатков модели Хольта

Здесь автокорреляционная и частная автокорреляционная функции выявляют яркую сезонность и коррелированность остатков, что свидетельствует о несоответствии модели двухпараметрического сглаживания наблюдаемому временному ряду.

Завершая рассмотрение вопроса о методе экспоненциального сглаживания, отметим, что, несмотря на субъективность выбора констант сглаживания и начальных значений параметров, а также отсутствие строгих процедур проверки адекватности модели, этот метод служит полезным инструментом для предсказания значений временного ряда, а также служит базисом для развития более эффективных адаптивных методов прогнозирования, описанных в работе [8].

2.3. СЕЗОННАЯ ДЕКОМПОЗИЦИЯ ВРЕМЕННЫХ РЯДОВ

2.3.1. КЛАССИЧЕСКАЯ СЕЗОННАЯ ДЕКОМПОЗИЦИЯ

Сезонная декомпозиция – это частный случай общей декомпозиции, когда ряд представляется в виде двух компонентов, один из которых легко прогнозируется, другой имеет ярко выраженный нерегулярный характер. Выделяя один за другим компоненты ряда, мы постепенно снимаем неопределенность исходного ряда. Выделенный тренд можно моделировать с использованием, например, линейной или нелинейной регрессии, после чего осуществляется прогноз.

В подразделе 2.1 были введены классическая мультипликативная и аддитивная модели (2.1) и (2.2), в которых выделяются четыре компонента: трендовый, сезонный, циклический и нерегулярный. Циклический компонент отличается от сезонного тем, что обычно длина цикла больше длины периода сезонности. Кроме того, различные циклы могут иметь разную длину. Поэтому выделение циклического компонента отдельно от трендового на основе только количественного анализа затруднительно. В методе сезонной декомпозиции вместо двух компонентов, трендового и циклического, рассматривается объединенный тренд-циклический компонент. Тогда мультипликативная модель принимает вид

$$Z_t = TC_t S_t I_t\,,\ t = 1,\ldots,n\,, \tag{2.19}$$

а аддитивная модель – соответственно

$$Z_t = TC_t + S_t + I_t\,,\ t = 1,\ldots,n\,, \tag{2.20}$$

где TC_t – тренд-циклический компонент, S_t – сезонный компонент, I_t – нерегулярный компонент.

Классическая техника выполнения такой декомпозиции известна как метод Census I. Рассмотрим, как этот метод реализован в системе *Statistica*. Алгоритм включает несколько последовательных этапов.

1. На первом шаге декомпозиции выполняется сглаживание исходного временного ряда Z_t центрированным скользящим средним по формуле

$$\tilde{Z}_t = \frac{1}{N} \sum_{i=(1-N)/2}^{(N-1)/2} \alpha_i Z_{t+i} ,$$

где N выбирается равным длине периода сезонности p, если p – нечетное, и $N = p + 1$, если p – четное. Если p – нечетное, то веса α_i выбираются одинаковыми и равными, например, единице. Если p – четное, то первый вес и последний вес $\alpha_{(1-N)/2}$ и $\alpha_{(N-1)/2}$ целесообразно заменить их средним, равным 1/2.

2. Так как предполагается, что ежегодная сумма сезонного временного ряда обладает маленькими сезонными вариациями, то сглаженный ряд $\tilde{Z}_t$ не содержит сезонного компонента. Поэтому можно выделить ряд

$$SI_t = Z_t / \tilde{Z}_t \text{ – для мультипликативной модели,}$$

или

$$SI_t = Z_t - \tilde{Z}_t \text{ – для аддитивной модели.}$$

Ряд SI_t содержит смесь сезонного и нерегулярного компонентов.

3. На этом шаге вычисляется ряд сезонных индексов, т.е. фактически из ряда SI_t выделяется сезонный компонент S_t. Сезонные индексы вычисляются как среднее арифметическое по всем точкам сезонного периода – для аддитивной модели, и как среднее арифметическое после удаления минимального и максимального значений по всем точкам сезонного периода – для мультипликативной модели. Например, если наблюдения снимаются ежемесячно и период сезонности равен 12, то среднее арифметическое вычисляется по наблюдениям в январе каждого года, затем по наблюдениям в феврале каждого года, …, декабре. В вычисленном таким образом ряде сезонных индексов S_t одни и те же значения повторяются через каждые 12 элементов.

4. На этом шаге выполняется сезонная корректировка исходного ряда. В результате получается сезонно скорректированный ряд

$$TCI_t = Z_t / S_t \text{ – для мультипликативной модели,}$$

или

$$TCI_t = Z_t - S_t \text{ – для аддитивной модели.}$$

Сезонно скорректированный ряд не содержит сезонного компонента, а представляет собой смесь тренд-циклического и нерегулярного компонентов.

5. Выделяется тренд-циклический компонент из сезонно скорректированного ряда TCI_t. Выделение производится применением к ряду TCI_t сглаживания 5-точечным центрированным симметричным взвешенным скользящим средним с весами 1, 2, 3, 2, 1:

$$Z_{t-2} + 2Z_{t-1} + 3Z_t + 2Z_{t+1} + Z_{t+2}.$$

Сглаженный ряд представляет собой тренд-циклический компонент TC_t.

6. Выделяется нерегулярный компонент I_t:

$$I_t = TCI_t / TC_t \text{ – для мультипликативной модели,}$$

или

$$I_t = TCI_t - TC_t \text{ – для аддитивной модели.}$$

В системе *Statistica* классическая сезонная декомпозиция выполняется в диалоге ***Seasonal decomposition (Census1)***, который становится доступным при нажатии одноименной кнопки. Выбрав закладку *Advanced*, можно выбрать тип модели (аддитивная или мультипликативная), длину сезонного периода, включить или выключить специальную опцию ***Centered moving average*** для четной длины периода сезонности, а также выбрать нужные для дальнейшей работы компоненты, получаемые на каждом шаге алгоритма декомпозиции (всего шесть компонентов, соответственно числу шагов). Выключенная опция ***Centered moving average*** соответствует сглаживанию скользящим средним с равными весами на первом шаге алгоритма, включенная опция – сглаживанию с неравными весами, когда первое и последнее наблюдения в формуле усредняются.

Применим метод сезонной декомпозиции для анализа временного ряда авиаперевозок. Зададим опции декомпозиции, как на рис. 2.24, и нажмем кнопку ***Summary:Seasonal decomposition***. Перед этим надо убедиться, что число зарезервированных мест для переменной ***Number of backups per variable*** не меньше числа выбранных для дальнейшей работы компонентов.

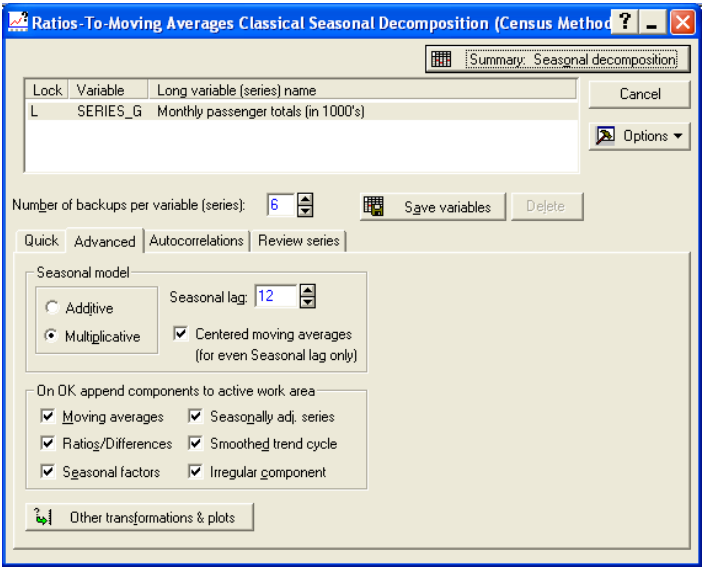


Рис. 2.24. Выбор параметров сезонной декомпозиции

В результате декомпозиции создается таблица с числовыми значениями семи временных рядов: исходного временного ряда и шести компонентов, выделенных на каждом из шести этапов декомпозиции. В дальнейшем для большей наглядности можно представить нужные компоненты на графике, выбрав в закладке ***Review Series*** кнопку ***Review Multiple Series*** и пометив те компоненты, которые необходимо изобразить графически. После выделения исходного ряда, сезонного, тренд-циклического и случайного компонентов получаем график, изображенный на рис. 2.25. На этом графике незаметны осцилляции случайного компонента, так как значения отклонений намного меньше уровня исходного ряда. Чтобы увидеть случайные колебания, необходимо построить отдельный график для этого компонента (рис. 2.26).

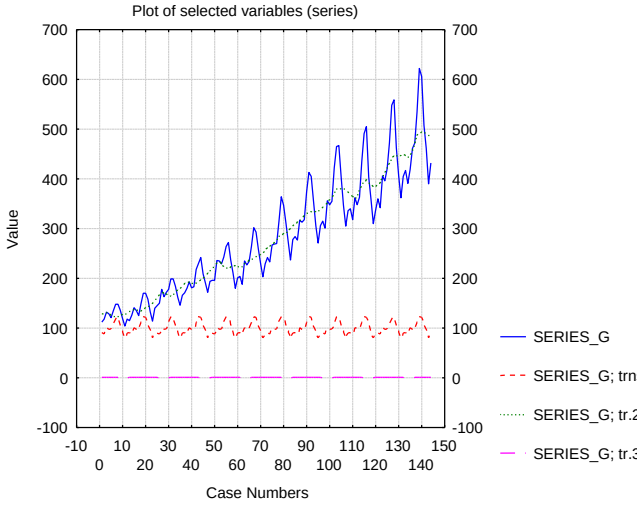


Рис. 2.25. Результаты декомпозиции временного ряда авиаперевозок

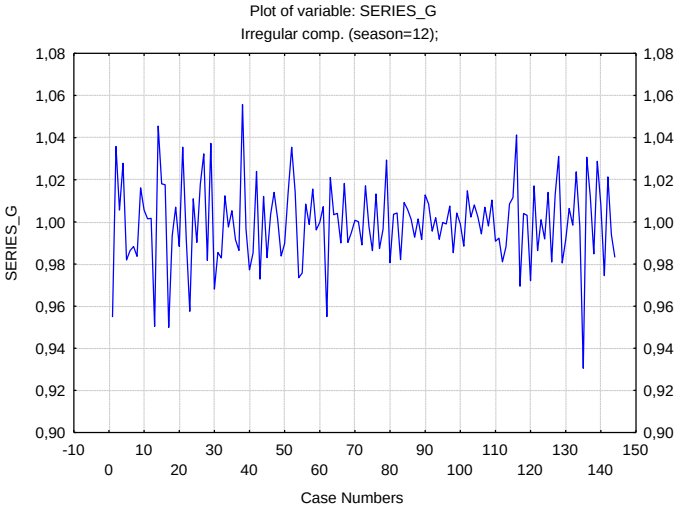


Рис. 2.26. График нерегулярного компонента декомпозиции ряда авиаперевозок

Для проверки качества модели рассмотрим графики автокорреляционной и частной автокорреляционной функций для выделенного нерегулярного (случайного) компонента, полученные в закладке *Autocorrelations* (рис. 2.27 и 2.28). Как видно из этих рисунков, имеются довольно значимые корреляции элементов нерегулярного компонента, т.е. с помощью метода сезонной декомпозиции не до конца удается выделить все содержащиеся в исходном ряде закономерности.

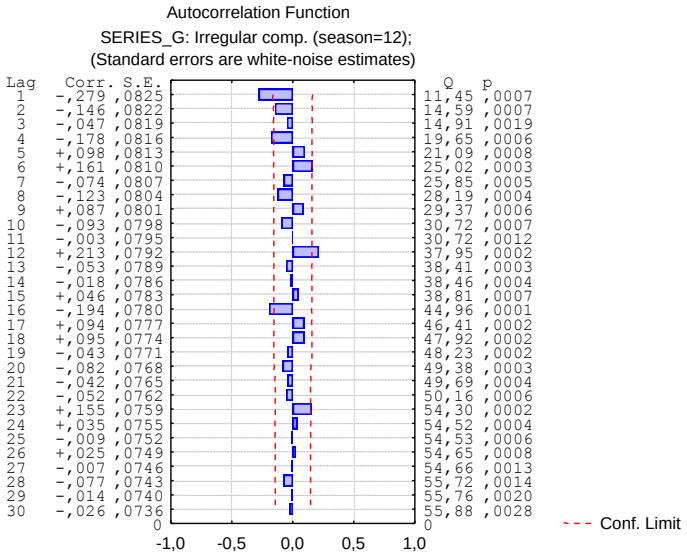


Рис. 2.27. Автокорреляционная функция нерегулярного компонента

После выделения компонентов временного ряда они могут быть использованы для прогнозирования основных тенденций процесса. В зависимости от используемого компонента декомпозиции ряда естественно возникает классификация прогнозов на долго-, средне- и краткосрочные. Для долгосрочных прогнозов наиболее важен трендциклический компонент, для среднесрочных – сезонный компонент, для краткосрочных – нерегулярный компонент. Для работы с трендциклическим компонентом могут оказаться полезными методы как линейного, так и нелинейного регрессионного анализа.

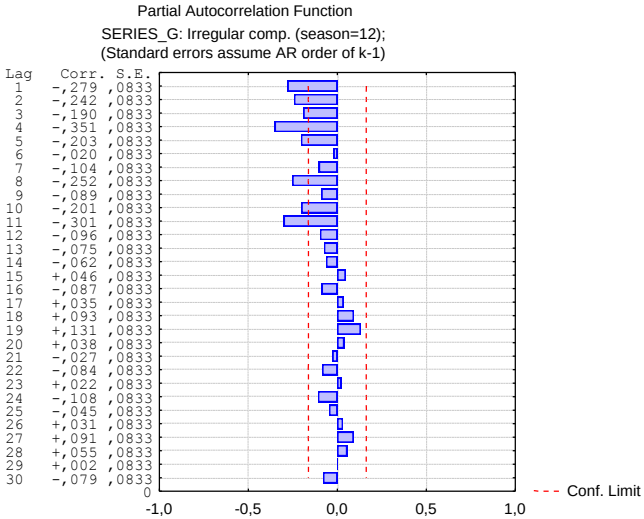


Рис. 2.28. Частная автокорреляционная функция нерегулярной компоненты

В системе *Statistica*, кроме метода сезонной декомпозиции Census 1, реализован и другой метод декомпозиции – Census 2 (XII-метод), разработанный в бюро переписей США и максимально приближенный к стандартам, применяемым в этой стране. Отличие Census 2 от Census 1 состоит в том, что в нем производится корректировка на разные дни недели и на разное число дней в месяце (месячная корректировка), а также на месяцы с экстремальными наблюдениями (квартальная корректировка).

2.3.2. СЕЗОННАЯ ДЕКОМПОЗИЦИЯ ВРЕМЕННОГО РЯДА ПРИМЕНИТЕЛЬНО К ИСХОДНЫМ ДАННЫМ, ХАРАКТЕРИЗУЮЩИМ РАБОТУ ТЕРМИНАЛА В СЕТИ ПЛАТЕЖНОЙ СИСТЕМЫ

Рассмотрим исходные данные, которые отражают работу терминала в сети платежной системы Новосибирска ежедневно с 1 декабря 2007 года по 15 мая 2008 (рис. 2.29). Нас интересует последняя графа приведенной таблицы данных, где указывается размер суммы, поступившей с данного терминала в пользу поставщиков услуг. Рассмотрим данные этой графы как временной ряд. График ряда изображен на рис. 2.30.

	1 Год	2 Месяц	3 День	4 Дата	5 ДеньНед	6 Выходной	7 :левыход	8 едвыход	9 :во Плате	**10** **а Постав**
1	2007	12	1	12/1/07	6	1	0	0	83	7003,7
2	2007	12	2	12/2/07	7	1	0	0	57	4194,3
3	2007	12	3	12/3/07	1	0	1	0	95	10122
4	2007	12	4	12/4/07	2	0	0	0	51	3924,6
5	2007	12	5	12/5/07	3	0	0	0	67	5719,5
6	2007	12	6	12/6/07	4	0	0	0	59	5375,4
7	2007	12	7	12/7/07	5	0	0	1	84	6054,3
8	2007	12	8	12/8/07	6	1	0	0	43	3180,6
9	2007	12	9	12/9/07	7	1	0	0	50	3283,3
10	2007	12	10	12/10/07	1	0	1	0	74	5291,7
11	2007	12	11	12/11/07	2	0	0	0	70	6919,2
12	2007	12	12	12/12/07	3	0	0	0	54	4296,6
13	2007	12	13	12/13/07	4	0	0	0	62	5133,6
14	2007	12	14	12/14/07	5	0	0	1	63	3943,2
15	2007	12	15	12/15/07	6	1	0	0	60	4464
16	2007	12	16	12/16/07	7	1	0	0	32	2232
17	2007	12	17	12/17/07	1	0	1	0	20	1887,9
18	2007	12	18	12/18/07	2	0	0	0	47	3338,7
19	2007	12	19	12/19/07	3	0	0	0	57	3627
20	2007	12	20	12/20/07	4	0	0	0	50	4408,2
21	2007	12	21	12/21/07	5	0	0	1	63	7319,1
22	2007	12	22	12/22/07	6	1	0	0	52	3524,7
23	2007	12	23	12/23/07	7	1	0	0	50	3989,7
24	2007	12	24	12/24/07	1	0	1	0	72	7170,3
25	2007	12	25	12/25/07	2	0	0	0	60	4715,1

Рис. 2.29. Исходные данные работы терминала в сети платежной системы

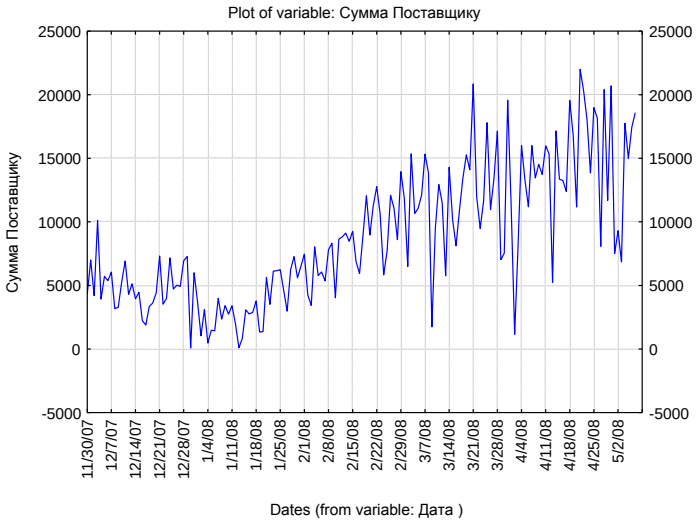


Рис. 2.30. График ряда суммы, поступившей в пользу поставщиков услуг

По графику можно предположить, что имеется недельная периодичность в данных, т.е. тенденция изменения повторяется каждую неделю. Рассмотрим и сравним мультипликативную и аддитивную модели декомпозиции данного временного ряда с периодом сезонности 7.

Начнем с мультипликативной модели. Выберем параметры декомпозиции, как показано на рис. 2.31, и нажмем кнопку ***Summary: Seasonal Decomposition***.

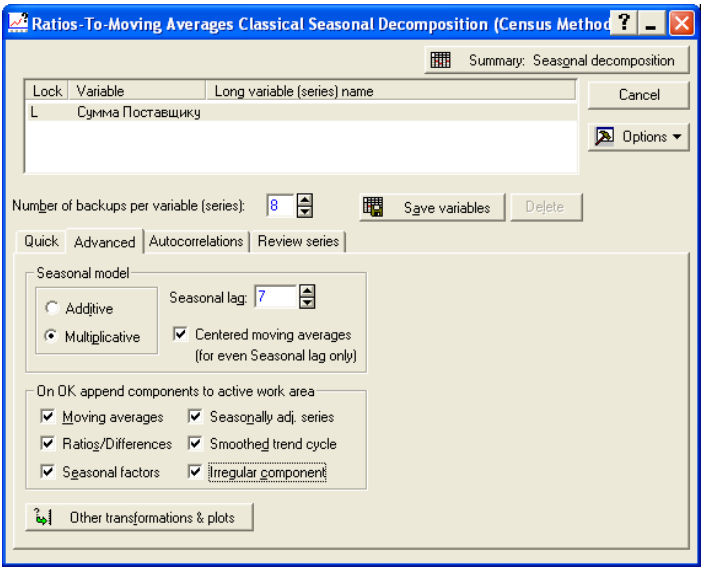


Рис. 2.31. Окно выбора модели декомпозиции ряда данных терминала

В результате имеем шесть желаемых производных временных рядов, полученных на каждом из шести этапов декомпозиции. На рис. 2.32 изображены исходный ряд данных и выделенный тренд-циклический компонент. Выделенный сезонный компонент изображен на рис. 2.33, а нерегулярный компонент – на рис. 2.34. График случайного (нерегулярного) компонента свидетельствует о некоторой неоднородности дисперсии. Автокорреляционная функция и частная автокорреляционная функции случайного компонента (рис. 2.35 и 2.36) свидетельствуют о существовании статистической зависимости между значениями.

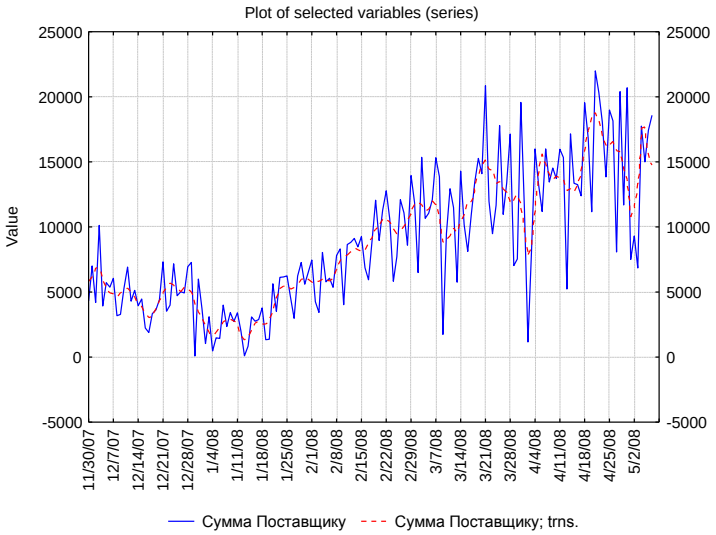


Рис. 2.32. Тренд-циклический компонент мультипликативной модели

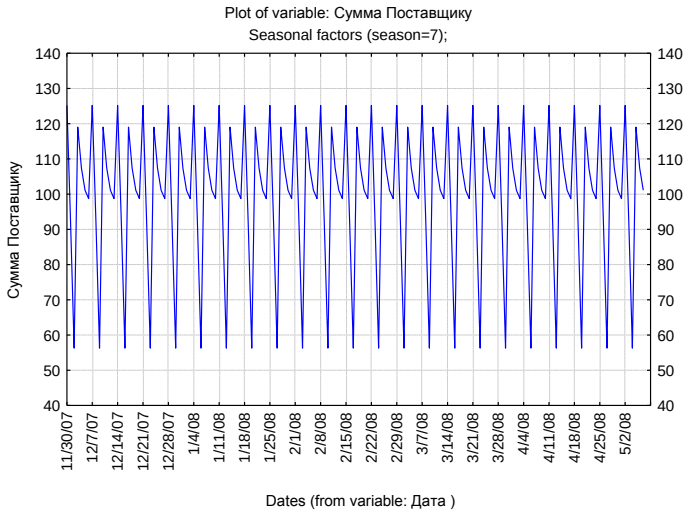


Рис. 2.33. Сезонный компонент мультипликативной модели

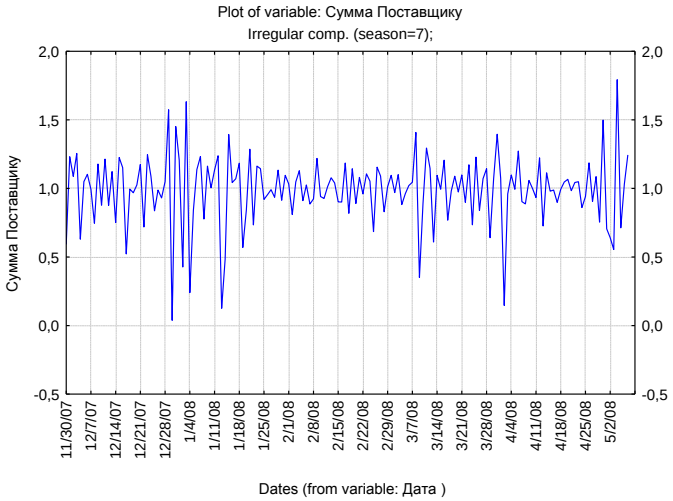


Рис. 2.34. Случайный компонент мультипликативной модели

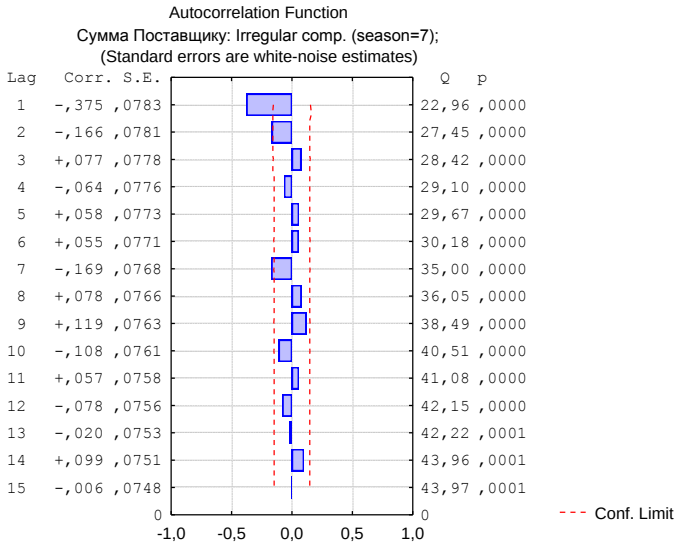


Рис. 2.35. Автокорреляционная функция случайного компонента мультипликативной модели

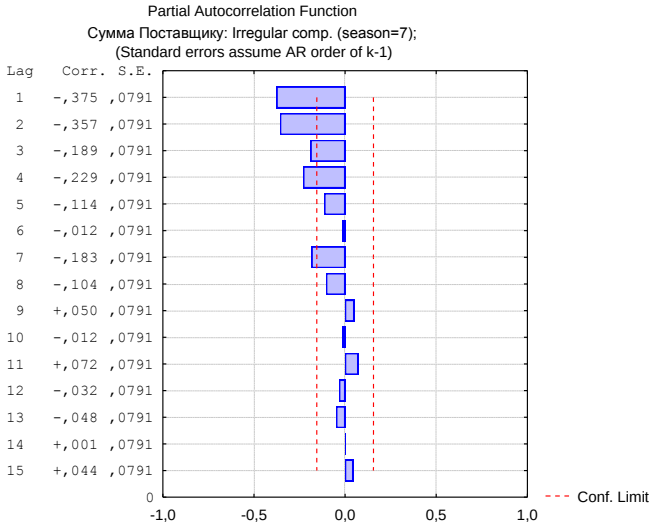


Рис. 2.36. Частная автокорреляционная функция случайного компонента мультипликативной модели

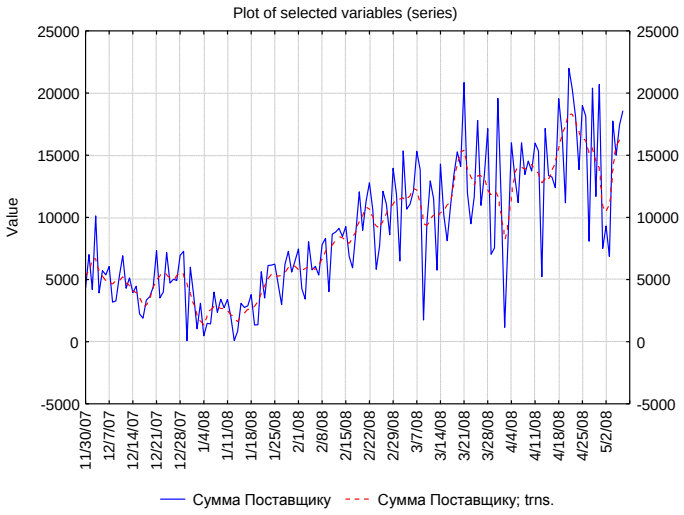


Рис. 2.37. Тренд-циклический компонент аддитивной модели

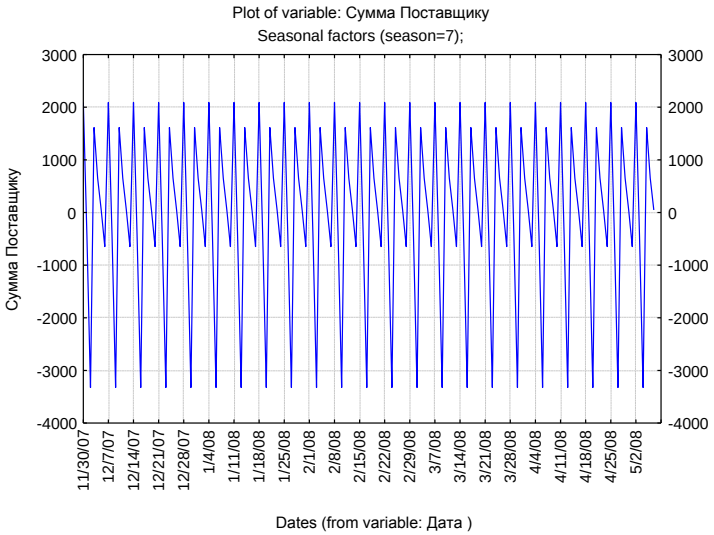


Рис. 2.38. Сезонный компонент аддитивной модели

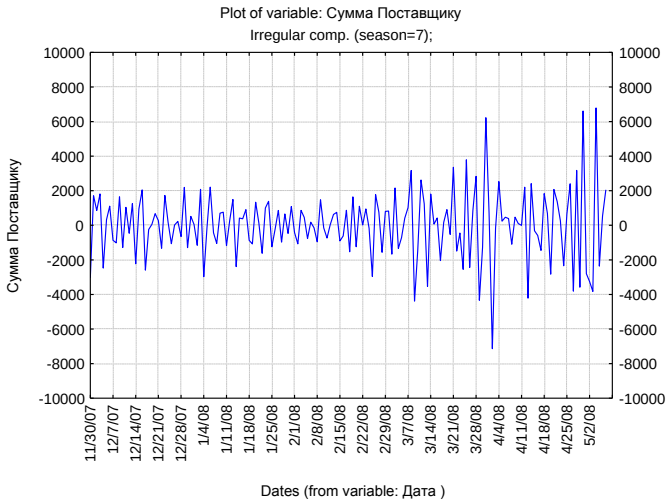


Рис. 2.39. Случайный компонент аддитивной модели

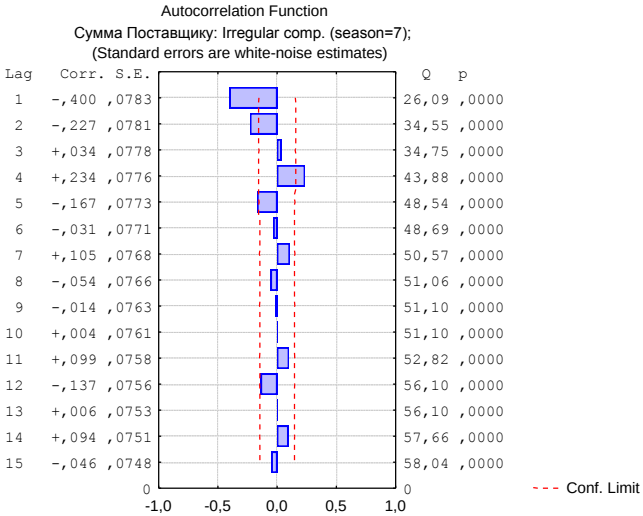


Рис. 2.40. Автокорреляционная функция случайного компонента аддитивной модели

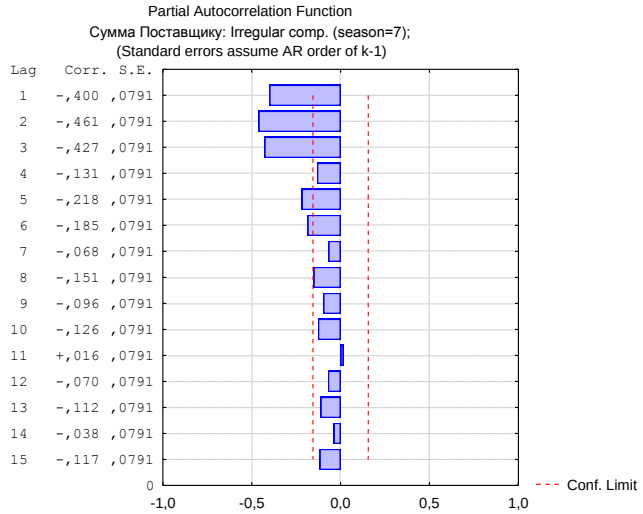


Рис. 2.41. Частная автокорреляционная функция случайного компонента аддитивной модели

Для сравнения рассмотрим аддитивную модель декомпозиции данного ряда. Графики тренд-циклического, сезонного и случайного компонентов для аддитивной модели изображены на рис. 2.37, 2.38 и 2.39 соответственно. Анализируя графики, можно сказать, что мультипликативная модель больше соответствует наблюдаемым данным. Неоднородность дисперсии случайного компонента аддитивной модели более ярко выражена. Значения автокорреляционной функции аддитивной модели более значимо отличаются от нуля, чем соответствующие значения автокорреляционной функции мультипликативной модели (рис. 2.40 и 2.41).

Используя выделенные компоненты временного ряда, можно осуществить анализ каждого из компонент для последующего долгосрочного, среднесрочного и краткосрочного прогнозов с использованием соответствующих компонентов.

Часть II

АНАЛИЗ ВРЕМЕННЫХ РЯДОВ С ИСПОЛЬЗОВАНИЕМ МЕТОДОЛОГИИ ARIMA

3. ВВЕДЕНИЕ В МЕТОДОЛОГИЮ ARIMA

3.1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ С ИСПОЛЬЗОВАНИЕМ МЕТОДОЛОГИИ ARIMA

Теоретической основой наиболее широко распространенной методологии анализа временных рядов и прогнозирования – методологии ARIMA (Бокса–Дженкинса) – является теория случайных процессов [3]. Согласно этой методологии временной ряд представляет собой реализацию некоторого случайного процесса. А задача моделирования заключается в восстановлении характеристик случайного процесса по одной его реализации. Такое восстановление в общем случае невозможно, однако вполне осуществимо при некоторых ограничениях на класс исследуемых процессов, чаще всего встречающихся в практической деятельности.

Итак, вспомним необходимые фундаментальные понятия. *Случайный процесс* – это семейство случайных переменных $Z(w,t)$, индексированных по времени, где w принадлежит выборочному пространству, а t – множеству индексов. Для фиксированного t $Z(w,t)$ является случайной переменной. Для заданного w $Z(w,t)$ есть функция t и называется выборочной функцией, или реализацией Z_t. Совокупность всех возможных реализаций называется ансамблем в теории случайных процессов. Таким образом, временной ряд есть не что иное, как *реализация* некоторого случайного процесса.

Рассмотрим конечное множество случайных переменных $\left\{Z_{t_1},...,Z_{t_2}\right\}$ из случайного процесса $\{Z(w,t) : t = 0, \pm 1, \pm 2, ...\}$. Тогда n-мерная функция распределения определяется следующим образом:

$$F(z_{t_1},...,z_{t_n}) = p\{w : Z(w,t_1) \le z_{t_1},...,Z(w,t_n) \le z_{t_n}\}.$$

Говорят, что процесс является стационарным первого порядка по распределению, если его одномерная функция распределения инвариантна по времени, т.е. $F(z_{t_1}) = F(z_{t_1+k})$ для любых целых t_1, k и $t_1 + k$; стационарным второго порядка по распределению, если $F(z_{t_1}, z_{t_2}) = F(z_{t_1+k}, z_{t_2+k})$ для любых целых t_1, t_2, k, $t_1 + k$ и $t_2 + k$; стационарным n-го порядка по распределению, если

$$F(z_{t_1},...,z_{t_n}) = F(z_{t_1+k},...,z_{t_n+k}) \tag{3.1}$$

для любого набора $(t_1,...,t_n)$ целых чисел и для любого целого k. Очевидно, что стационарность более высокого порядка влечет за собой стационарность более низкого порядка. Наконец, процесс является ***строго стационарным***, если соотношение (3.1) справедливо для $\forall n$, т.е. $n = 1,2,\ldots$.

Рассмотрим свойства строго стационарного процесса. Введем обозначение Z_t для реализации случайного процесса, подчеркивая факт зависимости только от времени. Для заданного вещественнозначного процесса $\{Z_t : t = 0, \pm 1, \pm 2,...\}$ определим функцию среднего $\mu_t = E(Z_t)$, функцию дисперсии $\sigma_t^2 = E(Z_t - \mu_t)^2$, ковариационную функцию между Z_{t_1} и Z_{t_2}:

$$\gamma(t_1,t_2) = E(Z_{t_1} - \mu_{t_1})(Z_{t_2} - \mu_{t_2})$$

и корреляционную функцию между Z_{t_1} и Z_{t_2}:

$$\rho(t_1,t_2) = \frac{\gamma(t_1,t_2)}{\sqrt{\sigma_{t_1}^2}\sqrt{\sigma_{t_2}^2}}.$$

Для строго стационарного процесса $\mu_t = \mu = \text{const}$ – константа, так как функция распределения одинакова для $\forall\, t$ (при условии $E(|Z_t|) < \infty$). Аналогично, если $E({Z_t}^2) < \infty$, то $\sigma_t^2 = \sigma^2 = \text{const}$

$\forall\, t$. Далее, так как $F(Z_{t_1}, Z_{t_2}) = F(Z_{t_1+k}, Z_{t_2+k})$ для $\forall$ целых t_1, t_2 и k, получаем

$$\gamma(t_1, t_2) = \gamma(t_1 + k, t_2 + k),$$

$$\rho(t_1, t_2) = \rho(t_1 + k, t_2 + k).$$

Положив $t_1 = t - k$ и $t_2 = t$, получаем

$$\gamma(t_1, t_2) = \gamma(t - k, t) = \gamma(t, t + k) = \gamma_k,$$

$$\rho(t_1, t_2) = \rho(t - k, t) = \rho(t, t + k) = \rho_k.$$

Таким образом, для строго стационарных процессов с конечными двумя первыми моментами ковариационная и корреляционная функции зависят только от разницы во времени k между моментами, в которые снимались наблюдения.

Простой пример строго стационарного процесса – последовательность одинаково распределенных независимых случайных переменных, которая редко встречается на практике, но используется для построения модели временного ряда.

Однако на практике очень трудно и часто даже невозможно исследовать функцию распределения процесса, который генерирует временной ряд. Поэтому при анализе временных рядов часто фигурирует стационарность в более слабом смысле, выраженная с использованием моментов процесса. Говорят, что процесс ***слабо стационарен*** n-го порядка, если все его моменты до n-го порядка включительно существуют и инвариантны во времени, т.е. независимы от начала отсчета времени. Таким образом, слабо стационарный процесс второго порядка имеет постоянные среднее и дисперсию, а ковариационная и корреляционная функции являются функциями лишь разницы во времени k между моментами.

Ясно, что строго стационарный процесс с конечными двумя первыми моментами – также слабо стационарный второго порядка. Очевидно также, что обратное неверно, т.е. «слабая стационарность второго порядка» является намного более слабым свойством, чем «строгая стационарность» и даже «стационарность по распределению второго порядка». Однако в силу простоты проверки на практике используется именно слабая стационарность второго порядка. Кроме того, существует класс процессов, для которых понятия слабой и строгой стационарности эквивалентны. Речь идет о *гауссовских процессах*, имеющих

нормальную функцию распределения. Нормальная функция распределения однозначно определяется двумя первыми моментами, что обусловливает в этом случае эквивалентность слабой и строгой стационарности. В дальнейшем будем использовать просто термин «стационарный», предполагая, что процессы, с которыми проводится анализ, являются гауссовскими.

3.2. АВТОКОВАРИАЦИОННАЯ И АВТОКОРРЕЛЯЦИОННАЯ ФУНКЦИИ

Основными инструментами идентификации стационарного временного ряда с использованием подхода во временной области являются автоковариационная и частная автоковариационная функции. Определим эти функции.

3.2.1. АВТОКОВАРИАЦИОННАЯ ФУНКЦИЯ

Предположим, что процесс $\{Z_t\}$ стационарен. Для стационарного гауссовского процесса имеем:

1) $E(Z_t) = \mu = \text{const}$;

2) $\text{Var}(Z_t) = E(Z_t - \mu_t)^2 = \sigma^2 = \text{const}$;

3) $\text{Cov}(Z_t, Z_s) = E(Z_t - \mu_t)(Z_s - \mu_s)$ – функции только модуля разности $|t - s|$.

В этом случае, очевидно, что ковариация между Z_t и Z_{t+k} является функцией только переменной k и не зависит от t:

$$\gamma_k = \text{Cov}(Z_t, Z_{t+k}) = E(Z_t - \mu)(Z_{t+k} - \mu).$$

Ковариационная функция имеет недостатки: с трудом поддается непосредственной интерпретации, так как зависит от единиц измерения Z_t. Поэтому определяется корреляционная функция между Z_t и Z_{t+k}:

$$\rho_k = \frac{\text{Cov}(Z_t, Z_{t+k})}{\sqrt{\text{Var}(Z_t)}\sqrt{\text{Var}(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0},$$

где $\gamma_0 = \text{Var}(Z_t) = \text{Var}(Z_{t+k})$.

В анализе временных рядов функция γ_k называется автоковариационной, а функция ρ_k – автокорреляционной. Такие названия объясняются тем, что они представляют собой ковариацию и корреляцию, измеренную для различных значений Z_t и Z_{t+k} одного и того же временного ряда. Легко убедиться, что для стационарного процесса ρ_k и γ_k удовлетворяют следующим свойствам:

1) $\gamma_0 = Var(Z_t)$, $\rho_0 = 1$;

2) $|\gamma_k| \le \gamma_0$, $|\rho_k| \le 1$;

3) $\gamma_k = \gamma_{-k}$ и $\rho_k = \rho_{-k}$.

Последнее свойство означает, что ρ_k и γ_k – четные функции и, следовательно, симметричны относительно точки $k = 0$. В силу этого обе функции рассматриваются только для неотрицательных значений аргумента.

График автокорреляционной функции (в дальнейшем будем называть ее АКФ) – важное средство идентификации модели стационарного временного ряда. Он строится только для неотрицательных значений дискретного аргумента k, называется коррелограммой, и может иметь вид, показанный на рис. 3.1.

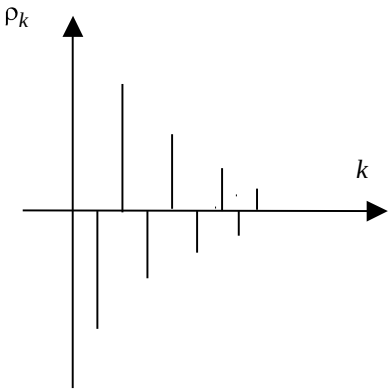


Рис. 3.1. График автокорреляционной функции

Другим свойством ρ_k и γ_k является их положительная полуопределенность, в том смысле, что выполняется

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\gamma_{|t_i-t_j|} \geq 0 \text{ и } \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\rho_{|t_i-t_j|} \geq 0$$

для любых моментов времени $t_1,...,t_n$ и для любых вещественных чисел $\alpha_1,...,\alpha_n$. Доказательство этих результатов очень простое. Определим случайную переменную $X = \sum_{i=1}^{n}\alpha_i Z_{t_i}$ и рассмотрим ее дисперсию:

$$\mathrm{Var}(X) = E\left\{\sum_{i=1}^{n}\alpha_i(Z_{t_i} - \mu)\right\}^2 =$$

$$= E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j(Z_{t_i} - \mu)(Z_{t_j} - \mu)\right\} = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\mathrm{Cov}\left(Z_{t_i}, Z_{t_j}\right).$$

Теперь легко получить первое неравенство:

$$0 \leq \mathrm{Var}(X) = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\mathrm{Cov}\left(Z_{t_i}, Z_{t_j}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\gamma_{|t_i-t_j|}.$$

Второй результат следует из первого неравенства делением на γ_0 ($\gamma_0 > 0$).

Таким образом, важно знать, что не всякая произвольная функция, удовлетворяющая свойствам 1–3, может быть автоковариационной или автокорреляционной. Необходимым условием для того, чтобы функция была автоковариационной или автокорреляционной для некоторого процесса, она должна быть положительно полуопределенной.

3.2.2. ЧАСТНАЯ АВТОКОРРЕЛЯЦИОННАЯ ФУНКЦИЯ

В ряде случаев значение коэффициента корреляции между двумя переменными, хотя и статистически значимое, может оказаться следствием не только некоторой причинной зависимости между этими переменными, но и результатом влияния какой-либо третьей перемен-

ной, которая так или иначе связана с каждой из первых двух. Для устранения влияния третьей искажающей переменной определяется так называемая частная корреляция.

В анализе временных рядов частная автокорреляционная функция (ЧАКФ) наряду с автокорреляционной функцией (АКФ) играет решающую роль в идентификации стационарного временного ряда. ЧАКФ измеряет корреляцию между случайными величинами Z_t и Z_{t+k}, «очищенную» от влияния случайных величин $Z_{t+1}, Z_{t+2},...,$ Z_{t+k-1}, находящихся между ними. Фактически ЧАКФ есть условная корреляция $\mathrm{Corr}(Z_t, Z_{t+k} \big| Z_{t+1}, Z_{t+2},..., Z_{t+k-1})$.

Рассмотрим два способа получения формулы для вычисления ЧАКФ. Пусть $\{Z_t\}$ – стационарный процесс. Предположим без потери общности, что $E(Z_t) = 0$. Определим $\hat{Z}_{t+k}$ – наилучшую линейную оценку Z_{t+k} в среднеквадратичном смысле как линейную функцию от $Z_{t+1}, Z_{t+2},...,Z_{t+k-1}$. То есть если $\hat{Z}_{t+k}$ есть наилучшая линейная оценка Z_{t+k}, то

$$\hat{Z}_{t+k} = \alpha_1 Z_{t+k-1} + \alpha_2 Z_{t+k-2} + ... + \alpha_{k-1} Z_{t+1},$$

где α_i $(1 \le i \le k-1)$ – коэффициенты, полученные в результате минимизации

$$E(Z_{t+k} - \hat{Z}_{t+k})^2 = E\left(Z_{t+k} - \alpha_1 Z_{t+k-1} - ... - \alpha_{k-1} Z_{t+1}\right)^2.$$

После применения процедуры дифференцирования и приравнивания производных к нулю получим следующую линейную систему уравнений:

$$\gamma_i = \alpha_1 \gamma_{i-1} + \alpha_2 \gamma_{i-2} + ... + \alpha_{k-1} \gamma_{i-k+1} \ \left(1 \le i \le k-1\right).$$

Разделив на γ_0, получаем

$$\rho_i = \alpha_1 \rho_{i-1} + \alpha_2 \rho_{i-2} + ... + \alpha_{k-1} \rho_{i-k+1} \ \left(1 \le i \le k-1\right).$$

Запишем последнюю систему для $1 \le i \le k-1$ в матричных обозначениях:

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{k-1} \end{bmatrix}.$$

Аналогично рассмотрим наилучшую линейную оценку

$$\hat{Z}_t = \beta_1 Z_{t+1} + \beta_2 Z_{t+2} + \ldots + \beta_{k-1} Z_{t+k-1},$$

где $\beta_i \left(1 \le i \le k-1\right)$ – коэффициенты, полученные в результате минимизации:

$$E(Z_t - \hat{Z}_t)^2 = E\left(Z_t - \beta_1 Z_{t+1} - \ldots - \beta_{k-1} Z_{t+k-1}\right)^2.$$

После дифференцирования по β_i и приравнивания производных к нулю получаем

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{k-1} \end{bmatrix}. \qquad (3.2)$$

Откуда следует $\alpha_i = \beta_i \left(1 \le i \le k-1\right)$.

Теперь логично определить частную автокорреляцию между Z_t и Z_{t+k} как корреляцию между $\left(Z_t - \hat{Z}_t\right)$ и $\left(Z_{t+k} - \hat{Z}_{t+k}\right)$:

$$P_k = \frac{\mathrm{Cov}\left[\left(Z_t - \hat{Z}_t\right), \left(Z_{t+k} - \hat{Z}_{t+k}\right)\right]}{\sqrt{\mathrm{Var}\left(Z_t - \hat{Z}_t\right)}\sqrt{\mathrm{Var}\left(Z_{t+k} - \hat{Z}_{t+k}\right)}}.$$

Распишем

$$\mathrm{Var}\left(Z_{t+k}-\hat{Z}_{t+k}\right)=E\left[\left(Z_{t+k}-\alpha_1 Z_{t+k-1}-\ldots-\alpha_{k-1}Z_{t+1}\right)^2\right]=$$

$$=E\left[Z_{t+k}\left(Z_{t+k}-\alpha_1 Z_{t+k-1}-\ldots-\alpha_{k-1}Z_{t+1}\right)\right]-$$

$$-\alpha_1 E\left[Z_{t+k-1}\left(Z_{t+k}-\alpha_1 Z_{t+k-1}-\ldots-\alpha_{k-1}Z_{t+1}\right)\right]-$$

$$\ldots-\alpha_{k-1}E\left[Z_{t+1}\left(Z_{t+k}-\alpha_1 Z_{t+k-1}-\ldots-\alpha_{k-1}Z_{t+1}\right)\right]=$$

$$=E\left[Z_{t+k}\left(Z_{t+k}-\alpha_1 Z_{t+k-1}-\ldots-\alpha_{k-1}Z_{t+1}\right)\right],$$

так как остальные члены равны 0.

Следовательно,

$$\mathrm{Var}\left(Z_{t+k}-\hat{Z}_{t+k}\right)=\mathrm{Var}\left(Z_t-\hat{Z}_t\right)=\gamma_0-\alpha_1\gamma_1-\ldots-\alpha_{k-1}\gamma_{k-1}.$$

Далее, используя факт, что $\alpha_i=\beta_i\ (1\le i\le k-1)$, получаем

$$\mathrm{Cov}\left[\left(Z_t-\hat{Z}_t\right),\left(Z_{t+k}-\hat{Z}_{t+k}\right)\right]=$$

$$=E\left[\ \left(Z_t-\alpha_1 Z_{t+1}-\ldots-\alpha_{k-1}Z_{t+k-1}\right)\times\right.$$

$$\times\left(Z_{t+k}-\alpha_1 Z_{t+k-1}-\ldots-\alpha_{k-1}Z_{t+1}\right)=$$

$$=E\left[\left(Z_t-\alpha_1 Z_{t+1}-\ldots-\alpha_{k-1}Z_{t+k-1}\right)Z_{t+k}\right]=$$

$$=\gamma_k-\alpha_1\gamma_{k-1}-\ldots-\alpha_{k-1}\gamma_1.$$

Таким образом,

$$P_k=\frac{\gamma_k-\alpha_1\gamma_{k-1}-\ldots-\alpha_{k-1}\gamma_1}{\gamma_0-\alpha_1\gamma_1-\ldots-\alpha_{k-1}\gamma_{k-1}}=$$

$$=\frac{\rho_k-\alpha_1\rho_{k-1}-\ldots-\alpha_{k-1}\rho_1}{1-\alpha_1\rho_1-\ldots-\alpha_{k-1}\rho_{k-1}}. \qquad (3.3)$$

Теперь найдем решение α_i системы (3.2), воспользовавшись правилом Крамера:

$$\alpha_i = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{i-2} & \rho_1 & \rho_i & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \cdots & \rho_{i-3} & \rho_2 & \rho_{i-1} & \cdots & \rho_{k-3} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \cdots & \rho_{k-i} & \rho_{k-1} & \rho_{k-i-2} & \cdots & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{i-2} & \rho_{i-1} & \rho_i & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \cdots & \rho_{i-3} & \rho_{i-2} & \rho_{i-1} & \cdots & \rho_{k-3} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \cdots & \rho_{k-i} & \rho_{k-i-1} & \rho_{k-i-2} & \cdots & 1 \end{vmatrix}}.$$

Подставляя α_i в выражение (3.3) и умножая в нем числитель и знаменатель на определитель

$$\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{vmatrix},$$

получаем выражение для частной автокорреляционной функции:

$$P_k = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}}. \qquad (3.4)$$

Второй способ вывода формулы для вычисления ЧАКФ заключается в следующем. Рассмотрим регрессионную модель с зависимой переменной Z_{t+k} (случайная переменная из стационарного процесса с нулевым средним) и k лаговыми переменными $Z_{t+k-1}, Z_{t+k-2},..., Z_t$, используемыми в качестве независимых переменных:

$$Z_{t+k} = \Phi_{k1}Z_{t+k-1} + \Phi_{k2}Z_{t+k-2} + \cdots + \Phi_{kk}Z_t + e_{t+k},$$

где Φ_{ki} – i-й регрессионный параметр, e_{t+k} – случайная ошибка, распределенная по нормальному закону, некоррелированная с Z_{t+k-j} для $j \geq 1$. Умножив на Z_{t+k-j} обе части данного регрессионного уравнения и применив оператор математического ожидания, получаем

$$\gamma_j = \Phi_{k1}\gamma_{j-1} + \Phi_{k2}\gamma_{j-2} + \cdots + \Phi_{kk}\gamma_{j-k},$$

откуда

$$\rho_j = \Phi_{k1}\rho_{j-1} + \Phi_{k2}\rho_{j-2} + \cdots + \Phi_{kk}\rho_{j-k}.$$

Расписывая последнее уравнение для $j = 1, 2, \ldots, k$, получим систему уравнений

$$\begin{aligned}
\rho_1 &= \Phi_{k1}\rho_0 + \Phi_{k2}\rho_1 + \cdots + \Phi_{kk}\rho_{k-1}, \\
\rho_2 &= \Phi_{k1}\rho_1 + \Phi_{k2}\rho_0 + \cdots + \Phi_{kk}\rho_{k-2}, \\
&\vdots \\
\rho_k &= \Phi_{k1}\rho_{k-1} + \Phi_{k2}\rho_{k-2} + \cdots + \Phi_{kk}\rho_0.
\end{aligned}$$

Решая полученную систему уравнений относительно $\Phi_{k1}, \Phi_{k2},...,$ Φ_{kk} по правилу Крамера, получаем

$$\Phi_{11} = \rho_1, \quad \Phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}},$$

$$\Phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}, \dots,$$

$$\Phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_1 \end{vmatrix}}. \qquad (3.5)$$

Сравнивая выражение (3.4) с выражением (3.5), видим, что Φ_{kk} равно P_k. Таким образом, частная автокорреляция между Z_t и Z_{t+k} может также быть получена как коэффициент регрессии при Z_t, если мы рассматриваем регрессию Z_{t+k} по k запаздывающим переменным $Z_{t+k-1}, Z_{t+k-2}, \dots, Z_{t+k}$.

В литературе принято стандартное обозначение Φ_{kk} для ЧАКФ между Z_t и Z_{t+k}.

3.3. ПРОЦЕСС БЕЛОГО ШУМА

Процесс $\{a_t\}$ называется процессом белого шума, если он является последовательностью некоррелированных случайных переменных, одинаково распределенных с постоянным средним $E(a_t) = \mu_a$, обычно предполагаемым равным нулю, с постоянными дисперсией

$\mathrm{Var}(a_t) = \sigma_a^2$ и ковариацией $\gamma_k = \mathrm{Cov}(a_t, a_{t+k}) = 0$ для любого $k \neq 0$.

Из определения следует, что процесс белого шума $\{a_t\}$ является стационарным и имеет автоковариационную функцию

$$\gamma_k = \begin{cases} \sigma_a^2, & k = 0, \\ 0, & k \neq 0, \end{cases}$$

автокорреляционную функцию – АКФ

$$\rho_k = \begin{cases} 1, & k = 0, \\ 0, & k \neq 0, \end{cases}$$

и частную автоковариационную функцию – ЧАКФ

$$\Phi_{kk} = \begin{cases} 1, & k = 0, \\ 0, & k \neq 0. \end{cases}$$

Коррелограммы АКФ и ЧАКФ для процесса белого шума $Z_t = \mu + a_t$ ($\mu = \mathrm{const}$) представлены на рис. 3.2.

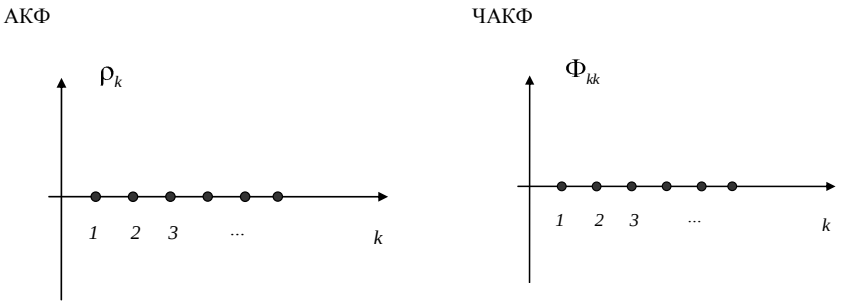


Рис. 3.2. Коррелограммы теоретических АКФ и ЧАКФ для процесса белого шума

Так как по определению $\rho_0 = \Phi_{00} = 1$ для любого процесса, то в дальнейшем будем рассматривать коррелограммы АКФ и ЧАКФ только при $k \neq 0$. Основной идентификационной характеристикой процессов белого шума является то, что АКФ и ЧАКФ для них тождественно равны нулю.

Хотя процессы белого шума в чистом виде едва ли встречаются на практике, они играют важную роль как базовые строительные блоки при построении моделей временных рядов, аналогичную роли функций sin и cos в анализе Фурье $\{\sin(nx), \cos(nx)\}$. Другими словами, процессы белого шума играют роль ортогонального базиса в общем векторном и функциональном анализе.

Процесс белого шума называется гауссовским, если его совместное распределение является нормальным. В дальнейшем будем полагать, что $\{a_t\}$ – процесс белого гауссовского шума с нулевым средним.

3.4. ОЦЕНИВАНИЕ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ, АВТОКОВАРИАЦИЙ И АВТОКОРРЕЛЯЦИЙ

Стационарный временной ряд характеризуется своим математическим ожиданием μ, дисперсией σ^2, автокорреляциями ρ_k и частными автокорреляциями Φ_{kk}. Точные значения этих параметров можно вычислить только в том случае, если известен ансамбль всех возможных реализаций. Если наблюдению доступно некоторое подмножество независимых реализаций, то можно получить оценки значений этих параметров. Однако в большинстве практических приложений трудно или даже невозможно получить даже множество реализаций. Чаще всего нам известна лишь одна реализация – конкретный временной ряд, что делает невозможным вычисление среднего по ансамблю траекторий. Однако для стационарных временных рядов имеется естественная альтернатива при определенных условиях заменить среднее по ансамблю средним по времени.

3.4.1. ВЫБОРОЧНОЕ СРЕДНЕЕ

При одной реализации естественной оценкой математического ожидания $\mu = E(Z_t)$ стационарного процесса является выборочное среднее – среднее из n наблюдений по времени:

$$\bar{Z} = \frac{1}{n}\sum_{t=1}^{n} Z_t .$$

Нетрудно показать [12], что выборочное среднее – это *несмещенная оценка* μ, т.е. $E(\bar{Z}) = \mu$. Важным асимптотическим свойством стационарного процесса является его *эргодичность* по отношению к среднему:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} Z_t = \mu,$$

т.е. сходимость к μ в среднеквадратическом. Эргодичность – важное условие, позволяющее использовать выборочные аналоги генеральных характеристик ряда для определения его свойств. Стационарные процессы не всегда эргодичны. Достаточным условием эргодичности является

$$\rho_k \xrightarrow[k\to\infty]{} 0, \tag{3.6}$$

из которого следует условие $\sum_{k=0}^{\infty} |\rho_k| < \infty$, из которого в свою очередь следует, что $\lim_{n\to\infty} \mathrm{Var}(\bar{Z}) = 0$.

Таким образом, при выполнении достаточного условия эргодичности (3.6) выборочное среднее $\bar{Z}$ является *состоятельной оценкой* μ.

3.4.2. ВЫБОРОЧНЫЕ АВТОКОВАРИАЦИОННАЯ И АВТОКОРРЕЛЯЦИОННАЯ ФУНКЦИИ

Располагая единственной реализацией случайного процесса, можно использовать усреднение по времени для оценивания автоковариационной функции γ_k:

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z})$$

или

$$\hat{\hat{\gamma}}_k = \frac{1}{n-k}\sum_{t=1}^{n-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z}).$$

Обе эти оценки смещенные, так как справедливы следующие приближенные равенства:

$$E(\hat{\gamma}_k) \simeq \gamma_k - \frac{k}{n}\gamma_k - \frac{n-k}{n}\mathrm{Var}(\bar{Z}) \text{ и } E(\hat{\gamma}_k) \simeq \gamma_k - \mathrm{Var}(\bar{Z}).$$

В общем случае оценка $\hat{\gamma}_k$ имеет большее смещение, чем $\hat{\hat{\gamma}}_k$, особенно если k велико относительно n. Последнее обстоятельство – причина того, что в анализе временных рядов предполагается вычисление оценок автоковариаций и автокорреляций для значений k не больше $n/4$.

При выполнении достаточного условия эргодичности по отношению к среднему (3.6) обе оценки $\hat{\gamma}_k$ и $\hat{\hat{\gamma}}_k$ являются асимптотически несмещенными. Однако на практике чаще используется оценка $\hat{\gamma}_k$, так как для определенных типов процессов она имеет меньшую среднеквадратичную ошибку, чем $\hat{\hat{\gamma}}_k$, и всегда является положительно полуопределенной, как и γ_k.

Выборочная автокорреляционная функция (выборочная АКФ) вычисляется на основе автоковариационной функции по формуле

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k}(Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^{n}(Z_t - \bar{Z})^2}. \tag{3.7}$$

Для процессов, в которых $\rho_k = 0$ для $k > m$, известна следующая аппроксимация Бартлетта:

$$\mathrm{Var}(\hat{\rho}_k) \simeq \frac{1}{n}(1 + 2\rho_1^2 + 2\rho_2^2 + ... + 2\rho_m^2).$$

На практике ρ_i, $i = 1, 2, \ldots, m$, неизвестны и заменяются выборочными оценками $\hat{\rho}_i$. Получаем следующую формулу для стандартной ошибки оценки $\hat{\rho}_k$:

$$S_{\hat{\rho}_k} \simeq \sqrt{\frac{1}{n}(1 + 2\hat{\rho}_1^2 + 2\hat{\rho}_2^2 + ... + 2\hat{\rho}_m^2)}\,.$$

Для процесса белого шума ($m = 0$) можно использовать формулу ошибки

$$S_{\hat{\rho}_k} = \sqrt{\frac{1}{n}}\,.$$

Наконец заметим, что выборочная АКФ $\hat{\rho}_k$ симметрична относительно начала $k = 0$, как и истинная АКФ ρ_k :

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k}\left(Z_t - \overline{Z}\right)\left(Z_{t+k} - \overline{Z}\right)}{\sum_{t=1}^{n}\left(Z_t - \overline{Z}\right)^2} = \frac{\sum_{t=k+1}^{n}\left(Z_t - \overline{Z}\right)\left(Z_{t-k} - \overline{Z}\right)}{\sum_{t=1}^{n}\left(Z_t - \overline{Z}\right)^2} = \hat{\rho}_{-k}\,.$$

Пример 1. Для иллюстрации вычисления выборочной АКФ рассмотрим следующие 10 значений временного ряда:

t	Z_t	Z_{t+1}	Z_{t+2}	Z_{t+3}	$\cdots$	Z_{t-1}	Z_{t-2}
1	13	8	15	4			
2	8	15	4	4		13	
3	15	4	4	12		8	13
4	4	4	12	11		15	8
5	4	12	11	7		4	15
6	12	11	7	14		4	4
7	11	7	14	12		12	4
8	7	14	12			11	12
9	14	12				7	11
10	12					14	7

Выборочное среднее этих 10 значений $\overline{Z} = 10$.

$$\hat{\rho}_2 = \frac{(13-10)(15-10)+(8-10)(4-10)+...+}{}$$

$$\frac{+(11-10)(14-10)+(7-10)(12-10)}{144} = \frac{-29}{144} = -0,201;$$

$$\hat{\rho}_3 = \frac{(13-10)(4-10)+(8-10)(4-10)+...+}{}$$

$$\frac{+(12-10)(14-10)+(11-10)(12-10)}{144} = \frac{26}{144} = 0,181.$$

3.4.3. ВЫБОРОЧНАЯ ЧАСТНАЯ АВТОКОРРЕЛЯЦИОННАЯ ФУНКЦИЯ

Выборочная ЧАКФ $\hat{\Phi}_{kk}$ может быть получена в результате подстановки $\hat{\rho}_i$ вместо ρ_i в формулу (3.5). В 1960 году Дж. Дарбин предложил рекурсивный метод для вычисления $\hat{\Phi}_{kk}$, позволяющий избежать сложного вычисления определителей матриц больших размерностей. Метод заключается в последовательном вычислении $\hat{\Phi}_{kk}$ по следующим формулам:

$$\hat{\Phi}_{11} = \hat{\rho}_1,$$

$$\hat{\Phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^{k} \hat{\Phi}_{kj}\hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^{k} \hat{\Phi}_{kj}\hat{\rho}_j}, \qquad (3.8)$$

$$\hat{\Phi}_{k+1,j} = \hat{\Phi}_{kj} - \hat{\Phi}_{k+1,k+1}\hat{\Phi}_{k,k+1-j}, \; j = 1,\ldots,k.$$

Этот метод справедлив и для вычисления теоретической ЧАКФ Φ_{kk} как функции ρ_i.

