



Ю.Е. Воскобойников

# ЭКОНОМЕТРИКА В EXCEL

## Часть 2 Анализ временных рядов

	A	B	C	D
1	Год	Спрос		#И/Д
2	1	213		#И/Д
3	2	171		#И/Д
4	3	291	225,00	#И/Д
5	4	309	257,00	#И/Д
6	5	317	305,67	48,950
7	6	362	329,33	36,053
8	7	351	343,33	20,448
9	8	361	358,00	19,450

	A	B	C	D
1	Год	Спрос		#И/Д
2	1	213		#И/Д
3	2	171		#И/Д
4	3	291	225,00	#И/Д
5	4	309	257,00	#И/Д
6	5	317	305,67	48,950
7	6	362	329,33	36,053
8	7	351	343,33	20,448
9	8	361	358,00	19,450
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				

**УЧЕБНОЕ ПОСОБИЕ**

**Скользящее среднее**

**Экспоненциальное сглаживание**

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ РОССИЙСКОЙ ФЕДЕРАЦИИ  
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ АРХИТЕКТУРНО-СТРОИТЕЛЬНЫЙ  
УНИВЕРСИТЕТ (СИБСТРИН)

**Ю.Е. Воскобойников**

# **ЭКОНОМЕТРИКА В EXCEL**

Часть 2

## **АНАЛИЗ ВРЕМЕННЫХ РЯДОВ**

**УЧЕБНОЕ ПОСОБИЕ**

НОВОСИБИРСК 2008

УДК 330.43(075.8)  
ББК 65.в6.я73  
В762

***Воскобойников Ю. Е.***

Эконометрика в Excel : учеб. пособие. Ч. 2. Анализ временных рядов / Ю. Е. Воскобойников ; Новосиб. гос. архитектур.-строит. ун-т. – Новосибирск : НГАСУ (Сибстрин), 2008. – 152 с.

**ISBN 978-5-7795-0366-2**

Учебное пособие содержит основные теоретические положения, необходимые для решения задач анализа временных рядов. Приводятся необходимые расчетные соотношения. Большое внимание уделяется реализации этих соотношений в табличном процессоре Excel. Пособие содержит большое количество примеров и копий фрагментов документов Excel, которые позволят студентам не только лучше понять и усвоить учебный материал, но и эффективно использовать Excel при выполнении дипломной и курсовых работ.

Учебное пособие рекомендуется студентам экономических специальностей вузов, а также будет полезно аспирантам и преподавателям по прикладной экономике и финансам.

Печатается по решению издательско-библиотечного совета НГАСУ (Сибстрин)

Рецензенты:

- В.З. Баликоев, д-р экон. наук, профессор,  
директор Института экономики и менеджмента  
НГАСУ (Сибстрин);
- А.С. Овсянников, д-р экон. наук, профессор, завкафедрой  
экономики труда и хозяйственной деятельности НГАВТ

**ISBN 978-5-7795-0366-2**

©

Воскобойников Ю.Е., 2008

©

Новосибирский государственный  
архитектурно-строительный университет  
(Сибстрин), 2008

## Оглавление

<b>ВВЕДЕНИЕ</b> .....	9
<b>ГЛАВА 1. ВРЕМЕННЫЕ РЯДЫ И ИХ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ</b> .....	12
1.1. Временной ряд и его модели .....	12
1.2. Числовые характеристики временного ряда.....	16
1.3. Проверка статистических гипотез о свойствах временного ряда.....	23
<b>ГЛАВА 2. ВЫДЕЛЕНИЕ НЕСЛУЧАЙНОЙ СОСТАВЛЯЮЩЕЙ ВРЕМЕННОГО РЯДА</b> .....	39
2.1. Выделение трендовой составляющей временного ряда .....	39
2.2. Выделение трендовой составляющей с помощью табличного процессора Excel .....	56
2.3. Выделение тригонометрической составляющей временного ряда.....	71
2.4. Проверка адекватности и качества построенной модели временного ряда.....	85
2.5. Прогнозирование трендовой составляющей временного ряда.....	90
<b>ГЛАВА 3. ВРЕМЕННЫЕ РЯДЫ С КОРРЕЛИРОВАННЫМИ ВОЗМУЩЕНИЯМИ. ОБОБЩЕННЫЙ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ</b> .....	104
3.1. Временные ряды с коррелированными возмущениями.....	104
3.2. Обобщенный метод наименьших квадратов .....	113
3.3. Выделения тренда временного ряда на основе обобщенного метода наименьших квадратов .....	116

<b>ГЛАВА 4. АВТОРЕГРЕССИОННЫЕ МОДЕЛИ ВРЕМЕННОГО РЯДА</b> .....	127
4.1. Определение авторегрессионной модели .....	127
4.2. Оценивание коэффициентов авторегрессионной модели.....	129
4.3. Оценивание коэффициентов авторегрессионной модели стационарного временного ряда .....	135
4.4. Тест на наличие автокорреляции .....	143
4.5. Определение порядка авторегрессионной модели временного ряда.....	146
<b>ЗАКЛЮЧЕНИЕ</b> .....	154
<b>БИБЛИОГРАФИЧЕСКИЙ СПИСОК</b> .....	155

## ВВЕДЕНИЕ

В последнее время специалисты, обладающие знаниями и навыками проведения прикладного экономического анализа с использованием современных математических и программных средств, пользуются спросом на рынке труда. Одной из центральных дисциплин в подготовке таких специалистов является «Эконометрика». Дословный перевод этого слова означает *экономические измерения*, но определение дисциплины «Эконометрика» гораздо шире этого перевода. Ниже приводятся два определения известных ученых, позволяющие получить представления о различном толковании эконометрики.

*Эконометрика – это раздел экономики, занимающийся разработкой и применением статистических методов для измерений взаимосвязей между экономическими переменными (С. Фишер).*

*Эконометрика – это самостоятельная научная дисциплина, объединяющая совокупность теоретических результатов, приемов, методов и моделей, предназначенных для того, чтобы на базе*

- экономической теории;*
- экономической статистики;*
- математико-статистического инструментария*

*придать конкретное количественное выражение общим качественным закономерностям, обусловленным экономической теорией (С.А. Айвазян).*

*Из этих определений можно сформулировать основную цель эконометрики – модельное описание конкретных количественных взаимосвязей, обусловленных общими качественными закономерностями, изучаемыми в экономической теории.*

Составленное модельное описание называется *эконометрической моделью*. В учебном пособии [5] было подробно рассмотрено построение регрессионных эконометрических моделей в Excel. В данном учебном пособии рассматривается построение эконометрических моделей для временных рядов (анализ временных рядов). Области применения таких моделей являются:

- прогноз экономических и социально-экономических показателей, характеризующих состояние и развитие анализируемой системы;

- имитация различных возможных сценариев социально-экономического развития анализируемой системы.

В качестве анализируемой системы могут выступать страна в целом, регионы, отрасли и корпорации, а также предприятия и фирмы.

Построение эконометрических моделей временных рядов обуславливает (особенно при большом объеме исходных данных) существенный объем вычислений. На этом этапе многие исследователи сталкиваются с проблемами численной реализации необходимого вычислительного алгоритма и графической интерпретации результатов решения. Этим вопросам в учебной литературе уделяется крайне мало внимания, что затрудняет использование современных алгоритмов решения задач анализа временных рядов на практике.

Поэтому основной целью данного пособия является *изложение численных методик решения основных задач анализа временных рядов в вычислительной среде табличного процессора Excel 2003.*

Для каждой из рассматриваемых задач анализа временных рядов эконометрики приводятся необходимый теоретический материал, математическая запись алгоритма решения (т.е. формулы или расчетные соотношения), а затем даются фрагменты документов Excel 2003, реализующих алгоритмы решения задачи.

При этом алгоритм решения может быть реализован путем программирования арифметических или логических выражений в ячейках электронной таблицы или путем обращения к *стандартным* функциям или модулям Excel 2003. Поэтому предполагается, что читатель знаком с адресацией ячеек (относительной, абсолютной и смешанной), арифметическими операциями и программированием простейших выражений в ячейках Excel.

Данное учебное пособие, хотя и содержит необходимый теоретический материал, но *не заменяет учебник по эконометрике, а является своеобразным справочником по численному решению за-*

дач эконометрики в Excel 2003. Учебное пособие можно также рассматривать как дополнение к основному учебнику по эконометрике, которое будет полезным при выполнении курсовых и дипломных работ, а также при самостоятельном решении практических задач эконометрики.

Кроме решения задач учебное пособие содержит набор лабораторных и контрольных работ по каждой теме, ориентированных на заочную и дистанционную формы обучения.

Предполагается, что читатель знаком с основными понятиями теории вероятностей и математической статистики. При необходимости он может обратиться к литературе [1–4]. Кроме этого, он знаком с основными методами, используемыми при построении и проверке регрессионных моделей [5–12].

Структура и содержание учебного пособия и рассматриваемых задач соответствуют требованиям государственного образовательного стандарта высшего профессионального образования для специальностей направления «Экономика и менеджмент».

## ГЛАВА 1. ВРЕМЕННЫЕ РЯДЫ И ИХ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ

### 1.1. Временной ряд и его модели

Широкий круг социально-экономических, технических и естественно-научных процессов часто представляются набором упорядоченных во времени случайных величин  $Y(\tau_1), Y(\tau_2), \dots, Y(\tau_n)$ , где  $\tau_i < \tau_{i+1}$ . Такая последовательность величин называется временным рядом или дискретным временным процессом. Набор наблюдений  $\{y_i\}$ ,  $i = 1, 2, \dots, n$  над случайной величинами  $\{Y(\tau_i)\}$  в моменты времени  $\tau_1, \tau_2, \dots, \tau_n$  называется *временной выборкой*. Заметим, что иногда в литературе понятие *временного ряда* подменяется понятием *временной выборки*. Принципиальная разница между этими двумя понятиями заключается в том, что временной ряд это *набор случайных величин*, а временная выборка – набор измеренных значений, которые являются конкретной реализацией случайных величин  $\{Y(\tau_i)\}$  в  $n$  экспериментах и по этой причине значения  $y_i = Y(\tau_i)$ ,  $i = 1, 2, \dots, n$  временной выборки уже *не являются случайными величинами*.

**Замечание 1.1.1.** Временной ряд  $\{Y(\tau_i)\}$  можно интерпретировать как наблюдения над непрерывным случайным процессом (случайной функцией)  $Y(\tau)$  в моменты времени  $\tau = \tau_i$ . Поэтому в дальнейшем наряду с временным рядом  $\{Y(\tau_i)\}$  будет рассматриваться и непрерывный случайный процесс  $Y(\tau)$  ♦

Изменения величины  $Y(\tau_i)$  во времени в реальной жизни обычно происходят под воздействием многочисленных причин, факторов. Поэтому в отношении временного ряда выдвигается предположение, что совокупное влияние этих факторов формирует некоторую закономерность в развитии временного ряда, что дает основание применить для описания динамики  $Y(\tau_i)$  эконометрическую модель из класса моделей временных рядов.

Модели временных рядов активно применяются в исследованиях значительного числа реальных процессов различной природы. Например, в исследованиях динамики пассажиропотоков, складских запасов, спроса на различные виды продукции, финансовых рынков, в анализе динамики финансовых показателей, прогнозировании цен на различные товары и т.д.

Одним из основных классов эконометрических моделей временных рядов является *класс аддитивных моделей* вида

$$Y(\tau_i) = q(\tau_i) + \varepsilon(\tau_i), \quad i = 1, 2, \dots, n, \quad (1.1.1)$$

где неслучайная (детерминированная) составляющая  $q(\tau_i)$  может включать одну или несколько из следующих компонент: трендовую  $t(\tau_i)$ , сезонную  $s(\tau_i)$  и периодическую  $p(\tau_i)$ . Часто компоненты  $s(\tau_i)$ ,  $p(\tau_i)$  называют *тригонометрическими составляющими* временного ряда.

*Тренд*, или *тенденция*  $t(\tau_i)$ , представляет собой устойчивую закономерность, наблюдаемую в течение длительного периода времени. Обычно тренд (тенденция) описывается с помощью той или иной неслучайной функции  $t(\tau)$  (аргументом которой является время), как правило, достаточно «гладкой» (часто монотонной).

*Сезонная компонента*  $s(\tau_i)$  связана с наличием факторов, действующих с заранее известной периодичностью. Это регулярные колебания, которые носят периодический или близкий к нему характер и заканчиваются в течение года. Типичные примеры сезонного эффекта: изменение загруженности автотрассы по временам года, пик продаж товаров для школьников в конце августа – начале сентября. Спрос на пластические операции сезонный: в осенне-зимний период обращений больше. Типичным примером являются сильные колебания объема товарно-материальных запасов в сезонных отраслях. Сезонная компонента со временем может меняться, либо иметь плавающий характер.

*Периодическая (циклическая) компонента*  $p(\tau_i)$  – неслучайная функция, описывающая длительные периоды (более одного

года) относительного подъема и спада и состоящая из циклов переменной длительности и амплитуды. Примерами периодической компоненты являются волны Кондратьева, демографические «ямы» и т.п. Подобная компонента весьма характерна для рядов макроэкономических показателей. Здесь циклические изменения обусловлены взаимодействием спроса и предложения, а также наложением таких факторов, как истощение ресурсов, погодные условия, изменения в налоговой политике и т.п. Отметим, что циклическую компоненту крайне трудно идентифицировать формальными методами, исходя только из данных изучаемого ряда.

*Случайная компонента*  $\varepsilon(\tau_i)$  – это составная часть временного ряда, оставшаяся после выделения систематических компонент. Она отражает воздействие многочисленных факторов случайного характера и представляет собой случайную, нерегулярную компоненту. Она является обязательной составной частью любого временного ряда в экономике, так как случайные отклонения неизбежно сопутствуют любому экономическому явлению.

**Замечание 1.1.2.** Если систематические компоненты временного ряда определены правильно, то остающаяся после выделения из временного ряда этих компонент так называемая остаточная последовательность (ряд остатков) будет случайной компонентой ряда. ♦

Модели, в которых временной ряд представлен как произведение некоторых из перечисленных компонент  $t(\tau_i)$ ,  $s(\tau_i)$ ,  $p(\tau_i)$ ,  $\varepsilon(\tau_i)$  относятся к *классу мультипликативных моделей*. Модели этого класса имеют вид

$$Y(\tau_i) = t(\tau_i) \cdot s(\tau_i) \cdot p(\tau_i) \cdot \varepsilon(\tau_i). \quad (1.1.2)$$

В процессе формирования значений временных рядов не всегда участвуют все четыре компоненты. Однако во всех случаях предполагается наличие *случайной составляющей*  $\varepsilon_i = \varepsilon(\tau_i)$ .

Другой класс моделей используется для описания временных рядов, у которых значение  $Y(\tau_i)$  в какой-то степени предопреде-

ляется значениями  $Y(\tau)$  в предыдущие моменты времени  $\tau_{i-1}, \tau_{i-2}, \dots$ . Модели такого класса можно записать в виде

$$Y(\tau_i) = f(Y(\tau_{i-1}), Y(\tau_{i-2}), \dots) + \varepsilon(\tau_i). \quad (1.1.3)$$

Функция  $f$  отражает характер взаимосвязи между последующим и предыдущими значениями величин  $Y(\tau_i)$ . Такие модели получили название *авторегрессионных моделей*.

Для широкого круга процессов функция  $f$  имеет линейный вид

$$Y(\tau_i) = \beta_1 Y(\tau_{i-1}) + \beta_2 Y(\tau_{i-2}) + \dots + \beta_k Y(\tau_{i-k}) + \varepsilon(\tau_i). \quad (1.1.4)$$

Такая модель получила название *линейной авторегрессионной модели  $k$ -го порядка*.

В дальнейшем будет считать, что моменты измерений  $\tau_i$  равноудалены друг от друга на величину  $\Delta_\tau$ , т.е. для любого  $i$  имеет место

$$\tau_{i+1} = \tau_i + \Delta_\tau. \quad (1.1.5)$$

Основная цель статистического анализа временных рядов – *изучение соотношения между закономерностью и случайностью в формировании значений ряда, оценка количественной меры их влияния*. Закономерности, объясняющие динамику показателя в прошлом, используются для *прогнозирования его значений в будущем*, а учет случайности позволяет *определить вероятность отклонения временного ряда от закономерного развития и возможную величину отклонения*.

Прогнозирование экономических процессов, представленных одномерными временными рядами, сводится к выполнению следующих основных этапов:

ЭТАП 1. Предварительный анализ данных.

ЭТАП 2. Построение моделей: формирование набора аппроксимирующих функций (кривых роста) и численное оценивание параметров моделей.

ЭТАП 3. Проверка адекватности моделей и оценка их точности.

ЭТАП 4. Выбор лучшей модели.

ЭТАП 5. Расчет точечного и интервального прогнозов.

## 1.2. Числовые характеристики временного ряда

Из определения временного ряда и моделей (1.1.1), (1.1.2) следует, что в каждый момент  $\tau_i$  величина  $Y(\tau_i)$  является случайной, подчиняющейся некоторому распределению, которое зависит от распределения случайной составляющей  $\varepsilon(\tau_i)$ . Математическое ожидание и дисперсия для модели (1.1.1) в момент  $\tau_i$  определяются выражениями

$$M(Y(\tau_i)) = q(\tau_i); \quad D(Y(\tau_i)) = D(\varepsilon(\tau_i)). \quad (1.2.1)$$

Для модели (1.1.2) эти характеристики определяются более сложными выражениями.

Временной ряд называется *стационарным в широком смысле*, если числовые характеристики случайных величин  $Y(\tau_i)$  не зависят от времени  $\tau_i$ . Так, для модели (1.1.1) справедливы соотношения:

$$M(Y(\tau_i)) = q; \quad D(Y(\tau_i)) = \sigma^2. \quad (1.2.2)$$

Для такого временного ряда в качестве оценок величин  $q$ ,  $\sigma^2$  используются выборочное среднее  $\bar{y}$  и выборочная дисперсия  $s^2$ :

$$\bar{y} = \frac{1}{n} \cdot \sum_{i=1}^n y_i; \quad s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (y_i - \bar{y})^2. \quad (1.2.3)$$

Временной ряд называется *стационарным в узком смысле*, если для каждого момента времени случайные величины  $Y(\tau_i)$

имеют одинаковые распределения. Очевидно, что из стационарности в узком смысле следует стационарность в широком смысле. Обратное, в общем случае, неверно. В дальнейшем рассматриваемые стационарные ряды являются стационарными в широком смысле. Введем еще некоторые характеристики временных рядов.

Степень статистической связи между последовательностями  $Y(\tau_1), Y(\tau_2), \dots, Y(\tau_n)$  и  $Y(\tau_{1+l}), Y(\tau_{2+l}), \dots, Y(\tau_{n+l})$  (сдвинутых относительно друг друга на  $l$  моментов времени, или, как говорят, с лагом  $l$ ) может быть определена с помощью *коэффициента автокорреляции*:

$$\rho(l) = \frac{M[(Y(\tau_i) - q)(Y(\tau_{i+l}) - q)]}{\sigma^2}. \quad (1.2.4)$$

Для стационарного временного ряда  $\rho(l)$  зависит только от лага  $l$  и для него справедливо следующее равенство:

$$\rho(-l) = \rho(l), \quad (1.2.5)$$

т.е. достаточно изучать  $\rho(l)$  только для положительных лагов  $l$ . Если  $l = 0$ , то  $\rho(0) = 1$ .

Оценкой для  $\rho(l)$  является выборочный коэффициент автокорреляции, определяемый по формуле:

$$r(l) = \frac{(n-l) \sum_{i=1}^{n-l} y_i y_{i+l} - \left( \sum_{i=1}^{n-l} y_i \right) \cdot \left( \sum_{i=1}^{n-l} y_{i+l} \right)}{\sqrt{(n-l) \sum_{i=1}^{n-l} y_i^2 - \left( \sum_{i=1}^{n-l} y_i \right)^2} \cdot \sqrt{(n-l) \sum_{i=1}^{n-l} y_{i+l}^2 - \left( \sum_{i=1}^{n-l} y_{i+l} \right)^2}}. \quad (1.2.6)$$

Заметим, что с увеличением  $l$  число пар наблюдений  $y_i, y_{i+l}$  уменьшается и поэтому число  $l$  не должно быть сравнительно большим (рекомендуют  $l \leq n/4$ ).

Стационарный временной ряд, у которого математическое ожидание равно 0, а величины  $\varepsilon(\tau_i)$  некоррелированы, часто называют *белым шумом*. Очевидно, что для белого шума

$$\rho(l) = \begin{cases} 1, & \text{если } l = 0; \\ 0, & \text{если } l \neq 0. \end{cases} \quad (1.2.7)$$

**Замечание 1.2.1.** Для стационарного временного ряда с увеличением лага  $l$  взаимосвязь членов  $Y(\tau_i), Y(\tau_{i+l})$  ослабевает и абсолютные величины коэффициента автокорреляции  $\rho(l)$  должны убывать. В то же время для выборочного коэффициента автокорреляции  $r(l)$  (особенно при небольших значениях  $n-l$ ) свойство монотонного убывания (по абсолютной величине) при возрастании  $l$  может нарушаться. ♦

Проиллюстрируем сделанное замечание.

**Пример 1.2.1.** В столбце А документа Excel, приведенного на рис. 1.1, представлены 20 значений стационарного временного ряда, являющегося белым шумом. Необходимо вычислить выборочное математическое ожидание, дисперсию и коэффициент автокорреляции  $\rho(l)$ ,  $l = 0, 1, 2, 3$ .

*Решение.* Первые две оценки вычисляются по формуле (1.2.3) с использованием стандартных функций Excel (обращение к ним показано на рис. 1.1), а выборочный коэффициент автокорреляции – по формуле (1.2.6), при этом используются предварительно вычисленные суммы:  $\sum_{i=1}^{n-l} y_i y_{i+l}$ ;  $\sum_{i=1}^{n-l} y_i$ ;  $\sum_{i=1}^{n-l} y_{i+l}$ ;  $\sum_{i=1}^{n-l} y_i^2$ ;  $\sum_{i=1}^{n-l} y_{i+l}^2$  (см. рис. 1.1).

Полученные значения оценок приведены в табл. 1.1 (вторая строка). Третья строка таблицы содержит точные значения искомым характеристик. Различие между оценками и точными значениями обусловлено малым объемом выборки. ●

	A	B	C	D	E	F	G	
1								
2	27,65		28,53		=СРЗНАЧ(A2:A21)			
3	32,07							
4	33,87		11,994		=ДИСП(A2:A21)			
5	24,83							
6	34,58							
7	24,59		$l=0$	$l=1$	$l=2$	$l=3$		
8	25,00	$\sum y_i$	570,60	541,01	515,38	488,60		
9	25,00	$\sum y_{i+l}$	570,60	542,95	510,89	477,01		
10	26,29	$\sum y_i^2$	16507,4	15631,6	14974,7	14257,4		
11	26,03	$\sum y_{i+l}^2$	16507,4	15742,7	14714,5	13567		
12	33,65	$\sum y_i \cdot y_{i+l}$	16507,4	15417,3	14658,4	13730,5		
13	25,38							
14	28,82	$r(l)$	1	-0,18911	0,14134	0,10467		
15	33,61							
16	27,85							
17	31,46	=(18*E12-E8*E9)/(КОРЕНЬ(18*E10-E8^2)*КОРЕНЬ(18*E11-E9^2))						
18	27,89							
19	26,78							
20	25,63							
21	29,59							

Рис. 1.1. Вычисление числовых характеристик стационарного ряда

Таблица 1.1

Характеристики	$M(Y)$	$D(Y)$	$\rho(0)$	$\rho(1)$	$\rho(2)$	$\rho(3)$
Оценка	28.5	12.0	1.0	-0.19	0.14	0.10
Точное значение	30	10	1	0	0	0

Отметим два важных свойства коэффициента автокорреляции.

**Во-первых**, коэффициент автокорреляции вычисляется по аналогии с линейным коэффициентом корреляции [5, гл. 2] и таким образом характеризует тесноту линейной зависимости между случайными величинами  $Y(\tau_i)$ ,  $Y(\tau_{i+l})$ . Поэтому по величине коэффициента автокорреляции можно судить о наличии линейной (или близкой к линейной) тенденции развития временного ряда.

**Во-вторых**, по знаку коэффициента автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции значений временного ряда. Многие временные ряды экономических данных имеют положительные величины коэффициентов автокорреляции, однако при этом наблюдается убывающая тенденция.

Последовательность коэффициентов автокорреляции  $\rho(0)$ ,  $\rho(1)$ ,  $\rho(2)$ , ... называют *автокорреляционной функцией временного ряда*, а график зависимости значений  $\rho(l)$  от величины лага  $l$  (или порядка коэффициента автокорреляции  $l$ ) – *коррелограммой*.

Анализ автокорреляционной функции позволяет выявить структуру временного ряда, т.е. наличие в нем составляющих  $t(\tau_i)$ ,  $p(\tau_i)$ ,  $s(\tau_i)$ .

Если наиболее высоким оказался коэффициент автокорреляции  $\rho(1)$ , то исследуемый ряд содержит только трендовую составляющую. Если наиболее высоким оказался коэффициент автокорреляции  $\rho(l)$ , то ряд содержит колебания с периодичностью  $l$  моментов времени, т.е. период колебания равен  $l \cdot \Delta_\tau$ . Если ни один из коэффициентов  $r(l)$  не является значимым (проверка значимости осуществляется точно так же, как и для коэффициента  $r_{xy}$  [5, с. 43]), то относительно структуры ряда можно сделать одно из двух предположений:

- временной ряд не содержит тренда и циклических колебаний, т.е. является белым шумом с  $\rho(l)$  (см. формулу (1.2.7));

- временной ряд содержит сильный нелинейный тренд, для выявления которого необходимо провести дополнительный анализ.

Поэтому коэффициент автокорреляции и автокорреляционную функцию целесообразно использовать для выявления во временном ряде трендовой составляющей и периодической, сезонных составляющих.

**Пример 1.2.2.** В табл. 1.2 приведены условные данные об объемах потребления электроэнергии жителями некоторого района за 16 кварталов.

Необходимо:

- вычислить выборочные коэффициенты корреляции  $r(l)$  для  $l=1, 2, \dots, 8$  и построить коррелограмму;
- провести анализ значений коэффициентов  $r(l)$  и сделать вывод о структуре данного временного ряда.

Таблица 1.2

$i$	$y_i$	$i$	$y_i$	$i$	$y_i$
1	6,0	7	6,0	13	9,0
2	4,4	8	10,0	14	6,6
3	5,0	9	8,0	15	7,0
4	9,0	10	5,6	16	10,8
5	7,2	11	6,4		
6	4,8	12	11,0		

*Решение.* На рис. 1.2а приведен фрагмент документа Excel, в столбец А которого введен номер квартала ( $\tau_i = i$ ,  $i=1, \dots, 16$ ), в столбец В – объем потребления электроэнергии в соответствующем квартале, в С, D, E, F, G, H, I, J – значения  $y_{i-1}, y_{i-2}, y_{i-3}, y_{i-4}, y_{i-5}, y_{i-6}, y_{i-7}, y_{i-8}$ . По заполненной таким образом таблице вычислялись коэффициенты автокорреляции  $r(l)$ ,  $l=1, 2, \dots, 8$ , приведенные на рис. 1.2б. Здесь же показаны примеры программирования вычисления  $\rho(1)$ ,  $\rho(8)$  с использованием функции Excel КОРРЕЛ. На рис. 1.3 приведены значения  $y_i$  временного ряда, а на рис. 1.4 – график коррелограммы.

	A	B	C	D	E	F	G	H	I	J
1	$\tau_i$	$y_i$	$y_{i-1}$	$y_{i-2}$	$y_{i-3}$	$y_{i-4}$	$y_{i-5}$	$y_{i-6}$	$y_{i-7}$	$y_{i-8}$
2	1	6								
3	2	4,4	6							
4	3	5	4,4	6						
5	4	9	5	4,4	6					
6	5	7,2	9	5	4,4	6				
7	6	4,8	7,2	9	5	4,4	6			
8	7	6	4,8	7,2	9	5	4,4	6		
9	8	10	6	4,8	7,2	9	5	4,4	6	
10	9	8	10	6	4,8	7,2	9	5	4,4	6
11	10	5,6	8	10	6	4,8	7,2	9	5	4,4
12	11	6,4	5,6	8	10	6	4,8	7,2	9	5
13	12	11	6,4	5,6	8	10	6	4,8	7,2	9
14	13	9	11	6,4	5,6	8	10	6	4,8	7,2
15	14	6,6	9	11	6,4	5,6	8	10	6	4,8
16	15	7	6,6	9	11	6,4	5,6	8	10	6
17	16	10,8	7	6,6	9	11	6,4	5,6	8	10

а)

	A	B	C	D	E	F	G	H	I	J
18		Порядок	1	2	3	4	5	6	7	8
19		$\rho(l)$	0,165	-0,567	0,114	0,983	0,119	-0,722	-0,003	0,974
20		$abs(\rho(l))$	0,165	0,567	0,114	0,983	0,119	0,722	0,003	0,974
21										
22			КОРРЕЛ(B3:B17;C3:C17)					КОРРЕЛ(B10:B17;J10:J17)		
23										

б)

Рис. 1.2. Вычисление коэффициентов автокорреляции

Анализ значений  $r(l)$  и коррелограммы позволяет сделать вывод о наличии в изучаемом временном ряде сезонных колебаний периодичностью в четыре квартала ( $r(4)=0,983$ ), а также небольшого линейного тренда. Эти выводы хорошо подтверждаются графиком значений временного ряда (см. рис. 1.3). Большое значение  $r(8)=0,974$  объясняется тем, что в 8 есть удвоенный период 4. Такое же большое значение будет у коэффициента автокорреляции  $r(12)$ . ●

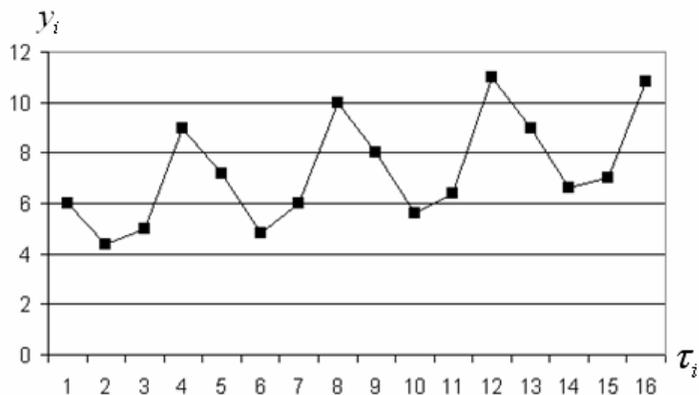


Рис. 1.3. Значения временного ряда

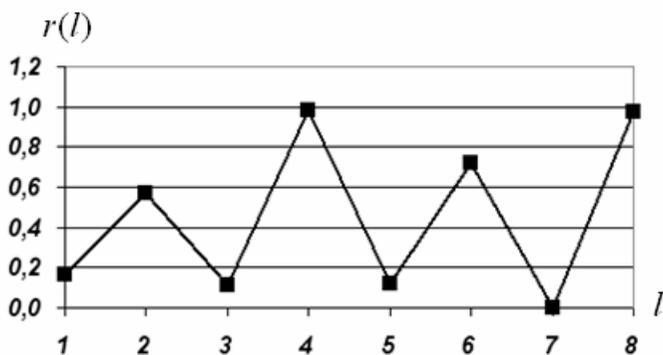


Рис. 1.4. Вычисленная коррелограмма временного ряда

### 1.3. Проверка статистических гипотез о свойствах временного ряда

#### 1.3.1. Проверка гипотезы о наличии аномальных наблюдений

Эта процедура выполняется на стадии предварительного анализа временного ряда и во многих случаях является обязательной процедурой. Для диагностики аномальных наблюдений используются различные статистические методы, одни из которых приводятся ниже.

**Метод Ирвина.** Для всех наблюдений или только «подозреваемых» (в аномальности) наблюдений формулируются следующие статистические гипотезы:

$$H_0: i\text{-е наблюдение не является аномальным;} \quad (1.3.1)$$

$$H_1: i\text{-е наблюдение является аномальным.} \quad (1.3.2)$$

Для проверки этих гипотез вычисляется значение критерия

$$I_i = \frac{|y_i - y_{i-1}|}{s_y}, \quad (1.3.3)$$

где

$$s_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (1.3.4)$$

Нетрудно видеть, что  $s_y$  есть выборочное среднее квадратическое отклонение, вычисленное по выборке объемом  $n$ .

Если вычисленная величина  $I_i$  превышает предельное значение  $I_{np}$  (т.е. попадает в критическую область), то с вероятностью  $\alpha$  ошибки первого рода отвергается гипотеза  $H_0$  и принимается альтернативная гипотеза  $H_1$ , т.е. наблюдение  $y_i$  является аномальным. Предельное значение  $I_{np}$  зависит от количества наблюдений  $n$  и для некоторых  $n$  значения  $I_{np}$  приведены в табл. 1.3.

Таблица 1.3

$n$	$\alpha = 0,05$	$\alpha = 0,01$
2	2,8	3,7
3	2,2	2,9
10	1,5	2,0
20	1,3	1,8
30	1,2	1,7
50	1,1	1,6
100	1,0	1,5
400	0,9	1,3

Обнаруженное аномальное наблюдение необходимо исключить из временного ряда и заменить *расчетным* значением, полученным с использованием соседних наблюдений. Самый простой способ замены – расчетное значение есть среднее двух соседних значений.

**Пример 1.3.1.** В табл. 1.4 приведен индекс потребительских цен.

Таблица 1.4

Дата	$i$	$y_i$	Дата	$i$	$y_i$
4 кв. 1994	1	100	2 кв. 1996	7	105
1 кв. 1995	2	143	3 кв. 1997	8	100
2 кв. 1995	3	124	4 кв. 1997	9	104
3 кв. 1995	4	115	1 кв. 1998	10	105
4 кв. 1995	5	113	2 кв. 1998	11	103
1 кв. 1996	6	110	3 кв. 1998	12	100

Необходимо проверить данный временной ряд на наличие аномальных измерений.

*Решение.* Введем в столбец А, начиная с ячейки А2, значения  $i = 1, 2, \dots, 12$ , а в столбец В – значения  $y_i$ , приведенные в табл. 1.4. (рис. 1.5). График значений  $y_i$  приведен на рис. 1.6.

В ячейке В14 вычислим  $s_y$  (см. (1.3.4)), используя функцию Excel ДИСП. После этого в столбце С запрограммируем вычисление значений  $I_i$ ,  $i = 1, 2, \dots, 12$  критерия (1.3.3). Для определения  $I_{np}$  выполним линейную интерполяцию третьей ( $n = 10$ ) и четвертой ( $n = 20$ ) строк табл. 1.4 для  $\alpha = 0.05$ . Получаем

$$I_{np} = \frac{1.3 - 1.5}{10} (12 - 10) + 1.5 = 1.46.$$

Видим, что неравенство

$$I_i > I_{np} = 1.46$$

выполняется для  $i = 2, 3$ . Следовательно, с вероятностью ошибки первого рода, равной  $\alpha$ , можно принять гипотезу о том, что  $y_2, y_3$  являются аномальными наблюдениями. ☺

	A	B	C	D	E	F
1	$i$	$y_i$		=ABS(B3-B2)/\$B\$14		
2	1	100				
3	2	143	3,406	Аномальное наблюдение		
4	3	124	1,505	Аномальное наблюдение		
5	4	115	0,713			
6	5	113	0,158			
7	6	110	0,238			
8	7	105	0,396			
9	8	100	0,396			
10	9	104	0,317			
11	10	105	0,079			
12	11	103	0,158			
13	12	100	0,238			
14	$S_y$	12,63		=(ДИСП(B2:B13))^(1/2)		

Рис. 1.5. Определение аномальных наблюдений

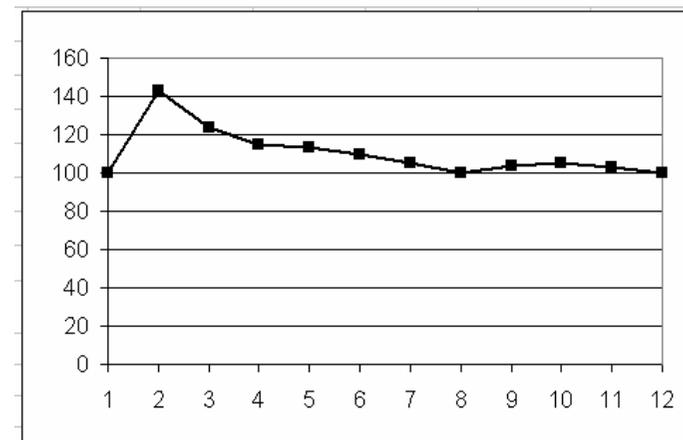


Рис. 1.6. Значения временного ряда

### 1.3.2. Проверка гипотез о наличии неслучайной составляющей временного ряда

Проверка наличия или отсутствия неслучайной составляющей  $q(\tau)$  (модель (1.1.1)) по существу состоит в проверке гипотезы о постоянстве среднего значения временного ряда. Поэтому сформулируем две статистические гипотезы:

$$H_0: M(Y(\tau_i)) = const; \quad (1.3.5)$$

$$H_1: M(Y(\tau_i)) \neq const. \quad (1.3.6)$$

Для проверки этих гипотез используются различные критерии. Здесь мы ограничимся двумя (достаточно простыми критериями).

**Критерий 1.** В этом критерии временной ряд разбивают на две примерно равные по числу значений части, каждая из которых рассматривается как некоторая самостоятельная выборочная совокупность, имеющая нормальное распределение. Если временной ряд имеет тенденцию к тренду, то средние, вычисленные для каждой совокупности, должны существенно (значимо) различаться между собой. Если же расхождение незначительно, не существенно (случайно), то временной ряд не имеет тенденции. Таким образом, проверка наличия тренда (т.е. неслучайной составляющей) в исследуемом временном ряду сводится к проверке гипотезы о равенстве средних двух нормально распределенных совокупностей. Пусть первая часть (обозначим ее  $Y^{(I)}$ ) содержит  $n_I$  наблюдений  $Y(\tau_i)$ ,  $i=1,2,\dots,n_I$ , а вторая часть –  $Y^{(II)}$  содержит  $n_{II}$  наблюдений  $Y(\tau_i)$ ,  $i=n_I+1,\dots,n_I+n_{II}$ .

Для каждой части временного ряда вычислим (используя формулы (1.2.3)) выборочное среднее  $\bar{y}_I$ ,  $\bar{y}_{II}$  и выборочные дисперсии  $s_I^2$ ,  $s_{II}^2$ :

$$s_I^2 = \frac{1}{n_I - 1} \cdot \sum_{i=1}^{n_I} (y_i - \bar{y}_I)^2; \quad s_{II}^2 = \frac{1}{n_{II} - 1} \cdot \sum_{i=n_I+1}^{n_{II}} (y_i - \bar{y}_{II})^2. \quad (1.3.7)$$

Далее рассчитаем значение критерия

$$K_S = \frac{\bar{y}_I - \bar{y}_{II}}{\sqrt{(n_I - 1)s_I^2 + (n_{II} - 1)s_{II}^2}} \cdot \sqrt{\frac{n_I \cdot n_{II} (n_I + n_{II} - 2)}{n_I + n_{II}}} \quad (1.3.8)$$

(часто называемого критерием Стьюдента). Если выполняется неравенство

$$K_S > t(1 - \alpha, n_I + n_{II} - 2), \quad (1.3.9)$$

то гипотеза о постоянстве математического ожидания отклоняется с уровнем значимости  $\alpha$ . Напомним, что значение  $t(1 - \alpha, n_I + n_{II} - 2)$  вычисляется с использованием следующей функции Excel:

$$t(1 - \alpha, n_I + n_{II} - 2) = \text{СТЮДРАСПОБР}(\alpha, n_I + n_{II} - 2).$$

Для использования критерия (1.3.8) необходимо убедиться, что дисперсии обеих частей ряда одинаковы. Для этого используем критерий Фишера:

$$F_S = \frac{\max(s_I^2, s_{II}^2)}{\min(s_I^2, s_{II}^2)}, \quad (1.3.10)$$

где  $s_I^2$ ,  $s_{II}^2$  – оценки дисперсии, вычисленные по первой (число измерений  $n_I$ ) и второй (число измерений  $n_{II}$ ) частям временного ряда. Если **не выполняется неравенство**

$$F_{\frac{\alpha}{2}; n_I - 1; n_{II} - 1} \leq F_S \leq F_{1 - \frac{\alpha}{2}; n_I - 1; n_{II} - 1}, \quad (1.3.11)$$

то гипотеза о постоянстве дисперсии отвергается с уровнем значимости  $\alpha$ . В этом случае критерий (1.3.8) не применим, и необходимо использовать другой критерий или принять гипотезу о наличии неслучайной составляющей временного ряда, так как это составляющая может сказаться и на других характеристиках временного ряда: дисперсии, коэффициенте автокорреляции и т.д.

Границы критической области при проверке гипотезы о равенстве дисперсий вычисляются с помощью следующей функции Excel:

$$F_{\frac{\alpha}{2}; n_I - 1; n_{II} - 1} = \text{ФРАСПОБР}\left(1 - \frac{\alpha}{2}; n_I - 1; n_{II} - 1\right). \quad (1.3.12)$$

**Пример 1.3.2.** Осуществить тестирование временного ряда, приведенного в столбце А на рис. 1.7, на наличие неслучайной составляющей.

*Решение.* Разобьем исходный временной ряд на две части по 10 измерений в каждой. Вычислим по каждой из этих частей выборочные оценки (см. рис. 1.7):

$$\bar{y}_I = 30.68, \quad \bar{y}_{II} = 30.14, \quad s_I^2 = 10.19, \quad s_{II}^2 = 8.16.$$

Затем определим значения критериев (1.3.8) и (1.3.10) (см. рис. 1.7):  $K_s = 0.40$ ;  $F_s = 1.249$ . Проверим выполнение неравенств (1.3.9) и (1.3.11). Неравенство (1.3.9) не выполняется, так как  $0.40 < 2.101$ , а неравенство (1.3.11) выполняется —  $0.248 < 1.249 < 4.026$ .

Следовательно, можно сделать вывод об отсутствии неслучайной составляющей рассматриваемого временного ряда. ●

	A	B	C	D	E	F	G
1							
2	27,17						
3	28,18						
4	32,21						
5	26,29						
6	33,84		$\bar{y}_I$	30,68			
7	34,61		$s_I$	10,19			
8	29,33		=СТЮДРАСПОБР(0,05;18)				
9	28,14			2,101			
10	34,44		Критерий K	0,40			
11	32,61						
12	27,10		Критерий F	1,249			
13	32,47		=ФРАСПОБР(0,975;9;9)		=ФРАСПОБР(0,025;9;9)		
14	32,58			0,248		4,026	
15	27,75						
16	30,28		$\bar{y}_{II}$	30,14			
17	30,21		$s_{II}$	8,16			
18	33,97						
19	26,10						
20	33,27						
21	27,64						

Рис. 1.7. Проверка гипотезы о стационарности ряда

**Критерий 2 (критерий серий).** Расположим члены анализируемого временного ряда в порядке возрастания, т.е. образуем вариационный ряд вида

$$y_{(1)} \leq y_{(2)} \leq y_{(3)} \dots \leq y_{(n)}.$$

Определим выборочную медиану по формуле

$$y_{med} = \begin{cases} y_{\left(\frac{n+1}{2}\right)}, & \text{если } n \text{ не четно;} \\ \frac{1}{2} \left( y_{\left(\frac{n}{2}\right)} + y_{\left(\frac{n}{2}+1\right)} \right), & \text{если } n \text{ четно.} \end{cases} \quad (1.3.13)$$

После этого образуем «серии» из плюсов и минусов, на статистическом анализе которых основана процедура проверки гипотезы о неизменности среднего значения временного ряда.

По исходному временному ряду, построим последовательность из плюсов и минусов следующим образом: переменной  $x_i$  ставим знак «+», если  $y_i > y_{med}$ , и знак «-», если  $y_i < y_{med}$  (члены временного ряда, равные  $y_{med}$  не учитываются).

Образованная последовательность плюсов и минусов характеризуется общим числом серий  $v(n)$  и протяженностью самой длинной серии  $\tau(n)$ . При этом под «серией» понимается последовательность подряд идущих плюсов и подряд идущих минусов. Если исследуемый ряд состоит из статистически независимых наблюдений, случайно варьирующих около некоторого постоянного уровня (т.е. справедлива гипотеза о неизменности среднего значения временного ряда), то чередование «+» и «-» в построенной последовательности должно быть случайным, т.е. эта последовательность не должна содержать слишком длинных серий подряд идущих «+» и «-», и, соответственно, общее число серий не должно быть слишком малым. Так что в данной критерии целесообразно рассмотреть одновременно пару критических статистик ( $v(n)$ ;  $\tau(n)$ ).

Справедлив следующий приближенный статистический критерий проверки гипотезы о неизменности среднего значения временного ряда:

если хотя бы одно из неравенств

$$\begin{cases} v(n) > \text{int} \left[ \frac{1}{2}(n+2-1,96\sqrt{n-1}) \right] \\ \tau(n) < \text{int} [1,43 \ln(n+1)] \end{cases} \quad (1.3.14)$$

окажется нарушенным, то гипотеза о неизменности среднего значения временного ряда отвергается с вероятностью ошибки  $\alpha$ , такой, что  $0.05 < \alpha < 0.0975$  и, тем самым, подтверждается

наличие зависящей от времени неслучайной составляющей в модели  $Y(\tau) = q(\tau) + \varepsilon(\tau)$ .

Функция  $\text{int}[z]$  означает взятие целой части числа  $z$ .

### 1.3.3. Проверка гипотезы о наличии неслучайной составляющей в Excel

Вернемся к критерию 1 (см. п. 1.3.2), который представляет собой проверку критериев Стьюдента и Фишера. Вычисление и проверку этих критериев можно оперативно осуществить используя соответствующие режимы пакета *Анализ данных* табличного процессора Excel [5, с. 137–142].

Проверку гипотезы о наличии неслучайной составляющей в Excel покажем на данных следующего примера.

**Пример 1.3.3.** В табл. 1.5 приведена урожайность ячменя в одной из областей среднего Поволжья, ц/га. Необходимо проверить этот временной ряд на наличие неслучайной составляющей, используя соответствующие режимы пакета *Анализ данных* табличного процессора Excel.

Таблица 1.5

Годы	1	2	3	4	5	6	7	8
Урожайность	14,1	9,3	19,4	19,7	5,4	24,2	13,8	24,5
Годы	9	10	11	12	13	14	15	
Урожайность	14,7	16,6	5,6	16,2	25,3	11,9	18,5	

*Решение.* Проверку на наличие неслучайной составляющей представим следующими шагами:

*Шаг 1.* Введем в столбец А (начиная с ячейки А2) значения  $i = 1, \dots, 15$  (номера годов), а в столбец В – значения  $y_i, i = 1, \dots, 15$ .

*Шаг 2.* Гипотезу о равенстве дисперсий проверим с помощью F-теста, для вызова которого необходимо обратиться к пункту **Сервис**, команде *Анализ данных* и в списке инструментов анализа выбрать *Двухвыборочный F-тест для дисперсий* (рис. 1.8).

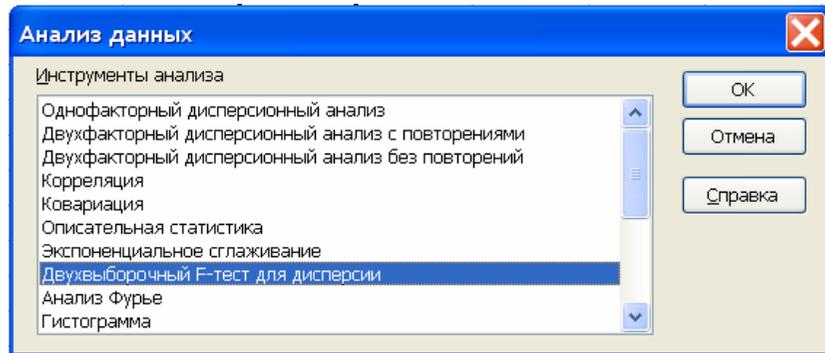


Рис. 1.8. Выбор F-теста

Шаг 3. Заполняем поля диалогового окна *Двухвыборочный F-тест для дисперсий* как показано на рис. 1.9.

Результат выполнения теста приведен в таблице, показанной на рис. 1.10.

Так как  $s_1^2 > s_1^2$ , то в качестве альтернативной принимается гипотеза

$$H_1: \sigma_1^2 > \sigma_1^2 \quad (1.3.15)$$

и в этом случае критическая область представляет собой интервал  $(x_{np,\alpha}, \infty)$ , где точка  $x_{np,\alpha}$  определяется из условия

$$P(F > x_{np,\alpha}) = \alpha, \quad (1.3.16)$$

где  $\alpha$  – вероятность ошибки первого рода. Из табл. рис. 1.10 находим:  $F = 1.022$ ,  $x_{np,\alpha} = 3.866$ .

Видно, что наблюдаемое значение  $F = 1.022$  не попадает в критическую область и принимается нулевая гипотеза

$$H_0: \sigma_1^2 = \sigma_1^2. \quad (1.3.17)$$

Шаг 4. Для проверки гипотезы о равенстве математических ожиданий в списке *Инструменты анализа* (см. рис. 1.8) выбираем режим *Двухвыборочный t-тест с одинаковыми дисперсиями* и вводим данные в соответствующие поля (см. рис. 1.11).

Результаты выполнения t-теста приведены в таблице, показанной на рис. 1.12.

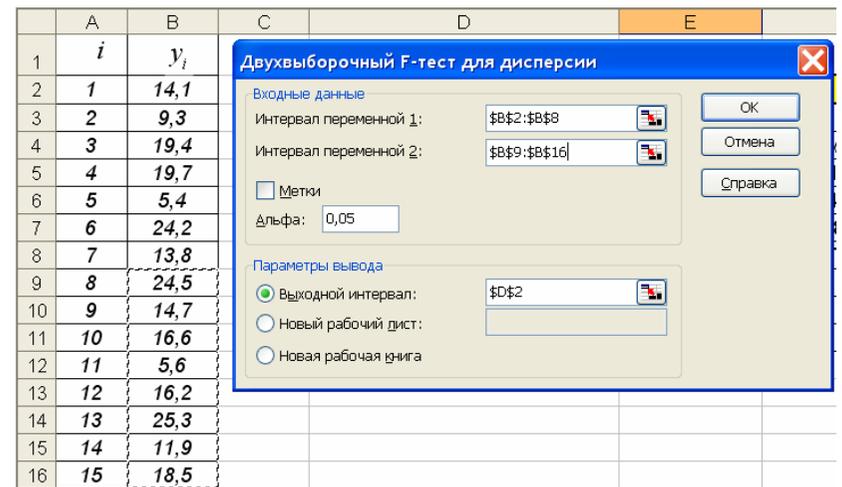


Рис. 1.9. Ввод данных для двухвыборочного F-теста

	D	E	F
1			
2	<b>Двухвыборочный F-тест для дисперсии</b>		
3			
4		<i>Переменная 1</i>	<i>Переменная 2</i>
5	<b>Среднее</b>	<b>15,129</b>	<b>16,663</b>
6	<b>Дисперсия</b>	<b>42,146</b>	<b>41,220</b>
7	<b>Наблюдения</b>	<b>7</b>	<b>8</b>
8	<b>df</b>	<b>6</b>	<b>7</b>
9	<b>F</b>	<b>1,022</b>	
10	<b>P(F&lt;=f) одностороннее</b>	<b>0,481</b>	
11	<b>F критическое одностороннее</b>	<b>3,866</b>	

Рис. 1.10. Результаты выполнения двухвыборочного теста

Рис. 1.11. Ввод данных для двухвыборочного t-теста

	D	E	F
1			
2	<b>Двухвыборочный t-тест с одинаковыми дисперсиями</b>		
3			
4		<i>Переменная 1</i>	<i>Переменная 2</i>
5	<b>Среднее</b>	15,129	16,663
6	<b>Дисперсия</b>	42,146	41,220
7	<b>Наблюдения</b>	7	8
8	<b>Объединенная дисперсия</b>	41,647	
9	<b>Гипотетическая разность средних</b>	0,000	
10	<b>df</b>	13,000	
11	<b>t-статистика</b>	-0,459	
12	<b>P(T&lt;=t) одностороннее</b>	0,327	
13	<b>t критическое одностороннее</b>	1,771	
14	<b>P(T&lt;=t) двухстороннее</b>	0,654	
15	<b>t критическое двухстороннее</b>	2,160	

Рис. 1.12. Результаты выполнения t-теста

Критическая область является объединением двух интервалов и имеет вид

$$(-\infty, -2.160] \cup [2.160, \infty).$$

Видно, что наблюдаемое значение критерия, равное  $-0.459$ , не попадает в эту область и поэтому принимается основная гипотеза  $H_0$  о равенстве математических ожиданий. Принятие этих двух гипотез (о равенстве дисперсий и равенстве математических ожиданий) позволяет принять гипотезу об отсутствии трендовой составляющей в данном временном ряду. ☺

### 1.3.4. Проверка гипотезы о стационарности временного ряда

Для стационарности временного ряда достаточно постоянства его числовых характеристик на всем интервале определения временного ряда. Наиболее часто в качестве таких характеристик берут математическое ожидание и дисперсию. Тогда ответ на вопрос стационарности дискретного временного ряда сводится к проверке следующей пары статистических гипотез:

$$\left. \begin{aligned} H_0 : M(Y(\tau_i)) = const; \\ H_1 : M(Y(\tau_i)) \neq const. \end{aligned} \right\} \begin{array}{l} \text{Постоянство} \\ \text{математического ожидания} \end{array} \quad (1.3.18)$$

$$\left. \begin{aligned} H_0 : D(Y(\tau_i)) = const; \\ H_1 : D(Y(\tau_i)) \neq const. \end{aligned} \right\} \begin{array}{l} \text{Постоянство дисперсии} \end{array} \quad (1.3.19)$$

Для проверки этих гипотез можно использовать критерий 1, описанный в п. 1.3.2. Действительно, используя критерий Фишера (1.3.10), проверяем гипотезу о постоянстве дисперсий. Если эта гипотеза принимается, то на следующем шаге проверяется гипотеза о постоянстве математического ожидания (критерий (1.3.8)). Если принимается гипотеза о постоянстве математического ожидания, то принимается гипотеза о стационарности (в широком смысле) временного ряда.

Очевидно, что для проверок гипотез (1.3.18), (1.3.19) можно использовать режим *Анализ данных* табличного процессора Excel, как это показано в п. 1.3.3.

## ЛАБОРАТОРНАЯ РАБОТА

*Исходные данные.* В таблице приведена урожайность гречи в Новосибирской области, ц/га. Необходимо проверить этот временной ряд на наличие неслучайной составляющей, используя критерий 1 п. 1.3.2.

Годы	1	2	3	4	5	6	7	8
Урожайность	14,1	11,3	19,4	19,7	8,4	21,2	16,8	24,5
Годы	9	10	11	12	13	14	15	
Урожайность	19,7	22,6	19,6	23,2	25,3	21,9	22,5	

*Рекомендации.*

1. При проверке статистической гипотезы временной ряд разбить на две части

$$n_I = 7; n_{II} = 8 \quad (n = n_I + n_{II} = 15).$$

2. При программировании вычислений использовать пример 1.3.2.

## КОНТРОЛЬНАЯ РАБОТА

В таблице приведены данные об урожайности озимой пшеницы за 10 лет, ц/га.

Годы $\tau_i$	1	2	3	4	5
$y_i$	16.3+N	20.2+N	17.1+N	9.2+N	15.3+N
Годы $\tau_i$	6	7	8	9	10
$y_i$	16.7+N	19.9+N	14.4+N	18.7+N	20.7+N

Примечание. N – последняя цифра в изменяющемся номере зачетной книжки.

*Необходимо:*

1. Построить график значений временного ряда.
2. Вычислить среднее значение и дисперсию.
3. Вычислить коэффициенты автокорреляции для лагов  $l = 1, 2, 3$  и построить коррелограмму.

4. Используя материал п. 1.2, сделать обоснованные выводы о структуре исследуемого временного ряда.

5. Выполнить исследования временного ряда на наличие аномальных наблюдений.

6. Выполнить исследования временного ряда на наличие неслучайной составляющей (используя соответствующие режимы пакета *Анализ данных* табличного процессора Excel).

7. Вставить в контрольную работу копии таблиц с результатами проверок п. 5, 6.

## ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Какой временной ряд называется стационарным в широком смысле?

2. Какой временной ряд называется стационарным в узком смысле?

3. Что характеризует коэффициент автокорреляции  $\rho(l)$ ?

4. Что характеризует трендовая составляющая временного ряда?

5. Что характеризует сезонная составляющая временного ряда?

6. Что характеризует периодическая составляющая временного ряда?

7. Что такое аномальное наблюдение?

8. Сущность метода Ирвина.

9. Сущность проверки гипотезы о наличии неслучайной составляющей с помощью критерия 1.

## ГЛАВА 2. ВЫДЕЛЕНИЕ НЕСЛУЧАЙНОЙ СОСТАВЛЯЮЩЕЙ ВРЕМЕННОГО РЯДА

### 2.1. Выделение трендовой составляющей временного ряда

#### 2.1.1. Задача выделения трендовой составляющей

Трендовая составляющая  $t(\tau)$  отражает влияние долговременных факторов и соответствует устойчивой и долговременной тенденции изменения временного ряда. Знание трендовой составляющей позволяет осуществлять долговременное прогнозирование. Поэтому возникает задача выделения тренда, т.е. построение оценки  $\hat{t}(\tau)$  для функции  $t(\tau)$  (или оценок  $\hat{t}(\tau_i)$  для значений  $t(\tau_i)$ ) по заданной временной выборке  $\{\tau_i, y_i\}$ . При этом предполагается, что остальные составляющие  $p(\tau)$ ,  $s(\tau)$  временного ряда отсутствуют.

Кроме прогнозирования задача выделения трендовой составляющей возникает в следующих ситуациях:

- при графическом отображении временного ряда тренд прослеживается недостаточно отчетливо. После выделения трендовой составляющей и нанесения значений тренда на график тенденция изменения временного ряда проявляется более четко;
- некоторые методы анализа и прогнозирования требуют в качестве предварительной обработки выделение тренда;
- выделение тренда используют для устранения аномальных наблюдений.

В литературе часто задачу выделения тренда называют *задачей сглаживания временного ряда* или *сглаживанием временного ряда*. В дальнейшем будут использоваться оба эти названия.

Существующие методы выделения тренда можно разделить на два класса:

- 1) методы парного регрессионного анализа;
- 2) сглаживающие методы.

Рассмотрим эти методы более подробно.

#### 2.1.2. Регрессионные методы выделения трендовой составляющей

Методы парного регрессионного анализа (или проще – регрессионные методы) основаны на следующей модели временного ряда:

$$Y(\tau_i) = t(\tau_i) + \varepsilon(\tau_i), \quad (2.1.1)$$

где случайные величины  $\xi(\tau_i)$  удовлетворяют условиям:

$$\begin{aligned} M(\varepsilon(\tau_i)) &= 0; \\ M(\varepsilon(\tau_i)\varepsilon(\tau_j)) &= \begin{cases} \sigma^2, & i = j; \\ 0, & i \neq j. \end{cases} \end{aligned} \quad (2.1.2)$$

Рассматривая время  $\tau$  как независимую переменную, функцию  $t(\tau)$  можно оценить, используя методы парной регрессии [5, гл. 2]. Поэтому здесь ограничимся только рассмотрением некоторых особенностей применения методов парной регрессии к решению задачи выделения трендовой составляющей.

Одна из особенностей заключается в том, что различный характер тренда (иногда достаточно сложный) обуславливает более широкое использование нелинейных функций. Так, наряду с линейной функцией  $t(\tau) = \beta_0 + \beta_1\tau$  гораздо чаще используются следующие нелинейные функции:

- полиномиальная

$$t(\tau) = \beta_0 + \beta_1\tau + \dots + \beta_p\tau^p, \quad (2.1.3)$$

где  $p$  – степень полинома (при  $p = 1$  получаем линейную функцию);

- экспоненциальная

$$t(\tau) = \beta_0 e^{\beta_1\tau}; \quad (2.1.4)$$

- логистическая

$$t(\tau) = \frac{\beta_0}{1 + \beta_1 e^{-\beta_2\tau}}. \quad (2.1.5)$$

Выбор вида функции  $t(\tau)$  часто основывается на анализе графического изображения ряда, т.е. на анализе диаграммы рассеяния, построенной по точкам  $\{\tau_i, y_i\}$  [5, п. 2.1].

При применении полиномиальной функции важно правильно определить степень полинома. Для этого можно использовать метод последовательных разностей, заключающийся в вычислении разностей:

- первого порядка

$$\Delta_i = y_i - y_{i-1}, \quad i = 1, 2, \dots, n-1;$$

- второго порядка

$$\Delta_i^2 = \Delta_i - \Delta_{i-1}, \quad i = 1, 2, \dots, n-2;$$

- $k$ -го порядка

$$\Delta_i^k = \Delta_i^{k-1} - \Delta_{i-1}^{k-1}, \quad i = 1, 2, \dots, n-k,$$

а также величин

$$d_k = \frac{1}{n-k} \cdot \frac{\sum_{i=1}^{n-k} (\Delta_i^{(k)})^2}{C_{2k}^k}, \quad (2.1.6)$$

где  $C_{2k}^k$  – сочетание, определяемое по формуле  $C_{2k}^k = \frac{(2k)!}{(k!)^2}$ . Величина  $d_k$  первоначально убывает с ростом  $k$ , а затем, начиная с некоторого значения  $k_0$ , стабилизируется, оставаясь приблизительно на одном уровне при дальнейшем росте  $k$ . Тогда степень полинома определяется по формуле  $p = k_0 - 1$ .