В предположении, что процесс является последовательностью белого шума, дисперсия $\widehat{\Phi}_{kk}$ может быть аппроксимирована как

$\mathrm{Var}(\hat{\Phi}_{kk}) \simeq \dfrac{1}{n}$. Следовательно, величины $\pm 2/\sqrt{n}$ можно использовать в качестве критических границ для Φ_{kk} при проверке гипотезы о том, что процесс является белым шумом.

Пример 2. Применим рекурсивный метод Дарбина к данным предыдущего примера 1:

$$\widehat{\Phi}_{11} = \hat{\rho}_1 = -0{,}188;$$

$$\widehat{\Phi}_{22} = \frac{\hat{\rho}_2 - \hat{\rho}_1^{\,2}}{1 - \hat{p}_1^{\,2}} = \frac{-0{,}201 - (-0{,}188)^2}{1 - (-0{,}188)^2} = -0{,}245\,;$$

$$\widehat{\Phi}_{21} = \widehat{\Phi}_{11} - \widehat{\Phi}_{22}\widehat{\Phi}_{11} = (-0{,}188) - (-0{,}245)(-0{,}188) = -0{,}234\,;$$

$$\widehat{\Phi}_{33} = \frac{\hat{\rho}_3 - \widehat{\Phi}_{21}\hat{\rho}_2 - \widehat{\Phi}_{22}\hat{\rho}_1}{1 - \widehat{\Phi}_{21}\hat{\rho}_1 - \widehat{\Phi}_{22}\hat{\rho}_2} =$$

$$= \frac{0{,}181 - (0{,}234)(-0{,}201) - (-0{,}245)(-0{,}188)}{1 - (0{,}234)(-0{,}188) - (-0{,}245)(-0{,}201)} = \frac{0{,}088}{0{,}907} = 0{,}097.$$

Остальные $\widehat{\Phi}_{kk}$ вычисляются аналогично.

3.5. ТЕОРЕТИЧЕСКИЕ ПРЕДСТАВЛЕНИЯ ПРОЦЕССОВ ВРЕМЕННЫХ РЯДОВ

3.5.1. ПРЕДСТАВЛЕНИЕ ПРОЦЕССОВ С ПОМОЩЬЮ МОДЕЛИ СКОЛЬЗЯЩЕГО СРЕДНЕГО

В анализе временных рядов существуют два полезных представления процессов. В первом случае процесс Z_t представляется в виде линейной комбинации последовательности некоррелированных случайных переменных, т.е.

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j}\,,$$

где $\psi_0 = 1$, $\{a_t\}$ – процесс белого гауссовского шума с нулевым средним и дисперсией σ_a^2, причем $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Данное представление называется представлением скользящего среднего (Moving Average), или МА-представлением.

Здесь и далее бесконечная сумма случайных переменных определяется как предел конечных частичных сумм в среднеквадратичном. Таким образом, Z_t в МА-представлении определяется таким образом, что

$$E\left[\left(\dot{Z}_t - \sum_{j=0}^{n} \psi_j a_{t-j}\right)^2\right] \to 0 \text{ при } n \to \infty,$$

где $\dot{Z}_t = Z_t - \mu$ – процесс с нулевым средним, полученный из Z_t.

Введем лаговый оператор (оператор запаздывания) B, который определяется следующим соотношением $B^j x_t = x_{t-j}$. Теперь можно записать МА-представление в следующей компактной операторной форме:

$$\dot{Z}_t = \psi(B) a_t,$$

где $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$.

Легко показать, что для процесса в МА-представлении выполняются следующие свойства:

1) $E(Z_t) = \mu$;

2) $\mathrm{Var}(Z_t) = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2 = \mathrm{const} < \infty$, так как по предположению $\sum_{j=0}^{\infty} \psi_j^2 < \infty$;

3) $\gamma_k = E\left(\dot{Z}_t \dot{Z}_{t+k}\right) = E\left(\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i\psi_j a_{t-i}a_{t+k-j}\right) = \sigma_a^2\sum_{i=0}^{\infty}\psi_i\psi_{i+k}$;

4) $\rho_k = \dfrac{\gamma_k}{\gamma_0} = \dfrac{\sum_{i=0}^{\infty}\psi_i\psi_{i+k}}{\sum_{i=0}^{\infty}\psi_i^2}$.

При выводе свойства 3 использовалась замена $j = i + k$ и свойство некоррелированности текущих ошибок с прошлыми значениями процесса:

$$E\left(a_t Z_{t-j}\right) = \begin{cases}\sigma_a^2, & \text{если} \quad j = 0,\\ 0, & \text{если} \quad j > 0.\end{cases}$$

Условие $\sum_{j=0}^{\infty}\psi_j^2 < \infty$ является достаточным для конечности γ_k и ρ_k, как и для конечности дисперсии процесса. Следовательно, оно является достаточным и для того, чтобы процесс, описываемый МА-моделью, был стационарным в силу независимости первых и вторых моментов от t (свойства 1–4).

Известно, что чисто недетерминированный стационарный процесс (т.е. процесс, не содержащий детерминированной составляющей, которую можно предсказать точно по ее прошлым значениям) всегда может быть представлен в виде МА-модели. Еще одно название недетерминированного процесса – линейный процесс.

Для вычисления автоковариаций МА-процесса полезное средство – так называемая *автоковариационная генерирующая функция*, определяемая следующим образом:

$$\gamma(B) = \sum_{k=-\infty}^{\infty}\gamma_k B^k ,$$

где дисперсия процесса γ_0 является коэффициентом при B^0, а автоковариация для лага k есть коэффициент при B^k и B^{-k}. Подставляя выражение $\gamma_k = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}$ в формулу для $\gamma(B)$, получим

$$\gamma(B) = \sigma_a^2 \sum_{k=-\infty}^{\infty} \sum_{i=0}^{\infty} \psi_i \psi_{i+k} B^k = \sigma_a^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j B^{j-i} =$$

$$= \sigma_a^2 \sum_{j=0}^{\infty} \psi_j B^j \sum_{i=0}^{\infty} \psi_i B^{-i} = \sigma_a^2 \psi(B) \psi(B^{-1}) .$$

Для вычислений автоковариаций для конкретного процесса необходимо подставить полином модели $\psi(B)$ в выражение $\gamma(B) = \sigma_a^2 \psi(B) \psi(B^{-1})$, перемножить полиномы и произвести группировку по степеням B. Коэффициенты при B^k и будут искомыми автоковариациями.

Соответствующая *автокорреляционная генерирующая функция* имеет вид

$$\rho(B) = \sum_{k=-\infty}^{\infty} \rho_k B^k = \frac{\gamma(B)}{\gamma_0} .$$

3.5.2 ПРЕДСТАВЛЕНИЕ ПРОЦЕССОВ С ПОМОЩЬЮ МОДЕЛИ АВТОРЕГРЕССИИ

Другая удобная форма записи процесса Z_t – авторегрессионое представление, или AR-представление (AutoRegression), в котором текущее значение Z в некоторый момент времени t задается функцией от прошлых значений самой этой переменной, т.е. $\dot{Z}_t = \pi_1 \dot{Z}_{t-1} + \pi_2 \dot{Z}_{t-2} + ... + a_t$, или, с использованием лагового оператора, $\pi(B)\dot{Z}_t = a_t$, где $\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$, причем $1 + \sum_{j=1}^{\infty} \left|\pi_j\right| < \infty$.

AR-представление полезно в понимании механизма предсказания. Процесс в AR-представлении называется обратимым. По мнению

Дж. Бокса и Г. Дженкинса [3], при решении задачи прогнозирования использование необратимого процесса не имеет смысла.

Не всякий стационарный процесс обратим. Для того чтобы линейный процесс в MA-представлении $Z_t = \psi(B)a_t$ был обратим (т.е. чтобы он мог быть записан в AR-представлении), корни уравнения $\psi(B) = 0$ должны лежать вне единичного круга. То есть если β – корень уравнения $\psi(B) = 0$, то $|\beta| > 1$, где $|\ |$ – стандартная евклидова метрика. Если β – вещественное число, то $|\beta|$ есть абсолютное значение β, если β – комплексное число $\beta = c + id$, то $|\beta| = \sqrt{c^2 + d^2}$.

Аналогично, не всякий обратимый процесс – стационарный. Для того чтобы процесс в AR-представлении был стационарным (он должен допускать запись в MA-представлении, т.е. $\dot{Z}_t = \frac{1}{\pi(B)}a_t = \psi(B)a_t$), корни уравнения $\pi(B) = 0$ должны лежать вне единичного круга.

4. МОДЕЛИ СТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ

Хотя теоретические AR- и MA-представления полезны, они не являются формами модели, которые могут быть использованы на первоначальных стадиях моделирования, так как содержат бесконечное число параметров, которое невозможно оценить, располагая конечным числом доступных наблюдений. Поэтому при моделировании явления мы конструируем аналоги рассмотренных теоретических моделей с конечным числом параметров.

4.1. МОДЕЛИ СТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ КОНЕЧНОГО ПОРЯДКА

Если в AR-представлении процесса лишь конечное число коэффициентов π является ненулевым, т.е. $\pi_1 = \phi_1, \pi_2 = \phi_2, ..., \pi_p = \phi_p$ и $\pi_k = 0$ для $k > p$, то получившийся процесс называют авторегрессионным порядка p (AR(p)), а модель записывается следующим образом:

$$Z_t - \phi_1 Z_{t-1} - ... - \phi_p Z_{t-p} = a_t .$$

Аналогично, если в МА-представлении лишь конечное число коэффициентов ψ является ненулевым, т.е. $\psi_1 = -\theta_1$, $\psi_2 = -\theta_2$,..., $\psi_q = -\theta_q$ и $\psi_k = 0$ для $k > q$, то получившийся процесс называют процессом скользящего среднего порядка q (МА(q)), а модель записывается следующим образом:

$$Z_t = a_t - \theta_1 a_{t-1} - ... - \theta_q a_{t-q} .$$

Однако если мы ограничимся для представления процессов только моделями AR(p) и МА(q), то число параметров в этих моделях может оказаться чрезмерно большим для получения приемлемой по точности модели. Естественное решение в такой ситуации – смешанная авторегрессионая модель со скользящим средним (ARMA(p, q)):

$$Z_t - \phi_1 Z_{t-1} - \phi_2 Z_{t-2} - ... - \phi_p Z_{t-p} = a_t - \theta_1 a_{t-1} - ... - \theta_q a_{t-q} .$$

Для фиксированного числа наблюдений, чем больше параметров в модели, тем меньше эффективность оценивания параметров. При прочих равных условиях выбирают более простую модель для описания явления. Это принцип носит название принципа экономии в построении моделей.

Модель ARMA(p, q), включающая как частные случаи модели AR(p), МА(q), описывает очень широкий класс экономичных процессов, полезных при анализе широкого разнообразия временных рядов.

4.2. МОДЕЛИ ПРОЦЕССОВ АВТОРЕГРЕССИИ

Как ранее было определено, модель авторегрессии порядка p имеет вид

$$\dot{Z}_t - \phi_1 \dot{Z}_{t-1} - ... - \phi_p \dot{Z}_{t-p} = a_t ,$$

или, с использованием лагового оператора, в операторной форме

$$\phi_p(B)\dot{Z}_t = a_t ,$$

где $\phi_p(B) = (1 - \phi_1 B - \ldots - \phi_p B^p)$.

Так как $\sum_{j=1}^{\infty}\left|\pi_j\right| = \sum_{j=1}^{p}\left|\phi_j\right| < \infty$, то процесс всегда обратим. Для того чтобы процесс был стационарным, т.е. допускал MA-представление, корни многочлена $\phi_p(B) = 0$ должны лежать вне единичного круга. Дж. Юл в 1927 году [15] использовал AR-модель для описания количества солнечных пятен и поведения простого маятника.

4.2.1. МОДЕЛЬ AR(1) – АВТОРЕГРЕССИЯ ПЕРВОГО ПОРЯДКА

Рассмотрим модель авторегрессии первого порядка:

$$(1-\phi_1 B)\dot{Z}_t = a_t, \text{ или } \dot{Z}_t - \phi_1\dot{Z}_{t-1} = a_t.$$

Как и все процессы авторегрессии конечного порядка, этот процесс всегда обратим. Для того чтобы он был стационарен, корень уравнения $1-\phi_1 B = 0$ должен лежать вне единичного круга. Таким образом, условием стационарности процесса AR(1) является $\left|\phi_1\right| < 1$. Процесс AR(1) иногда называют марковским процессом, так как значение $\dot{Z}_t$ полностью определяется значением $\dot{Z}_{t-1}$.

4.2.1.1. ТЕОРЕТИЧЕСКАЯ АВТОКОРРЕЛЯЦИОННАЯ ФУНКЦИЯ ДЛЯ ПРОЦЕССА AR(1)

Умножив обе части уравнения модели AR(1) на $\dot{Z}_{t-k}$ и применив оператор математического ожидания, получаем уравнение для нахождения автоковариации

$$E(\dot{Z}_{t-k}\dot{Z}_t) = E(\phi_1\dot{Z}_{t-k}\dot{Z}_{t-1}) + E(\dot{Z}_{t-k}a_t),$$

т.е. $\gamma_k = \phi_1\gamma_{k-1}$ для $k \geq 1$.

Отсюда теоретическая автокорреляционная функция (АКФ) определяется уравнением $\rho_k = \phi_1\rho_{k-1} = \phi_1^k$ для $k \geq 1$, учитывая начальное условие $\rho_0 = 1$.

Из полученного уравнения для АКФ вытекает, что для стационарного процесса AR(1), удовлетворяющего условию $\left|\phi_1\right| < 1$, АКФ экспоненциально уменьшается одним из двух способов в зависимости от

знака коэффициента ϕ_1. Если $0 < \phi_1 < 1$, все автокорреляции положительны, если $-1 < \phi_1 < 0$, то знаки автокорреляции чередуются, при этом абсолютная величина АКФ уменьшается в обоих случаях по экспоненте.

4.2.1.2. ЧАСТНАЯ АВТОКОРРЕЛЯЦИОННАЯ ФУНКЦИЯ ДЛЯ ПРОЦЕССА AR(1)

Выражение теоретической ЧАКФ для процесса AR(1) можно получить из формулы (3.5) для вычисления Φ_{kk}:

$$\Phi_{kk} = \begin{cases} \rho_1 = \phi_1, & \text{если } k = 1, \\ 0, & \text{если } k \geq 2. \end{cases}$$

Таким образом, для идентификации процесса AR(1) мы имеем два возможных набора коррелограмм (АКФ и ЧАКФ), изображенных на рис. 4.1.

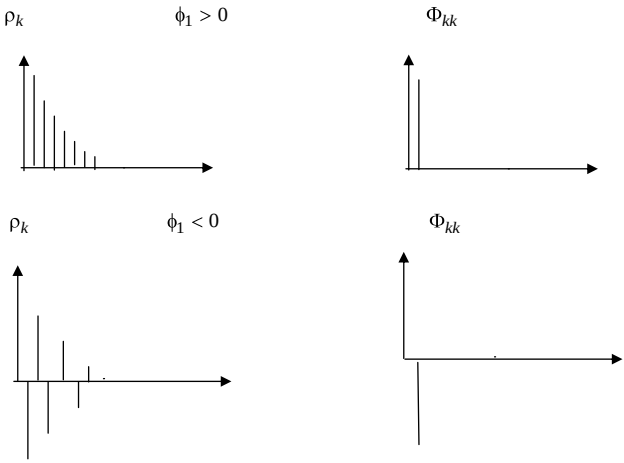


Рис. 4.1. Коррелограммы теоретических АКФ (ρ_k) и ЧАКФ (Φ_{kk}) для процесса AR(1)

Отметим, что коррелограммы выборочных АКФ и ЧАКФ, получаемые по формулам (3.7) и (3.8) по одной реализации случайного процесса – временному ряду, могут существенно отличаться от своих

теоретических аналогов из-за ошибки при получении оценок. Так, положительная экспоненциальная зависимость может быть значительно искажена и даже уходить в отрицательную область после некоторого лага, а строгое чередование знаков автокорреляций – нарушаться. Выборочная ЧАКФ после первого лага (при $k \geq 2$) принимает малые, но не обязательно равные нулю значения. В этом случае считается, что значения статистически незначимы (равны нулю), если они лежат внутри коридора $\pm 1{,}5$ или $2\,S_{\hat{\rho}_k}$ для АКФ и $\pm 1{,}5$ или $2\,S_{\hat{\Phi}_{kk}}$ для ЧАКФ соответственно.

Стандартная ошибка выборочной АКФ $\hat{\rho}_k$ вычисляется по формуле

$$S_{\hat{\rho}_k} \simeq \sqrt{\frac{1}{n}(1 + 2\hat{\rho}_1^2 + ... + 2\hat{\rho}_{k-1}^2)}\,. \qquad (4.1)$$

Стандартная ошибка выборочной ЧАКФ $\hat{\Phi}_{kk}$ устанавливается равной

$$S_{\hat{\Phi}_{kk}} \simeq \sqrt{\frac{1}{n}}\,. \qquad (4.2)$$

Формулы (4.1) и (4.2) для вычисления стандартных ошибок используются в большинстве пакетов статистического анализа данных.

4.2.2. МОДЕЛЬ AR(2) – АВТОРЕГРЕССИЯ ВТОРОГО ПОРЯДКА

Рассмотрим модель авторегрессии второго порядка:

$$(1 - \phi_1 B - \phi_2 B^2)\dot{Z}_t = a_t\,, \text{ или } \dot{Z}_t - \phi_1 \dot{Z}_{t-1} - \phi_2 \dot{Z}_{t-2} = a_t\,.$$

Как и все процессы авторегрессии конечного порядка, этот процесс всегда обратим. Для того чтобы он был стационарен, корни уравнения $\dot{Z}_t - \phi_1 \dot{Z}_{t-1} - \phi_2 \dot{Z}_{t-2} = a_t$ должны лежать вне единичного круга. Например, в модели

$$(1 - 1{,}5B + 0{,}56B^2)\dot{Z}_t = (1 - 0{,}7B)(1 - 0{,}8B) = a_t$$

корнями уравнения $1-\phi_1 B-\phi_2 B^2=0$ являются $B_1=1/0{,}7$ и $B_2=1/0{,}8$. Оба корня больше единицы по абсолютному значению, следовательно, процесс нестационарный.

В модели $(1-0{,}2B-0{,}8B^2)\dot{Z}_t=a_t$ один из корней $B_1=1$ лежит на единичном круге, другой $B_2=0{,}8$ меньше единицы по абсолютному значению. Такой процесс называется однородно нестационарным и будет рассматриваться нами дальше.

Более удобно иметь условие стационарности, выраженное не посредством корней полинома, а с использованием значений параметров. Выразим условие стационарности для модели AR(2) через значения параметров ϕ_1 и ϕ_2.

Пусть B_1 и B_2 – корни уравнения $1-\phi_1 B-\phi_2 B^2=0$, или, что эквивалентно, уравнения $\phi_2 B^2+\phi_1 B-1=0$. Легко вычислить корни данного квадратного уравнения:

$$B_1=\frac{-\phi_1+\sqrt{\phi_1{}^2+4\phi_2}}{2\phi_2},\quad B_2=\frac{-\phi_1-\sqrt{\phi_1{}^2+4\phi_2}}{2\phi_2}.$$

Отсюда

$$\frac{1}{B_1}=\frac{\phi_1+\sqrt{\phi_1{}^2+4\phi_2}}{2},\quad \frac{1}{B_2}=\frac{\phi_1-\sqrt{\phi_1{}^2+4\phi_2}}{2}.$$

Из условия стационарности $\left|\frac{1}{B_i}\right|<1,\ i=1,2$, получаем следующие два необходимые условия стационарности:

$$\left|\frac{1}{B_1}\frac{1}{B_2}\right|=|\phi_2|<1 \text{ и } \left|\frac{1}{B_1}+\frac{1}{B_2}\right|=|\phi_1|<2.$$

Таким образом, имеем следующие необходимые условия стационарности (без учета, являются ли корни действительными или комплексными):

$$-2<\phi_1<2 \text{ и } -1<\phi_2<1.$$

Условием вещественных корней является $\phi_1^2 + 4\phi_2 \geq 0$, что влечет следующую цепь неравенств:

$$-1 < \frac{1}{B_2} = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2} \leq \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2} = \frac{1}{B_1} < 1,$$

решив которые, получаем

$$\begin{cases} \phi_2 + \phi_1 < 1, \\ \phi_2 - \phi_1 < 1. \end{cases}$$

Для комплексных корней выполняется неравенство $\phi_1^2 + 4\phi_2 < 0$.

Таким образом, условием стационарности модели AR(2), выраженным в терминах значений параметров, является

$$\begin{cases} \phi_2 + \phi_1 < 1, \\ \phi_2 - \phi_1 < 1, \\ -1 < \phi_2 < 1. \end{cases} \tag{4.3}$$

Геометрически полученное условие стационарности представлено на рис. 4.2. Исходя из этих условий точки, соответствующие значениям параметров стационарного процесса, должны лежать внутри треугольника.

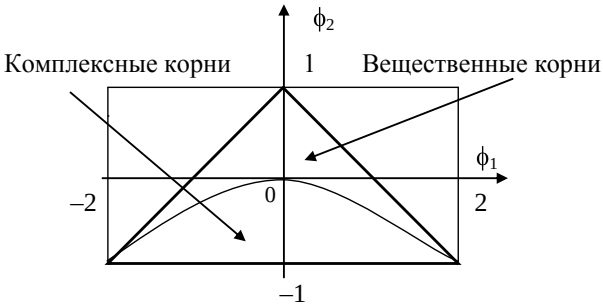


Рис. 4.2. Область стационарности процесса AR(2)

4.2.2.1. ТЕОРЕТИЧЕСКАЯ АВТОКОРРЕЛЯЦИОННАЯ ФУНКЦИЯ ДЛЯ ПРОЦЕССА AR(2)

Аналогично тому, как это делалось для модели AR(1), умножая обе части уравнения модели AR(2) на $\dot{Z}_{t-k}$ и применяя оператор математического ожидания, получаем уравнение для нахождения автоковариации:

$$E(\dot{Z}_{t-k}\dot{Z}_t) = \phi_1 E(\dot{Z}_{t-k}\dot{Z}_{t-1}) + \phi_2 E(\dot{Z}_{t-k}\dot{Z}_{t-2}) + E(\dot{Z}_{t-k}a_t),$$

откуда

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} \quad \text{для} \quad k \geq 1,$$

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} \quad \text{для} \quad k \geq 1.$$

Для того чтобы начать вычисления автокорреляций ρ_k по последней формуле, нужно знать два первых значения ρ_1 и ρ_2, которые определяются как решения следующей системы, получаемой после подстановки в разностное уравнение значений $k = 1$ и $k = 2$:

$$\begin{cases} \rho_1 = \phi_1 + \phi_2\rho_1, \\ \rho_2 = \phi_1\rho_1 + \phi_2. \end{cases}$$

Решив эту систему, получим значения ρ_1 и ρ_2, которые выражаются через коэффициенты ϕ_1 и ϕ_2 следующим образом:

$$\rho_1 = \frac{\phi_1}{1-\phi_2}, \ \rho_2 = \frac{\phi_1^2}{1-\phi_2} + \phi_2 = \frac{\phi_1^2 + \phi_2 - \phi_2^2}{1-\phi_2}.$$

Получив значения ρ_1 и ρ_2, далее ρ_k вычисляем рекурсивно для $k \geq 3$.

Чтобы сделать вывод о виде функции теоретической АКФ, рассмотрим однородное разностное уравнение, которому она удовлетворяет: $\rho_k - \phi_1\rho_{k-1} - \phi_2\rho_{k-2} = 0$. По теореме о решении однородных разностных уравнений имеем

$$\rho_k = b_1\left[\frac{1}{B_1}\right]^k + b_2\left[\frac{1}{B_2}\right]^k =$$

$$= b_1\left[\frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}\right]^k + b_2\left[\frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2}\right]^k = 0.$$

Константы b_1 и b_2 можно найти с использованием начальных условий для ρ_1 и ρ_2.

Таким образом, вид функции АКФ зависит от того, являются корни уравнения $\rho_k - \phi_1\rho_{k-1} - \phi_2\rho_{k-2} = 0$ вещественными или комплексными. Если оба корня – вещественные, то АКФ экспоненциально убывающая, а если корни комплексно-сопряженные, то АКФ ведет себя как затухающая синусоида. Модель AR(2) использовалась Дж. Юлом в 1921 году для описания поведения простого маятника [15], поэтому авторегрессионный процесс второго порядка иногда называется процессом Юла.

4.2.2.2. ЧАСТНАЯ АВТОКОРРЕЛЯЦИОННАЯ ФУНКЦИЯ ДЛЯ ПРОЦЕССА AR(2)

ЧАКФ для процесса AR(2) получаем, подставляя ρ_k из уравнения $\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$ в формулу (3.5):

$$\Phi_{11} = \rho_1 = \frac{\phi_1}{1-\phi_2},$$

$$\Phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} =$$

$$= \frac{\left(\dfrac{\phi_1^2 + \phi_2 - \phi_2^2}{1-\phi_2}\right) - \left(\dfrac{\phi_1}{1-\phi_2}\right)^2}{1 - \left(\dfrac{\phi_1}{1-\phi_2}\right)^2} = \frac{\phi_2\left[(1-\phi_2)^2 - \phi_1^2\right]}{(1-\phi_2)^2 - \phi_1^2} = \phi_2,$$

$$\Phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \rho_1 & \phi_1 + \phi_2\rho_1 \\ \rho_1 & 1 & \phi_2 + \phi_1\rho_1 \\ \rho_2 & \rho_1 & \phi_1\rho_2 + \phi_2\rho_1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = 0\,.$$

Аналогично можно показать, что $\Phi_{kk} = 0$ для всех $k \geq 3$, так как последний столбец матрицы числителя всегда является линейной комбинацией первых двух столбцов этой матрицы с коэффициентами ϕ_1 и ϕ_2. Таким образом, основной идентификационный показатель процесса AR(2) – эффект «обнуления» ЧАКФ при $k > 2$. Учитывая совместное поведение АКФ и ЧАКФ, можно представить четыре возможных сочетания этих характеристик, указывающих на принадлежность процесса к классу AR(2). Коррелограммы возможных вариантов АКФ и ЧАКФ для различных сочетаний знаков коэффициентов модели приведены на рис. 4.3.

4.2.3. ОБЩАЯ МОДЕЛЬ AR(*p*) АВТОРЕГРЕССИИ ПОРЯДКА *p*

Модель процесса AR(p) имеет вид

$$(1 - \phi_1 B - \ldots - \phi_p B^p)\dot{Z}_t = a_t\,,$$

или

$$\dot{Z}_t - \phi_1\dot{Z}_{t-1} - \ldots - \phi_p\dot{Z}_{t-p} = a_t\,.$$

Автокорреляционную функцию для AR(p) можно получить, умножив разностное уравнение системы на $\dot{Z}_{t-1}$:

$$\dot{Z}_{t-k}\dot{Z}_t = \phi_1\dot{Z}_{t-k}\dot{Z}_{t-1} + \phi_2\dot{Z}_{t-k}\dot{Z}_{t-2} + \ldots + \phi_p\dot{Z}_{t-k}\dot{Z}_{t-p} + \dot{Z}_{t-k}a_t\,.$$

Применив к обеим частям полученного равенства оператор математического ожидания с учетом условия $E(a_t Z_{t-k}) = 0$ для $k > 0$, получаем разностное уравнение для вычисления автоковариаций:

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \ldots\phi_p\gamma_{k-p} \text{ для } k > 0\,,$$

откуда

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots\phi_p\rho_{k-p} \text{ для } k > 0 .$$

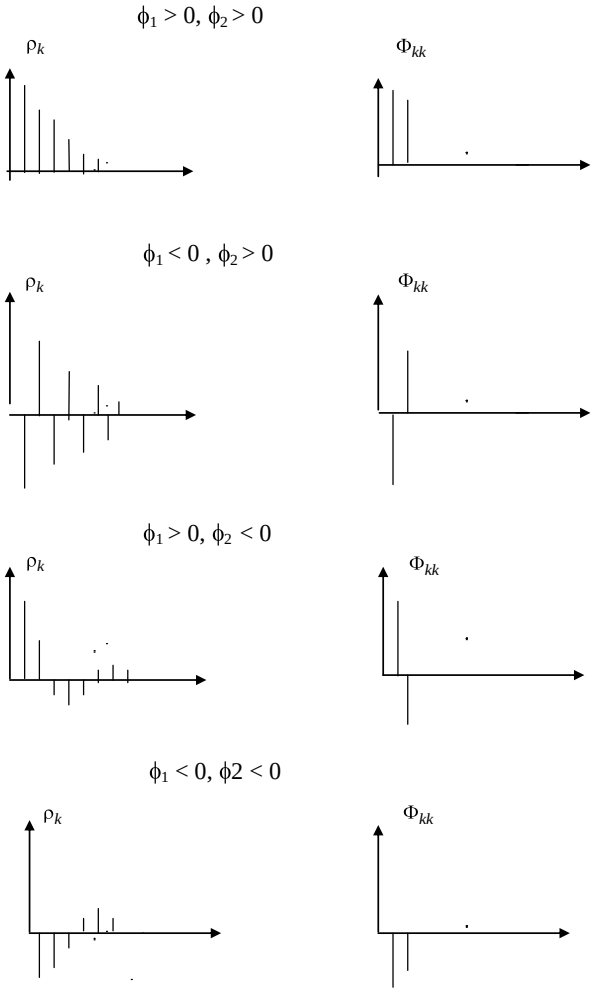


Рис. 4.3. Коррелограммы теоретических АКФ (ρ_k) и ЧАКФ (Φ_{kk}) для процесса AR(2)

Учитывая $\phi_p(B) = (1 - \phi_1 B - \ldots - \phi_p B^p)$, уравнение для АКФ можно записать как $\phi_p(B)\rho_k = 0$ для $k > 0$. Выполним разложение на множители:

$$\phi_p(B) = \prod_{i=1}^{m} (1 - G_i B)^{d_i} ,$$

где $\sum_{i=1}^{m} d_i = p$, а $G_i^{-1} = B_i$, $i = 1,\ldots,m$, – корни кратности d_i уравнения $\phi_p(B) = 0$. Теперь решение однородного разностного уравнения $\phi_p(B)\rho_k = 0$ можно записать следующим образом:

$$\rho_k = \sum_{i=1}^{m} G_i^k \sum_{j=0}^{d_i - 1} A_{ij} k^j .$$

Если $d_i = 1$ для всех i, то все G_i^{-1} различны и полученная формула модифицируется:

$$\rho_k = \sum_{i=1}^{p} A_i G_i^k , \quad k > 0 .$$

Для стационарного процесса $\left|G_i^{-1}\right| > 1$, т.е. $\left|G_i\right| < 1$. Следовательно, АКФ для процесса AR(p) есть смесь экспоненциальных затухающих функций и/или затухающих синусоидальных волн в зависимости от типа корней уравнения $\phi_p(B) = 0$ (действительные или комплексные). Синусоидальные волны появляются, когда имеются пары комплексно-сопряженных корней.

Применяя равенство $\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_p\rho_{k-p}$, $k > 0$, к формуле (3.5) для вычисления ЧАКФ, легко убедиться, что при $k > p$ последний столбец матрицы числителя в выражении (3.5) для Φ_{kk} может быть записан как линейная комбинация p первых столб-

цов этой матрицы. Следовательно, $\Phi_{kk} = 0$ при $k > p$, т.е. основной идентификационный показатель процесса AR(p) – эффект «обнуления» ЧАКФ после лага p.

4.3. МОДЕЛИ ПРОЦЕССОВ СКОЛЬЗЯЩЕГО СРЕДНЕГО

Модель процесса скользящего среднего порядка q (MA(q)) записывается следующим образом:

$$\dot{Z}_t = a_t - \theta_1 a_{t-1} - ... - \theta_q a_{t-q},$$

или, в операторной форме,

$$\dot{Z}_t = \theta(B) a_t,$$

где $\theta(B) = (1 - \theta_1 B - \ldots - \theta_q B^q)$.

Так как в данном случае $1 + \sum_{j=1}^{\infty} \psi_j^2 = 1 + \sum_{j=1}^{q} \theta_j^2 < \infty$, то процесс скользящего среднего конечного порядка всегда стационарен. Для того чтобы процесс был обратимым, т.е. допускал AR-представление, корни многочлена $\theta(B) = 0$ должны лежать вне единичного круга. Процессы скользящего среднего полезны при описании явлений, в которых эффект от событий проявляется немедленно и длится непродолжительное время. Для выявления свойств процесса рассмотрим сначала наиболее часто встречающиеся простые процессы MA(1) и MA(2), после чего обобщим эти свойства на общий процесс MA(q).

4.3.1. МОДЕЛЬ MA(1) – СКОЛЬЗЯЩЕЕ СРЕДНЕЕ ПЕРВОГО ПОРЯДКА

Рассмотрим модель процесса скользящего среднего первого порядка MA(1), положив $\theta(B) = (1 - \theta_1 B)$:

$$\dot{Z}_t = a_t - \theta_1 a_{t-1} = (1 - \theta_1 B) a_t,$$

где $\{a_t\}$ – процесс белого шума с нулевым математическим ожиданием и постоянной дисперсией σ_a^2. Математическое ожидание процесса $E(\dot{Z}_t) = 0$, $E(Z_t) = \mu$.

Так как $1+\theta_1^2<\infty$, то процесс МА(1) всегда стационарен. Для его обратимости корень уравнения $1-\theta_1 B=0$ должен лежать вне единичного круга. Следовательно, условием обратимости процесса МА(1) является $|\theta_1|<1$.

4.3.1.1. АКФ ПРОЦЕССА МА(1)

Для вычисления АКФ процессов скользящего среднего удобно использовать введенную в п. 3.5.1 для бесконечного (теоретического) процесса скользящего среднего автоковариационную генерирующую функцию

$$\gamma(B)=\sigma_a^2\theta(B)\theta(B^{-1})=\sigma_a^2(1-\theta_1 B)(1-\theta_1 B^{-1})=$$

$$=\sigma_a^2\left\{-\theta_1 B^{-1}+(1+\theta_1^2)-\theta_1 B\right\}.$$

Автоковариации γ_i получаются как коэффициенты при B^i:

$$\gamma_k=\begin{cases}(1+\theta_1^2)\sigma_a^2, \text{если } k=0,\\ -\theta_1\sigma_a^2, \qquad \text{если } k=1,\\ 0, \qquad\qquad \text{если } k>1.\end{cases}$$

Соответственно, так как $\gamma_0=\sigma_a^2(1+\theta_1^2)$, автокорреляции равны

$$\rho_k=\begin{cases}\dfrac{-\theta_1}{1+\theta_1^2}, \qquad \text{если } k=1,\\ 0, \qquad\qquad \text{если } k>1.\end{cases}$$

Таким образом, существенным идентификационным показателем для процесса МА(1) является эффект «обнуления» АКФ после лага 1.

4.3.1.2. ЧАКФ ПРОЦЕССА МА(1)

Частную автокорреляционную функцию для процесса МА(1) получим, подставляя выражения для ρ_k в формулу (3.5). В результате имеем

$$\Phi_{11} = \rho_1 = -\frac{\theta_1}{(1+\theta_1^2)} = -\frac{\theta_1(1-\theta_1^2)}{(1-\theta_1^4)};$$

$$\Phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{-\rho_1^2}{1-\rho_1^2} = \frac{-\theta_1^2}{1+\theta_1^2+\theta_1^4} = -\frac{\theta_1^2(1-\theta_1^2)}{1-\theta_1^6};$$

$$\Phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & 0 \\ 0 & \rho_1 & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & 0 \\ \rho_1 & 1 & \rho_1 \\ 0 & \rho_1 & 1 \end{vmatrix}} = \frac{\rho_1^3}{1-2\rho_1^2} = \frac{-\theta_1^3}{1+\theta_1^2+\theta_1^4+\theta_1^6} = \frac{-\theta_1^3(1-\theta_1^2)}{(1-\theta_1^8)}.$$

В общем случае формула для Φ_{kk} имеет вид

$$\Phi_{kk} = \frac{-\theta_1^k(1-\theta_1^2)}{(1-\theta_1^{2(k+1)})} \text{ для } k \geq 1.$$

Для обратимого процесса ($|\theta_1| < 1$) можно пренебречь зависимостью знаменателя этой дроби от k. Поэтому приближенно можно считать, что Φ_{kk} экспоненциально убывает в одной из двух форм (в зависимости от знака θ_1). Таким образом, для идентификации процесса МА(1) мы имеем два возможных набора коррелограмм (АКФ и ЧАКФ), изображенных на рис. 4.4.

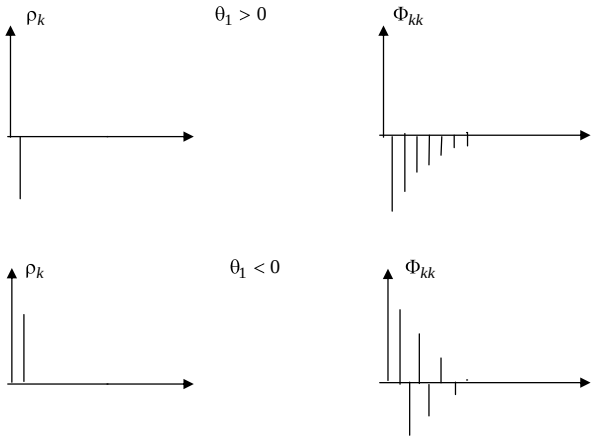


Рис. 4.4. Коррелограммы теоретических АКФ (ρ_k) и ЧАКФ (Φ_{kk}) для процесса МА(1)

4.3.2. МОДЕЛЬ МА(2) – СКОЛЬЗЯЩЕЕ СРЕДНЕЕ ВТОРОГО ПОРЯДКА

Рассмотрим модель процесса скользящего среднего второго порядка МА(2), полагая $\theta(B) = (1 - \theta_1 B - \theta_2 B^2)$:

$$\dot{Z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} = (1 - \theta_1 B - \theta_2 B^2) a_t ,$$

где $\{a_t\}$ – процесс белого шума. Математическое ожидание процесса $E(\dot{Z}_t) = 0$, $E(Z_t) = \mu$.

Как любой процесс скользящего среднего конечного порядка, процесс МА(2) всегда стационарен. Для его обратимости корни уравнения $(1 - \theta_1 B - \theta_2 B^2) = 0$ должны лежать вне единичного круга. Отсюда получаем условие обратимости (рис. 4.5), аналогичное условию стационарности (4.3) процесса AR(2):

$$\begin{cases} \theta_2 + \theta_1 < 1, \\ \theta_2 - \theta_1 < 1, \\ -1 < \theta_2 < 1. \end{cases}$$

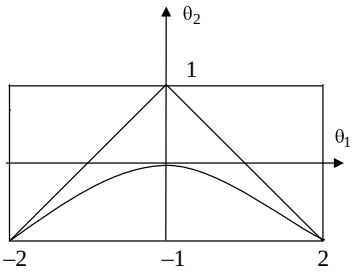


Рис. 4.5. Условие обратимости процесса МА(2)

4.3.2.1. АКФ ПРОЦЕССА МА(2)

Вычислим автоковариационную генерирующую функцию

$$\gamma(B) = \sigma_a^2\theta(B)\theta(B^{-1}) = \sigma_a^2(1-\theta_1 B-\theta_2 B^2)(1-\theta_1 B^{-1}-\theta_2 B^{-2}) =$$

$$= \sigma_a^2\left\{-\theta_2 B^{-2}-\theta_1(1-\theta_2)B^{-1}+(1+\theta_1^2+\theta_2^2)-\theta_1(1-\theta_2)B-\theta_2 B^2\right\}.$$

Автоковариации для модели МА(2) имеют вид

$$\gamma_k = \begin{cases} (1+\theta_1^2+\theta_2^2)\sigma_a^2, & \text{если } k=0, \\ -\theta_1(1-\theta_2)\sigma_a^2, & \text{если } k=1, \\ -\theta_2\sigma_a^2, & \text{если } k=2, \\ 0, & \text{если } k>2. \end{cases}$$

Соответственно автокорреляционная функция для модели МА(2) имеет вид

$$\rho_k = \begin{cases} \dfrac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}, & \text{если} \quad k=1, \\ \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2}, & \text{если} \quad k=2, \\ 0, & \text{если} \quad k>2, \end{cases}$$

т.е. существенным идентификационным показателем для процесса МА(2) является эффект «обнуления» АКФ после лага 2.

4.3.2.2. ЧАКФ ПРОЦЕССА МА(2)

Подставляя ρ_k в формулу (3.5) и учитывая, что $\rho_k = 0$ для $k \geq 3$, получим

$$\Phi_{11} = \rho_1,\ \Phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2},\ \Phi_{33} = \frac{\rho_1^3 - \rho_1\rho_2(2 - \rho_2)}{1 - \rho_2^2 - 2\rho_1^2(1 - \rho_2)}, \ldots$$

Процесс МА(2) содержит процесс МА(1) как частный случай. Следовательно, ЧАКФ для МА(2) – экспоненциальная убывающая функция или затухающая синусоидальная волна в зависимости от знаков и величины коэффициентов θ_1 и θ_2 или, что эквивалентно, от корней уравнения $(1 - \theta_1 B - \theta_2 B^2) = 0$. Так, ЧАКФ является затухающей синусоидальной волной, если корни этого уравнения – комплексно-сопряженные.

Четыре возможных сочетания коррелограмм АКФ и ЧАКФ для модели МА(2) представлены на рис 4.6. Их можно использовать при идентификации этого процесса.

4.3.3. ОБЩАЯ МОДЕЛЬ СКОЛЬЗЯЩЕГО СРЕДНЕГО МА(*q*)

Общий процесс скользящего среднего порядка q удовлетворяет уравнению

$$\dot{Z}_t = (1 - \theta_1 B - \ldots - \theta_q B^q) a_t .$$

Для этого общего процесса МА(q) дисперсия определяется

$$\gamma_0 = \sigma_a^2 \sum_{j=0}^{q} \theta_j^2 ,$$

где $\theta_0 = 1$;
автоковариации и автокорреляции определяются

$$\gamma_k = \begin{cases} \sigma_a^2(-\theta_k + \theta_1\theta_{k-1} + \ldots + \theta_{q-k}\theta_q), & k = 1, 2, \ldots, q, \\ 0, & k > q, \end{cases}$$

$$\rho_k = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k+1} + \ldots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \ldots + \theta_q^2}, & k = 1, 2, \ldots, q, \\ 0, & k > q. \end{cases}$$

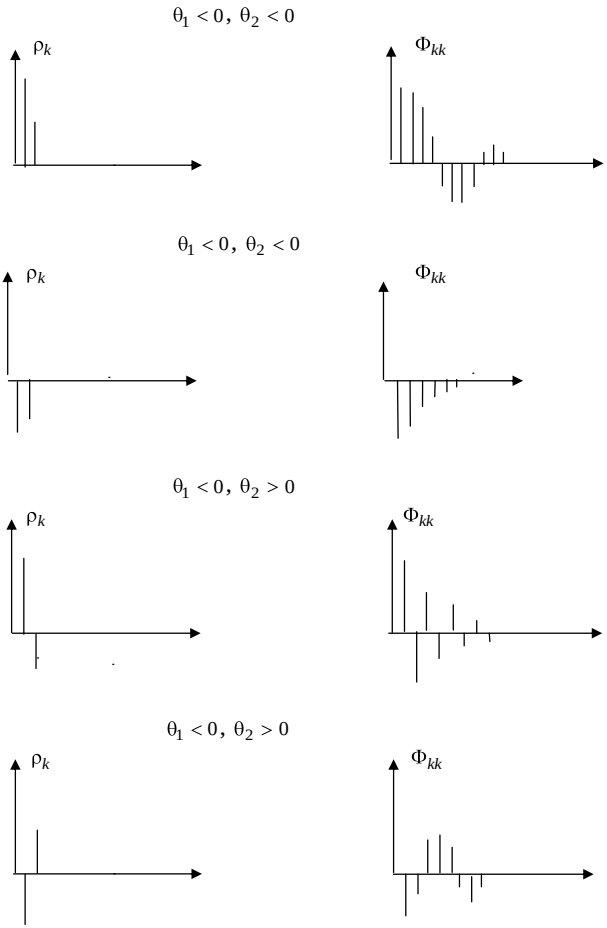


Рис. 4.6. Коррелограммы теоретических АКФ (ρ_k) и ЧАКФ (Φ_{kk}) для процесса МА(2)

Таким образом, можно идентифицировать процесс MA(q) по признаку статистической незначимости значений АКФ с номерами $k > q$. ЧАКФ содержит затухающие синусоидальные волны, если среди корней уравнения $(1 - \theta_1 B - \ldots - \theta_q B^q) = 0$ есть комплексно-сопряженные, и экспоненциально убывает, если все корни действительны.

4.4. АВТОРЕГРЕССИОННЫЕ ПРОЦЕССЫ СКОЛЬЗЯЩЕГО СРЕДНЕГО ARMA(*p, q*)

Стационарный и обратимый процесс может быть представлен как в форме скользящего среднего, так и в форме авторегрессии. Проблемой в обоих случаях может оказаться то, что для хорошей аппроксимации могут потребоваться модели высокого порядка, содержащие слишком много параметров. В общем случае большое число параметров уменьшает эффективность оценивания. Выход из такой ситуации – в использовании смешанной модели, включающей и компоненты авторегрессии, и компоненты скользящего среднего. Данная модель носит название ARMA(p, q) и имеет вид

$$\phi_p(B)\dot{Z}_t = \theta_q(B)a_t,$$

где $\phi_p(B) = 1 - \phi_1 B - ... - \phi_p B^p$, $\theta_q(B) = 1 - \theta_1 B - ... - \theta_q B^q$.

Чтобы процесс ARMA(p, q) был обратим, потребуем, чтобы корни уравнения $\theta_q(B) = 0$ лежали вне единичного круга. Для стационарности процесса ARMA(p, q) нужно, чтобы корни уравнения $\phi_p(B) = 0$ лежали вне единичного круга. Потребуем также, чтобы уравнения $\phi_p(B) = 0$ и $\theta_q(B) = 0$ не имели общих корней (в противном случае получаем неидентифицируемые параметры – коэффициенты общих множителей полиномов $\phi_p(B)$ и $\theta_q(B)$). Таким образом, p и q в модели ARMA(p, q) – порядки полиномов авторегрессии и скользящего среднего соответственно.

Стационарный и обратимый процесс ARMA(p, q) может быть записан в чисто авторегрессионном представлении бесконечного порядка:

$$\pi(B)\dot{Z}_t = a_t\,,$$

где $\pi(B) = \dfrac{\phi_p(B)}{\theta_q(B)} = (1 - \pi_1 B - \pi_2 B^2 - \ldots)\,.$

Стационарный и обратимый процесс ARMA(p, q) может быть также записан в представлении чисто скользящего среднего (бесконечного порядка):

$$\dot{Z}_t = \psi(B)a_t\,,$$

где $\psi(B) = \dfrac{\theta_q(B)}{\phi_p(B)} = (1 + \psi_1 B + \psi_2 B^2 + \ldots)\,.$

4.4.1. АКФ ДЛЯ ПРОЦЕССА ARMA(p, q)

Для получения АКФ запишем модель ARMA(p, q) в виде

$$\dot{Z}_t = \phi_1\dot{Z}_{t-1} + \phi_2\dot{Z}_{t-2} + \ldots + \phi_p\dot{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}\,,$$

умножим обе части на $\dot{Z}_{t-k}$ и применим оператор математического ожидания

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \ldots + \phi_p\gamma_{k-p} + E(\dot{Z}_{t-k}a_t) -$$

$$-\theta_1 E(\dot{Z}_{t-k}a_{t-1}) - \ldots - \theta_q E(\dot{Z}_{t-k}a_{t-q})\,.$$

Так как

$$E(\dot{Z}_{t-k}a_{t-i}) = 0 \text{ для } k > i\,,$$

то $\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \ldots + \phi_p\gamma_{k-p}$ для $k \geq q+1$.

Соответственно $\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \ldots + \phi_p\gamma_{k-p}$ для $k \geq q+1$.

Последнее уравнение для АКФ удовлетворяет однородному разностному уравнению порядка p для лагов $k \geq q+1$ точно так же, как и АКФ для AR-модели. Поэтому АКФ для процесса ARMA(p, q) после лага $k = q$ ведет себя точно так же, как АКФ для процесса AR(p) и зависит только от авторегрессионых параметров модели $\phi_1, \phi_2, \ldots, \phi_p$.

Однако первые q автокорреляций $\rho_q, \rho_{q-1}, \ldots, \rho_1$ зависят и от AR-коэффициентов $\phi_1, \phi_2, \ldots, \phi_p$, и от MA-коэффициентов θ_1, $\theta_2, \ldots, \theta_q$ и являются начальными значениями для модели. Это отличие может оказаться полезным при идентификации модели.

4.4.2. ЧАКФ ДЛЯ ПРОЦЕССА ARMA(*p*, *q*)

Поскольку особым случаем процесса ARMA(p, q) является процесс MA(q), ЧАКФ для процесса ARMA(p, q) будет, как и для процесса MA(q), смесью убывающих экспонент и/или затухающих синусоидальных волн в зависимости от корней уравнений $\phi(B) = 0$ и $\theta(B) = 0$.

4.4.3. ПРОЦЕСС ARMA(1, 1)

Рассмотрим самый простой процесс из класса смешанных процессов авторегрессии и скользящего среднего – процесс ARMA(1, 1):

$$(1 - \phi_1 B)\dot{Z}_t = (1 - \theta_1 B)a_t,$$

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + a_t - \theta_1 a_{t-1}.$$

Для стационарности предположим $|\phi_1| < 1$, для обратимости потребуем $|\theta_1| < 1$. Если $\phi_1 = 0$, то ARMA(1, 1) сводится к MA(1), если $\theta_1 = 0$, то ARMA(1,1) сводится к AR(1). Поэтому процессы AR(1) и MA(1) – частные случаи процесса ARMA(1, 1). Запишем модель ARMA(1, 1) в чисто авторегрессионном представлении (бесконечного порядка):

$$\pi(B)\dot{Z}_t = a_t,$$

где $\pi(B) = (1 - \pi_1 B - \pi_2 B^2 - \ldots) = \dfrac{(1 - \phi_1 B)}{(1 - \theta_1 B)}$,

т.е. $(1 - \theta_1 B)(1 - \pi_1 B - \pi_2 B^2 - \ldots) = (1 - \phi_1 B)$,

или

$$[1 - (\pi_1 + \theta_1)B - (\pi_2 - \pi_1\theta_1)B^2 -$$

$$-(\pi_3 - \pi_2\theta_1)B^3 - \ldots] = (1 - \phi_1 B).$$

Приравнивая коэффициенты при одинаковых степенях B, получаем

$$\pi_j = \theta_1^{j-1}(\phi_1 - \theta_1) \text{ для } j \geq 1 .$$

Теперь запишем модель ARMA(1, 1) в представлении чисто скользящего среднего (бесконечного порядка):

$$\dot{Z}_t = \psi(B)a_t = \frac{(1-\theta_1 B)}{(1-\phi_1 B)}a_t .$$

Получим коэффициенты МА-модели

$$(1-\phi_1 B)(1+\psi_1 B+\psi_2 B^2+\psi_3 B^3\ldots) = (1-\theta_1 B) ,$$

или

$$[1-(\psi_1-\phi_1)B+(\psi_2-\psi_1\phi_1)B^2+\ldots) = (1-\theta_1 B) .$$

В результате имеем $\psi_j = \phi_1^{j-1}(\phi_1 - \theta_1)$ для $j \geq 1$.

4.4.3.1. АКФ ДЛЯ ПРОЦЕССА ARMA(1,1)

Чтобы получить выражения для автоковариаций, умножим исходное уравнение модели на $\dot{Z}_{t-k}$ и возьмем математическое ожидание:

$$\gamma_k = \phi_1\gamma_{k-1} + E(\dot{Z}_{t-k}a_t) - \theta_1 E(\dot{Z}_{t-k}a_{t-1}) ,$$

в специальном случае $\gamma_0 = \phi_1\gamma_1 + E(\dot{Z}_t a_t) - \theta_1 E(\dot{Z}_t a_{t-1})$.

Справедливо $E(\dot{Z}_t a_t) = \sigma_a^2$:

$$E(\dot{Z}_t a_t) = \phi_1 E(\dot{Z}_{t-1}a_t) + E(a_t^2) - \theta_1 E(a_{t-1}a_t) = \sigma_a^2 ,$$

так как $E(\dot{Z}_{t-1}a_t) = 0$ и $E(a_{t-1}a_t) = 0$.

Справедливо также

$$E(\dot{Z}_t a_{t-1}) = \phi_1 E(\dot{Z}_{t-1}a_{t-1}) +$$

$$+E(a_t a_{t-1}) - \theta_1 E(a_{t-1}^2) = (\phi_1 - \theta_1)\sigma_a^2 .$$

Следовательно,

$$\gamma_0 = \phi_1\gamma_1 + \sigma_a^2 - \theta_1(\phi_1 - \theta_1)\sigma_a^2 . \qquad (4.4)$$

Для $k = 1$ получаем

$$\gamma_1 = \phi_1\gamma_0 - \theta_1\sigma_a^2. \tag{4.5}$$

Подставляя выражение (4.5) в выражение для γ_0 (4.4), получаем уравнение

$$\gamma_0 = \phi_1^2\gamma_0 - \phi_1\theta_1\sigma_a^2 + \sigma_a^2 - \phi_1\theta_1\sigma_a^2 + \theta_1^2\sigma_a^2,$$

решая которое, имеем $\gamma_0 = \dfrac{(1+\theta_1^2 - 2\phi_1\theta_1)}{(1-\phi_1^2)}\sigma_a^2$.

Далее,

$$\gamma_1 = \phi_1\gamma_0 - \theta_1\sigma_a^2 =$$

$$= \frac{\phi_1(1+\theta_1^2 - 2\phi_1\theta_1)}{(1-\phi_1^2)}\sigma_a^2 - \theta_1\sigma_a^2 = \frac{(\phi_1-\theta_1)(1-\phi_1\theta_1)}{(1-\phi_1^2)}\sigma_a^2.$$

Для $k \geq 2$ автоковариации удовлетворяют разностному уравнению $\gamma_k = \phi_1\gamma_{k-1}$ и вычисляются рекурсивно. Следовательно, модель ARMA(1, 1) имеет следующую автокорреляционную функцию:

$$\rho_k = \begin{cases} 1, & k = 0, \\ \dfrac{(\phi_1-\theta_1)(1-\phi_1\theta_1)}{1+\theta_1^2 - 2\phi_1\theta_1}, & k = 1, \\ \phi_1\rho_{k-1}, & k \geq 2. \end{cases}$$

Заметим, что АКФ для модели ARMA(1, 1) сочетает характеристики обоих процессов – AR(1) и MA(1). Параметр скользящего среднего θ_1 входит в выражение для ρ_1. Остальные автокорреляции (кроме ρ_1) модели ARMA(1, 1) ведут себя аналогично автокорреляциям модели AR(1).

4.4.3.2. ЧАКФ ДЛЯ ПРОЦЕССА ARMA(1, 1)

Общая формула ЧАКФ для смешанной модели сложна и не особенно нужна. Достаточно заметить, что, поскольку процесс ARMA(1, 1) содержит MA(1) как частный случай, ЧАКФ процесса

ARMA(1, 1) также убывает экспоненциально, как и АКФ этого процесса в форме, зависящей от знаков и соотношения коэффициентов ϕ_1 и θ_1 (с чередующимися или постоянными знаками). Таким образом, тот факт, что и АКФ, и ЧАКФ убывают экспоненциально (не «обнуляются» после какого-либо лага), свидетельствует о том, что для описания стационарного процесса требуется смешанная модель ARMA. На рис. 4.7 приведены некоторые возможные соотношения теоретических АКФ и ЧАКФ процесса ARMA(1, 1). Благодаря эффекту сочетания знаков и величин коэффициентов ϕ_1 и θ_1 ЧАКФ для процесса ARMA(1, 1) содержит намного больше различных форм, чем ЧАКФ для MA(1), которая содержит только две возможности.

В заключение рассмотрим модель с ненулевым средним ARMA(p, q):

$$(1-\phi_1 B-...-\phi_p B^p)(\dot{Z}_t-\mu)=(1-\theta_1 B-...-\theta_q B^q)a_t,$$

которая может быть записана в виде

$$(1-\phi_1 B-...-\phi_p B^p)\dot{Z}_t=\theta_0+(1-\theta_1 B-...-\theta_q B^q)a_t,$$

где $\theta_0=(1-\phi_1 B-...-\phi_p B^p)\mu=(1-\phi_1-...-\phi_p)\mu$.

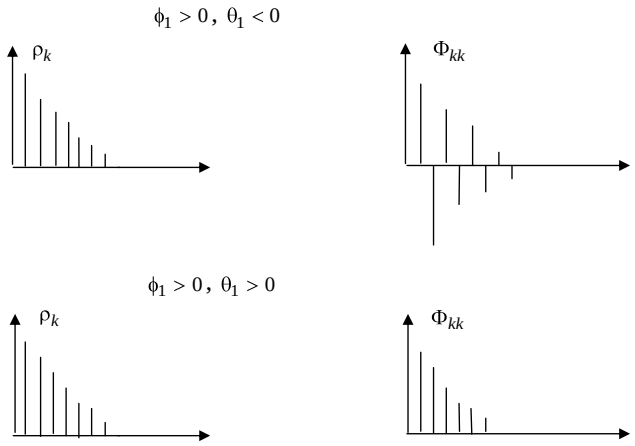


Рис. 4.7. Коррелограммы теоретических АКФ (ρ_k) и ЧАКФ (Φ_{kk}) для процесса ARMA(1, 1) (см. также на с. 112)

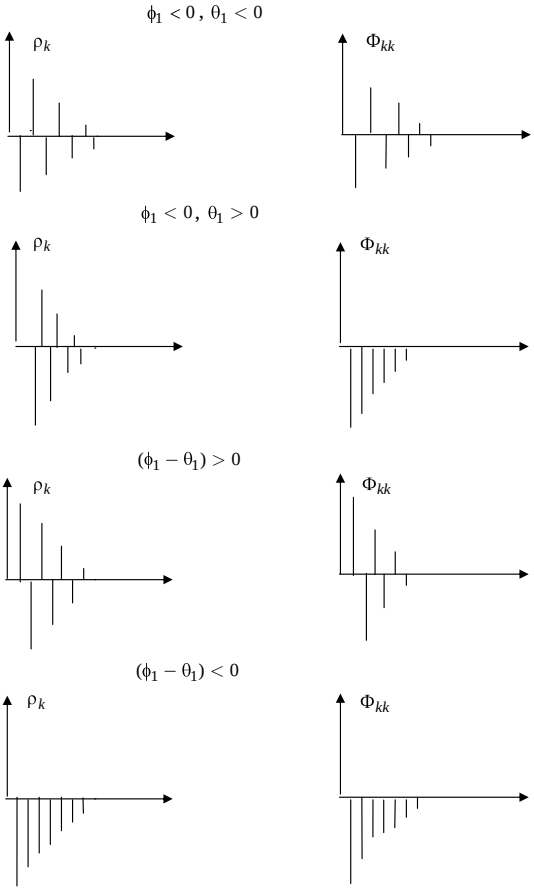


Рис. 4.7. Окончание

Для частного случая процесса AR(p) имеем:

$$(1 - \phi_1 B - ... - \phi_p B^p)\dot{Z}_t = \theta_0 + a_t ,$$

где $\theta_0 = (1 - \phi_1 - ... - \phi_p)\mu$.

Для частного случая процесса MA(q) имеем:

$$\dot{Z}_t = \theta_0 + (1 - \theta_1 B - ... - \theta_q B^q)a_t , \text{ где } \theta_0 = \mu .$$

5. МОДЕЛИ НЕСТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ

Рассмотренные в разделе 4 модели ARMA(p, q) позволяют описывать широкий спектр процессов, однако эти процессы являются стационарными. Но очень многие временные ряды, часто встречающиеся на практике, особенно в области экономики и бизнеса, являются нестационарными. Относительно рассматриваемого нами класса гауссовских стационарных процессов нестационарность бывает разного типа. Это может быть нестационарность по среднему, когда математическое ожидание процесса изменяется с течением времени, т.е. имеется тренд. С другой стороны, может иметь место нестационарность по моментам второго порядка, т.е. по дисперсиям и ковариациям. И, наконец, вполне типична ситуация, когда мы имеем нестационарность и по среднему, и по дисперсии и автоковариации. Так, очень часто с возрастанием уровня ряда в соответствии с имеющимся трендом возрастает и дисперсия, так как для больших значений, например, прибыли, отклонения естественным образом увеличиваются с увеличением уровня ряда. График ежемесячного производства табака (см. рис. 1.2) в США (1871–1979 годы) показывает, что и средний уровень зависит от времени, и дисперсия возрастает с увеличением уровня среднего.

В данном разделе мы вводим полезный класс *однородно-нестационарных процессов*, для моделирования которых используется класс моделей ARIMA – авторегрессионные модели с проинтегрированным скользящим средним. Класс ARIMA – это надстройка над классом моделей ARMA, применяемых для описания стационарных временных рядов.

5.1. НЕСТАЦИОНАРНОСТЬ ПО ОТНОШЕНИЮ К СРЕДНЕМУ

Процесс, нестационарный в среднем, может ставить очень серьезную проблему оценивания функции среднего, зависящей от времени, без наличия множества реализаций. К счастью, существует класс процессов, для которых можно построить модель по единственной реализации. В данном подразделе мы рассмотрим два таких класса моделей, часто оказывающихся очень полезными. Это модели с детерминированным трендом, построенные по классической методологии, и модели со стохастическим трендом, являющиеся расширением методологии ARIMA.

5.1.1. МОДЕЛИ С ДЕТЕРМИНИРОВАННЫМ ТРЕНДОМ

Функция среднего для нестационарного процесса может быть представлена в виде детерминированного тренда, зависящего от времени. В этом случае для описания явления могла бы быть полезна стандартная регрессия. Тип используемой детерминированной функции тренда зависит от вида зависимости уровня ряда от времени. Например, если зависимость уровня ряда от времени – линейная, то в качестве тренда можно использовать линейную функцию $\mu_t = \alpha_0 + \alpha_1 t$. Модель с таким трендом имеет вид

$$Z_t = \alpha_0 + \alpha_1 t + a_t,$$

где $\{a_t\}$ – процесс белого шума с нулевым средним.

Если зависимость уровня ряда от времени – квадратичная, то в качестве тренда можно использовать квадратичную функцию $\mu_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$, модель при этом становится иной:

$$Z_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + a_t.$$

Для полинома k-го порядка по времени модель имеет следующий вид:

$$Z_t = \alpha_0 + \alpha_1 t + \ldots + \alpha_k t^k + a_t.$$

Если детерминированный тренд представлен синусоидальной кривой, можно использовать модель

$$Z_t = \nu_0 + \nu\cos(wt + \theta) + a_t = \nu_0 + \alpha\cos(wt) + \beta\sin(wt) + a_t,$$

где $\alpha = \nu\cos\theta$, $\beta = -\nu\sin\theta$, $\nu = \sqrt{\alpha^2 + \beta^2}$, $\theta = \mathrm{arctg}(-\beta/\alpha)$.

В более общем случае периодичность может быть представлена следующей моделью, которая часто называется моделью скрытой периодичности:

$$Z_t = \nu_0 + \sum_{j=1}^{m}(\alpha_j \cos\omega_j t + \beta_j \sin\omega_j t) + a_t.$$

Представленные выше модели с детерминированным трендом можно рассматривать стандартными методами регрессионного анализа.