После выбора вида функции  $t(\tau)$  строят уравнение регрессии  $\hat{t}(\tau)$ , зависящее от коэффициентов  $b_0, b_1, \dots, b_k$ , которые являются оценками коэффициентов  $\beta_0, \beta_1, \dots, \beta_k$  функции тренда

$t(\tau)$ . Так, для полиномиального тренда (2.1.3) уравнение регрессии примет вид

$$\hat{t}(\tau) = b_0 + b_1\tau + \dots + b_k\tau^k. \quad (2.1.7)$$

Для вычисления коэффициентов  $b_0, b_1, \dots, b_k$  используется метод наименьших квадратов, т.е. коэффициенты находятся из условия минимума функционала

$$F(b_0, b_1, \dots, b_k) = \sum_{i=1}^n (t_i - \hat{t}(\tau_i))^2, \quad (2.1.8)$$

где  $\hat{t}(\tau)$  – значение уравнения тренда в точке  $\tau = \tau_i$ .

Использование нелинейных функций  $t(\tau)$  обуславливает следующие виды нелинейности уравнения регрессии: нелинейность по переменной и коэффициентам [5, п. 2.6]. Напомним, что в этих случаях используются два подхода для вычисления коэффициентов регрессии:

- 1) заменой переменной или нелинейными преобразованиями осуществляется линеаризация уравнения регрессии, к которому применяется метод наименьших квадратов;
- 2) непосредственное вычисление коэффициентов из условий минимума функционала (2.1.8).

Для иллюстрации **первого подхода** рассмотрим мультипликативную модель временного ряда:

$$Y(\tau) = \beta_0 \tau^{\beta_1} \cdot \varepsilon = t(\tau) \cdot \varepsilon. \quad (2.1.9)$$

После логарифмирования (2.1.9) получаем

$$\ln Y(\tau) = \ln \beta_0 + \beta_1 \ln \tau + \ln \varepsilon. \quad (2.1.10)$$

Введем новые величины:

$$Y'(\tau) = \ln Y(\tau); \beta'_0 = \ln \beta_0; \tau' = \ln \tau, \varepsilon' = \ln \varepsilon.$$

Относительно этих величин имеем линейную регрессионную модель

$$Y'(\tau') = \beta'_0 + \beta_1 \tau' + \varepsilon', \quad (2.1.11)$$

которой соответствует уравнение трендовой составляющей

$$\hat{t}'(\tau') = b'_0 + b_1 \tau'. \quad (2.1.12)$$

Коэффициенты  $b'_0, b_1$  вычисляются на основе МНК по формулам, приведенным в [5, п. 2.3]. Выполнив обратное преобразование  $b_0 = e^{b'_0}$ , получаем искомые оценки  $b_0, b_1$  для коэффициентов нелинейной регрессии (2.1.9).

**Замечание 2.1.1.** Эффективность оценок, получаемых методом наименьших квадратов, основана на допущении о том, что возмущения  $\varepsilon_i$  не коррелированы между собой и подчиняются нормальному распределению  $N(0, \sigma^2)$ , т.е. имеет одинаковую дисперсию  $\sigma^2$ . К сожалению, выполнение нелинейных преобразований приводит к нарушению этого допущения. Для иллюстрации этого вернемся к преобразованному уравнению регрессии (2.1.11). Коэффициенты этого уравнения будут являться эффективными оценками для  $\beta'_0, \beta_1$ , если  $\varepsilon' = \ln \varepsilon \sim N(0, \sigma^2)$ , т.е. возмущения  $\varepsilon_i$  исходной модели (2.1.9) должны иметь логарифмически нормальное распределение, что на практике встречается редко. Нарушение свойства гомоскедастичности приводит к тому, что вычисленные на основе МНК коэффициенты *будут несмещенными, состоятельными оценками* для соответствующих коэффициентов регрессионной модели, но *они не будут обладать свойством эффективности*, т.е. возможно вычислить (используя другие алгоритмы) оценки с меньшей дисперсией. ♦

**Второй подход** используется в случаях, когда невозможно подобрать преобразования для перехода к новой линейной регрессии. Для примера рассмотрим модель временного ряда

$$Y(\tau) = \beta_0 \cdot \tau^{\beta_1} + \varepsilon. \quad (2.1.13)$$

Логарифмирование этого уравнения не приводит к линейной модели  $\ln Y(\tau) = \ln(\beta_0 \cdot \tau^{\beta_1} + \varepsilon)$ .

В этих случаях оценки для коэффициентов уравнения тренда модели вычисляются на основе минимизации функционала некоторого функционала, например, функционала метода наименьших квадратов. Так, для модели (2.1.13) уравнение тренда имеет вид

$$\hat{t}(\tau) = b_0 \tau^{b_1}, \quad (2.1.14)$$

а минимизируемый функционал МНК определяется выражением:

$$F(b_0, b_1) = \sum_{i=1}^n (y_i - b_0 \tau_i^{b_1})^2 = \sum_{i=1}^n (y_i - \hat{t}_i)^2. \quad (2.1.15)$$

Существует достаточно большое число алгоритмов минимизации различных функционалов. Некоторые из этих алгоритмов реализованы в табличном процессоре Excel (команда *Поиск решения* пункта меню **Сервис** – подробнее см. п. 2.1.4).

После вычисления коэффициентов  $b_0, b_1, \dots, b_k$ , уравнение регрессии принимается в качестве оценки для функции тренда  $t(\tau)$  и может быть использовано для дальнейшего анализа временного ряда или его прогнозирования.

**Пример 2.1.1.** В табл. 2.1 приведены данные, отражающие спрос (в условных единицах) на некоторый товар за восьмилетний период.

Таблица 2.1

Год	1	2	3	4	5	6	7	8
Спрос	213	171	291	309	317	362	351	361

По этим данным (которые являются временной выборкой) найти оценку  $\hat{t}(\tau)$ , предполагая, что  $t(\tau)$  является квадратичной функцией. Выполнить прогноз временного ряда для десятого года.

Решение. При сделанном предположении оценка  $\hat{i}(\tau)$  имеет вид

$$\hat{i}(\tau) = b_0 + b_1\tau + b_2\tau^2 \quad (2.1.16)$$

и это уравнение регрессии нелинейно по переменным. Для перехода к линейному уравнению регрессии введем новые переменные  $x_1 = \tau$ ;  $x_2 = \tau^2$  и получим множественную линейную регрессию:

$$\hat{i}(x_1, x_2) = b_0 + b_1x_1 + b_2x_2.$$

Вектор коэффициентов  $b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}$  находим методом наименьших

квадратов, решая уже известную систему нормальных уравнений [5, п. 3.2]:

$$(X^T X) b = X^T y,$$

где  $X$  – матрица размера  $8 \times 3$ , а  $y$  – вектор наблюдений. Формирование матрицы  $X$  и решение системы показано на рис. 2.1. Вычисленный вектор коэффициентов (ячейки F16 – F18 выделенные

цветом) имеет следующие проекции :  $b = \begin{pmatrix} 132.3 \\ 55.09 \\ -3.26 \end{pmatrix}$ .

	A	B	C	D	E	F
1	Год	Спрос	Матрица X			$\hat{i}$
2	1	213	1	1	1	184,13
3	2	171	1	2	4	229,41
4	3	291	1	3	9	268,16
5	4	309	1	4	16	300,38
6	5	317	1	5	25	326,05
7	6	362	1	6	36	345,20
8	7	351	1	7	49	357,80
9	8	361	1	8	64	363,88
10	9					363,41
11	10					356,41
12		8	36	204		2375
13	$X^T \cdot X =$	36	204	1296	$X^T \cdot y =$	11766
14		204	1296	8772		69720
15						
16		1,9464	-0,9107	0,0893		132,3
17	$(X^T \cdot X)^{-1} =$	-0,9107	0,506	-0,0536	b =	55,089
18		0,0893	-0,0536	0,006		-3,2679

Рис. 2.1. Вычисление коэффициентов квадратичного тренда

Возвращаясь к уравнению (2.1.16), получаем следующую оценку для тренда временного ряда:

$$\hat{i}(\tau) = 132.3 + 55.09 \cdot \tau - 3.26 \cdot \tau^2. \quad (2.1.17)$$

На рис. 2.2 показана временная выборка  $y_i, i = 1, 2, \dots, 8$  (кривая 1, маркированная квадратиками) и график функции  $\hat{i}(\tau)$  (кривая 2, маркированная ромбами). Для выполнения прогноза достаточно в (2.1.16) подставить  $\tau = 10$ . Получаем значение  $\hat{i}(10) = 356.41$ . ●

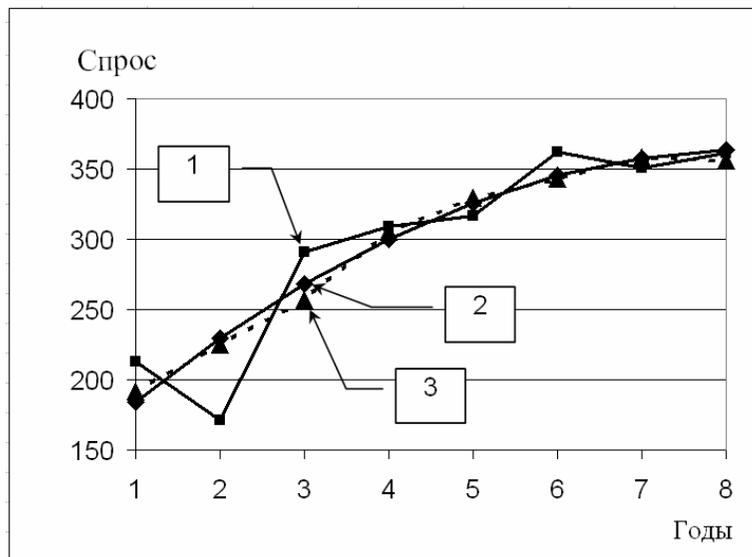


Рис. 2.2. Графики временной выборки и оценок тренда

### 2.1.3. Индекс детерминации и степень нелинейности трендовой составляющей

Введем суммы

$$Q_e = \sum_{i=1}^n (y_i - \hat{t}_i)^2; \quad Q = \sum_{i=1}^n (y_i - \bar{y})^2,$$

где  $\hat{t}_i$  – значение, вычисленное по уравнению тренда при  $\tau = \tau_i$ ,

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

Индексом детерминации называется величина

$$R_t^2 = 1 - \frac{Q_e}{Q}, \quad (2.1.18)$$

которая изменяется в пределах  $0 \leq R_t^2 \leq 1$  и показывает, какая часть (доля) изменения временного ряда обусловлена изменением переменной  $\tau$ , т.е. индекс детерминации имеет тот же смысл, что

и коэффициент детерминации  $R^2$  линейной регрессионной модели.

Если уравнение тренда является линейной функцией, то справедливо тождество

$$R_t^2 = R^2, \quad (2.1.19)$$

где  $R^2$  – коэффициент детерминации линейной регрессии. Это тождество является теоретическим обоснованием возможности замены нелинейного тренда линейной функцией. Заметим, что чем больше кривизна линии регрессии, тем величина коэффициента детерминации  $R^2$  меньше индекса детерминации  $R_t^2$ . Близость этих величин означает, что нет необходимости усложнять уравнения тренда и можно использовать для тренда линейную функцию.

Для проверки нулевой гипотезы  $H_0$  о возможности замены нелинейной регрессии линейной функцией определим следующий критерий:

$$T_{\text{нел}} = \frac{R_t^2 - R^2}{\delta_{\Delta}}, \quad (2.1.20)$$

где  $\delta_{\Delta}$  – ошибка разности  $\Delta = R_t^2 - R^2$ , определяемая по формуле

$$\delta_{\Delta} = 2 \cdot \sqrt{\frac{(R_t^2 - R^2) - (R_t^2 - R^2)^2 \cdot (2 - (R_t^2 + R^2))}{n}}. \quad (2.1.21)$$

Нулевая гипотеза  $H_0$  отвергается с уровнем значимости  $\alpha$ , если выполняется неравенство

$$T_{\text{нел}} > t(1 - \alpha, n - 2); \quad (2.1.22)$$

где  $t(1 - \alpha, n - 2) = \text{СТЫЮДРАСПОБР}(\alpha; n - 2)$ . Это говорит о существенном различии между  $R_t^2$  и  $R^2$  и невозможности замены нелинейного тренда линейной функцией.

**Пример 2.1.2.** Предположим, что построенное уравнение тренда имеет вид

$$\hat{t}(\tau) = 9.876 + 5.129 \ln(\tau) \quad (2.1.23)$$

и вычислен индекс детерминации  $R_t^2 = 0.99164$ . Необходимо проверить возможность замены этого нелинейного уравнения линейным уравнением вида

$$\hat{t}(\tau) = 9.28 + 1.777\tau, \quad (2.1.24)$$

для которого  $R^2 = 0.94898$ .

*Решение.* Вычислим следующие величины:

$$R_t^2 - R^2 = 0.04266; \quad R_t^2 + R^2 = 1.94063;$$

$$\delta_\Delta = 2 \cdot \sqrt{\frac{0.04266 - (0.04266)^2 \cdot (2 - 1.94063)}{6}} = 0.16841.$$

Определяем значение критерия  $T_{нел} = \frac{0.04266}{0.16841} = 0.25$ . Из неравенства (см. (2.1.22))  $0.25 < t(0.95, n-2) = 2$  следует вывод о возможности замены нелинейного уравнения тренда (2.1.23) линейным уравнением (2.1.24). ☉

Используя индекс детерминации  $R_t^2$ , можно выполнить проверку значимости построенной нелинейной регрессии. Для этого определим  $F$ -критерий:

$$F = \frac{R_t^2}{1 - R_t^2} \cdot \frac{n - k - 1}{k}, \quad (2.1.25)$$

где  $k$  – число коэффициентов уравнения тренда при переменной  $\tau$ . Тогда построенное уравнение нелинейной регрессии является значимым с уровнем значимости  $\alpha$ , если выполняется неравенство

$$F > F_{1-\alpha; k; n-k-1}. \quad (2.1.26)$$

Напомним, что квантиль  $F_{1-\alpha; k; n-k-1}$  можно вычислить в Excel с помощью выражения:

$$F_{1-\alpha; k; n-k-1} = \text{ФРАСПОБР}(\alpha; k; n - k - 1). \quad (2.1.27)$$

**Пример 2.1.3.** Определим значимость уравнения тренда  $\hat{t}(\tau) = 9.876 + 5.129 \cdot \ln(\tau)$ , используемого в примере 2.1.2.

*Решение.* Возьмем значение индекса детерминации из примера 2.1.2  $R_t^2 = 0.9916$  и вычислим значение критерия (2.1.25):

$$F = \frac{0.9916}{1 - 0.9916} \cdot (6 - 2) = 474.93.$$

Квантиль  $F_{0.95; 1; 4} = 7.70$ . Из выполнения неравенства (2.1.26):  $474.93 > 7.70$  следует вывод о *значимости построенного нелинейного тренда* с уровнем значимости  $\alpha = 0.05$ . ☉

**Замечание 2.1.2.** Индекс детерминации  $R_t^2$  характеризует близость построенного уравнения тренда к исходным данным, которые содержат «нежелательную» случайную составляющую  $\varepsilon$ . Очевидно, что, взяв полином порядка  $n-1$ , можно получить «идеальное» значение  $R_t^2 = 1$ , но такое уравнение содержит в себе не только независимую переменную  $\tau$ , но и составляющую  $\varepsilon$ , и это снижает точность использования построенного уравнения для прогноза. Поэтому при выборе уравнения регрессии надо учитывать не только величину  $R_t^2$ , но и «сложность» уравнения тренда, определяемую количеством коэффициентов уравнения. Такой учет удачно реализован в так называемом *приведенном индексе детерминации* (для линейной регрессии – *приведенный коэффициент детерминации*):

$$\hat{R}_t^2 = 1 - \frac{(n-1) \cdot Q_e}{(n-m) \cdot Q} = 1 - \frac{n-1}{n-m} \cdot (1 - R_t^2), \quad (2.1.28)$$

где  $m$  – количество коэффициентов регрессии. Видно, что при неизменных  $Q_e, Q$ , увеличение  $m$  вызывает уменьшение значения  $\hat{R}_t^2$ . Если количество коэффициентов у сравниваемых уравнений регрессии одинаково (например,  $m = 2$ ), то отбор наилучшей регрессии можно осуществлять по величине  $R_t^2$ . Если в

уравнениях регрессии меняется число коэффициентов, то отбор целесообразно проводить по величине  $\widehat{R}_t^2$ . ♦

### 2.1.5. Выделение трендовой составляющей сглаживающими методами

В отличие от регрессионных методов выделения трендовых составляющих эти методы оценивают трендовую составляющую временного ряда только для моментов времени  $\tau_i$ , т.е. для тех же моментов, для которых заданы значения временного ряда  $y_i$ ,  $i=1, \dots, n$ .

**Локальное сглаживание.** Определим оценку  $\widehat{t}_j$  для значений  $t(\tau_j)$  как взвешенное среднее тех исходных значений, которые находятся в некоторой близости от точки  $\tau_j$ , т.е.

$$\widehat{t}_j = \sum_{l=-L}^L c_l y_{j+l}, \quad j = L+1, L+2, \dots, n-L, \quad (2.1.29)$$

где  $c_l$  – весовые множители, удовлетворяющие условию

$$\sum_{l=-L}^L c_l = 1. \quad (2.1.30)$$

Видно, что суммируются  $L$  значений, находящихся левее точки  $\tau_j$ ,  $L$  значений – правее точки  $\tau_j$  и само значение  $y_j$ . Длина интервала суммирования равна  $(2L+1)$  и этот интервал «скользит» по исходным данным. Наиболее часто используют метод скользящего среднего, в котором множители задаются выражением

$$c_l = \frac{1}{2L+1}, \quad l = -L, \dots, 0, \dots, L. \quad (2.1.31)$$

Так, если  $L=1$ , то  $c_{-1} = c_0 = c_1 = 1/3$ , а сам метод скользящего среднего примет вид

$$\widehat{t}_j = \frac{1}{3}(y_{j-1} + y_j + y_{j+1}). \quad (2.1.32)$$

Очевидно, что чем больше величина  $L$ , тем меньше уровень «остаточных» возмущений в оценке  $\widehat{t}_j$ . Действительно, если  $M(\varepsilon(\tau_j)) = \sigma^2$ , то после алгоритма (2.1.32) дисперсия оценки  $\widehat{t}_j$  будет равна  $\sigma^2 / (2L+1) = \sigma^2 / 3$ . Следует помнить, что при возрастании  $L$  увеличивается систематическая ошибка.

Систематическая ошибка будет мала, если графическое изображение временного ряда напоминает прямую линию. Если же тренд имеет явно нелинейный характер, то фильтр скользящего среднего может привести к значительным искажениям (т.е. к большой систематической ошибке). В таких случаях предпочтительнее использовать метод экспоненциального сглаживания, описанный ниже.

К сожалению, выражение (2.1.29) не определяет «отфильтрованные» значения в первых и последних  $L$ -точках временного ряда. В этих случаях можно изменить алгоритм (2.1.29), используя под знаком суммирования только известные  $y_j$ . Например, в

точке  $\tau_1$  алгоритм (2.1.32) имеет вид  $\widehat{t}_1 = \frac{1}{2}(y_1 + y_2)$ , а в точке  $\tau_n$

определяется выражением  $\widehat{t}_n = \frac{1}{2}(y_{n-1} + y_n)$ .

**Пример 2.1.4.** По данным табл. 2.1, отражающим спрос (в условных единицах) на некоторый товар вычислить значение тренда в точках  $\tau_j$ ,  $j=1, 2, \dots, 8$ , используя алгоритм (2.1.32), т.е.  $L=1$ .

*Решение.* Фрагмент документа Excel, вычисляющий значения  $\widehat{t}_j$  по формуле (2.1.32) приведен на рис. 2.3, сами значения нанесены на рис. 2.2 (кривая 3, маркированная треугольниками, исходные значения – кривая 1). Сравнивая эти значения со значением оценки  $\widehat{t}(\tau) = 132.3 + 55.09\tau - 3.26\tau^2$  (кривая 2), построенной регрессионными методами в примере 2.1.1, видим некоторые отличия, которые можно объяснить использованием разных методов для выделения тренда временного ряда.

	A	B	C	D	E	F
1	Год	Спрос	$\hat{t}$			
2	1	213	192	←	=(B2+B3)/2	
3	2	171	225			
4	3	291	257			
5	4	309	305,67	←	=(B4+B5+B6)/3	
6	5	317	329,33			
7	6	362	343,33			
8	7	351	358			
9	8	361	356	←	=(B8+B9)/2	

Рис. 2.3. Выделение тренда методом скользящего среднего

**Экспоненциальное сглаживание.** В отличие от метода скользящего среднего в определении экспоненциальной средней участвуют все наблюдения исходного временного ряда, но с разными весовыми множителями. Алгоритм метода экспоненциального сглаживания определяется выражением

$$\hat{t}_j = (1 - \alpha)\hat{t}_{j-1} + \alpha y_j, \quad j = 1, 2, \dots, n, \quad (2.1.33)$$

где  $\alpha$  – коэффициент экспоненциального сглаживания, задаваемый как  $0 < \alpha < 1$ .

Можно доказать справедливость выражения

$$\hat{t}_j = \alpha \cdot \sum_{i=0}^{j-1} (1 - \alpha)^i y_{j-i}, \quad (2.1.34)$$

из которого следует, что каждое «старое» измерение  $y_i$  входит в оценку  $\hat{t}_j$  ( $i \leq j$ ) с весом  $\alpha(1 - \alpha)^i$ , т.е. по мере удаления от точки  $\tau_j$  вес измерения  $y_i$  уменьшается. В качестве начального значения  $\hat{t}_0$  может быть принято среднее арифметическое всей временной выборки или только ее части, например, среднее трех первых значений.

Значение  $\hat{t}_j$  можно рассматривать как прогнозное значение тренда в момент  $\tau_j$  и представить как

$$\hat{t}_j = \hat{t}_{j-1} + \alpha(y_j - \hat{t}_{j-1}). \quad (2.1.35)$$

Из этого выражения видно, что прогноз в момент  $\tau_j$  состоит из двух слагаемых: прогнозного значения  $\hat{t}_{j-1}$  в предыдущий момент и невязки (ошибки) прогнозирования  $y_j - \hat{t}_{j-1}$ , взятой с весом  $\alpha$ .

Из выражения (2.1.34) видно, что, уменьшая величину  $\alpha$ , увеличивается степень сглаживания (за счет увеличения числа «значимых» слагаемых). Рекомендуется  $\alpha$  определять по формуле

$$\alpha = \frac{2}{n+1}, \quad (2.1.36)$$

где  $n$  – объем выборки.

**Пример 2.1.5.** По данным табл. 2.1, отражающим спрос (в условных единицах) на некоторый товар, вычислить значение тренда в точках  $\tau_j$ ,  $j = 1, 2, \dots, 8$ , используя алгоритм (2.1.33) при  $\alpha = 0.2, 0.5$ .

*Решение.* Фрагмент документа Excel, вычисляющий значения  $\hat{t}_j$  по формуле (2.1.33) при  $\alpha = 0.5$ , приведен на рис. 2.4. В качестве  $\hat{t}_0$  принято среднее значение  $\bar{y}$  (ячейка B10). На рис. 2.5 приведены графики:

- кривая 1 – исходные значения  $y_i$  (маркированная квадратиками);
- кривая 2 – значения  $\hat{t}_j$ , вычисленные при  $\alpha = 0.5$  (маркированная треугольниками);
- кривая 3 – значения  $\hat{t}_j$ , вычисленные при  $\alpha = 0.2$  (маркированная крестиками).

Видим, что меньшему значению параметра  $\alpha$  соответствует большая степень сглаживания исходных значений  $y_i$ .

	A	B	C	D	E	F	G
1	Год	Спрос	$\hat{t}$	$\alpha =$	0,5		
2	1	213	254,938		← $= (1-\$E\$1)*B10+\$E\$1*B2$		
3	2	171	212,969				
4	3	291	251,984		← $= (1-\$E\$1)*C2+\$E\$1*B3$		
5	4	309	280,492				
6	5	317	298,746				
7	6	362	330,373				
8	7	351	340,687				
9	8	361	350,843		← $= (1-\$E\$1)*C8+\$E\$1*B9$		
10	$\bar{y} =$	296,875					

Рис. 2.4. Выделение тренда методом экспоненциального сглаживания

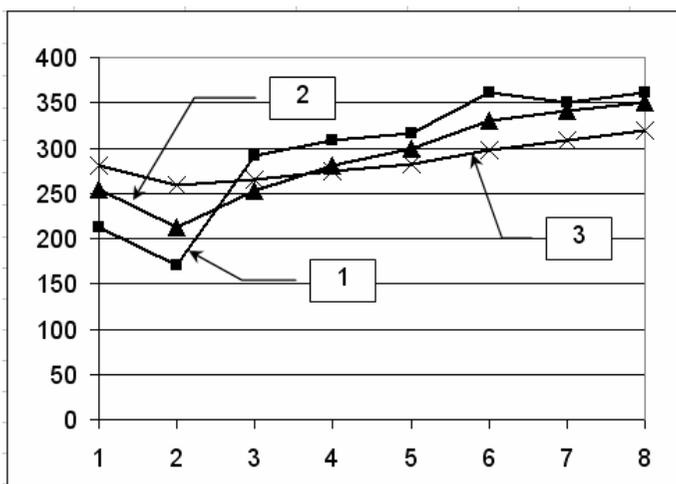


Рис. 2.5. Результаты экспоненциального сглаживания при различных значениях  $\alpha = 0.2, 0.5$

Проанализировав примеры выделения тренда разными методами можно сделать вывод, что: *наиболее эффективным является метод, основанный на построении парной регрессии*. Этот метод достаточно универсален, позволяет непосредственно решать

задачи прогнозирования, лишен недостатка, присущего методам сглаживания при вычислении значений на концах временного интервала.

## 2.2. Выделение трендовой составляющей с помощью табличного процессора Excel

В п. 2.1 расчеты, необходимые для выделения трендовой составляющей осуществлялись в Excel путем программирования соответствующих выражений. Автоматизация таких вычислений заключается в том, что табличный процессор Excel позволяет реализовать все рассмотренные в п. 2.1 методы выделения трендовой составляющей с использованием следующих операций:

- команды *Добавить линию тренда*;
- команды *Поиск решения*;
- режима *Скользящее среднее* модуля *Анализ данных*;
- режима *Экспоненциальное сглаживание* модуля *Анализ данных*.

Кратко рассмотрим эти возможности Excel.

### 2.2.1. Выделение трендовой составляющей регрессионными методами

Выделить тренд (т.е. вычислить коэффициенты  $b_0, b_1, \dots, b_k$  уравнения регрессии) с помощью табличного процессора Excel можно используя команду *Добавить линию тренда*; используя команду *Поиск решения*.

Команда *Добавить линию тренда*. При использовании этой команды переменная  $x$  играет роль временного аргумента  $\tau$ .

Эта команда позволяет построить следующие регрессии:

- линейную  $\hat{t} = b_0 + b_1x$ ;
- полиномиальную  $\hat{t} = b_0 + b_1x + \dots + b_kx^k$  ( $k \leq 6$ );
- логарифмическую  $\hat{t} = b_0 + b_1 \ln x$ ;
- степенную  $\hat{t} = b_0x^{b_1}$ ;
- экспоненциальную  $\hat{t} = b_0e^{b_1x}$ .

Для построения одной из перечисленных регрессий необходимо выполнить следующие шаги:

*Шаг 1.* В выбранном листе Excel ввести по столбцам исходные данные  $\{x_i, y_i\}, i = 1, 2, \dots, n$  (рис. 2.6).



Рис. 2.6. Построение графика по исходным данным

*Шаг 2.* По этим данным построить график в декартовой системе координат (см. рис. 2.6).

*Шаг 3.* Установить курсор на построенном графике, сделать щелчок правой кнопкой и в появившемся контекстном меню выполнить команду *Добавить линию тренда* (см. рис. 2.6).

*Шаг 4.* В появившемся диалоговом окне (рис. 2.7) активизировать закладку **Тип** и выбрать нужное уравнение регрессии.

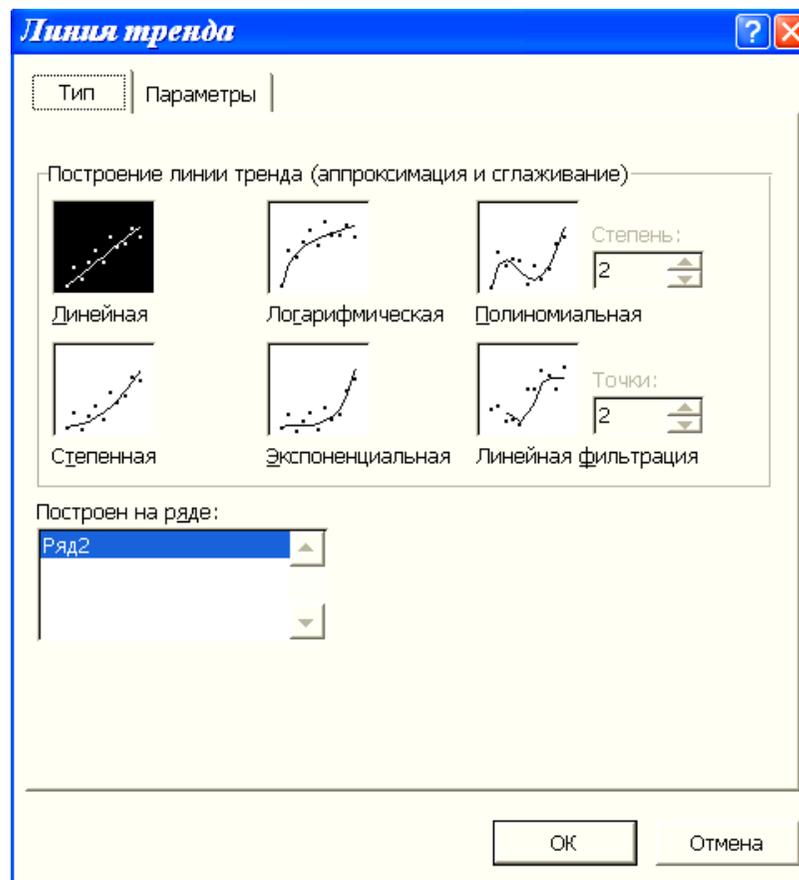


Рис. 2.7. Выбор вида уравнения регрессии

*Шаг 5.* Активизировать закладку **Параметры** (рис. 2.8) и «включить» необходимые для нас опции:

*Показать уравнение на диаграмме* – на диаграмме будет показано выбранное уравнение тренда с вычисленными коэффициентами;

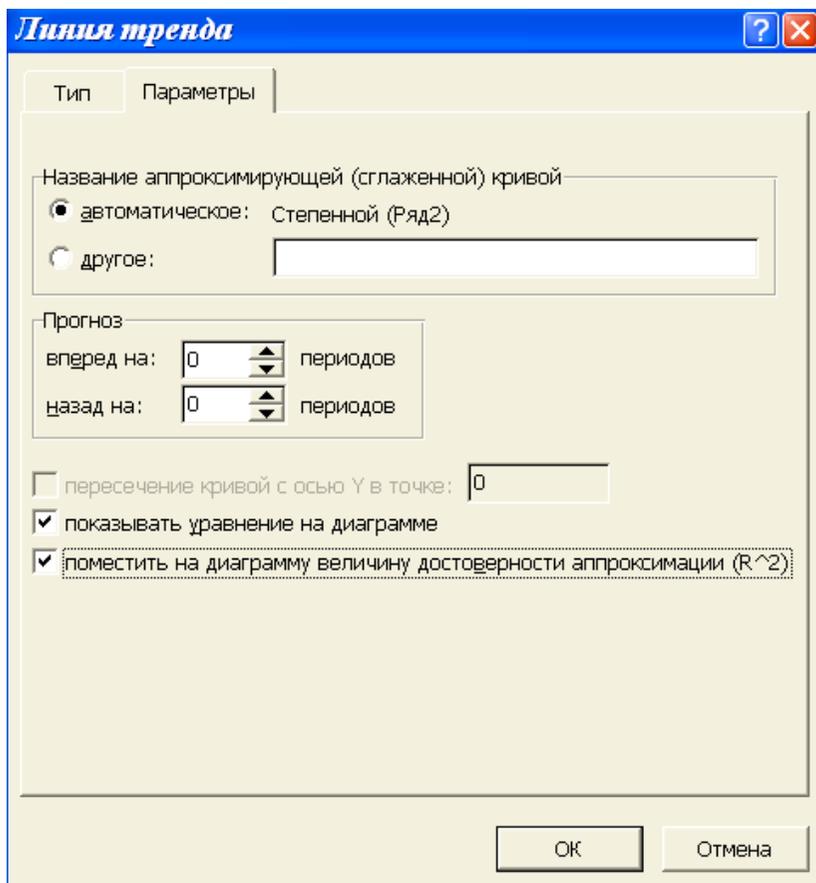


Рис. 2.8. Задание опций вывода информации

Поместить на диаграмму величину достоверности аппроксимации ( $R^2$ ) – на диаграмме будет показано значение индекса детерминации  $R_t^2$  (см. (2.1.18)), которое можно использовать для проверки значимости построенной регрессии с помощью  $F$ -критерия (2.1.25). Если по построенному уравнению регрессии необходимо выполнить прогноз, то нужно указать число периодов прогноза (см. рис. 2.8).

Назначение других опций понятно из их названий.

*Шаг 6.* После задания всех перечисленных опций щелкнуть ОК и на диаграмме появятся формула построенного уравнения регрессии и значение индекса детерминации  $R_t^2$ .

**Пример 2.2.1.** Используя команду *Добавить линию тренда*, по данным табл. 2.1 найти оценку  $\hat{t}(\tau)$ , предполагая, что  $t(\tau)$  является квадратичной функцией.

*Решение.* При сделанном предположении оценку  $\hat{t}(\tau)$  будем искать в виде

$$\hat{t}(\tau) = b_0 + b_1\tau + b_2\tau^2.$$

Первоначально в документ Excel вводим данные из табл. 4.1 (рис. 2.9, кривая 1). Затем по этим данным строим график и вызываем команду *Добавить линию тренда*. В появившемся диалоговом окне зададим необходимые параметры и щелкнем на ОК. На экране появится график  $\hat{t}(\tau)$  (кривая 2), уравнение функции  $\hat{t}(\tau)$  и значение индекса детерминации (2.1.18), равное 0.849. Заметим, что коэффициенты  $b_0 = 132.3$ ,  $b_1 = 55.09$ ,  $b_2 = -3.26$  совпадают с коэффициентами, вычисленными в примере 2.1.1. ●

**Пример 2.2.2.** По данным табл. 2.2 построить уравнения регрессии (предусмотренные командой *Добавить линию тренда*) и по значению индекса детерминации  $R_t^2$  выбрать наилучшее уравнение.

Таблица 2.2

$x_i$	1	2	3	4	5	6
$y_i$	10	13.4	15.4	16.5	18.6	19.1

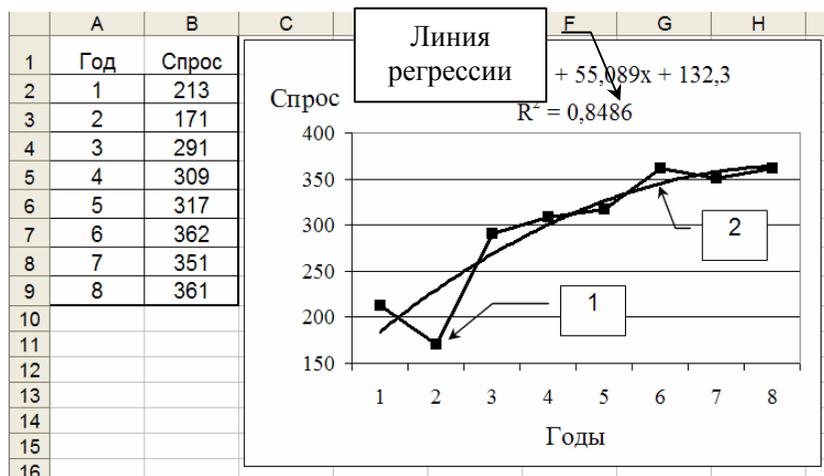


Рис. 2.9. Построение уравнения тренда (пример 2.2.1)

*Решение.* Построение каждого из пяти уравнений осуществляем по описанным выше шагам. Для уравнения  $\hat{y} = b_0 \cdot x^{b_1}$  выполнение шагов иллюстрируют рис. 2.6–2.8. В табл. 2.3 заносим регрессионное уравнение и соответствующее значение  $R_t^2$ . Сравнивая величину индекса детерминации  $R_t^2$  для этих уравнений, в качестве «наилучшего» уравнения выбираем степенную регрессию  $\hat{y} = 10,18x^{0,3226}$  (номер 5), для которой индекс детерминации  $R_t^2 = 0,9921$ .

Таблица 2.3

№	Уравнение	$R_t^2$	$\bar{R}_t^2$
1	2	3	4
1	$\hat{t} = 9,28 + 1,777x$	0,949	0,938
2	$\hat{t} = 9,8759 + 5,1289 \cdot \ln x$	0,9916	0,9895
3	$\hat{t} = 6,93 + 3,5396x - 0,2518x^2$ (полиномиальная, $m = 3$ )	0,9896	0,9827

Окончание табл. 2.3

1	2	3	4
4	$\hat{t} = 5,8333 + 4,9192x - 0,7087x^2 - 0,0435x^3$ (полиномиальная, $m = 4$ )	0,9917	0,9792
5	$\hat{t} = 10,18x^{0,3226}$	0,9921	0,9901
6	$\hat{t} = 9,8675 \cdot e^{0,1225x}$	0,9029	0,8786

В этой же таблице приведены значения *приведенного индекса детерминации*, определяемого по формуле (2.1.28). Видно, что по величине  $\bar{R}_t^2$  наилучшей регрессией также является степенная регрессия. Полиномиальная регрессия третьей степени имеет  $\bar{R}_t^2$  значительно меньше коэффициента  $R_t^2$ . ☹

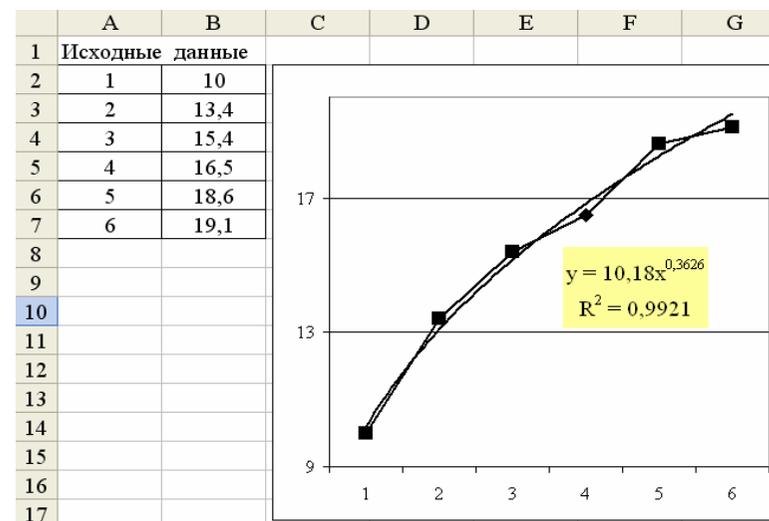


Рис. 2.10. График и уравнение построенной регрессии (пример 2.2.2)

Команда *Поиск решения* (пункт меню **Сервис**). Используется для вычисления параметров (коэффициентов) при которых некоторый функционал, зависящий от этих параметров, достигает наименьшего или наибольшего значения. Эта команда позволяет также решать задачи *условной оптимизации*, т.е. когда ищется минимум или максимум функционала с учетом дополнительных ограничений (линейных или нелинейных) на значения искомых параметров. Например, искомый параметр  $b$  должен удовлетворять ограничению  $0.2 \leq b < 1$ . Эта возможность обуславливает существенное преимущество рассматриваемого подхода по сравнению с командой *Добавить линию тренда*. К недостатку следует отнести необходимость программировать «вручную» вычисление индекса детерминации  $R_i^2$ .

Применение команды *Поиск решения* для вычисления коэффициентов нелинейной регрессии на основе метода наименьших квадратов покажем на следующем примере.

**Пример 2.2.3.** По данным табл. 2.2 построить уравнения степенной регрессии, используя команду *Поиск решения*.

*Решение.* Первоначально на листе Excel введем исходные данные: значения  $\tau_i$  в ячейках A2÷A7; значения  $y_i$  в ячейках B2÷B7. Затем в ячейку B9 введем произвольное значение коэффициента  $b_0$ , а в ячейку B10 – произвольное значение коэффициента  $b_1$ . На рис. 2.11 показан фрагмент документа Excel с введенными данными.

Следующим шагом является вычисление по уравнению регрессии значений  $\hat{y}_i = b_0 \cdot x_i^{b_1}$ ,  $i = 1, \dots, 6$ . Так, для вычисления значения  $\hat{y}_1$  в ячейке C2 программируется выражение  $=B\$9*A2^{\$B\$10}$ . Использование абсолютных адресов для ячеек B9, B10 позволяет «размножить» это выражение на ячейки C3÷C7. Далее в ячейках D2÷D7 вычисляется квадрат невязки при соответствующем значении  $x_i$ . Так, в ячейку D2 вводится выражение  $=(C2-B2)^2$ , «размножаемое» в ячейках D3÷D7. Значение минимизируемого функционала МНК вычисляется в ячейке D9

(см. рис. 2.11). На этом подготовка необходимой для команды *Поиск решения* информации завершается.

Для выполнения команды *Поиск решения* необходимо обратиться к пункту основного меню **Сервис** и в появившемся меню щелкнуть мышью на команде *Поиск решения*.

	A	B	C	D	E	F
1	Исходные данные		$\hat{y}_i$	$(\hat{y}_i - y_i)^2$		
2	1	10	1	81,000		
3	2	13,4	1,231	148,081		
4	3	15,4	1,390	196,269		
5	4	16,5	1,516	224,529		
6	5	18,6	1,621	288,298		
7	6	19,1	1,712	302,351		
8						
9	$b_0$	1	$F(b_0, b_1)$	1240,528		
10	$b_1$	0,3		=СУММ(D2:D7)		
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						

Рис. 2.11. Задание параметров команды *Поиск решения*

Затем в появившемся диалоговом окне выполнить следующие действия (см. рис. 2.11):

- в поле ввода *Установить целевую ячейку*: ввести адрес ячейки, в которой вычисляется значение минимизируемого функционала (в нашем примере D9);
- включить опцию *Равной: минимальному значению*: (ищутся значения коэффициентов, при которых функционал достигает своего минимального значения);
- в поле ввода *Изменяя значения*: ввести адреса ячеек, в которых находятся значения искомых коэффициентов (в нашем примере это ячейки B9, B10);
- щелкнув мышью на кнопке *Добавить*, сформировать ограничения на значения искомых коэффициентов (в нашем примере это требование неотрицательности искомых коэффициентов).

	A	B	C	D	E
1	Исходные данные		$\hat{y}_i$	$(\hat{y}_i - y_i)^2$	
2	1	10	10,28299	0,080	
3	2	13,4	13,147	0,064	
4	3	15,4	15,179	0,049	
5	4	16,5	16,809	0,096	
6	5	18,6	18,193	0,166	
7	6	19,1	19,408	0,095	
8					
9	$b_0$	10,28299	$F(b_0, b_1)$	0,549	
10	$b_1$	0,354496		=СУММ(D2:D7)	

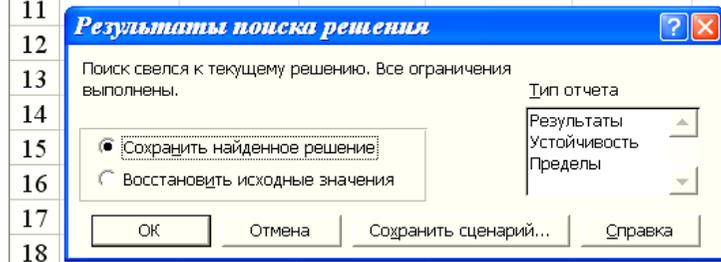


Рис. 2.12. Результаты выполнения команды *Поиск решения*

После выполнения этих операций щелкнуть на кнопке *Выполнить*. Начинается поиск решения введенной оптимизационной задачи. Спустя некоторое время на экране появится новое диалоговое окно *Результаты поиска решения* (рис. 2.12). Для сохранения найденных значений коэффициентов в соответствующих ячейках необходимо включить опцию *Сохранить найденное решение* и щелкнуть на кнопке ОК.

Из рис. 2.12 видно, что вычисленные значения коэффициентов находятся в ячейках B9, B10 и равны  $b_0 = 10.28299$ ,  $b_1 = 0.354496$ . Ячейка D9 содержит значение минимизируемого функционала. Найденные значения коэффициентов незначительно отличаются от значений, вычисленных в примере 2.2.2 с помощью команды *Добавить линию тренда*. ☉

### 2.2.2. Выделение трендовой составляющей с использованием режима *Скользящее среднее*

В режиме *Скользящее среднее* реализуется алгоритм

$$\hat{t}_j = \frac{1}{2L+1} \cdot \sum_{l=-2L}^0 y_{j+l}, \quad j = 2L+1, 2L+2, \dots, n \quad (2.2.1)$$

с одинаковыми весами  $c_l = \frac{1}{2L+1}$ . Для вызова режима *Скользящее среднее* необходимо обратиться к пункту меню **Сервис**, команде *Анализ данных*, выбрать в списке режимов *Скользящее среднее* и щелкнуть на кнопке ОК.

В появившемся диалоговом окне *Скользящее среднее* задать следующие параметры (рис. 2.13):

*Входной интервал*: диапазон адресов ячеек, содержащих значения  $y_i$ .

*Метки в первой строке*: включается, если первая строка во входном интервале содержит заголовки. В этом случае автоматически будут созданы стандартные названия.

*Интервал*: задается размер окна  $(2L+1)$ . По умолчанию размер 3, т.е.  $L = 1$ .

*Выходной интервал: Новый рабочий лист: : / Новая рабочая книга:* содержит адрес верхней ячейки, начиная с которой выводятся вычисленные сглаженные значения.

*Вывод графика* – включает вывод графика заданных и сглаженных значений временного ряда.

*Стандартные погрешности* – включает вычисление и вывод в виде столбца стандартных погрешностей  $\sigma_{\hat{t}_j}$ , которые вычисляются по формуле

$$\sigma_{\hat{t}_j} = \sqrt{\frac{1}{2L+1} \cdot \sum_{i=-2L}^0 (y_{j+i} - \hat{t}_{j+i})^2}. \quad (2.2.2)$$

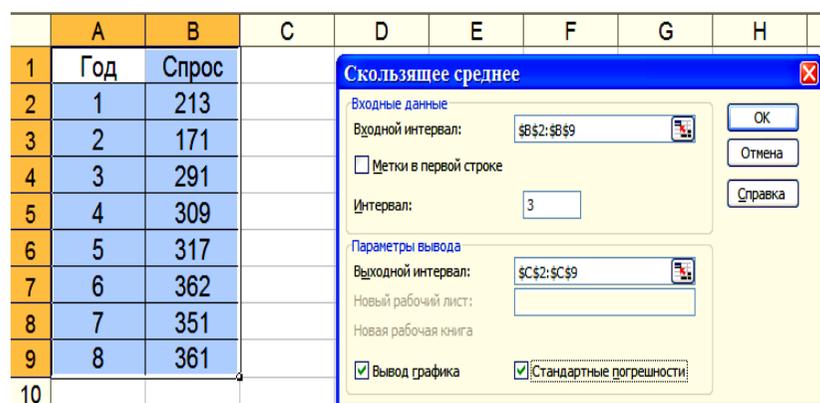


Рис. 2.13. Диалоговое окно режима *Скользящее среднее*

**Пример 2.2.4.** По данным табл. 2.1 вычислить значения тренда  $\hat{t}(\tau_j)$ , используя режим *Скользящее среднее*.

*Решение.* Введем в документ Excel исходные данные (рис. 2.13), а затем вызовем режим *Скользящее среднее* и зададим необходимые параметры (см. рис. 2.13). На рис. 2.14 показаны графики значений  $y_j$  (маркированные ромбами) и  $\hat{t}_j$  (маркированные квадратами). ♦

**Замечание 2.2.1.** Из формулы (2.2.1) следует, что в первых  $2L$  ячейках значения  $\hat{t}_j$  не определены, а из выражения (2.2.2) следует, что значения  $\sigma_{\hat{t}_j}$  не вычисляются в первых  $4L$  ячейках. ♦

Пример 2.2.4 хорошо иллюстрирует это замечание. Так в ячейках C2, C3 документа на рис. 2.14 не определены сглаженные значения, в ячейках D2=D5 не определены значения стандартных погрешностей  $\sigma_{\hat{t}_j}$ .



Рис. 2.14. Результаты работы режима *Скользящее среднее*

**Задание.** Вычислите значение тренда, задав  $(2L+1)=5$ , сравните с предыдущими результатами.

### 2.2.3. Выделение трендовой составляющей с использованием режим Экспоненциальное сглаживание

Режим *Экспоненциальное сглаживание* реализует алгоритм (2.1.33). Для вызова режима необходимо обратиться к пункту основного меню **Сервис**, выполнить команду *Анализ данных*, а затем в появившемся списке режимов работы выделить *Экспоненциальное сглаживание* и щелкнуть на ОК. В появившемся диалоговом окне (см. рис. 2.15) задать необходимые параметры. Параметры режима *Экспоненциальное сглаживание* совпадают с параметрами режима *Скользящее среднее* за исключением одного параметра. Вместо параметра *Интервал* необходимо задать *Фактор затухания*, равный величине  $\alpha$  в формуле (2.1.33), который может меняться в интервале (0,1) (см. соотношение (2.1.36)).

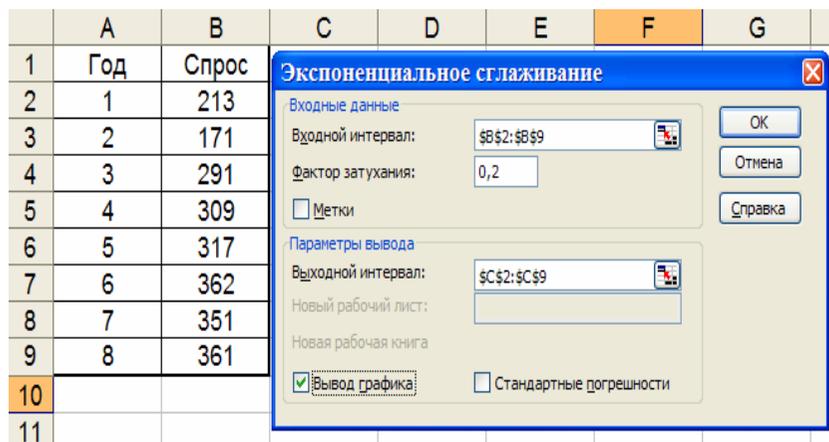


Рис. 2.15. Диалоговое окно режима *Экспоненциальное сглаживание*

**Пример 2.2.5.** По данным табл. 2.1 вычислить значение тренда  $\hat{t}_j = \hat{t}(\tau_j)$ , используя режим *Экспоненциальное сглаживание*.

**Решение.** Введем в документ Excel исходные данные (рис. 2.16), вызовем режим *Экспоненциальное сглаживание* и

зададим необходимые параметры ( $\alpha = 0.2$ ) (см. рис. 2.15). На рис. 2.16 показаны графики значений  $y_j$  (маркированные ромбами) и  $\hat{t}_j$  (маркированные квадратами).

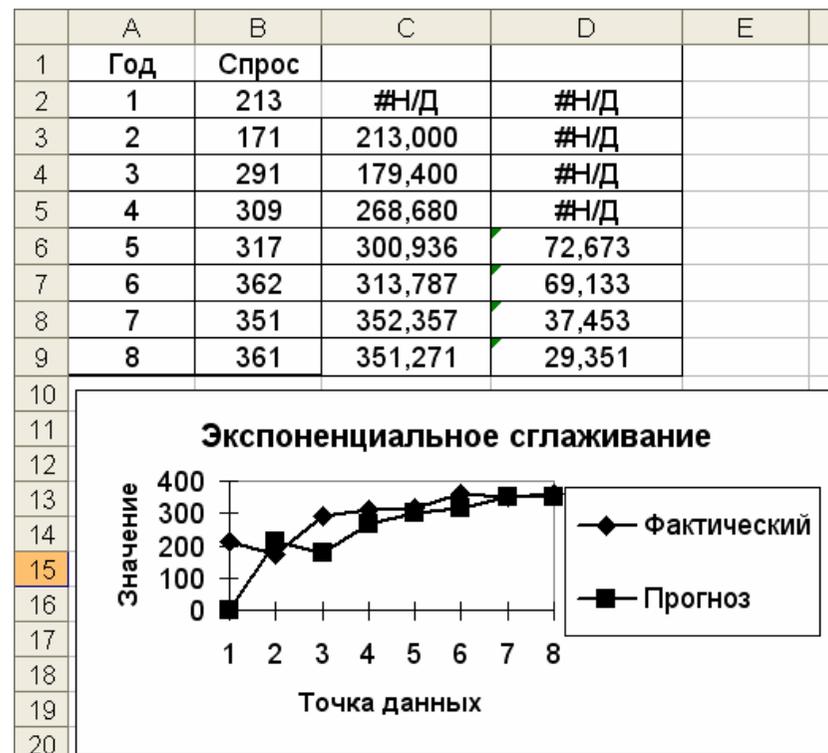


Рис. 2.16. Результаты работы режима *Экспоненциальное сглаживание*

Видно, что:

- первое значение  $\hat{t}_1$  не вычисляется (ячейка C2);
- чем меньше значение  $\alpha$ , тем больше степень сглаживания значений  $y_i$  временной выборки.

В столбце D (ячейки D5÷D8) содержатся значения  $\sigma_{\hat{t}_j}$  – среднеквадратические отклонения значений  $\hat{t}_j$ , которые вычисляются по формуле

$$\sigma_{\hat{t}_j} = \sqrt{\frac{\sum_{i=-2}^0 (y_{j+i} - \hat{t}_{j+i})^2}{3}}. \quad (2.2.3)$$

**Задание.** Вычислите значение тренда, задав  $\alpha = 0.6$ . Сравните полученные результаты с результатами примера и объясните отличия.

### 2.3. Выделение тригонометрической составляющей временного ряда

К тригонометрической составляющей  $c(\tau)$  временного ряда относятся:

- *сезонная составляющая*  $s(t)$ , отражающая повторяемость экономических процессов в течение не очень длительного периода (года, иногда месяца);
- *периодическая составляющая*  $p(\tau)$ , отражающая повторяемость экономических процессов в течение длительных периодов (например, волны экономической активности Кондратьева).

Решение о наличии тригонометрической составляющей в структуре временного ряда принимается на основе анализа автокорреляционной функции и коррелограммы временного ряда (см. п. 1.2).

Для выделения и анализа тригонометрической составляющей временного ряда используются методы гармонического анализа периодических функций.

#### 2.3.1. Основы гармонического анализа периодических функций

Гармонический анализ позволяет представить периодическую функцию линейной комбинацией косинусов и синусов.

Предположим, что функция  $\varphi(\tau)$  является непрерывной функцией с периодом  $T$ . Тогда функцию  $\varphi(\tau)$  можно представить рядом Фурье вида

$$\varphi(\tau) = a_0 + \sum_{k=1}^{\infty} [a_k \cos(\frac{2\pi k}{T}\tau) + b_k \sin(\frac{2\pi k}{T}\tau)], \quad 0 \leq \tau \leq T, \quad (2.3.1)$$

где  $k$  – номер гармоники. Видно, что при увеличении номера уменьшается период функций  $\cos(\frac{2\pi k}{T}\tau)$ ,  $\sin(\frac{2\pi k}{T}\tau)$ . Коэффициенты разложения определяются формулами:

$$\begin{aligned} a_0 &= \frac{1}{T} \int_0^T \varphi(\tau) d\tau; \\ a_k &= \frac{2}{T} \int_0^T \varphi(\tau) \cos(\frac{2\pi k}{T}\tau) d\tau; \\ b_k &= \frac{2}{T} \int_0^T \varphi(\tau) \sin(\frac{2\pi k}{T}\tau) d\tau. \end{aligned}$$

Аргументы тригонометрических функций  $\cos$ ,  $\sin$  можно трактовать как частоты  $\omega_k$ , определяемые соответствующим номером гармоники, т.е.

$$\omega_k = \frac{2\pi}{T} k. \quad (2.3.2)$$

Величины  $S_k = a_k^2 + b_k^2$  характеризуют «энергетический вклад»  $k$ -й гармоники в функцию  $\varphi(\tau)$ . Зависимость величины  $S_k$  от номера гармоники  $k$  (или от частоты  $\omega_k$  (2.3.2)) характеризует спектральный состав (или спектр) функции  $\varphi(\tau)$ . Сравнительно большие величины  $S_k$  определяют частоты, на которых сосредоточена основная энергия функции  $\varphi(\tau)$ .

Под аппроксимацией функции  $\varphi(\tau)$  рядом Фурье понимают новую функцию  $\hat{\varphi}(\tau)$ , полученную суммированием первых членов ряда (2.3.1), число которых обозначим  $K_0$ , т.е.

$$\widehat{\varphi}(\tau) = a_0 + \sum_{k=1}^{K_0} [a_k \cos(\frac{2\pi k}{T} \tau) + b_k \sin(\frac{2\pi k}{T} \tau)]. \quad (2.3.3)$$

Видно, что в функции  $\widehat{\varphi}(\tau)$  отсутствуют «высокочастотные» гармоники с номерами  $k > K_0$ , которые присутствовали в исходной функции  $\varphi(\tau)$ . Такой способ получения функции  $\widehat{\varphi}(\tau)$  часто называют низкочастотной фильтрацией функции  $\varphi(\tau)$ .

По аналогии можно построить новую функцию  $\widehat{\varphi}(\tau)$ , содержащую только заданные гармоники, например, гармоники с наиболее значимым спектром  $S_k$ . Предположим, что такие гармоники имеют номера  $k = 3, 8$ . Тогда функция  $\widehat{\varphi}(\tau)$ , содержащая только эти гармоники, записывается в виде

$$\begin{aligned} \widehat{\varphi}(\tau) = & a_3 \cos(\frac{2\pi 3}{T} \tau) + b_3 \sin(\frac{2\pi 3}{T} \tau) + \\ & + a_8 \cos(\frac{2\pi 8}{T} \tau) + b_8 \sin(\frac{2\pi 8}{T} \tau). \end{aligned}$$

Такой способ построения функции широко используется для выделения тригонометрической составляющей временного ряда.

**Пример 2.3.1.** Дана функция

$$\varphi(\tau) = 0.1 + 0.4\tau + 0.5\tau^2 + 3\sin(\frac{2\pi}{3.2} 5\tau), \quad (2.3.4)$$

определенная на интервале  $[0, 3.2]$ . График функции показан сплошной линией на рис. 2.17. Необходимо вычислить спектр  $S_k$  этой функции и выделить из функции  $\varphi(\tau)$  основную (имеющую наибольшее значение спектра) тригонометрическую составляющую.