5.1.2. МОДЕЛИ СТОХАСТИЧЕСКОГО ТРЕНДА И ОПЕРАТОР ВЗЯТИЯ РАЗНОСТИ

Хотя многие временные ряды нестационарны, благодаря некоторым уравновешивающим силам различные части этих рядов ведут себя очень одинаково, различие заключается лишь в локальных изменениях уровней. Дж. Бокс и Г. Дженкинс [3] назвали этот тип нестационарного поведения *однородной нестационарностью*. Однородная нестационарность встречается во временных рядах, с которыми мы имеем дело в экономике, в бизнесе, в социологических исследованиях.

В терминологии ARMA-моделей процесс нестационарен, если некоторые корни AR-полиномов не лежат вне единичного круга. Однородная нестационарность предполагает, что локальное поведение временного ряда не зависит от уровня, на котором происходят осцилляции, т.е. механизм, порождающий случайную составляющую процесса, не связан с возможным изменением уровня в зависимости от тренда.

Предположим, что поведение однородно-нестационарного процесса описывается с помощью оператора авторегрессии $\Psi(B)$. Свойство однородной нестационарности можно математически сформулировать в следующем виде:

$$\Psi(B)(Z_t + C) = \Psi(B)(Z_t) \quad \forall C,$$

где C – некоторая константа, отражающая факт изменения уровня ряда. Из данного свойства следует, что оператор $\Psi(B)$ должен содержать множитель $(1 - B)$. Только в этом случае справедливо вышеприведенное равенство

$$(1 - B)(Z_t + C) = (1 - B)Z_t,$$

так как $(1 - B)C = C - C = 0$.

Таким образом, однородная нестационарность получается, если один из корней оператора авторегрессии в точности равен единице: $B = 1$. Отсюда следует, что оператор $\Psi(B)$ можно представить в виде

$$\Psi(B) = \phi(B)(1 - B)^d \text{ для некоторого } d > 0, \qquad (5.1)$$

где $\phi(B)$ – стационарный авторегрессионный оператор, корни которого лежат вне единичного круга.

Декомпозиция (5.1) свидетельствует о том, что однородно-нестационарный ряд может быть сведен к стационарному ряду применением *оператора взятия разности* $(1-B)$. Другими словами, если ряд $\{Z_t\}$ – однородно-нестационарен, причем кратность корня $B=1$ равна d, то ряд, полученный взятием разности d раз, т.е. $W_t = (1-B)^d Z_t$ для некоторого целого $d \geq 1$, является стационарным.

Например, если ряд, полученный взятием разности d раз, удовлетворяет процессу белого шума, мы имеем $(1-B)^d Z_t = a_t$. При $d=1$ получаем процесс случайного блуждания, удовлетворяющий уравнению

$$(1-B)Z_t = Z_t - Z_{t-1} = a_t.$$

При заданных прошлых значениях Z_{t-1}, Z_{t-2}, … средний уровень процесса случайного блуждания в момент t равен $\mu_t = E(Z_t \mid Z_{t-1},\ldots) = Z_{t-1}$. Таким образом, средний уровень процесса Z_t в модели $(1-B)^d Z_t = a_t$, $d \geq 1$, изменяется во времени стохастически, т.е. случайным образом. Поэтому модель, в которой содержится авторегресионный оператор с корнем $B=1$, называется моделью со стохастическим трендом, в отличие от рассмотренной ранее модели с детерминированным трендом, в которой средний уровень процесса в момент t является чисто детерминированной функцией времени.

5.1.3. МОДЕЛЬ АВТОРЕГРЕССИИ С ПРОИНТЕГРИРОВАННЫМ СКОЛЬЗЯЩИМ СРЕДНИМ

Стационарный процесс, полученный в результате взятия разностей однородно-нестационарного процесса, не обязательно является белым шумом. В общем случае ряд $W_t = (1-B)^d Z_t$ удовлетворяет общей модели ARMA(p, q), описывающей поведение стационарного ряда. Таким образом, для описания однородно-нестационарного ряда имеем следующую модель:

$$\Phi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)a_t, \qquad (5.2)$$

где стационарный AR-оператор $\Phi_p(B) = (1 - \Phi_1 B - \ldots - \Phi_p B^p)$ и обратимый MA-оператор $\theta_q(B) = (1 - \theta_1 B - \ldots - \theta_q B^q)$ не имеют общих корней. Параметр θ_0 играет различную роль в случае $d = 0$ и $d > 0$. При $d = 0$ исходный процесс Z_t стационарен и θ_0 связан со средним процесса уравнением $\theta_0 = (1 - \phi_1 - ... - \phi_p)\mu$ и может быть опущен (учтен неявно). С другой стороны, при $d \geq 1$ параметр θ_0 называется детерминированным трендом и, как будет показано далее, часто исключается из модели, если в его включении действительно нет необходимости.

Модель (5.2) называется моделью авторегрессии с проинтегрированным скользящим средним ARIMA(p, d, q). При $p = 0$ мы имеем модель IMA(d, q). Далее рассмотрим некоторые наиболее известные модели ARIMA.

5.1.3.1. МОДЕЛЬ СЛУЧАЙНОГО БЛУЖДАНИЯ

Если в модели ARIMA(p, d, q) положить $p = 0$, $d = 1$, $q = 0$, мы имеем хорошо известную модель случайного блуждания:

$$(1-B)Z_t = a_t, \text{ или } Z_t = Z_{t-1} + a_t.$$

Эта модель широко используется для описания поведения временного ряда при ежедневных изменениях цен на акции. В модели случайного блуждания значение Z в момент t равно ее значению в момент $(t-1)$ плюс случайное возмущение. Такое поведение подобно поведению нетрезвого человека, чье положение в момент t есть позиция при $(t-1)$ плюс шаг в случайном направлении в момент t.

Для того чтобы понять, как ведет себя АКФ для модели случайного блуждания, заметим, что эта модель может рассматриваться как предельный случай при $\Phi \to 1$ модели AR(1), описываемой уравнением $(1-\Phi B)Z_t = a_t$. Поскольку АКФ для процесса AR(1) ведет себя по экспоненциальному закону $\rho_k = \Phi^k$, то для модели случайного блуждания мы получаем в пределе $\rho_k \to 1$. Таким образом, можно охарак-

теризовать модель случайного блуждания тем, что она имеет большие, очень слабо уменьшающиеся с увеличением k значения выборочной АКФ для исходного ряда Z_t.

Выборочная ЧАКФ для процесса случайного блуждания также есть предельный случай ЧАКФ для процесса AR(1). В результате ЧАКФ имеет близкое к единице значение на первом лаге (при $k = 1$) и очень маленькие, близкие к нулю значения на остальных лагах.

Теперь рассмотрим очень простую модификацию модели случайного блуждания – модель случайного блуждания со сдвигом, описываемую следующим уравнением:

$$(1 - B)Z_t = \theta_0 + a_t, \text{ или } Z_t = Z_{t-1} + \theta_0 + a_t.$$

Произведем $(t - k)$ последовательных подстановок $Z_{t-1} = Z_{t-2} + \theta_0 + a_{t-1}$, $Z_{t-2} = Z_{t-3} + \theta_0 + a_{t-2}$, … в исходное уравнение модели, в результате чего имеем

$$Z_t = Z_{t-1} + \theta_0 + a_t = Z_{t-2} + 2\theta_0 + a_t + a_{t-1} =$$

$$= Z_k + (t - k)\theta_0 + \sum_{j=k+1}^{t} a_j \text{ для } t > k.$$

Из полученного уравнения видно, что Z_t содержит детерминированный тренд с наклоном θ_0.

В общем случае, для модели, включающей взятие разности порядка d, т.е. $\left\{(1 - B)^d Z_t\right\}$, ненулевой параметр θ_0 в исходной модели соответствует коэффициенту α_d при t^d в детерминированном тренде $\alpha_0 + \alpha_1 t + \ldots + \alpha_d t^d$. Поэтому при $d > 0$ θ_0 называется членом детерминированного тренда. Для больших t член детерминированного тренда может оказаться очень большим, и ряд будет удовлетворять чисто детерминированной модели. Поэтому при $d > 0$ полагают $\theta_0 = 0$, если из данных или из природы процесса не ясно, что детерминированный компонент действительно необходим.

Для модели случайного блуждания со сдвигом при известных Z_{t-1}, Z_{t-2}, … средним значением ряда Z_t в момент времени t является

$$\mu_t = Z_{t-1} + \theta_0 .$$

На среднее μ_t в данном случае влияет случайное возмущение в момент $(t-1)$ через Z_{t-1}, а также детерминированный компонент через наклон θ_0. При $\theta_0 = 0$ мы имеем модель чисто стохастического тренда – модель случайного блуждания без детерминированного тренда. Эта модель характеризуется коррелограммами выборочных АКФ и ЧАКФ, представленными на рис. 5.1.

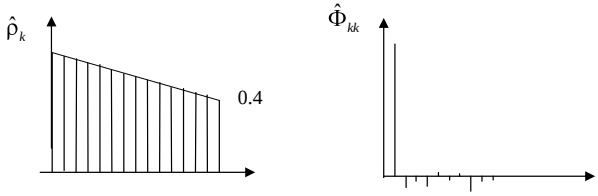


Рис. 5.1. Коррелограммы АКФ ($\hat{\rho}_k$) и ЧАКФ ($\hat{\Phi}_{kk}$) процесса случайного блуждания

После взятия разности получаем процесс белого шума в чистом виде:

$$W_t = a_t = (1-B)Z_t ,$$

характеризующийся нулевыми значениями АКФ и ЧАКФ, как показано на рис. 5.2.

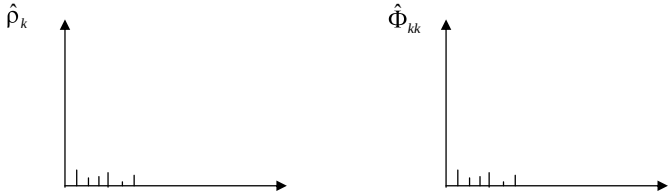
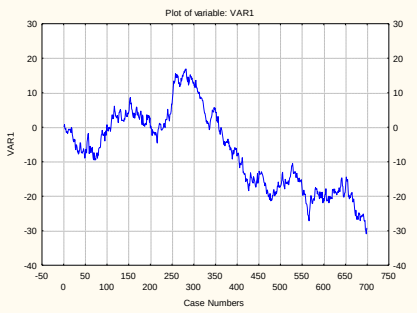


Рис. 5.2. Коррелограммы процесса случайного блуждания после взятия разности

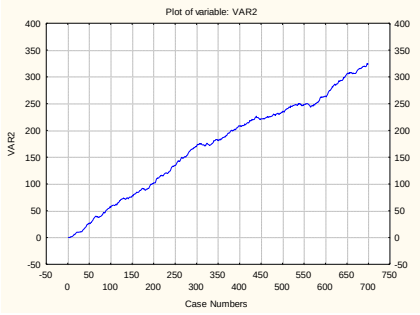
Рассмотрим модель случайного блуждания со сдвигом $\theta_0 = 4$:

$$(1-B)Z_t = 4 + a_t ,$$

для которой коррелограммы имеют вид, аналогичный модели случайного блуждания без сдвига, представленный на рис. 5.1, только АКФ убывает несколько медленнее. Возникает вопрос, как же различить при идентификации эти два типа моделей? Ответ на этот вопрос очевиден, если взглянуть на исходные графики обоих временных рядов и сравнить их между собой (рис. 5.3).



а



б

Рис. 5.3. Графики временных рядов для моделей случайного блуждания:

а – модель без сдвига – доминирует случайная составляющая, значения случайно блуждают; *б* – доминирует детерминированная составляющая – детерминированный линейный тренд с наклоном 0,5

5.1.3.2. МОДЕЛЬ ARIMA(0, 1, 1), ИЛИ IMA(1, 1)

При $p = 0$, $d = 1$, $q = 1$ мы получаем модель ARIMA(0, 1, 1), или IMA(1, 1):

$$(1 - B)Z_t = (1 - \theta B)a_t \text{ или } Z_t = Z_{t-1} + a_t - \theta a_{t-1},$$

где $-1 < \theta < 1$.

После взятия первой разности эта модель сводима к стационарной модели MA(1). Модель случайного блуждания есть специальный случай модели IMA(1, 1), получаемый при $\theta = 0$.

Если посмотреть на коррелограммы выборочных АКФ и ЧАКФ этого процесса, то они имеют вид, аналогичный коррелограммам, изображенным на рис. 5.1. Более того, такой вид имеют коррелограммы любого однородно-нестационарного процесса. На основе данного признака можно установить или подтвердить однородную нестационар-

ность случайного процесса, реализацией которого является тот или иной временной ряд.

Однако возникает вопрос, если коррелограммы всех однородно-нестационарных моделей аналогичны и принципиально неотличимы друг от друга, то как различить эти процессы? Как идентифицировать модель IMA(1, 1) и отличить от модели случайного блуждания или какой-либо другой однородно-нестационарной модели?

Ответ на поставленный вопрос очевиден. В случае модели IMA(1, 1) надо просто один раз применить к исходному ряду оператор взятия разности. Полученный в результате такого преобразования временной ряд $W_t = (1 - B)Z_t$ является стационарным и удовлетворяет модели MA(1). Идентификация такого преобразованного ряда не представляет трудности по коррелограммам, которые могут иметь один из двух видов, изображенных на рис. 4.4.

В общем случае после взятия разности мы можем получить любой стационарный процесс, описываемой моделью ARMA(p, q). Идентификация такого процесса осуществляется по виду коррелограмм, характерный вид которых для наиболее типичных ситуаций был описан в разделе 3.

Может оказаться также, хотя и намного реже, что после взятия разности полученный ряд $W_t = (1 - B)Z_t$ все еще будет нестационарным, что можно проверить по коррелограммам, сохраняющим вид, изображенный на рис. 5.1. В этом случае необходимы еще одно взятие разности и переход к ряду $\tilde{W}_t = (1 - B)W_t = (1 - B)^2 Z_t$. Если этого будет достаточно, т.е. ряд $\tilde{W}_t$ окажется стационарным, то в модели ARIMA(p, d, q) значение показателя оператора взятия разности равно $d = 2$. В очень редких случаях может потребоваться дополнительное взятие разностей. После достижения стационарности процесс идентифицируется моделью ARMA(p, q).

5.2. НЕСТАЦИОНАРНОСТЬ ПО ОТНОШЕНИЮ К ДИСПЕРСИИ И АВТОКОВАРИАЦИИ

5.2.1. НЕСТАЦИОНАРНОСТЬ ПО ДИСПЕРСИИ И АВТОКОВАРИАЦИИ ПРОЦЕССА ARIMA

Процесс, стационарный в среднем, не обязательно стационарен по дисперсии и автокорреляции. Однако процесс, нестационарный в среднем, чаще всего будет нестационарным по дисперсии и автокорре-

ляции. Как было показано в предыдущем разделе, математическое ожидание ARIMA-модели зависит от времени. Покажем теперь, что ARIMA-модель генерирует процесс, являющийся нестационарным также по дисперсии и автоковариации.

Отметим сначала одно фундаментальное свойство ARIMA-модели. Хотя модель нестационарна, процесс для любого момента времени полностью определяется лишь конечным числом параметров, а именно: ϕ_i, θ_j и σ_a^2. Таким образом, все будущее развитие процесса может быть предсказано на основе ARIMA-модели, удовлетворяющей исходному множеству данных $\{Z_1, Z_2, \ldots, Z_n\}$, составляющему временной ряд.

Например, пусть мы подгоняем данные под модель IMA(1, 1):

$$(1-B)Z_t = (1-\theta B)a_t,$$

или

$$Z_t = Z_{t-1} + a_t - \theta a_{t-1},$$

имея временной ряд из n_0 наблюдений. Относительно начала времени n_0, для $t > n_0$ можно записать

$$Z_t = Z_{t-1} + a_t - \theta a_{t-1} = Z_{t-2} + a_t + (1-\theta)a_{t-1} - \theta a_{t-2} =$$

$$= Z_{n_0} + a_t + (1-\theta)a_{t-1} + \ldots + (1-\theta)a_{n_0+1} - \theta a_{n_0},$$

где Z_{n_0} и a_{n_0} – известные значения относительно начала времени n_0.

Значит, для начала времени n_0 имеем следующие выражения для дисперсии и ковариации:

$$\mathrm{Var}(Z_t) = [1 + (t - n_0 - 1)(1-\theta)^2]\sigma_a^2,$$

$$\mathrm{Var}(Z_{t-k}) = [1 + (t - k - n_0 - 1)(1-\theta)^2]\sigma_a^2,$$

$$\mathrm{Cov}(Z_{t-k}, Z_t) = [1 + (t - k - n_0 - 1)(1-\theta)^2]\sigma_a^2,$$

$$\mathrm{Corr}(Z_{t-k}, Z_t) = \frac{\mathrm{Cov}(Z_{t-k}, Z_t)}{\sqrt{\mathrm{Var}(Z_{t-k})\mathrm{Var}(Z_t)}} =$$

$$= \frac{(1 + (t - k - n_0 - 1)(1-\theta)^2}{\sqrt{[1 + (t - k - n_0 - 1)(1-\theta)^2][1 + (t - n_0 - 1)(1-\theta)^2]}}.$$

На основании полученных формул можно сделать несколько важных наблюдений.

1. Дисперсия $\mathrm{Var}(Z_t)$ процесса ARIMA зависит от времени и $\mathrm{Var}(Z_t) \neq \mathrm{Var}(Z_{t-k})$ для $k \neq 0$.

2. Дисперсия $\mathrm{Var}(Z_t)$ неограниченна при $t \to \infty$.

3. Автоковариация $\mathrm{Cov}(Z_{t-k}, Z_t)$ и автокорреляция $\mathrm{Corr}(Z_{t-k}, Z_t)$ зависят от времени (т.е. они являются функциями не только k, но и t, и исходной точки n_0).

4. Если n_0 велико относительно k, то $\mathrm{Corr}(Z_{t-k}, Z_t) \approx 1$.

Поскольку $\left|\mathrm{Corr}(Z_{t-k}, Z_t)\right| \leq 1$, то автокорреляционная функция убывает медленно при увеличении k.

В общем случае по одной реализации трудно или невозможно сделать статистический вывод для процесса, нестационарного и по среднему, и по автоковариационной (автокорреляционной) функции. К счастью, для однородно-нестационарных процессов мы можем применить взятие разности и свести их к стационарным. И хотя исходный ряд Z_t нестационарен, ряд $W_t = (1-B)^d Z_t$ стационарен и может быть представлен моделью ARMA(p, q):

$$\phi(B)W_t = \theta(B)a_t,$$

где $\phi(B) = (1 - \phi_1 B - \ldots - \phi_p B^p)$ и $\theta(B) = (1 - \theta_1 B - \ldots - \theta_q B^q)$, и все их корни лежат вне единичного круга. Таким образом, параметры ϕ_i, θ_j и σ_a^2, которые контролируют эволюцию процесса Z_t, могут быть оценены из W_t точно таким же способом, как это делается для стационарных процессов.

5.2.2. ПРЕОБРАЗОВАНИЯ, СТАБИЛИЗИРУЮЩИЕ ДИСПЕРСИЮ

Не все нестационарные проблемы могут быть исправлены взятием разности. Существует много временных рядов, стационарных в среднем, но нестационарных по дисперсии. Чтобы учесть подобный вид нестационарности, используют так называемые преобразования, стабилизирующие дисперсию.

Типичная ситуация для нестационарного процесса – изменение дисперсии с изменением уровня μ_t ряда. Таким образом,

$$\mathrm{Var}(Z_t) = Cf(\mu_t)$$

для некоторой положительной константы C и функции f. Задача, следовательно, состоит в том, чтобы найти преобразование T исходного нестационарного по дисперсии процесса Z_t, при этом процесс $T(Z_t)$ должен иметь постоянную дисперсию.

Для нахождения такого преобразования аппроксимируем желаемую функцию рядом Тейлора в точке μ_t, ограничившись членом первого порядка:

$$T(Z_t) \simeq T(\mu_t) + T'(\mu_t)(Z_t - \mu_t),$$

где $T'(\mu_t)$ – первая производная функции $T(Z_t)$, вычисленная в точке μ_t. Далее, вычислим дисперсию

$$\mathrm{Var}[T(Z_t)] = [T'(\mu_t)]^2 \mathrm{Var}(Z_t) = C[T'(\mu_t)]^2 f(\mu_t).$$

Таким образом, чтобы дисперсия преобразованного ряда $T(Z_t)$ была постоянной (равной C), необходимо выбрать T-преобразование так, чтобы

$$T'(\mu_t) = \frac{1}{\sqrt{f(\mu_t)}};$$

после интегрирования получаем $T(\mu_t) = \int \frac{1}{\sqrt{f(\mu_t)}} d\mu_t$.

Рассмотрим несколько конкретных примеров преобразований, стабилизирующих дисперсию для заданных зависимостей дисперсии исходного ряда от уровня μ_t.

1. Если, например, $\mathrm{Var}(Z_t) = C^2\mu_t^2$, то

$$T(\mu_t) = \int \frac{1}{\sqrt{\mu_t^2}} d\mu_t = \ln(\mu_t).$$

В этом случае логарифмическое преобразование даст постоянную дисперсию преобразованного ряда $\ln(Z_t)$.

2. Если $\mathrm{Var}(Z_t) = C\mu_t$, то

$$T(\mu_t) = \int \frac{1}{\sqrt{\mu_t}} d\mu_t = 2\sqrt{\mu_t},$$

т.е. преобразование $\sqrt{Z_t}$ стабилизирует дисперсию ряда, являющуюся линейной функцией тренда.

3. Если $\mathrm{Var}(Z_t) = C^2\mu_t^4$, то

$$T(\mu_t) = \int \frac{1}{\sqrt{\mu_t{}^4}} d\mu_t = -\frac{1}{\mu_t},$$

т.е. в данном случае для стабилизации дисперсии применяется преобразование $\frac{1}{Z_t}$.

В общем случае для стабилизации дисперсии можно использовать следующее степенное преобразование, известное под названием преобразование Бокса–Кокса:

$$T(Z_t) = Z_t^{(\lambda)} = \frac{Z_t^{\lambda} - 1}{\lambda},$$

где λ – параметр преобразования.

Некоторые наиболее часто используемые значения λ и связанные с ними преобразования даны в табл. 5.1.

Таблица 5.1

Основные преобразования, стабилизирующие дисперсию

Значение λ	Преобразование
–1,0	$\frac{1}{Z_t}$
–0,5	$\frac{1}{\sqrt{Z_t}}$
0,0	$\ln(Z_t)$
0,5	$\sqrt{Z_t}$
1,0	Z_t (нет преобразования)

Чтобы убедиться, что значение $\lambda = 0$ соответствует логарифмическому преобразованию, заметим, что

$$\lim_{\lambda \to 0} T(Z_t) = \lim_{\lambda \to 0} Z_t^{(\lambda)} = \lim_{\lambda \to 0} \frac{Z_t^{\lambda} - 1}{\lambda} = \lim_{\lambda \to 0} \frac{Z_t^{\lambda} \ln(Z_t)}{1} = \ln Z_t .$$

Завершая этот раздел, сделаем несколько важных замечаний.

1. Преобразования, стабилизирующие дисперсию, определены только для положительных рядов. Однако это не так важно, как может показаться, потому что к ряду всегда можно прибавить константу без влияния на структуру корреляции ряда.

2. Преобразование, стабилизирующее дисперсию, должно быть выполнено, если необходимо, до любого другого анализа (например, до взятия разности).

3. Параметр λ в степенном преобразовании можно рассматривать как параметр модели, оцениваемый по наблюдаемому ряду. Оценка максимального правдоподобия параметра λ минимизирует остаточную сумму квадратов. Для любого заданного значения λ остаточная сумма квадратов вычисляется из построенной модели. Оценка максимального правдоподобия дает минимальное значение суммы квадратов остатков среди всех значений λ. В статистических пакетах вычисление этих остаточных сумм часто производится на сетке значений λ (–1; –0,5; 0; 0,5; 1).

4. Очень часто преобразование Бокса–Кокса не только стабилизирует дисперсию, но и производит приближение процесса к нормальному распределению. Оптимальное значение параметра λ может быть найдено как минимизирующее $S(\lambda)$:

$$S(\lambda) = \sum_{t=1}^{n} (Z_t^{(\lambda)} - \hat{\mu})^2 ,$$

где $\hat{\mu}$ – выборочное среднее.

Значение $S(\lambda)$ вычисляется для λ_i = –1; –0,5; 0; 0,5; 1. Оптимальное λ_i, определяющее искомое преобразование, соответствует минимуму $S(\lambda)$.

6. ИДЕНТИФИКАЦИЯ МОДЕЛИ ARIMA

В анализе временных рядов самые ответственные этапы – идентификация и построение модели, описывающей имеющиеся данные. Предположим, что модель ищем в классе ARIMA(p, d, q). Идентификация – это наиболее решающий шаг при построении модели, заключающийся в определении конкретных значений p, d и q, выделяющих пробную модель из класса ARIMA(p, d, q). Идентификация требует хорошего знания особенностей однородно нестационарных моделей ARIMA и стационарных моделей ARMA в терминах характеристик АКФ и ЧАКФ. На практике АКФ и ЧАКФ неизвестны и оцениваются по данному временному ряду наблюдений $Z_1, Z_2, \ldots, Z_n$. Таким образом, при идентификации модели нам надо сравнить графики выборочных АКФ и ЧАКФ с теоретическими АКФ и ЧАКФ для ARMA-моделей конкретных порядков, чтобы выбрать порядки p, d и q, наиболее соответствующие наблюдаемым данным.

6.1. ЭТАПЫ ИДЕНТИФИКАЦИИ МОДЕЛИ ARIMA

Рассмотрим общую модель ARIMA(p,d,q):

$$(1 - \phi_1 B - \ldots - \phi_p B^p)(1 - B)^d Z_t =$$

$$= \theta_0 + (1 - \theta_1 B - \ldots - \theta_q B^q) a_t .$$

Для получения пробной модели, параметры которой в дальнейшем подвергаются оцениванию, необходимо пройти несколько этапов.

6.1.1. ЭТАП 1. ПОСТРОЕНИЕ ГРАФИКА РЯДА И ИСПОЛЬЗОВАНИЕ ПРЕОБРАЗОВАНИЯ, СТАБИЛИЗИРУЮЩЕГО ДИСПЕРСИЮ

В самом начале исследования рекомендуется построить график временного ряда. Из графика, тщательно изучив его, можно извлечь очень ценную информацию и предположить, что ряд содержит тренд, сезонность, выбросы, непостоянные дисперсии и другие ненормальные и нестационарные явления. Часто наблюдение графика помогает определить возможное преобразование данных.

В анализе временных рядов чаще всего используются два типа преобразований: преобразование, стабилилизируещее дисперсию, и взятие

конечного числа разностей. Так как взятие разности может привести к отрицательным значениям ряда, для которых такие преобразования, как логарифм и квадратный корень, не могут быть применены, то сначала применяют преобразования, стабилизирующие дисперсию. Ряд с непостоянной дисперсией часто требует логарифмирования. В более общем случае используется степенное преобразование Бокса–Кокса, определяемое следующей формулой:

$$Z_t^{(\lambda)} = \frac{Z_t^{\lambda} - 1}{\lambda}.$$

Для различных значений $-1 \le \lambda \le 1$ мы получаем различные преобразования. Самое сильное – обратное преобразование $\frac{1}{Z_t}$ – соответствует значению $\lambda = -1$; самое слабое – тождественное преобразование (отсутствие преобразования) – соответствует $\lambda = 1$. Промежуточные значения $\lambda = -0{,}5; 0; 0{,}5$ соответствуют преобразованиям $\frac{1}{\sqrt{Z_t}}$, $\ln Z_t$ и $\sqrt{Z_t}$ соответственно.

Нестационарность дисперсии чаще всего проявляется в увеличении разброса относительно среднего значения при увеличении уровня ряда. В этом случае, чем больше проявляется указанная нестационарность, тем более сильное преобразование, стабилизирующее дисперсию, необходимо использовать. Усиление преобразования соответствует уменьшению значения λ от 1 до –1. Так, преобразование $\sqrt{Z_t}$ – более слабое, чем $\ln Z_t$, а $\ln Z_t$ – более слабое, чем $\frac{1}{\sqrt{Z_t}}$.

Сила используемого преобразования контролируется визуально. Для начала можно попробовать использовать логарифмическое преобразование. Ряд, полученный в результате преобразования, вновь визуализируется. Если оказывается, что преобразование привело к некоторому уменьшению разброса с увеличением уровня ряда, необходимо попробовать применить более слабое преобразование. Если, наоборот, преобразованный ряд еще выявляет признаки увеличения дисперсии с увеличением уровня, необходимо ослабить преобразование.

Пусть на первом этапе выбрано некоторое преобразование $T(Z_t)$, в результате которого дисперсия стабилизирована, т.е. $\mathrm{Var}(T(Z_t)) = \mathrm{const}$. С этим преобразованным рядом переходим ко второму этапу идентификации.

6.1.2. ЭТАП 2. ПРИВЕДЕНИЕ ОДНОРОДНО-НЕСТАЦИОНАРНОГО РЯДА К СТАЦИОНАРНОМУ С ПОМОЩЬЮ ОПЕРАЦИИ ВЗЯТИЯ РАЗНОСТИ

Если ряд однородно нестационарный, на этом этапе его необходимо привести к стационарному ряду с использованием операции взятия разности. Подтверждением однородной нестационарности, т.е. необходимости взятия разности, служит определенное поведение выборочных АКФ и ЧАКФ исходного ряда. Поэтому на данном этапе идентификации производится несколько итераций, на каждой из которых вычисляются выборочные АКФ и ЧАКФ для дальнейшего подтверждения числа взятия разностей.

Итак, если в построенных коррелограммах выборочная АКФ убывает очень медленно (при этом отдельные значения АКФ могут не быть велики), а выборочная ЧАКФ «обнуляется» уже после первого лага, т.е. для $k > 1$ (как на рис. 5.1), то имеет место однородная нестационарность и необходимо взятие разности. В результате получаем ряд $(1 - B)T(Z_t)$.

В более общем случае, чтобы убрать нестационарность, может понадобиться взятие разности более высокого порядка $(1 - B)^d T(Z_t)$ для $d > 1$. В большинстве случаев при идентификации временных рядов d принимает значение 0, 1 или 2. Тем не менее некоторые авторы считают, что взять разность более чем нужное число раз менее опасно, чем взять разность меньшее, чем необходимо, число раз. Однако даже в этом случае следует избегать излишнего усложнения параметризации.

Таким образом, в результате проведения второго этапа мы определяем значение первичного параметра d и получаем стационарный ряд $W_t = (1 - B)^d T(Z_t)$, который в дальнейшем исследуется на третьем этапе.

6.1.3. ЭТАП 3. ИДЕНТИФИКАЦИЯ СТАЦИОНАРНОГО ВРЕМЕННОГО РЯДА

На данном этапе необходимо произвести идентификацию стационарного временного ряда W_t, т.е. определить порядки p и q полино-

мов авторегрессии $\phi_p(B) = 1 - \phi_1 B - ... - \phi_p B^p$ и скользящего среднего $\theta_q(B) = 1 - \theta_1 B - ... - \theta_q B^q$ модели ARMA (p,q):

$$\phi_p(B)\dot{W}_t = \theta_q(B)a_t .$$

Идентификация порядков p и q производится опять с использованием выборочной АКФ и ЧАКФ ряда W_t. На практике требуемые значения этих параметров обычно не превышают 3. В табл. 6.1 суммированы основные идентификационные характеристики стационарных процессов авторегрессии, скользящего среднего и смешанного процесса.

Анализируя табл. 6.1, полезно и интересно отметить строгую двойственность между AR- и MA-моделями через свойственные им характеристики АКФ и ЧАКФ. Заметим также, что свойство «обнуления» выборочных АКФ и ЧАКФ, соответствующее равенству нулю теоретических АКФ и ЧАКФ, устанавливается в результате непревышения выборочными аналогами некоторых пороговых значений, ниже которых они считаются статистически незначимыми. Пороговые значения рекомендуется устанавливать $\pm 1{,}5\ S_{\hat{\rho}_k}$ для АКФ и $\pm 1{,}5\ S_{\hat{\Phi}_{kk}}$ для ЧАКФ, где стандартные ошибки выборочной АКФ $\hat{\rho}_k$ и выборочной ЧАКФ $\hat{\Phi}_{kk}$ определяются по формулам:

$$S_{\hat{\rho}_k} \simeq \sqrt{\frac{1}{n}\left(1 + 2\hat{\rho}_1^2 + ... + 2\hat{\rho}_{k-1}^2\right)},$$

$$S_{\hat{\Phi}_{kk}} \simeq \sqrt{\frac{1}{n}} .$$

Таблица 6.1

Идентификация стационарных временных рядов

Процесс	АКФ	ЧАКФ
AR(p)	Экспонента или затухающая синусоида	«Обнуляется» при $k > p$
MA(q)	«Обнуляется» при $k > q$	Экспонента или затухающая синусоида
ARMA(p, q)	«Имеет хвост» после лага $k = (q - p)$	«Имеет хвост» после лага $k > (p - q)$

Чтобы построить приемлемую ARIMA-модель, в идеале необходимо иметь выборку как минимум $n = 50$ наблюдений. При этом число значений выборочных АКФ и ЧАКФ не превышает $n/4$, так как последующие значения вычисляются с большой погрешностью. Хотя иногда для данных хорошего качества может оказаться возможным идентифицировать адекватную модель и с меньшим объемом выборки.

Более точная идентификация порядков p и q может быть проведена из сравнения вычисленных по выборке и построенных в виде коррелограмм выборочных АКФ и ЧАКФ с теоретическими АКФ и ЧАКФ известных моделей. Идентификация осуществляется по принципу схожести выборочных и теоретических коррелограмм. Искусство аналитика по идентификации временных рядов очень похоже на расследование преступлений. Чаще всего преступники маскируются, чтобы не быть узнанными. Точно так же выборочные АКФ и ЧАКФ часто маскируют теоретические АКФ и ЧАКФ, скрывающиеся за случайным компонентом временного ряда.

На начальном этапе идентификации мы концентрируемся на самых общих характеристиках АКФ и ЧАКФ процесса, не вдаваясь в детали, поэтому можем сразу не распознать истинную модель. Усовершенствование модели достигается на более поздних итерациях построения модели путем ее последовательного уточнения и диагностической проверки.

6.1.4. ЭТАП 4. ПРОВЕРКА НАЛИЧИЯ ДЕТЕРМИНИРОВАННОГО ТРЕНДА

Последний этап идентификации временного ряда – проверка необходимости включения детерминированного тренда θ_0 при $d > 0$ в модель, полученную к данному этапу:

$$\phi_p(B)(1-B)^d T(Z_t) = \theta_0 + \theta_q(B)a_t .$$

Как отмечалось ранее, параметр θ_0 обычно опускается, чтобы модель допускала представление временных рядов со случайными изменениями в уровне, наклоне или в тренде. Однако если есть основание считать, что стационарный ряд W_t, полученный взятием разности, содержит ненулевое среднее, можно проверить включение детерминированного тренда в модель, если сравнить выборочное среднее $\bar{W}$ ряда $W_t = (1-B)^d T(Z_t)$ с его приближенной стандартной ошибкой $S_{\bar{W}}$.

Если $\bar{W} < S_{\bar{W}}$, то включение детерминированного тренда не требуется. В противном случае есть основание включить θ_0 в модель.

Оценки $S_{\bar{W}}$ получаются для отдельных классов моделей. Например, для модели ARIMA(1, d, 0), имеющей вид $(1-\phi_1 B)W_t = a_t$, требуемая стандартная ошибка определяется выражением $S_{\bar{W}} = \sqrt{\frac{\hat{\gamma}_0}{n}\left(\frac{1+\hat{\rho}_1}{1-\hat{\rho}_1}\right)}$. Выражения $S_{\bar{W}}$ для других моделей можно получить аналогично. Однако на этапе идентификации, так как вид модели пока неизвестен, большинство программного обеспечения по анализу данных использует следующую аппроксимацию:

$$S_{\bar{W}} = \left[\frac{\hat{\gamma}_0}{n}(1 + 2\hat{\rho}_1 + 2\hat{\rho}_2 + ... + 2\hat{\rho}_k)\right]^{\frac{1}{2}},$$

где $\hat{\gamma}_0$ – выборочная дисперсия, $\hat{\rho}_1,..., \hat{\rho}_k$ – первые k значимых выборочных АКФ, вычисленных для стационарного процесса $\{W_t\}$.

В настоящее время в связи с общедоступностью и распространенностью программного обеспечения анализа данных для определения необходимости включения в модель детерминированного тренда используется альтернативный, существенно более простой, подход. Первоначально детерминированный тренд θ_0 включается в модель, после чего оцениваются ее параметры. Если оценка $\hat{\theta}_0$ оказывается статистически незначимой, то детерминированный тренд исключается из модели и проводятся повторное оценивание, проверка адекватности модели и т.д.

7. РАСШИРЕНИЕ МЕТОДОЛОГИИ ARIMA ДЛЯ АНАЛИЗА СЕЗОННЫХ ВРЕМЕННЫХ РЯДОВ

Многие временные ряды в социальной и экономической сферах содержат явление сезонности, которое проявляется в повторении тенденции изменения данных через некоторый регулярный период време-

ни. Наименьший период времени s, через который тенденция повторяется, называется периодом сезонного ряда. Например, ежеквартальный временной ряд продажи мороженого имеет наивысший уровень летом ($s = 4$); ежемесячные объемы продаж автомобилей каждый год уменьшаются в августе, сентябре в связи с переходом к новым моделям, а ежемесячный объем продаж игрушек увеличивается к декабрю. Период сезонности в последних двух случаях равен $s = 12$. Сезонное явление может быть следствием таких факторов, как, например, погодные изменения в течение года, влияющие на экономическую деятельность.

Пусть ряд Z_t – сезонный с периодом сезонности s. Чтобы проанализировать данный ряд, полезно организовать данные в таблицу Бэйса–Бэлота, имеющую вид табл. 7.1. Здесь $T_{j.}$ – сумма для j-го сезона, $j = \overline{1,s}$; $T_{.i}$ – сумма для i-го периода, $i = \overline{1,n}$; $\bar{Z}_{j.}$ – среднее для j-го сезона, $j = \overline{1,s}$; $\bar{Z}_{.i}$ – среднее для i-го периода, $i = \overline{1,n}$; T – общая сумма ряда. Располагая данные временного ряда в таблицу Бэйса–Бэлота, можно увидеть зависимости не только в строках этой таблицы, но и в столбцах. Зависимости в столбцах как раз и указывают на присутствующую сезонность временного ряда.

Таблица 7.1

Таблица Бэйса–Бэлота для представления сезонного ряда

	1	2	3	...	s	Сумма	Среднее
1	Z_1	Z_2	Z_3	...	Z_s	$T_{.1}$	$\bar{Z}_{.1}$
2	Z_{s+1}	Z_{s+2}	Z_{s+3}	...	Z_{2s}	$T_{.2}$	$\bar{Z}_{.2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
n	$Z_{(n-1)s+1}$	$Z_{(n-1)s+2}$	$Z_{(n-1)s+3}$	...	Z_{ns}	$T_{.n}$	$\bar{Z}_{.n}$
Сумма	$T_{1.}$	$T_{2.}$	$T_{3.}$	...	$T_{s.}$	T	T/s
Среднее	$\bar{Z}_{1.}$	$\bar{Z}_{2.}$	$\bar{Z}_{3.}$	...	$\bar{Z}_{s.}$	T/n	T/ns

7.1. ТРАДИЦИОННЫЕ МЕТОДЫ УЧЕТА СЕЗОННОСТИ ВРЕМЕННОГО РЯДА. РЕГРЕССИОННЫЙ МЕТОД

Традиционно временной ряд, содержащий сезонность, рассматривают как смесь тренд-циклического (P_t), сезонного (S_t) и нерегулярного компонентов (e_t). Предположим, что эти компоненты независимы и аддитивны, тогда можно записать

$$Z_t = P_t + S_t + e_t .$$

Для оценивания этих компонентов в литературе предлагаются некоторые методы декомпозиции.

Для *регрессионного метода* аддитивный сезонный временной ряд представляется в виде следующей регрессионной модели:

$$Z_t = P_t + S_t + e_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i U_{it} + \sum_{j=1}^{k} \beta_j V_{jt} + e_t ,$$

где $P_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i U_{it}$ и U_{it} – тренд-циклические переменные, $S_t = \sum_{j=1}^{k} \beta_j V_{jt}$ и V_{jt} – сезонные переменные. Например, линейный тренд-циклический компонент P_t может быть записан так:

$$P_t = \alpha_0 + \alpha_1 t .$$

В более общем случае тренд-циклический компонент P_t может быть представлен полиномом m-го порядка по времени:

$$P_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i t^i .$$

Аналогичным образом, сезонный компонент S_t описывается как линейная комбинация сезонных фиктивных (индикаторных) переменных или как линейная комбинация функций синус и косинус с различными частотами. Например, сезонный компонент с периодом s может быть записан в виде

$$S_t = \sum_{j=1}^{s-1} \beta_j D_{jt} ,$$

где $D_{jt} = \begin{cases} 1, \text{ если } t \text{ соответствует сезонному периоду } j, \\ 0, \text{ в противном случае.} \end{cases}$

Заметим, что, если сезонный период равен s, требуется лишь $s-1$ сезонных фиктивных переменных. Другими словами, β_s полагается равным нулю, так что коэффициент β_j, $j \neq s$, представляет сезонный эффект j-го периода по сравнению с периодом s.

Альтернативно сезонный компонент S_t может быть записан в виде

$$S_t = \sum_{j=1}^{s/2} \left[\beta_j \sin\left(\frac{2\pi j}{s}\right) + \gamma_j \cos\left(\frac{2\pi j}{s}\right) \right],$$

где $[s/2]$ – целая часть $s/2$. С учетом введенного вида сезонного компонента модель временного ряда принимает одну из следующих форм:

$$Z_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i t^i + \sum_{j=1}^{s-1} \beta_j D_{jt} + e_t ,$$

или $$Z_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i t^i + \sum_{j=1}^{[s/2]} \left[\beta_j \sin\left(\frac{2\pi j}{s}\right) + \gamma_j \cos\left(\frac{2\pi j}{s}\right) \right] + e_t .$$

Для заданного набора данных Z_t и заданных значений m и s при получении оценок $\hat{\alpha}_i$, $\hat{\beta}_j$, $\hat{\gamma}_j$ неизвестных параметров α_i, β_j, γ_j можно применять *стандартный метод наименьших квадратов*. Оценки тренд-циклического, сезонного и нерегулярного компонентов для модели с фиктивными сезонными переменными задаются следующим образом:

$$\hat{P}_t = \hat{\alpha}_0 + \sum_{i=1}^{m} \hat{\alpha}_i t^i ,$$

$$\hat{S}_t = \sum_{j=1}^{s-1} \hat{\beta}_j D_{jt}$$

и

$$\hat{e}_t = Z_t - \hat{P}_t - \hat{S}_t .$$

Для модели с сезонными переменными синус и косинус соответствующие оценки тренд-циклического, сезонного и нерегулярного компонентов задаются так:

$$\hat{P}_t = \hat{\alpha}_0 + \sum_{i=1}^{m} \hat{\alpha}_i t^i ,$$

$$\hat{S}_t = \sum_{j=1}^{s/2} \left[\hat{\beta}_j \sin\left(\frac{2\pi j}{s}\right) + \hat{\gamma}_j \cos\left(\frac{2\pi j}{s}\right) \right],$$

и

$$\hat{e}_t = Z_t - \hat{P}_t - \hat{S}_t .$$

7.2. СЕЗОННЫЕ ARIMA-МОДЕЛИ

Традиционные методы, представленные в предыдущем разделе, основаны на предположении о том, что сезонный компонент является детерминированным (неслучайным) и не зависит от других несезонных компонентов. Однако многие временные ряды не ведут себя так хорошо. Более вероятно, что сезонный компонент является стохастическим и коррелированным с несезонными компонентами. В этом случае лучшее решение – расширение ARIMA-модели, введенной в предыдущих разделах, на случай сезонного временного ряда.

Расположение данных в таблице Бэйса–Бэлота (табл. 7.1) позволяет заметить, что в общем случае сезонного временного ряда $\{Z_t\}$ имеются связи не только внутри периодов (по строкам таблицы), но и между периодами (по столбцам таблицы). Внутри периодов связи представлены корреляциями между значениями в строке $\ldots, Z_{t-2}, Z_{t-1}, Z_t, Z_{t+1}, Z_{t+2}, \ldots$, а между периодами – корреляциями между значениями в столбце $\ldots, Z_{t-2s}, Z_{t-s}, Z_t, Z_{t+s}, Z_{t+2s}, \ldots$.

Предположим, мы не знаем, что ряд $\{Z_t\}$ содержит сезонные вариации между периодами, и подгоним для этого ряда несезонную ARIMA-модель, введенную ранее, т.е.

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B) b_t . \qquad (7.1)$$

Очевидно, что ряд $\{b_t\}$ не будет белым шумом, так как он содержит неучтенную (сезонную) корреляцию между периодами. Пусть

$$\rho_{js} = \frac{E(b_{t-js} - \mu_b)(b_t - \mu_b)}{\sigma_b^2}, \quad j = 1,2,3,\ldots$$

есть автокорреляционная функция для ряда $\{b_t\}$, представляющая неучтенные связи между периодами. Представим эти связи такой же ARIMA-моделью, но отличающейся длиной запаздывания (B^s вместо B):

$$\Phi_P(B^s)(1-B^s)^D b_t = \Theta_Q(B^s) a_t , \qquad (7.2)$$

где

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps},$$

$$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \ldots - \Theta_Q B^{Qs}$$

есть полиномы от аргумента B^s, не имеющие общих корней. Корни этих полиномов лежат вне единичного круга, и $\{a_t\}$ – процесс белого шума с нулевым математическим ожиданием.

Для иллюстрации предположим, что в модели (7.2) $P = 1$, $s = 12$, $D = 0$, $Q = 0$. В результате получаем модель

$$(1 - \Phi B^{12}) b_t = a_t .$$

Данная модель похожа на модель авторегрессии первого порядка AR(1), имеющую вид $(1 - \phi B) Z_t = a_t$ с той лишь разницей, что оператор запаздывания B заменяется оператором B^{12}. Поэтому АКФ и ЧАКФ для данного процесса ведут себя аналогичным модели AR(1) образом (с поправкой на лаговый оператор B^{12}): АКФ экспоненциально убывает по лагам, кратным 12, а ЧАКФ везде равна нулю, кроме лага $k = 12$. Например, для $\Phi = 0.8$ теоретические АКФ и ЧАКФ процесса $\{b_t\}$ представлены на рис. 7.1 и 7.2.

Если же процесс $\{b_t\}$ удовлетворяет сезонной модели, для которой $P = 0$, $s = 12$, $D = 0$, $Q = 1$:

$$b_t = (1 - \Theta B^{12})a_t,$$

то она похожа на модель скользящего среднего первого порядка МА(1), имеющую вид $Z_t = (1 - \theta B)a_t$ с заменой B на B^{12}. Для этого процесса, наоборот, ЧАКФ экспоненциально убывает по лагам, кратным 12, а АКФ везде равна нулю, кроме лага $k = 12$.

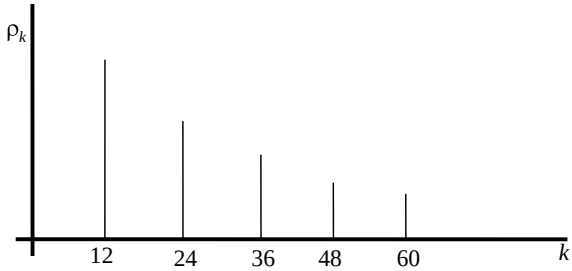


Рис. 7.1. АКФ процесса $(1 - 0.8B^{12})b_t = a_t$

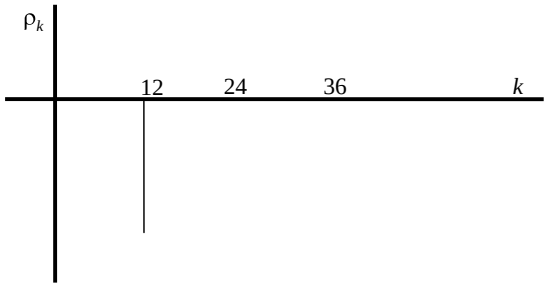


Рис. 7.2. ЧАКФ процесса $(1 - 0.8B^{12})b_t = a_t$

Умножая обе части модели (7.1) на $\Phi_P(B^s)(1 - B^s)^D$ и заменяя в правой части полученного уравнения $\Phi_P(B^s)(1 - B^s)^D b_t$ на $\Theta_Q(B^s)a_t$ согласно уравнению (7.2), получаем мультипликативную сезонную ARIMA-модель Бокса–Дженкинса

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D\dot{Z}_t = \theta_q(B)\Theta_Q(B^s)a_t, \qquad (7.3)$$

где $\dot{Z}_t = \begin{cases} Z_t - \mu, \text{ если } d = D = 0, \\ Z_t, \text{ в противном случае.} \end{cases}$

Полиномы $\phi_p(B)$ и $\theta_q(B)$ называются регулярными полиномами авторегрессии и скользящего среднего, а полиномы $\Phi_P(B^s)$ и $\Theta_Q(B^s)$ – сезонными полиномами авторегрессии и скользящего среднего соответственно. Мультипликативную модель (7.3) обозначают ARIMA $(p,d,q)\times(P,D,Q)_s$, где индекс s – сезонный период.

Идентификация сезонных временных рядов – более сложная процедура, чем идентификация несезонных временных рядов. Она базируется в основном на анализе АКФ, так как ЧАКФ для сезонной модели имеет существенно более сложный вид. В общем случае сезонные, как и несезонные компоненты авторегрессии дают эффект «обнуления» для сезонных и несезонных лагов ЧАКФ. С другой стороны, сезонные и несезонные компоненты скользящего среднего дают экспоненциальный или затухающий синусоидальный вид ЧАКФ для сезонных и несезонных запаздываний.

Так как идентификация сезонной модели построена в основном на анализе АКФ, рассмотрим вид АКФ для классической модели авиаперевозок.

7.3. ВЫЧИСЛЕНИЕ АВТОКОРРЕЛЯЦИОННОЙ ФУНКЦИИ ДЛЯ МОДЕЛИ АВИАПЕРЕВОЗОК

Рассмотрим модель ARIMA $(0,1,1)\times(0,1,1)_{12}$

$$(1-B)(1-B^{12})\dot{Z}_t = (1-\theta B)(1-\Theta B^{12})a_t .$$

Эта модель оказывается применимой для представления множества сезонных временных рядов, таких как данные авиалиний, временные ряды в торговле и т.п. Модель была впервые введена Дж. Боксом и Г. Дженкинсом для представления данных международных авиаперевозок. Отсюда в литературе ее часто называют моделью авиаперевозок.

Если мы построим АКФ временного ряда по данным, сгенерированным данной моделью, то она будет выявлять ярко выраженную нестационарность. Причем нестационарность имеет не только регуляр-

ный, но и сезонный характер. Как отличить один вид нестационарности от другого, будет рассмотрено нами позже. Сейчас же предположим, что мы применили оператор взятия разности $(1-B)$, затем – оператор взятия сезонной разности $(1-B^{12})$, после чего получили ряд

$$W_t = (1-B)(1-B^{12})\dot{Z}_t,$$

соответствующий модели скользящего среднего:

$$W_t = (1-\theta B)(1-\Theta B^{12})a_t.$$

Для полученного процесса W_t, включающего только скользящее среднее, легко найти выражение АКФ, используя автоковариационную генерирующую функцию:

$$\gamma\ B = \sigma_a^2\theta(B)\Theta(B^{12})\theta(B^{-1})\Theta(B^{-12}) =$$

$$= \sigma_a^2(1-\theta B)(1-\Theta B^{12})(1-\theta B^{-1})(1-\Theta B^{-12}) =$$

$$= \sigma_a^2\{\theta\Theta B^{-13} - \Theta(1+\theta^2)B^{-12} + \theta\Theta B^{-11} -$$

$$-\theta(1+\Theta^2)B^{-1} + (1+\theta)(1+\Theta^2) -$$

$$-\theta(1+\Theta^2)B + \theta\Theta B^{11} - \Theta(1+\theta^2)B^{12} + \theta\Theta B^{13}\}.$$

Отсюда получаем выражения для автоковариаций:

$$\gamma_0 = (1+\theta^2)(1+\Theta^2)\sigma_a^2,$$

$$\gamma_1 = -\theta(1+\Theta^2)\sigma_a^2,$$

$$\gamma_{11} = \theta\Theta\sigma_a^2,$$

$$\gamma_{12} = -\Theta(1+\theta^2)\sigma_a^2,$$

$$\gamma_{13} = \theta\Theta\sigma_a^2,$$

$\gamma_j = 0$, в противном случае.

Следовательно, автокорреляции имеют вид

$$\rho_1 = \frac{-\theta}{(1+\theta^2)},$$

$$\rho_{11} = \frac{\theta\Theta}{(1+\theta^2)(1+\Theta^2)} = \rho_{13},$$

$$\rho_{12} = \frac{-\Theta}{(1+\Theta^2)},$$

$\rho_j = 0$, в противном случае.

Отсюда следует, что для рассмотренной модели авиаперевозок ненулевыми являются только автоковариации с лагами 1, 11, 12, 13, остальные автоковариации равны нулю. Для более общего случая сезонной модели с периодом s ненулевыми оказываются автоковариации с лагами $1, s-1, s, s+1$, остальные автоковариации равны нулю. Этот факт может лежать в основе идентификации временного ряда, удовлетворяющего данной модели.

7.4. ОБЩАЯ СЕЗОННАЯ ARIMA-МОДЕЛЬ

В более общем случае может быть использована общая сезонная ARIMA-модель вида

$$\prod_{j=1}^{M} \phi_j(B) \prod_{i=1}^{K} (1 - B^{s_i})^{d_i} \dot{Z}_t = \theta_0 + \prod_{k=1}^{N} \theta_k(B) a_t .$$

Приведенная модель может содержать K факторов взятия разности, M факторов авторегрессии и N факторов скользящего среднего. Это расширение ARIMA-модели полезно при описании многих нестандартных временных рядов, которые, например, содержат смесь сезонных рядов с различными периодами.

В общей модели i-й фактор взятия разности $1 - B^{s_i}{}^{d_i}$ с порядком s_i (степенью B) и степенью d_i. Если $K = 0$, то $\dot{Z}_t = Z_t - \mu$. В противном случае $\dot{Z}_t = Z_t$. Параметр θ_0 представляет собой детер-

минированный тренд и рассматривается только при $K \neq 0$. Далее, j-й фактор авторегрессии

$$\phi_j(B) = 1 - \phi_{j1}B - \phi_{j2}B^2 - ... - \phi_{jp_j}B^{p_j}$$

содержит один или более параметров авторегрессии ϕ_{jm}, а k-й фактор скользящего среднего

$$\theta_k(B) = 1 - \theta_{k1}B - \theta_{k2}B^2 - ... - \theta_{kq_k}B^{q_k}$$

содержит один или более параметров скользящего среднего θ_{kn}.

В большинстве приложений параметры K, M и N обычно меньше либо равны 2.

Поскольку сезонная мультипликативная ARIMA-модель есть расширенная форма ARIMA-модели, оценивание параметров и прогнозирование можно осуществлять теми же методами, которые рассмотрены ранее.

8. ОЦЕНИВАНИЕ ПАРАМЕТРОВ МОДЕЛИ ARIMA

После идентификации модели ARIMA мы определяем значения первичных параметров p, d, q и, таким образом, получаем пробную модель, которую необходимо далее подогнать под данные. Поэтому следующий после идентификации шаг – оценивание параметров модели.

Модель ARIMA(p, d, q) для процесса Z_t

$$(1 - \phi_1 B - \ldots - \phi_p B^p)(1 - B)^d Z_t =$$

$$= \theta_0 + (1 - \theta_1 B - ... - \theta_q B^q)a_t \qquad (8.1)$$

не содержит дополнительных параметров по сравнению с моделью ARMA(p, q)

$$(1 - \phi_1 B - \ldots - \phi_p B^p)W_t = \theta_0 + (1 - \theta_1 B - ... - \theta_q B^q)a_t \qquad (8.2)$$

для процесса $W_t = (1-B)^d Z_t$. Поэтому оценивание параметров без потери общности может проводиться для процесса W_t, полученного в результате применения d операций взятия разности к исходному процессу Z_t.

Таким образом, методы оценивания параметров могут без потери общности годиться для модели $\mathrm{ARMA}(p,q)$. Рассмотрим оценивание параметров $\phi = (\phi_1, \phi_2, \ldots, \phi_p)^T$, $\mu = E(Z_t)$, $\theta = (\theta_1, \theta_2, \ldots, \theta_q)^T$ и $\sigma_a^2 = E(a_t^2)$ модели

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + \ldots + \phi_p \dot{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q},$$

где $\dot{Z}_t = Z_t - \mu$, $\{a_t\}$ – i.i.d. (identically and independently distributed – одинаково и независимо распределенный) процесс белого шума, подчиняющийся нормальному распределению $N(0, \sigma_a^2)$, по выборке, содержащей n наблюдений стационарного (изначально или в результате преобразований) временного ряда Z_t, $t = 1, \ldots, n$.

В данном разделе мы рассмотрим несколько широко используемых процедур оценивания параметров рассматриваемых моделей. Первый метод – метод моментов – применяется на начальных этапах исследования для получения приближенных оценок неизвестных параметров модели. Второй метод – метод максимального правдоподобия – считается наиболее точным методом оценивания временных параметров и может использоваться для получения конечных оценок параметров.

После применения процедуры оценивания параметров необходимо проверить адекватность модели для рассматриваемого временного ряда. Очень часто сразу несколько моделей могут адекватно представлять заданный временной ряд. В этом случае проводится диагностическая проверка, для которой в литературе предлагается несколько критериев, широко используемых для выбора модели временного ряда.

8.1. МЕТОД МОМЕНТОВ

Метод моментов заключается в подстановке выборочного среднего $\bar{Z}$, выборочной дисперсии $\hat{\gamma}_0$, значений выборочной АКФ $\hat{\rho}_i$ вместо их теоретических значений и в решении полученной в результате

системы уравнений относительно неизвестных параметров. Например, для процесса $\mathrm{AR}(p)$, заданного моделью

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + \ldots + \phi_p \dot{Z}_{t-p} + a_t,$$

оценкой математического ожидания $\mu = E(Z_t)$ является выборочное среднее $\bar{Z}$. При оценивании вектора параметров ϕ мы используем уравнение для автокорреляций $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots + \phi_p \rho_{k-p}$, полученное нами при рассмотрении модели $\mathrm{AR}(p)$ и справедливое для $k > 1$. Расписывая данное уравнение для $k = 1, 2, \ldots, p$, получаем систему уравнений Юла–Уокера:

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2 \rho_1 + \phi_3 \rho_2 + \ldots + \phi_p \rho_{p-1}, \\
\rho_2 &= \phi_1 \rho_1 + \phi_2 + \phi_3 \rho_1 + \ldots + \phi_p \rho_{p-2}, \\
&\vdots \\
\rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \phi_3 \rho_{p-3} + \ldots + \phi_p.
\end{aligned}$$

Заменяя в данной системе ρ_k на $\hat{\rho}_k$ и решая систему линейных уравнений, мы получаем оценки параметров $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p$ методом моментов:

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{bmatrix} = \begin{bmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \cdots & \hat{\rho}_{p-2} & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-3} & \hat{\rho}_{p-2} \\ \vdots & \vdots & & & & \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \hat{\rho}_{p-3} & \cdots & \hat{\rho}_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{bmatrix}.$$

Эти оценки обычно называют оценками Юла–Уокера.

После получения оценок $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p$ можно использовать уравнение

$$\gamma_0 = E(\dot{Z}_t \dot{Z}_t) = E[\dot{Z}_t (\phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + \ldots + \phi_p \dot{Z}_{t-p} + a_t)] =$$

$$= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \ldots + \phi_p \gamma_p + \sigma_a^2$$

для получения оценки дисперсии σ_a^2. Выражение для оценки дисперсии имеет вид

$$\hat{\sigma}_a^2 = \hat{\gamma}_0(1 - \hat{\phi}_1\hat{\rho}_1 - \hat{\phi}_2\hat{\rho}_2 - ... - \hat{\phi}_p\hat{\rho}_p).$$

Рассмотрим простые примеры применения метода моментов. Например, для модели авторегрессии AR(1), имеющей вид

$$(Z_t - \mu) = \phi_1(Z_{t-1} - \mu) + a_t,$$

оценка Юла–Уокера для параметра ϕ_1 определяется как

$$\hat{\phi}_1 = \hat{\rho}_1.$$

Оценки для μ и σ_a^2 задаются следующим образом:

$$\hat{\mu} = \bar{Z}, \quad \hat{\sigma}_a^2 = \hat{\gamma}_0(1 - \hat{\phi}_1\hat{\rho}_1).$$

Теперь рассмотрим простейшую модель скользящего среднего MA(1).

$$\dot{Z}_t = a_t - \theta_1 a_{t-1}.$$

В этом случае оценкой μ также является выборочное среднее $\bar{Z}$. Для получения оценки параметра скользящего среднего θ_1 можно использовать выведенное нами ранее выражение для автокорреляции ρ_1:

$$\rho_1 = \frac{-\theta_1}{1 + \theta_1^2}.$$

Заменяя в данном квадратичном уравнении ρ_1 на $\hat{\rho}_1$, θ_1 на $\hat{\theta}_1$ и решая его, получаем оценку $\hat{\theta}_1$:

$$\hat{\theta}_1 = \frac{-1 \pm \sqrt{1 - 4\hat{\rho}_1^2}}{2\hat{\rho}_1}.$$

Для $\hat{\rho}_1 = \pm 0{,}5$ мы получаем единственное решение $\hat{\theta}_1 = \pm 1$, которое приводит к модели, не удовлетворяющей условию обратимости. Если $|\hat{\rho}_1| > 0{,}5$, то вещественного значения оценки $\hat{\theta}_1$ не существует.

При $|\hat{\rho}_1| < 0,5$ существует два различных вещественнозначных решения для $\hat{\theta}_1$ и из них выбирается одно – то, которое удовлетворяет условию обратимости. После получения $\hat{\theta}_1$ вычисляется оценка для σ_a:

$$\hat{\sigma}_a^2 = \frac{\hat{\gamma}_0}{1+\hat{\theta}_1^2}.$$

Приведенный пример модели MA(1) показывает, что оценки по методу моментов для модели скользящего среднего MA и смешанной модели ARMA сложны для получения. Вообще, и для моделей авторегрессии AR, и для моделей скользящего среднего MA, и для смешанных моделей ARMA оценки по методу моментов оказываются очень чувствительными к округлению ошибок. Они используются обычно как начальные оценки для последующего более эффективного оценивания.

Последнее замечание особенно относится к моделям MA и ARMA. Оценки, полученные по методу моментов, не рекомендуются в качестве результатов окончательного оценивания и не должны использоваться, если процесс близок к тому, чтобы быть нестационарным или необратимым.

8.2. МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

8.2.1. МЕТОД УСЛОВНОГО МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Для общей стационарной ARMA(p, q)-модели

$$\dot{Z}_t = \phi_1\dot{Z}_{t-1} + \ldots + \phi_p\dot{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}, \qquad (8.3)$$

где $\dot{Z}_t = Z_t - \mu$, $\{a_t\}$ – i.i.d. процесс белого шума, подчиняющийся нормальному распределению $N(0, \sigma_a^2)$. Совместная плотность распределения вектора ошибок $\mathbf{a} = (a_1, \ldots, a_n)^T$ задается следующей формулой:

$$P\ \mathbf{a} \mid \phi, \mu, \theta, \sigma_a^2 = 2\pi\sigma_a^2\ ^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma_a^2}\sum_{t=1}^{n} a_t^2\right].$$

Переписывая исходное уравнение модели следующим образом:

$$a_t = \theta_1 a_{t-1} + ... + \theta_q a_{t-q} + \dot{Z}_t - \phi_1 \dot{Z}_{t-1} - ... - \phi_p \dot{Z}_{t-p}, \qquad (8.4)$$

мы можем записать функцию правдоподобия от параметров $(\phi, \mu, \theta, \sigma_a^2)$.

Введем обозначения для вектора исходных данных $\mathbf{Z} = (Z_1, Z_2, ..., Z_n)^T$, а также для векторов, задающих начальные условия $\mathbf{Z}_* = (Z_{1-p}, ..., Z_{-1}, Z_0)^T$, $\mathbf{a}_* = (a_{1-q}, ..., a_{-1}, a_0)^T$. Логарифмическую функцию условного правдоподобия можем записать следующим образом:

$$\ln L_* \ \phi, \mu, \theta, \sigma_a^2 \ = -\frac{n}{2} \ln 2\pi\sigma_a^2 - \frac{S_*(\phi, \mu, \theta)}{2\sigma_a^2},$$

где $S_*(\phi, \mu, \theta) = \sum_{t=1}^{n} a_t^2(\phi, \mu, \theta \mid \mathbf{Z}_*, \mathbf{a}_*, \mathbf{Z})$ – функция условной суммы квадратов.