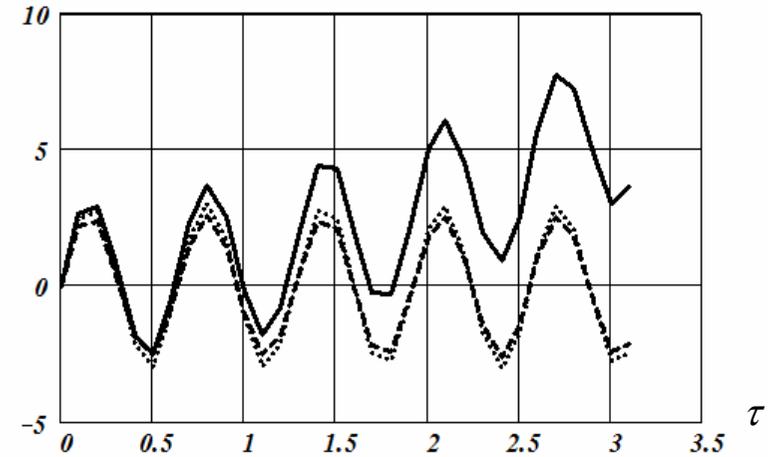


Рис. 2.17. Тригонометрическая составляющая функции  $\varphi(\tau)$

*Решение.* Из аналитического задания  $\varphi(\tau)$  (2.3.4) следует, что тригонометрическая составляющая этой функции обусловлена слагаемым  $3\sin(\frac{2\pi}{3.2} \cdot 5\tau)$  и соответствует гармонике с номером 5. Так как функция задана на интервале  $[0, 3.2]$ , то период этой функции задаем  $T = 3.2$ . Используя приведенные выше формулы, вычисляем коэффициенты  $a_0, a_k, b_k, k = 1, \dots, 20$  и определяем спектры  $S_0 = a_0^2, S_k = a_k^2 + b_k^2$ . Значения  $S_k$  приведены на рис. 2.18 (кривая 1). Большие значения  $S_0, S_1$  обусловлены наличием в функции  $\varphi(\tau)$  квадратичного тренда (первые три слагаемых в (2.3.4)), большое значение  $S_5$  обусловлено присутствием в  $\varphi(\tau)$  тригонометрической составляющей, для которой вычислены коэффициенты  $a_5 = 0.021, b_5 = 2.593$ . Построим функцию  $\widehat{\varphi}(\tau) = 0.021 \cdot \cos(\frac{2\pi}{3.2} 5\tau) + 2.593 \cdot \sin(\frac{2\pi}{3.2} 5\tau)$ , которая соответствует этой гармонике. График этой функции приведен на рис. 2.17 (штриховая кривая), здесь же приведен график функции

$3\sin(\frac{2\pi}{3.2} \cdot 5\tau)$  (точечная кривая), которая входит в исходную функцию  $\varphi(\tau)$ . Из хорошего совпадения этих графиком можно сделать вывод об эффективности применения методов гармонического анализа для выделения тригонометрических составляющих периодических функций. ☺

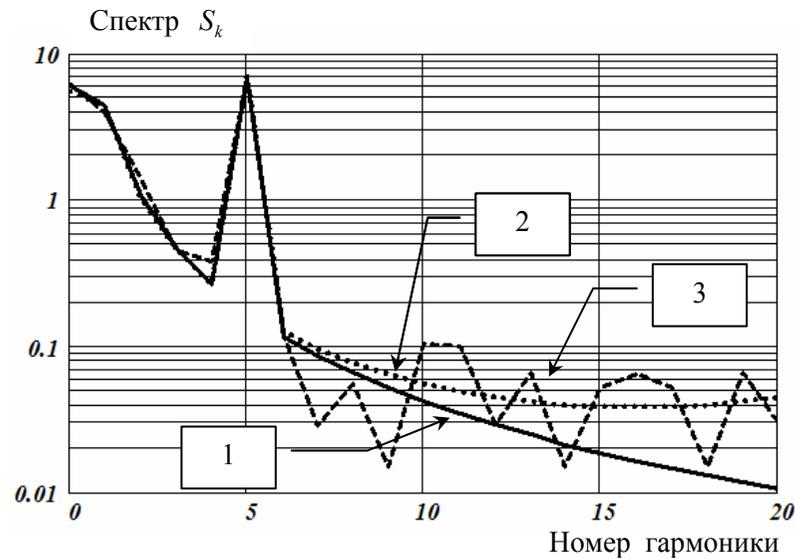


Рис. 2.18. Спектральный состав функции  $\varphi(\tau)$

### 2.3.2. Выделение тригонометрической составляющей временного ряда методами гармонического анализа

Использование рядов Фурье для выделения тригонометрической составляющей временного ряда отличается от рассмотренного выше следующим:

1. Значения временного ряда заданы в дискретные моменты времени  $\tau_i$  и чаще всего эти моменты представляют собой арифметическую прогрессию с шагом  $\Delta_\tau$ , т.е.

$$\tau_i = \tau_{нач} + (i-1) \cdot \Delta_\tau, \quad i=1,2,\dots,n. \quad (2.3.5)$$

Тогда в качестве периода принимается величина

$$T = \Delta_\tau \cdot n, \quad (2.3.6)$$

а условие периодичности значений временного ряда имеет вид

$$y_{n+i} = y_i, \quad i=1,\dots,n.$$

В дальнейшем полагается, что  $\tau_i$  образуют арифметическую прогрессию (2.3.5) при  $\tau_{нач} = 0$ .

2. Временной ряд  $Y(\tau)$ , кроме тригонометрической составляющей, содержит случайную составляющую  $\varepsilon(\tau)$ , которую необходимо отделить от тригонометрической составляющей.

Первое отличие обуславливает замену интегралов, определяющих значения  $a_0, a_k, b_k$  квадратурными формулами, в которые входят значения  $y_i, i=1,2,\dots,n$ . В качестве примера примем формулу левых прямоугольников и тогда получим следующие выражения для вычисления интегралов:

$$a_0^* = \frac{1}{n} \cdot \sum_{i=1}^n y_i; \quad (2.3.7)$$

$$a_k^* = \frac{2}{n} \cdot \sum_{i=1}^n y_i \cos\left(\frac{2\pi k}{T} \tau_i\right); \quad b_k^* = \frac{2}{n} \cdot \sum_{i=1}^n y_i \sin\left(\frac{2\pi k}{T} \tau_i\right).$$

Символ «\*» означает, что коэффициенты вычислены по дискретным значениям временного ряда, или иначе – по временной выборке.

Если период  $T = n \cdot \Delta_\tau$ , то две последние формулы принимают вид

$$a_k^* = \frac{2}{n} \cdot \sum_{i=1}^n y_i \cos\left(\frac{2\pi k}{n} i\right), \quad (2.3.8)$$

$$b_k^* = \frac{2}{n} \cdot \sum_{i=1}^n y_i \sin\left(\frac{2\pi k}{n} i\right). \quad (2.3.9)$$

По аналогии со спектром  $S_k$  определим дискретный спектр как  $S_k^* = (a_k^*)^2 + (b_k^*)^2$ . Дискретность задания значений  $y_i$  обуславливает симметричность спектра  $S_k^*$  относительно точки  $n/2$ , т.е.  $S_{n/2+j}^* = S_{n/2-j}^*$ ,  $j = 1, \dots, n/2 - 1$ . Поэтому имеет смысл вычислить коэффициенты ряда Фурье для гармоник с номерами  $k = 0, 1, 2, \dots, n/2$ , а саму функцию  $p(\tau)$  аппроксимировать рядом

$$\hat{c}(\tau) = a_0^* + \sum_{k=1}^{\frac{n-1}{2}} a_k^* \cdot \cos\left(\frac{2\pi k}{T} \tau\right) + \sum_{k=1}^{\frac{n-1}{2}} b_k^* \cdot \sin\left(\frac{2\pi k}{T} \tau\right) + a_{\frac{n}{2}}^*, \quad (2.3.10)$$

где

$$a_{\frac{n}{2}}^* = \frac{1}{n} \cdot \sum_{i=1}^n (-1)^i y_i. \quad (2.3.11)$$

Учет второго отмеченного момента основан на следующем предположении: амплитуда случайной составляющей  $\varepsilon(\tau)$  намного меньше амплитуды  $c(\tau)$  и спектр  $\varepsilon(\tau)$  более менее равномерно «распределен» по гармоникам с различными номерами, т.е. сигнал  $\varepsilon(\tau)$  имеет «широкий» спектр, но его вклад в спектр каждой гармоники тригонометрической составляющей сравнительно мал.

**Пример 2.3.2.** Функция  $p(\tau)$  задается формулой

$$c(\tau) = 3 \sin\left(\frac{2\pi}{3.2} 5\tau\right),$$

а значения временного ряда формируются как

$$y_i = 0.1 + 0.4\tau_i + 0.5\tau_i^2 + p(\tau_i) + \varepsilon(\tau_i),$$

где  $\tau_i = (i-1) \cdot 0.1$ ,  $i = 1, 2, \dots, 32$ ;  $\varepsilon(\tau_i)$  – нормально распределенная величина с нулевым средним и дисперсией  $\sigma^2 = 0.11$  (что соответствует относительному уровню 0.10). Необходимо вычислить спектр  $S_k^{**}$  временной выборки и выделить тригонометрическую составляющую.

*Решение.* Первоначально были получены значения  $y_i$  при  $\delta = 0$  (т.е. отсутствует случайная составляющая  $\varepsilon(\tau_i)$ ) и по формулам (2.3.7)–(2.3.9) вычислены коэффициенты  $a_0^*$ ,  $a_k^*$ ,  $b_k^*$  и определен спектр  $S_k^*$ , значения которых отображены на рис. 2.18 (кривая 2). Затем были получены значения  $y_i$  при  $\delta = 1$  (т.е. присутствует случайная составляющая  $\varepsilon(\tau_i)$ ), вычислены коэффициенты  $a_0^{**}$ ,  $a_k^{**}$ ,  $b_k^{**}$  и определен спектр  $S_k^{**}$ , значения которых отображены на рис. 2.18 (кривая 3). Анализ спектров, изображенных на рис. 2.18, позволяет сделать следующие выводы:

- Спектры  $S_k^*$ ,  $S_k^{**}$ , вычисленные по значениям дискретного временного ряда, являются *симметричными функциями* относительно точки  $k = n/2 = 16$ .

- Во всех трех спектрах *присутствует максимум, соответствующий*  $k = 5$ , что говорит о наличии во временном ряду тригонометрической составляющей вида

$$\hat{c}(\tau) = a_5^{**} \cos\left(\frac{2\pi}{3.2} 5\tau\right) + b_5^{**} \sin\left(\frac{2\pi}{3.2} 5\tau\right),$$

где коэффициенты  $a_5^{**} = -0.119$ ;  $b_5^{**} = 2.625$ , вычисленные по значениям дискретного временного ряда, приближенно равны коэффициентам, вычисленным по непрерывной функции (2.3.4) (см. пример 2.3.1). ☉

Таким образом, *используя разложение временного ряда в ряд Фурье, удастся достаточно точно выделить тригонометрическую составляющую временного ряда, «отфильтровав» тренд  $t(\tau)$  (в нашем примере это полином второй степени  $0.1 + 0.4\tau_i + 0.5\tau_i^2$ ) и случайную составляющую  $\varepsilon(\tau)$ .*

### 2.3.3. Вычисление коэффициентов ряда Фурье в Excel

В Excel вычислять коэффициенты разложения в ряд Фурье можно двумя способами:

- программированием в документе Excel формул (2.3.7)–(2.3.9);

- используя режим *Анализ Фурье* модуля *Анализ данных*.

**Первый способ** достаточно громоздок, и его можно рекомендовать при сравнительно небольших объемах временной выборки с небольшим числом вычисляемых коэффициентов ряда Фурье.

**Пример 2.3.3.** В табл. 2.4 приведены помесичные удои в Новосибирской области в 1975, 1978, 1983 годах.

Таблица 2.4

Месяц $\tau_i$	Помесичные надои по годам			Среднее $\bar{y}_i$
	1975	1978	1983	
1	140	143	133	<b>138.7</b>
2	147	148	135	<b>143.3</b>
3	196	196	183	<b>191.7</b>
4	210	208	203	<b>208.0</b>
5	259	240	254	<b>251.0</b>
6	288	290	294	<b>290.7</b>
7	271	278	276	<b>275.0</b>
8	244	245	264	<b>247.7</b>
9	190	195	196	<b>193.7</b>
10	136	136	144	<b>138.7</b>
11	104	110	115	<b>109.7</b>
12	116	120	124	<b>120.0</b>
<b>Среднее</b>	<b>191.8</b>	<b>192.4</b>	<b>192.6</b>	<b>192.2</b>

В эту таблицу вошли данные только тех лет, которые характеризуются практически одинаковыми среднегодовыми удоями. Это означает, что отсутствует смещение, отличающее один год от другого, и имеют место только сезонные циклические колебания. В последнем столбце приведены средние значения месячных удоев, вычисленных по трем годам. Усреднение месячных надоев по

трем годам выполнено для уменьшения уровня случайной составляющей  $\varepsilon(\tau)$  временного ряда.

Рассматривая  $\bar{y}_i$  как временную выборку ряда необходимо исследовать структуру сезонных циклических колебаний месячных надоев молока.

*Решение.* Занесем в документ Excel значения  $\{\tau_i, \bar{y}_i\}$ ,  $i = 1, 2, \dots, 12$  (рис. 2.19) и по этим данным построим график значений  $\bar{y}_i$  временного ряда (рис. 2.20, кривая 1). Из графика виден колебательный характер изменения значений  $\bar{y}_i$  и поэтому для выяснения структуры этого временного ряда обратимся к методам гармонического анализа (см. п. 2.3.1, 2.3.2).

	A	B	C	D	E	F
1	$\tau_i$	$\bar{y}_i$		=B2*СOS(ПИ()/6*A2)		
2	1	138,7	120,12	69,35	69,35	120,118
3	2	143,3	71,65	124,101	-71,65	124,101
4	3	191,7	0,00	191,7	-191,70	2,3E-14
5	4	208,0	-104,00	180,133	-104,00	-180,13
6	5	251,0	-217,37	125,5	125,50	-217,37
7	6	290,7	-290,70	3,6E-14	290,70	-7E-14
8	7	275,0	-238,16	-137,5	137,50	238,157
9	8	247,7	-123,85	-214,51	-123,85	214,514
10	9	193,7	0,00	-193,7	-193,70	7,1E-14
11	10	138,7	69,35	-120,12	-69,35	-120,12
12	11	109,7	95,00	-54,85	54,85	-95,003
13	12	120,0	120,00	-3E-14	120,00	-6E-14
14		192,4	-82,993	-4,983	7,275	14,044
15		$a_0$	$a_1$	$b_1$	$a_2$	$b_2$
16			=СУММ(C2:C13)/6			

Рис. 2.19. Вычисление коэффициентов Фурье

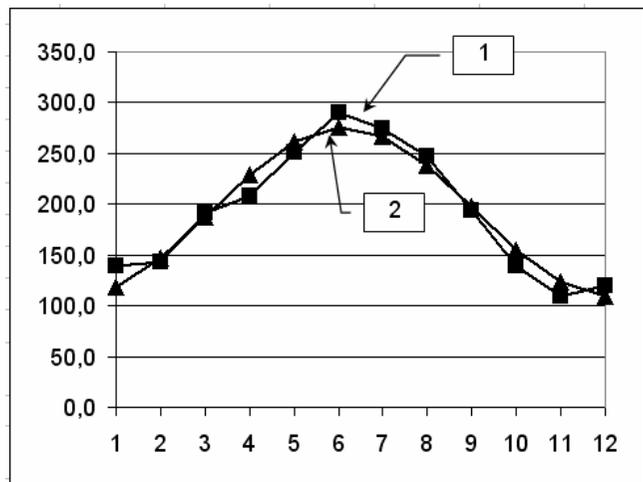


Рис. 2.20. Значения  $y_i$  и  $\hat{p}(\tau_i)$  (пример 2.3.3)

Коэффициенты ряда Фурье будем вычислять по формулам (2.3.7), (2.3.8), (2.3.9) при  $T = 12$ ,  $\Delta_\tau = 1$ ,  $n = 12$ . Тогда имеем следующие соотношения:

$$a_0^* = \frac{1}{12} \sum_{i=1}^{12} \bar{y}_i; \quad a_k^* = \frac{1}{6} \sum_{i=1}^{12} \bar{y}_i \cos\left(\frac{\pi k}{6} i\right); \quad b_k^* = \frac{1}{6} \sum_{i=1}^{12} \bar{y}_i \sin\left(\frac{\pi k}{6} i\right).$$

Вычислим коэффициенты  $a_0$ ,  $a_k^*$ ,  $a_2^*$ ,  $b_1^*$ ,  $b_2^*$ ,  $a_{n/2}^* = a_6^*$ . На рис. 2.19 показан фрагмент документа Excel, реализующий приведенные выше формулы вычисления коэффициентов ряда Фурье. Здесь же приведены вычисленные значения коэффициентов, по которым определим:

$$S_0^* = a_0^{*2} = 36998.5; \quad S_1^* = a_1^{*2} + b_1^{*2} = 6912.7; \quad S_2^* = a_2^{*2} + b_2^{*2} = 250.2.$$

Видно, что первая гармоника вносит существенно больший энергетический вклад по сравнению со второй гармоникой

( $S_1^*/S_2^* \approx 27.6$ ). Поэтому в качестве модели временного ряда примем первые три члена ряда Фурье, а именно, функцию

$$\begin{aligned} \hat{p}(\tau) &= a_0^* + a_1^* \cos\left(\frac{2\pi}{12} \tau\right) + b_1^* \sin\left(\frac{2\pi}{12} \tau\right) = \\ &= 192.4 - 82.99 \cos\left(\frac{\pi}{6} \tau\right) - 4.98 \sin\left(\frac{\pi}{6} \tau\right). \end{aligned}$$

Вычисления этой функции показано на рис. 2.21.

На рис. 2.20 приведен график этой функции и исходных значений  $\bar{y}_i$ . Видно достаточно хорошее совпадение  $\bar{y}_i$  и  $\hat{p}(\tau_i)$ . Об этом же говорит почти нулевое ( $-1.18 \cdot 10^{-14}$ ) среднее значение остатков  $\bar{y}_i - \hat{p}(\tau_i)$  (столбец D на рис. 2.21). ☉

	A	B	C	D
1	$\tau_i$	$\bar{y}_i$	$\hat{p}(\tau_i)$	$e(\tau_i)$
2	1	138,7	118,0	20,72
3	2	143,3	146,5	-3,24
4	3	191,7	187,4	4,33
5	4	208,0	229,5	-21,53
6	5	251,0	261,7	-10,73
7	6	290,7	275,3	15,36
8	7	275,0	266,7	8,28
9	8	247,7	238,2	9,54
10	9	193,7	197,3	-3,63
11	10	138,7	155,2	-16,47
12	11	109,7	123,0	-13,27
13	12	120,0	109,4	10,64
14			$\bar{e} = -1,18E-14$	
15		192,35	-82,9931	-4,982915
16		$a_0$	$a_1$	$b_1$

Рис. 2.21. Вычисление аппроксимирующей функции

**Второй способ** основан на дискретном преобразовании Фурье. Кратко остановимся на этом преобразовании.

Пара дискретных преобразований, определяемая формулами:

$$Z(k) = \sum_{j=0}^{N-1} z(j)e^{-i\frac{2\pi jk}{N}} \quad (\text{прямое ДПФ});$$

$$z(k) = \frac{1}{N} \sum_{j=0}^{N-1} Z(j)e^{i\frac{2\pi jk}{N}} \quad (\text{обратное ДПФ}),$$

где  $i = -\sqrt{-1}$  мнимая единица, называется дискретным преобразованием Фурье (ДПФ). Исходная дискретная последовательность  $z(j)$  является периодической (с периодом  $N$ ). Последовательность  $Z(k)$ , называемая коэффициентами ДПФ также является периодической (с периодом  $N$ ).

Если коэффициенты  $Y(k)$  вычислены по значениям временного ряда  $y_j, j = 1, \dots, n$ , то связь между коэффициентами ДПФ и коэффициентами разложения в ряд определяется как:

$$a_0^* = \frac{1}{n}Y(0), \quad a_k^* = \frac{2}{n}\text{Re}[Y(k)], \quad b_k^* = \frac{2}{n}\text{Im}[Y(k)], \quad k = 1, \dots, n/2,$$

где  $\text{Re}[A]$ ,  $\text{Im}[A]$  означают вещественную и мнимую части комплексного числа  $A$ .

Прямое и обратное ДПФ вычисляются в режиме **Анализ Фурье**. Для вызова этого режима необходимо обратиться к пункту **Сервис** команде **Анализ данных** и выбрать в списке режимов **Анализ Фурье** и щелкнуть мышью ОК.

Затем в новом диалоговом окне задать следующие параметры (рис. 2.22):

**Входной интервал:** – диапазон ячеек, содержащих вещественные данные, к которым применяется ДПФ.

**Метки в первой строке** – включается, если первая строка содержит заголовки.

**Выходной интервал:** – вводится адрес левой верхней ячейки выходного диапазона.

**Инверсия** – включается, если необходимо вычислить обратное ДПФ.

**Замечание 2.3.1.** Используемый для вычисления ДПФ алгоритм, называемый алгоритмом *быстрого преобразования Фурье* (БПФ) требует, чтобы  $n$  – число значений временного ряда, должно быть обязательно равным степени 2 (т.е. 8, 16, 32, 64, ...), что является существенным ограничением. Один из путей преодоления этого недостатка – добавление в конец временной выборки нулей до тех пор, пока длина «новой» временной выборки не станет равной степени 2. Однако такой способ, применяемый при цифровой обработке сигналов, далеко не всегда пригоден для обработки данных, характеризующих экономические процессы. Поэтому перед применением режима **Анализ Фурье** необходимо сформировать выборку длиной  $n$ , равной степени числа 2. ♦

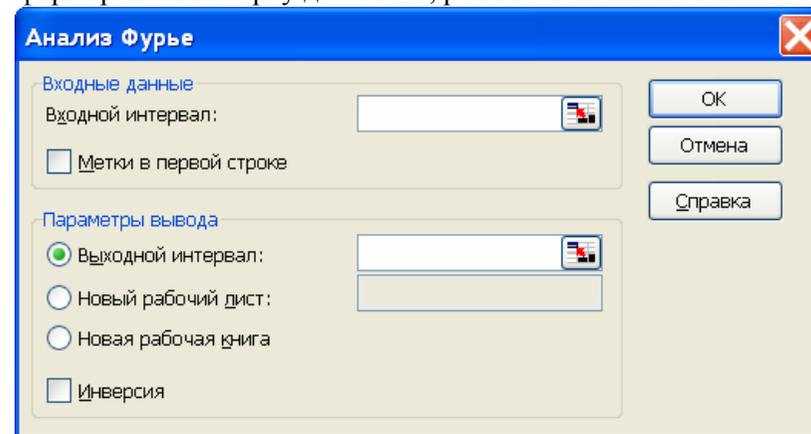


Рис. 2.22. Диалоговое окно режима **Анализ Фурье**

Так как результатом работы режима **Анализ Фурье** будут комплексные числа вида  $a_k + ib_k$ , то для вычисления вещественной и мнимой части можно использовать следующие функции Excel (категория *Инженерные*): ВЕЩ(), МНИМ().

## 2.4. Проверка адекватности и качества построенной модели временного ряда

Проверка адекватности построенной модели реальному явлению или процессу является важным этапом прогнозирования исследуемых процессов. Исходными данными для такой проверки является ряд остатков (невязок)

$$e_i = y_i - \hat{q}(\tau_i), \quad i=1,2,\dots,n, \quad (2.4.1)$$

где  $\hat{q}(\tau_i)$  – значение, вычисленное по построенной модели при  $\tau = \tau_i$ . Заметим, что  $\hat{q}(\tau)$  может содержать не только тренд  $\hat{t}(\tau)$ , но и тригонометрическую составляющую.

Проверка значимости трендовой составляющей рассматривалась с использованием  $F$ -критерия и индексов детерминации  $R_t^2$ ,  $\hat{R}_t^2$  в п. 2.1.3. Ниже будут рассмотрены способы проверки важных свойств и числовых характеристик ряда остатков.

### 2.4.1. Проверка математического ожидания ряда остатков

Если трендовая составляющая  $t(\tau)$  оценена достаточно точно и  $M(e(\tau_i)) \equiv 0$ , то математическое ожидание остатков  $e(\tau_i)$  должно быть равно нулю. Поэтому формулируем следующие статистические гипотезы:

$$H_0: M(e(\tau_i)) = 0; \quad (2.4.2)$$

$$H_1: M(e(\tau_i)) \neq 0, \quad i=1,2,\dots,n. \quad (2.4.3)$$

Для проверки этих гипотез обратимся к критерию, применяемому в математической статистике, а именно, введем величину

$$T_e = \frac{\bar{e}}{s_e} \cdot \sqrt{n}, \quad (2.4.4)$$

где

$$\bar{e} = \frac{1}{n} \sum_{i=1}^n e_i; \quad s_e = \left[ \frac{\sum_{i=1}^n (e_i - \bar{e})^2}{n-1} \right]^{1/2}.$$

Если гипотеза  $H_0$  справедлива, то критерий  $T_e$  подчиняется распределению Стьюдента с  $(n-1)$  степенями свободы. Поэтому критическая область при альтернативной гипотезе (2.4.3) имеет вид

$$(-\infty, x_{лев,\alpha/2}] \cup [x_{пр,\alpha/2}, \infty), \quad (2.4.5)$$

где  $x_{лев,\alpha/2} = -x_{пр,\alpha/2}$ ,  $x_{пр,\alpha/2} = t(1-\alpha, n-1)$ . (2.4.6)

Значение  $t(1-\alpha, n-1)$  можно вычислить, используя функцию Excel

$$t(1-\alpha, n-1) = \text{РАСПСТЬЮДОБР}(\alpha; n-1). \quad (2.4.7)$$

Таким образом, если вычисленное значение критерия (2.4.4) попадает в критическую область (2.4.5), то гипотеза  $H_0$  отвергается с уровнем значимости  $\alpha$  и принимается гипотеза о том, что математические ожидания остатков отличаются от нуля. Это означает, что выделенная неслучайная составляющая  $\hat{q}(\tau)$  содержит ненулевую систематическую (методическую) ошибку. Для устранения этой ошибки необходимо изменить уравнение тренда.

### 2.4.2. Проверка случайности ряда остатков

При правильно подобранной модели неслучайной составляющей временного ряда ряд остатков  $e_i$ ,  $i=1,2,\dots,n$ , должен представлять собой набор случайных величин. Для проверки свойства случайности значений  $e_i$  воспользуемся критерием «поворотных точек». Значение  $e_i$  назовем поворотной точкой, если выполняется одно из следующих систем неравенств:

$$\text{а) } e_i < e_{i-1}; e_i < e_{i+1},$$

$$\text{б) } e_i > e_{i-1}; e_i > e_{i+1}.$$

Если остатки случайны, то одна поворотная точка будет приходиться примерно на 1,5 наблюдения. Обозначим через  $n_p$  – фактическое количество поворотных точек в ряде остатков. Если выполняется неравенство

$$n_p > \text{ent} \left[ \frac{2}{3}(n-2) - 1.96 \sqrt{\frac{16n-29}{90}} \right], \quad (2.4.8)$$

где  $\text{ent}[\cdot]$  – целая часть, то ряд остатков  $e_i$  можно считать случайным и, следовательно, построенная модель не содержит систематическую ошибку и является адекватной данному временному ряду.

### 2.4.3. Проверка независимости значений ряда остатков

Зависимость между значениями ряда остатков (другими словами наличие автокорреляции между этими значениями) также свидетельствует о неполноте построенной модели, в которой не была учтена автокорреляция между значениями исходного временного ряда (более подробно об авторегрессионных моделях см. п. 4.1).

Достаточно простым критерием, определяющим наличие автокорреляции между соседними наблюдениями, является следующий тест.

**Тест Дарбина–Уотсона.** Этот тест основан на простой идее: если корреляция между  $\varepsilon_i$  и  $\varepsilon_{i+1}$  не равна нулю, то она присутствует и в остатках (невязках)  $e_i = y_i - \hat{y}_i$  регрессионной модели, где  $\hat{y}_i = \hat{q}(\tau_i)$  – оценка объясненной части временного ряда, построенная обычным методом наименьших квадратов. Определим статистику:

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=2}^n e_i^2}. \quad (2.4.9)$$

Между этой статистикой и выборочным коэффициентом корреляции  $r$  имеется связь:

$$d \approx 2(1-r). \quad (2.4.10)$$

В случае отсутствия автокорреляции (т.е.  $r \approx 0$ ) значение статистики близко к двум. Близость статистики к нулю должна означать наличие положительной автокорреляции, к четырем – отрицательной автокорреляции. К сожалению, неопределена пороговая точка для статистики  $d$  при принятии или отвержении нулевой гипотезы  $H_0$ : автокорреляция отсутствует. Поэтому весь диапазон значений  $d$  делится на ряд интервалов. Если наблюдается значение:

а)  $d_B < d < 4 - d_B$  – гипотеза  $H_0$  принимается;

б)  $d_H < d < d_B$  или  $4 - d_B < d < 4 - d_H$  – вопрос о принятии или отвержении гипотезы  $H_0$  остается открытым;

в)  $0 < d < d_H$  – гипотеза  $H_0$  отвергается и принимается альтернативная гипотеза о положительной автокорреляции;

г)  $4 - d_H < d < 4$  – гипотеза  $H_0$  отвергается и принимается альтернативная гипотеза о наличии отрицательной автокорреляции.

Пороговые значения  $d_H, d_B$  зависят от числа наблюдений, числа объясняющих переменных в функции  $q(\tau)$  и уровня значимости. Эти значения приводятся в специальной таблице [5] и определены для  $n \geq 15$ . Это ограничение является определенным недостатком теста. В табл. 2.5 приведены некоторые значения  $d_H, d_B$  для уровня значимости  $\alpha = 0.05$ , где  $k$  – число объясняющих переменных. Используя данные табл. 2.5, можно экстраполировать  $d_H, d_B$  на меньшее число наблюдений.

**Пример 2.4.1.** Выявить на уровне значимости  $\alpha = 0.05$  наличие автокорреляции ряда остатков уравнения тренда, построенного для временного ряда (значения приведены в табл. 2.1).

*Решение.* В качестве оценки  $\hat{q}(\tau)$  для неслучайной составляющей  $q(\tau)$  временного ряда возьмем квадратичный полином (см. пример 2.1.1)  $\hat{q}(\tau) = 132.3 + 55.09\tau - 3.26\tau^2$ , что соответствует  $k = 2$  (объясняющие переменные модели  $\tau, \tau^2$ ). Вычислим остатки  $e_i, i = 1, 2, \dots, 8$  и статистику  $d$ , как показано на фрагменте документа Excel, приведенного на рис. 2.23.

Таблица 2.5

n	k = 1		k = 2		k = 3		k = 4	
	$d_H$	$d_B$	$d_H$	$d_B$	$d_H$	$d_B$	$d_H$	$d_B$
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72

	Н	l	J	К	L	М	N
1	Год	Спрос	$\hat{q}$	$e_i$	$e_i - e_{i-1}$	$e_i^2$	$(e_i - e_{i-1})^2$
2	1	213	184,13	28,87		833,77	
3	2	171	229,41	-58,41	-87,29	3411,81	7618,80
4	3	291	268,16	22,84	81,25	521,63	6601,56
5	4	309	300,38	8,62	-14,21	74,39	202,05
6	5	317	326,05	-9,05	-17,68	81,97	312,53
7	6	362	345,20	16,80	25,86	282,36	668,59
8	7	351	357,80	-6,80	-23,61	46,29	557,30
9	8	361	363,88	-2,88	3,93	8,27	15,43
10				Суммы		5260,48	15976,26
11							
12	Статистика d			3,037			
13						=СУММ(N3:N9)	
14							
15		=N10/M10					

Рис. 2.23. Вычисление статистики Дарбина–Уотсона

Найденное значение  $d = 3.037$ . Используя данные табл. 2.5, выполним линейную экстраполяцию значений  $d_H, d_B$  при  $k = 2$  для числа наблюдений  $n = 8$ . Получим  $d_H \approx 0.74, d_B \approx 1.54$ . Видно, что вычисленное значение  $d$  находится в пределах от  $d_B = 1.54$  до  $4 - d_H = 3.26$ , что соответствует принятию гипотезы  $H_0$  об отсутствии автокорреляции с уровнем значимости  $\alpha = 0.05$ . ●

Положительные результаты описанных выше проверок свойств ряда остатков, а именно:

- нулевое математическое ожидание остатков;
- случайность значений ряда остатков;
- независимость значений ряда остатков,

дают уверенность в полноте и адекватности построенной модели и в успешном использовании модели для решения задач прогнозирования временного ряда.

## 2.5. Прогнозирование трендовой составляющей временного ряда

Идея социально-экономического прогнозирования базируется на предположении, что закономерность развития, действовавшая в прошлом (внутри ряда экономической динамики), сохранится и в прогнозируемом будущем. В этом смысле прогноз основан на *экстраполяции*. Экстраполяция, проводимая в будущее, называется *перспективной*, а в прошлое – *ретроспективной*.

Прогнозирование метода экстраполяции базируется на следующих предположениях:

- развитие исследуемого явления в целом описывается плавной кривой;
- общая тенденция развития явления в прошлом и настоящем не указывает на серьезные изменения в будущем;
- учет случайности позволяет оценить вероятность отклонения от закономерного развития.

Поэтому надежность и точность прогноза зависят от того, насколько близкими к действительности окажутся эти предположения и насколько точно удалось охарактеризовать выявленную в прошлом закономерность.