Величины $\hat{\phi}$, $\hat{\mu}$ и $\hat{\theta}$, максимизирующие $\ln L_* \ \phi, \mu, \theta, \sigma_a^2$, называются оценками условного максимального правдоподобия. Так как $\ln L_* \ \phi, \mu, \theta, \sigma_a^2$ зависит от исходной выборки $\mathbf{Z}$ только через зависимость $S_*(\phi, \mu, \theta)$ от $\mathbf{Z}$, то оценки $\hat{\phi}$, $\hat{\mu}$ и $\hat{\theta}$ являются также оценками условного метода наименьших квадратов и могут быть получены минимизацией условной функции суммы квадратов $S_*(\phi, \mu, \theta)$, которая, однако, не содержит оцениваемого параметра σ_a. Этот параметр должен в таком случае быть оценен отдельно.

8.2.2. СПОСОБЫ ОПРЕДЕЛЕНИЯ НАЧАЛЬНЫХ УСЛОВИЙ ДЛЯ МЕТОДА УСЛОВНОГО МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ МЕТОД ОБРАТНОГО ПРОГНОЗА

Для того чтобы начать выполнять вычисления условной суммы квадратов

$$S_*(\phi, \mu, \theta) = \sum_{t=1}^{n} a_t^2(\phi, \mu, \theta \mid \mathbf{Z}_*, \mathbf{a}_*, \mathbf{Z}),$$

необходимо определить векторы начальных условий $\mathbf{Z}_* = (Z_{1-p},..., Z_{-1}, Z_0)^T$ и $\mathbf{a}_* = (a_{1-q},..., a_{-1}, a_0)^T$. Здесь возникает проблема, так как, например, значения индексов компонент вектора $\mathbf{Z}_*$ неположительны, а у нас нет значений временного ряда $\{Z_t\}$ с такими индексами.

Существует несколько альтернативных способов определения векторов начальных условий $\mathbf{Z}_*$ и $\mathbf{a}_*$. Первый способ заключается в следующем. Основываясь на предположениях, что ряд $\{Z_t\}$ стационарный, а $\{a_t\}$ – i.i.d. процесс белого шума, подчиняющийся нормальному распределению $N(0,\sigma_a^2)$, мы можем заменить неизвестные Z_t выборочным средним $\bar{Z}$:

$$Z_{1-p} = Z_{2-p} = ... = Z_{-1} = Z_0 = \bar{Z},$$

а неизвестные a_t – их математическим ожиданием, равным 0:

$$a_{1-q} = a_{2-q} = ... = a_{-1} = a_0 = 0.$$

Вторым способом нахождения начальных значений $\mathbf{Z}_*$ и $\mathbf{a}_*$, применяемых в большинстве программных продуктов, является использование для этих целей части выборки. В этом случае для исходной модели ARMA(p, q) можно положить $a_p = a_{p-1} = ... = a_{p+1-q} = 0$ и вычислять a_t для $t \geq (p+1)$, используя исходное уравнение (8.4), в котором теперь не требуются значения Z_t с отрицательными индексами. В результате условная сумма квадратов имеет вид

$$S_*(\phi, \mu, \theta) = \sum_{t=p+1}^{n} a_t^2(\phi, \mu, \theta \mid \mathbf{Z}). \qquad (8.5)$$

После получения оценок параметров $\hat{\phi}$, $\hat{\mu}$ и $\hat{\theta}$ оценка $\hat{\sigma}_a^2$ дисперсии σ_a^2 вычисляется по формуле

$$\hat{\sigma}_a^2 = \frac{S_*(\hat{\phi}, \hat{\mu}, \hat{\theta})}{\text{d.f.}},$$

где d.f. – число степеней свободы, равное числу членов, используемых в сумме $S_*(\hat{\phi}, \hat{\mu}, \hat{\theta})$, минус число оцениваемых параметров. Если формула (8.5) используется для вычисления суммы квадратов $S_*(\hat{\phi}, \hat{\mu}, \hat{\theta})$, то

$$\text{d.f.} = (n - p) - (p + q + 1) = n - (2p + q + 1).$$

Для первого способа вычисления начальных значений

$$\text{d.f.} = n - (p + q + 1).$$

Третий, более точный, способ определения начальных значений $\mathbf{Z}_*$ и $\mathbf{a}_*$ предложен в книге [3], он называется *методом обратного прогноза*. Этот метод основывается на теории прогнозирования временных рядов, изложенной в следующих разделах. Основная идея метода заключается в том, что мы прогнозируем не будущие значения временного ряда, а прошлые значения рядов $\{Z_t\}$ и $\{a_t\}$, составляющих векторы $\mathbf{Z}_* = (Z_{1-p},..., Z_{-1}, Z_0)^T$ и $\mathbf{a}_* = (a_{1-q},..., a_{-1}, a_0)^T$ и необходимых для вычисления суммы квадратов и функции правдоподобия. Такое прогнозирование возможно, поскольку любая модель ARMA может быть записана как в прямой форме:

$$(1 - \phi_1 B - ... - \phi_p B^p)\dot{Z}_t = (1 - \theta_1 B - ... - \theta_q B^q)a_t,$$

где $B^j Z_t = Z_{t-j}$, так и в обратной:

$$(1 - \phi_1 F - ... - \phi_p F^p)\dot{Z}_t = (1 - \theta_1 F - ... - \theta_q F^q)e_t,$$

где $F^j Z_t = Z_{t+j}$ $(F = B^{-1})$.

В силу стационарности процесса эти два представления должны иметь в точности одинаковую структуру автоковариации, так как генерирующие автоковариационные функции автоковариации совпадают. Следовательно e_t – белый шум с нулевым математическим ожиданием и дисперсией σ_e^2. Поэтому таким же образом, как мы используем прямую форму для прогнозирования неизвестных будущих значений Z_{n+j} $(j > 0)$ на основе выборочных данных $(Z_1, Z_2, \ldots, Z_n)$, мы можем использовать и обратную форму для прогноза неизвестных про-

шлых значений Z_j и последующего вычисления a_j $(j \le 0)$ на основе данных $(Z_n, Z_{n-1}, \ldots, Z_1)$.

В результате получаем следующую безусловную логарифмическую функцию правдоподобия:

$$\ln L\ \phi, \mu, \theta, \sigma_a^2\ = -\frac{n}{2}\ln 2\pi\sigma_a^2 - \frac{S(\phi, \mu, \theta)}{2\sigma_a^2},$$

где $S(\phi, \mu, \theta)$ – функция безусловной суммы квадратов, задаваемая выражением

$$S(\phi, \mu, \theta) = \sum_{t=-\infty}^{n} E(a_t \mid \phi, \mu, \theta, \mathbf{Z})^{\ 2}, \tag{8.6}$$

где $E(a_t \mid \phi, \mu, \theta, \mathbf{Z})$ есть условное математическое ожидание a_t при заданных ϕ, $\mu, \theta, \mathbf{Z}$. Некоторые из слагаемых должны быть вычислены с использованием метода обратного прогноза, как подробно объясняется в [3].

Величины $\hat{\phi}, \hat{\mu}, \hat{\theta}$, максимизирующие $\ln L\ \phi, \mu, \theta, \sigma_a^2$, называются безусловными оценками максимального правдоподобия. Как и ранее, так как $\ln L\ \phi, \mu, \theta, \sigma_a^2$ зависит от выборочных данных только через $S(\phi, \mu, \theta)$, безусловные оценки максимального правдоподобия эквивалентны безусловным оценкам наименьших квадратов, получаемым в результате минимизации $S(\phi, \mu, \theta)$. На практике бесконечная сумма (8.6) аппроксимируется конечной суммой:

$$S(\phi, \mu, \theta) = \sum_{t=-M}^{n} E(a_t \mid \phi, \mu, \theta, \mathbf{Z})^{\ 2},$$

где M – достаточно большое целое – такое, что

$$\left|E(Z_t \left|\phi, \mu, \theta, \mathbf{Z}\right.) - E(Z_{t-1}\left|\phi, \mu, \theta, \mathbf{Z}\right.)\right| < \varepsilon \tag{8.7}$$

для заданного произвольно малого ε и $t \le -(M+1)$. Выполнение условия (8.7) предполагает, что $E(Z_t \left|\phi, \mu, \theta, \mathbf{Z}\right.) \simeq \mu$, следовательно, значение $E(a_t \left|\phi, \mu, \theta, \mathbf{Z}\right.)$ пренебрежимо мало для $t \le -(M+1)$.

После получения оценок параметров $\hat{\phi}$, $\hat{\mu}$ и $\hat{\theta}$ оценка $\hat{\sigma}_a^2$ дисперсии σ_a^2 может быть вычислена для метода безусловного правдоподобия по формуле

$$\hat{\sigma}_a^2 = \frac{S(\hat{\phi}, \hat{\mu}, \hat{\theta})}{n}.$$

Особо отметим эффективность использования метода обратных прогнозов при оценивании параметров сезонных моделей, для временных рядов, близких к нестационарным, и особенно для относительно коротких временных рядов.

8.2.3. ТОЧНЫЕ ФУНКЦИИ ПРАВДОПОДОБИЯ

Отметим, что рассмотренные ранее функции правдоподобия $\left(\text{условная } \ln L_* \ \phi, \mu, \theta, \sigma_a^2 \right.$ и безусловная $\left. \ln L \ \phi, \mu, \theta, \sigma_a^2 \right)$ являются аппроксимациями. Их использование задает приближенный метод максимального правдоподобия. Точный метод максимального правдоподобия предполагает использование не аппроксимаций, а точных функций. Точные функции правдоподобия $\ln L \ \dot{Z}_1, \dot{Z}_2, \ldots, \dot{Z}_n \middle| \phi, \mu, \theta, \sigma_a^2$ не могут быть определены так просто, как приближенные, поскольку $\dot{Z}_t$ в выборке $(\dot{Z}_1, \dot{Z}_2, \ldots, \dot{Z}_n)$ сильно коррелируют друг с другом в отличие от a_t, которые по предположению независимы. Поэтому вид функции правдоподобия $\ln L \ \dot{Z}_1, \dot{Z}_2, \ldots, \dot{Z}_n \middle| \phi, \mu, \theta, \sigma_a^2$ для общего случая модели ARMA(p, q) довольно сложен.

Рассмотрим для примера получение точной функции правдоподобия для наиболее простого случая процесса AR(1), который может быть представлен в виде

$$(1 - \phi B)\dot{Z}_t = a_t,$$

или, эквивалентно,

$$\dot{Z}_t = \phi \dot{Z}_{t-1} + a_t,$$

где $\dot{Z}_t = (Z_t - \mu)$, $|\phi| < 1$ и a_t являются i.i.d. $N \ 0, \sigma_a^2$.

Для дальнейших целей перепишем процесс AR(1) в представлении скользящего среднего MA бесконечного порядка:

$$\dot{Z}_t = \sum_{j=0}^{\infty} \phi^j a_{t-j} .$$

Очевидно, что $\dot{Z}_t$ распределен по нормальному закону как $N\ 0, \sigma_a^2 / (1-\phi^2)$. Однако, как указывалось ранее, значения процесса $\dot{Z}_t$ сильно коррелированны друг с другом, поэтому нельзя записать совместную функцию плотности распределения как произведение плотностей отдельных $\dot{Z}_t$.

Чтобы получить совместную функцию плотности распределения $P(\dot{Z}_1, \dot{Z}_2, ..., \dot{Z}_n)$ для выборки $(\dot{Z}_1, \dot{Z}_2, ..., \dot{Z}_n)$ и, следовательно, функцию правдоподобия для оценивания параметров, рассмотрим следующую последовательность:

$$e_1 = \sum_{j=0}^{\infty} \phi^j a_{1-j} = \dot{Z}_1,$$

$$a_2 = \dot{Z}_2 - \phi \dot{Z}_1,$$

$$a_3 = \dot{Z}_3 - \phi \dot{Z}_2,$$

$$\dots$$

$$a_n = \dot{Z}_n - \phi \dot{Z}_{n-1}.$$

Заметим, что e_1 подчиняется нормальному распределению $N\ 0, \sigma_a^2 / (1-\phi^2)$ и a_t, $2 \le t \le n$, также подчиняются нормальному распределению $N\ 0, \sigma_a^2$, причем все члены последовательности независимы друг от друга. Следовательно, совместная плотность распределения $(e_1, a_2, ..., a_n)$ определяется таким образом:

$$P(e_1, a_2, ..., a_n) = .$$

$$= \left[\frac{1-\phi^2}{2\pi\sigma_a^2} \right]^{1/2} \exp\left[\frac{-e_1^2 (1-\phi^2)}{2\sigma_a^2} \right] \left[\frac{1}{2\pi\sigma_a^2} \right]^{(n-1)/2} \exp\left[-\frac{1}{2\sigma_a^2} \sum_{t=2}^{n} a_t^2 \right].$$

Чтобы перейти к переменным $\dot{Z}_t$, рассмотрим следующее преобразование:

$$\dot{Z}_1 = e_1,$$

$$\dot{Z}_2 = \phi\dot{Z}_1 + a_2,$$

$$\dot{Z}_3 = \phi\dot{Z}_2 + a_3,$$

$$\ldots$$

$$\dot{Z}_n = \phi\dot{Z}_{n-1} + a_n.$$

Якобиан этого преобразования равен единице:

$$J = \frac{D(\dot{Z}_1, \dot{Z}_2, ..., \dot{Z}_n)}{D(e_1, a_2, ..., a_n)} = \frac{1}{\dfrac{D(e_1, a_2, ..., a_n)}{D(\dot{Z}_1, \dot{Z}_2, ..., \dot{Z}_n)}} =$$

$$= 1/\det\begin{bmatrix} 1 & 0 & 0 & 0 & . & . & 0 & 0 & 0 \\ -\phi & 1 & 0 & 0 & . & . & 0 & 0 & 0 \\ 0 & -\phi & 1 & 0 & . & . & 0 & 0 & 0 \\ \vdots & & \vdots & & . & . & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & . & . & 0 & -\phi & 1 \end{bmatrix} = 1.$$

Следовательно,

$$P(\dot{Z}_1, \dot{Z}_2, ..., \dot{Z}_n) = P(e_1, a_2, ..., a_n) =$$

$$= \left[\frac{1-\phi^2}{2\pi\sigma_a^2}\right]^{1/2} \exp\left[\frac{\dot{Z}_1^2(1-\phi^2)}{2\sigma_a^2}\right]\left[\frac{1}{2\pi\sigma_a^2}\right]^{(n-1)/2} \times$$

$$\times \exp\left[-\frac{1}{2\sigma_a^2}\sum_{t=2}^{n} \dot{Z}_t - \phi\dot{Z}_{t-1}\ ^2\right].$$

Отсюда для заданного ряда $(\dot{Z}_1, \dot{Z}_2, ..., \dot{Z}_n)$ мы получаем следующую точную логарифмическую функцию правдоподобия при оценивании параметров модели AR(1):

$$\ln \quad \dot{Z}_1, \dot{Z}_2, ..., \dot{Z}_n \mid \phi, \mu, \sigma_a^2 \quad =$$

$$= -\frac{n}{2}\ln 2\pi + \frac{1}{2}\ln(1-\phi^2) - \frac{n}{2}\ln\sigma_a^2 - \frac{S(\phi,\mu)}{2\sigma_a^2},$$

где $S(\phi,\mu) = (Z_1 - \mu)^2(1-\phi^2) + \sum_{t=2}^{n} (Z_t - \mu) - \phi(Z_{t-1} - \mu)^2$ – сумма квадратов, которая является функцией только параметров ϕ и μ.

8.2.4. ОПИСАНИЕ ПРОЦЕДУРЫ НЕЛИНЕЙНОГО ОЦЕНИВАНИЯ

Оценивание по методу максимального правдоподобия и методу наименьших квадратов включает в себя процедуру минимизации условной суммы квадратов $S_*(\phi,\mu,\theta)$ или безусловной суммы квадратов $S(\phi,\mu,\theta)$, представляющих собой суммы квадратов ошибок a_t. Для процесса авторегрессии AR(p) ошибки a_t линейны по параметрам:

$$a_t = \dot{Z}_t - \phi_1\dot{Z}_{t-1} - \phi_2\dot{Z}_{t-2} - ... - \phi_p\dot{Z}_{t-p}.$$

Однако для модели, содержащей фактор скользящего среднего, ошибки a_t нелинейны по параметрам. Это легко видно на простом примере модели ARMA(1,1):

$$\dot{Z}_t - \phi_1\dot{Z}_{t-1} = a_t - \theta_1 a_{t-1}.$$

Выразим a_t, затем подставим $a_{t-1} = \dot{Z}_{t-1} - \phi_1\dot{Z}_{t-2} + \theta_1 a_{t-2}$, в результате получим

$$a_t = \dot{Z}_t - \phi_1\dot{Z}_{t-1} + \theta_1 a_{t-1} =$$

$$= \dot{Z}_t - \phi_1\dot{Z}_{t-1} + \theta_1(\dot{Z}_{t-1} - \phi_1\dot{Z}_{t-2} + \theta_1 a_{t-2}) =$$

$$= \dot{Z}_t - (\phi_1 - \theta_1)Z_{t-1} - \phi\theta_1 Z_{t-2} + \theta_1^2 a_{t-2} = ...$$

Теперь очевидно, что в этом простом случае a_t нелинейно зависит от параметров модели. Следовательно, для оценивания параметров смешанной модели ARMA необходимо использовать нелинейную процедуру оценивания.

Нелинейная процедура оценивания наименьших квадратов (минимизируется условная или безусловная сумма квадратов ошибок) включает итерационный метод поиска решения, базирующийся на процедуре линейного оценивания. Рассмотрим для начала следующую линейную регрессионную модель:

$$Y_t = E(Y_t \mid X_{t1}, X_{t2}, ..., X_{tp}) + e_t =$$

$$= \alpha_1 X_{t1} + \alpha_2 X_{t2} + ... + \alpha_p X_{tp} + e_t, \quad t = 1, 2, ..., n, \tag{8.8}$$

где ошибки e_i являются i.i.d. $N\ 0, \sigma_a^2$ и независимы от всех X_{ti}. Пусть $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^T$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_p)^T$, а $\bar{\mathbf{X}}$ – соответствующая матрица для независимых переменных X_{ti}. Из теории линейного регрессионного анализа известно, что оценки наименьших квадратов определяются следующим образом:

$$\hat{\boldsymbol{\alpha}} = \ \bar{\mathbf{X}}^T \bar{\mathbf{X}}\ ^{-1} \bar{\mathbf{X}}^T \mathbf{Y} \tag{8.9}$$

и подчиняются многомерному нормальному распределению $MN(\alpha, V(\hat{\alpha}))$, где

$$V(\hat{\boldsymbol{\alpha}}) = \sigma_a^2\ \bar{\mathbf{X}}^T \bar{\mathbf{X}}\ ^{-1}.$$

При этом минимальная остаточная сумма квадратов определяется

$$S(\hat{\boldsymbol{\alpha}}) = \sum_{t=1}^{n} (Y_t - \hat{\alpha}_1 X_{t1} - \hat{\alpha}_2 X_{t2} - ... - \hat{\alpha}_p X_{tp})^2. \tag{8.10}$$

Оценки наименьших квадратов в (8.9) могут также быть получены с использованием следующей двухшаговой процедуры. Пусть $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \tilde{\alpha}_2, ..., \tilde{\alpha}_p)^T$ – начальные предполагаемые значения вектора

параметров $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_p)^T$. Перепишем сумму квадратов $S(\hat{\boldsymbol{\alpha}})$ в (8.10) следующим образом:

$$S(\hat{\boldsymbol{\alpha}}) = \sum_{t=1}^{n} \Big[Y_t - \tilde{\alpha}_1 X_{t1} - ... - \tilde{\alpha}_p X_{tp} -$$

$$-(\hat{\alpha}_1 - \tilde{\alpha}_1) X_{t1} - \ldots - (\hat{\alpha}_p - \tilde{\alpha}_p) X_{tp} \Big]^2$$

или

$$S(\boldsymbol{\delta}) = S(\hat{\boldsymbol{\alpha}}) = \sum_{t=1}^{n} (\tilde{e}_t - \delta_1 X_{t1} - ... - \delta_p X_{tp})^2,$$

где $\boldsymbol{\delta} = (\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}})$, а $\tilde{e}_t$ – оцененные остатки для начальных значений параметров $\tilde{\boldsymbol{\alpha}}$. Так как $S(\boldsymbol{\alpha})$ и $S(\boldsymbol{\delta})$ записаны в одинаковой форме, можно определить оценку наименьших квадратов для вектора приращения $\boldsymbol{\delta}$:

$$\boldsymbol{\delta} = \left(\bar{\mathbf{X}}^T \bar{\mathbf{X}} \right)^{-1} \bar{\mathbf{X}}^T \tilde{\mathbf{e}}, \tag{8.11}$$

где $\tilde{\mathbf{e}} = (\tilde{e}_1, ... \tilde{e}_n)^T$, $\tilde{e}_t = Y_t - \tilde{Y}_t$, $\tilde{Y}_t = \tilde{\alpha}_1 X_{t1} + ... + \tilde{\alpha}_p X_{tp}$.

Таким образом, после задания начальных значений $\tilde{\boldsymbol{\alpha}}$ можно вычислить вектор остатков $\tilde{\mathbf{e}}$, затем вектор $\boldsymbol{\delta} = (\delta_1, \delta_2, ..., \delta_p)^T$ по формуле (8.11). После этого вектор оценок наименьших квадратов $\hat{\boldsymbol{\alpha}}$ модели (8.8) может быть определен таким образом:

$$\hat{\alpha} = \tilde{\alpha} + \delta,$$

после чего процедура линейного оценивания завершается.

Теперь рассмотрим аналогичную процедуру оценивания для нелинейной, в общем случае, модели:

$$Y_t = f(\mathbf{X}_t, \boldsymbol{\alpha}) + e_t, \; t = 1, 2, ..., n, \tag{8.12}$$

где $\mathbf{X}_t = (X_{t1}, X_{t2}, ..., X_{tp})$ – множество независимых переменных, соответствующих наблюдениям; $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_p)^T$ – вектор парамет-

ров, а e_t – процесс белого шума с нулевым математическим ожиданием и постоянной дисперсией σ_e^2, не зависимый от $\mathbf{X}_t$. Пусть $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^T$, $\mathbf{f}(\boldsymbol{\alpha}) = f(\mathbf{X}_1, \boldsymbol{\alpha}), f(\mathbf{X}_2, \boldsymbol{\alpha}), \ldots, f(\mathbf{X}_n, \boldsymbol{\alpha})^T$.

Изложенная выше техника линейного оценивания лежит в основе следующей итерационной процедуры получения оценок наименьших квадратов нелинейной в общем случае регрессии. Процедура состоит из повторяющихся двух шагов.

Шаг 1. Задав вектор начальных оценок параметров $\tilde{\alpha}$, вычисляем остатки $\tilde{\mathbf{e}} = (\mathbf{Y} - \tilde{\mathbf{Y}})$ и остаточную сумму квадратов

$$S(\tilde{\boldsymbol{\alpha}}) = \tilde{\mathbf{e}}^T \tilde{\mathbf{e}} = (\mathbf{Y} - \tilde{\mathbf{Y}})^T (\mathbf{Y} - \tilde{\mathbf{Y}}),$$

где $\tilde{\mathbf{Y}} = \mathbf{f}(\tilde{\boldsymbol{\alpha}})$ – вектор, полученный из $\mathbf{f}(\boldsymbol{\alpha})$ с помощью замены неизвестных параметров начальным вектором $\tilde{\boldsymbol{\alpha}}$. Далее аппроксимируем модель $f(\mathbf{X}_t, \tilde{\boldsymbol{\alpha}})$ разложением в ряд Тейлора первого порядка в окрестности начальной точки $\tilde{\boldsymbol{\alpha}}$:

$$\mathbf{f}(\boldsymbol{\alpha}) = \mathbf{f}(\tilde{\boldsymbol{\alpha}}) + \bar{\mathbf{X}}_{\tilde{\boldsymbol{\alpha}}} \boldsymbol{\delta},$$

где $\boldsymbol{\delta} = (\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}})$, а $\bar{\mathbf{X}}_{\tilde{\boldsymbol{\alpha}}} = \{X_{tj}\}$ – $n \times p$ – матрицы частных производных, вычисленных в точке $\tilde{\boldsymbol{\alpha}}$, т.е.

$$X_{tj} = \left. \frac{\partial f(\mathbf{X}_t, \boldsymbol{\alpha})}{\partial \alpha_j} \right|_{\boldsymbol{\alpha} = \tilde{\boldsymbol{\alpha}}}, \; t = 1, 2, \ldots, n, \; j = 1, 2, \ldots, p.$$

Оценки $\boldsymbol{\delta}$ можно вычислить по формуле

$$\boldsymbol{\delta} = \left(\bar{\mathbf{X}}_{\tilde{\boldsymbol{\alpha}}}^T \bar{\mathbf{X}}_{\tilde{\boldsymbol{\alpha}}} \right)^{-1} \bar{\mathbf{X}}_{\tilde{\boldsymbol{\alpha}}}^T \tilde{\mathbf{e}} = (\delta_1, \delta_2, \ldots, \delta_p)^T.$$

Заметим, что для линейной модели матрица производных $\bar{\mathbf{X}}_{\tilde{\boldsymbol{\alpha}}}$ фиксирована и равна $\bar{\mathbf{X}}$. Для нелинейной модели матрица $\bar{\mathbf{X}}_{\tilde{\boldsymbol{\alpha}}}$ меняется от

итерации к итерации в зависимости от новых значений вектора параметров.

Шаг 2. Получаем обновленные оценки наименьших квадратов

$$\hat{\boldsymbol{\alpha}} = \tilde{\boldsymbol{\alpha}} + \boldsymbol{\delta},$$

и соответствующую остаточную сумму квадратов $S(\hat{\boldsymbol{\alpha}})$. Для линейной модели шаг 2 дает окончательные оценки наименьших квадратов. Для нелинейной модели шаг 2 лишь ведет к новым начальным значениям для дальнейших итераций.

Итак, для заданной общей модели ARMA(p,q) мы можем использовать итерационную процедуру нелинейного оценивания для получения результатов, минимизирующих ошибочную сумму квадратов $S_*(\boldsymbol{\varphi},\mu,\boldsymbol{\theta})$ или $S(\boldsymbol{\varphi},\mu,\boldsymbol{\theta})$. Процедура нелинейного оценивания начинается с начальных значений параметров, полученных, например, с использованием метода моментов. На шаге 2 эти значения улучшаются в соответствии с минимизацией суммы квадратов. Итерации продолжаются до тех пор, пока не будет достигнут некоторый критерий сходимости. В качестве такого критерия может быть использовано, например, относительное уменьшение суммы квадратов, или максимальное изменение значений параметров, менее заданного уровня, или число итераций, больше некоторого числа.

Чтобы достигнуть хорошей и быстрой сходимости, разработано множество алгоритмов поиска. Один из часто используемых алгоритмов – алгоритм Марквардта. Он является компромиссом между методом Ньютона–Гаусса и методом наискорейшего спуска.

8.2.5. СВОЙСТВА ОЦЕНОК ПАРАМЕТРОВ

Пусть $\boldsymbol{\alpha} = (\boldsymbol{\varphi},\mu,\boldsymbol{\theta})$, а $\hat{\boldsymbol{\alpha}}$ – оценка $\boldsymbol{\alpha}$, и $\bar{\mathbf{X}}_{\tilde{\boldsymbol{\alpha}}}$ – матрица частных производных на последней итерации нелинейной процедуры оценивания наименьших квадратов. Известно, что $\hat{\boldsymbol{\alpha}}$ имеет многомерное нормальное распределение $MN(\boldsymbol{\alpha},V(\hat{\boldsymbol{\alpha}}))$. Оценка ковариационной матрицы $\hat{\mathrm{V}}(\hat{\boldsymbol{\alpha}})$ оценки $\hat{\boldsymbol{\alpha}}$ определяется по формуле

$$\hat{V}(\hat{\boldsymbol{\alpha}}) = \hat{\sigma}_a^2 \left(\bar{X}_{\hat{\boldsymbol{\alpha}}}^T \bar{X}_{\hat{\boldsymbol{\alpha}}} \right)^{-1} = (\hat{\sigma}_{\hat{\alpha}_i \hat{\alpha}_j}),$$

где $\hat{\sigma}_a^2 = \dfrac{S_*(\hat{\boldsymbol{\phi}}, \hat{\mu}, \hat{\boldsymbol{\theta}})}{\text{d.f.}}$ или $\hat{\sigma}_a^2 = \dfrac{S(\hat{\boldsymbol{\phi}}, \hat{\mu}, \hat{\boldsymbol{\theta}})}{n}$, а $\hat{\sigma}_{\hat{\alpha}_i \hat{\alpha}_j}$ – выборочная ковариация между $\hat{\alpha}_i$ и $\hat{\alpha}_j$. Соответственно, оценка корреляционной матрицы полученных оценок определяется следующим образом:

$$\hat{R}(\alpha) = (\rho_{\hat{\alpha}_i \hat{\alpha}_j}),$$

где $\rho_{\hat{\alpha}_i \hat{\alpha}_j} = \dfrac{\hat{\sigma}_{\hat{\alpha}_i \hat{\alpha}_j}}{\sqrt{\hat{\sigma}_{\hat{\alpha}_i \hat{\alpha}_i} \hat{\sigma}_{\hat{\alpha}_j \hat{\alpha}_j}}}$.

Высокая корреляция между оценками свидетельствует о перепараметризации (излишне большом числе параметров) модели. В этом случае рекомендуется попробовать уменьшить число параметров, так как перепараметризация часто вызывает трудности в сходимости процедуры нелинейного оценивания наименьших квадратов. Число параметров можно уменьшить, проверяя гипотезу о незначимости параметров.

Гипотезу о равенстве значений параметров некоторым значениям

$$H_0 : \alpha_i = \alpha_{i0}$$

можно проверить, используя следующую *t*-статистику:

$$t = \frac{\hat{\alpha}_i - \alpha_{i0}}{\sqrt{\hat{\sigma}_{\hat{\alpha}_i \hat{\alpha}_i}}}$$

с числом степеней свободы $n - (p + q + 1)$ для модели ARMA (8.3) (в общем случае число степеней свободы равно размеру выборки, используемой при оценивании, минус число оцениваемых параметров). Если положить $\alpha_{i0} = 0$, то получаем гипотезу о незначимости оцениваемых параметров:

$$H_0 : \alpha_i = 0,$$

которая проверяется с использованием *t*-статистики:

$$t = \frac{\hat{\alpha}_i}{\sqrt{\hat{\sigma}_{\hat{\alpha}_i \hat{\alpha}_i}}},$$

вычисляемой для каждого из оцениваемых параметров. Если получаемое для данной статистики *p*-значение меньше заданного уровня значимости (например, 0,05), то гипотеза о незначимости параметра α_i отвергается. В противном случае, если *p*-значение больше заданного уровня значимости, надо рассмотреть возможность исключения рассматриваемого параметра из модели.

8.3. ОЦЕНИВАНИЕ ОБЫКНОВЕННЫМ МНК В АНАЛИЗЕ ВРЕМЕННЫХ РЯДОВ

Регрессионный анализ – наиболее часто используемый статистический метод в анализе данных. Для стандартных регрессионных моделей разработан обыкновенный метод наименьших квадратов (ОМНК), являющийся самой используемой процедурой вообще в статистике. В данном подразделе покажем, что обыкновенный МНК – не самый лучший метод оценивания в анализе временных рядов.

Рассмотрим простую линейную парную регрессионную модель:

$$Z_t = \phi X_t + e_t\,,\ t = 1, 2, \ldots, n.$$

Сделаем следующие основные стандартные предположения об ошибочном члене e_t:

1) нулевое математическое ожидание: $E(e_t) = 0$;

2) постоянная дисперсия: $E(e_t^2) = \sigma_e^2$;

3) некоррелированность: $E(e_t\, e_k) = 0$ для $t \neq k$;

4) некоррелированность с объясняющими переменными X_t: $E(X_t\, e_k) = 0$.

Хорошо известно, что при этих предположениях ОМНК-оценка

$$\hat{\phi} = \frac{\sum_{t=1}^{n} X_t Z_t}{\sum_{t=1}^{n} X_t^2}$$

состоятельна и является наилучшей линейной оценкой параметра ϕ. Отметим, однако, особую важность условия 4. Это условие выполняет-

ся автоматически, если объясняющие переменные X_t являются фиксированными (не стохастическими).

Однако в анализе временных рядов объясняющие переменные – случайны. Рассмотрим следующую простейшую модель временного ряда:

$$Z_t = \phi Z_{t-1} + e_t, \quad t = 1, 2, ..., n.$$

ОМНК-оценка параметра ϕ этой модели определяется по формуле

$$\hat{\phi} = \frac{\sum_{t=2}^{n} Z_{t-1} Z_t}{\sum_{t=2}^{n} Z_{t-1}^2}. \tag{8.13}$$

Чтобы увидеть, являются ли оценки в этом случае состоятельными и несмещенными, перепишем $\hat{\phi}$ в виде

$$\hat{\phi} = \frac{\sum_{t=2}^{n} Z_{t-1} Z_t}{\sum_{t=2}^{n} Z_{t-1}^2} = \frac{\sum_{t=2}^{n} Z_{t-1}(\phi Z_{t-1} + e_t)}{\sum_{t=2}^{n} Z_{t-1}^2} = \phi + \frac{\sum_{t=2}^{n} Z_{t-1} e_t}{\sum_{t=2}^{n} Z_{t-1}^2}.$$

Отсюда видно, что свойства полученной оценки зависят от стохастических свойств ошибочного члена e_t. В зависимости от этих свойств ошибочного члена можно выделить два важных (с точки зрения свойств оценок) случая.

Случай 1: $e_t = a_t$, т.е. e_t – белый шум с нулевым математическим ожиданием и дисперсией σ_e^2. В этом случае значение $E(\hat{\phi}) = \phi$, т.е. $\hat{\phi}$ – несмещенная оценка параметра ϕ. Кроме того, легко показать, что $\hat{\phi}$ – состоятельная оценка ϕ. Действительно, очевидно, что оценка $\hat{\phi}$ в (8.13) эквивалентна первой выборочной автокорреляции $\hat{\rho}_1$ для ряда Z_t. При $|\phi| < 1$ мы получаем AR(1) – процесс с абсолютно суммируемой автокорреляционной функцией, для которого $\hat{\rho}_1$ – состоятельная оценка ρ_1. Таким образом, $\hat{\phi}$ – состоятельная оценка ϕ.

Случай 2: $e_t = (1-\theta B)a_t$, где a_t – белый шум с нулевым математическим ожиданием и дисперсией σ_a^2. Таким образом, получается, что e_t – процесс скользящего среднего первого порядка MA(1), а Z_t – смешанный процесс ARMA(1,1):

$$Z_t = \phi Z_{t-1} + a_t - \theta a_{t-1}.$$

Проверим выполнение свойств 3 и 4 о некоррелированности ошибок друг с другом и объясняющими переменными для модели регрессионного анализа:

$$E(e_{t-1}e_t) = E\ (a_{t-1} - \theta a_{t-2})(a_t - \theta a_{t-1}\ = -\theta\sigma_a^2,$$

$$E(Z_{t-1}e_t) = E\ Z_{t-1}(a_t - \theta a_{t-1}\ = 0 - \theta\sigma_a^2 = -\theta\sigma_a^2.$$

Отсюда видно, что автокорреляция в ошибочном члене не только нарушает предположение 3 о некоррелированности ошибок друг с другом, но и вызывает нарушение условия 4 о некоррелированности ошибок с объясняющими переменными, когда объясняющие переменные содержат лаговую зависимую переменную. В этом случае $\hat{\phi}$ больше не является несмещенной оценкой параметра ϕ. Хуже того, $\hat{\phi}$ не является даже состоятельной оценкой ϕ, так как $\hat{\phi} \simeq \hat{\rho}_1$ – состоятельная оценка для автокорреляции ρ_1, а для процесса ARMA(1,1)

$$\rho_1 = \frac{(\phi-\theta)(1-\phi\theta)}{1+\theta^2-2\phi\theta} \neq \phi.$$

Итак, получается, что ОМНК-оценки параметров при объясняющих переменных в регрессионной модели будут несостоятельными, за исключением лишь того случая, когда член ошибки некоррелирован с объясняющими переменными. Для процесса ARMA(p,q) это условие не выполняется, кроме случая $q = 0$. Поэтому в этом случае обосновано использование условного или безусловного метода максимального правдоподобия, описанного ранее.

8.4. ДИАГНОСТИКА МОДЕЛИ

Построение модели временного ряда представляет собой итерационную процедуру, которая начинается с этапа идентификации модели, в результате чего мы получаем пробную модель. Далее следует этап

оценивания параметров, в результате которого параметры пробной модели получают конкретные значения и мы имеем оцененную модель. После обсуждения методов оценивания параметров модели временного ряда мы выбрали метод оценивания – это метод условного или безусловного максимального правдоподобия. Следующий этап – диагностика оцененной модели, в результате которой мы должны определить адекватность модели, т.е. выяснить, насколько адекватно построенная модель соответствует имеющимся данным.

Первое, что надо сделать после получения оценок параметров – проверить, значимы ли полученные оценки. Если оценки незначимы, это может свидетельствовать о том, что пробная модель не соответствует описываемым данным и надо предложить альтернативную модель. Если незначимы только оценки параметров при высших степенях полиномов AR и MA, это может свидетельствовать о том, что надо сократить порядки указанных полиномов, придя к модели с меньшим количеством параметров. Однако перед тем, как менять пробную модель, рекомендуется довести ее диагностику до конца, – дальнейший анализ может подсказать направление, в котором надо осуществлять модификацию.

Дальнейшая диагностика модели сводится к проверке того, выполняются ли исходные предположения модели. Основное предположение ARMA-модели: процесс $\{a_t\}$ является белым шумом, т.е. a_t – некоррелированные случайные возмущения с нулевым математическим ожиданием и постоянной дисперсией. Для любой оцененной модели остатки $\hat{a}_t$ являются оценками этих ненаблюдаемых случайных возмущений a_t. Следовательно, диагностика модели выполняется посредством тщательного анализа ряда остатков $\{\hat{a}_t\}$. Вычислить ряд остатков оцененной модели можно, подставляя, например, в формулу (8.4) полученные оценки параметров:

$$\hat{a}_t = \hat{\theta}_1\hat{a}_{t-1} + \ldots + \hat{\theta}_q\hat{a}_{t-q} + \dot{Z}_t - \hat{\phi}_1\dot{Z}_{t-1} - \ldots - \hat{\phi}_p\dot{Z}_{t-p},$$

и используя в качестве начальных условий $\mathbf{Z}_* = (Z_{1-p},\ldots,Z_{-1},Z_0)^T$ и $\hat{\mathbf{a}}_* = (\hat{a}_{1-q},\ldots,\hat{a}_{-1},\hat{a}_0)^T$, которые использовались на этапе оценивания параметров.

Обычно используют два критерия для проверки остатков. Первый критерий – соответствие ряда остатков нормальному распределению.

Второй критерий – проверка некоррелированности ряда остатков, постоянства дисперсии и математического ожидания. Чтобы проверить, распределены ли ошибки по нормальному закону, используется гистограмма стандартизированных остатков $\hat{a}_t / \hat{\sigma}_a$ в явном виде или график остатков на нормальной вероятностной бумаге, в результате чего стандартизированные остатки сравниваются со стандартным нормальным распределением.

Чтобы проверить, постоянна ли дисперсия остатков, проще всего исследовать график остатков, на котором обычно этот дефект хорошо заметен. Более тонкая процедура – эффект различных значений λ методом Бокса–Кокса (см. преобразования, стабилизирующие дисперсию).

Стандартная процедура проверки некоррелированности остатков – вычисление для ряда остатков выборочных АКФ и ЧАКФ. Для ряда, представляющего собой белый шум, как было показано ранее, обе эти характеристики должны быть статистически незначимыми (должны лежать внутри коридора полутора-двух стандартных отклонений, при уровне значимости $\alpha = 0.05$).

На основе результатов анализа остатков, если пробная модель неадекватна, можно легко получить новую модель. Например, пусть применяемая модель AR(1)

$$(1 - \phi_1 B)(Z_t - \mu) = b_t$$

дает в качестве ряда остатков не белый шум, а процесс, который идентифицируется как процесс скользящего среднего MA(1):

$$b_t = (1 - \theta_1 B) a_t .$$

В этом случае необходимо вернуться к этапу идентификации и идентифицировать исходный ряд моделью ARMA(1,1):

$$(1 - \phi_1 B)(Z_t - \mu) = (1 - \theta_1 B) a_t ,$$

после чего повторить последующие этапы построения модели. И так до тех пор, пока не будет получена удовлетворительная модель. Описанная процедура использования остатков для модификации моделей обычно работает очень хорошо.

8.5. КРИТЕРИИ ВЫБОРА АЛЬТЕРНАТИВНЫХ МОДЕЛЕЙ

В анализе временных рядов и вообще в анализе данных может существовать несколько адекватных альтернативных моделей, описывающих исходные данные. Альтернативные модели порождают проблему выбора. Иногда сделать наилучший выбор легко, в других случаях выбор оказывается очень сложным. Поэтому в литературе по анализу данных приводится множество критериев сравнения моделей. Эти критерии отличаются от методов идентификации моделей, таких как АКФ, ЧАКФ и др. Эти средства используются только для идентификации адекватных моделей, но не применимы в качестве критериев выбора среди адекватных моделей. Остатки всех адекватных моделей являются белым шумом и, в общем, не различимы по этим функциям. Поэтому стандартный анализ остатков также не может быть критерием выбора среди адекватных моделей.

Если имеется несколько адекватных моделей для конкретного набора данных, то выбор модели обычно основывается на суммарных статистиках остатков, вычисленных по оцененной модели, или на ошибках прогноза, вычисленных по вневыборочным прогнозам. Последний метод часто выполняется путем использования части выборки для конструирования модели, а оставшейся части – для оценки качества прогноза.

Далее приведем некоторые наиболее широко известные критерии выбора моделей, основанные на суммарных статистиках остатков. Все они строятся на принципе «экономии мышления», суть которого сводится к выбору среди нескольких адекватных моделей той, которая содержит меньшее количество параметров.

Информационный критерий Акаике

Общий вид критерия Акаике (AIC – Akaike Information Criterion) для модели ARMA:

$$\mathrm{AIC}(M) = n\ln\hat{\sigma}_a^2 + 2M\ ,$$

где M – число оцениваемых параметров модели, n – число наблюдений в выборке, $\hat{\sigma}_a^2$ – оценка максимального правдоподобия для σ_a^2.

Критерий Акаике – это эвристическая попытка свести в один показатель два требования: уменьшение числа параметров модели и качество подгонки модели. Согласно этому критерию из двух моделей следует выбрать модель с меньшим значением AIC.

AIC-критерий имеет тенденцию переоценивать порядок авторегрессии. Чтобы устранить этот недостаток, введено байесовское расширение этого критерия – критерий BIC (Bayesian Information Criterion):

$$BIC(M) = n\ln\hat{\sigma}_a^2 - (n-M)\ln\left(1-\frac{M}{n}\right) +$$

$$+M\ln n + M_z \ln\left[\left(\frac{\hat{\sigma}_Z^2}{\hat{\sigma}_a^2} - 1\right)\Big/ M\right],$$

где σ_Z^2 – выборочная дисперсия исходного ряда, остальные величины те же, что и в критерии AIC.

Информационный критерий Шварца

На основе байесовского критерия Дж. Шварц предложил свой критерий (SBC – Schwartz's Bayesian Criterion):

$$\mathrm{SBC}(M) = n\ln\hat{\sigma}_a^2 + M\ln n.$$

Отличие критерия Шварца от критерия AIC заключается в большем штрафе за количество параметров.

Информационный критерий Парзена

Е. Парзен предложил свой критерий (Parzen's CAT criterion) для проверки модели CAT (Criterion for Autoregressive Transfer functions):

$$CAT(p) = \begin{cases} -\left(1+\dfrac{1}{n}\right), & p = 0, \\ \dfrac{1}{n}\displaystyle\sum_{j=1}^{p}\frac{1}{\hat{\sigma}_j^2} - \frac{1}{\hat{\sigma}_p^2}, & p = 1,2,3,\ldots, \end{cases}$$

где $\hat{\sigma}_j^2$ – несмещенная оценка σ_a^2, если авторегрессионная модель $\mathrm{AR}(j)$ подгоняется под данные; n – число наблюдений. Оптимальный порядок p авторегрессионной модели $\mathrm{AR}(p)$ выбирается так, чтобы $\mathrm{CAT}(p)$ был минимальным.

9. АНАЛИЗ ВРЕМЕННЫХ РЯДОВ С ИСПОЛЬЗОВАНИЕМ МЕТОДОЛОГИИ ARIMA

В данном разделе проведем анализ нескольких временных рядов и построим для них модели класса ARIMA. При построении моделей последовательно осуществляются этапы идентификации, оценивания параметров, диагностической проверки. Если проверка показывает неадекватность модели, указанные этапы повторяются до тех пор, пока не получим приемлемую модель.

9.1. АНАЛИЗ ВРЕМЕННОГО РЯДА ВУЛЬФА ЕЖЕГОДНОГО ЧИСЛА СОЛНЕЧНЫХ ПЯТЕН, СНЯТЫЙ с 1700 по 1983 год

Рассмотрим временной ряд, содержащий ежегодные данные наблюдений о числе солнечных пятен с 1700 по 1983 год. Для работы с временным рядом сначала необходимо либо создать таблицу с указанными данными самостоятельно, либо открыть или импортировать соответствующий файл. Далее на стартовой панели модуля ***Time Series Analysis*** необходимо выбрать соответствующую переменную – столбец, содержащий необходимые данные.

Первый этап идентификации данного временного ряда – построение и исследование графика (рис. 9.1). Из полученного графика на рис. 9.1 видно, что имеет место некоторая нестационарность дисперсии. Использование логарифмического преобразования, график которого представлен на рис. 9.2, не дает желаемого эффекта, так как это преобразование является слишком сильным. Из рисунка данный факт не столь очевиден. Однако вспомним, что преобразование Бокса–Кокса помимо стабилизации дисперсии еще и приближает исходное распределение к нормальному.

Анализ гистограмм исходного и логарифмированного рядов, а также графиков на нормальной вероятностной бумаге, изображенных на рис. 9.3–9.6, свидетельствует о несимметричном характере распределения.

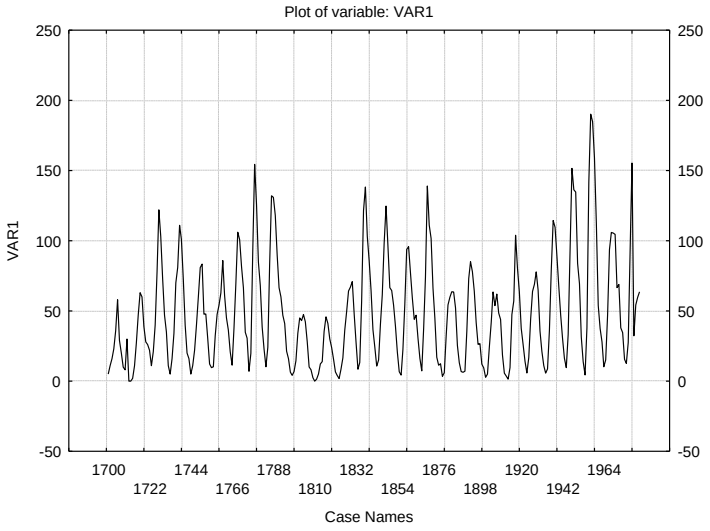


Рис. 9.1. График временного ряда ежегодного числа солнечных пятен

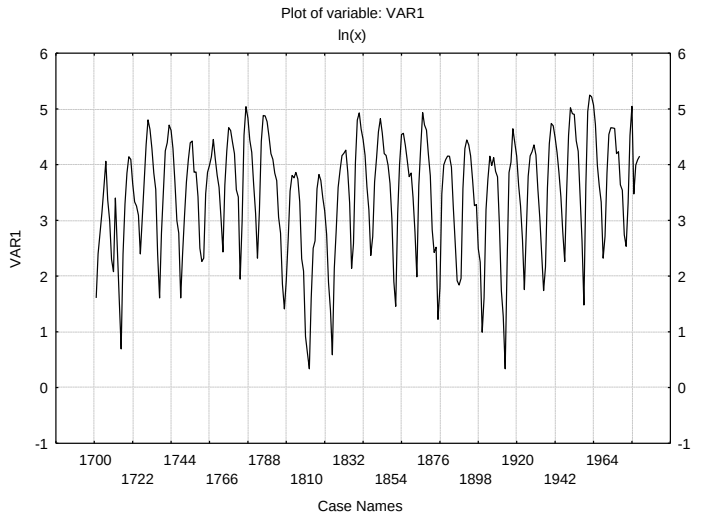


Рис. 9.2. Использование логарифмического преобразования

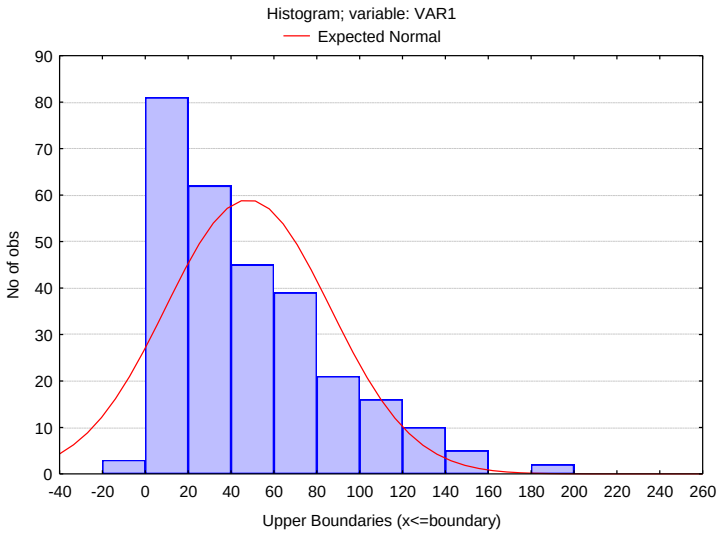


Рис. 9.3. Гистограмма исходного распределения

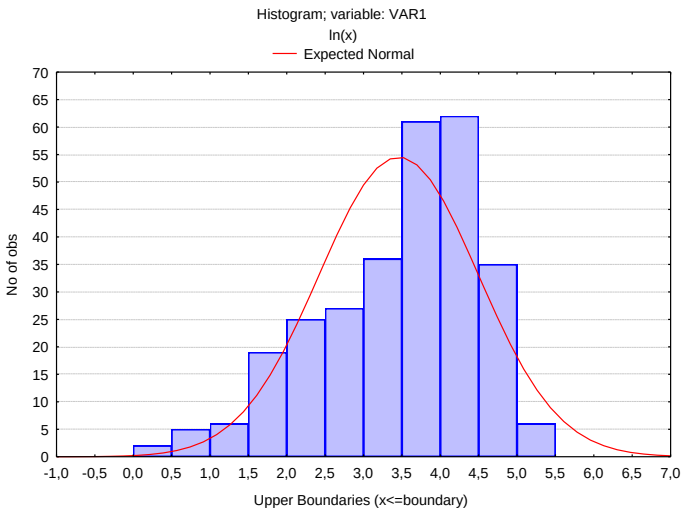


Рис. 9.4. Гистограмма распределения, полученного после логарифмирования

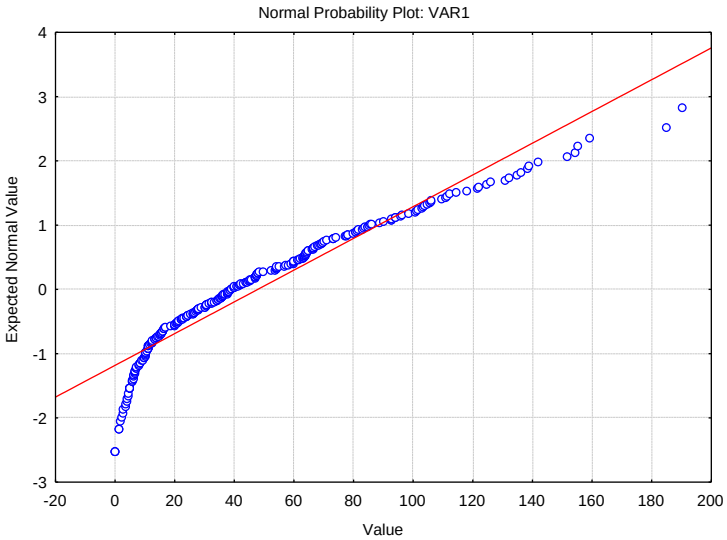


Рис. 9.5. Нормальный вероятностный график исходного распределения

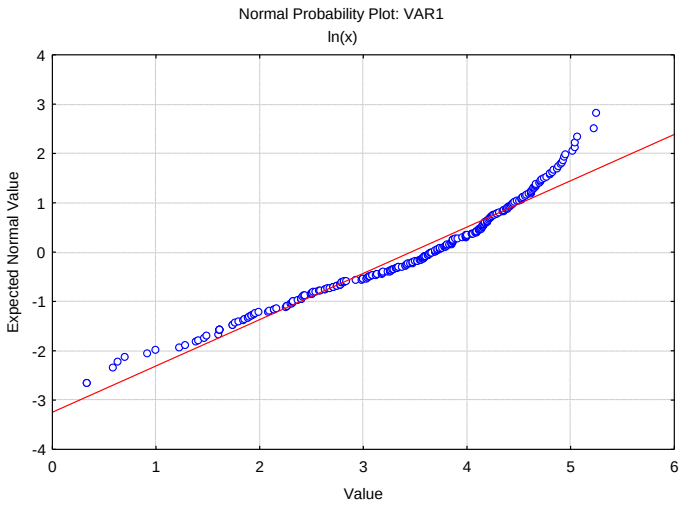


Рис. 9.6. Нормальный вероятностный график распределения после логарифмирования

Попробуем применить более слабое преобразование – квадратный корень. График, гистограмма и нормальный график преобразованного ряда, изображенные на рис 9.7–9.9, подтверждают, что это преобразование хорошо стабилизирует дисперсию, и его распределение является более симметричным и близким к нормальному, чем исходное.

Теперь, после стабилизации дисперсии с использованием преобразования,

$$T(Z_t) = \sqrt{Z_t}\ ,$$

перейдем *ко второму этапу* идентификации. Суть этого этапа – переход от нестационарного ряда к стационарному с помощью операции взятия разности.

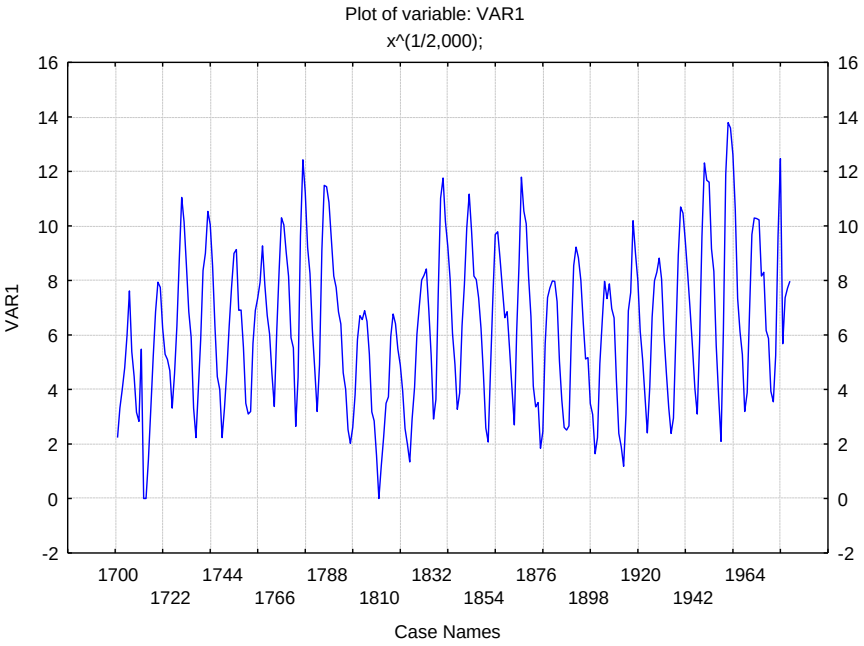


Рис. 9.7. Использование преобразования «корень квадратный»

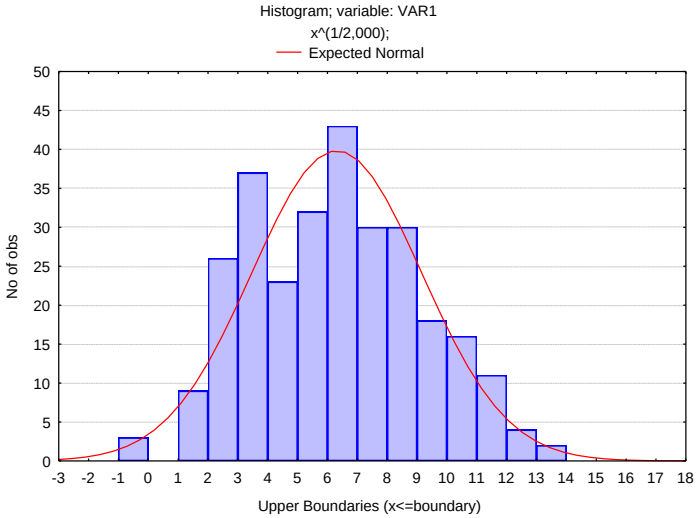


Рис. 9.8. Гистограмма после преобразования «корень квадратный»

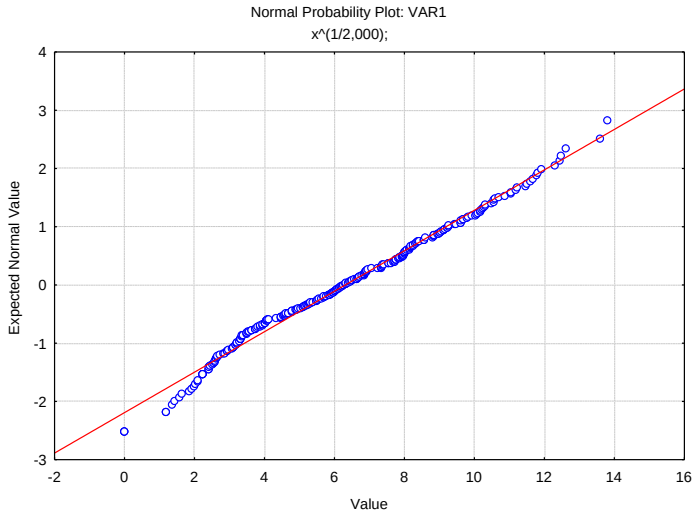


Рис. 9.9. Нормальный вероятностный график после преобразования «корень квадратный»

Для начала заметим, что из графика на рис. 9.7 можно сделать вывод о том, что ряд стационарен по среднему. Для того чтобы подтвердить этот вывод, построим графики АКФ и ЧАКФ. Результаты, приведенные на рис. 9.10 и 9.11, позволяют предположить, что ряд стационарен, а предположительная модель – модель авторегрессии AR порядка более двух. Основанием для такого предположения служит то, что АКФ представляет собой затухающую синусоиду, а ЧАКФ выявляет значимые автокорреляции до $k = 9$. Нестационарный же ряд выявляет слабо убывающую АКФ и одну значимую ЧАКФ при $k = 1$. Порядок авторегрессии больше двух, так как для AR(1) АКФ экспоненциальна и не проявляет синусоидальный характер.

Так как ряд $\sqrt{Z_t}$ стационарен, то перед подгонкой модели ARMA необходимо вычесть среднее $\dfrac{\sum_{t=1}^{n}\sqrt{Z_t}}{n} = 6{,}311$. В результате получаем ряд $\sqrt{Z_t} - 6{,}311$.

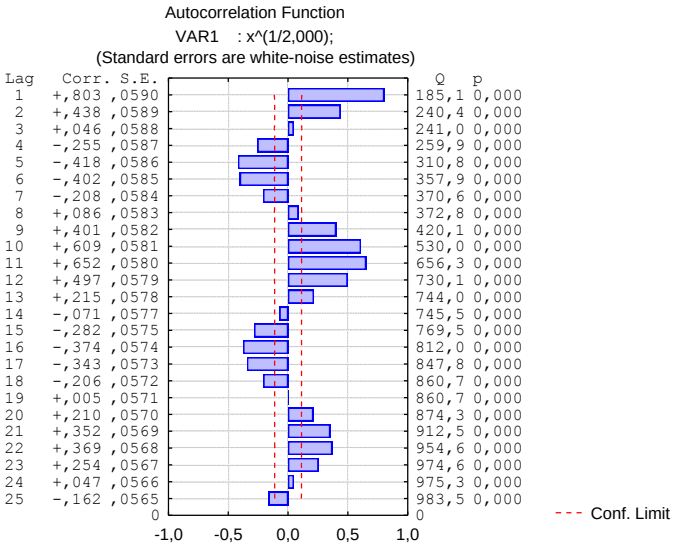


Рис. 9.10. АКФ ряда $T(Z_t) = \sqrt{Z_t}$

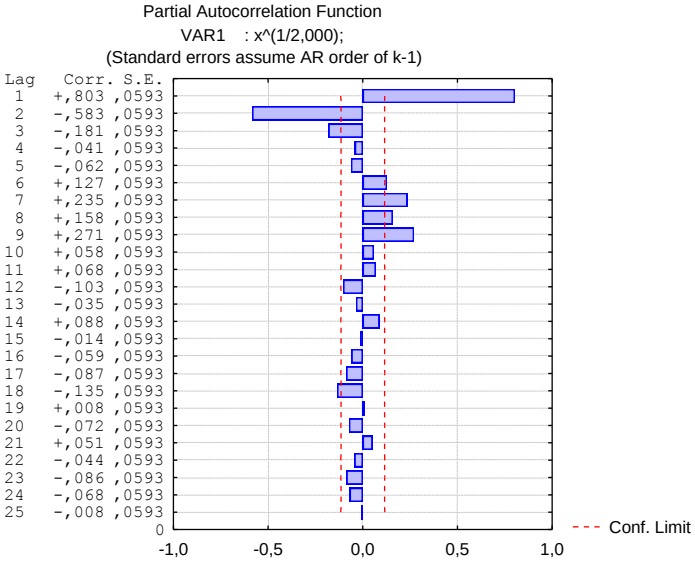


Рис. 9.11. ЧАКФ ряда $T(Z_t) = \sqrt{Z_t}$

Сначала предположим, что мы имеем три значимые ЧАКФ, а выходящие за допустимый интервал значения с номерами 7–9 получаются в результате больших ошибок оценивания выборочной ЧАКФ. Выберем в окне *Single Series ARIMA* (рис. 9.12), появляющемся после нажатия кнопки ARIMA & autocorrelation functions, значения первичных параметров модели ARMA(p, q): $p = 3$, $q = 0$ и точный метод оценивания Exact.

После нажатия кнопки ***OK* (*Begin parameter estimation*)** получаем окно результатов (рис. 9.13). Более подробные результаты оценивания, которые можно получить после нажатия кнопки ***Summary: Parameter estimation***, приведены в табл. 9.1 с указанием в каждом столбце точечных оценок, стандартных ошибок оценивания, значений t-статистики с соответствующим уровнем значимости, а также 95%-х доверительных интервалов (интервалов, в которых с вероятностью 0,95 расположены точные значения параметров). Из анализа полученных результатов очевидно, что все три оценки значимы, и это говорит в пользу построенной модели.

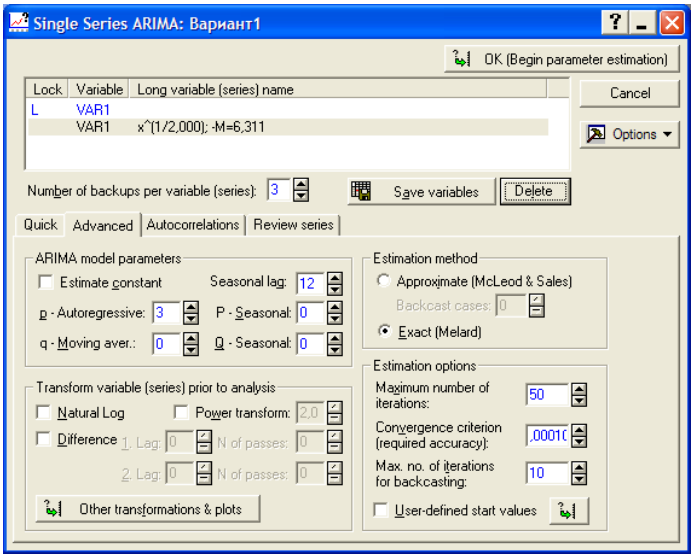


Рис. 9.12. Выбор значений первичных параметров и метода оценивания

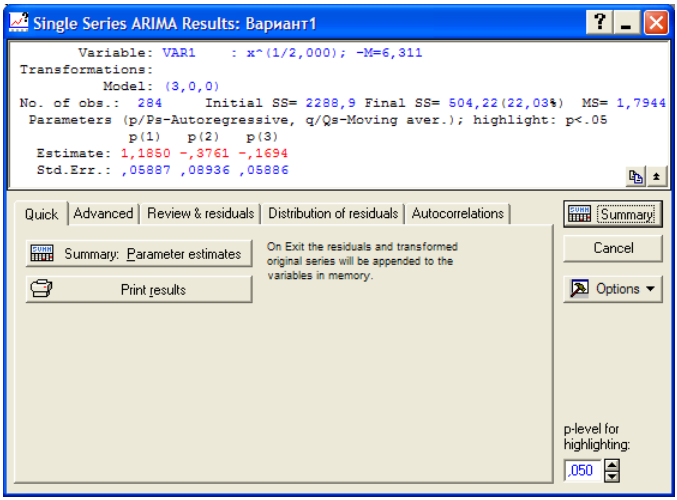


Рис. 9.13. Окно результатов оценивания параметров модели AR(3)

Таблица 9.1

Результаты оценивания параметров модели AR(3)

Параметры	Значение параметров	Асимптотическая стандартная ошибка	Асимптотическое значение *t*-статистики *t*(281)	*p*-значение	Нижняя граница 95%-го доверительного интервала	Верхняя граница 95%-го доверительного интервала
p(1)	1,1850	0,058869	20,12971	0,0000	1,0691	1,3008
p(2)	–0,3760	0,089362	–4,20849	0,0000	–0,5519	–0,2001
p(3)	–0,1693	0,058863	–2,87767	0,0043	–0,2852	–0,0535

Третий этап построения модели (после идентификации и оценивания параметров) состоит в проверке адекватности очередной пробной модели. При проверке адекватности моделей временных рядов большое значение имеет анализ остатков. Все те погрешности идентификации, которые не были учтены ранее, можно выявить при анализе остатков: например, неоднородность дисперсии, ненормальность распределения. Однако при анализе самое важное – проверить, является ли полученный временной ряд остатков белым шумом, как и постулируется при определении модели ARIMA. Лучшее средство для этого – посмотреть коррелограммы выборочных АКФ и ЧАКФ остаточного ряда. Для процесса белого шума теоретические значения коррелограмм равны нулю.

Для построенной модели ARMA(3) мы получаем АКФ и ЧАКФ остатков, приведенные на рис. 9.14 и 9.15. Значимое превышение корреляциями пороговых значений на лагах 9–12 говорит о неадекватности построенной модели. Выход из данной ситуации – в построении альтернативной модели, в которой учитывается значимость этих корреляций, – модели AR(9). Результаты оценивания параметров этой модели представлены на рис. 9.16. Как видим, значения оценок параметров $\phi_3 - \phi_6$ не являются значимыми. Однако мы не можем отбросить эти члены, так как значимыми являются следующие за ними оценки параметров $\phi_7 - \phi_9$.

Анализ АКФ и ЧАКФ остатков, изображенных на рис. 9.17 и 9.18, говорит о том, что остатки можно считать белым шумом. Нормальность остатков подтверждается гистограммой и графиком на нормальной вероятностной бумаге (рис. 9.19 и 9.20). Следовательно, модель адекватна.

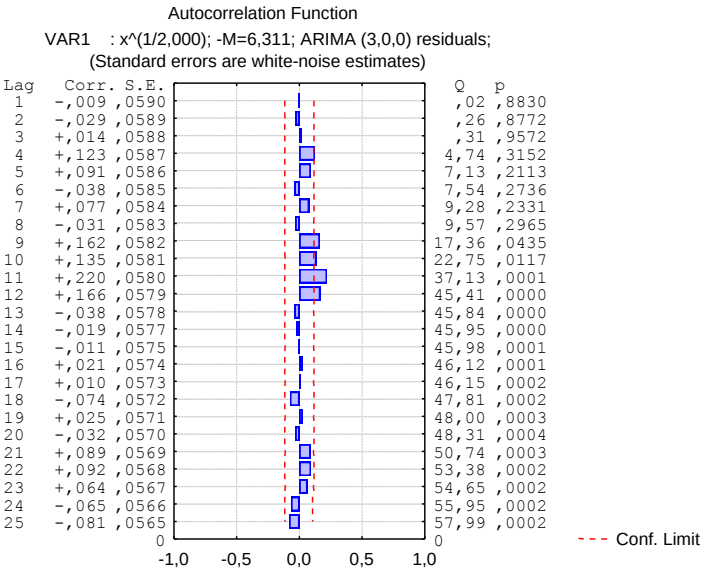


Рис. 9.14. АКФ остатков модели AR(3)

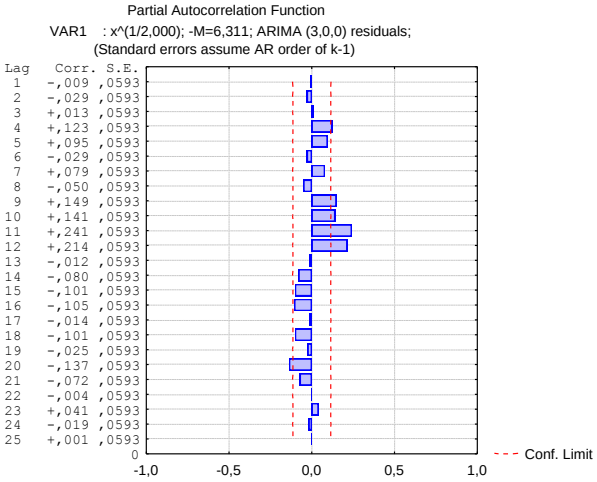


Рис. 9.15. ЧАКФ остатков модели AR(3)

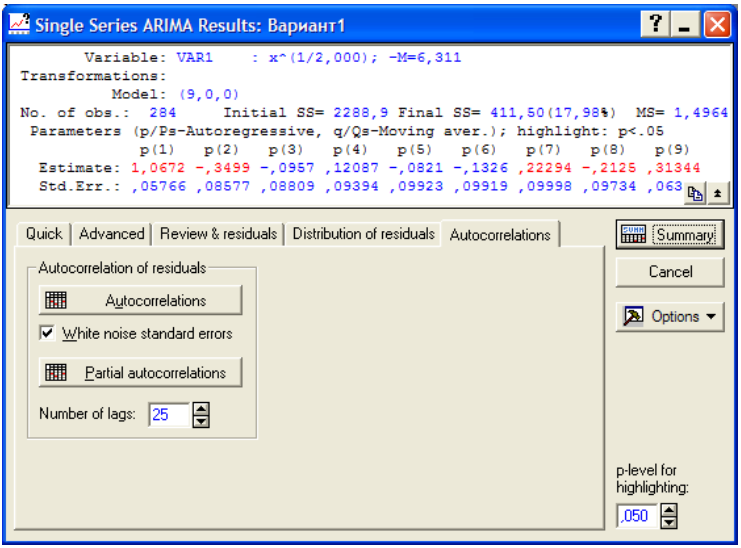


Рис. 9.16. Окно результатов оценивания параметров модели AR(9)

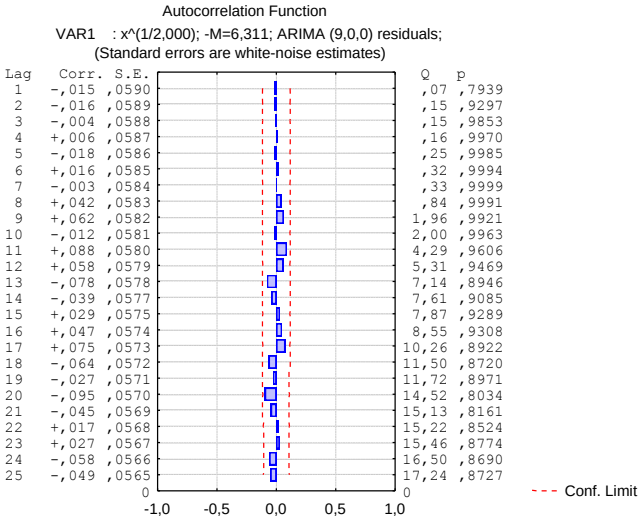


Рис. 9.17. АКФ остатков модели AR(9)

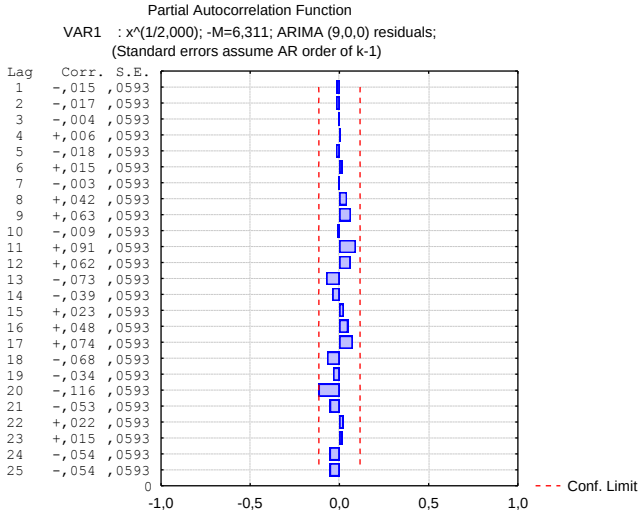


Рис. 9.18. ЧАКФ остатков модели AR(9)

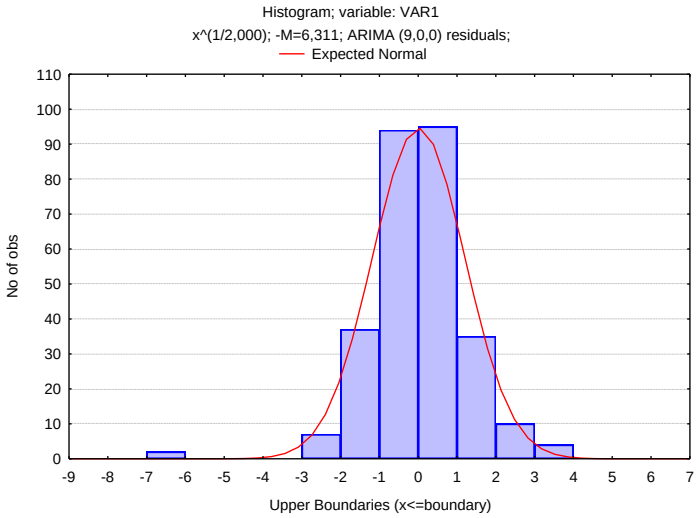


Рис. 9.19. Гистограмма остатков модели AR(9)

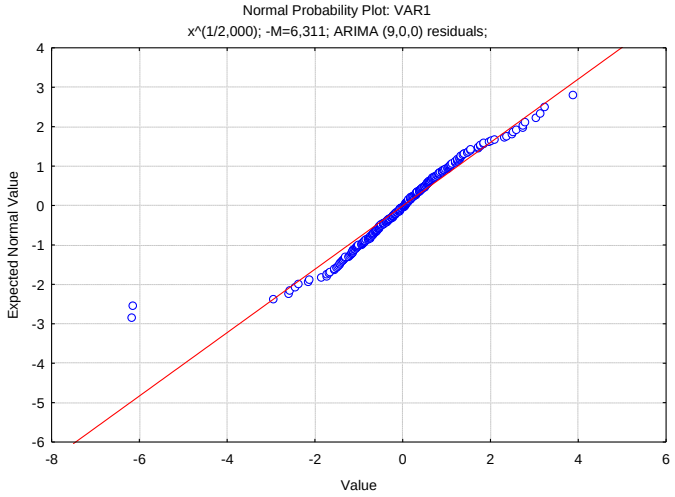


Рис. 9.20. Нормальный вероятностный график остатков модели AR(9)

В итоге построенная адекватная модель для описания количества солнечных пятен имеет вид

$$(1-1{,}067B+0{,}350B^2+0{,}096B^3-0{,}121B^4+0{,}082B^5+$$

$$+0{,}132B^6-0{,}222B^7+0{,}212B^8-0{,}313B^9)(\sqrt{Z_t}-6{,}311)=a_t.$$

9.2. АНАЛИЗ ВРЕМЕННОГО РЯДА КОЛЕБАНИЙ ИНДЕКСА ДОУ-ДЖОНСА в 1980-х годах

Рассмотрим временной ряд, содержащий данные о ежемесячных средних значениях индекса Доу-Джонса в период с 1981 по 1990 год. Сначала построим и проанализируем график временного ряда (рис. 9.21). Ряд, очевидно, имеет тренд, являющийся следствием нестационарности среднего. Что касается нестационарности дисперсии, то для начала «забудем» о ее стабилизации и посмотрим, к чему это приведет.

Попытаемся учесть нестационарность среднего и дисперсии с помощью операции взятия разности. Но сначала надо убедиться в обоснованности этой операции и построить коррелограммы АКФ и ЧАКФ (рис. 9.22 и 9.23).

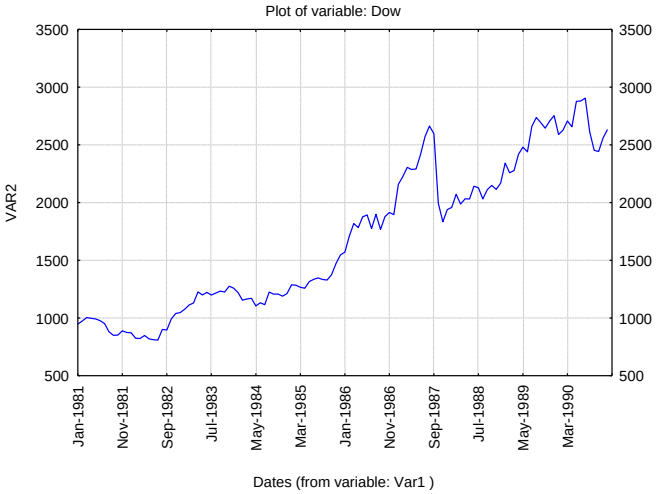


Рис. 9.21. График временного ряда колебаний индекса Доу-Джонса

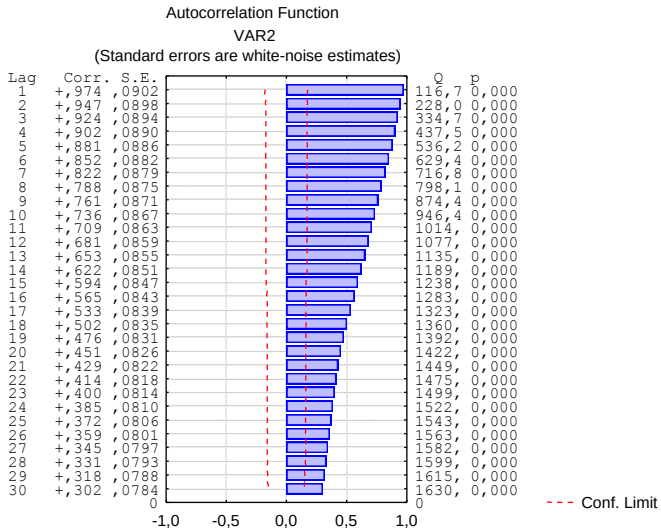


Рис. 9.22. Коррелограмма АКФ временного ряда колебаний индекса Доу-Джонса

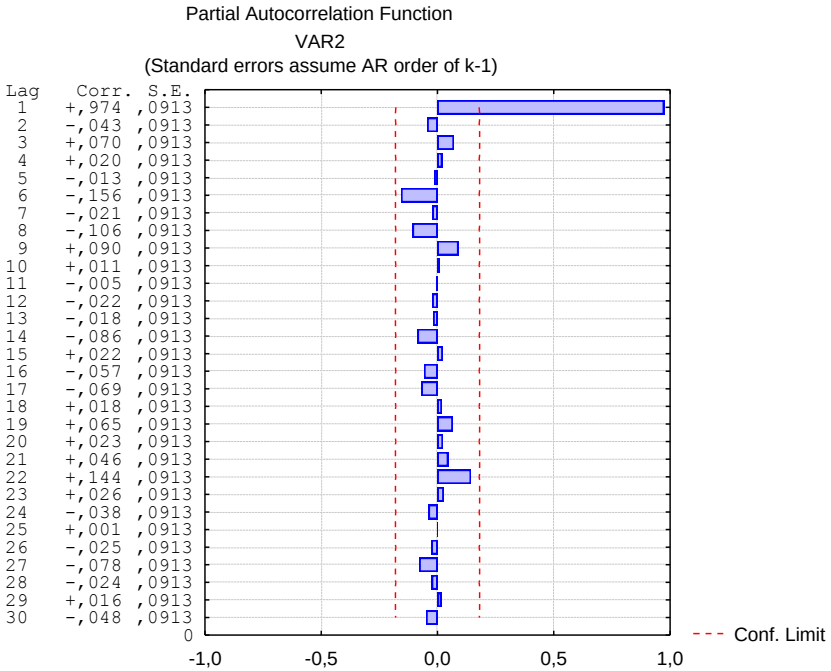


Рис. 9.23. Коррелограмма ЧАКФ временного ряда колебаний индекса Доу-Джонса

Слабо убывающий характер АКФ и единственная значимая ЧАКФ на лаге $k = 1$ убеждают в необходимости взятия разности. График и коррелограммы преобразованного ряда представлены на рис. 9.24–9.26. Анализ коррелограмм АКФ и ЧАКФ позволяет сделать вывод о том, что полученный после взятия разности ряд приближенно можно считать белым шумом, т.е. по этому признаку получена адекватная модель. Однако обратим внимание на то, что разброс значений нового ряда ощутимо увеличивается с течением времени, т.е. проявляется нестационарность дисперсии. Причем на графике ряда после взятия разности это увеличение становится более очевидным, чем на исходном графике.

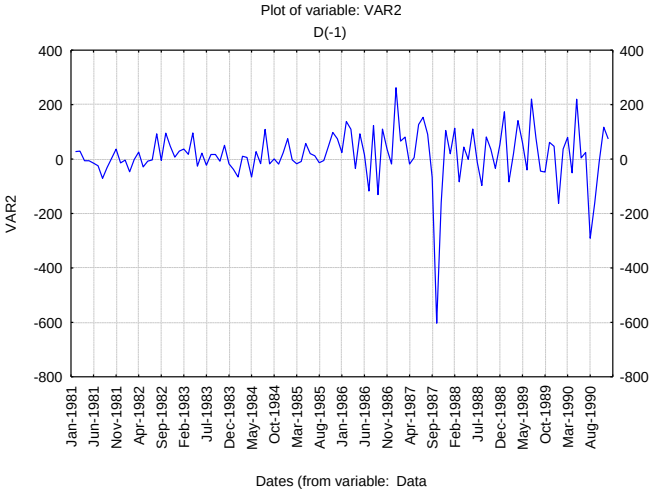


Рис. 9.24. График разности временного ряда колебаний индекса Доу-Джонса

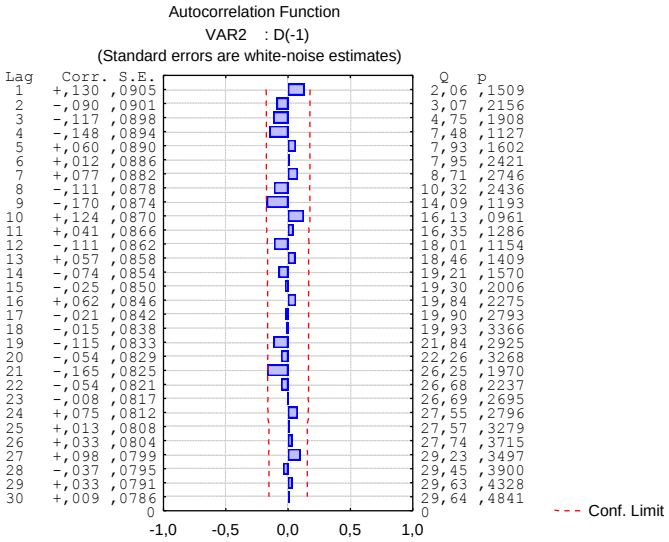


Рис. 9.25. АКФ разности временного ряда колебаний индекса Доу-Джонса

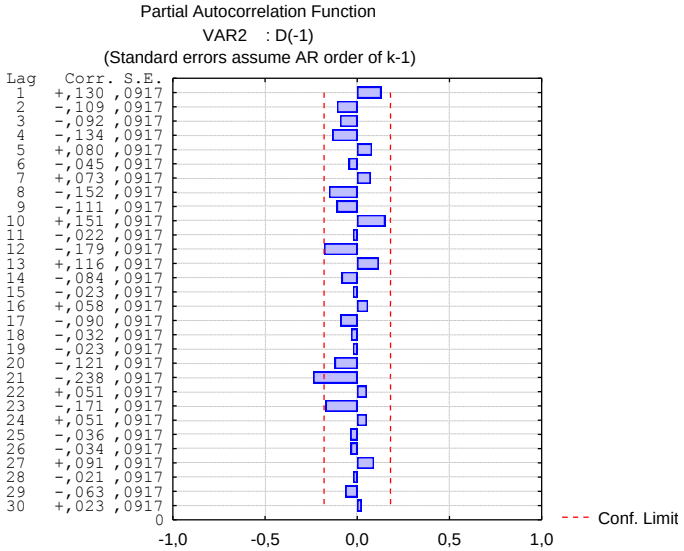


Рис. 9.26. ЧАКФ разности временного ряда колебаний индекса Доу-Джонса

В данном случае «невыявленная» на первом этапе идентификации нестационарность дисперсии отчетливо проявилась после взятия разности. Вернемся к исходной переменной, содержащей данные временного ряда, и определим преобразование, стабилизирующее дисперсию. В данном случае для большей наглядности рекомендуется подбирать наилучшее преобразование, глядя на график ряда после проведенного преобразования и последующего взятия разности (приведения к стационарному ряду). В этом случае нестационарность дисперсии особенно очевидна.

Поступая таким образом, можно определить, что в данном случае логарифмирование является искомым преобразованием, стабилизирующим дисперсию. График ряда и коррелограммы после логарифмирования и последующего взятия разности представлены на рис. 9.27–9.29. По их виду можно сделать вывод об адекватности модели $(1-B)\ln Z_t = a_t$ для описания данных. Эту модель можно назвать моделью случайного блуждания после логарифмирования.

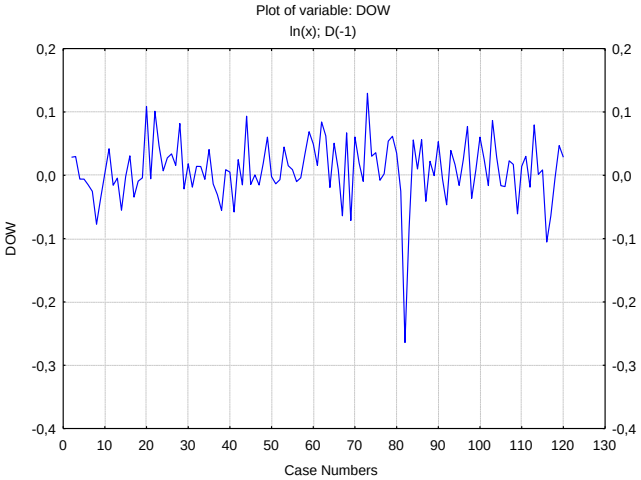


Рис. 9.27. График первой разности прологарифмированного временного ряда

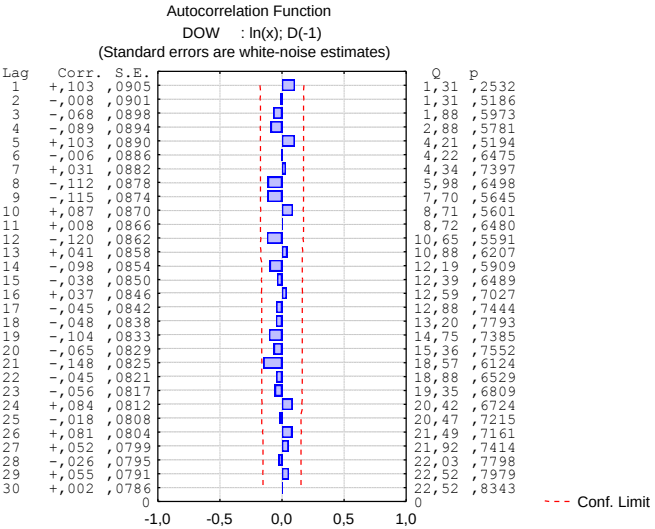


Рис. 9.28. АКФ первой разности прологарифмированного временного ряда

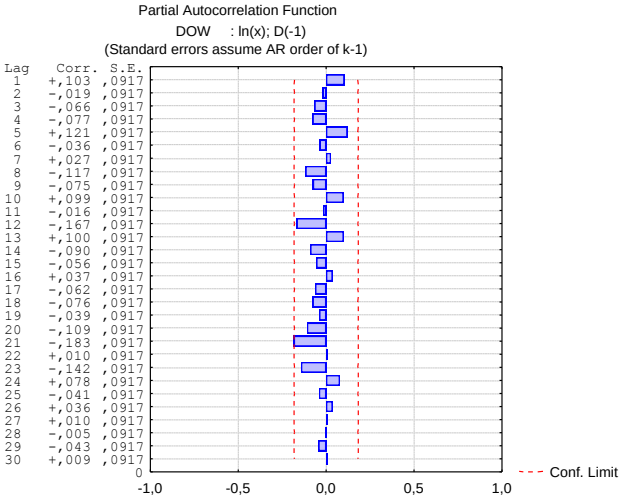


Рис. 9.29. ЧАКФ первой разности прологарифмированного временного ряда

9.3. АНАЛИЗ СЕЗОННОГО ВРЕМЕННОГО РЯДА

В качестве примера идентификации сезонных временных рядов выберем переменную ***Sparkling*** из файла **WINE.STA**. Для анализа данных, содержащихся в файле **WINE.STA**, выберем процедуру ARIMA (рис. 9.30).

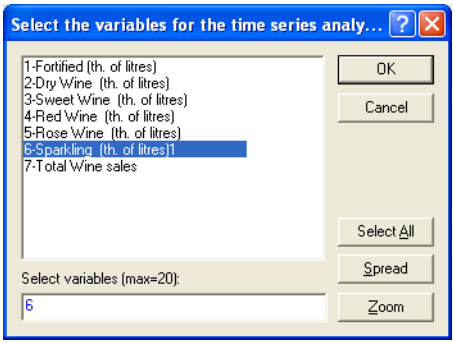


Рис. 9.30. Выбор переменной временного ряда для анализа

Прежде всего посмотрим данные по переменной ***Sparkling*** в графической интерпретации (рис. 9.31).

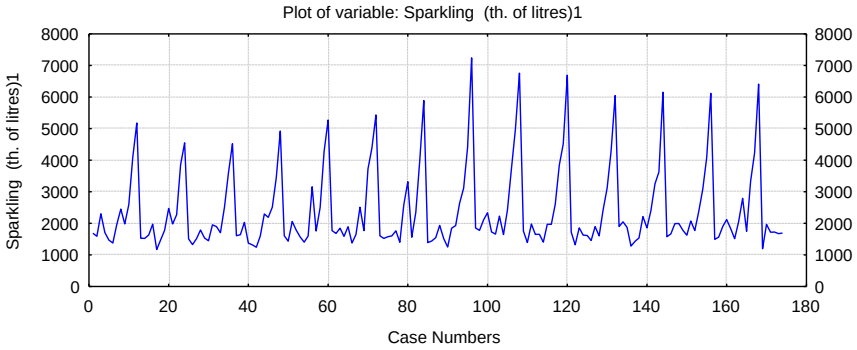


Рис. 9.31. График реализации игристых вин на территории Австралии с января 1980 по июнь 1994 года

Выявим имеющиеся особенности ряда (нестационарность, тренд и т.п.). Очевидно, что на данном графике присутствуют отчетливые годовые периоды, имеются выраженные пики, амплитуда колебаний непостоянна. В ряде имеется незначительный тренд – среднее значение объема продаж постепенно увеличивается, тренд подтверждается видом АКФ и ЧАКФ. Для построения графиков автокорреляционной и частной автокорреляционной функции в диалоговом окне ***ARIMA*** выберем закладку ***Autocorrelations***. Коррелограммы изображены на рис. 9.32–9.33. По виду коррелограмм можно предположить не только регулярную нестационарность, но и сезонную нестационарность, которые могут быть устранены операцией взятия сначала первой, а затем сезонной разности.

Устраним выявленные особенности, применив подходящие преобразования. Выполним трансформацию выходных данных, для стабилизации дисперсии – логарифмирование, в результате получим график преобразованного временного ряда (рис. 9.34).

Далее возьмем разность первого порядка. Данное преобразование позволяет избавиться от линейного тренда. Получим график временного ряда без линейного тренда, изображенный на рис. 9.35.

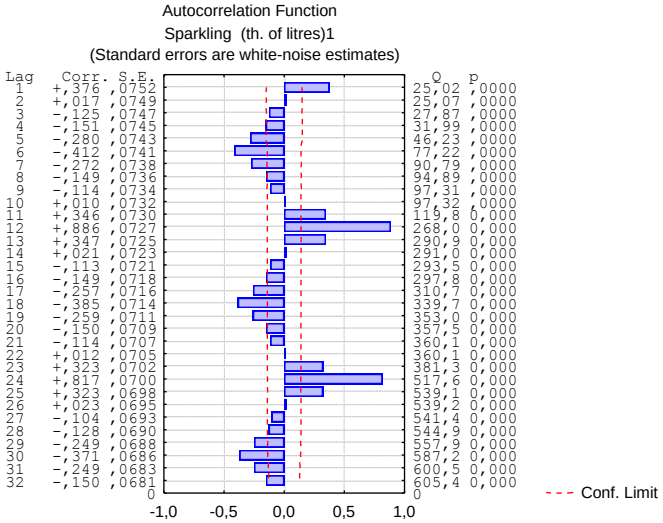


Рис. 9.32. График функции автокорреляции исходного ряда

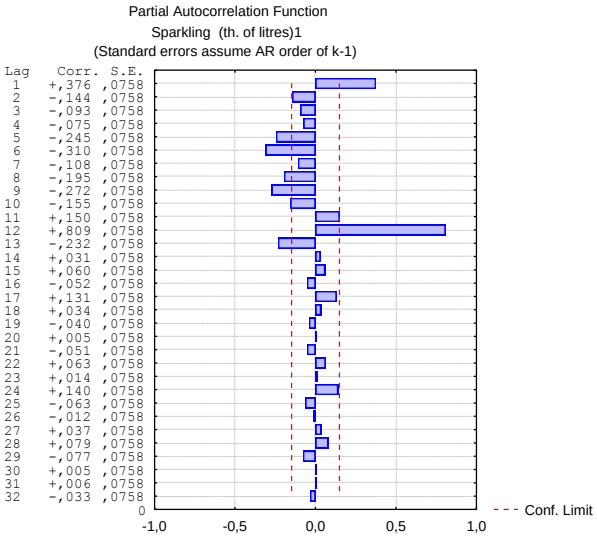


Рис. 9.33. График функции частной автокорреляции исходного ряда

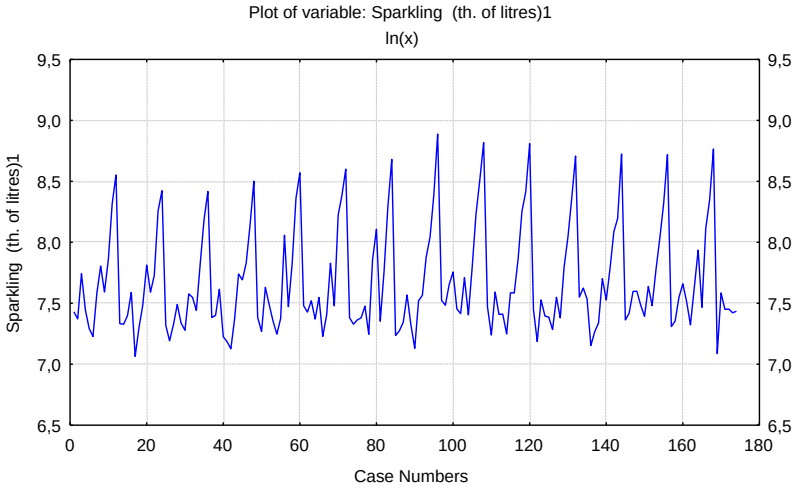


Рис. 9.34. График реализации игристых вин после логарифмирования

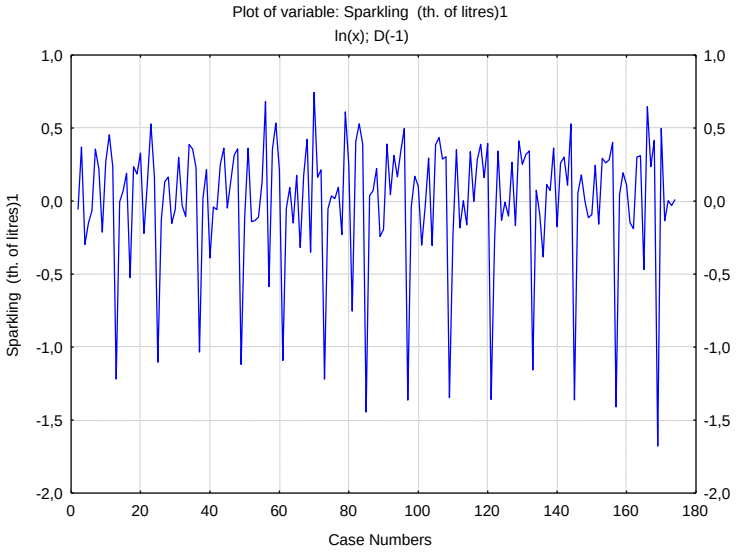


Рис. 9.35. График реализации игристых вин после логарифмирования и взятия разности

Далее возьмем сезонную разность. В результате получим ряд, очень похожий на стационарный (рис. 9.36).

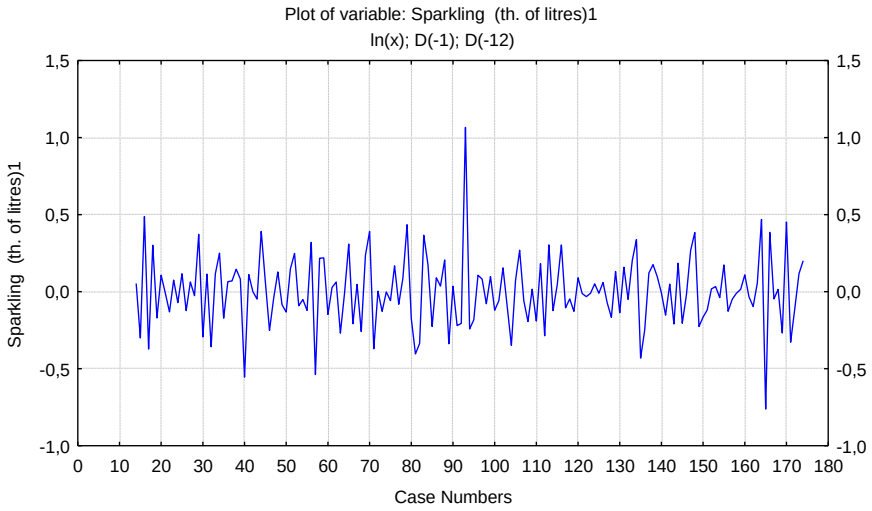


Рис. 9.36. График реализации игристых вин после преобразований: Natural log, Differencing с лагом 1, Differencing с лагом 12

Каждый раз, возвращаясь после очередного преобразования в окно ***Transformation of variables***, смотрим автокорреляции и частные автокорреляции преобразованного ряда, стараясь увидеть закономерности в характере их изменения.

После устранения нестационарности переходим в главное окно ***ARIMA***, закрыв окно ***Transformation of variables.*** Осуществляем установки для оценивания параметров последовательности, указанной на рис. 9.37, после чего нажимаем кнопку ***OK*** (***Begin parameter estimation***). Результаты оценивания приведены в окнах, изображенных на рис. 9.38 и 9.39. Полученные оценки параметров значимы.

Проанализируем остатки построенной модели. Гистограмма, график на нормальной вероятностной бумаге, АКФ и ЧАКФ остатков, приведенные на рис. 9.40–9.43, говорят о приемлемой близости распределения остатков к нормальному, а также о их некоррелированности остатков.

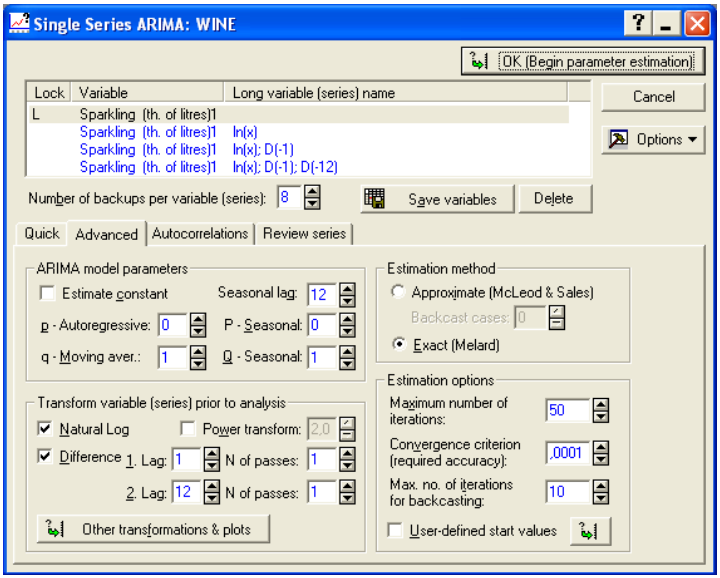


Рис. 9.37. Стартовая панель ***ARIMA***: установка параметров

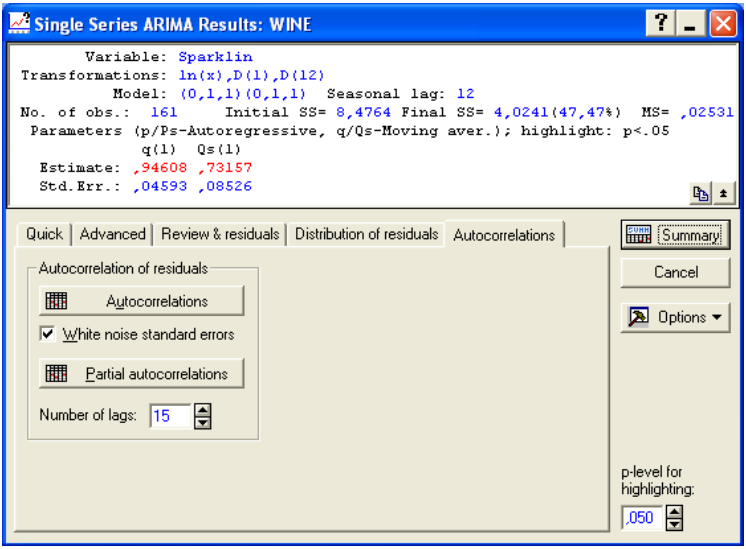


Рис. 9.38. Окно ***Результаты ARIMA***

Paramet.	Input: Sparkling (th. of litres)1 (WINE.sta) Transformations: ln(x),D(1),D(12) Model:(0,1,1)(0,1,1) Seasonal lag: 12 MS Residual=,02531					
	Param.	Asympt. Std.Err.	Asympt. t(159)	p	Lower 95% Conf	Upper 95% Conf
q(1)	0,946076	0,045927	20,59942	0,000000	0,855369	1,036782
Qs(1)	0,731572	0,085256	8,58092	0,000000	0,563192	0,899951

Рис. 9.39. Таблица оценок параметров ARIMA

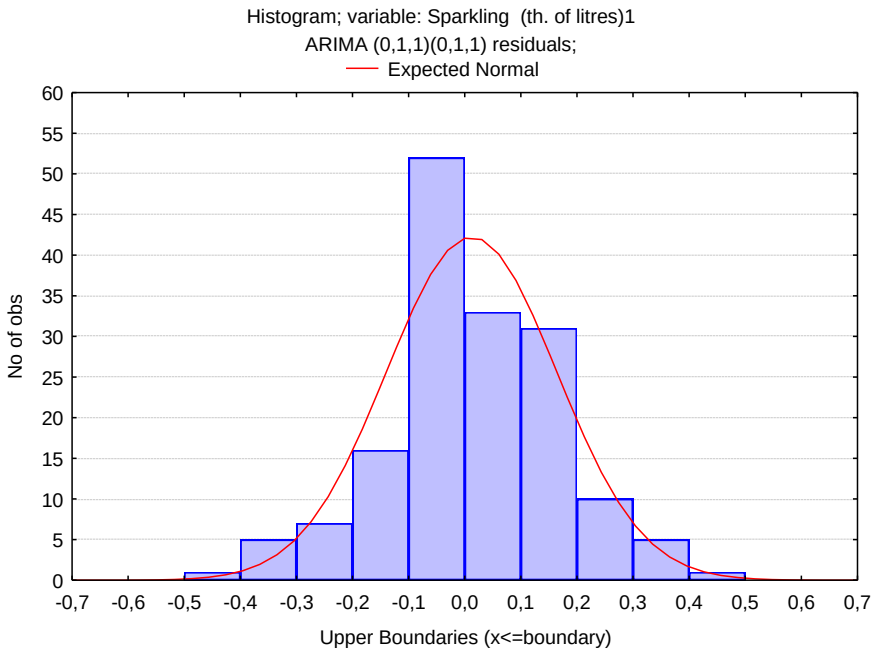


Рис. 9.40. Гистограмма распределения остатков

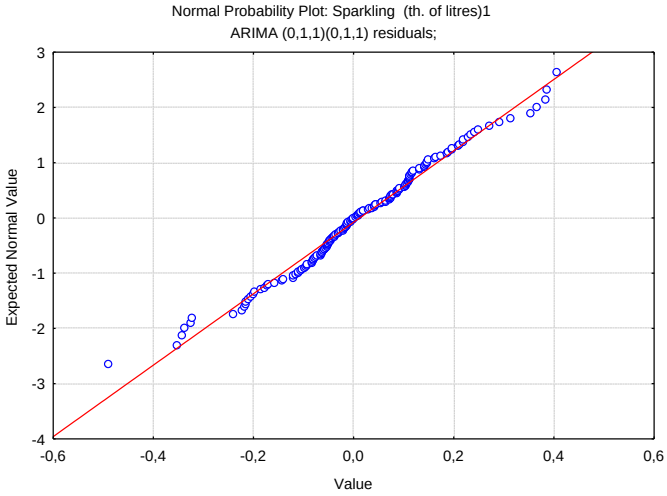


Рис. 9.41. График остатков на нормальной вероятностной бумаге

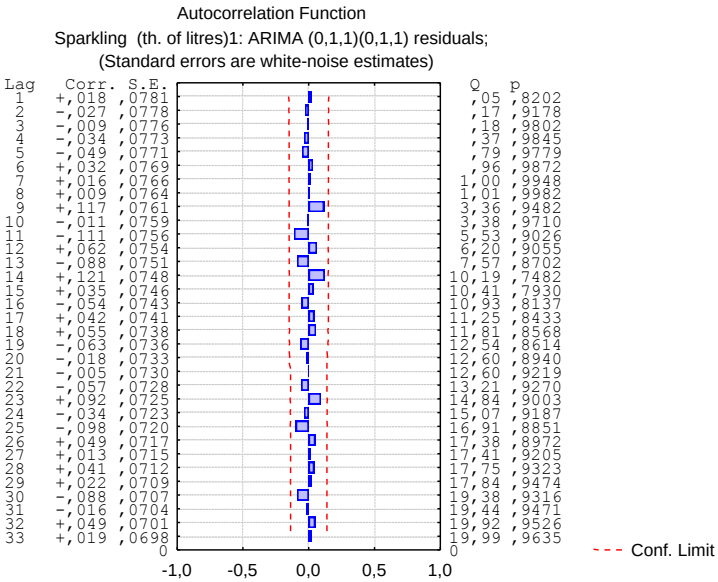


Рис. 9.42. График автокорреляционной функции остатков

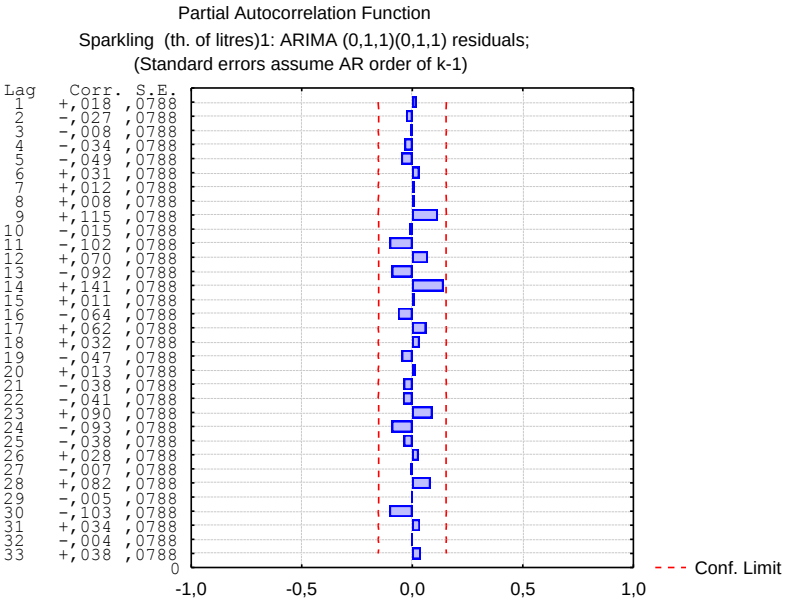


Рис. 9.43. График частной автокорреляционной функции остатков

Проведенный анализ показывает, что построенная ARIMA-модель достаточно адекватно описывает наблюдаемый временной ряд:

$$(1-B^{12})(1-B)\ln Z_t = (1-0{,}95B)(1-0{,}73B^{12})\ .$$

10. АНАЛИЗ ИНТЕРВЕНЦИЙ И ОПРЕДЕЛЕНИЕ ВЫБРОСОВ

Временные ряды часто нарушаются под влиянием некоторых внешних событий, таких как праздники, забастовки, распродажи и др. Такие внешние события называют *интервенциями.* В данном разделе вводится специальная методология, называемая *анализом интервенций*, позволяющая оценить влияние внешних событий на временной ряд. С помощью этой методологии можно, например, исследовать, улучшились ли значимо экономические показатели после внедрения той или иной новой экономической политики? Или повлияло ли введение нового уголовного закона на изменение уровня преступности?

Сначала рассмотрим случай, когда известно, в какой именно момент времени произошло внешнее событие (момент интервенции известен). Этот тип анализа называется анализом интервенций. Затем данный метод анализа обобщается для изучения влияния внешних событий, для которых момент интервенции неизвестен. Таким образом мы приходим к общей методологии анализа выбросов временных рядов.

10.1. АНАЛИЗ ИНТЕРВЕНЦИЙ

10.1.1. МОДЕЛЬ ИНТЕРВЕНЦИЙ

В данном подразделе представим самую общую методологию анализа моделей интервенций, как она рассматривается в [14], а в следующем подразделе приведем ее усечение до методологии, реализованной в системе *Statistica*. Предположим, что в момент времени T происходит интервенция – некоторое событие, предположительно влияющее на временной ряд. Нас интересует, есть ли подтверждение того, что данное событие действительно повлияло на временной ряд, например, увеличило уровень ряда. И если влияние есть, то насколько оно сильное?

Предположим, что для анализа этой проблемы использован двухвыборочный *t*-критерий. В таком случае можно сравнить две выборки: одну, состоящую из данных, собранных до интервенции, вторую, состоящую из данных, собранных после интервенции. Однако использование *t*-критерия предполагает нормальность и независимость элементов выборки. И хотя *t*-критерий известен своей робастностью к предположению нормальности, он чрезвычайно чувствителен к нарушению предположения независимости. Поэтому для анализа временных рядов *t*-критерий фактически не применим. В данном подразделе будет описана методология анализа интервенций как расширение описанной ранее методологии построения моделей ARIMA.

Определим переменные интервенции двух типов. Одна из них представляет собой интервенцию, произошедшую в момент времени T и продолжающую действовать и после момента возникновения. В этом случае интервенция есть ступенчатая функция

$$S_t^{(T)} = \begin{cases} 0, & t < T, \\ 1, & t \geq T. \end{cases} \tag{10.1}$$

Другая переменная действует только в момент интервенции и далее прекращает свое действие. Она представляется импульсной функцией

$$P_t^{(T)} = \begin{cases} 0, & t = T, \\ 1, & t \neq T. \end{cases} \tag{10.2}$$

Заметим, что импульсная функция $P_t^{(T)}$ может быть получена путем взятия разности от ступенчатой функции $S_t^{(T)}$: $P_t^{(T)} = S_t^{(T)} - S_{t-1}^{(T)} = (1 - B)S_t^{(T)}$. Поэтому модель интервенций можно строить на основе любой из этих двух переменных. Выбор переменной обычно основывается на удобстве интерпретации.

После выбора типа интервенции надо определить, как будет реагировать система, генерирующая временной ряд, на эту интервенцию. Существует много возможных откликов на ступенчатую и импульсную интервенции. Приведем некоторые, наиболее часто встречающиеся.

1. Фиксированное неизвестное воздействие интервенции чувствуется спустя b периодов после интервенции. Таким образом, в зависимости от типа переменной интервенции, отклик на нее выражается

$$\omega B^b S_t^{(T)} \tag{10.3}$$

или

$$\omega B^b P_t^{(T)}. \tag{10.4}$$

2. Воздействие чувствуется спустя b периодов после интервенции, но отклик постепенный:

$$\frac{\omega B^b}{1 - \delta B} S_t^{(T)} \tag{10.5}$$

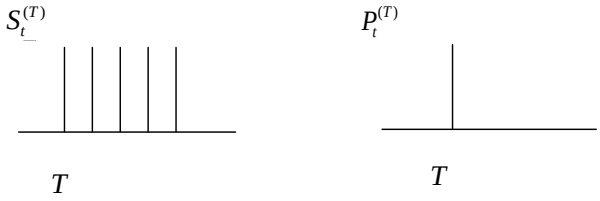
или

$$\frac{\omega B^b}{1 - \delta B} P_t^{(T)}, \tag{10.6}$$

где $0 \le \delta \le 1$. Для $\delta = 0$ отклики (10.5) и (10.6) сводятся к откликам (10.3) и (10.4) соответственно. При $\delta = 1$ отклик возрастает линейно неограниченно. Для большинства случаев справедливо $0 < \delta < 1$, когда отклик постепенный.

Проиллюстрируем приведенные отклики с помощью графиков для $0 < \delta < 1$ (рис. 10.1).

Интервенция



Отклик на интервенцию

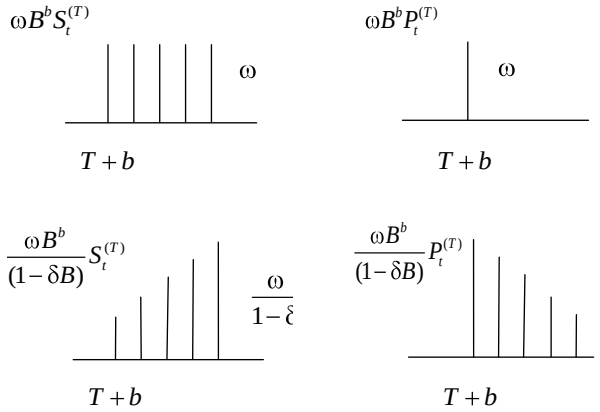


Рис. 10.1. Отклики на ступенчатую и импульсную интервенции

Более разнообразные отклики могут быть получены с использованием различных комбинаций ступенчатой и импульсной функций, например:

$$\frac{\omega_0 B}{1-\delta B} P_t^{(T)} + \omega_1 B S_t^{(T)}. \qquad (10.7)$$

Однако, так как $P_t^{(T)} = (1-B) S_t^{(T)}$, этот комбинированный отклик (10.7) может быть также записан в виде

$$\left[\frac{\omega_0 B}{1-\delta B} + \frac{\omega_1 B}{1-B}\right] P_t^{(T)}.$$

Данная модель полезна для представления явлений, в которых интервенция генерирует отклик, постепенно уменьшающийся до какого-то уровня, а затем сохраняющийся постоянным в течение длительного времени. Это, например, отклик, описывающий воздействие рекламы на объем продаж (график его изображен на рис. 10.2, *а*), или отклик, описывающий влияние увеличения цены (или налога) на импорт (график изображен на рис. 10.2, *б*).

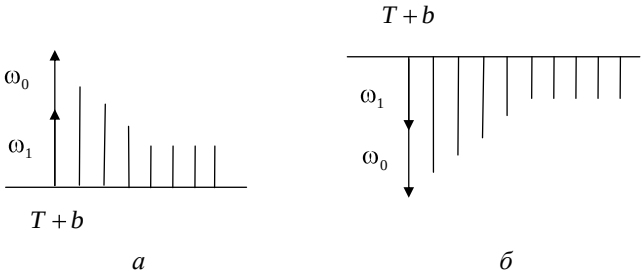


Рис. 10.2. Отклики на комбинированную интервенцию:
а – $\omega_0 > 0$, $\omega_1 > 0$; *б* – $\omega_0 < 0$, $\omega_1 < 0$

В более общем виде отклик на интервенцию можно представить как рациональную функцию

$$\frac{\omega(B)B^b}{\delta(B)},$$

где $\omega(B) = \omega_0 - \omega_1 B - ... - \omega_s B^s$ и $\delta(B) = 1 - \delta_1 B - ... - \delta_r B^r$ – полиномы относительно оператора запаздывания B; b – временная задержка эффекта интервенции. Вес ω_j в полиноме $\omega(B)$ часто представляет ожидаемые начальные эффекты интервенции. Полином $\delta(B)$, с другой стороны, определяет долговременный эффект интервенции. Предполагается, что корни уравнения $\delta(B) = 0$ лежат на единичном круге или вне его. Единичный корень соответствует линейному отклику, а корни вне единичного круга – постепенному отклику.

Для моделирования случая нескольких интервенций имеется следующий общий класс моделей:

$$Z_t = \sum_{j=1}^{k} \frac{\omega_j(B)B^{b_j}}{\delta_j(B)} I_{jt} + \frac{\theta(B)}{\psi(B)} a_t, \qquad (10.8)$$

где I_{jt}, $j = 1, 2, \ldots, k$ – переменные интервенции. Эти переменные бывают ступенчатыми или импульсными функциями. В общем случае они могут быть подходящими индикаторными переменными. Конкретный вид $\frac{\omega_j(B)B^{b_j}}{\delta_j(B)}$ для j-й интервенции постулируется на основе ожидаемой формы отклика на известную интервенцию.

По отношению к переменным интервенции I_{jt} ряд, свободный от интервенций, называют шумовым рядом и обозначают N_t, а соответствующую ему модель называют шумовой моделью. Шумовая модель имеет вид

$$N_t = \frac{\theta(B)}{\psi(B)} a_t. \qquad (10.9)$$

Ее идентификация осуществляется в классе моделей ARIMA(p, d, q):

$$\psi(B)N_t = \theta(B)a_t$$

с помощью стандартной процедуры, описанной в предыдущих разделах, на основе шумового временного ряда N_t, представляющего собой одномерный временной ряд Z_t до момента интервенции: $N_t = Z_t : t < T$.

После получения адекватной шумовой модели приступают к идентификации модели интервенций (10.8). При этом полиномы $\psi(B)$ и $\theta(B)$ берутся из шумовой модели, а вид отклика на интервенцию выбирается в соответствии с имеющимися данными. После идентификации модели оцениваются параметры методом условного или безусловного максимального правдоподобия – так, как это было описано ранее. Потом проводится диагностика модели. Указанные три этапа (идентификация, оценивание параметров, проверка адекватности) повторяются до получения адекватной модели.

10.1.2. АНАЛИЗ МОДЕЛЕЙ С ИНТЕРВЕНЦИЕЙ В СИСТЕМЕ *STATISTICA*

В системе *Statistica* имеется возможность проводить анализ моделей ARIMA с интервенцией по описанной выше методологии. Для этого в модуле ***TimeSeries/Forecasting*** имеется клавиша ***Interrupted time series analysis*** (*Анализ прерванных временных рядов*, иначе называемый анализом интервенций), нажатие которой выдает окно, представленное на рис. 10.3. Цель анализа интервенций – оценить влияние одного или нескольких дискретных событий на последующие значения временного ряда.

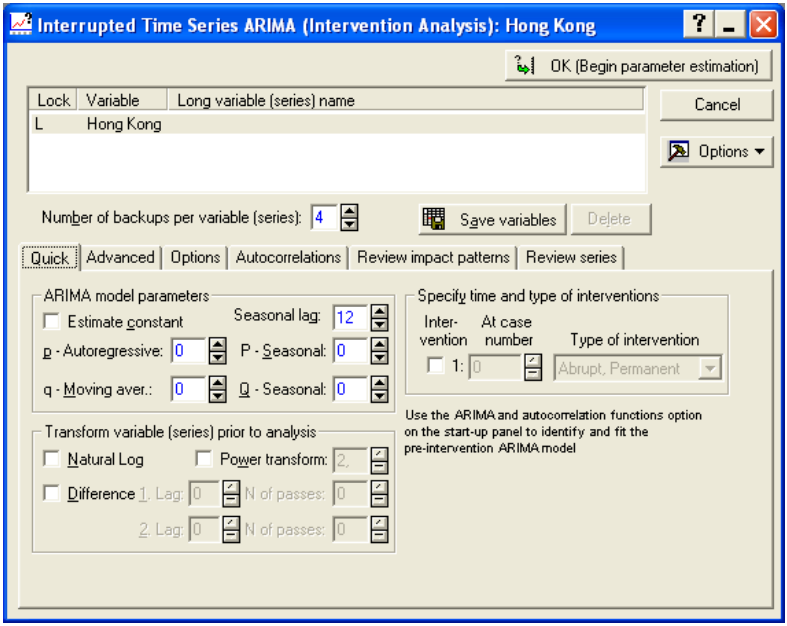


Рис. 10.3. Окно диалога *Interrupted Time Series Analysis*

Нажатие указанной клавиши приводит к диалогу, аналогичному диалогу *ARIMA&Autocorrelations functions* и содержащему шесть закладок: ***Quick, Advanced, Options, Autocorrelations, Review impact patterns*** и ***Review series.*** Изменения касаются, по сути, только добавки к ARIMA-модели, связанной с откликом на интервенцию, заключающейся в дополнении к авторегрессионным параметрам и параметрам

скользящего среднего параметров интервенции, отвечающих за краткосрочное или долгосрочное изменение общего «уровня» временного ряда.

Выбор закладки ***Review impact patterns*** позволяет просмотреть типы откликов на интервенции, предлагаемые системой *Statistica*. Всего в системе *Statistica* можно выбрать только три разных типа внешнего воздействия на временной ряд из всех возможностей, рассмотренных нами в предыдущем подразделе. Первый тип – *Abrupt/Permanent* – с резким скачком ряда на новый уровень в момент интервенции и последующим долговременным сохранением ряда на этом новом уровне, изображен на рис. 10.4.

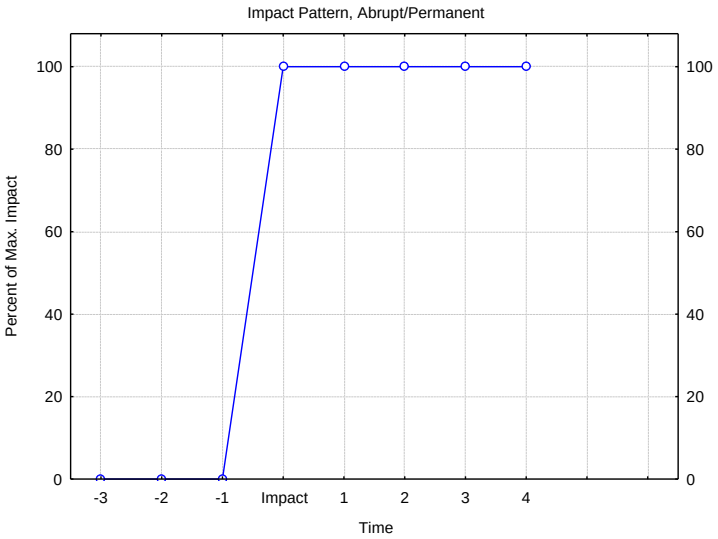


Рис. 10.4. Первый тип интервенции – *Abrupt/Permanent*

Данный простейший тип интервенции моделируется с использованием отклика (10.3) при $b = 0$, т.е.

$$\text{Impact}_t = \omega S_t^{(T)},$$

где, как и прежде, $S_t^{(T)}$ – ступенчатая функция, описываемая формулой (10.1).

Если мы ограничиваемся только одной интервенцией первого типа в модели, то модель ARIMA с интервенцией в данном случае имеет вид

$$Z_t = \omega S_t^{(T)} + \frac{\theta(B)}{\psi(B)} a_t .$$

И тогда к параметрам модели ARIMA добавляется всего один параметр ω, смысл которого – изменение уровня ряда в результате интервенции.

Второй тип интервенции – *Gradual/Permanent* – с постепенным переходом ряда на новый уровень с момента интервенции и последующим длительным сохранением ряда на этом новом уровне, изображен на рис. 10.5. Этот тип интервенции моделируется с использованием отклика (10.5) при $b = 0$, т.е.

$$\mathrm{Impact}_t = \frac{\omega}{1 - \delta B} S_t^{(T)},$$

или, эквивалентно,

$$(1 - \delta B)\mathrm{Impact}_t = \mathrm{Impact}_t - \delta \mathrm{Impact}_{t-1} = \omega S_t^{(T)}.$$

Если мы ограничиваемся только одной интервенцией второго типа в модели, то модель ARIMA с интервенцией в данном случае имеет вид

$$Z_t = \frac{\omega}{1 - \delta B} S_t^{(T)} + \frac{\theta(B)}{\psi(B)} a_t .$$

И тогда к параметрам модели ARIMA добавляются два параметра, ω и δ. Постепенность перехода регулируется параметром δ, который в случае, изображенном на рис. 10.5, принимает значение 0,3. Параметр ω отвечает за итоговое изменение уровня ряда: асимптотическое (получаемое в пределе) изменение уровня ряда равно $\frac{\omega}{1 - \delta}$.

При анализе результатов оценивания параметров важно обратить внимание на то, чтобы оба эти параметра были значимыми. В противном случае можно получить парадоксальные выводы. Например, предположим, что параметр ω не отличается значимо от нуля, а параметр δ, наоборот, оказался статистически значимым. Это означало бы парадок-

сальную ситуацию: интервенция вызывает значимое постепенное изменение, итоговый результат которого значимо не отличается от нуля!

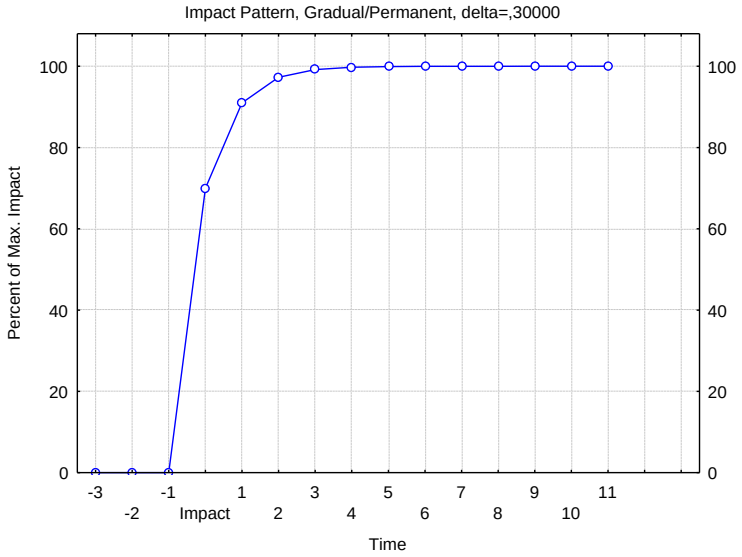


Рис. 10.5. Второй тип интервенции – *Gradual/Permanent*

Третий тип интервенции, изображенный на рис. 10.6, – *Abrupt/Temporary* – с резким переходом ряда на новый уровень с момента интервенции и последующим постепенным возвращением ряда на прежний уровень.