На основе построенной модели рассчитываются точечные и интервальные прогнозы. Точечный прогноз на основе временных моделей получается подстановкой в модель (уравнение тренда) соответствующего значения фактора времени, т.е.  $\tau_{n+1} = (n+1)\Delta_\tau$ ,  $\tau_{n+2} = (n+2)\Delta_\tau$  и т.д.

Интервальные прогнозы строятся на основе точечных прогнозов. Напомним, что *доверительным* называется *интервал*, относительно которого можно с заранее выбранной вероятностью утверждать, что он содержит значение прогнозируемого показателя. Ширина интервала зависит от качества модели, т.е. степени ее близости к фактическим данным, числа наблюдений и выбранного пользователем доверительной вероятности.

Нижняя  $\hat{t}_H(l)$  и верхняя  $\hat{t}_B(l)$  границы доверительного интервала для «индивидуального» значения трендовой составляющей [5, с. 39–40] вычисляются по следующим формулам

$$\hat{t}_H(\tau_l) = \hat{t}(\tau_l) - \delta_i(\tau_l); \quad (2.5.1)$$

$$\hat{t}_B(\tau_l) = \hat{t}(\tau_l) + \delta_i(\tau_l), \quad (2.5.2)$$

где  $\hat{t}(\tau_l)$  – значение оценки трендовой составляющей в момент времени  $\tau_l$ . Если  $\tau_l > \tau_n$ , то  $\hat{t}(\tau_l)$  – прогнозируемое значение трендовой составляющей.

Величина  $\delta_f(\tau_l)$  для линейного тренда

$$\hat{t}(l) = b_0 + b_1\tau_l \quad (2.5.3)$$

определяется выражением

$$\delta_f(\tau_l) = t(\gamma, n-2)s \sqrt{1 + \frac{1}{n} + \frac{(\tau_l - \bar{\tau})^2}{\sum_{i=1}^n (\tau_i - \bar{\tau})^2}}, \quad (2.5.4)$$

где

$$s = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n-2}}; \quad \bar{\tau} = \frac{1}{n} \sum_{i=1}^n \tau_i. \quad (2.5.5)$$

Напомним, что [5, с. 37–40]:

- величина  $s^2$  является оценкой для дисперсии  $\sigma^2$  случайной составляющей временного ряда;
- величина  $\gamma$  определяют доверительную вероятность (или надежность интервальной оценки) и обычно  $\gamma = 0.95$ ;
- $t(\gamma, n-2)$  – квантиль распределения Стьюдента с  $(n-2)$  степенями свободы, определяемый функцией Excel:

$$t(\gamma, n-2) = \text{СТБЮДРАСПОБР}(1-\gamma; n-2). \quad (2.5.6)$$

Для нелинейных моделей величина  $\delta_f(\tau_l)$  также зависит от  $s$  и разности  $\tau_l - \bar{\tau}$ , но эта зависимость имеет более сложный характер.

Если построенная модель адекватна, то с вероятностью  $\gamma$  можно утверждать, что при сохранении сложившихся закономерностей развития прогнозируемая величина попадает в интервал, образованный границами формул (2.5.1), (2.5.2).

**Пример 2.5.1.** В табл. 2.6 приведены величины нетто-платежей ежемесячного финансирования инвестиционного проекта.

Таблица 2.6

Месяц	1	2	3	4	5	6	7	8	9	10	11	12
Платеж, тыс. р.	45	40	43	48	42	47	51	55	50	57	60	62

Требуется определить:

- линейную модель зависимости объемов платежей от сроков (времени);
- адекватность построенной модели;
- размеры платежей на три последующих месяца, построив точечный и интервальный прогноз.

*Решение.* Введем в столбцы А, В листа Excel данные табл. 2.6 (рис. 2.24), затем построим диаграмму рассеяния, как геометрическое место точек  $(i, y_i)$ ,  $i = 1, \dots, 12$ . Анализ диаграммы показывает наличие небольшого тренда.

Используя команду *Добавить линию тренда*, строим линейную модель тренда, описываемую уравнением

$$\hat{i}(\tau) = 38.23 + 1.81\tau \quad (2.5.7)$$

с коэффициентом детерминации  $R^2 = 0.823$ . Для проверки значимости коэффициентов уравнения тренда обратимся к модулю *Анализ данных* режима *Регрессия* [5, с. 94–101] и зададим необходимые параметры в появившемся диалоговом окне. Результаты выполнения режима *Регрессия* показаны на рис. 2.25, 2.26.

Видно, что коэффициент детерминации равен 0.823, приведенный коэффициент детерминации  $\hat{R}^2 = 0.805$  (см. рис. 2.25). Все коэффициенты  $b_0$ ,  $b_1$  являются значимыми (при уровне значимости  $\alpha = 0.05$ ) и построенное уравнение линейного тренда также является значимым (см. рис. 2.25, 2.26), так как выполняются неравенства (более подробно см. [5, с. 41–46]):

$$T_{b_0} = 19.554 > t(1 - \alpha, n - 2) = 2.228;$$

$$T_{b_1} = 6.818 > t(1 - \alpha, n - 2) = 2.228;$$

$$F = 46.490 > F_{1-\alpha;1;n-2} = 4.965.$$

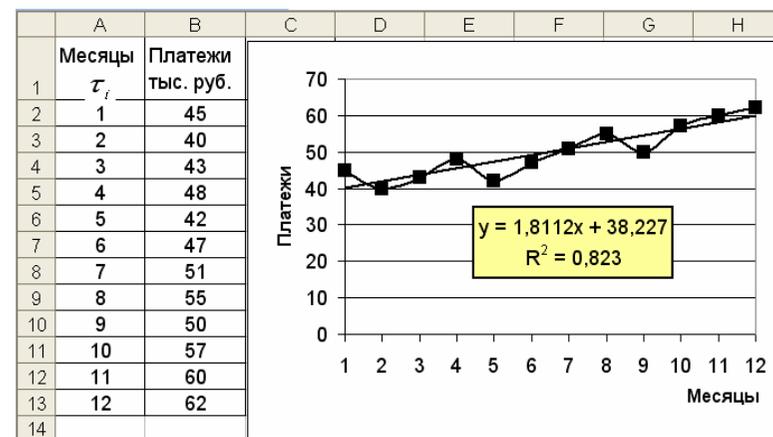


Рис. 2.24. Построение линейного тренда (пример 2.5.1)

ВЫВОД ИТОГОВ					
Регрессионная статистика					
Множественный R	0,907				
R-квадрат	0,823				
Нормированный R-квадрат	0,805				
Стандартная ошибка	3,177				
Наблюдения	12				
Дисперсионный анализ					
	df	SS	MS	F	Значимость F
Регрессия	1,000	469,098	469,098	46,490	0,00005
Остаток	10,000	100,902	10,090		
Итого	11,000	570,000			

Рис. 2.25. Результаты выполнения режима *Регрессия*

17			Стандартная ошибка	t-статистика
18	Y-пересечение	38,227	1,955	19,554
19	Переменная X 1	1,811	0,266	6,818

Коэффициенты	Нижние 95%	Верхние 95%
38,227	33,871	42,583
1,811	1,219	2,403

Рис. 2.26. Результаты выполнения режима *Регрессия*

Выполним проверку статистической гипотезы (2.4.2) о равенстве математического ожидания ряда остатков  $e_i$  нулю, график которых приведен на рис. 2.27.

Подставляя значения

$$\bar{e} = 0.002; s_e = \sqrt{\frac{100.902 - (0.002)^2}{12 - 1}} = 3.03$$

в выражение (2.4.4), получаем

$$T_e = \frac{0.002}{3.03} \sqrt{12} = 0.0023.$$

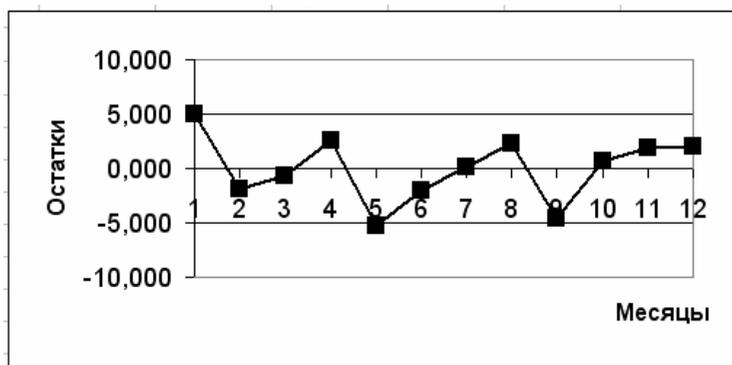


Рис. 2.27. График остатков  $e_i$

Используя функцию (2.4.6), вычисляем  $x_{np,\alpha/2} = 2.179$ . Видно, что вычисляемое значение критерия не попадает в критическую область (2.4.5)

$$(-\infty, -2.179] \cup [2.179, \infty),$$

и поэтому принимается гипотеза о равенстве нулю математического ожидания значений остатков  $e_i$ .

Проверку случайности значений остатков  $e_i$  проведем на основе критерия поворотных точек (см. п. 2.4.2). Как видно из рис. 2.28 (столбец G), количество поворотных точек  $n_p = 5$ . Тогда неравенство (2.4.8) выполняется

$$5 > \text{ent} \left[ \frac{2}{3}(12 - 2) - 1.96 \sqrt{\frac{16 \cdot 12 - 29}{90}} \right] = \text{ent} [4.029] = 4.$$

Поэтому имеет место свойство случайности остатков  $e_i$ .

Проверку независимости остатков  $e_i$  (т.е. отсутствие автокорреляции) выполним с использованием теста Дарбина–Уотсона (см. п. 2.4.3). На рис. 2.28 показан фрагмент документа Excel, в котором вычисляется по формуле (2.4.10) наблюдаемое значение критерия  $d = 2.116$ . Используя табл. 2.5 выполним линейную экстраполяцию значений  $d_H$ ,  $d_B$  при  $k = 1$  и числе наблюдений  $n = 12$ . Получаем  $d_H \approx 1.008$ ,  $d_B \approx 1.33$ . Видно, что вычисляемое значение  $d$  находится внутри интервала

$$[d_B, 4 - d_B] = [1.33, 2.67].$$

Следовательно, можно говорить о независимости значений остатков  $e_i$ .

*Проверенные свойства ряда остатков позволяют считать, что построенная линейная модель тренда (2.5.7) адекватно описывает исследуемый временной ряд и ее можно использовать для решения задач прогнозирования.*

	A	B	C	D	E	F	G
	Месяцы $\tau_i$	Платежи тыс. руб.	$\hat{t}_i$	$e_i$	$e_i^2$	$(e_i - e_{i-1})^2$	Точки поворота
1							
2	1	45	40,038	4,962	24,619		
3	2	40	41,849	-1,849	3,420	46,392	***
4	3	43	43,661	-0,661	0,436	1,413	
5	4	48	45,472	2,528	6,392	10,168	***
6	5	42	47,283	-5,283	27,910	61,015	***
7	6	47	49,094	-2,094	4,386	10,168	
8	7	51	50,905	0,095	0,009	4,791	
9	8	55	52,717	2,283	5,214	4,791	***
10	9	50	54,528	-4,528	20,501	46,392	***
11	10	57	56,339	0,661	0,437	26,924	
12	11	60	58,150	1,850	3,422	1,413	
13	12	62	59,961	2,039	4,156	0,036	
14			Итого	0,002	100,902	213,504	5
15							
16		d=	2,116		=F14/E14		
17						=СУММ(F3:F13)	

Рис. 2.28. Вычисление значения  $d$  (пример 2.5.1)

Вычислим точечный прогноз по формулам (2.5.1), (2.5.2) при  $\tau = 13, 14, 15$ :

$$\hat{t}(13) = 38.23 + 1.81 \cdot 13 \approx 61.77;$$

$$\hat{t}(14) = 38.23 + 1.81 \cdot 14 \approx 63.58;$$

$$\hat{t}(15) = 38.23 + 1.81 \cdot 15 \approx 65.40.$$

Результаты вычисления интервального прогноза приведены в табл. 2.7.

С вероятностью  $\gamma = 1 - \alpha = 0.95$  можно утверждать, что величина ежемесячного финансирования будет находиться в интервале, определяемом вычисленными нижними и верхними границами.

Таблица 2.7

Время, $\tau_i$	$\delta_f(\tau_i)$	Точечный прогноз	Нижняя граница	Верхняя граница
13	6.80	61.77	54.97	68.57
14	7.04	63.58	56.55	70.62
15	7.29	65.40	58.10	72.69

### ЛАБОРАТОРНАЯ РАБОТА № 1

#### Выделение трендовой составляющей временного ряда

*Исходные данные.* В таблице приведены значения урожайности зерновых совхоза «Путь к коммунизму» Черепановского района за 14 лет, ц/га.

Год	1972	1973	1974	1975	1976	1977	1978
Урожайность	17.9	19.5	30.1	26.7	31.0	28.4	20.0
Год	1979	1980	1981	1982	1983	1984	1985
Урожайность	24.7	26.9	21.3	28.6	27.7	30.0	22.9

*Необходимо:*

1. Используя команду *Добавить линию тренда* (см. п. 2.2.1) построить следующие модели для трендовой составляющей:

$$\hat{t}_1(\tau) = b_0 + b_1\tau;$$

$$\hat{t}_2(\tau) = b_0 + b_1\tau + b_2\tau^2;$$

$$\hat{t}_3(\tau) = b_0 + b_1\tau + b_2\tau^2 + b_3\tau^3.$$

2. Для каждого уравнения вычислить коэффициенты, индекс детерминации  $R_t^2$ , приведенный индекс детерминации  $\hat{R}_t^2$ .

3. Занести в следующую таблицу (на том же листе Excel) уравнение тренда и вычисленные значения  $R_t^2$ ,  $\hat{R}_t^2$ .

Уравнение	$R_t^2$	$\hat{R}_t^2$

4. Анализируя величины  $R_t^2$ ,  $\widehat{R}_t^2$ , найти наилучшую модель  $\widehat{t}_{opt}(\tau)$  трендовой составляющей.

5. Построить графики значений  $y_i$ ,  $\widehat{t}_{opt}(\tau)$ ,  $e_i$ .

6. Используя критерий (2.1.20), проверить возможность замены нелинейного тренда  $\widehat{t}_3(\tau)$  линейной функцией  $\widehat{t}_1(\tau)$ .

7. Используя наилучшую модель тренда, осуществить прогноз урожайности на следующие годы: 1986, 1988, 1990.

*Рекомендации:*

1. Перед выполнением работы ознакомьтесь с примером 2.2.2, в котором решается подобная задача.

2. Вычисление величин  $R_t^2$ ,  $\widehat{R}_t^2$  осуществлять по формулам (2.1.18), (2.1.28).

3. При выполнении задания 6 использовать пример 2.1.2.

## ЛАБОРАТОРНАЯ РАБОТА № 2

### Исследование модели трендовой составляющей временного ряда

*Исходные данные.* В лабораторной работе № 1 по выборке, представленной в таблице, построена наилучшая модель тренда  $\widehat{t}_{opt}(\tau)$  (для этой модели приведенный индекс детерминации  $\widehat{R}_t^2$  имеет максимальное значение среди всех остальных рассмотренных моделей).

*Необходимо:*

1. Вычислить для модели  $\widehat{t}_{opt}(\tau)$  ряд остатков  $e_i = y_i - \widehat{t}_{opt}(\tau_i)$ ,  $i = 1, 2, \dots, n$ .

2. Построить на одном рисунке графики  $y_i$ ,  $\widehat{t}_{opt}(\tau_i)$ ,  $e_i$ .

3. Используя критерий (2.4.4), проверить гипотезу о нулевом математическом ожидании остатков  $e_i$ .

4. Используя критерий поворотных точек, проверить гипотезу о случайности значений остатков  $e_i$ .

5. Используя тест Дарбина–Уотсона, проверить гипотезу о независимости между собой остатков  $e_i$ .

6. В зависимости от результатов проверки статистических гипотез п. 3–5 сделать аргументированный вывод об адекватности построенной модели и ее пригодности для решения задач прогнозирования.

*Рекомендации:*

1. Проверку гипотезы о нулевом математическом ожидании остатков  $e_i$  осуществить следуя п. 2.4.1 и примеру 2.5.1.

2. Проверку гипотезы о случайности значений остатков  $e_i$  осуществить следуя п. 2.4.2 и примеру 2.5.1.

3. Проверку гипотезы о независимости остатков  $e_i$  осуществить следуя п. 2.4.3 и примеру 2.5.1.

## ЛАБОРАТОРНАЯ РАБОТА № 3

### Прогнозирование трендовой составляющей временного ряда

*Исходные данные.* В таблице лабораторной работы № 1 приведены значения урожайности зерновых за 14 лет, ц/га.

*Необходимо:*

1. По данным таблицы построить модель линейного тренда вида  $\widehat{t}(\tau) = b_0 + b_1\tau$ .

2. Вычислить по этой модели точечный прогноз для 1986–1992 годов.

3. Вычислить интервальный прогноз для 1986–1992 годов.

4. На одном рисунке построить графики  $y_i$ ,  $\widehat{t}(\tau_i)$  точечного и интервального прогноза.

*Рекомендации:*

1. При выполнении задания 1 можно использовать команду *Добавить линию тренда*, описанную в п. 2.2.1.

2. При выполнении задания 3 использовать материал п. 2.5 и примера 2.5.1.

## КОНТРОЛЬНАЯ РАБОТА

*Исходные данные.* В таблице приведены значения удоев молока от одной коровы одного из хозяйств Новосибирской области, кг.

Номер года	1	2	3	4	5	6	7
Год	1971	1972	1973	1974	1975	1976	1977
Удои	3201	3192	3156	3364	3489	3587	3648
Номер года	8	9	10	11	12	13	14
Год	1978	1979	1980	1981	1982	1983	1984
Удои	3475	3481	3579	3543	3685	3701	3966

*Необходимо:*

- По данным таблицы построить диаграмму рассеяния и визуально убедиться в наличии трендовой составляющей.
- Проверить гипотезу о наличии неслучайной составляющей в исследуемом временном ряду.
- Используя команду *Добавить линию тренда* (см. п. 2.2.1) построить следующие модели для трендовой составляющей:

$$\hat{t}_1(\tau) = b_0 + b_1\tau;$$

$$\hat{t}_2(\tau) = b_0 + b_1\tau + b_2\tau^2;$$

$$\hat{t}_3(\tau) = b_0 + b_1\tau + b_2\tau^2 + b_3\tau^3.$$

- Для каждого уравнения вычислить коэффициенты, индекс детерминации  $R_t^2$ , приведенный индекс детерминации  $\hat{R}_t^2$  и занести эти значения в следующую таблицу

Уравнение	$R_t^2$	$\hat{R}_t^2$

5. Анализируя величины  $R_t^2$ ,  $\hat{R}_t^2$  для разных уравнений найти наилучшую модель  $\hat{t}_{opt}(\tau)$  трендовой составляющей.

6. Построить графики значений  $y_i$ ,  $\hat{t}_{opt}(\tau)$ ,  $e_i$ .

7. Используя наилучшую модель тренда осуществить прогноз удоев на следующие годы: 1986, 1988, 1990.

8. Используя критерий (2.4.4) проверить гипотезу о нулевом математическом ожидании остатков  $e_i$ .

9. Используя критерий поворотных точек проверить гипотезу о случайности значений остатков  $e_i$ .

10. Используя тест Дарбина–Уотсона проверить гипотезу о независимости между собой остатков  $e_i$ .

11. В зависимости от результатов проверки статистических гипотез п. 3–5 сделать аргументированный вывод об адекватности построенной модели и ее пригодности для решения задач прогнозирования.

12. Для модели линейного тренда осуществить точечный прогноз удоев на следующие годы: 1986, 1988, 1990. Сравнить сделанный прогноз с прогнозом, построенном в задании 7.

13. Для модели линейного тренда осуществить интервальный прогноз удоев на следующие годы: 1986, 1988, 1990.

## ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

- Что понимается под терминами «выделить трендовую составляющую» или «выделить тренд»?
- Перечислите основные подходы, используемые для выделения трендовой составляющей.
- Сущность регрессионного метода выделения тренда временного ряда.
- Сущность метода скользящего среднего выделения тренда временного ряда.
- Сущность метода экспоненциального сглаживания для выделения тренда временного ряда.

6. Реализация регрессионного метода выделения тренда временного ряда в табличном процессоре Excel.

7. Что такое периодическая функция и ее представление тригонометрическим рядом?

8. Сущность выделения тригонометрической составляющей методами гармонического анализа.

9. Построение точечного прогноза для трендовой составляющей временного ряда.

10. Построение интервального прогноза для трендовой составляющей временного ряда.

11. Различие между точечным и интервальным прогнозом трендовой составляющей временного ряда.

12. Как нужно изменить дисперсию  $\sigma^2$  для уменьшения «ширины» интервала прогнозирования трендовой составляющей временного ряда?

### Глава 3. ВРЕМЕННЫЕ РЯДЫ С КОРРЕЛИРОВАННЫМИ ВОЗМУЩЕНИЯМИ. ОБОБЩЕННЫЙ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

#### 3.1. Временные ряды с коррелированными возмущениями

##### 3.1.1. Коррелированные возмущения

Упорядоченность наблюдений в пространственной выборке временного ряда оказывается существенной в тех случаях, когда присутствуют механизмы влияния результатов предыдущих наблюдений на результаты последующих. Такое влияние имело место при наличии в модели временного ряда лаговых переменных  $Y_{i-1}, Y_{i-2}, \dots, Y_{i-p}$  (см. уравнение (1.1.4)). Авторегрессионные модели с такими переменными будут рассмотрены в п. 4.1. При этом возмущения  $\varepsilon_i$  временного ряда удовлетворяли условиям Гаусса–Маркова P2÷P4 (см. п. 4.2), т.е.  $\varepsilon_i$  подчинялась нормальному распределению с нулевым средним, дисперсией  $\sigma^2$  и возмущения  $\varepsilon_i, \varepsilon_j$  при  $i \neq j$  были некоррелированными.

Влияние результатов предыдущих наблюдений на последующие также имеет место в случаях, когда отсутствуют лаговые переменные, но возмущения  $\varepsilon_i$  в следующей модели наблюдений временного ряда

$$y_i = q(\tau_i) + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (3.1.1)$$

оказываются зависимыми случайными величинами, т.е. корреляционный момент  $\mu_{i,j} = M(\varepsilon_i \varepsilon_j)$  не равен нулю при  $i \neq j$ . Очевидно, что и коэффициент корреляции  $\rho_{\varepsilon_i \varepsilon_j}$  между величинами  $\varepsilon_i, \varepsilon_j$  отличается от нуля при  $i \neq j$ .

Регрессионные модели временного ряда, в которых условие  $\rho_{\varepsilon_i \varepsilon_j} = 0$  при  $i \neq j$  не выполняется, называются *моделями с коррелированными возмущениями*. Возможны два вида корреляции:

положительная и отрицательная, определяемые знаком коэффициента корреляции между соседними  $\varepsilon_i, \varepsilon_{i+1}$ .

Положительная корреляция проявляется в чередовании временных интервалов, где наблюдаемые значения временного ряда  $y_i$  оказываются выше или ниже значений  $q(\tau_i)$  объясненной части регрессионной модели. Пример положительной корреляции (коэффициент  $\rho_{\varepsilon_i, \varepsilon_{i+1}} = 0.4$ ) случайных возмущений  $\varepsilon_i$  приведен на рис 3.1 (кривая 1 – значения  $\varepsilon_i$ , кривая 2 – значения  $q(\tau_i)$ , кривая 3 – значения  $y_i$ ).

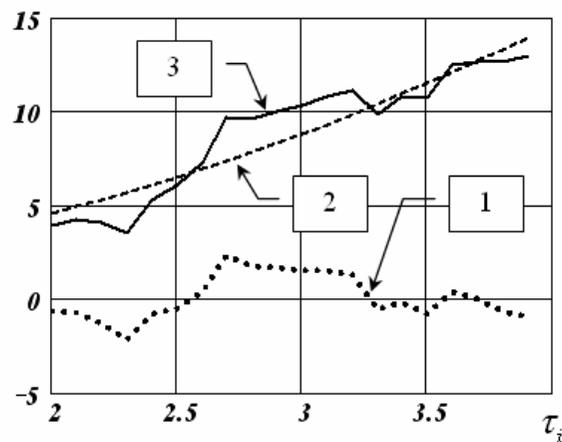


Рис. 3.1. Временной ряд с положительной автокорреляцией

Отрицательная корреляция характеризуется тем, что наблюдения действуют друг на друга по *принципу маятника* – завышенные значения в предыдущих наблюдениях приводят к занижению последующих значений, т.е. наблюдения  $y_i$  слишком часто перескакивают через график объясненной части  $q(\tau_i)$ . Пример отрицательной корреляции (коэффициент  $\rho_{\varepsilon_i, \varepsilon_{i+1}} = -0.4$ ) случайных возмущений  $\varepsilon_i$  приведен на рис 3.2 (кривая 1 – значения  $\varepsilon_i$ , кривая 2 – значения  $q(\tau_i)$ , кривая 3 – значения  $y_i$ ).

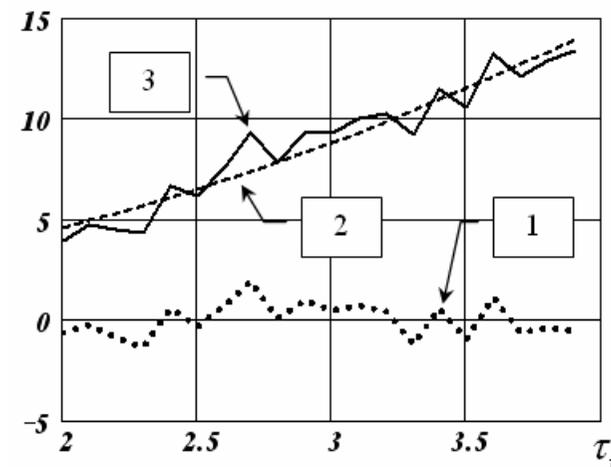


Рис. 3.2. Временной ряд с отрицательной автокорреляцией

Возникает вопрос: к чему приводит наличие корреляции случайной составляющей временного ряда? Использование обычного (классического) метода наименьших квадратов для оценивания коэффициентов объясненной части в этом случае также дает *несмещенные и состоятельные оценки*, но эти оценки *не являются эффективными* (т.е. существуют оценки с меньшей дисперсией). Более того, оценки  $s_{b_j}^2$  дисперсий коэффициентов  $b_j$  являются смещенными и несостоятельными, что приводит к недостоверным результатам проверки гипотез о значимости вычисленных коэффициентов  $b_j$ .

Как же определить наличие автокорреляции в наблюдаемых значениях временного ряда? Достаточно простой критерий, дающий ответ на наличие корреляции между соседними наблюдениями дает тест Дарбина–Уотсона, изложенный в п. 2.4.3. Этот тест основан на простой идее: если корреляция между  $\varepsilon_i$  и  $\varepsilon_{i+1}$  не равна нулю, то она присутствует и в остатках (невязках)  $e_i = y_i - \hat{y}_i$  регрессионной модели, где  $\hat{y}_i = \hat{q}(\tau_i)$  оценка объяс-

ненной части временного ряда, построенная обычным методом наименьших квадратов.

**Замечание 3.1.1.** Тест Дарбина–Уотсона неприменим в случаях, когда объясненная часть модели временного ряда  $q(\tau_i)$  коррелирована со случайными возмущениями  $\varepsilon_j$ . ♦

Допустим, что с помощью теста Дарбина–Уотсона (или другого теста) установлено наличие корреляции. Кроме этого справедливы следующие предположения о числовых характеристиках возмущений  $\varepsilon_i$ :  $M(\varepsilon_i) = 0$ ,  $D(\varepsilon_i) = \sigma^2$ . При этих предположениях возмущения  $\varepsilon_i$  можно рассматривать как *стационарный дискретный случайный процесс* (другими словами – *стационарный временной ряд*) с коррелированными значениями.

В качестве математических моделей для описания такого процесса используют модели двух классов:

- авторегрессионные модели;
- модели скользящего среднего.

Рассмотрим некоторые модели из этих классов.

### 3.1.2. Модель авторегрессии первого порядка (модель AR(1))

Эта модель описывает так называемый марковский процесс, состояние которого в каждый следующий момент времени определяется только состоянием в настоящий момент времени и не зависит от того, каким путем процесс достиг этого состояния. Модель AR(1) определяется соотношением

$$\varepsilon_i = \mu\varepsilon_{i-1} + \xi_i, \quad (3.1.2)$$

где  $\mu$  – коэффициент модели, часто называемый коэффициентом авторегрессии ( $|\mu| < 1$ );  $\xi_i$  – последовательность случайных величин, образующих «белый шум» (см. п. 1.2) с характеристиками:

$$M(\xi_i) = 0; \quad M(\xi_i \xi_{i \pm j}) = \begin{cases} \sigma^2, & \text{если } j = 0; \\ 0, & \text{если } j \neq 0. \end{cases}$$

Тогда из (3.1.2) следуют выражения для числовых характеристик процесса  $\varepsilon_i$ :

$$M(\varepsilon_i) = 0, \quad \sigma_{\varepsilon_i}^2 = D(\varepsilon_i) = \frac{\sigma^2}{1 - \mu^2}, \quad \rho(l) = \rho_{\varepsilon_i \varepsilon_{i \pm l}} = \mu^l, \quad (3.1.3)$$

где  $\rho(l)$  – коэффициент автокорреляции (1.2.3).

Условие стационарности ряда  $\varepsilon_i$  имеет вид

$$|\mu| < 1. \quad (3.1.4)$$

Из выражений (3.1.3) видно, что при  $|\mu|$  близком к 1 дисперсия случайной величины  $\varepsilon_i$  будет намного больше дисперсии  $\sigma^2$  и между соседними значениями имеется сильная корреляция (коэффициент корреляции  $\rho$  равен  $\mu$ ).

Для идентификации параметров  $\mu, \sigma^2$  модели AR(1), используются остатки (невязки)  $e_i = y_i - \hat{y}_i$ ,  $i = 1, 2, \dots, n$ , где  $y_i = \hat{q}(\tau_i)$  – оценка детерминированной составляющей временного ряда (см. главу 2). Оценки параметров определяются выражениями:

$$\hat{\mu} = \frac{1}{n-1} \sum_{i=1}^{n-1} (e_i - \bar{e})(e_{i+1} - \bar{e}) / s_e^2; \quad (3.1.5)$$

$$\hat{\sigma}^2 = (1 - \hat{\mu}^2) \cdot s_e^2, \quad (3.1.6)$$

где  $\bar{e} = \frac{1}{n} \sum_{i=1}^n e_i$ ;  $s_e^2 = \frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2$  – выборочные среднее и дисперсия остатков  $e_i$ .

### 3.1.3. Модель авторегрессии второго порядка (модель AR(2))

Модель определяется соотношением

$$\varepsilon_i = \mu_1 \varepsilon_{i-1} + \mu_2 \varepsilon_{i-2} + \xi_i, \quad (3.1.7)$$

где  $\xi_i$  – последовательность некоррелированных величин с характеристиками

$$M(\xi_j) = 0; \quad M(\xi_i \xi_{i+j}) = \begin{cases} \sigma^2, & \text{если } j = 0; \\ 0, & \text{если } j \neq 0. \end{cases} \quad (3.1.8)$$

Математическое ожидание и дисперсия процесса  $\varepsilon_i$  определяются соотношениями:

$$M(\varepsilon_i) = 0; \quad \sigma_{\varepsilon_i}^2 = D(\varepsilon_i) = \frac{\sigma^2}{\frac{1+\mu_2}{1-\mu_2} \left[ (1-\mu_2)^2 - \mu_1^2 \right]}, \quad (3.1.9)$$

а значение коэффициента автокорреляции определяется выражениями:

$$\begin{aligned} \rho(1) &= \frac{\mu_1}{1-\mu_2}; \quad \rho(2) = \mu_2 + \frac{\mu_1^2}{1-\mu_2}; \\ \rho(l) &= \mu_1 \rho(l-1) + \mu_2 \rho(l-2), \quad l = 3, 4, \dots \end{aligned} \quad (3.1.10)$$

Условие стационарности процесса  $\varepsilon_i$  имеет вид

$$\begin{cases} -1 < \frac{\mu_1}{1-\mu_2} < 1, \\ -1 < \mu_2 + \frac{\mu_1^2}{1-\mu_2} < 1 \end{cases} \Rightarrow \begin{cases} |\mu_1| < 2, \\ \mu_2 < 1 - |\mu_1|. \end{cases}$$

Оценки параметров  $\mu_1, \mu_2, \sigma^2$  модели AR(2) находятся по формулам:

$$\hat{\mu}_1 = \frac{r(1)(1-r(2))}{1-r^2(1)}; \quad (3.1.11)$$

$$\hat{\mu}_2 = \frac{r(2) - r^2(1)}{1-r^2(1)}; \quad (3.1.12)$$

$$\hat{\sigma}^2 = s_e^2 \cdot \frac{1+\hat{\mu}_2}{1-\hat{\mu}_1} \cdot \left[ (1-\hat{\mu}_2)^2 - \hat{\mu}_1^2 \right], \quad (3.1.13)$$

где

$$r(l) = \frac{1}{n-l} \sum_{i=1}^{n-l} (e_i - \bar{e})(e_{i+l} - \bar{e}), \quad l = 1, 2; \quad (3.1.14)$$

$$\bar{e} = \frac{1}{n} \sum_{i=1}^n e_i; \quad s_e^2 = \frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2.$$

### 3.1.4. Модель скользящего среднего первого порядка (или модель MA(1))

Модель определяется соотношением:

$$\varepsilon_i = \xi_i - \theta \xi_{i-1}, \quad (3.1.15)$$

где  $\xi_i$  – последовательность некоррелированных случайных величин с характеристиками:

$$M(\xi_i) = 0; \quad M(\xi_i \xi_{i+l}) = \begin{cases} \sigma^2, & \text{если } l = 0; \\ 0, & \text{если } l \neq 0. \end{cases}$$

Коэффициент автокорреляции процесса  $\varepsilon_i$ :

$$\rho(l) = \begin{cases} -\frac{\theta}{1+\theta^2}, & \text{если } l = 1; \\ 0, & \text{если } l \geq 2, \end{cases}$$

а дисперсия процесса  $\varepsilon_i$ :

$$\sigma_{\varepsilon_i}^2 = D(\varepsilon_i) = (1+\theta^2) \cdot \sigma^2.$$

Процесс стационарен при любом значении  $\theta$ .

Оценка параметра  $\theta$  равна одному из корней  $\theta_1, \theta_2$  квадратного уравнения

$$\theta^2 + \frac{1}{r(1)} \theta + 1 = 0, \quad (3.1.16)$$

который меньше 1 ( $\theta_1 \cdot \theta_2 = 1$ ), где  $r(l)$  вычисляется по формуле (3.1.14) при  $l=1$ . Такое значение корня обозначим  $\hat{\theta}$ . Оценка дисперсии вычисляется по формуле

$$\hat{\sigma}^2 = \frac{s_e^2}{1 + \hat{\theta}^2}, \quad (3.1.17)$$

где  $s_e^2 = \frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2$ ,  $\bar{e} = \frac{1}{n} \sum_{i=1}^n e_i$ .

### 3.1.5. Идентификация параметров моделей в Excel

Из приведенных выше выражений видно, что в оценку параметров моделей входят три величины:

$$\bar{e} = \frac{1}{n} \sum_{i=1}^n e_i, \quad s_e^2 = \frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2, \quad \text{cov}(l) = \frac{1}{n-l} \cdot \sum_{i=1}^{n-l} (e_i - \bar{e})(e_{i+l} - \bar{e}).$$

Значения этих величин удобно вычислять, используя стандартные функции Excel. Для конкретности предположим, что значения остатков  $e_i$  находятся в диапазоне ячеек A3÷A22 (т.е.  $n = 20$ ). Тогда

$$\bar{e} = \text{СРЗНАЧ}(A3:A22), \quad (3.1.18)$$

$$s_e^2 = \text{ДИСПР}(A3:A22), \quad (3.1.19)$$

$$\text{cov}(l) = \text{КОВАР}(A3:A22 - l; A3 + l:A22). \quad (3.1.20)$$

Подставляя вычисленные значения в соответствующие формулы, получаем нужные оценки параметров моделей.

**Пример 3.1.1.** В ячейках A3÷A22 документа Excel, приведенного на рис. 4.1, находятся значения  $\varepsilon_i$ , соответствующие авторегрессионной модели первого порядка (3.1.2) с параметрами  $\mu = 0.4$ ,  $\sigma^2 = 0.144$ . Эти значения показаны на графике в документе Excel (рис. 3.3). Необходимо по приведенной выборке оценить значения параметров модели и сравнить их с известными параметрами.

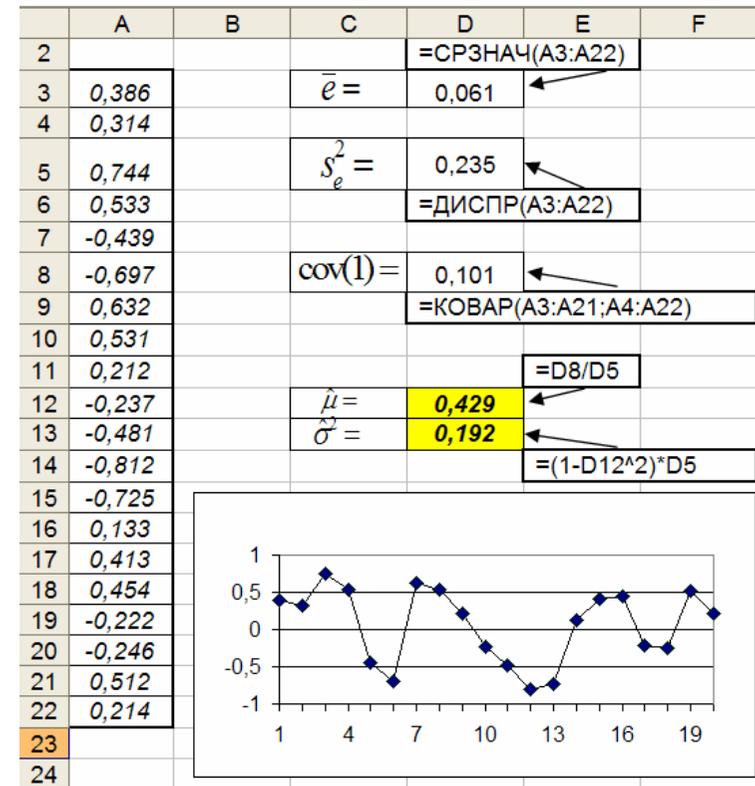


Рис. 3.3. Оценивание параметров модели  $AR(1)$

*Решение.* Будем считать, что  $e_i = \varepsilon_i$ . Тогда, используя выражения (3.1.18)–(3.1.20), вычисляем значения  $\bar{e} = 0.061$ ;  $s_e^2 = 0.235$ ;  $\text{cov}(1) = 0.101$ , в ячейках D3, D5, D8 соответственно. Затем, подставляя эти значения в формулы (3.1.5), (3.1.6), находим оценки параметров  $\hat{\mu} = 0.429$ ;  $\hat{\sigma}^2 = 0.192$ . Сравнивая эти значения с «точными» параметрами  $\mu = 0.4$ ,  $\sigma^2 = 0.144$ , отмечаем невысокую точность оценивания (порядка 20–30%). Это объясняется маленьким объемом выборки  $n = 20$ . ●

### 3.2. Обобщенный метод наименьших квадратов

При моделировании реальных экономических процессов допущения P2, P3 (см. п. 4.2) классической линейной регрессионной модели оказываются нарушенными. В частности, могут не выполняться предположения о том, что случайные возмущения  $\varepsilon_i$  имеют одинаковую дисперсию и некоррелированы между собой. Такая ситуация характерна для временных рядов с коррелированными возмущениями  $\varepsilon_i$  (см. п. 3.1). Использование обычного метода наименьших квадратов приводит к неэффективным оценкам коэффициентов функции регрессии, т.е. к оценкам, имеющим неминимальную дисперсию. Возникает вопрос: можно ли и в этих случаях построить эффективные оценки? Для ответа на этот вопрос рассмотрим новую модель множественной регрессии.

#### 3.2.1. Обобщенная линейная модель множественной регрессии

Пусть модель наблюдений линейной множественной регрессии имеет вид

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (3.2.1)$$

или в матричном виде

$$y = X\beta + \varepsilon, \quad (3.2.2)$$

где  $X, y, \beta, \varepsilon$  – матрица и векторы, определенные следующими формулами:

$$X = \begin{vmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{vmatrix}; \quad y = \begin{vmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{vmatrix}; \quad \beta = \begin{vmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{vmatrix}; \quad \varepsilon = \begin{vmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{vmatrix}.$$

Относительно  $X, \varepsilon$  сделаем следующие предположения:

P1. Матрица  $X$  размера  $n \times (k+1)$  является неслучайной матрицей.

P2. Вектор  $\varepsilon$  имеет нулевое среднее, т.е.  $M(\varepsilon) = \bar{0}$ , где  $\bar{0}$  – нулевой вектор.

P3. Ковариационная матрица  $V_\varepsilon = M(\varepsilon\varepsilon^T)$  вектора  $\varepsilon$  не является диагональной (напомним, что недиагональные элементы  $\mu_{ij}$  характеризуют степень коррелированности случайных величин  $\varepsilon_i$  и  $\varepsilon_j$ ).

P4. Ранг матрицы  $X$   $rank(x) = p + 1 \leq n$ .

Множественная регрессия (3.2.1), удовлетворяющая предположениям P1÷P4, получила название *обобщенной линейной модели множественной регрессии*. Сравнивая эту модель с классической [5, п. 3.1]), видим, что она отличается от классической только видом ковариационной матрицы (в классической  $V_\varepsilon = \sigma^2 I$ , где  $I$  – единичная матрица размера  $n \times n$ ).

Оценка  $b = (X^T X)^{-1} X^T y$  для обобщенной линейной модели также является несмещенной, состоятельной, но неэффективной. Ковариационная матрица  $V_b$  этой оценки

$$V_b = (X^T X)^{-1} X^T V_\varepsilon X (X^T X)^{-1}, \quad (3.2.3)$$

в то время как для классической модели

$$V_b = \sigma^2 (X^T X)^{-1}.$$

Для получения эффективной оценки будем использовать другую оценку, получаемую так называемым *обобщенным методом наименьших квадратов*.

#### 3.2.2. Обобщенный метод наименьших квадратов

Можно показать, что при выполнении предположений P1÷P4 обобщенной линейной множественной регрессии линейная оценка

$$b^* = (X^T V_\varepsilon^{-1} X)^{-1} X^T V_\varepsilon^{-1} y \quad (3.2.4)$$

имеет наименьшую ковариационную матрицу среди всех линейных несмещенных оценок для вектора коэффициентов  $\beta$ . Другими словами, дисперсии оценок  $b_j^*, j = 0, 1, \dots, k$  минимальны, т.е.

вектор  $b^*$  является эффективной оценкой вектора коэффициентов  $\beta$  модели обобщенной линейной множественной регрессии. Ковариационная матрица оценки  $b^*$  имеет вид (сравнить с выражением (3.2.3)):

$$V_{b^*} = (X^T V_\varepsilon^{-1} X)^{-1}. \quad (3.2.5)$$

Отметим, что оценка  $b^*$  минимизирует функционал

$$F^*(b) = (y - Xb)^T V_\varepsilon^{-1} (y - Xb), \quad (3.2.6)$$

называемый функционалом обобщенного метода наименьших квадратов.

**Замечание 3.2.1.** Для обобщенной линейной регрессионной модели, в отличие от классической, коэффициент детерминации, вычисленный по формуле (см. (2.1.18)):

$$R^2 = 1 - \frac{(y - Xb^*)^T (y - Xb^*)}{(y - \bar{y})^T (y - \bar{y})},$$

где  $\bar{y}$  – вектор размерности  $n$ , составленный из средних значений  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ , не является удовлетворительной мерой качества

построенного уравнения регрессии. В общем случае  $R^2$  можно выводить за пределы интервала  $[0;1]$ . Поэтому для обобщенной модели коэффициент детерминации  $R^2$  может использоваться лишь как приближенная характеристика качества построенного уравнения регрессии. ♦

Применение обобщенного метода наименьших квадратов требует знания элементов ковариационной матрицы  $V_\varepsilon$ , что в практике эконометрического моделирования встречается очень редко. Поэтому для практической реализации обобщенного МНК необходимо вводить дополнительные условия на структуру матрицы  $V_\varepsilon$  таким образом, чтобы элементы матрицы  $V_\varepsilon$  зависели от нескольких параметров. Рассмотрим пример такого задания  $V_\varepsilon$ .

### 3.3. Выделения тренда временного ряда на основе обобщенного метода наименьших квадратов

Вернемся к задаче выделения тренда  $t(\tau)$  временного ряда, представленного моделью (см. п. 2.1)

$$y(\tau_i) = t(\tau_i) + \varepsilon(\tau_i), \quad i = 1, 2, \dots, n. \quad (3.3.1)$$

В качестве уравнения тренда примем полином  $p$ -го порядка вида

$$t(\tau) = \beta_0 + \beta_1 \tau + \beta_2 \tau^2 + \dots + \beta_p \tau^p. \quad (3.3.2)$$

Тогда, введя новые переменные

$$x_1 = \tau, \quad x_2 = \tau^2, \quad \dots, \quad x_p = \tau^p,$$

приходим к следующей модели для измеренных значений временного ряда:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (3.3.3)$$

Далее предполагается, что возмущения  $\varepsilon_i$  имеют нулевое математическое ожидание, но коррелированы между собой, т.е. ковариационная матрица  $V_\varepsilon$  не является диагональной. Таким образом, приходим к модели обобщенной линейной множественной регрессии. В рамках этой модели несмещенная, состоятельная и эффективная оценка коэффициентов  $\beta_0, \beta_1, \dots, \beta_p$  определяется на основе обобщенного МНК и имеет вид (3.2.4). Но для построения такой оценки необходимо задать ковариационную матрицу  $V_\varepsilon$ . Определим матрицу  $V_\varepsilon$  для двух следующих моделей возмущений  $\varepsilon_i$ .

#### 3.3.1. Модель авторегрессии первого порядка

Эта модель возмущений  $\varepsilon_i$  определяется соотношением (3.1.2). С учетом числовых характеристик возмущений  $\varepsilon_i$  (см. (3.1.3)) получаем следующую ковариационную матрицу  $V_\varepsilon$  вектора возмущений  $\varepsilon$ :

$$V_\varepsilon = \sigma_\varepsilon^2 \begin{vmatrix} 1 & \mu & \mu^2 & \cdots & \mu^{n-1} \\ \mu & 1 & \mu & \cdots & \mu^{n-2} \\ \mu^2 & \mu & 1 & \cdots & \mu^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu^{n-1} & \mu^{n-2} & \mu^{n-3} & \cdots & 1 \end{vmatrix} = \sigma_\varepsilon^2 \cdot \Omega, \quad (3.3.4)$$

где матрица  $\Omega$  имеет элементы  $[\Omega]_{i,j} = \mu^{|i-j|}$ ;

$$\sigma_\varepsilon^2 = \frac{\sigma^2}{1-\mu^2}. \quad (3.3.5)$$

Можно показать, что обратная матрица  $\Omega^{-1}$  имеет трехдиагональную структуру:

$$\Omega^{-1} = \frac{1}{1-\mu^2} \begin{vmatrix} 1 & -\mu & 0 & \cdots & 0 & 0 \\ -\mu & (1+\mu^2) & -\mu & \cdots & 0 & 0 \\ 0 & -\mu & (1+\mu^2) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & (1+\mu^2) & -\mu \\ 0 & 0 & 0 & \cdots & -\mu & \mu \end{vmatrix} \quad (3.3.6)$$

и тогда  $V_\varepsilon^{-1}$  определяется как

$$V_\varepsilon^{-1} = \frac{1}{\sigma_\varepsilon^2} \cdot \Omega^{-1}. \quad (3.3.7)$$

Тогда оценка  $b^*$  обобщенного метода наименьших квадратов является решением следующих систем уравнений:

$$(X^T \Omega^{-1} X) b = X^T \Omega^{-1} y,$$

или

$$b^* = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} y. \quad (3.3.8)$$

Для вычисления вектора  $b^*$  достаточно знания параметра  $\mu$ , а для вычисления ковариационной матрицы  $V_{b^*}$ , определяемой соотношением

$$V_{b^*} = \sigma_\varepsilon^2 (X^T \Omega^{-1} X)^{-1},$$

необходимо знание дисперсии  $\sigma^2$  авторегрессионной модели возмущений временного ряда. В большинстве практических случаев параметры  $\mu, \sigma^2$  априори неизвестны и их приходится оценивать в процессе построения уравнения регрессии, используя следующую итерационную процедуру (обозначенную *IPAR(1)*).

### 3.3.2. Итерационная процедура *IPAR(1)*

Алгоритм процедуры *IPAR(1)*:

*Шаг 0.* По заданной выборке вычисляем оценку простого метода наименьших квадратов:

$$\hat{b}^{(0)} = (X^T X)^{-1} X^T y$$

и полагаем номер итерации  $l = 0$ .

*Шаг 1.* Вычисляется вектор невязки на  $l$ -й итерации:

$$e^{(l)} = y - X \hat{b}^{(l)}.$$

*Шаг 2.* По формуле (3.1.5) находим оценку  $\hat{\mu}^{(l)}$  на  $l$ -й итерации и формируем матрицу  $(\Omega^{(l)})^{-1}$  согласно (3.3.6).

*Шаг 3.* Вычисляем вектор

$$\hat{b}^{(l+1)} = (X^T (\Omega^{(l)})^{-1} X)^{-1} X^T (\Omega^{(l)})^{-1} y$$

и полагаем  $l = l + 1$ .

Шаги 1–3 повторяют до тех пор пока различие между  $\hat{\mu}^{(l)}$  и  $\hat{\mu}^{(l+1)}$  будут малы. Значение  $\hat{\mu}^{(l+1)}$ , при котором итерационная процедура закончилась, обозначим  $\hat{\mu}^*$ .

*Шаг 4.* По найденному значению  $\hat{\mu}^*$  вычисляем  $\hat{\sigma}^2$  (формула (3.1.6)) матрицу  $\hat{\Omega}^{-1}$  (формула (3.3.6)),  $\hat{\sigma}_\varepsilon^2$  (формула (3.3.5)) и находим вектор

$$\hat{b}^* = (X^T \hat{\Omega}^{-1} X)^{-1} X^T \hat{\Omega}^{-1} y \quad (3.3.9)$$

и ковариационную матрицу

$$\widehat{V}_{b^*} = \widehat{\sigma}_\varepsilon^2 \cdot (X^T \widehat{\Omega}^{-1} X)^{-1}. \quad (3.3.10)$$

Заметим, что часто ограничиваемся только одной или двумя итерациями описанной процедуры как это сделано в следующем примере.

**Пример 3.3.1.** В ячейках D3÷D17 документа Excel (рис. 3.4) приведены значения временного ряда  $y_i$ ,  $i = 1, 2, \dots, 15$  (т.е.  $n = 15$ ). Предполагается, что возмущения  $\varepsilon_i$  соответствуют авторегрессионной модели первого порядка (3.1.2). Необходимо выделить трендовую составляющую временного ряда, аппроксимируя ее полиномом второй степени

$$\widehat{t}(\tau) = b_0 + b_1\tau + b_2\tau^2.$$

*Решение.* Введем новые переменные  $x_1 = \tau$ ;  $x_2 = \tau^2$  и сформируем матрицу  $X$  размером  $15 \times 3$ , элементы которой находятся в ячейках A3÷C17 (см. рис. 3.4). Выполним одну итерацию итерационной процедуры *IPAR*(1) и вычислим оценки параметров авторегрессионной модели (3.1.2):  $\widehat{\mu} = 0.403$ ,  $\widehat{\sigma}^2 = 0.106$ . Этот шаг в документе на рис. 3.4 не отражен из-за недостатка места.

Сформируем матрицу  $V_\varepsilon$ , элементы которой вычисляются по формуле

$$\{\widehat{V}_\varepsilon\}_{i,j} = \frac{\widehat{\sigma}^2}{1 - \widehat{\mu}^2} \cdot \widehat{\mu}^{|i-j|}, \quad i, j = 1, 2, \dots, n$$

и разместим их в диапазоне I3:W17 (на рисунке не показаны). Затем вычисляем: матрицу  $X^T \widehat{V}_\varepsilon^{-1} X$  размером  $3 \times 3$  (ячейки E3÷G5), обратную матрицу  $(X^T \widehat{V}_\varepsilon^{-1} X)^{-1}$  размером  $3 \times 3$  (ячейки E8÷G10),

вектор  $X^T \widehat{V}_\varepsilon^{-1} y$ , содержащий три проекции (ячейки E14÷E16). Далее по формуле

$$\widehat{b}^* = (X^T \widehat{V}_\varepsilon^{-1} X)^{-1} X^T \widehat{V}_\varepsilon^{-1} y$$

вычислим вектор  $\widehat{b}^*$  (ячейки G14÷G16)

$$\widehat{b}^* = \begin{vmatrix} \widehat{b}_0^* \\ \widehat{b}_1^* \\ \widehat{b}_2^* \end{vmatrix} = \begin{vmatrix} 1.486 \\ 1.434 \\ -0.336 \end{vmatrix}.$$

Таким образом, уравнение трендовой составляющей принимает вид

$$\widehat{t}(\tau) = 1.486 + 1.434\tau - 0.336\tau^2.$$

На рис. 3.5 показаны исходные значения  $y_i$  (кривая 1 отмечена квадратными маркерами) и значения  $\widehat{t}(\tau_i)$  (кривая 2 – треугольные маркеры). ☺

Заметим, что применение простого метода наименьших квадратов приводит к уравнению

$$\widehat{t}(\tau) = 2.674 + 0.118\tau - 0.010\tau^2,$$

коэффициенты которого существенно отличаются от коэффициентов, найденных обобщенным методом наименьших квадратов.

	A	B	C	D	E	F	G
1	=МУМНОЖ(ТРАНСП(А3:С17);МУМНОЖ(МОБР(І3:W17);А3:С17))						
2							
3	1	1	1,000	2,113	52,310	125,544	345,605
4	1	1,2	1,440	2,908	125,544	347,631	1051,811
5	1	1,4	1,960	3,119	345,605	1051,811	3382,906
6	1	1,6	2,560	3,567			
7	1	1,8	3,240	3,401			
8	1	2	4,000	3,284	0,893	-0,784	0,152
9	1	2,2	4,840	2,706	-0,784	0,737	-0,149
10	1	2,4	5,760	3,144	0,152	-0,149	0,031
11	1	2,6	6,760	2,707			
12	1	2,8	7,840	2,339	=МУМНОЖ(Е8:G10;Е14:Е16)		
13	1	3	9,000	2,348			
14	1	3,2	10,240	2,225	141,800		1,486
15	1	3,4	11,560	2,616	332,162	$\hat{b}^*$	1,434
16	1	3,6	12,960	2,664	886,888		-0,335
17	1	3,8	14,440	2,241			
18							
19	=МУМНОЖ(ТРАНСП(А3:С17);МУМНОЖ(МОБР(І3:W17);D3:D17))						

Рис. 3.4. Вычисление коэффициентов обобщенным МНК

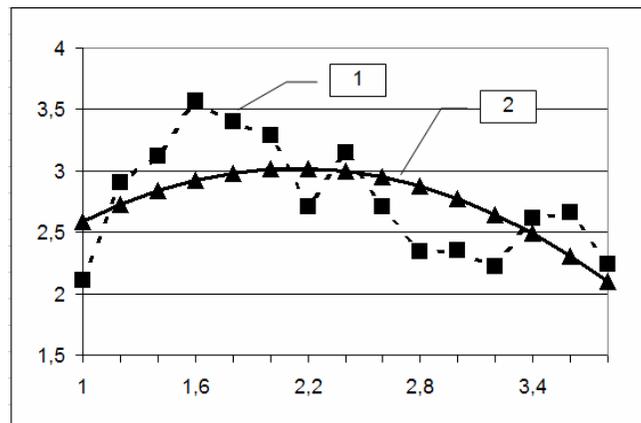


Рис. 3.5. Исходные данные и трендовая составляющая

### 3.3.3. Модель скользящего среднего первого порядка

Эта модель возмущений  $\varepsilon_i$  определяется соотношением (3.1.17). Ковариационная матрица  $V_\varepsilon$  вектора возмущений  $\varepsilon$  имеет вид

$$V_\varepsilon = \sigma_\varepsilon^2 \cdot \begin{pmatrix} 1 & -\lambda & 0 & \dots & 0 \\ -\lambda & 1 & -\lambda & \dots & 0 \\ 0 & -\lambda & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & -\lambda \\ 0 & 0 & 0 & -\lambda & 1 \end{pmatrix}, \quad (3.3.11)$$

где  $\sigma_\varepsilon^2 = (1 + \theta) \cdot \sigma^2$ ,  $\lambda = \frac{\theta}{1 + \theta^2}$ . Матрица  $V_\varepsilon$  имеет трехдиагональную структуру, и ее элементы зависят от параметров  $\theta, \sigma^2$  модели скользящего среднего. Если эти параметры априори неизвестны, то для их оценивания можно использовать итерационную процедуру, аналогичную описанной *IPAR(1)*. Ограничимся только записью одной итерации такой процедуры:

*Шаг 0.* По заданной выборке вычисляется оценка

$$\hat{b} = (X^T X)^{-1} X^T y.$$

*Шаг 1.* Вычисляется вектор невязки  $e = y - X\hat{b}$ .

*Шаг 2.* По формулам (3.1.16), (3.1.17) вычисляются оценки  $\hat{\theta}, \hat{\sigma}^2$  и формируются элементы матрицы  $\hat{V}_\varepsilon$ :

$$\{\hat{V}_\varepsilon\}_{i,j} = (1 + \hat{\theta}^2) \cdot \hat{\sigma}^2 \cdot \psi_{i,j}, \quad (3.3.12)$$

$$\text{где } \psi_{i,j} = \begin{cases} 1, & \text{если } i = j; \\ -\frac{\hat{\theta}}{1 + \hat{\theta}^2}, & \text{если } |i - j| = 1; \\ 0, & \text{во всех остальных случаях.} \end{cases}$$

*Шаг 3.* Подставляя матрицу  $\hat{V}_\varepsilon$  в (3.2.4), получаем оценку

$$\hat{b}^* = (X^T \hat{V}_\varepsilon^{-1} X)^{-1} X^T \hat{V}_\varepsilon^{-1} y$$

для вектора  $\beta$  коэффициентов полиномиального тренда. Ковариационная матрица вектора  $\hat{b}^*$  определяется выражением

$$V_{\hat{b}^*} = (X^T \hat{V}_\varepsilon^{-1} X)^{-1}. \quad (3.3.13)$$

Использование обобщенного метода наименьших квадратов при значительной корреляции возмущений  $\varepsilon_i$  (например,  $\rho_{\varepsilon_i, \varepsilon_{i+1}} > 0.4$ ) позволяет более точно оценить коэффициенты уравнения трендовой составляющей временного ряда.

### ЛАБОРАТОРНАЯ РАБОТА

#### Вычисление параметров модели AR(1) возмущений временного ряда

*Исходные данные.* Значения временного ряда  $\{\tau_i, y_i\}$ ,  $i=1, \dots, 15$  приведены в столбцах А, В фрагмента документа, показанного на рисунке.

*Необходимо:* Оценить параметры  $\mu, \sigma^2$  авторегрессионной модели возмущений (3.1.2).

*Рекомендации:*

1. Первоначально по данным  $\{\tau_i, y_i\}, i=1, \dots, 15$ , используя режим **Регрессия**, выделить трендовую составляющую (см. п. 2.2), которая описывается уравнением

$$\hat{t}(\tau) = -29 + 21.61\tau - 2.86\tau^2.$$

2. По этому уравнению вычислить значения  $\hat{t}_i = \hat{t}(\tau_i)$  (столбец С) и значения  $e_i = y_i - \hat{t}_i$  (столбец D).

3. По формулам (3.1.5), (3.1.6) вычислить следующие оценки  $\hat{\mu}, \hat{\sigma}^2$  (ячейки В18, В19):

$$\hat{\mu} = 0.479, \quad \hat{\sigma}^2 = 0.678.$$

	A	B	C	D	E	F	G
1	$\tau_i$	$y_i$	$\hat{t}_i$	$=-29+21,61*A2-2,86*A2^2$			
2	2	3,92	2,78	1,14		1,30	
3	2,1	4,2	3,77	0,43	$=D2^2$	0,19	
4	2,2	4,06	4,70	-0,64		0,41	
5	2,3	3,51	5,57	-2,06		4,26	
6	2,4	5,26	6,39	-1,13		1,28	
7	2,5	6,05	7,15	-1,10		1,21	
8	2,6	7,3	7,85	-0,55		0,31	
9	2,7	9,71	8,50	1,21		1,47	
10	2,8	9,57	9,09	0,48		0,23	
11	2,9	9,99	9,62	0,37		0,14	
12	3	10,34	10,09	0,25		0,06	
13	3,1	10,84	10,51	0,33		0,11	
14	3,2	11,12	10,87	0,25		0,06	
15	3,3	9,84	11,17	-1,33		1,76	
16	3,4	10,78	11,41	-0,63		0,40	
17				-0,198	0,421	0,879	
18	$\hat{\mu} =$	0,479				$=\text{СУММ}(F2:F16)/15$	
19	$\hat{\sigma}^2 =$	0,678	$=\text{СУММ}(D2:D16)/15$		$=\text{КОВАР}(D2:D15;D3:D16)$		

### КОНТРОЛЬНАЯ РАБОТА

*Исходные данные.* В таблице приведены значения временного ряда  $\{\tau_i, y_i\}, i=1, \dots, 14$ . Известно, что возмущения  $\varepsilon_i$  коррелированы и описываются моделью скользящего среднего первого порядка (3.1.15).