Этот тип интервенции моделируется с использованием отклика (10.6) при $b = 0$, т.е.

$$\text{Impact}_t = \frac{\omega}{1 - \delta B} P_t^{(T)},$$

или, эквивалентно,

$$(1 - \delta B)\text{Impact}_t = \text{Impact}_t - \delta\text{Impact}_{t-1} = \omega P_t^{(T)},$$

где, как и прежде, $P_t^{(T)}$ – импульсная функция, описываемая формулой (10.2).

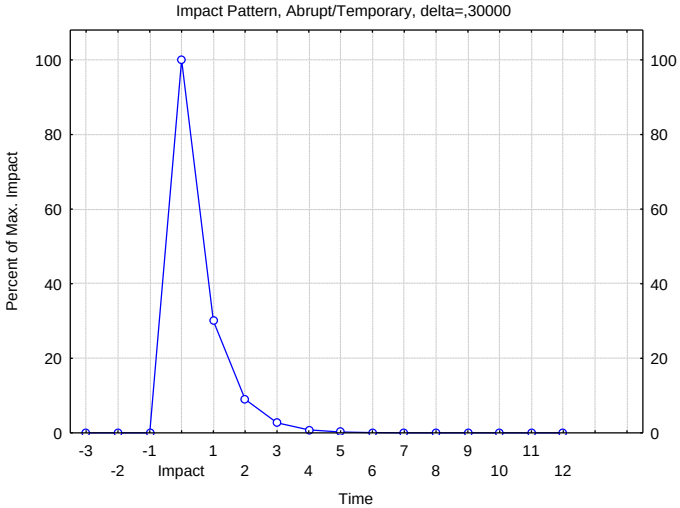


Рис. 10.6. Третий тип интервенции – *Abrupt/Temporary*

Если мы ограничиваемся только одной интервенцией третьего типа в модели, то модель ARIMA с интервенцией в данном случае имеет вид

$$Z_t = \frac{\omega}{1-\delta B} P_t^{(T)} + \frac{\theta(B)}{\psi(B)} a_t .$$

И тогда к параметрам модели ARIMA, как и в предыдущем случае, добавляются два параметра: ω и δ. Параметр ω означает начальное (в момент интервенции) максимальное изменение уровня ряда, которое в последующие моменты времени постепенно снижается в асимптотике до нуля. Постепенность снижения регулируется параметром δ, который в случае, изображенном на рис. 10.6, принимает значение 0,3.

Как и в предыдущем случае, важно обратить внимание на то, чтобы оценки обоих этих параметров интервенции были значимыми. Снова для примера рассмотрим ситуацию, когда параметр ω не отличается значимо от нуля, а параметр δ, наоборот, оказался статистически значимым. Это означало бы, что интервенция не вызывает значимо отличного от нуля начального скачкообразного изменения уровня ряда, который далее выявляет значимо отличное от нуля угасание, что, конечно же, абсурдно!

Анализ временного ряда с интервенцией в системе *Statistica* состоит, как и прежде, из этапов идентификации, оценивания параметров и диагностики. Однако теперь процедура идентификации модели проводится по более сложной схеме и сама состоит из двух этапов.

Первый этап идентификации заключается в построении модели шумового ряда N_t, представляющего собой исходный временной ряд Z_t до момента интервенции: $N_t = Z_t : t < T$. Так как момент интервенции T известен, получить ряд N_t весьма нетрудно. В системе *Statistica* для этого предусмотрена дополнительная возможность – кнопка ***Select cases***, нажатие которой выводит окно, изображенное на рис. 10.7. Включив опцию *Enable Selection Conditions* и выбрав в разделе *Include cases* (Включить случаи) опцию *Specific, selected By case number*, нужно набрать диапазон номеров наблюдений до интервенции, например: 1–160. Заметим, что если рассматривается несколько интервенций, то шумовая модель строится по данным до самой ранней интервенции. Построение модели для шумового ряда проводится по той же методике, по которой строится модель ARIMA в диалоге *ARIMA&Autocorrelations functions*. В результате идентификации мы выдвигаем пробную модель, которая может оказаться стационарной, нестационарной или даже сезонной, далее оцениваем неизвестные параметры, после чего проводим диагностику модели. Процесс продолжается до тех пор, пока мы не построим адекватную шумовую модель.

После построения адекватной шумовой модели проводится второй этап идентификации модели с интервенциями, теперь уже непосредственно в диалоге *Interrupted Time Series Analysis.* Здесь надо указать сначала все те преобразования и все первичные параметры (p, q, P, Q, s), которые были определены для адекватной шумовой модели. Далее в подокне *Specify times and types of intervensions* выбираются моменты и типы интервенций. В закладке ***Quick*** имеется возможность выбрать лишь одну интервенцию, а в закладке ***Advanced*** – до шести различных интервенций! Здесь для каждой интервенции указывается момент времени, в который произошло воздействие (шумовой ряд строился по наблюдениям, предшествующим моменту интервенции), а также один из трех вышеописанных типов откликов на интервенцию. Выбор типа интервенции проводится в соответствии с предварительным анализом графика ряда или исходя из содержательной интерпретации имеющихся данных. График ряда можно посмотреть, выбрав закладку ***Review Series***, как показано на рис. 10.8.

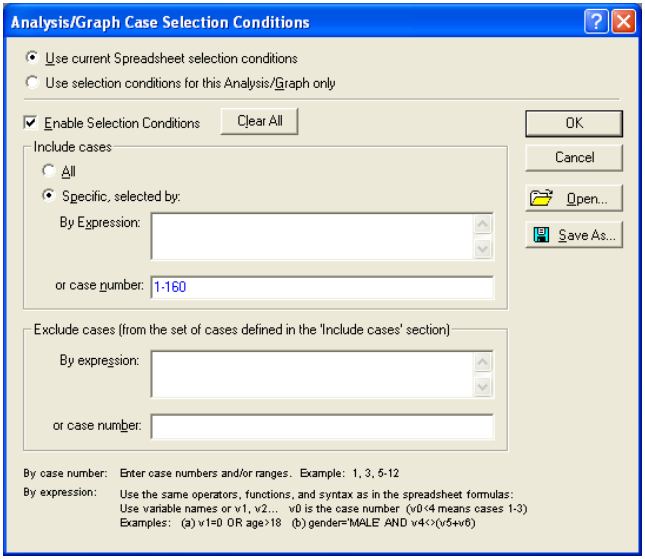


Рис. 10.7. Окно, выдаваемое по кнопке ***Select cases***

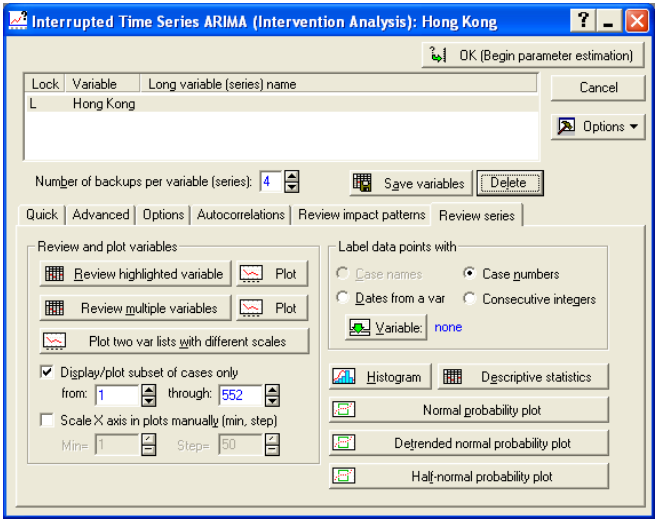


Рис. 10.8. Окно просмотра графика ряда – закладка ***Review Series***

Следующий этап после идентификации модели с интервенциями – оценивание параметров. Метод оценивания определяется при выборе закладки ***Options***, как показано на рис. 10.9. Как и в диалоге *ARIMA&Autocorrelations functions*, для оценивания параметров предлагается три модификации метода условного максимального правдоподобия: приближенный метод, приближенный с использованием обратного прогноза и точный метод, а также предоставляются возможности для выбора параметров сходимости итерационного процесса. Параметры оцениваются после нажатия кнопки ***OK* (*Begin parameter estimation*)**, в результате чего выдается окно результатов, полностью аналогичное соответствующему окну результатов диалога *ARIMA& Autocorrelations functions*.

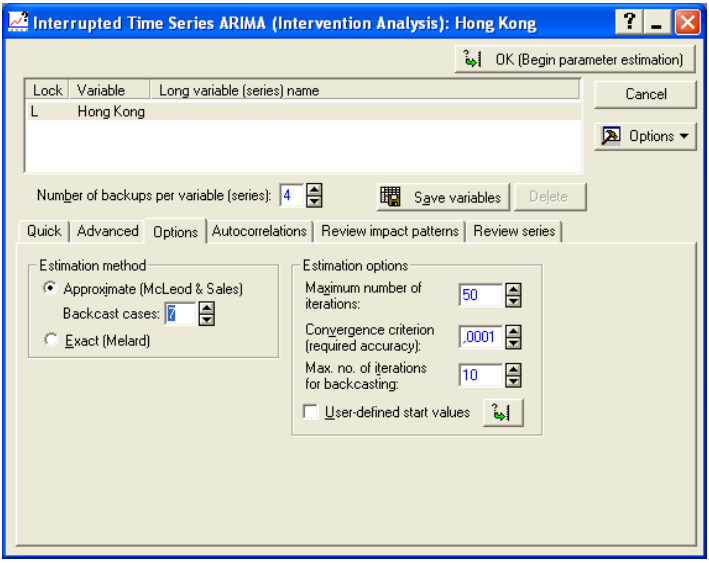


Рис. 10.9. Окно выбора метода оценивания и параметров сходимости

Данное окно используется для третьего этапа построения модели – диагностики модели, который полностью аналогичен соответствующему этапу диагностики модели ARIMA, описанному ранее. При анализе следует обратить особое внимание на оценки параметров интервенции ω и δ, чтобы избежать вышеописанных абсурдных ситуаций.

10.1.3. ПРИМЕРЫ АНАЛИЗА ИНТЕРВЕНЦИЙ

В этом подразделе приведем примеры анализа временных рядов с интервенциями.

10.1.3.1. АНАЛИЗ ВРЕМЕННОГО РЯДА РЕАЛИЗАЦИИ СЛАДКИХ ВИН В АВСТРАЛИИ

Моделью **Interrupted AR1MA — Прерванная ARIMA** описываются временные ряды, имеющие на графике следующий типичный вид. На рис. 10.10 показана динамика реализации вина сладкого сорта на территории Австралии с января 1980 по июнь 1994 года (данные взяты из файла *sweet wine.sta).* В какой-то момент времени резко увеличивается уровень продаж рассматриваемого временного ряда.

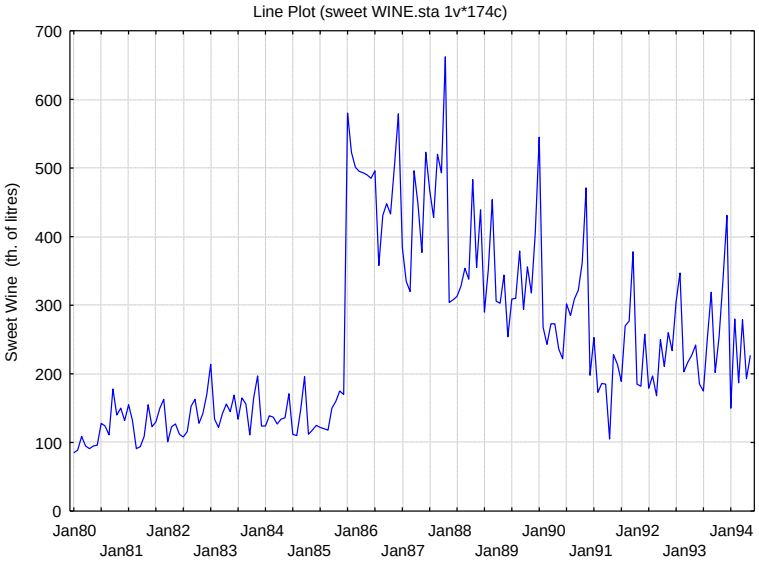


Рис. 10.10. Пример временного ряда с интервенцией

В более общей ситуации в определенный момент времени на ряд оказывается сильное внешнее воздействие – существенно более сильное, чем флуктуации ряда. Это воздействие может быть мгновенным, начавшимся и почти тут же закончившимся, или, начавшись, очень быстро достигнуть определенного уровня и на этом уровне остано-

виться, или после начала постепенно достигнуть определенного уровня и на нем останавливаться.

На рис. 10.10 видно мгновенное воздействие, резко поднимающее временной ряд вверх. Случайное поведение ряда после воздействия похоже на поведение ряда до воздействия. Поэтому воспользуемся диалогом *Interrupted ARIMA*.

При построении шумовой модели ряда с использованием данных до интервенции было выяснено, что ряд нестационарный как по среднему, так и по дисперсии. В результате последовательности логарифмического преобразования, взятия первой разности и взятия сезонной разности с лагом 12 временной ряд преобразуется к стационарному виду. В результате построена модель шумового ряда: $(1-B^{12})(1-B)\ln N_t = (1-\Theta B)(1-\Theta B^{12})a_t$. Следующий шаг – построение модели с интервенцией по полному набору данных.

В стартовом окне **Time Series Analysis** после нажатия на кнопку ***Interrupted time series analysis*** появляется стартовая панель ***Interrupted Time Series ARIMA – Прерванная ARIMA*** (рис. 10.11). В левой части окна задаются характеристики построенной шумовой модели, в правой – время и тип интервенции. Всего можно выбрать до шести интервенций из трех различных типов.

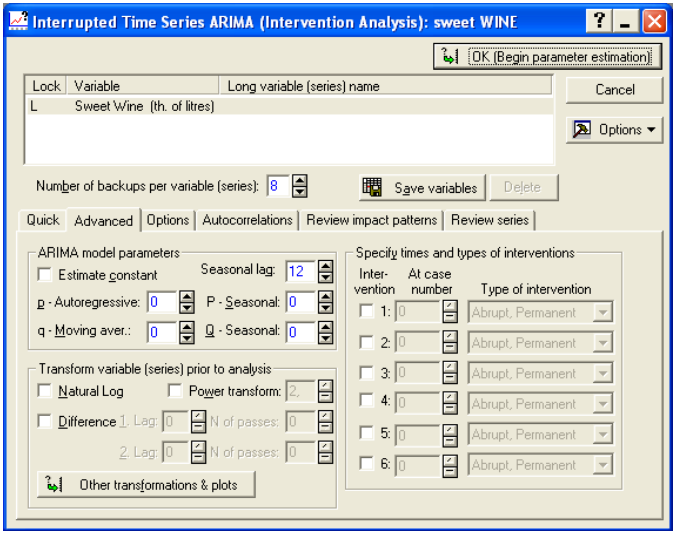


Рис. 10.11. Стартовая панель *Прерванная ARIMA*

Введем значения параметров для оценивания временного ряда согласно проведенным исследованиям ряда (рис. 10.12) и нажмем кнопку ***OK*** **(*Begin parameter estimation*)**. Результаты оценивания представлены на рис. 10.13 и 10.14. Как видим, все параметры модели высокозначимые.

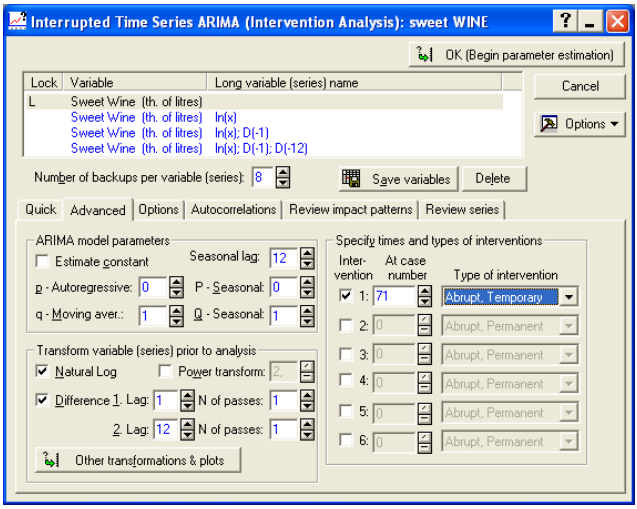


Рис. 10.12. Стартовая панель *Прерванная ARIMA*: установка параметров

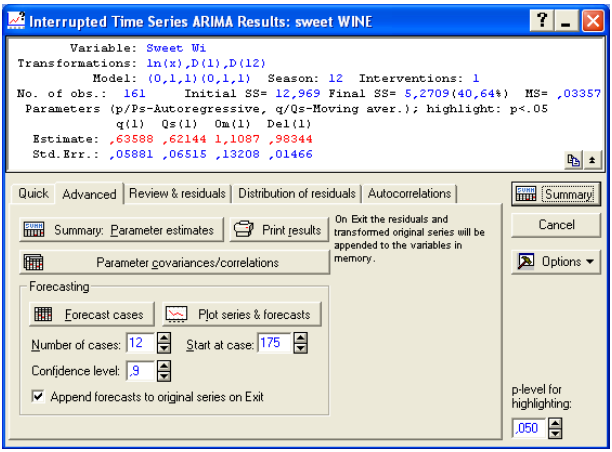


Рис. 10.13. Окно *Результаты Прерванной ARIMA*

Paramet.	Input: Sweet Wine (th. of litres) (sweet WINE.sta) Transformations: ln(x),D(1),D(12) (Interrupted ARIMA) Model:(0,1,1)(0,1,1) Seasonal lag: 12 MS Residual=,03357							
	Param.	Asympt. Std.Err.	Asympt. t(157)	p	Lower 95% Conf	Upper 95% Conf	Interv. Case No.	Interv. Type
q(1)	0,635878	0,058815	10,81155	0,000000	0,519708	0,752048		
Qs(1)	0,621441	0,065147	9,53901	0,000000	0,492762	0,750119		
Omega(1)	1,108731	0,132079	8,39444	0,000000	0,847849	1,369612	71	Abr/Temp
Delta(1)	0,983438	0,014661	67,07898	0,000000	0,954480	1,012396	71	Abr/Temp

Рис. 10.14. Таблица оценок параметров прерванной ARIMA

Произведем анализ остатков, как это делалось ранее. Графики на рис. 10.15 и 10.16 подтверждают близость распределения остатков к нормальному.

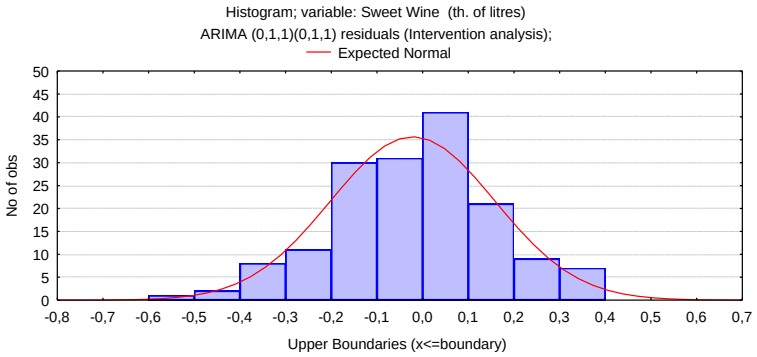


Рис. 10.15. Гистограмма распределения остатков

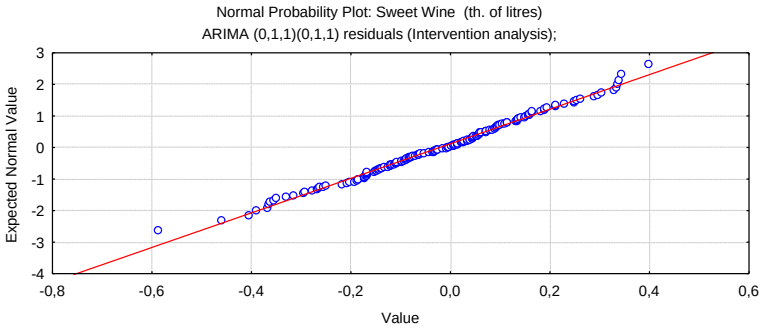


Рис. 10.16. График распределения остатков

Графики АКФ и ЧАКФ остатков построенной модели (рис. 10.17 и 10.18) не выявляют каких-либо закономерностей и коррелированности остатков (лежат внутри коридора ошибки).

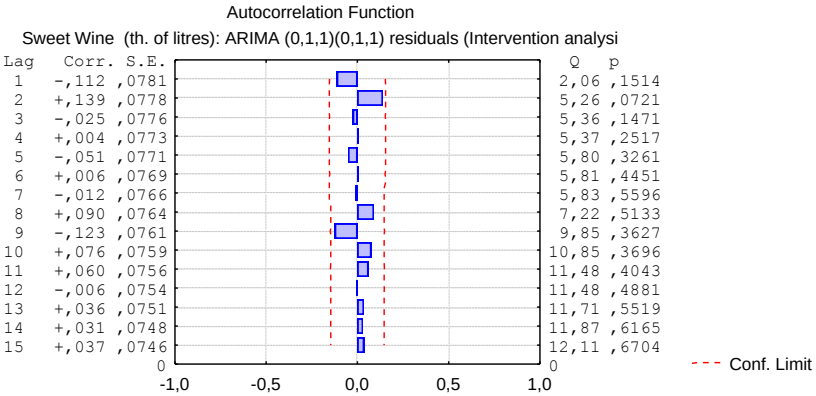


Рис. 10.17. График автокорреляционной функции остатков

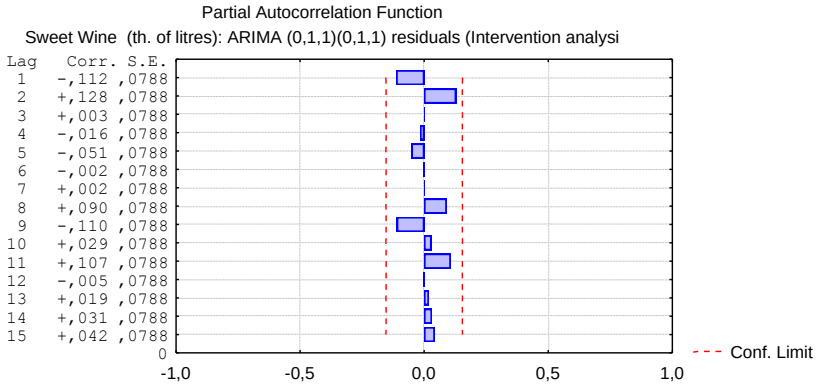


Рис. 10.18. График ЧАКФ остатков

Таким образом, можно сделать вывод, что следующая модель достаточно адекватно описывает наблюдаемый временной ряд:

$$(1-B)(1-B^{12})\ln Z_t = \frac{1{,}11}{1-0{,}98B}P_t^{(T)} + (1-0{,}64B)(1-0{,}62B^{12})a_t\,.$$

10.1.3.2. АНАЛИЗ ВРЕМЕННОГО РЯДА АНАЛИЗА ИНВЕСТИЦИЙ ГОНКОНГА

Рассмотрим пример анализа интервенций, приведенный в книге [14].

Сначала обратимся к небольшой исторической справке.

Гонконг – Британская королевская колония с первой Опиумной войны (1839–1842) между Китаем и Британией. Колония включает остров Гонконг и близкие к нему земли – полуостров Коулун и Новые Территории. Остров Гонконг был передан Британии по Нанкинскому договору как непосредственный результат Опиумной войны. Так как Китай в XIX веке потерял свое былое могущество, Британия приобрела полуостров Коулун в 1860 году и свободную от уплаты 99-летнюю аренду на Новые Территории в 1898 году.

Ныне Гонконг обладает одним из трех самых крупных в мире контейнерных портов. Аэропорт Гонконга обслуживает более 30 международных авиалиний. Твердо придерживаясь экономической политики свободного порта, Гонконг стал процветающим торговым и финансовым центром, притягивающим огромные зарубежные инвестиции, особенно из США. В 1982 году Британия начала переговоры с китайским коммунистическим правительством об окончании в 1997 году договора аренды Новых Территорий. Утверждая, что договор о передаче всей территории Гонконга был неравным и несправедливым, китайское правительство опубликовало 16 июля 1982 года свое первое заявление (проект) о возможном возвращении Гонконгу суверенитета. Это заявление вызвало огромное беспокойство у жителей и инвесторов Гонконга. Гонконгский доллар резко упал до небывало низкой отметки на зарубежных валютных рынках. Рынок недвижимости стал сокращаться. Фондовый рынок упал ниже, чем во время глобального экономического кризиса в 1973 году.

Авторами книги [14] проведен анализ интервенций для оценивания влияния этого заявления китайского руководства на Гонконгский фондовый рынок. На рис. 10.19 построен график временного ряда ежедневного индекса курса цен акций в период с 16 июля 1981 года по 31 сентября 1983 года, выпускаемого Экономическим и исследовательским центром Гонконга в конце каждого торгового дня. Индекс Ханг Сенг включает котировки 33 компаний Гонконга.

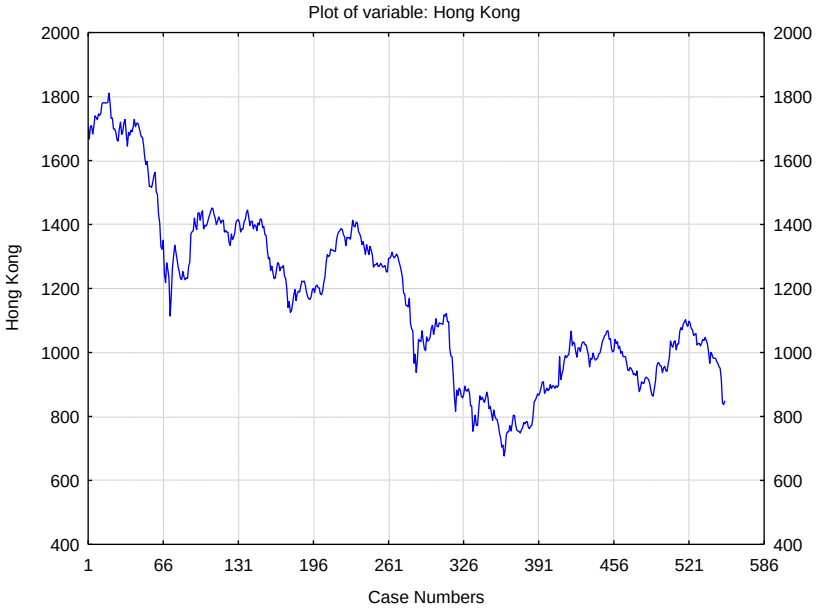


Рис. 10.19. Индекс Ханг Сенг курса акций Гонконга с 16 июля 1981 года по 31 сентября 1983 года

Период между 16 июля 1981 года и 15 июля 1982 года будем считать шумовым рядом N_t, не содержащим интервенции, влияющей на индекс. Этот ряд можно выделить для анализа с помощью кнопки ***Select Cases***, и далее выбрав *Include Cases by Case Numbers*, указав диапазон 1–259 (рис. 10.20).

В книге [14] для шумового ряда выбрана модель случайного блуждания без сдвига:

$$(1-B)N_t = a_t .$$

Действительно, исходные АКФ и ЧАКФ для шумового ряда имеют вид (рис. 10.21 и 10.22), указывающий на ярко выраженный тренд, который можно устранить взятием разности $(1-B)N_t$.

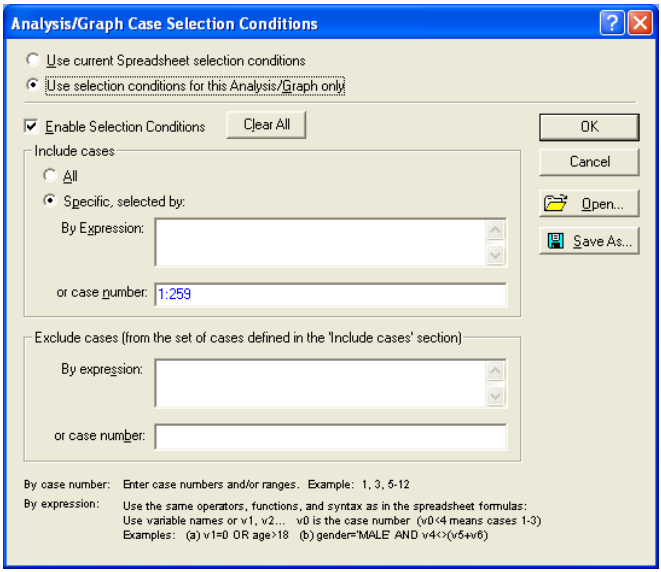


Рис. 10.20. Определение шумового ряда

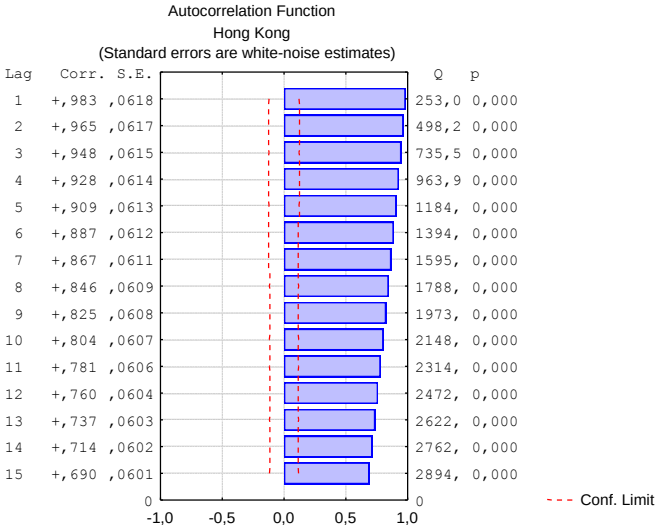


Рис. 10.21. АКФ шумового ряда

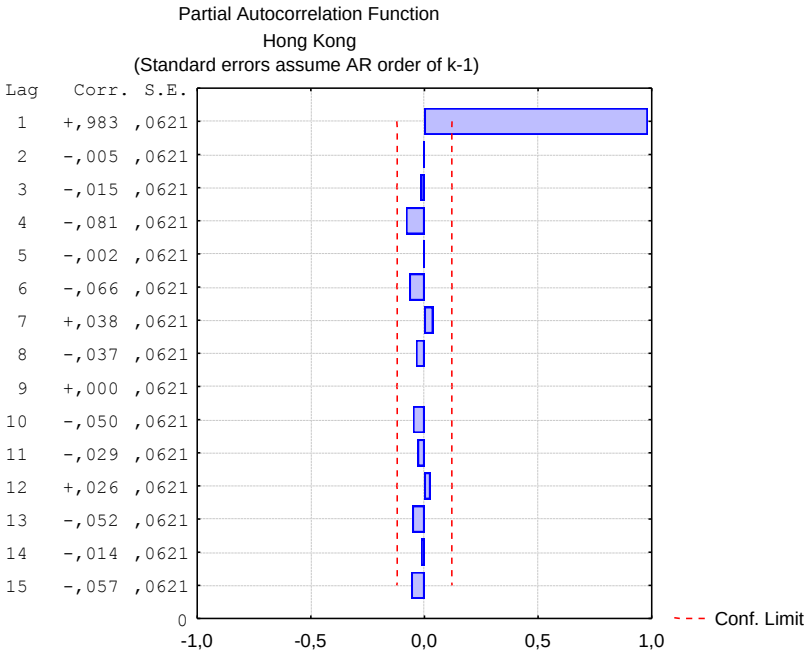


Рис. 10.22. ЧАКФ шумового ряда

После взятия разности характеристики АКФ и ЧАКФ приобретают вид (рис. 10.23 и 10.24), указывающий на то, что полученный ряд является белым шумом. Однако заметим, что ряд, полученный после взятия разности, выявляет некоторую неравномерность дисперсии, соответствующую первому участку времени с большими значениями индекса, что видно из графика преобразованного ряда, изображенного на рис. 10.25. Логарифмирование позволяет несколько уменьшить эту неравномерность (рис. 10.26), хотя в данном случае разница не столь существенна.

Таким образом, для временного ряда индекса курса акций Гонконга мы получили модель случайного блуждания, что в целом согласуется с теорией, согласно которой подобные ряды описываются именно таким типом модели. Следующий шаг – построить модель с интервенцией.

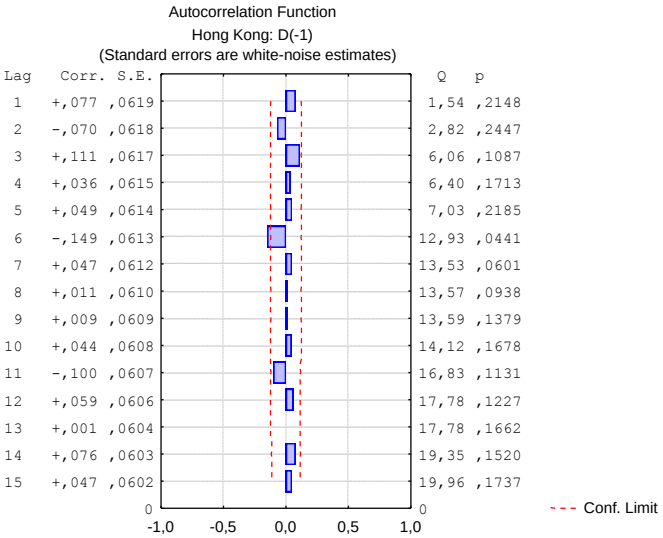


Рис. 10.23. АКФ шумового ряда после взятия разности

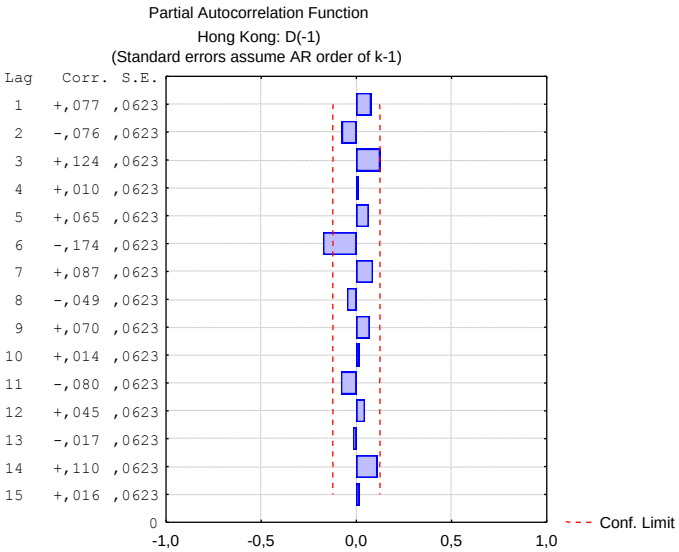


Рис. 10.24. ЧАКФ шумового ряда после взятия разности

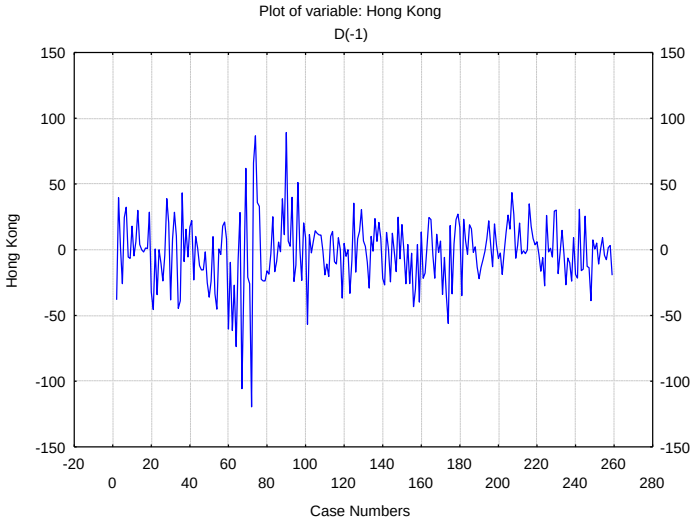


Рис. 10.25. Шумовой ряд после взятия разности

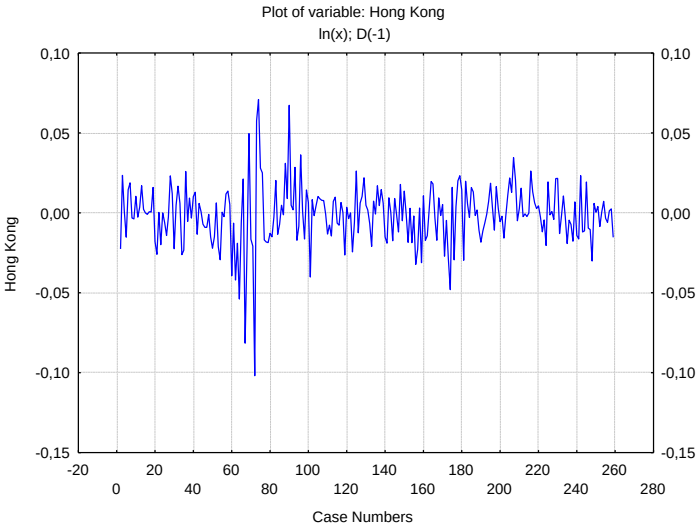


Рис. 10.26. Шумовой ряд после логарифмирования и взятия разности

Предполагая, что заявление китайского правительства должно вызвать немедленное изменение курса акций, можно предложить и следующую функцию отклика:

$$\omega I_t^{(260)},$$

где параметр ω представляет эффект влияния интервенции, $I_t^{(260)}$ – переменная интервенции

$$I_t^{(260)} = \begin{cases} 0, & t < 260, \\ 1, & t \geq 260. \end{cases}$$

Модель интервенции, таким образом, имеет вид

$$Z_t = \omega I_t^{(260)} + \frac{a_t}{1 - B}. \qquad (10.10)$$

В [14] приведены результаты оценивания для этой модели:

Параметр	Оценка	Стандартная ошибка
ω	–1,55	1,05

Таким образом, в терминах модели интервенции (10.10), хотя заявление китайского правительства и вызвало некоторый эффект падения на рынке акций, результат оказался статистически не значимым. Поскольку ряд Z_t имеет большую дисперсию, этот результат не является совсем уж неожиданным.

Повторить оценивание модели (10.10) в системе *Statistica* не удается. При выборе модели, в которой задается только время интервенции и ее тип, но не указан ни один из параметров авторегрессии и скользящего среднего, система выдает сообщение, представленное на рис. 10.27.

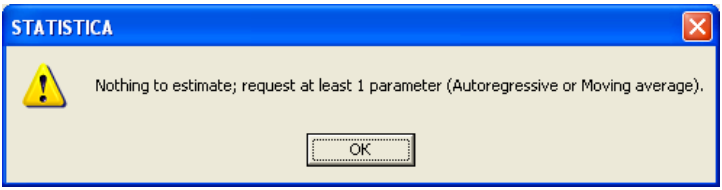


Рис. 10.27. Сообщение о необходимости задания хотя бы одного из параметров авторегрессии и скользящего среднего

Однако вспомним, что модель случайного блуждания – предельный случай модели авторегрессии первого порядка AR(1), и зададим значение порядка авторегрессионного полинома $p = 1$. Кроме того, в данном случае оказывается целесообразным включить в модель детерминированный тренд, поставив галочку на ***Estimate constant***. В качестве единственной интервенции выбираем интервенцию в момент $t = 260$ первого типа ***Abrupt,Permanent***, моделирующую взвешенную единичную ступеньку, как и в исследовании [14]. Итоговое окно задания параметров оценивания изображено на рис. 10.28. Заметим, что в данном случае мы не задаем операции взятия разности $(1 - B)$. Вместо этого у нас задается авторегрессионный оператор $(1 - \phi B)$, ожидаемое значение параметра ϕ должно быть очень близким к единице. Таким образом, мы задаем модель

$$Z_t = \omega I_t^{(260)} + \frac{\theta_0 + a_t}{1 - \phi B}. \qquad (10.11)$$

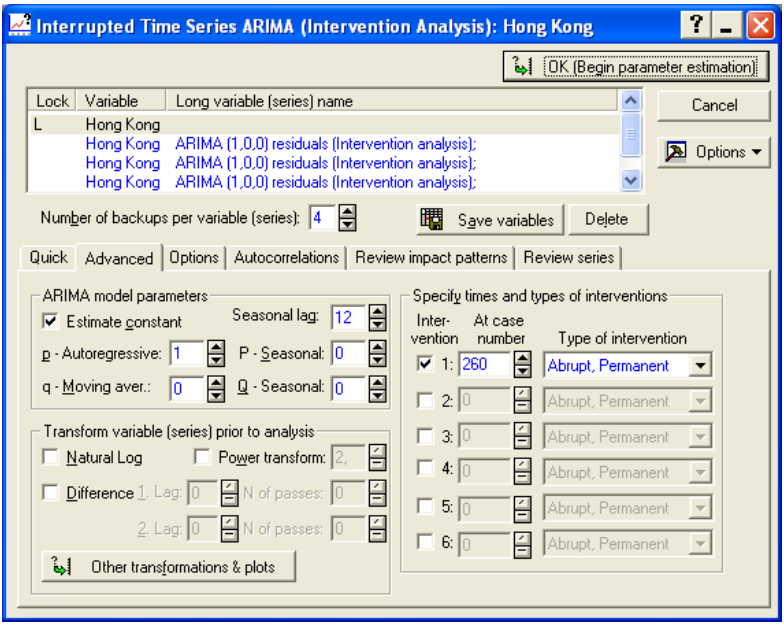


Рис. 10.28. Вид окна для задания параметров модели интервенций

Результаты оценивания представлены на рис. 10.29. Заметим, что в данном случае в закладке ***Options*** мы выбрали точный метод оценивания ***Exact***, так как обе модификации приближенного метода – простая и с обратным прогнозом – не приводили к желаемому решению, выдавая сообщение о плохой обусловленности матрицы. В данном случае это было предсказуемо, так как модель однородно нестационарна и параметр авторегрессии близок к единице, что повышает неустойчивость процедуры оценивания.

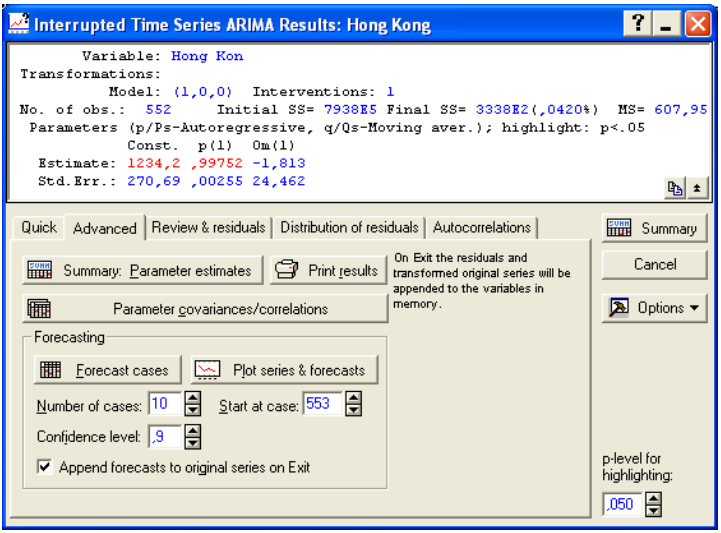


Рис. 10.29. Результаты оценивания модели интервенций

Как видно из рис. 10.29 и табл. 10.1, мы действительно получили значимое, близкое к единице значение параметра ϕ и значимый детерминированный тренд. Однако, как и в исследовании [14], эффект интервенции в рассмотренной нами модели (10.11) оказался незначимым, на что указывает *p*-значение 0,941731 для критерия Стьюдента параметра ω. Данному результату можно доверять, так как он согласуется с результатом, полученным в [14]. Кроме того, полученные оценки являются некоррелированными, как показано в табл. 10.2. И, наконец, графики АКФ и ЧАКФ остатков соответствуют белому шуму, что показано на рис. 10.30 и 10.31, а сам ряд остатков имеет распределение, близкое к нормальному.

Таблица 10.1

Результаты оценивания модели интервенций.

Параметр	Оценка	Стандартная ошибка	*t*-тест	*p*-значение для *t*-теста
θ_0	1234,225	270,6940	4,5595	0,000006
ϕ	0,998	0,0026	390,8149	0,000000
ω	–1,813	24,7950	–0,0731	0,941731

Таблица 10.2

Корреляционная матрица оценок модели интервенций

	θ_0	ϕ	ω
θ_0	1,000000	0,092359	–0,039905
ϕ	0,092359	1,000000	0,075299
ω	–0,039905	0,075299	1,000000

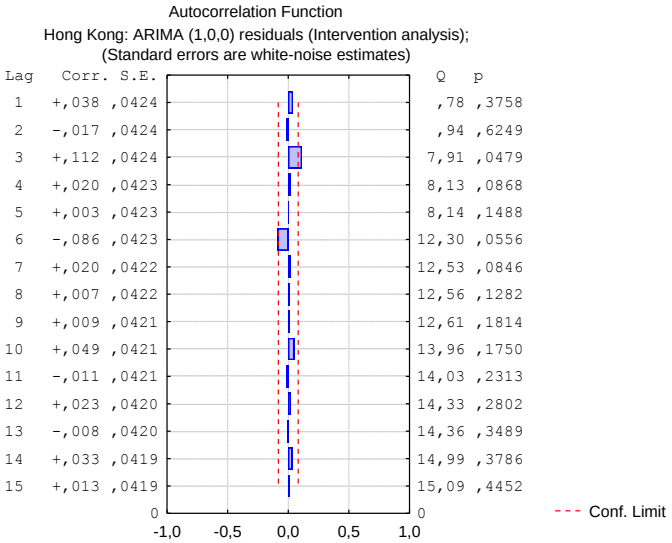


Рис. 10.30. АКФ остатков модели интервенций (10.11)

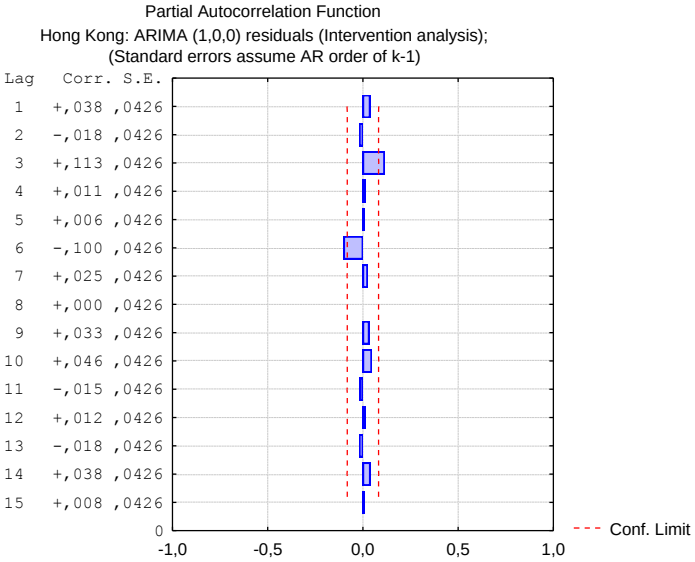


Рис. 10.31. АКФ остатков модели интервенций (10.11)

10.2. ВЫБРОСЫ ВРЕМЕННЫХ РЯДОВ

Наблюдения в выборках временных рядов иногда подвергаются влиянию неожиданных событий, таких как забастовки, развязывание войны, внезапные политические или экономические кризисы, неожиданные колебания температур или даже незамеченные ошибки печати или регистрации. Последствия таких внезапных событий создают ложные наблюдения, которые не согласуются с остальной частью временного ряда. Такие наблюдения, как уже упоминалось, называются *выбросами*. Когда моменты возникновения и причины прерываний известны, их влияние может быть учтено с помощью моделей интервенций, рассмотренных нами ранее. Однако на практике моменты возникновения выбросов и их причины часто неизвестны. Так как выбросы проявляются в анализе данных своим сильным влиянием на результаты, что делает результирующие выводы ненадежными или даже безосновательными, очень важным средством аналитика становятся процедуры, позволяющие определять и удалять такие выбросы. В данном подразделе и будут описаны такие процедуры.

10.2.1. АДДИТИВНЫЕ И ИННОВАЦИОННЫЕ ВЫБРОСЫ

Для заданного или полученного в результате преобразований (стабилизации дисперсии и/или взятия разности) стационарного процесса положим, что Z_t – наблюдаемый временной ряд, а X_t – ряд, свободный от выброса. Пусть временной ряд $\{X_t\}$ удовлетворяет следующей общей модели ARMA(p,q):

$$\phi(B)X_t = \theta(B)a_t, \tag{10.12}$$

где $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$, $\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$ – стационарный и обратимый операторы авторегрессии и скользящего среднего, не имеющие общих множителей, $\{a_t\}$ – i.i.d. процесс белого шума, подчиняющийся нормальному распределению $N(0, \sigma_a^2)$.

Рассмотрим два вида выбросов – аддитивные и инновационные. Аддитивный выброс определяется следующим образом:

$$Z_t = \begin{cases} X_t, & t \neq T, \\ X_t + \omega, & t = T \end{cases}$$

или

$$Z_t = X_t + \omega I_t^{(T)} = \frac{\theta(B)}{\phi(B)} a_t + \omega I_t^{(T)}, \tag{10.13}$$

где $I_t^{(T)} = \begin{cases} 1, & t = T, \\ 0, & t = T \end{cases}$ – индикаторная переменная, обозначающая присутствие или отсутствие выброса в момент времени T.

Инновационный выброс определяется следующим образом:

$$Z_t = X_t + \frac{\theta(B)}{\phi(B)} \omega I_t^{(T)} = \frac{\theta(B)}{\phi(B)} a_t + \omega I_t^{(T)} . \tag{10.14}$$

Следовательно, аддитивный выброс влияет только на значение наблюдения в момент времени T, а инновационный выброс – на все наблюдения Z_T, Z_{T+1}, …, следующие после момента T посредством памяти системы, описываемой полиномом $\theta(B)/\phi(B)$.

В общем случае временной ряд может содержать несколько, скажем, k, выбросов различного типа. Тогда мы имеем следующую об-

щую модель, связывающую реальные наблюдения Z_t с рядом X_t, очищенным от выбросов:

$$Z_t = \sum_{j=1}^{k} \omega_j \nu_j(B) I_t^{(T_j)} + X_t ,$$

где $X_t = \dfrac{\theta(B)}{\phi(B)} a_t$, $\nu_j(B) = 1$ – для аддитивного выброса, $\nu_j(B) = \dfrac{\theta(B)}{\phi(B)}$ для инновационного выброса в момент $t = T_j$.

10.2.2. ОЦЕНИВАНИЕ ЭФФЕКТА ВЫБРОСА ПРИ ИЗВЕСТНОМ ВРЕМЕНИ ВЫБРОСА

Для начала рассмотрим самый простой случай, когда момент времени T и все параметры модели (10.12) известны. Определим

$$\pi(B) = \frac{\phi(B)}{\theta(B)} = (1 - \pi_1 B - \pi_2 B^2 - \ldots) ,$$

$$e_t = \pi(B) Z_t .$$

Тогда для аддитивного выброса из (10.13) получаем

$$e_t = \omega \pi(B) I_t^{(T)} + a_t , \tag{10.15}$$

а для инновационного из (10.14) –

$$e_t = \omega I_t^{(T)} + a_t . \tag{10.16}$$

Расписывая аддитивную модель (10.15) для выборки, состоящей из n наблюдений, получаем

$$\begin{bmatrix} e_1 \\ \vdots \\ e_{T-1} \\ e_T \\ e_{T+1} \\ e_{T+2} \\ \vdots \\ e_n \end{bmatrix} = \omega \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ -\pi_1 \\ -\pi_2 \\ \vdots \\ -\pi_{n-T} \end{bmatrix} + \begin{bmatrix} a_1 \\ \vdots \\ a_{T-1} \\ a_T \\ a_{T+1} \\ a_{T+2} \\ \vdots \\ a_n \end{bmatrix} . \tag{10.17}$$

Пусть $\hat{\omega}_{AT}$ – оценка по методу наименьших квадратов параметра ω модели (10.17):

$$\hat{\omega}_{AT} = \frac{e_T - \sum_{j=1}^{n-T} \pi_j e_{T+j}}{\sum_{j=0}^{n-T} \pi_j^{\,2}} = \frac{\pi^*(F) e_T}{\tau^2},$$

где $\pi^*(F) = (1 - \pi_1 F - \pi_2 F^2 - \ldots - \pi_{n-T} F^{n-T})$, F – оператор сдвига вперед, $Fe_t = e_{t+1}$, $\tau^2 = \sum_{j=0}^{n-T} \pi_j^{\,2}$, $\pi_0 = 1$. Дисперсия оценки параметра определяется следующим образом:

$$\mathrm{Var}(\hat{\omega}_{AT}) = \mathrm{Var}\left(\frac{\pi^*(F) e_T}{\tau^2}\right) = \frac{1}{\tau^4} \mathrm{Var}[\pi^*(F) a_T] = \frac{\sigma_a^2}{\tau^2}.$$

Аналогично, пусть $\hat{\omega}_{IT}$ – оценка по методу наименьших квадратов параметра ω модели (10.16) для инновационного выброса:

$$\hat{\omega}_{IT} = e_T,$$

ее дисперсия:

$$\mathrm{Var}(\hat{\omega}_{IT}) = \mathrm{Var}(e_T) = \mathrm{Var}(\omega I_t^{(T)} + a_T) = \sigma_a^2.$$

Таким образом, наилучшая оценка эффекта влияния инновационного выброса в момент времени T – остаточная e_T, а наилучшая оценка эффекта влияния аддитивного выброса – линейная комбинация e_T, $e_{T+1}, \ldots, e_n$ с весами, зависящими от типа стационарного процесса, генерирующего временной ряд. Очевидно, что $\mathrm{Var}(\hat{\omega}_{AT}) \le \mathrm{Var}(\hat{\omega}_{IT}) = \sigma_a^2$ (так как $\tau^2 \ge 1$), а в некоторых случаях дисперсия $\mathrm{Var}(\hat{\omega}_{AT})$ может быть много меньше σ_a^2.

При анализе выбросов полезны следующие гипотезы:

$H_0 : Z_t$ не является ни аддитивным, ни инновационным выбросом;

$H_1 : Z_t$ является аддитивным выбросом;

$H_2 : Z_t$ является инновационным выбросом.

Статистики отношения правдоподобия $\lambda_{1,T}$ и $\lambda_{2,T}$ для аддитивного и инновационного выбросов соответственно могут использоваться для проверки следующих гипотез:

H_1 против H_0: $\lambda_{1,T} = \tau\hat{\omega}_{AT} / \sigma_a$;

H_2 против H_0: $\lambda_{2,T} = \hat{\omega}_{IT} / \sigma_a$.

При справедливости нулевой гипотезы H_0 статистики $\lambda_{1,T}$ и $\lambda_{2,T}$ распределены по стандартному нормальному закону $N(0,1)$.

10.2.3. ОПРЕДЕЛЕНИЕ ВЫБРОСОВ С ИСПОЛЬЗОВАНИЕМ ИТЕРАЦИОННОЙ ПРОЦЕДУРЫ

Если момент выброса T неизвестен, а параметры временного ряда известны, можно вычислить статистики $\lambda_{1,T}$ и $\lambda_{2,T}$ для каждого $t = 1,2,\ldots,n$ и затем принять решение на основе проверки вышеприведенных статистических гипотез. Однако на практике параметры временного ряда ϕ_j, θ_j, π_j и σ_a^2 обычно неизвестны и должны быть предварительно оценены по имеющейся выборке Z_t, содержащей выбросы. Мы знаем, что существование выбросов приводит к сильно смещенным оценкам параметров. В частности, σ_a^2 имеет тенденцию к завышенной оценке.

Рассмотрим итерационную процедуру определения аддитивных и/или инновационных выбросов, описанную в [10].

Шаг 1. Строим модель временного ряда по выборке $\{Z_t\}$, предполагая, что выбросов нет. Вычисляем остатки оцененной модели, т.е.

$$\hat{e}_t = \hat{\pi}(B)Z_t = \frac{\hat{\phi}(B)}{\hat{\theta}(B)} Z_t,$$

где $\hat{\phi}(B) = (1 - \hat{\phi}_1 B - \ldots - \hat{\phi}_p B^p)$ и $\hat{\theta}(B) = (1 - \hat{\theta}_1 B - \ldots - \hat{\theta}_q B^q)$.

Пусть $\hat{\sigma}_a^2 = \frac{1}{n}\sum_{t=1}^{n} \hat{e}_t^2$ – начальная оценка σ_a^2.

Шаг 2. Вычисляем $\hat{\lambda}_{1,t}$ и $\hat{\lambda}_{2,t}$ для $t = 1,2,\ldots,n$, используя оцененную модель. Определяем:

$$\hat{\lambda}_T = \max_t \max_i \{|\hat{\lambda}_{i,t}|\},$$

где T – момент времени, соответствующий максимуму.

Если $\hat{\lambda}_T = |\hat{\lambda}_{1,T}| > C$, где C – заданная положительная константа, обычно задаваемая в диапазоне $3 < C < 4$, то существует аддитивный выброс в момент времени T, а его эффект оценивается параметром $\hat{\omega}_{AT}$. Затем можно модифицировать данные с учетом формулы (10.13):

$$\tilde{Z}_t = Z_t - \hat{\omega}_{AT} I_t^{(T)} \tag{10.18}$$

и определить новые остатки с учетом формулы (10.15):

$$\tilde{e}_t = \hat{e}_t - \hat{\omega}_{AT} \hat{\pi}(B) I_t^{(T)}.$$

Аналогично, если $\hat{\lambda}_T = |\lambda_{2,T}| > C$, то существует инновационный выброс в момент времени T, а его эффект оценивается параметром $\hat{\omega}_{IT}$. Этот эффект можно убрать, модифицируя данные с учетом формулы (10.14):

$$\tilde{Z}_t = Z_t - \frac{\hat{\theta}(B)}{\hat{\phi}(B)} \hat{\omega}_{IT} I_t^{(T)}, \tag{10.19}$$

и определить новые остатки с учетом формулы (10.16):

$$\tilde{e}_t = \hat{e}_t - \hat{\omega}_{IT} I_t^{(T)}.$$

Затем вычисляется новая оценка $\tilde{\sigma}_a^2 = \frac{1}{n}\sum_{t=1}^{n} \tilde{e}_t^2$ для модифицированных остатков.

Шаг 3. Вычисляем новые $\hat{\lambda}_{1,t}$ и $\hat{\lambda}_{2,t}$ на основании модифицированных остатков и $\tilde{\sigma}_a^2$ и повторяем шаг 2 до тех пор, пока все выбросы не будут идентифицированы. Начальные оценки полинома $\pi(B)$ при этом остаются неизменными.

Шаг 4. К началу этого шага новые выбросы, получаемые на шаге 2, уже не выявляются. Пусть всего экспериментально идентифицировано k выбросов в моменты $T_1, T_2, \ldots, T_k$. Используем эти моменты време-

ни как известные и вычислим оценки параметров ω_1, $\omega_2, \ldots,$ ω_k и параметров временного ряда одновременно, используя модель

$$Z_t = \sum_{j=1}^{k} \omega_j \nu_j(B) I_j^{(T_j)} + \frac{\theta(B)}{\phi(B)} a_t ,$$

где $\nu_j(B) = 1$ – для аддитивного выброса, $\nu_j(B) = \theta(B)/\phi(B)$ – для инновационного выброса в момент времени T_j. Новые оценки параметров приводят к новым остаткам:

$$\hat{e}_t^1 = \hat{\pi}^{(1)}(B)\left[Z_t - \sum_{j=1}^{k} \hat{\omega}_j \nu_j(B) I_t^{(T_j)} \right],$$

и соответственно к новой оценке σ_a^2.

Шаги 2–4 повторяются до тех пор, пока не выявлены все выбросы и их совместные эффекты не оценены. В итоге получаем следующую модель с учетом выбросов:

$$Z_t = \sum_{j=1}^{k} \hat{\omega}_j \nu_j(B) I_t^{(T_j)} + \frac{\hat{\theta}(B)}{\hat{\phi}(B)} a_t , \qquad (10.20)$$

где $\hat{\omega}_j$, $\hat{\phi}(B) = 1 - \hat{\phi}_1 B - \ldots - \hat{\phi}_p B^p$ и $\hat{\theta}(B) = 1 - \hat{\theta}_1 B - \ldots - \hat{\theta}_q B^q$ – оценки, полученные на последней итерации.

10.3. ЗАМЕЧАНИЯ ПО ИСПОЛЬЗОВАНИЮ МЕТОДОЛОГИИ АНАЛИЗА ИНТЕРВЕНЦИЙ И ОПРЕДЕЛЕНИЯ ВЫБРОСОВ

1. После того как выбросы идентифицированы, можно очистить данные от выбросов в соответствии с формулой (10.18) или (10.19). Однако более эффективный подход, возможно, состоит в поиске причин идентифицированных выбросов и в дальнейшем усовершенствовании подогнанной модели. Это справедливо не только для оценивания параметров, но и для диагностики модели, и для прогнозирования. Например, некоторые выбросы могут оказаться важными переменными интервенций благодаря каким-то изменениям в политике, о которых аналитику неизвестно и которые были упущены на предыдущих стадиях анализа данных. Таким образом, вместо того, чтобы очищать

данные от выбросов, аналитик должен включить в модель информацию о соответствующих переменных и откликах интервенций, описанных в подразделе 10.1. Эта явная форма модели интервенции может оказаться более полезной при решении задач прогнозирования и управления, чем очищенные от выбросов данные.

2. Вышеизложенная процедура определения выбросов основывается на предположении, что для очищенного от выбросов ряда модель или известна, или может быть идентифицирована. Однако на практике такая модель обычно неизвестна и должна быть идентифицирована с использованием стандартного анализа выборочных статистик, таких как АКФ, ЧАКФ и некоторых других. Таким образом, процедура определения выбросов обычно работает хорошо только при условии умеренности влияния выбросов, когда их эффект не затемняет или даже не скрывает закономерности, содержащиеся в статистиках ряда, свободного от выбросов. В худшем случае загрязнение выбросами может привести к невозможности идентификации модели. Например, может оказаться, что характеристики АКФ и ЧАКФ исходного временного ряда выявляют нулевой паттерн, т.е. указывают на то, что исходный ряд является белым шумом. В этом случае скорее всего дело в том, что имеется очень сильный выброс в исходных данных, который нужно устранить до начала анализа. Такой выброс легко определяется на графике временного ряда. Вообще, нужно подчеркнуть, что визуальный анализ – это средство первой помощи при анализе выбросов.

11. ПРОГНОЗИРОВАНИЕ

Неопределенность является фактом жизни как отдельных людей, так и организаций. Прогнозирование очень важно для планирования и оперативного контроля в различных областях, таких как управление производством, система инвентаризации, контроль качества, финансовое планирование и инвестиционный анализ. Поэтому одна из наиболее важных целей анализа таких существенно динамических данных, как временные ряды, – прогноз будущих значений. Даже если главной целью моделирования является управление системой, в конечном счете любое управление основано на качественном прогнозировании

В данном разделе рассматриваются прогнозы с минимальной среднеквадратической ошибкой для стационарных и нестационарных моделей ARIMA, рассмотренных в предыдущих разделах. Эти же модели

могут быть использованы и для процедуры подправления (обновления) прогнозов, также рассмотренной в данном разделе. Такое подправление прогнозов бывает необходимо, если по истечении некоторого времени после первого прогноза становятся доступными новые наблюдения временного ряда. Большинство результатов по прогнозированию основывается на теории линейного предсказания, разработанной в работах А.Н. Колмогорова, Н. Винера, Р.Е. Калмана и других исследователей.

Рассмотрим общую модель в классе ARIMA (p,d,q):

$$\phi(B)(1-B)^d Z_t = \theta(B)a_t, \tag{11.1}$$

где $\phi(B) = (1-\phi_1 B - ... - \phi_p B^p)$ и $\theta(B) = (1-\theta_1 B - ... - \theta_q B^q)$, $\{a_t\}$ – i.i.d. процесс белого шума, подчиняющийся нормальному распределению $N(0,\sigma_a^2)$. Параметр θ_0 детерминированного тренда опущен для простоты, но без потери общности. Уравнение (11.1) наиболее часто используется в приложениях по прогнозированию.

Рассмотрим прогноз с минимальной среднеквадратической ошибкой для модели (11.1) отдельно для двух случаев: если $d = 0$ (модель ARMA) и $d \neq 0$ (модель ARIMA).

11.1. ПРОГНОЗИРОВАНИЕ ДЛЯ МОДЕЛИ ARMA

Итак, рассмотрим модель (11.1) для случая $d = 0$:

$$\phi(B)Z_t = \theta(B)a_t \tag{11.2}$$

Так как модель стационарна, ее можно представить в виде модели скользящего среднего бесконечного порядка:

$$Z_t = \psi(B)a_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + ..., \tag{11.3}$$

где $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j = \dfrac{\theta(B)}{\phi(B)}$, $\psi_0 = 1$.

Для $t = n + l$ получаем

$$Z_{n+l} = \sum_{j=0}^{\infty} \psi_j a_{n+l-j}. \tag{11.4}$$

Предположим, что в момент времени $t = n$ мы имеем наблюдения $Z_n, Z_{n-1}, Z_{n-2}...$ и хотим предсказать значение Z_{n+l} как линейную комбинацию наблюдений $Z_n, Z_{n-1}, Z_{n-2}...$. Так как значения Z_t для $t = n, n-1, n-2,...$ могут быть записаны в форме (11.3), будем искать прогноз с минимальной среднеквадратической ошибкой в следующем виде:

$$\hat{Z}_n(l) = \psi_l^* a_n + \psi_{l+1}^* a_{n-1} + \psi_{l+2}^* a_{n-2} + ...$$

Необходимо определить значения ψ_j^*, минимизирующие среднеквадратическую ошибку прогноза

$$E(Z_{n+l} - \hat{Z}_n(l))^2 = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2 + \sigma_a^2 \sum_{j=0}^{\infty} \left[\psi_{l+j} - \psi_{l+j}^* \right]^2 .$$

Очевидно, что эта ошибка минимальна, если $\psi_{l+j}^* = \psi_{l+j}$. Следовательно,

$$\hat{Z}_n(l) = \psi_l a_n + \psi_{l+1} a_{n-1} + \psi_{l+2} a_{n-2} + ...$$

С другой стороны, из (11.4) и того, что

$$E(a_{n+j} \mid Z_n, Z_{n-1,...}) = \begin{Bmatrix} 0, & j > 0 \\ a_{n+j}, & j \le 0 \end{Bmatrix},$$

имеем

$$E(Z_{n+l} \mid Z_n, Z_{n-1},...) = \psi_l a_n + \psi_{l+1} a_{n-1} + \psi_{l+2} a_{n-2} + ...$$

Получается, что прогноз с минимальной среднеквадратической ошибкой для значения Z_{n+l} задается условным математическим ожиданием, т.е.

$$\hat{Z}_n(l) = E(Z_{n+l} \mid Z_n, Z_{n-1,...}),$$

где $\hat{Z}_n(l)$ обычно понимается как l-шаговый прогноз вперед значения Z_{n+l} из начальной точки n.

Запишем выражение для ошибки прогноза и вычислим его математическое ожидание и дисперсию:

$$e_n(l) = Z_{n+l} - \hat{Z}_n(l) = \sum_{j=0}^{l-1} \psi_j a_{n+l-j}\,, \tag{11.5}$$

$$E(e_n(l) \mid Z_t, t \le n) = \sum_{j=0}^{l-1} \psi_j E(a_{n+l-j}) = 0\,, \tag{11.6}$$

$$\mathrm{Var}(\hat{Z}_n(l)) = \mathrm{Var}(e_n(l)) = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2\,. \tag{11.7}$$

Так как $E(e_n(l) \mid Z_t, t \le n) = 0$, то прогноз является несмещенным. Для гауссовского процесса $(1-\alpha)\cdot 100\,\%$ пределы прогноза определяются следующим образом:

$$\hat{Z}_n(l) \pm N_{\alpha/2} \left[1 + \sum_{j=1}^{l-1} \psi_j^2 \right]^{1/2} \sigma_a\,, \tag{11.8}$$

где $N_{\alpha/2}$ – квантиль стандартного нормального распределения: $P(N > N_{\alpha/2}) = \dfrac{\alpha}{2}$.

Таким образом, в соответствии с (11.5), ошибка прогноза $e_n(l)$ есть линейная комбинация будущих случайных возмущений, поступающих в систему после момента времени n. В частном случае ошибка одношагового прогноза:

$$e_n(1) = Z_{n+1} - \hat{Z}_n(1) = a_{n+1}\,, \tag{11.9}$$

т.е. ошибки одношагового прогноза являются независимыми. Отсюда следует, что $\hat{Z}_n(1)$ – действительно наилучший прогноз Z_{n+1}. В противном случае, если ошибки одношагового прогноза коррелированы, то можно вычислить прогноз $\hat{a}_{n+1}$ для значения a_{n+1} на основе доступных ошибок $a_n, a_{n-1}, a_{n-2}, \ldots$ и, следовательно, улучшить прогноз Z_{n+1}, просто используя $\hat{Z}_n(1) + \hat{a}_{n+1}$ в качестве прогноза. Однако для более продолжительных прогнозов ошибки прогноза коррелированы.

Например, ошибки

$$e_n(l) = Z_{n+l} - \hat{Z}_n(l) = a_{n+l} + \psi_1 a_{n+l-1} + ... + \psi_{l-1} a_{n+1}$$

и

$$e_{n-j}(l) = Z_{n+l-j} - \hat{Z}_{n-j}(l) =$$

$$= a_{n+l-j} + \psi_1 a_{n+l-j-1} + ... + \psi_{l-1} a_{n-j+1}$$

прогнозов для одинаковой продолжительности прогноза l, но разных моментов начала прогноза n и $n-j$ коррелированы для $j < l$.

Ошибки также коррелированы для различных продолжительностей l и одинаковых моментов начала прогноза n. Например:

$$\mathrm{Cov}(e_n(2), e_n(1)) = E[a_{n+2} + \psi_1 a_{n+1})(a_{n+1})] = \psi_1 \sigma_a^2 . \qquad (11.10)$$

11.2. ПРОГНОЗИРОВАНИЕ ДЛЯ МОДЕЛИ ARIMA

Рассмотрим теперь общую нестационарную модель ARIMA p, d, q, для которой $d \neq 0$, т.е.

$$\phi(B)(1-B)^d Z_t = \theta(B) a_t , \qquad (11.11)$$

где $\phi(B) = (1 - \phi_1 B - ... - \phi_p B^p)$ – стационарный оператор авторегрессии и $q(B) = (1 - q_1 B - ... - q_q B^q)$ – обратимый оператор скользящего среднего.

Запишем модель (11.11) в момент времени $(t+l)$ в авторегрессионном представлении. Это можно сделать в силу обратимости процесса. Таким образом, получаем модель в операторном виде

$$\pi(B) Z_{t+l} = a_{t+l} ,$$

где

$$\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j = \frac{\phi(B)(1-B)^d}{\theta(B)} ,$$

или в разностном виде

$$Z_{t+l} = \sum_{j=1}^{\infty} \pi_j Z_{t+l-j} + a_{t+l} . \qquad (11.12)$$

Применяя оператор $1+\psi_1 B+...+\psi_{l-1}B^{l-1}$ к модели (11.12), получаем

$$\sum_{j=0}^{\infty}\sum_{k=0}^{l-1}\pi_j\psi_k Z_{t+l-j-k}+\sum_{k=0}^{l-1}\psi_k a_{t+l-k}=0,\ \ \pi_0=-1,\ \ \psi_0=1.$$

Легко показать, что

$$\sum_{j=0}^{\infty}\sum_{k=0}^{l-1}\pi_j\psi_k Z_{t+l-j-k}=\pi_0 Z_{t+l}+\sum_{m=1}^{l-1}\sum_{i=0}^{m}\pi_{m-i}\psi_i Z_{t+l-m}+$$

$$+\sum_{j=1}^{\infty}\sum_{i=0}^{l-1}\pi_{l-1+j-i}\psi_i Z_{t-j+1}. \qquad (11.13)$$

Если выбрать вес ψ таким образом, чтобы для $m=1,\ 2,\ldots,\ l-1$ выполнялось

$$\sum_{i=0}^{m}\pi_{m-i}\psi_i=0, \qquad (11.14)$$

то второе слагаемое в сумме (11.13) справа от знака равенства обнуляется, и мы получаем

$$Z_{t+l}=\sum_{j=1}^{\infty}\pi_j^{(l)}Z_{t-j+1}+\sum_{i=0}^{l-1}\psi_i a_{t+l-i},$$

где $\pi_j^{(l)}=\sum_{i=0}^{l-1}\pi_{l-1+j-i}\psi_i$.

Таким образом, имея выборку Z_t для $t\le n$, получаем выражение для прогноза с упреждением l:

$$\hat{Z}_n(l)=E(Z_{n+l}\mid Z_t, t\le n)=\sum_{j=1}^{\infty}\pi_j^{(l)}Z_{n-j+1},$$

так как $E(a_{n+j}\mid Z_t, t\le n)=0$ для $j>0$. Ошибка прогноза определяется следующим образом:

$$e_n(l) = Z_{n+l} - \hat{Z}_n(l) = \sum_{j=0}^{l-1} \psi_j a_{n+l-j} \,. \qquad (11.15)$$

Вес ψ_j может быть вычислен рекурсивно по формуле

$$\psi_j = \sum_{i=0}^{j-1} \pi_{j-i}\psi_i \,, \; j = 1,\ldots, \; l-1 \,, \qquad (11.16)$$

являющейся следствием формулы (11.14). Заметим, что ошибка прогноза (11.15) для модели ARIMA совпадает с ошибкой прогноза (11.5) для модели ARMA, поэтому все последующие результаты для модели ARMA (11.6)–(11.10) остаются справедливыми и для модели ARIMA.

Для стационарного процесса пределы прогноза, вычисляемые по формуле (11.8), с увеличением l приближаются к константе, т.е. доверительный интервал с возрастанием l не увеличивается. То есть пределы прогноза для стационарного процесса представляют собой две горизонтальные параллельные линии, как показано на рис. 11.1.

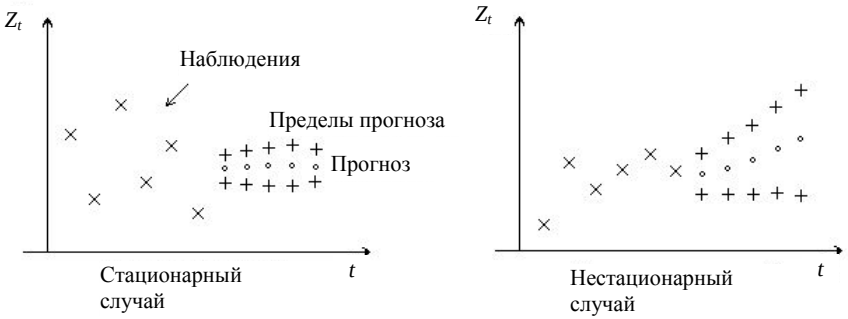


Рис. 11.1. Зависимость пределов прогноза от упреждения

В нестационарном прогнозе с увеличением l пределы прогноза увеличиваются. Таким образом, в нестационарном случае прогноз становится все менее определенным с увеличением упреждения l, т.е. точность прогноза падает.

11.3. ВЫЧИСЛЕНИЕ ПРОГНОЗОВ

Можно получить формулы для вычисления прогноза, используя исходную модель в виде разностного уравнения. Рассмотрим общую модель ARIMA(p,d,q):

$$\Psi(B)Z_t = \theta(B)a_t,$$

где

$$\Psi(B) = \phi(B)(1-B)^d = (1 - \Psi_1 B - ... - \Psi_{p+d}B^{p+d}),$$

$$\theta(B) = (1 - \theta_1 B - ... - \theta_q B^q).$$

Для $t = n + l$ получаем следующее разностное уравнение:

$$Z_{n+l} = \Psi_1 Z_{n+l-1} + \Psi_2 Z_{n+l-2} + ... + \Psi_{p+d}Z_{n+l-p-d} + a_{n+l} -$$

$$-\theta_1 a_{n+l-1} - \theta_2 a_{n+l-2} - ... - \theta_q a_{n+l-q}.$$

Для получения формулы вычисления прогноза используем тот факт, что прогнозы $\hat{Z}_n(l)$ с минимальной среднеквадратичной ошибкой величины Z_{n+l} из начала прогноза в точке n задаются условным математическим ожиданием:

$$\hat{Z}_n(l) = E(Z_{n+l} \mid Z_n, Z_{n-1}, \ldots).$$

Взяв условное математическое ожидание в момент времени n, получаем:

$$\hat{Z}_n(l) = E(Z_{n+l} \mid Z_n, Z_{n-1}, \ldots) =$$

$$= \Psi_1 \hat{Z}_n(l-1) + ... + \Psi_{p+d}\hat{Z}_n(l-p-d) +$$

$$+\hat{a}_n(l) - \theta_1 \hat{a}_n(l-1) - \ldots - \theta_q \hat{a}_n(l-q),$$

где $\hat{Z}_n(j) = E(Z_{n+j} \mid Z_n, Z_{n-1,}...)$, $\quad j \geq 1$,

$\hat{Z}_n(j) = Z_{n+j}$, $\quad j \leq 0$,

$\hat{a}_n(j) = 0$, $\quad j \geq 1$,

$\hat{a}_n(j) = Z_{n+j} - \hat{Z}_{n+j-1}(1) = a_{n+j}$, $\quad j \leq 0$.

Проиллюстрируем описанный выше способ вычисления прогнозов для стационарной модели ARIMA(1,0,1), эквивалентной модели ARMA(1,1), а затем для нестационарной модели ARIMA(0,1,1), эквивалентной IMA(1,1).

Рассмотрим сначала стационарную модель:

$$(1-\phi B)(Z_t-\mu)=(1-\theta B)a_t\,. \qquad (11.17)$$

1. Для $t=n+l$ можно записать исходную модель в виде разностного уравнения:

$$Z_{n+l}=\mu+\phi(Z_{n+l-1}-\mu)+a_{n+l}-\theta a_{n+l-1}\,.$$

Отсюда получаем формулу для вычисления прогноза $\hat{Z}_n(l)$ как условного математического ожидания:

$$\hat{Z}_n(1)=E(Z_{n+1}\mid Z_n,Z_{n-1},...)=\mu+\phi(Z_n-\mu)-\theta a_n\,,$$

$$\hat{Z}_n(l)=E(Z_{n+l}\mid Z_n,Z_{n-1},\ldots)=\mu+\phi[\hat{Z}_n(l-1)-\mu]=$$

$$=\mu+\phi^l(Z_n-\mu)-\phi^{l-1}\theta a_n,\ \text{для } l\ge 2\,. \qquad (11.18)$$

2. Вычисляем дисперсию прогноза:

$$\mathrm{Var}(\hat{Z}_n(l))=\sigma_a^2\sum_{j=0}^{l-1}\psi_j^2\,.$$

При $|\phi|<1$ мы можем вычислить вес ψ_j из представления скользящего среднего (11.3), в котором $\phi(B)=(1-\phi B)$ и $\theta(B)=(1-\theta B)$:

$$(1-\phi B)(1+\psi_1 B+\psi_2 B^2+...)=(1-\theta B)\,.$$

Приравнивая коэффициенты при B^j по обе стороны равенства, получаем

$$\psi_j=\phi^{j-1}(\phi-\theta)\,,\ j\ge 1\,.$$

Следовательно, дисперсия прогноза вычисляется таким образом:

$$\mathrm{Var}(\hat{Z}_n(l))=\sigma_a^2\left\{1+\sum_{j=1}^{l-1}\left[\phi^{j-1}(\phi-\theta)\right]^2\right\}. \qquad (11.19)$$

Заметим, что дисперсия прогноза при неограниченном увеличении l сходится к константе $\sigma_a^2\left[1+\frac{(\phi-\theta)^2}{1-\phi^2}\right]$.

При $\phi=1$ предыдущая модель превращается в однородно нестационарную модель IMA(1,1):

$$(1-B)Z_t=(1-\theta B)a_t,$$

причем $(1-B)\mu=0$.

Так как представление скользящего среднего (11.3) для нестационарной модели больше не существует, то мы не можем воспользоваться тем же методом вычисления веса ψ, необходимого для вычисления дисперсии прогноза, как в стационарных моделях. В этом случае сначала перепишем уравнение модели в форме авторегрессии в момент $t+l$, т.е.

$$\pi(B)Z_{t+l}=a_{t+l},$$

где

$$\pi(B)=1-\pi_1B-\pi_2B^2-\ldots=\frac{1-B}{1-\theta B},$$

или

$$1-B=1-(\pi_1+\theta)B-(\pi_2-\pi_1\theta)B^2-(\pi_3-\pi_2\theta)B^3-\ldots$$

Приравнивая коэффициенты при B^j, получаем

$$\pi_j=(1-\theta)\theta^{j-1},\ j\geq 1.$$

Теперь, применяя формулу (11.16), получаем

$$\psi_1=\pi_1=(1-\theta),$$

$$\psi_2=\pi_2+\pi_1\psi_1=(1-\theta)\theta+(1-\theta)^2=(1-\theta),$$

$$\psi_3=\pi_3+\pi_2\psi_1+\pi_1\psi_2=(1-\theta)\theta^2+(1-\theta)\theta(1-\theta)+$$

$$+(1-\theta)^2=(1-\theta),$$

$$\vdots$$

В итоге получаем, что $\psi_j = (1-\theta)$, $1 \le j \le l-1$. Подставляя данный вес в выражение (11.7), получаем, что дисперсия прогноза $\hat{Z}_n(l)$ равна

$$\mathrm{Var}(\hat{Z}_n(l)) = \sigma_a^2 \quad 1 + (l-1)(1-\theta)^2 \quad . \tag{11.20}$$

Заметим, что в данном случае дисперсия прогноза стремится к $+\infty$ при $l \to \infty$. Как и следовало ожидать, дисперсия (11.20) получается как предельный случай (11.19) при $\phi \to 1$.

Таким образом, если ϕ близко к единице, выбор между стационарной моделью ARIMA(1,1) и нестационарной моделью IMA(1,1) имеет очень сильные последствия для прогнозирования. Для $|\phi| < 1$ прогноз $\hat{Z}_n(l)$ приближается к среднему μ процесса при $l \to \infty$. Если $\phi \to 1$, то из выражения прогноза (11.18) при $\phi = 1$ получаем справедливость для любого l выражения:

$$\hat{Z}_n(l) = \hat{Z}_n(l-1).$$

В этом случае говорят, что прогноз «блуждает», без тенденции группирования вокруг некоторого фиксированного значения.

11.4. ARIMA-ПРОГНОЗ КАК ВЗВЕШЕННОЕ СРЕДНЕЕ НАБЛЮДЕНИЙ ВРЕМЕННОГО РЯДА

В данном подразделе будет описана форма прогноза, основанная на авторегрессионном представлении процесса. Такое представление всегда можно получить и для стационарных, и для нестационарных процессов при условии их обратимости. Так как условие обратимости всегда предполагается выполненным, то можно представить Z_t в AR-форме как бесконечную сумму предыдущих наблюдений плюс случайное возмущение, т.е.

$$Z_{n+l} = \sum_{i=1}^{\infty} \pi_j Z_{n+l-j} + a_{n+l} \tag{11.21}$$

или, эквивалентно,

$$\pi(B) Z_{n+l} = a_{n+l},$$

где $\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j = \frac{\phi(B)(1-B)^d}{\theta(B)}$.

Применив к (11.21) оператор условного математического ожидания, получаем

$$\hat{Z}_n(l) = \sum_{j=1}^{\infty} \pi_j \hat{Z}_n(l-j), \quad l \geq 1. \tag{11.22}$$

Подставляя в (11.22) вместо прогнозов $\hat{Z}_n(l-j)$, $1 \leq j \leq l-1$, их выражения (11.22), получаем, что прогноз $\hat{Z}_n(l)$ можно выражать как взвешенную сумму текущих и прошлых наблюдений исходной выборки Z_t, $t \leq n$. Например,

$$\hat{Z}_n(1) = \pi_1 Z_n + \pi_2 Z_{n-1} + \pi_3 Z_{n-2} + \ldots = \sum_{j=1}^{\infty} \pi_j Z_{n+1-j},$$

$$\hat{Z}_n(2) = \pi_1 \hat{Z}_n(1) + \pi_2 Z_n + \pi_3 Z_{n-1} + \ldots =$$

$$= \pi_1 \sum_{j=1}^{\infty} \pi_j Z_{n+1-j} + \sum_{j=1}^{\infty} \pi_{j+1} Z_{n+1-j} = \sum_{j=1}^{\infty} \pi_j^{(2)} Z_{n+1-j},$$

где $\pi_j^{(2)} = \pi_1 \pi_j + \pi_{j+1}, \quad j \geq 1$.

Для общего случая можно показать, что

$$\hat{Z}_n(l) = \sum_{j=1}^{\infty} \pi_j^{(l)} Z_{n+1-j}, \tag{11.23}$$

где $\pi_j^{(l)} = \pi_{j+l-1} + \sum_{i=1}^{l-1} \pi_i \pi_j^{(l-i)}$, $l > 1$ и $\pi_j^{(1)} = \pi_j$.

Выражение (11.23) позволяет утверждать, что многие методы сглаживания, например, метод скользящего среднего и метод экспоненциального сглаживания, являются специальными случаями ARIMA-прогнозирования. Модели ARIMA обеспечивают очень естественный и оптимальный способ получения требуемых весов для прогнозирования. Определив модель ARIMA, пользователь сам не должен

определять число необходимых коэффициентов ψ_j и их вид, что требуется в методах скользящего среднего и экспоненциального сглаживания. Необходимо также заметить, что ARIMA-прогнозы являются прогнозами с минимальной среднеквадратичной ошибкой. Это свойство оптимальности в общем случае не выполняется для методов скользящего среднего и экспоненциального сглаживания.

Для обратимого процесса вес π в (11.22) или (11.23) образует сходящийся ряд. Следовательно, для заданной степени точности прогноз $\hat{Z}_n(l)$ зависит только от конечного числа последних наблюдений. В совокупности веса π обеспечивают очень полезную информацию для многих важных управленческих решений.

В качестве примера рассмотрим прогноз как взвешенное среднее наблюдений временного ряда для модели ARMA(1,1). Для нахождения весов π_j авторегрессионного представления этой модели запишем следующее уравнение:

$$(1-\phi B) = (1-\pi_1 B - \pi_2 B^2 - \ldots)(1-\theta B) =$$

$$= 1-(\pi_1+\theta)B - (\pi_2 - \pi_1\theta)B^2 - (\pi_3 - \pi_2\theta)B^3 - \ldots,$$

$$(1-\phi B) = 1-(\pi_1+\theta)B - (\pi_2 - \pi_1\theta)B^2 - (\pi_3 - \pi_2\theta)B^3 - \ldots$$

Приравнивая коэффициенты при B^j с обеих сторон равенства, имеем

$$\pi_j = (\phi-\theta)\theta^{j-1}, \ j \ge 1.$$

Положив $\mu = 0$, из (11.22) имеем

$$\hat{Z}_n(l) = \sum_{j=1}^{\infty}(\phi-\theta)\theta^{j-1}\hat{Z}_n(l-j).$$

Получим выражение прогноза как взвешенной суммы текущих и прошлых наблюдений исходной выборки Z_t, $t \le n$.

При $l = 1$ имеем $\hat{Z}_n(1) = (\phi-\theta)\sum_{j=1}^{\infty}\theta^{j-1}Z_{n+1-j}$.

При $l = 2$ получаем

$$\hat{Z}_n(2) = \sum_{j=1}^{\infty} \pi_j^{(2)} Z_{n+1-j} = \sum_{j=1}^{\infty} \left[\pi_{j+1} + \pi_1 \pi_j \right] Z_{n+1-j} =$$

$$= \sum_{j=1}^{\infty} \left[(\phi - \theta)\theta^j + (\phi - \theta)^2 \theta^{j-1} \right] Z_{n+1-j} = \phi(\phi - \theta) \sum_{j=1}^{\infty} \theta^{j-1} Z_{n+1-j} \cdot$$

Снова $\hat{Z}_n(2)$ – взвешенное среднее предыдущих наблюдений ($Z_n, Z_{n-1}, \ldots$) с весами $\pi_j^{(2)} = \phi(\phi - \theta)\theta^{j-1}$ для $j \geq 1$. Сравнивая $\hat{Z}_n(2)$ и $\hat{Z}_n(1)$, видим:

$$\hat{Z}_n(2) = \phi \hat{Z}_n(1),$$

что согласуется с результатом (11.18) при $l = 2$ и $\mu = 0$ для этой модели, полученным методом вычисления условного математического ожидания.

Чтобы установить смысл полученных весов π_j, рассмотрим, например, одношаговый прогноз $\hat{Z}_n(1)$ более подробно. Для $|\theta| < 1$ веса $\pi_j = (\phi - \theta)\theta^{j-1}$ сходятся к 0 при $j \to \infty$. Следовательно, более свежие наблюдения в выборке имеют большее влияние на прогноз. Для $|\theta| > 1$, хотя модель все еще стационарна, но она необратима, авторегрегрессионного представления не существует. Неприятность заключается в том, что при этом вес π_j быстро стремится к $-\infty$ или к $+\infty$ при увеличении j. Следовательно, в этом случае более удаленные по времени наблюдения имеют большее влияние на прогноз, чем недавние. При $|\theta| = 1$ получаем

$$\pi_j = \begin{cases} (\phi - 1), & \theta = 1, \\ (-1)^{j-1}(1 + \phi), & \theta = -1, \end{cases}$$

т.е. π_j имеют одинаковые абсолютные значения для любого j. Следовательно, все прошлые и настоящие наблюдения в этом случае оди-

наково влияют на прогноз. Таким образом, прогноз, имеющий смысл, может быть получен только для обратимого процесса, удовлетворяющего условию $|\theta| < 1$.

11.5. ОБНОВЛЕНИЕ ПРОГНОЗОВ

Пусть известен прогноз значения Z_{n+l}, сделанный по выборке Z_1, $Z_2, \ldots, Z_n$, содержащей n наблюдений. Предположим, что в следующий момент времени получена новая информация о процессе в форме наблюдения Z_{n+1}. Понятно, что эта информация может уточнить значение прежнего прогноза значения Z_{n+l}. Возникает вопрос, как использовать эту новую информацию для уточнения прогноза. Существенным моментом является то, что мы хотим не вычислять прогноз заново, а использовать уже рассчитанный ранее прогноз. Такая процедура называется обновлением прогноза.

Начнем рассмотрение данного вопроса с выражения ошибки прогноза с минимальной среднеквадратичной ошибкой (11.5) для общей модели ARIMA:

$$e_n(l) = Z_{n+l} - \hat{Z}_n(l) = \sum_{j=0}^{l-1} \psi_j a_{n+l-j} .$$

В частности, ошибка одношагового прогноза:

$$e_n(1) = Z_{n+1} - \hat{Z}_n(1) = a_{n+1} .$$

Очевидно, этот результат можно записать для момента времени $t = n$:

$$Z_n - \hat{Z}_{n-1}(1) = a_n .$$

Из выражения для ошибки прогноза $e_n(l)$, очевидно, выполняется

$$e_{n-1}(l+1) = e_n(l) + \psi_l a_n, \qquad (11.24)$$

где

$$e_{n-1}(l+1) = Z_{n+l} - \hat{Z}_{n-1}(l+1) = \sum_{j=0}^{l} \psi_j a_{n+l-j}$$

и

$$e_n(l) = Z_{n+l} - \hat{Z}_n(l) = \sum_{j=0}^{l-1} \psi_j a_{n+l-j} \ .$$

Заметим, что прогнозы $\hat{Z}_n(l)$ и $\hat{Z}_{n-1}(l+1)$ предсказывают одно и то же значение Z_{n+l}, но из разных точек – начала прогноза. Теперь, подставляя в (11.24) ошибки и выражение $a_n = Z_n - \hat{Z}_{n-1}(1)$, получаем следующее уравнение для обновления прогноза:

$$\hat{Z}_n(l) = \hat{Z}_{n-1}(l+1) + \psi_l \ \ Z_n - \hat{Z}_{n-1}(1) \ ,$$

или, что эквивалентно,

$$\hat{Z}_{n+1}(l) = \hat{Z}_n(l+1) + \psi_l \ \ Z_{n+1} - \hat{Z}_n(1) \ . \qquad (11.25)$$

Таким образом, подправленный прогноз $\hat{Z}_{n+1}(l)$ получается прибавлением к предыдущему прогнозу $\hat{Z}_n(l+1)$ ошибки одношагового прогноза $a_{n+1} = Z_{n+1} - \hat{Z}_n(1)$, умноженной на коэффициент ψ_l.

Формула (11.25) для обновления прогноза определенно согласуется со здравым смыслом. Например, если значение Z_{n+1} становится доступным и превышает предыдущий прогноз (т.е. ошибка $a_{n+1} = Z_{n+1} - \hat{Z}_n(1)$ положительна), мы, естественно, модифицируем прогноз $\hat{Z}_n(l+1)$, сделанный ранее, в сторону увеличения, прибавляя к старому прогнозу константу, умноженную на величину ошибки.

11.6. ЭВЕНТУАЛЬНЫЕ ФУНКЦИИ ПРОГНОЗА

11.6.1. ОБЩАЯ ФОРМУЛА ВЫЧИСЛЕНИЯ ЭВЕНТУАЛЬНОЙ ФУНКЦИИ ПРОГНОЗА

Рассмотрим снова общую модель ARIMA:

$$\Psi(B)Z_t = \theta(B)a_t \, ,$$

где $\Psi(B) = \phi(B)(1-B)^d$.

Вспомним следующее выражение для вычисления прогнозов, полученное применением оператора математического ожидания:

$$\hat{Z}_n(l) = \Psi_1\hat{Z}_n(l-1) + \Psi_2\hat{Z}_n(l-2) + \ldots + \Psi_{p+d}\hat{Z}_n(l-p-d) +$$

$$+\hat{a}_n(l) - \theta_1\hat{a}_n(l-1) - \ldots - \theta_q\hat{a}_n(l-q)\,. \qquad (11.26)$$

При $l > q$ соотношение (11.26) принимает вид

$$\hat{Z}_n(l) = \Psi_1\hat{Z}_n(l-1) + \Psi_2\hat{Z}_n(l-2) + \ldots + \Psi_{p+d}\hat{Z}_n(l-p-d)$$

или

$$\Psi(B)\hat{Z}_n(l) = 0\,, \qquad (11.27)$$

т.е. прогноз $\hat{Z}_n(l)$ в случае $l > q$ удовлетворяет однородному разностному уравнению (11.27) порядка $p + d$.

Пусть $\Psi(B) = \prod_{i=1}^{N}(1 - R_iB)^{m_i}$, где $\sum_{i=1}^{N} m_i = (p+d)$.

Тогда общее решение разностного уравнения (11.27):

$$\hat{Z}_n(l) = \sum_{i=1}^{N} R_i^{\,l}\left(\sum_{j=0}^{m_i-1} C_{ij}^{(n)} l^{\,j}\right), \qquad (11.28)$$

справедливое для $l > (q-p-d)$, где $C_{ij}^{(n)}$ – константы, зависящие от начального времени прогноза n. Для заданного n они являются фиксированными константами для любого упреждения l. Константы изменяются, если изменяется начало прогноза.

Решение для $\hat{Z}_n(l)$ называется *эвентуальной функцией* прогноза, так как оно справедливо только для $l > (q-p-d)$, т.е. определяет тенденцию развития в будущем. При $q < (p+d)$ эвентуальная функция справедлива для любого упреждения $l > 0$. В общем, эта функция есть единственная кривая, проходящая через $(p+d)$ значений

$$\hat{Z}_n(q), \hat{Z}_n(q-1), \ldots, \hat{Z}_n(q-p-d+1)\,,$$

где $\hat{Z}_n(-j) = Z_{n-j}$ для $j \geq 0$. Для модели ARIMA(p, d, q), у которых $q = 0$, т.е. модели ARIMA $(p, d, 0)$, функция проходит через точки исходной выборки

$$Z_n, Z_{n-1}, \ldots, Z_{n-p-d+1}.$$

11.6.2. ПРИМЕРЫ ВЫЧИСЛЕНИЯ ЭВЕНТУАЛЬНОЙ ФУНКЦИИ ПРОГНОЗА

Данный подраздел посвящен нахождению эвентуальной функции прогноза для трех моделей.

Рассмотрим сначала модель ARIMA(1,0,1):

$$(1 - \phi B)(Z_t - \mu) = (1 - \theta B)a_t.$$

Прогноз $\hat{Z}_n(l)$ для этой модели удовлетворяет уравнению

$$(1 - \phi B)(\hat{Z}_t(l) - \mu) = 0 \text{ для } l > 1,$$

решением которого является эвентуальная функция прогноза

$$\hat{Z}_n(l) - \mu = C_1^{(n)}\phi^l,$$

или $\hat{Z}_n(l) = \mu + C_1^{(n)}\phi^l$ для $l > (q - p - d) = 0$ и константы $C_1^{(n)}$. Эвентуальная функция проходит через точку одношагового прогноза $\hat{Z}_n(1)$, откуда может быть найдена константа $C_1^{(n)} = \dfrac{\hat{Z}_n(l) - \mu}{\phi^l}$. При $l \to \infty$ прогноз $\hat{Z}_n(l)$, как и ожидалось, приближается к математическому ожиданию μ стационарного процесса.

Рассмотрим теперь модель ARIMA(1,1,1):

$$(1 - \phi B)(1 - B)Z_t = (1 - \theta B)a_t.$$

Сначала составляем однородное разностное уравнение прогноза

$$(1 - \phi B)(1 - B)\hat{Z}_t(l) = 0 \text{ для } l > 1.$$

Решением этого уравнения является эвентуальная функция прогноза

$$\hat{Z}_n(l) = C_1^{(n)} + C_2^{(n)}\phi^l \text{ для } l > (q - p - d) = -1.$$

Функция проходит через точку одношагового прогноза $\hat{Z}_n(1)$ и последнее наблюдение Z_n, следовательно, константы определяются из системы уравнений

$$Z_n = C_1^{(n)} + C_2^{(n)},$$

$$\hat{Z}_n(1) = C_1^{(n)} + C_2^{(n)}\phi.$$

И в заключение рассмотрим модель ARIMA(0,2,1):

$$(1-B)^2 Z_t = (1-\theta B)a_t.$$

Однородное уравнение прогноза для этой модели имеет вид

$$(1-B)^2 \hat{Z}_n(l) = 0 \text{ для } l > 1.$$

Решением является эвентуальная функция прогноза

$$\hat{Z}_n(l) = C_1^{(n)} + C_2^{(n)} l \text{ для } l > (q-p-d) = -1.$$

В данном случае эвентуальная функция прогноза – прямая линия, проходящая через точки $\hat{Z}_n(1)$ и Z_n. Константы определяются следующим образом:

$$C_1^{(n)} = Z_n,$$

$$C_2^{(n)} = \hat{Z}_n(1) - Z_n.$$

УПРАЖНЕНИЯ

ПРАКТИЧЕСКАЯ РАБОТА № 1

ИССЛЕДОВАНИЕ ARMA И ARIMA МОДЕЛЕЙ ВРЕМЕННЫХ РЯДОВ

ЗАДАНИЕ

ЧАСТЬ I. ИССЛЕДОВАНИЕ СТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ

Для каждой модели своего варианта:

1) смоделировать значения стационарного временного ряда при принадлежности параметров модели различным интервалам значений (например, для модели AR(1) нужно рассмотреть случаи, когда $\phi_1 < 0, \quad \phi_1 > 0$, для модели ARMA значения параметров подбираются эмпирическим путем);

для каждого сочетания значений параметров:

2) построить график значений временного ряда;

3) вычислить (если для заданной модели существуют такие теоретические результаты) значения теоретических автокорреляционной функции (АКФ) и частной автокорреляционной функции (ЧАКФ); для ЧАКФ можно вычислить только первые несколько значений ряда;

4) построить графики выборочных АКФ и ЧАКФ (с соответствующими доверительными интервалами для тех моделей и функций, для которых это целесообразно);

5) сравнить полученные выборочные АКФ и ЧАКФ с теоретическими, сделать выводы.

ЧАСТЬ II. ИССЛЕДОВАНИЕ ОДНОРОДНО-НЕСТАЦИОНАРНЫХ ВР

Для каждой модели своего варианта:

1) смоделировать значения однородно-нестационарного временного ряда в соответствии с вариантом задания;

2) построить график значений временного ряда, убедиться в его нестационарности;

3) построить графики выборочных АКФ и ЧАКФ, сравнить с теоретическими;

4) привести ряд к стационарному виду путем взятия разности;

5) для преобразованного ряда построить графики выборочных АКФ и ЧАКФ (с соответствующими доверительными интервалами для тех моделей и функций, для которых это целесообразно) и по ним определить вид модели (AR, MA, ARMA) преобразованного ряда;

6) сравнить выборочные АКФ и ЧАКФ для стационарного и нестационарного рядов; сделать выводы.

ВАРИАНТЫ ИНДИВИДУАЛЬНЫХ ЗАДАНИЙ

1) для исследования стационарных ВР: AR(1), MA(3), ARMA(2,2); для исследования нестационарных ВР: модель вида $(1-B)(1-\phi_1 B)z_t = a_t$ с подходящим значением ϕ_1;

2) для исследования стационарных ВР: AR(3), MA(2), ARMA(1,2); для исследования нестационарных ВР: модель вида $(1-B)(1-\phi_1 B - \phi_2 B^2)z_t = a_t$ с подходящими значениями ϕ_1 и ϕ_2;

3) для исследования стационарных ВР: AR(3), MA(1), ARMA(2,1); для исследования нестационарных ВР: модель вида $(1-B)z_t = (1-\theta_1 B)a_t$ с подходящим значением θ_1;

4) для исследования стационарных ВР: AR(3), MA(1), ARMA(1,1); для исследования нестационарных ВР: модель вида $(1-B)z_t = (1-\theta_1 B - \theta_2 B^2)a_t$ с подходящими значениями θ_1 и θ_2;

5) для исследования стационарных ВР: AR(1), MA(3), ARMA(2,1); для исследования нестационарных ВР: модель вида $(1-B)(1-\phi_1 B)z_t = (1-\theta_1 B)a_t$ с подходящими значениями ϕ_1 и θ_1;

6) для исследования стационарных ВР: AR(3), MA(1), ARMA(1,2); для исследования нестационарных ВР: модель вида $(1-B)(1-\phi_1 B - \phi_2 B^2)z_t = (1-\theta_1 B)a_t$ с подходящими значениями ϕ_1, ϕ_2 и θ_1;

7) для исследования стационарных ВР: AR(1), MA(2), ARMA(2,1); для исследования нестационарных ВР: модель вида $(1-B)(1-\phi_1 B)z_t = (1-\theta_1 B - \theta_2 B^2)a_t$ с подходящими значениями ϕ_1, θ_1 и θ_2;

8) для исследования стационарных ВР: AR(1), MA(3), ARMA(1,2); для исследования нестационарных ВР: модель вида $(1-B)(1-\phi_1 B)z_t = a_t$ с подходящим значением ϕ_1;

9) для исследования стационарных ВР: AR(2), MA(3), ARMA(1,2); для исследования нестационарных ВР: модель вида $(1-B)(1-\phi_1 B - \phi_2 B^2)z_t = a_t$ с подходящими значениями ϕ_1 и ϕ_2;

10) для исследования стационарных ВР: AR(3), MA(1), ARMA(2,2); для исследования нестационарных ВР: модель вида $(1-B)z_t = (1-\theta_1 B)a_t$ с подходящим значением θ_1;

11) для исследования стационарных ВР: AR(3), MA(1), ARMA(1,1); для исследования нестационарных ВР: модель вида $(1-B)z_t = (1 - \theta_1 B - \theta_2 B^2)a_t$ с подходящими значениями θ_1 и θ_2;

12) для исследования стационарных ВР: AR(2), MA(1), ARMA(1,2); для исследования нестационарных ВР: модель вида $(1-B)(1-\phi_1 B)z_t = (1-\theta_1 B)a_t$ с подходящими значениями ϕ_1 и θ_1.

ПРАКТИЧЕСКАЯ РАБОТА № 2

ПОСТРОЕНИЕ ARMA И ARIMA МОДЕЛЕЙ ВРЕМЕННЫХ РЯДОВ

ЗАДАНИЕ

Для временного ряда своего варианта провести идентификацию, включающую следующие действия:

1) по виду графиков временных рядов, АКФ, ЧАКФ выявить имеющиеся особенности ряда (нестационарность, тренд и т.п.);

2) устранить выявленные особенности применением подходящих преобразований;

3) для преобразованного ряда по виду графиков временных рядов, АКФ, ЧАКФ определить тип модели, оценить параметры и их значимость;

4) проверить адекватность полученной модели (посредством анализа остатков).

ВАРИАНТЫ

Вариант 1. Ежегодные данные о количестве солнечных пятен в 1700–1983 годах

1700 5,00
5,00
1701 11,00
1701 11,00
11,00
1702 16,00
1702 16,00
16,00
1703 23,00
1703 23,00
23,00
1704 36,00
1704 36,00
36,00

1705 58,00
1705 58,00
58,00
1706 29,00
1706 29,00
29,00
1707 20,00
1707 20,00
20,00
1708 10,00
1708 10,00
10,00
1709 8,00
1709 8,00
8,00
1710 30,00
1710 30,00
30,00
1711 0,00
1711 0,00
0,00
1712 0,00
1712 0,00
0,00
1713 2,00
1713 2,00
2,00
1714 11,00
1714 11,00
11,00
1715 27,00
1715 27,00
27,00
1716 4

7,00
1716 47,
00
47,00
1717 6
3,00
1717 63,
00
63,00
1718 6
0,00
1718 60,
00
60,00
1719 3
9,00
1719 39,
00
39,00
1720 2
8,00
1720 28,
00
28,00
1721 2
6,00
1721 26,
00
26,00
1722 2
2,00
1722 22,
00
22,00
1723 1
1,00
1723 11,
00
11,00
1724 2
1,00
1724 21,
00
21,00
1725 4
0,00
1725 40,
00
40,00
1726 7
8,00
1726 78,
00
78,00
1727 1
22,00

1727 12
2,00
122,00
1728 1
03,00
1728 10
3,00
103,00
1729 7
3,00
1729 73,
00
73,00
1730 4
7,00
1730 47,
00
47,00
1731 3
5,00
1731 35,
00
35,00
1732 1
1,00
1732 11,
00
11,00
1733 5
,00
1733 5,0
0
5,00
1734 1
6,00
1734 16,
00
16,00
1735 3
4,00
1735 34,
00
34,00
1736 7
0,00
1736 70,
00
70,00
1737 8
1,00
1737 81,
00
81,00
1738 1
11,00
1738 11
1,00

111,00
1739 101,00
1739 101,00
101,00
1740 73,00
1740 73,00
73,00
1741 40,00
1741 40,00
40,00
1742 20,00
1742 20,00
20,00
1743 16,00
1743 16,00
16,00
1744 5,00
1744 5,00
5,00
1745 11,00
1745 11,00
11,00
1746 22,00
1746 22,00
22,00
1747 40,00
1747 40,00
40,00
1748 60,00
1748 60,00
60,00
1749 80,90
1749 80,90
80,90
1750 83,40

1750 83,40
1750 83,40
83,40
1751 47,70
1751 47,70
47,70
1752 47,80
1752 47,80
47,80
1753 30,70
1753 30,70
30,70
1754 12,20
1754 12,20
12,20
1755 9,60
1755 9,60
9,60
1756 10,20
1756 10,20

10,20
1757 3
2,40
1757 32,
40
32,40
1758 4
7,60
1758 47,
60
47,60
1759 5
4,00
1759 54,
00
54,00
1760 6
2,90
1760 62,
90
62,90
1761 8
5,90
1761 85,
90
85,90
1762 6
1,20
1762 61,
20
61,20
1763 4
5,10
1763 45,
10
45,10
1764 3
6,40
1764 36,
40
36,40
1765 2
0,90
1765 20,
90
20,90
1766 1
1,40
1766 11,
40
11,40
1767 3
7,80
1767 37,
80
37,80

1768 6
9,80
1768 69,
80
69,80
1769 1
06,10
1769 10
6,10
106,10
1770 1
00,80
1770 10
0,80
100,80
1771 8
1,60
1771 81,
60
81,60
1772 6
6,50
1772 66,
50
66,50
1773 3
4,80
1773 34,
80
34,80
1774 3
0,60
1774 30,
60
30,60
1775 7
,00
1775 7,0
0
7,00
1776 1
9,80
1776 19,
80
19,80
1777 9
2,50
1777 92,
50
92,50
1778 1
54,40
1778 15
4,40
154,40
1779 1

25,90
1779 12
5,90
125,90
1780 8
4,80
1780 84,
80
84,80
1781 6
8,10
1781 68,
10
68,10
1782 3
8,50
1782 38,
50
38,50
1783 2
2,80
1783 22,
80
22,80
1784 1
0,20
1784 10,
20
10,20
1785 2
4,10
1785 24,
10
24,10
1786 8
2,90
1786 82,
90
82,90
1787 1
32,00
1787 13
2,00
132,00
1788 1
30,90
1788 13
0,90
130,90
1789 1
18,10
1789 11
8,10
118,10
1790 8
9,90
1790 89,

90
89,90
1791 66,60
1791 66,60
66,60
1792 60,00
1792 60,00
60,00
1793 46,90
1793 46,90
46,90
1794 41,00
1794 41,00
41,00
1795 21,30
1795 21,30
21,30
1796 16,00
1796 16,00
16,00
1797 6,40
1797 6,40
6,40
1798 4,10
1798 4,10
4,10
1799 6,80
1799 6,80
6,80
1**800** 14,50

1**800** 14,50
1800 14,50
14,50
1801 34,00
1801 34,00
34,00
1802 45,00
1802 45,00
45,00
1803 43,10
1803 43,10
43,10
1804 47,50
1804 47,50
47,50
1805 42,20
1805 42,20
42,20
1806 28,10
1806 28,10
28,10
1807 10,10
1807 10,10
10,10
1808 8,10

1808 8,10
8,10
1809 2,50
1809 2,50
2,50
1810 0,00
1810 0,00
0,00
1811 1,40
1811 1,40
1,40
1812 5,00
1812 5,00
5,00
1813 12,20
1813 12,20
12,20
1814 13,90
1814 13,90
13,90
1815 35,40
1815 35,40
35,40
1816 45,80
1816 45,80
45,80
1817 41,10
1817 41,10
41,10
1818 30,10
1818 30,10
30,10
1819 23,90
1819 23,

90
23,90
1820 1
5,60
1820 15,
60
15,60
1821 6
,60
1821 6,6
0
6,60
1822 4
,00
1822 4,0
0
4,00
1823 1
,80
1823 1,8
0
1,80
1824 8
,50
1824 8,5
0
8,50
1825 1
6,60
1825 16,
60
16,60
1826 3
6,30
1826 36,
30
36,30
1827 4
9,60
1827 49,
60
49,60
1828 6
4,20
1828 64,
20
64,20
1829 6
7,00
1829 67,
00
67,00
1830 7
0,90
1830 70,
90

70,90
1831 4
7,80
1831 47,
80
47,80
1832 2
7,50
1832 27,
50
27,50
1833 8
,50
1833 8,5
0
8,50
1834 1
3,20
1834 13,
20
13,20
1835 5
6,90
1835 56,
90
56,90
1836 1
21,50
1836 12
1,50
121,50
1837 1
38,30
1837 13
8,30
138,30
1838 1
03,20
1838 10
3,20
103,20
1839 8
5,70
1839 85,
70
85,70
1840 6
4,60
1840 64,
60
64,60
1841 3
6,70
1841 36,
70
36,70
1842 2

4,20
1842 24,20
24,20
1843 10,70
1843 10,70
10,70
1844 15,00
1844 15,00
15,00
1845 40,10
1845 40,10
40,10
1846 61,50
1846 61,50
61,50
1847 98,50
1847 98,50
98,50
1848 124,70
1848 124,70
124,70
1849 96,30
1849 96,30
96,30
1**850** 66,60

1**850** 66,60
1850 66,60
66,60
1851 64,50
1851 64,50
64,50
1852 54,10
1852 54,10
54,10
1853 39,00
1853 39,00
39,00
1854 20,60
1854 20,60
20,60
1855 6,70
1855 6,70
6,70
1856 4,30
1856 4,30
4,30
1857 22,70
1857 22,70
22,70
1858 54,80
1858 54,80
54,80
1859 93,80
1859 93,80
93,80

1860 9
5,80
1860 95,
80
95,80
1861 7
7,20
1861 77,
20
77,20
1862 5
9,10
1862 59,
10
59,10
1863 4
4,00
1863 44,
00
44,00
1864 4
7,00
1864 47,
00
47,00
1865 3
0,50
1865 30,
50
30,50
1866 1
6,30
1866 16,
30
16,30
1867 7
,30
1867 7,3
0
7,30
1868 3
7,60
1868 37,
60
37,60
1869 7
4,00
1869 74,
00
74,00
1870 1
39,00
1870 13
9,00
139,00
1871 1

11,20
1871 11
1,20
111,20
1872 1
01,60
1872 10
1,60
101,60
1873 6
6,20
1873 66,
20
66,20
1874 4
4,70
1874 44,
70
44,70
1875 1
7,00
1875 17,
00
17,00
1876 1
1,30
1876 11,
30
11,30
1877 1
2,40
1877 12,
40
12,40
1878 3
,40
1878 3,4
0
3,40
1879 6
,00
1879 6,0
0
6,00
1880 3
2,30
1880 32,
30
32,30
1881 5
4,30
1881 54,
30
54,30
1882 5
9,70

1882 59,70
59,70
1883 63,70
1883 63,70
63,70
1884 63,50
1884 63,50
63,50
1885 52,20
1885 52,20
52,20
1886 25,40
1886 25,40
25,40
1887 13,10
1887 13,10
13,10
1888 6,80
1888 6,80
6,80
1889 6,30
1889 6,30
6,30
1890 7,10
1890 7,10
7,10
1891 35,60
1891 35,60
35,60
1892 73,00
1892 73,00
73,00
1893 85,10
1893 85,10

85,10
1894 78,00
1894 78,00
78,00
1895 64,00
1895 64,00
64,00
1896 41,80
1896 41,80
41,80
1897 26,20
1897 26,20
26,20
1898 26,70
1898 26,70
26,70
1899 12,10
1899 12,10
12,10
1900 9,50
1900 9,50
1900 9,50
9,50
1901 2,70
1901 2,70
2,70
1902 5,00
1902 5,00
5,00
1903 24,40
1903 24,40
24,40
1904 42,00
1904 42,00
42,00
1905 63,50
1905 63,50
63,50
1906 53,80
1906 53,80
53,80
1907 62,00
1907 62,00
62,00
1908 48,50
1908 48,50
48,50
1909 43,90
1909 43,90
43,90
1910 18,60
1910 18,60
18,60
1911 5,70
1911 5,70
5,70

1912 3
,60
1912 3,6
0
3,60
1913 1
,40
1913 1,4
0
1,40
1914 9
,60
1914 9,6
0
9,60
1915 4
7,40
1915 47,
40
47,40
1916 5
7,10
1916 57,
10
57,10
1917 1
03,90
1917 10
3,90
103,90
1918 8
0,60
1918 80,
60
80,60
1919 6
3,60
1919 63,
60
63,60
1920 3
7,60
1920 37,
60
37,60
1921 2
6,10
1921 26,
10
26,10
1922 1
4,20
1922 14,
20
14,20
1923 5

,80
1923 5,8
0
5,80
1924 1
6,70
1924 16,
70
16,70
1925 4
4,30
1925 44,
30
44,30
1926 6
3,90
1926 63,
90
63,90
1927 6
9,00
1927 69,
00
69,00
1928 7
7,80
1928 77,
80
77,80
1929 6
4,90
1929 64,
90
64,90
1930 3
5,70
1930 35,
70
35,70
1931 2
1,20
1931 21,
20
21,20
1932 1
1,10
1932 11,
10
11,10
1933 5
,70
1933 5,7
0
5,70
1934 8
,70

1934 8,70
8,70
1935 36,10
1935 36,10
36,10
1936 79,70
1936 79,70
79,70
1937 114,40
1937 114,40
114,40
1938 109,60
1938 109,60
109,60
1939 88,80
1939 88,80
88,80
1940 67,80
1940 67,80
67,80
1941 47,50
1941 47,50
47,50
1942 30,60
1942 30,60
30,60
1943 16,30
1943 16,30
16,30
1944 9,60
1944 9,60
9,60
1945 33,20
1945 33,20

33,20
1946 92,60
1946 92,60
92,60
1947 151,60
1947 151,60
151,60
1948 136,30
1948 136,30
136,30
1949 134,70
1949 134,70
134,70
1**950** 83,90

1**950** 83,90
1950 83,90
83,90
1951 69,40
1951 69,40
69,40
1952 31,50
1952 31,50
31,50
1953 13,90
1953 13,90
13,90
1954 4,40
1954 4,40
4,40
1960 112,30
1960 112,30
112,30
1961 53,90
1961 53,90
53,90
1962 37,50
1962 37,50
37,50
1963 27,90
1963 27,90
27,90
1964 10,20
1964 10,20
10,20
1965 15,10
1965 15,10
15,10
1966 47,00
1966 47,00
47,00
1967 93,80
1967 93,80
93,80
1968 105,90
1968 105,90

105,90
1969 1
05,50
1969 10
5,50
105,50
1970 1
04,50
1970 10
4,50
104,50
1971 6
6,60
1971 66,
60
66,60
1972 6
8,90
1972 68,
90
68,90
1973 3
8,00
1973 38,
00
38,00
1974 3
4,50
1974 34,
50
34,50
1975 1
5,50
1975 15,
50
15,50
1976 1
2,60
1976 12,
60
12,60
1977 2
7,50
1977 27,
50
27,50
1978 9
2,50
1978 92,
50
92,50
1979 1
55,40
1979 15
5,40
155,40

1980 3
2,27
1980 32,
27
32,27
1981 5
4,25
1981 54,
25
54,25
1982 5
9,65
1982 59,
65
59,65
1983 6
3,62
1983 63,
62
63,62

Вариант 2. Ежегодные данные о производстве табака в 1871–1984 годах

1871 32
7
327
1872
1872 38
5
385
1873
1873 38
2
382
1874
1874 21
7
217
1875
1875 60
9
609
1876
1876 46
6
466

1877
1877 62
1
621
1878
1878 45
5
455
1879
1879 47
2
472
1880
1880 46
9
469
1881
1881 42
6
426
1882
1882 57
9
579
1883
1883 50
9
509
1884
1884 58
0
580
1885
1885 61
1
611
1886
1886 60
9
609
1887
1887 46
9
469
1888
1888 66
1
661
1889
1889 52
5
525
1890
1890 64
8
648

1891
1891 74
7
747
1892
1892 75
7
757
1893
1893 76
7
767
1894
1894 76
7
767
1895
1895 74
5
745
1**896**
760
1**896**
1896 76
0
760
1897
1897 70
3
703
1898
1898 90
9
909
1899
1899 87
0
870
1900
1900 85
2
852
1901
1901 88
6
886
1902
1902 96
0
960
1903
1903 97
6
976
1904
1904 85

7
857
1905
1905 93
9
939
1906
1906 97
3
973
1907
1907 88
6
886
1908
1908 83
6
836
1909
1909 10
54
1054
1910
1910 11
42
1142
1911
1911 94
1
941
1912
1912 11
17
1117
1913
1913 99
2
992
1914
1914 10
37
1037
1915
1915 11
57
1157
1916
1916 12
07
1207
1917
1917 13
26
1326
1918
1918 14

45
1445
1919
1919 14
44
1444
1920
1920 15
09
1509
1**921**
1005
1**921**
1921 10
05
1005
1922
1922 12
54
1254
1923
1923 15
18
1518
1924
1924 12
45
1245
1925
1925 13
76
1376
1926
1926 12
89
1289
1927
1927 12
11
1211
1928
1928 13
73
1373
1929
1929 15
33
1533
1930
1930 16
48
1648
1931
1931 15
65
1565

1932
1932 10
18
1018
1933
1933 13
72
1372
1934
1934 10
85
1085
1935
1935 13
02
1302
1936
1936 11
63
1163
1937
1937 15
69
1569
1938
1938 13
86
1386
1939
1939 18
81
1881
1940
1940 14
60
1460
1941
1941 12
62
1262
1942
1942 14
08
1408
1943
1943 14
06
1406
1944
1944 19
51
1951
1945
1945 19
91
1991

1946
1315
1946
1946 13
15
1315
1947
1947 21
07
2107
1948
1948 19
80
1980
1949
1949 19
69
1969
1950
1950 20
30
2030
1951
1951 23
32
2332
1952
1952 22
56
2256
1953
1953 20
59
2059
1954
1954 22
44
2244
1955
1955 21
93
2193
1956
1956 21
76
2176
1957
1957 16
68
1668
1958
1958 17
36
1736
1959
1959 17

96
1796
1960
1960 19
44
1944
1961
1961 20
61
2061
1962
1962 23
15
2315
1963
1963 23
44
2344
1964
1964 22
28
2228
1965
1965 18
55
1855
1966
1966 18
87
1887
1967
1967 19
68
1968
1968
1968 17
10
1710
1969
1969 18
04
1804
1970
1970 19
06
1906
1971
1705
1971
1971 17
05
1705
1972
1972 17
49
1749

1973
1973 17
42
1742
1974
1974 19
90
1990
1975
1975 21
82
2182
1976
1976 21
37
2137
1977
1977 19
14
1914
1978
1978 20
25
2025
1979
1979 15
27
1527
1980
1980 17
86
1786
1981
1981 20
64
2064
1982
1982 19
94
1994
1983
1983 14
29
1429
1984
1984 17
28
1728

Вариант 3. Ежедневные данные о числе дефектов на грузовик при сборке грузовиков

1 1,2
1,2
2 1,5
2 1,5
1,5
3 1,54
3 1,54
1,54
4 2,7
4 2,7
2,7
5 1,95
5 1,95
1,95
6 2,4
6 2,4
2,4
7 3,44
7 3,44
3,44
8 2,83
8 2,83
2,83
9 1,76
9 1,76
1,76
10 2
10 2
2
1**1** 2,09
1**1** 2,09
11 2,09
2,09
12 1,89
12 1,89
1,89
13 1,8
13 1,8
1,8
14 1,25
14 1,25
1,25
15 1,58
15 1,58
1,58
16 2,25
16 2,25
2,25
17 2,5
17 2,5
2,5
18 2,05
18 2,05
2,05
19 1,46

19 1,46

1,46
20 1,5
4
20 1,54

1,54
2**1** 1,
42
2**1** 1,4
2
21 1,42

1,42
22 1,5
7
22 1,57

1,57
23 1,4

23 1,4
1,4
24 1,5
1
24 1,51

1,51
25 1,0
8
25 1,08

1,08
26 1,2
7
26 1,27

1,27
27 1,1
8
27 1,18

1,18
28 1,3
9
28 1,39

1,39
29 1,4
2
29 1,42

1,42
30 2,0
8

30 2,08

2,08
3**1** 1,
85
3**1** 1,8
5
31 1,85

1,85
32 1,8
2
32 1,82

1,82
33 2,0
7
33 2,07

2,07
34 2,3
2
34 2,32

2,32
35 1,2
3
35 1,23

1,23
36 2,9
1
36 2,91

2,91
37 1,7
7
37 1,77

1,77
38 1,6
1
38 1,61

1,61
39 1,2
5
39 1,25

1,25
40 1,1
5
40 1,15

1,15
4**1** 1,

37
41 1,3
7
41 1,37

1,37
42 1,7
9
42 1,79

1,79
43 1,6
8
43 1,68

1,68
44 1,7
8
44 1,78

1,78
45 1,8
4
45 1,84

1,84

Вариант 4. Ежемесячное количество молодых людей, ищущих работу

Jan/1961
375
375
Feb/196
1 384
Feb/1961
384
384
Mar/19
61 383
Mar/1961
383
383
Apr/19
61 326
Apr/1961
326
326

May/19
61 344
May/196
1 344
344
Jun/19
61 375
Jun/1961
375
375
Jul/196
1 419
Jul/1961
419
419
Aug/19
61 424
Aug/1961
424
424
Sep/196
1 429
Sep/1961
429
429
Oct/196
1 399
Oct/1961
399
399
Nov/19
61 376
Nov/1961
376
376
Dec/19
61 288
Dec/1961
288
288
Jan/196
2 360
Jan/1962
360
360
Feb/19
62 376
Feb/1962
376
376
Mar/19
62 360
Mar/196
2 360
360
Apr/19

62 381
Apr/1962 381
381
May/1962 354

May/1962 354
354
Jun/1962 301

Jun/1962 301
301
Jul/1962 333

Jul/1962 333
333
Aug/1962 339

Aug/1962 339
339
Sep/1962 346

Sep/1962 346
346
Oct/1962 352

Oct/1962 352
352
Nov/1962 378

Nov/1962 378
378
Dec/1962 360

Dec/1962 360
360
Jan/1963 388

Jan/1963 388
388
Feb/1963 398

Feb/1963 398
398
Mar/1963 377

Mar/1963 377
377
Jan/1966 42
1
Jan/1966 421

Jan/1966 421
421
Feb/1966 374

Feb/1966 374

374
Mar/1966 401
Mar/1966 401
401
Apr/1966 451
Apr/1966 451
451
May/1966 465
May/1966 465
465
Jun/1966 456
Jun/1966 456
456
Jul/1966 469
Jul/1966 469
469
Aug/1966 466
Aug/1966 466
466
Sep/1966 412
Sep/1966 412
412
Oct/1966 427
Oct/1966 427
427
Nov/1966 414
Nov/1966 414
414
Dec/1966 384
Dec/1966 384
384
Jan/1967 328
Jan/1967 328

328
Feb/19
67 395
Feb/1967
395
395
Mar/19
67 381
Mar/196
7 381
381
Apr/19
67 360
Apr/1967
360
360
May/19
67 383
May/196
7 383
383
Jun/19
67 383
Jun/1967
383
383
Jul/196
7 403
Jul/1967
403
403
Aug/19
67 425
Aug/1967
425
425
Sep/196
7 422
Sep/1967
422
422
Oct/196
7 414
Oct/1967
414
414
Nov/19
67 382
Nov/1967
382
382
Dec/19
67 390
Dec/1967
390

390
Jan/196
8 320
Jan/1968
320
320
Feb/19
68 412
Feb/1968
412
412
Mar/19
68 437
Mar/196
8 437
437
Jan/1
971 587

Jan/197
1 587
Jan/1971
587
587
Feb/19
71 560
Feb/1971
560
560
Mar/19
71 590
Mar/197
1 590
590
Apr/19
71 556
Apr/1971
556
556
May/19
71 582
May/197
1 582
582
Jun/19
71 527
Jun/1971
527
527
Jul/197
1 585
Jul/1971
585
585
Aug/19
71 556

Aug/1971 556
556
Sep/1971 574

Sep/1971 574
574
Oct/1971 556

Oct/1971 556
556
Nov/1971 582

Nov/1971 582
582
Dec/1971 583

Dec/1971 583
583
Jan/1972 644

Jan/1972 644
644
Feb/1972 620

Feb/1972 620
620
Mar/1972 618

Mar/1972 618
618
Apr/1972 623

Apr/1972 623
623
May/1972 546

May/1972 546
546
Jun/1972 568

Jun/1972 568
568
Jul/1972 595

Jul/1972 595
595
Aug/1972 605

Aug/1972 605
605
Sep/1972 598

Sep/1972 598
598
Oct/1972 592

Oct/1972
592
592
Nov/19
72 558
Nov/1972
558
558
Dec/19
72 595
Dec/1972
595
595
Jan/197
3 549
Jan/1973
549
549
Feb/19
73 637
Feb/1973
637
637
Mar/19
73 568
Mar/197
3 568
568
Jan/1
976 760

Jan/197
6 760
Jan/1976
760
760
Feb/19
76 781
Feb/1976
781
781
Mar/19
76 769
Mar/197
6 769
769
Apr/19
76 766
Apr/1976
766
766
May/19
76 752
May/197
6 752

752
Jun/19
76 751
Jun/1976
751
751
Jul/197
6 761
Jul/1976
761
761
Aug/19
76 873
Aug/1976
873
873
Sep/197
6 750
Sep/1976
750
750
Oct/197
6 758
Oct/1976
758
758
Nov/19
76 772
Nov/1976
772
772
Dec/19
76 791
Dec/1976
791
791
Jan/197
7 813
Jan/1977
813
813
Feb/19
77 781
Feb/1977
781
781
Mar/19
77 797
Mar/197
7 797
797
Apr/19
77 802
Apr/1977
802

802
May/19
77 782
May/197
7 782
782
Jun/19
77 838
Jun/1977
838
838
Jul/197
7 756
Jul/1977
756
756
Aug/19
77 764
Aug/1977
764
764
Sep/197
7 796
Sep/1977
796
796
Oct/197
7 781
Oct/1977
781
781
Nov/19
77 780
Nov/1977
780
780
Dec/19
77 679
Dec/1977
679
679
Jan/197
8 748
Jan/1978
748
748
Feb/19
78 759
Feb/1978
759
759
Mar/19
78 749
Mar/197
8 749
749

Jan/1981 76
2
Jan/1981 762

Jan/1981 762
762
Feb/1981 783

Feb/1981 783
783
Mar/1981 796

Mar/1981 796
796
Apr/1981 803

Apr/1981 803
803
May/1981 806

May/1981 806
806
Jun/1981 765

Jun/1981 765
765
Jul/1981 781

Jul/1981 781
781
Aug/1981 768

Aug/1981 768
768
Sep/1981 812

Sep/1981 812
812
Oct/1981 854

Oct/1981 854
854
Nov/1981 858

Nov/1981 858
858
Dec/1981 818

Dec/1981 818
818
Jan/1982 856

Jan/1982 856
856
Feb/1982 897

Feb/1982
897
897
Mar/1982 817
Mar/1982 817
817
Apr/1982 872
Apr/1982
872
872
May/1982 895
May/1982 895
895
Jun/1982 825
Jun/1982
825
825
Jul/1982 922
Jul/1982
922
922
Aug/1982 915
Aug/1982
915
915
Sep/1982 902
Sep/1982
902
902
Oct/1982 908
Oct/1982
908
908
Nov/1982 911
Nov/1982
911
911
Dec/1982 919
Dec/1982
919
919
Jan/1983 861

Jan/1983
861
861
Feb/1983 827
Feb/1983
827
827
Mar/1983 855
Mar/1983 855
855

Apr/1963
383
383
May/1963 449
May/1963
449
449
Jun/1963 415
Jun/1963
415
415
Jul/1963 429
Jul/1963
429
429
Aug/1963 369
Aug/1963
369
369
Sep/1963 414
Sep/1963
414
414
Oct/1963 462
Oct/1963
462
462
Nov/1963 447

Nov/1963
447
447
Dec/1963 403
Dec/1963
403
403
Jan/1964 409
Jan/1964
409
409
Feb/1964 390
Feb/1964
390
390
Mar/1964 380
Mar/1964 380
380
Apr/1964 438
Apr/1964
438
438
May/1964 431
May/1964 431
431
Jun/1964 426
Jun/1964
426
426
Jul/1964 348
Jul/1964
348
348
Aug/1964 394
Aug/1964
394
394
Sep/1964 396
Sep/1964
396
396
Oct/1964 451

Oct/1964
451
451
Nov/1964 384
Nov/1964
384
384
Dec/1964 491
Dec/1964
491
491
Jan/1965 466
Jan/1965
466
466
Feb/1965 454
Feb/1965
454
454
Mar/1965 442
Mar/1965
442
442
Apr/1965 475
Apr/1965
475
475
May/1965 401
May/1965
401
401
Jun/1965 406
Jun/1965
406
406
Jul/1965 385
Jul/1965
385
385
Aug/1965 380
Aug/1965
380
380
Sep/1965 422

Sep/1965
422
422
Oct/196
5 397
Oct/1965
397
397
Nov/19
65 430
Nov/1965
430
430
Dec/19
65 433
Dec/1965
433
433
Apr/1
968 421

Apr/19
68 421
Apr/1968
421
421
May/19
68 450
May/196
8 450
450
Jun/19
68 **4**42
Jun/1968
442
442
Jul/196
8 450
Jul/1968
450
450
Aug/19
68 412
Aug/1968
412
412
Sep/196
8 422
Sep/1968
422
422
Oct/196
8 372
Oct/1968

372
372
Nov/1968 375
Nov/1968
375
375
Dec/1968 392
Dec/1968
392
392
Jan/1969 356
Jan/1969
356
356
Feb/1969 392
Feb/1969
392
392
Mar/1969 426
Mar/1969 426
426
Apr/1969 442
Apr/1969
442
442
May/1969 426
May/1969 426
426
Jun/1969 406
Jun/1969
406
406
Jul/1969 392
Jul/1969
392
392
Aug/1969 426
Aug/1969
426
426
Sep/1969 445
Sep/1969

445
445
Oct/196
9 464
Oct/1969
464
464
Nov/19
69 379
Nov/1969
379
379
Dec/19
69 409
Dec/1969
409
409
Jan/197
0 497
Jan/1970
497
497
Feb/19
70 459
Feb/1970
459
459
Mar/19
70 513
Mar/197
0 513
513
Apr/19
70 549
Apr/1970
549
549
May/19
70 447
May/197
0 447
447
Jun/19
70 445
Jun/1970
445
445
Jul/197
0 432
Jul/1970
432
432
Aug/19
70 514
Aug/1970

514
514
Sep/197
0 565
Sep/1970
565
565
Oct/197
0 557
Oct/1970
557
557
Nov/19
70 601
Nov/1970
601
601
Dec/197
0 582
Dec/1970
582
582
Apr/1
973 605

Apr/19
73 605
Apr/1973
605
605
May/19
73 594
May/1973
594
594
Jun/197
3 567
Jun/1973
567
567
Jul/197
3 545
Jul/1973
545
545
Aug/19
73 545
Aug/1973
545
545
Sep/197
3 592
Sep/1973
592
592

Oct/197
3 576
Oct/1973
576
576
Nov/19
73 593
Nov/1973
593
593
Dec/19
73 603
Dec/1973
603
603
Jan/197
4 631
Jan/1974
631
631
Feb/19
74 614
Feb/1974
614
614
Mar/19
74 617
Mar/197
4 617
617
Apr/19
74 546
Apr/1974
546
546
May/19
74 632
May/197
4 632
632
Jun/19
74 673
Jun/1974
673
673
Jul/197
4 732
Jul/1974
732
732
Aug/19
74 593
Aug/1974
593
593

Sep/1974 693
Sep/1974 693
693
Oct/1974 730
Oct/1974 730
730
Nov/1974 731
Nov/1974 731
731
Dec/1974 733
Dec/1974 733
733
Jan/1975 802
Jan/1975 802
802
Feb/1975 755
Feb/1975 755
755
Mar/1975 805
Mar/1975 805
805
Apr/1975 751
Apr/1975 751
751
May/1975 855
May/1975 855
855
Jun/1975 769
Jun/1975 769
769
Jul/1975 800
Jul/1975 800
800

Aug/19
75 825
Aug/1975
825
825
Sep/197
5 799
Sep/1975
799
799
Oct/197
5 802
Oct/1975
802
802
Nov/19
75 765
Nov/1975
765
765
Dec/19
75 827
Dec/1975
827
827
Apr/1
978 756

Apr/19
78 756
Apr/1978
756
756
May/19
78 802
May/197
8 802
802
Jun/19
78 754
Jun/1978
754
754
Jul/197
8 792
Jul/1978
792
792
Aug/19
78 772
Aug/1978
772
772
Sep/197

8 769
Sep/1978
769
769
Oct/1978
731
Oct/1978
731
731
Nov/1978 746
Nov/1978
746
746
Dec/1978
741
Dec/1978
741
741
Jan/1979 712
Jan/1979
712
712
Feb/1979 723
Feb/1979
723
723
Mar/1979 698
Mar/1979
698
698
Apr/1979 746
Apr/1979
746
746
May/1979 754
May/1979
754
754
Jun/1979 735
Jun/1979
735
735
Jul/1979 722
Jul/1979
722
722
Aug/19

79 737
Aug/1979
737
737
Sep/197
9 728
Sep/1979
728
728
Oct/197
9 773
Oct/1979
773
773
Nov/19
79 723
Nov/1979
723
723
Dec/19
79 741
Dec/1979
741
741
Jan/198
0 738
Jan/1980
738
738
Feb/19
80 765
Feb/1980
765
765
Mar/19
80 748
Mar/198
0 748
748
Apr/19
80 707
Apr/1980
707
707
May/19
80 808
May/198
0 808
808
Jun/19
80 746
Jun/1980
746
746
Jul/198

0 773
Jul/1980
773
773
Aug/19
80 751
Aug/1980
751
751
Sep/198
0 721
Sep/1980
721
721
Oct/198
0 731
Oct/1980
731
731
Nov/19
80 735
Nov/1980
735
735
Dec/19
80 701
Dec/1980
701
701
Apr/1
983 867

Apr/19
83 867
Apr/1983
867
867
May/19
83 836
May/198
3 836
836
Jun/19
83 **9**16
Jun/1983
916
916
Jul/198
3 828
Jul/1983
828
828
Aug/19
83 835

Aug/1983
835
835
Sep/1983 792
Sep/1983
792
792
Oct/1983 771
Oct/1983
771
771
Nov/1983 757
Nov/1983
757
757
Dec/1983 756
Dec/1983
756
756
Jan/1984 712
Jan/1984
712
712
Feb/1984 733
Feb/1984
733
733
Mar/1984 746
Mar/1984 746
746
Apr/1984 728
Apr/1984
728
728
May/1984 707
May/1984 707
707
Jun/1984 666
Jun/1984
666
666
Jul/1984 636

Jul/1984
636
636
Aug/1984 676
Aug/1984
676
676
Sep/1984 696
Sep/1984
696
696
Oct/1984 654
Oct/1984
654
654
Nov/1984 613
Nov/1984
613
613
Dec/1984 677
Dec/1984
677
677
Jan/1985 705
Jan/1985
705
705
Feb/1985 680
Feb/1985
680
680
Mar/1985 699
Mar/1985
699
699
Apr/1985 650
Apr/1985
650
650
May/1985 687
May/1985
687
687
Jun/1985 638

Jun/1985
638
638
Jul/1985 670
Jul/1985
670
670
Aug/1985 555
Aug/1985
555
555
Sep/1985 631
Sep/1985
631
631
Oct/1985 676
Oct/1985
676
676
Nov/1985 659
Nov/1985
659
659
Dec/1985 689
Dec/1985
689
689

Вариант 5. Ежегодное число несчастных случаев в Пенсильвании на 100 тыс. человек

1950 54,2
54,2
1951 55
1951 55
55
1952 53,9
1952 53,9
53,9
1953 53,1

1953 53,
1
53,1
1954 5
0,2
1954 50,
2
50,2
1955 5
0,8
1955 50,
8
50,8
1956 4
8,8
1956 48,
8
48,8
1**957**
48,1
1**957** 4
8,1
1957 48,
1
48,1
1958 4
5,5
1958 45,
5
45,5
1959 4
4,5
1959 44,
5
44,5
1960 4
3
1960 43

43
1961 4
1,5
1961 41,
5
41,5
1962 4
4,2
1962 44,
2
44,2
1963 4
5,1
1963 45,
1
45,1

1**964**
45,1
1**964** 4
5,1
1964 45,
1
45,1
1965 4
6,4
1965 46,
4
46,4
1966 4
4,9
1966 44,
9
44,9
1967 4
6,9
1967 46,
9
46,9
1968 4
7,5
1968 47,
5
47,5
1969 4
5,4
1969 45,
4
45,4
1970 4
5,2
1970 45,
2
45,2
1**971**
45,3
1**971** 4
5,3
1971 45,
3
45,3
1972 4
4,1
1972 44,
1
44,1
1973 4
3,2
1973 43,
2
43,2
1974 4

0,6
1974 40,
6
40,6
1975 3
9,8
1975 39,
8
39,8
1976 3
8,2
1976 38,
2
38,2
1977 3
9,5
1977 39,
5
39,5
1**978**
38,2
1**978** 3
8,2
1978 38,
2
38,2
1979 4
0,7
1979 40,
7
40,7
1980 3
8,7
1980 38,
7
38,7
1981 3
6,9
1981 36,
9
36,9
1982 3
4,6
1982 34,
6
34,6
1983 3
4,9
1983 34,
9
34,9
1984 3
5,1
1984 35,
1

35,1

Вариант 6. Ежегодное количество проданных компанией шкур рыси

1857 23362
23362
1858 31642
1858 31642
31642
1859 33757
1859 33757
33757
1860 23226
1860 23226
23226
1861 15178
1861 15178
15178
1862 7272
1862 7272
7272
1863 4448
1863 4448
4448
1864 4926
1864 4926
4926
1865 5437
1865 5437
5437
1866 16498
1866 16498
16498
1867 35971
1867 35971
35971
1**868** 76556

1**868** 76556
1868 76556
76556
1869 68392
1869 68392
68392
1870 37447
1870 37447
37447
1871 45686

1871 45
686
45686
1872 7
942
1872 79
42
7942
1873 5
123
1873 51
23
5123
1874 7
106
1874 71
06
7106
1875 1
1250
1875 11
250
11250
1876 1
8774
1876 18
774
18774
1877 3
0508
1877 30
508
30508
1878 4
2834
1878 42
834
42834
1**879**
27345
1**879** 2
7345
1879 27
345
27345
1880 1
7834
1880 17
834
17834
1881 1
5386
1881 15
386
15386

1882 9
443
1882 94
43
9443
1883 7
599
1883 75
99
7599
1884 8
061
1884 80
61
8061
1885 2
7187
1885 27
187
27187
1886 5
1511
1886 51
511
51511
1887 7
4050
1887 74
050
74050
1888 7
8773
1888 78
773
78773
1889 3
3899
1889 33
899
33899
1**890**
18886
1**890** 1
8886
1890 18
886
18886
1891 1
1520
1891 11
520
11520
1892 8
352
1892 83

52
8352
1893 8
660
1893 86
60
8660
1894 1
2902
1894 12
902
12902
1895 2
0331
1895 20
331
20331
1896 3
6853
1896 36
853
36853
1897 5
6407
1897 56
407
56407
1898 3
9437
1898 39
437
39437
1899 2
6761
1899 26
761
26761
1900 1
5185
1900 15
185
15185
1**901**
4473
1**901** 4
473
1901 44
73
4473
1902 5
781
1902 57
81
5781
1903 9

117
1903 91
17
9117
1904 1
9267
1904 19
267
19267
1905 3
6116
1905 36
116
36116
1906 5
8850
1906 58
850
58850
1907 6
1478
1907 61
478
61478
1908 3
6300
1908 36
300
36300
1909 9
704
1909 97
04
9704
1910 3
410
1910 34
10
3410
1911 3
774
1911 37
74
3774

Вариант 7. Смоделированный временной ряд

1 -2,401
-2,401

2 -
0,574
2 -
0,574
-0,574
3 0,38
2
3 0,382

0,382
4 -
0,535
4 -
0,535
-0,535
5 -
1,639
5 -
1,639
-1,639
6 -
0,96
6 -
0,96
-0,96
7 -
1,118
7 -
1,118
-1,118
8 -
0,719
8 -
0,719
-0,719
9 -
1,236
9 -
1,236
-1,236
10 0,1
17
10 0,117

0,117
11 -
0,493
11 -
0,493
-0,493
12 -
2,282
12 -
2,282
-2,282

13 -
1,823
13 -
1,823
-1,823
14 0,6
45
14 0,645

0,645
15 -
0,179
15 -
0,179
-0,179
16 0,5
89
16 0,589

0,589
17 1,4
13
17 1,413

1,413
18 0,3
7
18 0,37

0,37
19 0,0
82
19 0,082

0,082
20 -
0,531
20 -
0,531
-0,531
21 -
1,891
21 -
1,891
21 -
1,891
-1,891
22 -
0,961
22 -
0,961
-0,961
23 -
0,856
23 -

0,856
-0,856
24 -
0,79
24 -
0,79
-0,79
25 -
1,476
25 -
1,476
-1,476
26 -
2,491
26 -
2,491
-2,491
27 -
4,479
27 -
4,479
-4,479
28 -
2,809
28 -
2,809
-2,809
29 -
2,154
29 -
2,154
-2,154
30 -
1,532
30 -
1,532
-1,532
31 -
2,119
31 -
2,119
-2,119
32 -
3,349
32 -
3,349
-3,349
33 -
1,588
33 -
1,588
-1,588
34 0,7
4
34 0,74

0,74
35 0,9
07
35 0,907

0,907
36 1,5
4
36 1,54

1,54
37 0,5
57
37 0,557

0,557
38 2,2
59
38 2,259

2,259
39 2,6
22
39 2,622

2,622
40 0,7
01
40 0,701

0,701
4**1** 2,
463
4**1** 2,4
63
41 2,463

2,463
42 2,7
14
42 2,714

2,714
43 2,0
89
43 2,089

2,089
44 3,7
5
44 3,75

3,75
45 4,3

22
45 4,322

4,322
46 3,1
86
46 3,186

3,186
47 3,1
92
47 3,192

3,192
48 2,9
39
48 2,939

2,939
49 3,2
63
49 3,263

3,263
50 3,2
79
50 3,279

3,279
51 0,2
95
51 0,295

0,295
52 0,2
27
52 0,227

0,227
53 1,3
56
53 1,356

1,356
54 1,9
12
54 1,912

1,912
55 1,0
6
55 1,06

1,06
56 0,3

7
56 0,37

0,37
57 -
0,195
57 -
0,195
-0,195
58 0,3
4
58 0,34

0,34
59 1,0
84
59 1,084

1,084
60 1,2
37
60 1,237

1,237
6**1** 0,
61
6**1** 0,6
1
61 0,61

0,61
62 2,1
26
62 2,126

2,126
63 3,9
6
63 3,96

3,96
64 3,3
17
64 3,317

3,317
65 2,1
67
65 2,167

2,167
66 1,2
92
66 1,292

1,292
67 0,5
95
67 0,595

0,595
68 0,1
4
68 0,14

0,14
69 -
0,082
69 -
0,082
-0,082
70 -
0,769
70 -
0,769
-0,769
71 0,8
7
71 0,87

0,87
72 1,5
51
72 1,551

1,551
73 2,6
1
73 2,61

2,61
74 2,1
93
74 2,193

2,193
75 1,3
53
75 1,353

1,353
76 -
0,6
76 -
0,6
-0,6
77 -
0,455
77 -
0,455

-0,455
78 0,2
03
78 0,203

0,203
79 1,4
72
79 1,472

1,472
80 1,3
67
80 1,367

1,367
8**1** 1,
875
8**1** 1,8
75
81 1,875

1,875
82 2,0
82
82 2,082

2,082
83 1,6
04
83 1,604

1,604
84 2,0
33
84 2,033

2,033
85 3,7
46
85 3,746

3,746
86 2,9
54
86 2,954

2,954
87 0,6
76
87 0,676

0,676
88 1,1
63

88 1,163

1,163
89 1,3
68
89 1,368

1,368
90 0,3
43
90 0,343

0,343
91 -
0,334
91 -
0,334
-0,334
92 1,0
41
92 1,041

1,041
93 1,3
28
93 1,328

1,328
94 1,3
25
94 1,325

1,325
95 0,9
68
95 0,968

0,968
96 1,9
7
96 1,97

1,97
97 2,2
96
97 2,296

2,296
98 2,8
96
98 2,896

2,896
99 1,9
18

99 1,918

1,918
100 1,569
100 1,569
1,569

Вариант 8. Смоделированный временной ряд

1 -1,453
-1,453
2 0,867
2 0,867
0,867
3 0,727
3 0,727
0,727
4 -0,765
4 -0,765
-0,765
5 -1,317
5 -1,317
-1,317
6 0,024
6 0,024
0,024
7 -0,542
7 -0,542
-0,542
8 -0,048
8 -0,048
-0,048
9 -0,805
9 -0,805
-0,805
10 0,858
10 0,858
0,858
11 -0,563
11 -0,563
-0,563
12 -1,986
12 -1,986
-1,986
13 -0,454
13 -0,454
-0,454
14 1,738

14 1,738

1,738
15 -
0,566
15 -
0,566
-0,566
16 0,6
97
16 0,697

0,697
17 1,0
6
17 1,06

1,06
18 -
0,478
18 -
0,478
-0,478
19 -
0,14
19 -
0,14
-0,14
20 -
0,581
20 -
0,581
-0,581
2**1** -
1,572
2**1** -
1,572
21 -
1,572
-1,572
22 0,1
74
22 0,174

0,174
23 -
0,289
23 -
0,289
-0,289
24 -
0,27
24 -
0,27
-0,27

25 -
1,002
25 -
1,002
-1,002
26 -
1,605
26 -
1,605
-1,605
27 -
2,984
27 -
2,984
-2,984
28 -
0,122
28 -
0,122
-0,122
29 0,4
69
29 0,469

0,469
30 -
0,239
30 -
0,239
-0,239
31 -
1,2
31 -
1,2
-1,2
32 -
2,077
32 -
2,077
-2,077
33 0,4
21
33 0,421

0,421
34 1,6
93
34 1,693

1,693
35 0,4
63
35 0,463

0,463

36 0,996
36 0,996

0,996
37 -0,367
37 -0,367
-0,367
38 1,925
38 1,925

1,925
39 1,267
39 1,267

1,267
40 -0,872
40 -0,872
-0,872
41 2,043
41 2,043
41 2,043

2,043
42 1,236
42 1,236

1,236
43 0,461
43 0,461

0,461
44 2,497
44 2,497

2,497
45 2,072
45 2,072

2,072
46 0,593
46 0,593

0,593
47 1,2
81
47 1,281

1,281
48 1,0
23
48 1,023

1,023
49 1,5

49 1,5
1,5
50 1,3
21
50 1,321

1,321
51 -
1,673
51 -
1,673
-1,673
52 0,0
5
52 0,05

0,05
53 1,2
19
53 1,219

1,219
54 1,0
98
54 1,098

1,098
55 -
0,087
55 -
0,087
-0,087
56 -
0,266
56 -
0,266
-0,266
57 -
0,417
57 -
0,417

-0,417
58 0,4
57
58 0,457

0,457
59 0,8
8
59 0,88

0,88
60 0,5
86
60 0,586

0,586
6**1** -
0,132
6**1** -
0,132
61 -
0,132
-0,132
62 1,7
6
62 1,76

1,76
63 2,6
84
63 2,684

2,684
64 0,9
41
64 0,941

0,941
65 0,1
77
65 0,177

0,177
66 -
0,008
66 -
0,008
-0,008
67 -
0,18
67 -
0,18
-0,18
68 -
0,217

68 -
0,217
-0,217
69 -
0,165
69 -
0,165
-0,165
70 -
0,72
70 -
0,72
-0,72
71 1,3
32
71 1,332

1,332
72 1,0
29
72 1,029

1,029
73 1,6
79
73 1,679

1,679
74 0,6
27
74 0,627

0,627
75 0,0
38
75 0,038

0,038
76 -
1,412
76 -
1,412
-1,412
77 -
0,095
77 -
0,095
-0,095
78 0,4
76
78 0,476

0,476
79 1,3
5

79 1,35

1,35
80 0,4
84
80 0,484

0,484
81 1,
055
81 1,0
55
81 1,055

1,055
82 0,9
57
82 0,957

0,957
83 0,3
55
83 0,355

0,355
84 1,0
71
84 1,071

1,071
85 2,5
26
85 2,526

2,526
86 0,7
07
86 0,707

0,707
87 -
1,096
87 -
1,096
-1,096
88 0,7
57
88 0,757

0,757
89 0,6
7
89 0,67

0,67

90 -
0,477
90 -
0,477
-0,477
91 -
0,54
91 -
0,54
-0,54
92 1,2
41
92 1,241

1,241
93 0,7
04
93 0,704

0,704
94 0,5
28
94 0,528

0,528
95 0,1
73
95 0,173

0,173
96 1,3
89
96 1,389

1,389
97 1,1
15
97 1,115

1,115
98 1,5
19
98 1,519

1,519
99 0,1
8
99 0,18

0,18
100 0,
419
100 0,41
9
0,419

Вариант 9. Смоделированный временной ряд

1 0,315

0,315
2 -
0,458
2 -
0,458
-0,458
3 -
0,488
3 -
0,488
-0,488
4 -
0,17
4 -
0,17
-0,17
5 0,56
5
5 0,565

0,565
6 -
0,344
6 -
0,344
-0,344
7 -
1,176
7 -
1,176
-1,176
8 -
1,054
8 -
1,054
-1,054
9 -
0,826
9 -
0,826
-0,826
10 0,7
1
10 0,71

0,71
11 -
0,341
11 -
0,341
-0,341
12 -
1,809
12 -
1,809
-1,809
13 -
1,242
13 -
1,242
-1,242
14 -
0,667
14 -
0,667
-0,667
15 -
0,999
15 -
0,999
-0,999
16 2,8
12
16 2,812

2,812
17 1,2
86
17 1,286

1,286
18 -
1,084
18 -
1,084
-1,084
19 -
1,505
19 -
1,505
-1,505
20 -
2,556
20 -
2,556
-2,556
21 -
0,144
21 -
0,144

21 -
0,144
-0,144
22 -
1,749
22 -
1,749
-1,749
23 -
3,032
23 -
3,032
-3,032
24 -
2,958
24 -
2,958
-2,958
25 -
2,827
25 -
2,827
-2,827
26 -
3,392
26 -
3,392
-3,392
27 -
2,431
27 -
2,431
-2,431
28 -
2,757
28 -
2,757
-2,757
29 -
2,822
29 -
2,822
-2,822
30 -
3,314
30 -
3,314
-3,314
31 -
2,738
31 -
2,738
-2,738
32 -
1,979
32 -

1,979
-1,979
33 -1,671
33 -1,671
-1,671
34 -2,977
34 -2,977
-2,977
35 -0,709
35 -0,709
-0,709
36 0,718
36 0,718
0,718
37 0,736
37 0,736
0,736
38 0,879
38 0,879
0,879
39 1,642
39 1,642
1,642
40 2,18
40 2,18
2,18
41 1,963
41 1,963
41 1,963
1,963
42 0,716
42 0,716
0,716
43 0,769
43 0,769
0,769
44 0,973
44 0,973
0,973
45 0,334
45 0,334
0,334
46 1,309
46 1,309
1,309
47 0,878
47 0,878
0,878
48 0,062
48 0,062
0,062
49 0,169
49 0,169
0,169
50 0,677
50 0,677

0,677
51 1,8
51
51 1,851

1,851
52 0,2
42
52 0,242

0,242
53 0,8
28
53 0,828

0,828
54 -
0,317
54 -
0,317
-0,317
55 -
1,042
55 -
1,042
-1,042
56 -
2,093
56 -
2,093
-2,093
57 0,6
53
57 0,653

0,653
58 0,2
61
58 0,261

0,261
59 2,0
2
59 2,02

2,02
60 2,1
36
60 2,136

2,136
6**1** 1,
635
6**1** 1,6
35

61 1,635

1,635
62 -
0,141
62 -
0,141
-0,141
63 -
1,747
63 -
1,747
-1,747
64 -
2,047
64 -
2,047
-2,047
65 -
0,752
65 -
0,752
-0,752
66 -
0,211
66 -
0,211
-0,211
67 -
1,062
67 -
1,062
-1,062
68 -
1,565
68 -
1,565
-1,565
69 0,2
32
69 0,232

0,232
70 0,0
15
70 0,015

0,015
71 -
0,935
71 -
0,935
-0,935
72 -
0,338

72 -
0,338
-0,338
73 0,8
53
73 0,853

0,853
74 0,8
88
74 0,888

0,888
75 3,0
69
75 3,069

3,069
76 3,3
64
76 3,364

3,364
77 3,8
54
77 3,854

3,854
78 4,4
19
78 4,419

4,419
79 2,1
45
79 2,145

2,145
80 2,2
91
80 2,291

2,291
8**1** 1,
753
8**1** 1,7
53
81 1,753

1,753
82 1,0
58
82 1,058

1,058
83 1,0

48
83 1,048
1,048
84 0,2
84 0,2
0,2
85 1,424
85 1,424
1,424
86 0,59
86 0,59
0,59
87 0,356
87 0,356
0,356
88 0,476
88 0,476
0,476
89 0,684
89 0,684
0,684
90 -2,26
90 -2,26
-2,26
91 -0,569
91 -0,569
-0,569
92 -1,014
92 -1,014
-1,014
93 -0,207
93 -0,207
-0,207
94 0,638
94 0,638
0,638
95 -0,664
95 -0,664
-0,664
96 -0,469
96 -0,469
-0,469
97 -0,215
97 -0,215
-0,215
98 -0,296
98 -0,296
-0,296
99 -1,561
99 -1,561
-1,561
100 0,246
100 0,246
0,246

ПРАКТИЧЕСКАЯ РАБОТА № 3

ПОСТРОЕНИЕ ARIMA-МОДЕЛЕЙ СЕЗОННЫХ ВРЕМЕННЫХ РЯДОВ

ЗАДАНИЕ

Для временного ряда своего варианта провести идентификацию, включающую следующие действия:

1) по виду графиков временного ряда, АКФ, ЧАКФ выявить имеющиеся особенности ряда (нестационарность, тренд и т.п.);
2) устранить выявленные особенности путем применения подходящих преобразований;
3) выделить и устранить сезонный компонент;
4) для преобразованного ряда по виду графиков временного ряда, АКФ, ЧАКФ определить тип модели, оценить параметры и их значимость;
5) проверить адекватность полученной модели, выполнив анализ остатков.

ВАРИАНТЫ

Вариант 1. Ежедневные данные фондового рынка

1 1705,4
5
1705,45
2 1667
,36
2 1667,3
6
1667,36
3 1706
,95
3 1706,9
5
1706,95
4 1708
,9
4 1708,9

1708,9
5 1683
,09
5 1683,0
9
1683,09
6 1707
,57
6 1707,5
7
1707,57
7 1739
,94
7 1739,9
4
1739,94
8 1734
,36
8 1734,3
6
1734,36
9 1727
,78
9 1727,7
8
1727,78
10 174
5,72
10 1745,
72
1745,72
11 174
1,04
11 1741,
04
1741,04
12 174
7,2

12 1747,2
1747,2
13 1777,13
13 1777,13
1777,13
14 1781,13
14 1781,13
1781,13
15 1781,07
15 1781,07
1781,07
16 1779,37
16 1779,37
1779,37
17 1780,69
17 1780,69
1780,69
18 1781,62
18 1781,62
1781,62
19 1810,2
19 1810,2
1810,2
20 1777,92
20 1777,92
1777,92
21 1732,38
21 1732,38
1732,38
22 1732,94
22 1732,94
1732,94
23 1698,73
23 1698,

73
1698,73
24 1698,73
24 1698,73
1698,73
25 1688,56
25 1688,56
1688,56
26 1664,81
26 1664,81
1664,81
27 1660,52
27 1660,52
1660,52
28 1699,39
28 1699,39
1699,39
29 1719,84
29 1719,84
1719,84
30 1681,65
30 1681,65
1681,65
31 1689,67
31 1689,67
1689,67
32 1718,15
32 1718,15
1718,15
33 1728,45
33 1728,45
1728,45
34 1683,82
34 1683,82
1683,82

35 1644,94
35 1644,94
1644,94
36 1688,15
36 1688,15
1688,15
37 1679,15
37 1679,15
1679,15
38 1694,75
38 1694,75
1694,75
39 16S9,27
39 16S9,27
16S9,27
40 1706,54
40 1706,54
1706,54
41 1728,71
41 1728,71
1728,71
42 1705,73
42 1705,73
1705,73
43 1715,84
43 1715,84
1715,84
44 1717,27
44 1717,27
1717,27
45 1705, 16
45 1705, 16
1705, 16
46 1689,78
46 1689,78
1689,78
47 1674,47
47 1674,47
1674,47
48 1672,34
48 1672,34
1672,34
49 1647,65
49 1647,65
1647,65
50 1611,57
50 1611,57
1611,57
51 1537,63
51 1537,63
1537,63
52 1597,54
52 1597,54
1597,54
53 1564,12
53 1564,12

1564,12
54 151
8,9
54 1518,
9
1518,9
55 151
9,49
55 1519,
49
1519,49
56 151
5,88
56 1515,
88
1515,88
57 153
3,85
57 1533,
85
1533,85
58 155
4,92
58 1554,
92
1554,92
59 156
3,09
59 1563,
09
1563,09
60 150
2,79
60 1502,
79
1502,79
6**1** 1
493,15
6**1** 149
3,15
61 1493,
15
1493,15
62 143
1,68
62 1431,
68
1431,68
63 140
4,71
63 1404,
71
1404,71
64 133
1,01
64 1331,

01
1331,01
65 132
2,69
65 1322,
69
1322,69
66 135
1,01
66 1351,
01
1351,01
67 124
5,26
67 1245,
26
1245,26
68 121
8,2
68 1218,
2
1218,2
69 128
0,13
69 1280,
13
1280,13
70 125
9,16
70 1259,
16
1259,16
71 123
3,26
71 1233,
26
1233,26
72 111
3,77
72 1113,
77
1113,77
73 117
9,83
73 1179,
83
1179,83
74 126
6,52
74 1266,
52
1266,52
75 130
2,42
75 1302,
42

1302,42
76 1335,42
76 1335,42
1335,42
77 1312,97
77 1312,97
1312,97
78 1289,24
78 1289,24
1289,24
79 1265,47
79 1265,47
1265,47
80 1249,47
80 1249,47
1249,47
81 1230,87
81 1230,87
1230,87
82 1227,72
82 1227,72
1227,72
83 1252,81
83 1252,81
1252,81
84 1236,12
84 1236,12
1236,12
85 1227,24
85 1227,24
1227,24
86 1233,06
86 1233,06
1233,06
87 123

1,54
87 1231,54
1231,54
88 1270,28
88 1270,28
1270,28
89 1281,71
89 1281,71
1281,71
90 1370,9
90 1370,9
1370,9
91 1377,51
91 1377,51
1377,51
92 1380,16
92 1380,16
1380,16
93 1420,05
93 1420,05
1420,05
94 1395,92
94 1395,92
1395,92
95 1384,29
95 1384,29
1384,29
96 1435,47
96 1435,47
1435,47
97 1436,73
97 1436,73
1436,73
98 1413,43
98 1413,43
1413,43
99 1433,82
99 1433,82
1433,82
100 1443,06
100 1443,06
1443,06
101 1386,31
101 1386,31
1386,31
102 1397,8
102 1397,8
1397,8
103 1395,39
103 1395,39
1395,39
104 1401,75
104 1401,75
1401,75
105 1416,18
105 1416,18

1416,18
106 14
28,68
106 142
8,68
1428,68
107 14
39,95
107 143
9,95
1439,95
108 14
51,07
108 145
1,07
1451,07
109 14
50,22
109 145
0,22
1450,22
110 14
31,1
110 143
1,1
1431,1
111 14
20,19
111 142
0,19
1420,19
112 13
99,82
112 139
9,82
1399,82
113 14
09,92
113 140
9,92
1409,92
114 14
23,82
114 142
3,82
1423,82
115 14
14,71
115 141
4,71
1414,71
116 14
03,86
116 140
3,86
1403,86

117 14
12,99
117 141
2,99
1412,99
118 14
13,19
118 141
3,19
1413,19
119 13
76,39
119 137
6,39
1376,39
120 13
81,18
120 138
1,18
1381,18
1**21**
1**21** 13
76,15
121 137
6,15
1376,15
122 13
76,15
122 137
6,15
1376,15
123 13
42,98
123 134
2,98
1342,98
124 13
34,75
124 133
4,75
1334,75
125 13
70,13
125 137
0,13
1370,13
126 13
53,11
126 135
3,11
1353,11
127 13
61,48
127 136
1,48
1361,48

128 1375,59
128 1375,59
1375,59
129 1406,11
129 1406,11
1406,11
130 1412,78
130 1412,78
1412,78
131 1415,31
131 1415,31
1415,31
132 1405,82
132 1405,82
1405,82
133 1376,65
133 1376,65
1376,65
134 1386,78
134 1386,78
1386,78
135 1385,72
135 1385,72
1385,72
136 1409,4
136 1409,4
1409,4
137 1415,79
137 1415,79
1415,79
138 1436,57
138 1436,57
1436,57
139 1445,32

139 1445,32
1445,32
140 1423,57
140 1423,57
1423,57
141 1396,82
141 1396,82
1396,82
142 1410
142 1410
1410
143 1410,46
143 1410,46
1410,46
144 1386,04
144 1386,04
1386,04
145 1398,57
145 1398,57
1398,57
146 1397,08
146 1397,08
1397,08
147 1330,46
147 1330,46
1330,46
148 1405,23
148 1405,23
1405,23
149 1398,3
149 1398,3
1398,3
150 1417,42
150 1417,42
1417,42
151 1416,02
151 1416,02
1416,02
152 1390,15
152 1390,15
1390,15
153 1394,15
153 1394,15
1394,15
154 1368,38
154 1368,38
1368,38
155 1565,7
155 1565,7
1565,7
156 1322,39
156 1322,39
1322,39
157 1292,47
157 1292,47
1292,47

158 12
96,4
158 129
6,4
1296,4
159 12
56,57
159 125
6,57
1256,57
160 12
70,04
160 127
0,04
1270,04
161 12
48,35
161 124
8,35
1248,35
162 12
30,62
162 123
0,62
1230,62
163 12
33,46
163 123
3,46
1233,46
164 12
57,94
164 125
7,94
1257,94
165 12
80,92
165 128
0,92
1280,92
166 12
77,01
166 127
7,01
1277,01
167 12
55,33
167 125
5,33
1255,33
168 12
67,12
168 126
7,12
1267,12
169 12

65
169 126
5
1265
170 12
71,6
170 127
1,6
1271,6
171 12
37,61
171 123
7,61
1237,61
172 12
31,77
172 123
1,77
1231,77
173 11
96,69
173 119
6,69
1196,69
174 11
40,58
174 114
0,58
1140,58
175 11
5S,92
175 115
S,92
115S,92
176 11
25,36
176 112
5,36
1125,36
177 11
29,83
177 112
9,83
1129,83
178 11
52,69
178 115
2,69
1152,69
179 11
79,78
179 117
9,78
1179,78
180 11
96,9

180 119
6,9
1196,9
1**81**
1**81** 11
61,93
181 116
1,93
1161,93
182 11
85,13
182 118
5,13
1185,13
183 11
91,76
183 119
1,76
1191,76
184 11
88,29
184 118
8,29
1188,29
185 12
07,29
185 120
7,29
1207,29
186 12
23,19
186 122
3,19
1223,19
187 12
21,18
187 122
1,18
1221,18
188 12
23,38
188 122
3,38
1223,38
189 12
10,18
189 121
0,18
1210,18
190 11
88,18
190 118
8,18
1188,18
191 11
74,72
191 117

4,72
1174,72
192 1167,16
192 1167,16
1167,16
193 1165,96
193 1165,96
1165,96
194 1174,3
194 1174,3
1174,3
195 1196,27
195 1196,27
1196,27
196 1200,12
196 1200,12
1200,12
197 1187,22
197 1187,22
1187,22
198 1206,86
198 1206,86
1206,86
199 1210,45
199 1210,45
1210,45
200 1203,81
200 1203,81
1203,81
201 1201,26
201 1201,26
1201,26
202 1182,48
202 1182,48
1182,48
203 1130,65
203 1130,65
1130,65
204 1193,77
204 1193,77
1193,77
205 1220,13
205 1220,13
1220,13
206 1235,84
206 1235,84
1235,84
207 1279,27
207 1279,27
1279,27
208 1306,08
208 1306,08
1306,08
209 1299,65
209 1299,65
1299,65

210 13
03,09
210 130
3,09
1303,09
211 13
23,36
211 132
3,36
1323,36
212 13
20,2
212 132
0,2
1320,2
213 13
19,36
213 131
9,36
1319,36
214 13
16,36
214 131
6,36
1316,36
215 13
16,2
215 131
6,2
1316,2
216 13
51,09
216 135
1,09
1351,09
217 13
68,97
217 136
8,97
1368,97
218 13
77,55
218 137
7,55
1377,55
219 13
81,26
219 138
1,26
1381,26
220 13
87,35
220 138
7,35
1387,35
221 13

83,85
221 138
3,85
1383,85
222 13
67,53
222 136
7,53
1367,53
223 13
61,5
223 136
1,5
1361,5
224 13
34,05
224 133
4,05
1334,05
225 13
60,1
225 136
0,1
1360,1
226 13
58,37
226 135
8,37
1358,37
227 13
59,51
227 135
9,51
1359,51
228 13
53,93
228 135
3,93
1353,93
229 13
83,32
229 138
3,32
1383,32
230 14
13,47
230 141
3,47
1413,47
231 13
95,29
231 139
5,29
1395,29
232 13
92,81

232 139
2,81
1392,81

Вариант 2. Ежеквартальное производство пива

1975-
1 36,14

36,14
1975-
2 44,6
1975-
2 44,6
44,6
1975-
3 44,15

1975-
3 44,15

44,15
1975-
4 35,72

1975-
4 35,72

35,72
1976-
1 36,19

1976-
1 36,19

36,19
1976-
2 44,63

1976-
2 44,63

44,63
1976-
3 46,95

1976-
3 46,95

46,95
1976-
4 36,9

1976-4 36,9
36,9
1977-1
39,66
1977-1
39,66
1977-1
39,66
39,66
1977-2
49,72
1977-2
49,72
49,72
1977-3
44,49
1977-3
44,49
44,49
1977-4
36,54
1977-4
36,54
36,54
1978-1
41,44
1978-1
41,44
41,44
1978-2
49,07
1978-2
49,07
49,07
1978-3
48,98
1978-3
48,98
48,98
1978-4
39,59
1978-4
39,59
39,59
1979-1
44,29
1979-1
44,29
1979-1
44,29
44,29
1979-2
50,09
1979-2

50,09
50,09
1979-3
48,42
1979-3
48,42
48,42
1979-4
41,39
1979-4
41,39
41,39
1980-1
46,11
1980-1
46,11
46,11
1980-2
53,44
1980-2
53,44
53,44
1980-3
53
1980-3
53
53
1980-4
42,52
1980-4
42,52
42,52
1**981-1**
44,61
1**981-1**
44,61
1981-1
44,61
44,61
1981-2
55,18
1981-2
55,18
55,18
1981-3
52,24
1981-3
52,24
52,24
1981-4
41,66
1981-4
41,66
41,66
1982-1
47,84

1982-1
47,84
47,84
1982-2
54,27
1982-2
54,27
54,27
1982-3
52,31
1982-3
52,31
52,31
1982-4
42,03
1982-4
42,03
42,03

Вариант 3. Смоделированный сезонный ряд

1 97.323

97.323
2 99.3
86
2 99.386

99.386
3 100.
581
3 100.58
1
100.581
4 95.6
71
4 95.671

95.671
5 102.
601
5 102.60
1
102.601
6 101.
751
6 101.75
1
101.751
7 103.
672

7 103.672
103.672
8 100.169
8 100.169
100.169
9 104.475
9 104.475
104.475
10 105.712
10 105.712
105.712
11 109.055
11 109.055
109.055
12 103.344
12 103.344
103.344
13 109.132
13 109.132
109.132
14 110.472
14 110.472
110.472
15 110.752
15 110.752
110.752
16 108.949
16 108.949
108.949
31 126.888
31 126.888
31 126.888
126.888
32 122

.581
32 122.5
81
122.581
33 124
.873
33 124.8
73
124.873
34 130
.316
34 130.3
16
130.316
35 129
.496
35 129.4
96
129.496
36 125
.721
36 125.7
21
125.721
37 130
.557
37 130.5
57
130.557
38 133
.554
38 133.5
54
133.554
39 134
.013
39 134.0
13
134.013
40 130
.221
40 130.2
21
130.221
41 133
.251
41 133.2
51
133.251
42 137
.640
42 137.6
40
137.640
43 136
.152

43 136.1
52
136.152
44 134
.049
44 134.0
49
134.049
45 136
.034
45 136.0
34
136.034
46 141
.636
46 141.6
36
141.636
61 1
56.866
61 156
.866
61 156.8
66
156.866
62 161
.653
62 161.6
53
161.653
63 158
.831
63 158.8
31
158.831
64 157
.871
64 157.8
71
157.871
65 159
.769
65 159.7
69
159.769
66 166
.352
66 166.3
52
166.352
67 163
.872
67 163.8
72
163.872
68 162
.214

68 162.214
162.214
69 165.239
69 165.239
165.239
70 170.421
70 170.421
170.421
71 169.188
71 169.188
169.188
72 167.274
72 167.274
167.274
73 170.048
73 170.048
170.048
74 175.631
74 175.631
175.631
75 174.562
75 174.562
174.562
76 172.450
76 172.450
172.450
91 192.726

91 192.726
91 192.726
192.726
92 189.861
92 189.861
189.861
93 191.439
93 191.439
191.439
94 200.030
94 200.030
200.030
95 196.391
95 196.391
196.391
96 196.406
96 196.406
196.406
97 198.099
97 198.099
198.099
98 204.684
98 204.684
204.684
99 203.344
99 203.344
203.344
100 197.990

100 197.990
197.990
101 202.407
101 202.407
202.407
102 210.073
102 210.073
210.073
103 207.403
103 207.403
207.403
104 203.713
104 203.713
203.713
105 208.789
105 208.789
208.789
106 213.626
106 213.626
213.626
121
121 225.794
121 225.794
225.794
122 231.908
122 231.908
231.908
123 228.497
123 228.497
228.497
124 225.202
124 225.202
225.202
125 229.665

125 229.665
229.665
126 236.350
126 236.350
236.350
127 233.512
127 233.512
233.512
128 230.428
128 230.428
230.428
129 233.996
129 233.996
233.996
130 240.955
130 240.955
240.955
131 237.421
131 237.421
237.421
132 234.174
132 234.174
234.174
133 236.984
133 236.984
236.984
134 245.057
134 245.057
245.057
135 241.983
135 241.983
241.983
136 237.094
136 237.

094
237.094

17 111.7
13
111.713
18 113
.635
18 113.6
35
113.635
19 115
.535
19 115.5
35
115.535
20 111
.439
20 111.4
39
111.439
21 115
.798
21 115.7
98
115.798
22 119
.851
22 119.8
51
119.851
23 120
.605
23 120.6
05
120.605
24 115
.752
24 115.7
52
115.752
25 119
.380
25 119.3
80
119.380
26 122
.423
26 122.4
23
122.423
27 122

.320
27 122.3
20
122.320
28 118
.610
28 118.6
10
118.610
29 121
.799
29 121.7
99
121.799
30 127
.540
30 127.5
40
127.540
47 1
40.264
47 140
.264
47 140.2
64
140.264
48 137
.761
48 137.7
61
137.761
49 140
.608
49 140.6
08
140.608
50 145
.765
50 145.7
65
145.765
51 145
.491
51 145.4
91
145.491
52 143
.401
52 143.4
01
143.401
53 145
.873
53 145.8
73
145.873

54 151
.227
54 151.2
27
151.227
55 149
.660
55 149.6
60
149.660
56 148
.268
56 148.2
68
148.268
57 152
.012
57 152.0
12
152.012
58 156
.293
58 156.2
93
156.293
59 156
.306
59 156.3
06
156.306
60 151
.636
60 151.6
36
151.636
77 1
74.456
77 174
.456
77 174.4
56
174.456
78 181
.318
78 181.3
18
181.318
79 179
.960
79 179.9
60
179.960
80 177
.068
80 177.0
68

177.068
81 179
.141
81 179.1
41
179.141
82 183
.945
82 183.9
45
183.945
83 183
.056
83 183.0
56
183.056
84 181
.098
84 181.0
98
181.098
85 183
.128
85 183.1
28
183.128
86 191
.285
86 191.2
85
191.285
87 188
.676
87 188.6
76
188.676
88 185
.391
88 185.3
91
185.391
89 187
.751
89 187.7
51
187.751
90 194
.459
90 194.4
59
194.459
107
107 21
1.808
107 211.
808
211.808

108 209.169
108 209.169
209.169
109 211.793
109 211.793
211.793
110 219.515
110 219.515
219.515
111 216.688
111 216.688
216.688
112 213.734
112 213.734
213.734
113 216.468
113 216.468
216.468
114 223.600
114 223.600
223.600
115 219.971
115 219.971
219.971
116 216.864
116 216.864
216.864
117 221.199
117 221.199
221.199
118 227.439
118 227.439
227.439
119 224.872
119 224.872
224.872
120 222.034
120 222.034
222.034
137 242.367

137 242.367
137 242.367
242.367
138 249.358
138 249.358
249.358
139 245.153
139 245.153
245.153
140 240.533
140 240.533
240.533
141 245.504
141 245.504
245.504

142 251.616
142 251.616
251.616
143 250.315
143 250.315
250.315
144 244.670
144 244.670
244.670
145 249.887
145 249.887
249.887
146 256.824
146 256.824
256.824
147 254.345
147 254.345
254.345
148 251.052
148 251.052
251.052
149 253.507
149 253.507
253.507
150 260.501
150 260.501
260.501

Вариант 4. Объемы продаж ликера

Jan/1970
580

580
Feb/19
70 514
Feb/1970
514
514
Mar/19
70 555
Mar/197
0 555
555
Apr/19
70 563
Apr/1970
563
563
May/19
70 627
May/197
0 627
627
Jun/19
70 596
Jun/1970
596
596
Jul/197
0 632
Jul/1970
632
632
Aug/19
70 639
Aug/1970
639
639
Sep/197
0 577
Sep/1970
577
577
Oct/197
0 611
Oct/1970
611
611
Nov/19
70 639
Nov/1970
639
639
Dec/19
70 875
Dec/1970
875
875

Jan/197
1 650
Jan/1971
650
650
Feb/197
1 594
Feb/1971
594
594
Mar/19
71 650
Mar/1971
650
650
Apr/19
71 668
Apr/1971
668
668
May/19
71 712
May/1971
712
712
Jun/197
1 731
Jun/1971
731
731
Jul/197
1 779
Jul/1971
779
779
Aug/19
71 712
Aug/1971
712
712
Sep/197
1 708
Sep/1971
708
708
Jan/1
973 734

Jan/197
3 734
Jan/1973
734
734
Feb/197
3 707
Feb/1973

707
707
Mar/19
73 785
Mar/197
3 785
785
Apr/19
73 762
Apr/1973
762
762
May/19
73 838
May/197
3 838
838
Jun/19
73 876
Jun/1973
876
876
Jul/197
3 878
Jul/1973
878
878
Aug/19
73 871
Aug/1973
871
871
Sep/197
3 807
Sep/1973
807
807
Oct/197
3 834
Oct/1973
834
834
Nov/19
73 877
Nov/1973
877
877
Dec/19
73 1236

Dec/1973
1236
1236
Jan/197
4 789
Jan/1974

789
789
Feb/1974 744
Feb/1974
744
744
Mar/1974 827
Mar/1974
827
827
Apr/1974 831
Apr/1974
831
831
May/1974 895
May/1974
895
895
Jun/1974 889
Jun/1974
889
889
Jul/1974 955
Jul/1974
955
955
Aug/1974 983
Aug/1974
983
983
Sep/1974 976
Sep/1974
976
976
Jan/1976 908

Jan/1976 908
Jan/1976
908
908
Feb/1976 849
Feb/1976
849
849
Mar/19

76 916
Mar/197
6 916
916
Apr/19
76 958
Apr/1976
958
958
May/19
76 1008

May/197
6 1008
1008
Jun/19
76 1033

Jun/1976
1033
1033
Jul/197
6 1129
Jul/1976
1129
1129
Aug/19
76 1019

Aug/1976
1019
1019
Sep/197
6 984
Sep/1976
984
984
Oct/197
6 1045
Oct/1976
1045
1045
Nov/19
76 1049

Nov/1976
1049
1049
Dec/19
76 1455

Dec/1976
1455
1455
Jan/197
7 910
Jan/1977

910
910
Feb/1977 927
Feb/1977
927
927
Mar/1977 981
Mar/1977
981
981
Apr/1977 1011

Apr/1977
1011
1011
May/1977 1041

May/1977
1041
1041
Jun/1977 1080
Jun/1977
1080
1080
Jul/1977 1138
Jul/1977
1138
1138
Aug/1977 1072

Aug/1977
1072
1072
Sep/1977 1033
Sep/1977
1033
1033
Jan/1979 1071
1
Jan/1979 1071
Jan/1979
1071
1071
Feb/1979 1044
Feb/1979
1044

1044
Mar/1979 1158

Mar/1979 1158
1158
Apr/1979 1122

Apr/1979 1122
1122
May/1979 1209

May/1979 1209
1209
Jun/1979 1334

Jun/1979 1334
1334
Jul/1979 1360
Jul/1979 1360
1360
Aug/1979 1368

Aug/1979 1368
1368
Sep/1979 1297
Sep/1979 1297
1297
Oct/1979 1283
Oct/1979 1283
1283
Nov/1979 1375

Nov/1979 1375
1375
Dec/1979 1973

Dec/1979 1973

1973
Jan/1980 1294
Jan/1980
1294
1294
Feb/1980 1258
Feb/1980
1258
1258
Mar/1980 1301

Mar/1980
1301
1301
Apr/1980 1297

Apr/1980
1297
1297
May/1980 1425

May/1980
1425
1425
Jun/1980 1378
Jun/1980
1378
1378
Jul/1980 1429
Jul/1980
1429
1429
Aug/1980 1452

Aug/1980
1452
1452
Sep/1980 1305
Sep/1980
1305
1305

Oct/1971
738

738
Nov/19
71 758
Nov/1971
758
758
Dec/197
1 1073
Dec/1971
1073
1073
Jan/197
2 669
Jan/1972
669
669
Feb/197
2 652
Feb/1972
652
652
Mar/19
72 743
Mar/1972
743
743
Apr/19
72 709
Apr/1972
709
709
May/19
72 751
May/1972
751
751
Jun/197
2 774
Jun/1972
774
774
Jul/197
2 803
Jul/1972
803
803
Aug/19
72 760
Aug/1972
760
760
Sep/197
2 749
Sep/1972
749
749
Oct/197

2 757
Oct/1972 757
757
Nov/1972 779

Nov/1972 779
779
Dec/1972 106
6
Dec/1972 1066

1066
Oct/1974 92
9
Oct/1974 929

Oct/1974 929
929
Nov/1974 989

Nov/1974 989
989
Dec/1974 129
3
Dec/1974 1293

1293
Jan/1975 860

Jan/1975 860
860
Feb/1975 799

Feb/1975 799
799
Mar/1975 899

Mar/1975 899
899
Apr/1975 866

Apr/1975 866
866
May/1975 101
6
May/1975 1016

1016
Jun/1975 978

Jun/1975 978
978
Jul/1975 1042

Jul/1975 1042
1042

Aug/1975 1026

Aug/1975 1026
1026
Sep/1975 944
Sep/1975 944
944
Oct/1975 1002
Oct/1975 1002
1002
Nov/1975 1009

Nov/1975 1009
1009
Dec/1975 1365

Dec/1975 1365
1365
Oct/1977 1072
Oct/1977 1072
1072
Nov/1977 1111

Nov/1977 1111
1111
Dec/1977 1575

Dec/1977 1575
1575
Jan/1978 950
Jan/1978 950
950
Feb/1978 932
Feb/1978 932

932
Mar/1978 1049

Mar/1978 1049
1049
Apr/1978 1021

Apr/1978
1021
1021
May/1978 1097

May/1978 1097
1097
Jun/1978 1151

Jun/1978
1151
1151
Jul/1978 1194
Jul/1978
1194
1194
Aug/1978 1174

Aug/1978
1174
1174
Sep/1978 1160
Sep/1978
1160
1160
Oct/1978 1135
Oct/1978
1135
1135
Nov/1978 1209

Nov/1978
1209
1209
Dec/1978 1643

Dec/1978
1643

1643
Oct/1980 1377
7
Oct/1980 1377
Oct/1980
1377
1377
Nov/1980 1439

Nov/1980
1439
1439
Dec/1980 1989
Dec/1980
1989
1989

Вариант 5. Потребление нефти, млрд долларов

I 1967 16,6
16,6
II - 1967 16,9
II 1967 16,9
16,9
III - 1967 17,1
III 1967 17,1
17,1
IV - 1967 17,5
IV 1967 17,5
17,5
I - 1968 18,1
I

1968 18,
1
18,1
II -
1968 18,
3
II
1968 18,
3
18,3
III -
1968 19,
0
III
1968 19,
0
19,0
IV -
1968 19,
1
IV
1968 19,
1
19,1
I -
1969 19,
8
I
1969 19,
8
19,8
II -
1969 20,
8
II
1969 20,
8
20,8
III -
1969 20,
9
III
1969 20,
9
20,9
IV -
1969 21,
4
IV
1969 21,
4
21,4
I -
1970 21,
8
I
1970 21,
8

I - 1970 21,8
21,8
II - 1970 22,3
II - 1970 22,3
22,3
III - 1970 22,5
III - 1970 22,5
22,5
IV - 1970 23,1
IV - 1970 23,1
23,1
I - 1971 23,5
I - 1971 23,5
23,5
II - 1971 23,4
II - 1971 23,4
23,4
III - 1971 24,0
III - 1971 24,0
24,0
IV - 1971 24,6
IV - 1971 24,6
24,6
I - 1972 24,8
I - 1972 24,

8
24,8
II - 1972 24,7
II - 1972 24,7
24,7
III - 1972 25,5
III - 1972 25,5
25,5
IV - 1972 26,6
IV - 1972 26,6
26,6
I - **1973** 27,6
I - **1973** 27,6
I - 1973 27,6
27,6
II - 1973 27,9
II - 1973 27,9
27,9
III - 1973 28,4
III - 1973 28,4
28,4
IV - 1973 30,6
IV - 1973 30,6
30,6
I - 1974 32,4
I - 1974 32,

4
32,4
II -
1974 36,9
II - 1974 36,9
36,9
III -
1974 38,2
III - 1974 38,2
38,2
IV -
1974 38,9
IV - 1974 38,9
38,9
I -1975 38,6
I -1975 38,6
38,6
II -
1975 39,3
II - 1975 39,3
39,3
III -
1975 41,3
III - 1975 41,3
41,3
IV -
1975 42,4
IV - 1975 42,4
42,4
I -
1976 42,8
I **-1976** 42,8
I -1976 42,8
42,8
II -
1976 43,0
II - 1976 43,0
43,0
III -
1976 44,1
III - 1976 44,1
44,1
IV -
1976 45,9
IV - 1976 45,9
45,9
I -1977 47,2
I -1977 47,2
47,2
II -
1977 48,6
II - 1977 48,6
48,6
III -
1977 48,3
III - 1977 48,3
48,3

IV - 1977 48,5

IV 1977 48,5

48,5

I - 1978 49,3

I 1978 49,3

49,3

II - 1978 49,9

II 1978 49,9

49,9

III - 1978 51,5

III 1978 51,5

51,5

IV - 1978 54,3

IV 1978 54,3

54,3

I - **1979** 57,7

I **1979** 57,7

I 1979 57,7

57,7

II - 1979 62,2

II 1979 62,2

62,2

III - 1979 70,5

III 1979 70,5

70,5
IV - 1979 75,9
IV - 1979 75,9
75,9
I - 1980 80,7
I - 1980 80,7
80,7
II - 1980 84,4
II - 1980 84,4
84,4
III - 1980 85,1
III - 1980 85,1
85,1
IV - 1980 88,9
IV - 1980 88,9
88,9
I - 1981 94,1
I - 1981 94,1
94,1
II - 1981 95,1
II - 1981 95,1
95,1
III - 1981 94,6
III - 1981 94,6
94,6
IV -

1981	94,7
IV - 1981	94,7
94,7	
I - 1982	93,4
I - 1982	93,4
93,4	
II - 1982	88,6
II - 1982	88,6
88,6	
III - 1982	89,9
III - 1982	89,9
89,9	
IV - 1982	89,6
IV - 1982	89,6
89,6	

Вариант 6. Количество безработных молодых людей

Jan/1971	707	**Jan/1974**	714	**Jan/1977**	896	**Jan/1980**	840
Feb/1971	655	**Feb/1974**	715	**Feb/1977**	936	**Feb/1980**	847
Mar/1971	638	**Mar/1974**	672	**Mar/1977**	902	**Mar/1980**	774
Apr/1971	574	**Apr/1974**	588	**Apr/1977**	765	**Apr/1980**	720
May/1971	552	**May/1974**	567	**May/1977**	735	**May/1980**	848
Jun/1971	980	**Jun/1974**	1057	**Jun/1977**	1234	**Jun/1980**	1240
Jul/1971	926	**Jul/1974**	949	**Jul/1977**	1052	**Jul/1980**	1168
Aug/1971	680	**Aug/1974**	683	**Aug/1977**	868	**Aug/1980**	936

Sep/1971	597	**Sep/1974**	771	**Sep/1977**	798	**Sep/1980**	853
Oct/1971	637	**Oct/1974**	708	**Oct/1977**	751	**Oct/1980**	910
Nov/1971	660	**Nov/1974**	824	**Nov/1977**	820	**Nov/1980**	953
Dec/1971	704	**Dec/1974**	835	**Dec/1977**	725	**Dec/1980**	874
Jan/1972	758	**Jan/1975**	980	**Jan/1978**	821	**Jan/1981**	1026
Feb/1972	835	**Feb/1975**	969	**Feb/1978**	895	**Feb/1981**	1030
Mar/1972	747	**Mar/1975**	931	**Mar/1978**	851	**Mar/1981**	946
Apr/1972	617	**Apr/1975**	892	**Apr/1978**	734	**Apr/1981**	860
May/1972	554	**May/1975**	828	**May/1978**	636	**May/1981**	856
Jun/1972	929	**Jun/1975**	1350	**Jun/1978**	994	**Jun/1981**	1190
Jul/1972	815	**Jul/1975**	1218	**Jul/1978**	990	**Jul/1981**	1038
Aug/1972	702	**Aug/1975**	977	**Aug/1978**	750	**Aug/1981**	883
Sep/1972	640	**Sep/1975**	863	**Sep/1978**	727	**Sep/1981**	843
Oct/1972	588	**Oct/1975**	838	**Oct/1978**	754	**Oct/1981**	857
Nov/1972	669	**Nov/1975**	866	**Nov/1978**	792	**Nov/1981**	1016
Dec/1972	675	**Dec/1975**	877	**Dec/1978**	817	**Dec/1981**	1003
Jan/1973	610	**Jan/1976**	1007	**Jan/1979**	856		
Feb/1973	651	**Feb/1976**	951	**Feb/1979**	886		
Mar/1973	605	**Mar/1976**	906	**Mar/1979**	833		
Apr/1973	592	**Apr/1976**	911	**Apr/1979**	733		
May/1973	527	**May/1976**	812	**May/1979**	675		
Jun/1973	898	**Jun/1976**	1172	**Jun/1979**	1004		
Jul/1973	839	**Jul/1976**	1101	**Jul/1979**	956		
Aug/1973	614	**Aug/1976**	900	**Aug/1979**	777		
Sep/1973	594	**Sep/1976**	841	**Sep/1979**	761		
Oct/1973	576	**Oct/1976**	853	**Oct/1979**	709		
Nov/1973	672	**Nov/1976**	922	**Nov/1979**	*777*		
Dec/1973	651	**Dec/1976**	886	**Dec/1979**	771		

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Андерсон Т.* Статистический анализ временных рядов: пер. с англ. – М: Мир, 1976. – 755 с.

2. *Берк К.Н., Кэйри П.* Анализ данных с помощью Microsoft Excel: пер. с англ. – М.: Вильямс, 2005. – 560 с.

3. *Бокс Дж., Дженкинс Г.* Анализ временных рядов. Прогноз и управление: пер. с англ. – М.: Мир, 1974. – Вып. 1. – 406 с.

4. *Боровиков В.П., Боровиков И.П.* Statistica – статистический анализ и обработка данных в среде Windows. – М.: Филинъ, 1998. – 608 с.

5. *Боровиков В.П., Ивченко Г.И.* Прогнозирование в системе STATISTICA в среде Windows. – М.: Финансы и статистика, 2006. – 368 с.

6. *Дубнов П.Ю.* Обработка статистической информации с помощью SPSS. – М.: АСТ; НТ Пресс, 2004. – 221 с.

7. *Канторович Г.Г.* Анализ временных рядов//Экономический журнал Высшей школы экономики. – Т. 6. – № 2.

8. *Лукашин Ю.П.* Адаптивные методы краткосрочного прогнозирования временных рядов: учеб. пособие. – М.: Финансы и статистика, 2003. – 416 с.

9. *Магнус Я.Р., Катышев П.К., Пересецкий А.А.* Эконометрика. Начальный курс: учебник. – М.: Дело, 2005. – 504 с.

10. *Chang I., Tiao G., Chen C.* Estimation of time series parameters in the presence of ontliers // Technometric, 1988, v. 30, N 2, pp. 193–204.

11. *Durbin J.* The fitting of time series models // Review of the Institute of International Statistics, 1960, v. 28, pp. 233–244.

12. *Gardner E. S., Jr.* Exponential smoothing: The state of the art // Journal of Forecasting. – 1985. – Vol. 4. – P. 1–28.

13. *Hamilton, J. D.* Time Series Analysis. – Princeton University Press, 1994.

14. *Wei W.S.* Time Series Analysis (Univariate and multivariate methods). – Addison-Wesley Publishing Company, 1990. – 504 p.

15. *Yule G.V.* On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers // Philos. Trans. Roy. Soc. London, Ser. A, 1927, v. 226, pp. 267–298.

Авдеенко Татьяна Владимировна

КОМПЬЮТЕРНЫЕ МЕТОДЫ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ И ПРОГНОЗИРОВАНИЯ

Учебное пособие

Редактор *Н.А. Лукашова*
Выпускающий редактор *И.П. Бровановa*
Корректор *И.Е. Семенова*
Дизайн обложки *А.В. Ладыжская*
Компьютерная верстка *Н.В. Гаврилова*

Подписано в печать 26.11.2008. Формат 60 × 84 1/16. Бумага офсетная. Тираж 150 экз.
Уч.-изд. л. 14,8. Печ. л. 16,0. Изд. № 214. Заказ № 1530. Цена договорная

Отпечатано в типографии
Новосибирского государственного технического университета
630092, г. Новосибирск, пр. К. Маркса, 20