$\tau_i$	0	0.1	0.2	0.3	0.4	0.5	0.6
$y_i$	0.4+N	0.8+N	0.3+N	N	2.0+N	1.4+N	2.3+N
$\tau_i$	0.7	0.8	0.9	1.0	1.1	1.2	1.3
$y_i$	3.9+N	2.2+N	3.3+N	3.4+N	3.9+N	4.1+N	2.9+N

Примечание. N – последняя цифра в изменяющемся номере зачетной книжке.

*Необходимо:*

1. Выделить трендовую составляющую временного ряда в классе полиномов до третьего порядка включительно. Отбор наилучшего полинома  $\hat{t}_{opt}(\tau)$  осуществить по величине приведенного индекса детерминации (см. замечание 2.1.2).

2. Вычислить остатки  $e_i = y_i - \hat{t}_{opt}(\tau_i), i = 1, \dots, 14$ .

3. Построить графики значений  $y_i, \hat{t}_{opt}(\tau_i), e_i$ .

4. Вычислить оценки для параметров  $\theta, \sigma^2$  модели (3.1.15) (см. п. 3.1.4).

5. В контрольную работу включить фрагменты документов Excel, содержащие

- вычисление коэффициентов всех полином до третьего порядка включительно;
- вычисление приведенного индекса детерминации;
- вычисление значений  $y_i, \hat{t}_{opt}(\tau_i), e_i$  и их график;
- вычисление оценок для параметров  $\theta, \sigma^2$ .

*Рекомендации:*

1. При выборе наилучшего полинома использовать пример 2.2.2. Следует отметить, что иногда коэффициенты полинома, вычисляемые с помощью команды *Добавить линию тренда*, являются ошибочными и не соответствуют кривой, график которой приводится на соответствующем рисунке. В этом случае коэффициенты следует вычислять с использованием режима **Регрессия**, команды *Анализ данных* пункта меню **Сервис**.

2. При вычислении оценок для параметров  $\theta, \sigma^2$  программирование соответствующих выражений выполнить по аналогии с лабораторной работой третьей главы.

## ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Что такое коррелированные возмущения временного ряда?

2. Как проявляется положительная и отрицательная автокорреляции возмущений временного ряда?

3. Как записывается модель авторегрессии первого порядка?

4. Как записывается модель авторегрессии второго порядка?

5. Как записывается модель скользящего среднего первого порядка?

6. Какие недостатки имеют оценки, вычисленные классическим методом наименьших квадратов, когда возмущения  $\varepsilon_i$  коррелированы между собой?

7. В каких случаях применяется обобщенный метод наименьших квадратов?

8. Сущность применение обобщенного метода наименьших квадратов для выделения трендовой составляющей при коррелированных возмущениях временного ряда.

## Глава 4. АВТОРЕГРЕССИОННЫЕ МОДЕЛИ ВРЕМЕННОГО РЯДА

### 4.1. Определение авторегрессионной модели

Ранее, при выделении тренда (см. главу 2), в качестве объясняющей переменной (или регрессора) модели временного ряда выступало время  $\tau$ . Однако в эконометрике достаточно широкое распространение получили регрессионные модели, в которых регрессорами выступают *лаговые переменные*, влияние которых характеризуется некоторым «запаздыванием». Наиболее часто в качестве такой модели используются *авторегрессионные модели* вида (1.1.4). Использование авторегрессионных моделей для описания временного ряда основано на предположении о том, что текущее значение временного ряда может быть выражено в виде линейной комбинации некоторого количества предыдущих его значений и случайной величины, обладающей свойствами «белого» шума.

Общий вид *модели авторегрессии  $p$ -го порядка* (или модель AR(p)) имеет вид

$$y_i = \beta_0 + \beta_1 y_{i-1} + \beta_2 y_{i-2} + \dots + \beta_p y_{i-p} + \varepsilon_i, i = 1, 2, \dots, n, \quad (4.1.1)$$

где  $\beta_0, \beta_1, \dots, \beta_p$  – коэффициенты модели,  $y_{i-j}$  – лаговые переменные, определяющие зависимость значения  $y_i$  временного ряда в момент  $\tau_i$  от значений в предыдущие моменты времени. Эти лаговые переменные и выступают в роли регрессоров (объясняющих переменных). Возмущения  $\varepsilon_i$  удовлетворяют условиям Гаусса–Маркова:

$$\mathbf{P1.} \quad M(\varepsilon_i) = 0. \quad (4.1.2)$$

$\mathbf{P2.}$  Корреляционные моменты случайных величин  $\varepsilon_i, \varepsilon_j$  удовлетворяют условию

$$M(\varepsilon_i \varepsilon_j) = \begin{cases} \sigma^2, & \text{если } i = j; \\ 0, & \text{если } i \neq j. \end{cases} \quad (4.1.3)$$

Первая строка означает *постоянство дисперсии возмущений  $\varepsilon_i$* , и это свойство называют *гомоскедастичностью*. Зависимость дисперсии возмущения  $\varepsilon_i$  от номера наблюдения  $i$  называется *гетероскедастичностью*.

$\mathbf{P3.}$  Случайные величины  $\varepsilon_i$  подчиняются нормальному распределению с нулевым средним и дисперсией  $\sigma^2$ .

Для вектора

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (4.1.4)$$

условия Гаусса–Маркова примут вид:

$$\mathbf{P1.} \quad M[\varepsilon] = 0_n, \quad (4.1.5)$$

где  $0_n$  – вектор, все  $n$  проекций которого равны нулю (т.е. нулевой вектор).

$$\mathbf{P2.} \quad V_\varepsilon = M[\varepsilon \varepsilon^T] = \sigma^2 I, \quad (4.1.6)$$

где  $V_\varepsilon$  – ковариационная матрица размера  $n \times n$ ;  $I$  – единичная матрица размера  $n \times n$ . Напомним, что  $i, i$ -й элемент ковариационной матрицы  $V_\varepsilon$  определяет дисперсию  $i$ -й проекции вектора  $\varepsilon$ , а  $i, j$ -й элемент равен корреляционному моменту  $\mu_{i,j} = M[\varepsilon_i \cdot \varepsilon_j]$ . Если проекции  $\varepsilon_i$  и  $\varepsilon_j$  статистически независимы, то  $\mu_{i,j} = 0$  и матрица  $V_\varepsilon$  является диагональной.

$\mathbf{P3.}$  Случайный вектор  $\varepsilon$  подчиняется нормальному распределению  $N(0_n, \sigma^2 I)$ .

Наиболее часто используются:

• *авторегрессионная модель первого порядка* (или модель AR(1)):

$$y_i = \beta_0 + \beta_1 y_{i-1} + \varepsilon_i, i = 1, 2, \dots, n; \quad (4.1.7)$$

• авторегрессионная модель второго порядка (или модель AR(2)):

$$y_i = \beta_0 + \beta_1 y_{i-1} + \beta_2 y_{i-2} + \varepsilon_i, i=1, 2, \dots, n. \quad (4.1.8)$$

Построение модели типа (4.1.1) предполагает решение двух взаимосвязанных задач:

- определение порядка  $p$  авторегрессионной модели временного ряда;
- оценивание коэффициентов  $\beta_i$ .

Рассмотрим сначала задачу оценивания коэффициентов  $\beta_i$ .

#### 4.2. Оценивание коэффициентов авторегрессионной модели

Так как лаговые переменные по своей сути являются объясняющими переменными, то модель (4.1.1) можно рассматривать как линейную множественную регрессию, коэффициенты которой можно оценить на основе классического метода наименьших квадратов, записав для модели (4.1.1) следующее уравнение регрессии:

$$\hat{y}_i = b_0 + b_1 y_{i-1} + b_2 y_{i-2} + \dots + b_p y_{i-p}, \quad i=1, 2, \dots, n, \quad (4.2.1)$$

коэффициенты которого являются оценками для  $\beta_0, \beta_1, \dots, \beta_p$  соответственно.

Тогда матрица  $X$  размера  $n \times (p+1)$ , входящая в матричную запись системы нормальных уравнений, имеет вид

$$X = \begin{pmatrix} 1 & y_0 & y_{-1} & \dots & y_{-p+1} \\ 1 & y_1 & y_0 & \dots & y_{-p+2} \\ 1 & y_2 & y_1 & \dots & y_{-p+3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & y_{n-1} & y_{n-2} & \dots & y_{n-p} \end{pmatrix}. \quad (4.2.2)$$

Для моделей AR(1), AR(2) матрицы  $X$  содержат элементы:

$$X = \begin{pmatrix} 1 & y_0 \\ 1 & y_1 \\ 1 & y_2 \\ \vdots & \vdots \\ 1 & y_{n-1} \end{pmatrix}; \quad X = \begin{pmatrix} 1 & y_0 & y_{-1} \\ 1 & y_1 & y_0 \\ 1 & y_2 & y_1 \\ \vdots & \vdots & \vdots \\ 1 & y_{n-1} & y_{n-2} \end{pmatrix}. \quad (4.2.3)$$

**Замечание 4.2.1.** Начальные значения  $y_0, y_{-1}, \dots, y_{-p+1}$ , входящие в матрицу  $X$ , как правило, неизвестны, и поэтому их необходимо доопределить. Например,  $y_0 = \bar{y}$  – среднее значение наблюдаемых значений,  $y_{-1} = y_{-2} = \dots = y_{-p+1} = 0$ . Этот способ используется в примере 4.2.1. Возможны и другие способы задания этих начальных значений. Например, сдвинуть точку отсчета временного ряда на несколько значений вправо. Так, если  $p = 2$ , то

$$y_{-1} = y_1, y_0 = y_2, y_1 = y_3, \dots, y_{n-p} = y_n, \quad (4.2.4)$$

что позволяет устранить «произвол» в определении начальных значений. ♦

Определив вектор наблюдений  $y$  как  $y = |y_1, y_2, \dots, y_n|^T$ , приходим к системе нормальных уравнений:

$$X^T X b = X^T y,$$

где вектор  $b = |b_0, b_1, \dots, b_p|$  содержит искомые оценки для коэффициентов  $\beta_0, \beta_1, \dots, \beta_p$  авторегрессионной модели. Предполагая, что обратная матрица  $(X^T X)^{-1}$  существует, находим вектор  $b$ :

$$b = (X^T X)^{-1} X^T y. \quad (4.2.5)$$

После этого уравнение регрессии (4.2.1) можно использовать для прогнозирования временного ряда, описываемого авторегрессионной моделью.

**Замечание 4.2.2.** Существенным предположением, при котором вектор  $b$  является несмещенной и состоятельной оценкой

для вектора коэффициентов  $\beta$  является предположение, что матрица  $X$  является неслучайной [5, с. 72]. В нашем случае матрица  $X$  (4.2.2) является случайной, так как элементами ее являются случайные величины – лаговые переменные  $y_{i-1}, \dots, y_{i-p}$ . Нарушение этого предположения приводит к тому, что вектор  $b$  уже **не является несмещенной и состоятельной оценкой** для вектора  $\beta$ .

**Пример 4.2.1.** В табл. 4.1 представлены данные, отражающие динамику курса акций некоторой компании (в условных единицах).

Таблица 4.1

$i$	1	2	3	4	5	6	7
$y_i$	971	1166	1044	907	957	727	752
$i$	8	9	10	11	12	13	14
$y_i$	1019	972	815	823	1112	1386	1428
$i$	15	16	17	18	19	20	21
$y_i$	1364	1241	1145	1351	1325	1226	1189

Необходимо по этим данным построить модель AR(1), AR(2) и определить их значимость при уровне значимости  $\alpha = 0.05$ , а также построить полиномиальные тренды  $\hat{t}(\tau)$  первого, второго порядка и проверить их значимость. Построить прогноз для  $i = 22, 23$ .

*Решение.* Первоначально, используя команду *Добавить линию тренда* (см. п. 2.2.1), определим коэффициенты уравнений:

линейного тренда ( $p = 1, m = p + 1 = 2$ )

$$\hat{t}(\tau) = b_0 + b_1\tau; \quad (4.2.6)$$

квадратичного тренда ( $p = 2, m = 3$ )

$$\hat{t}(\tau) = b_0 + b_1\tau + b_2\tau^2 \quad (4.2.7)$$

и вычислим для каждого уравнения тренда значения величин  $R^2, \hat{R}^2, F$  по следующим формулам:

$$R^2 = 1 - \frac{Q_e}{Q} = 1 - \frac{(y - Xb)^T (y - Xb)}{(y - \bar{y})^T (y - \bar{y})} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}; \quad (4.2.8)$$

$$\hat{R}^2 = 1 - \frac{n-1}{n-m} \cdot (1 - R^2); \quad (4.2.9)$$

$$F = \frac{Q_r \cdot (n-m)}{Q_e \cdot (m-1)}, \quad (4.2.10)$$

где  $\hat{y}_i$  – значение, вычисленное по модели временного ряда при  $\tau = \tau_i$ ,  $\bar{y}$  – вектор размерности  $n$ , составленный из средних значений  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ , а суммы  $Q_e, Q_r$  определяются выражениями

$$Q_r = b^T X^T y - n(\bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2;$$

$$Q_e = y^T y - b^T X^T y = \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

Значения  $R^2, \hat{R}^2, F$  и соответствующие уравнения трендов приведены в табл. 4.2.

Таблица 4.2

Уравнение тренда	$R^2$	$\hat{R}^2$	$F$	Кван- тиль
$\hat{t}(\tau) = 854.66 + 21.52\tau$ ( $p = 1, m = 2$ )	0.379	0.346	11.62	4.38
$\hat{t}(\tau) = 943.83 - 1.74\tau + 1.06\tau^2$ ( $p = 2, m = 3$ )	0.406	0.373	12.99	3.55

Определим квантиль  $F_{0.95;m-1;m-n}$   $F$ -распределения по известной формуле:

$$F_{0.95;m-1;m-n} = \text{FRАСПОБР}(0.05; m-1; m-n)$$

при  $n = 21$ . Значение квантиля, играющего роль границы критической области при проверке гипотезы о значимости построенной регрессии также приведены в табл. 4.2. Из этой табл. видно, что неравенство

$$F > F_{0.95;m-1;m-n} \quad (4.2.11)$$

выполняется как для линейного, так и для квадратичного тренда. Поэтому можно сказать, что построенные модели тренда значимы при уровне значимости 0.05.

На рис. 4.1 нанесены значения временного ряда  $y_i$  (отображены маркерами в форме квадратов – кривая 1) и значения квадратичного тренда (кривая 2). Видно, что, несмотря на принятие гипотезы о значимости, уравнение тренда не отображает динамику рассматриваемого временного ряда, а представляет «усредненную линию» его поведения. Поэтому перейдем к построению авторегрессионных моделей для этого временного ряда.

Уравнение регрессии для модели AR(1), определенной выражением (4.1.7) имеет вид

$$\hat{y}_i = b_0 + b_1 y_{i-1}. \quad (4.2.12)$$

Для вычисления коэффициентов  $b_0, b_1$ , которые являются оценками для  $\beta_0, \beta_1$  модели (4.1.7) используем классический метод наименьших квадратов, реализованный в режиме **Регрессия** модуля *Анализ данных* [5, с. 94]. Вычисленные значения:  $b_0 = 280.90$ ,  $b_1 = 0.746$ . Значения величин  $R^2, \hat{R}^2, F$  приведены в табл. 4.3 (вторая строка).

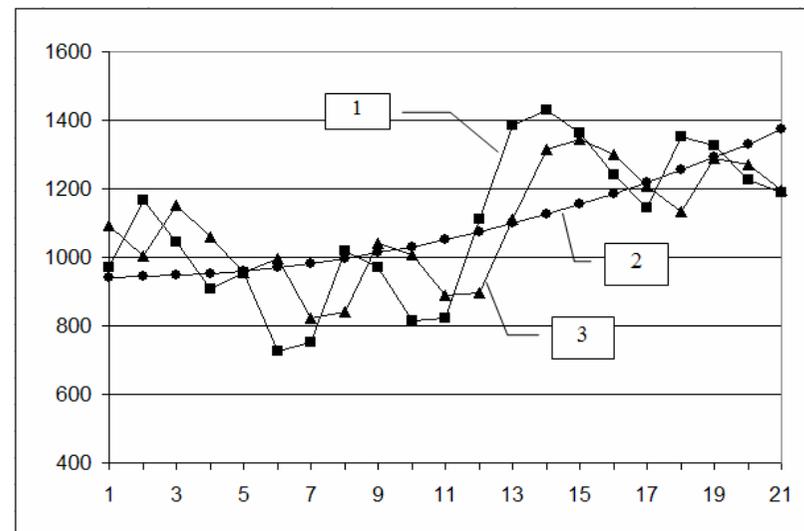


Рис. 4.1. Графики значений  $y_i, \hat{y}_i$  (пример 4.2.1)

Уравнение регрессии для модели AR(2), определенной формулой (4.1.8) имеет вид

$$\hat{y}_i = b_0 + b_1 y_{i-1} + b_2 y_{i-2}. \quad (4.2.13)$$

Используя режим **Регрессия** модуля *Анализ данных*, вычисляем коэффициенты  $b_0 = 275.61$ ,  $b_1 = 0.730$ ,  $b_2 = 0.022$ . В табл. 4.3 приведены значения  $R^2, \hat{R}^2, F$  для уравнения (4.2.13).

Из табл. 4.3 видно:

- а) уравнения (4.2.12), (4.2.13) имеют существенно большие значения  $R^2, \hat{R}^2, F$  по сравнению с квадратичным трендом;
- б) уравнения (4.2.12), (4.2.13) при вычисленных коэффициентах  $b_0, b_1, b_2$  являются значимыми при  $\alpha = 0.05$  и практически имеют одни и те же характеристики.

Таблица 4.3

Модель авторегрессии	$R^2$	$\widehat{R}^2$	$F$	Кван- тиль
$\widehat{y}_i = 280.90 + 0.746y_{i-1}$	0.553	0.528	23.26	4.38
$\widehat{y}_i = 275.61 + 0.730y_{i-1} + 0.022y_{i-2}$	0.554	0.525	23.22	3.55

Поэтому, исходя из принципа минимальной сложности, для прогнозирования значений временного ряда будет использоваться уравнение:

$$\widehat{y}_i = 280.90 + 0.746y_{i-1}.$$

Для  $i = 22$  прогнозное значение

$$\widehat{y}_{22} = 280.90 + 0.746 \cdot 1189 = 1168.$$

Для  $i = 23$  прогнозное значение

$$\widehat{y}_{23} = 280.90 + 0.746\widehat{y}_{22} = 280.90 + 0.746 \cdot 1168 = 1152. \bullet$$

### 4.3. Оценивание коэффициентов авторегрессионной модели стационарного временного ряда

Предположим, что рассматриваемый временной ряд  $\{Y(\tau_i)\}$  представлен выборкой  $y_i$ ,  $i = 1, 2, \dots, n$ . Этот ряд является стационарным с нулевым математическим ожиданием

$$M(Y(\tau_i)) = 0, \quad i = 1, 2, \dots, n. \quad (4.3.1)$$

Если это условие не выполняется, то следует рассмотреть *центрированные* значения

$$z_i = y_i - \bar{y}, \quad i = 1, 2, \dots, n, \quad (4.3.2)$$

где  $\bar{y}$  – оценка (1.2.3) математического ожидания  $M(Y(\tau_i))$ .

Обратимся к уравнению (4.1.1), в котором  $\beta_0$  следует положить равным 0 (из-за выполнения условия (4.3.1)):

$$y_i = \beta_1 y_{i-1} + \beta_2 y_{i-2} + \dots + \beta_p y_{i-p} + \varepsilon_i \quad (4.3.3)$$

Умножим левую и правую части на  $y_{i-m}$  и возьмем математическое ожидание от получившихся произведений

$$M[y_i y_{i-m}] = \beta_1 M[y_{i-1} y_{i-m}] + \beta_2 M[y_{i-2} y_{i-m}] + \dots + \beta_m M[y_{i-m} y_{i-m}] + \dots + M[y_{i-m} \varepsilon_i] \quad (4.3.4)$$

Так как случайные величины  $\varepsilon_i$  удовлетворяют условию Гаусса–Маркова, то

$$M[y_{i-m} \varepsilon_i] = 0, \quad m = 1, 2, \dots \quad (4.3.5)$$

Далее в силу стационарности временного ряда

$$M[y_{i-k} y_{i-m}] = M[y_i y_{i-(m-k)}]. \quad (4.3.6)$$

Обозначим ковариационный момент

$$\mu_k = M[y_i y_{i-k}] \quad (4.3.7)$$

и с учетом (4.3.5), (4.3.6), (4.3.7) перепишем (4.3.4) в виде

$$\mu_m = \beta_1 \mu_{m-1} + \beta_2 \mu_{m-2} + \beta_m \mu_0 + \dots + \beta_k \mu_{m-k}, \quad m = 1, 2, \dots, k. \quad (4.3.8)$$

Величина  $\mu_0 = M[y_m y_m]$  есть дисперсия  $\sigma_y^2$  значений временного ряда и поэтому поделив левую и правую части (4.3.8) на  $\sigma_y^2 = \mu_0$ , получаем систему уравнений в терминах коэффициентов автокорреляции:

$$\rho(m) = \beta_1 \rho(m-1) + \beta_2 \rho(m-2) + \dots + \beta_m + \beta_k \rho(m-k), \quad m = 1, 2, \dots, p. \quad (4.3.9)$$

Эта система содержит  $p$  уравнений относительно  $p$  неизвестных  $\beta_1, \beta_2, \dots, \beta_p$ . Перепишем систему уравнений (4.3.9) в матричном виде

$$Y\beta = \rho, \quad (4.3.10)$$

где  $Y$  – матрица размера  $p \times p$  имеет структуру

$$Y = \begin{pmatrix} 1 & \rho(1) & \dots & \rho(p-1) \\ \rho(1) & 1 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \rho(p-1) & \rho(p-2) & \dots & 1 \end{pmatrix}, \quad (4.3.11)$$

а векторы  $\beta$ ,  $\rho$  состоят из  $\beta$  проекций

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \rho = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{pmatrix}. \quad (4.3.12)$$

Решая систему (4.3.10), вычисляем вектор  $\beta$ , а следовательно, коэффициенты  $\beta_1, \beta_2, \dots, \beta_p$ , входящие в уравнение (4.3.3).

К сожалению, на практике коэффициенты автокорреляции  $\rho(m)$ ,  $m = 1, 2, \dots, p$  точно вычислить нельзя и вместо них в матрицу  $Y$  и вектор  $\rho$  подставляют выборочные коэффициенты автокорреляции  $r(m)$ , вычисляемые по формуле (1.2.5). В этом случае приходим к системе уравнений

$$Rb = r, \quad (4.3.13)$$

где

$$R = \begin{pmatrix} 1 & r(1) & \dots & r(p-1) \\ r(1) & 1 & & r(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & \dots & 1 \end{pmatrix}; \quad r = \begin{pmatrix} r(1) \\ r(2) \\ \vdots \\ r(p) \end{pmatrix}; \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix}. \quad (4.3.14)$$

Эту систему уравнений часто называют уравнениями Юла–Уокера.

Проекции  $b_1, b_2, \dots, b_p$  вектора  $b$  являются только оценками для коэффициентов  $\beta_1, \beta_2, \dots, \beta_p$ . Так как выборочные коэффициенты автокорреляции  $r(l)$  являются случайными величинами, то матрица  $R$  содержит случайные элементы (так же, как и матрица  $X$  (4.2.2)) и поэтому вектор  $b$  является уже смещенной оценкой для вектора  $\beta$ . Неэффективность оценки  $b$  может быть вызвана (при большом значении порядка  $p$ ) плохой обусловленностью матрицы  $R$  (т.е. матрица  $R$  имеет большое число обусловленности).

Напомним [5, с. 113], что числом обусловленности квадратной обратимой матрицы  $R$  называется величина

$$\text{cond}(R) = \|R\| \cdot \|R^{-1}\|, \quad (4.3.15)$$

где евклидова норма матрицы  $\|R\|$  определяется выражением

$$\|R\| = \left( \sum_{i=1}^p \sum_{j=1}^p r_{i,j}^2 \right)^{1/2}. \quad (4.3.16)$$

Число обусловленности  $\text{cond}(R)$  может изменяться от 1 до  $+\infty$  (вырожденная матрица) и входит в неравенство

$$\frac{\|\tilde{b} - b\|}{\|b\|} \leq \text{cond}(R) \frac{\|\tilde{r} - r\|}{\|r\|}, \quad (4.3.17)$$

где  $\tilde{b}$  – решение системы  $R\tilde{b} = \tilde{r}$  при неточно заданной правой части  $\tilde{r}$ . Следовательно, чем больше число обусловленности, тем больше относительная ошибка получаемого решения при неточно заданной правой части системы уравнений.

**Пример 4.3.1.** Предположим, что матрица  $R$  имеет число обусловленности, равное 100, а относительная погрешность вектора правой части  $r$  равна  $\frac{\|\tilde{r} - r\|}{\|r\|} = 0.01$  (или 1%). Тогда имеет место неравенство для относительной ошибки решения  $\tilde{b}$ :

$$\frac{\|\tilde{b} - b\|}{\|b\|} \leq 100 \cdot 0.01 = 1 \text{ (или 100 \%)} . \quad (4.3.18)$$

Очевидно, что такая точность оценивания коэффициентов  $\beta_m$  является неудовлетворительной. В этом случае для повышения точности решения необходимо использовать специальные методы решения системы (4.3.13) при неточно заданном векторе правой части – методы регуляризации [13]. ●

Рассмотрим частные случаи уравнения (4.3.3) при  $p = 1, 2$ .

Оценка  $b_1$  для коэффициента  $\beta_1$  модели авторегрессии первого порядка определяется как

$$b_1 = r(1), \quad (4.3.19)$$

а само уравнения авторегрессии временного ряда имеет вид

$$\hat{y}_i = b_1 y_{i-1}. \quad (4.3.20)$$

Оценки  $b_1, b_2$  для коэффициента  $\beta_1, \beta_2$  модели авторегрессии второго порядка определяются соотношениями:

$$b_1 = \frac{r(1)[1 - r(2)]}{1 - r^2(1)}; \quad (4.3.21)$$

$$b_2 = \frac{r(2) - r^2(1)}{1 - r^2(1)}, \quad (4.3.22)$$

а уравнение авторегрессии имеет вид

$$\hat{y}_i = b_1 y_{i-1} + b_2 y_{i-2}. \quad (4.3.23)$$

Заметим, что при небольшом порядке авторегрессионной модели ( $p \leq 3$ ) и достаточно большом объеме выборки  $n$  оценки  $b_m$ , получаемые из решения системы (1.1.10), являются достаточно «хорошими» (т.е. имеют приемлемую точность).

Поэтому возникает вопрос: можно ли изложенный метод оценивания коэффициентов авторегрессионной модели использовать в случаях, когда не выполняются условия (4.3.1) и  $M(Y(\tau))$  есть некоторая функция  $t(\tau)$ ? Ответ положителен, если авторегрессионную модель вида (4.3.3) строить для функции

$$Z(\tau) = Y(\tau) - t(\tau). \quad (4.3.24)$$

Это иллюстрируется следующим примером

**Пример 4.3.2.** В табл. 4.1 представлены данные, отражающие динамику курса акций некоторой компании (в условных единицах).

Необходимо по этим данным построить модель AR(2) с коэффициентами  $b_1, b_2$ , определяемыми выражениями (4.3.21), (4.3.22).

*Решение.* По графику значений  $y_i$  (рис. 4.1, кривая 1) видно, что условие (4.3.1) не выполняется. Поэтому *отцентрируем* значения  $y_i$  следующим способом:

- по значениям  $y_i$  построим уравнения тренда  $\hat{t}(\tau)$  в виде полинома степени  $p = 2$ . В табл. 4.2 приведено уравнение

$$\hat{t}(\tau) = 943.83 - 1.74\tau + 1.06\tau^2, \quad (4.3.25)$$

для которого  $R^2 = 0.406$ ,  $\widehat{R}^2 = 0.373$ ;

- вычислим значение функции  $\widehat{z}(\tau)$  (которая является оценкой для функции (4.3.24)) при  $\tau = \tau_i$

$$\widehat{z}_i = y_i - \widehat{t}(\tau_i), \quad i=1,2,\dots,n. \quad (4.3.26)$$

График значений  $\widehat{z}_i$  приведен на рис. 4.2. Используя критерий  $T_e$  (см. п. 2.4.1) проверим гипотезу о равенстве математического ожидания случайных величин  $\widehat{Z}(\tau_i) = Y(\tau_i) - \widehat{t}(\tau_i)$  нулю. Вычисленное значение критерия  $T_e = 0.86$  не попадает в критическую область  $[2.38, \infty)$  и поэтому принимается гипотеза  $M(\widehat{Z}(\tau_i)) = 0, i=1,2,\dots,n$ .

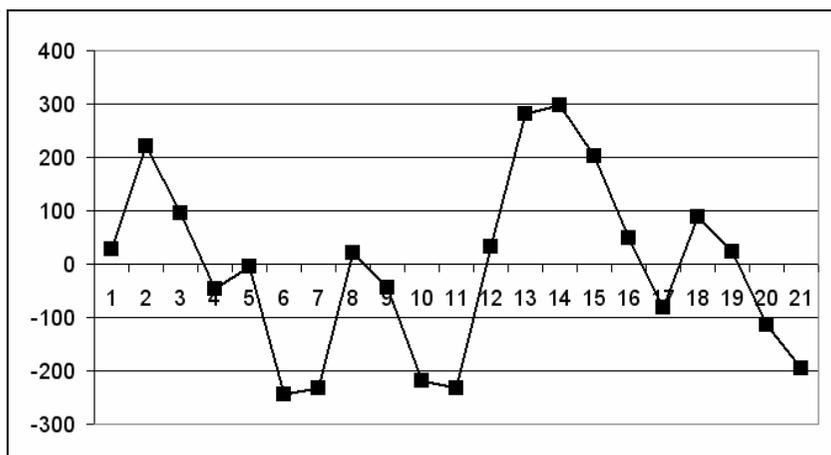


Рис. 4.2. График значений  $\widehat{z}_i$  (пример 4.3.2)

Далее используя функцию Excel КОРРЕЛ (см. пример 1.2.2) вычисляем выборочные коэффициенты автокорреляции:  $r(1) = 0.456$ ,  $r(2) = -0.004$ . Подставляя эти значения в формулы (4.3.21), (4.3.22), определяем  $b_1 = 0.783$ ,  $b_2 = -0.443$ . Таким обра-

зом, временной ряд  $Z(\tau_i)$  описывается уравнением авторегрессии

$$\widehat{z}_i = 0.783 z_{i-1} - 0.443 z_{i-2}. \quad (4.3.27)$$

Учитывая приведенное центрирование (4.3.26) исходного временного ряда, получаем следующую *комбинированную* модель временного ряда  $Y(\tau_i)$ :

$$\widehat{y}_i = 943.83 - 1.74 \tau_i + 1.06 \tau_i^2 + 0.783 \widehat{z}_{i-1} - 0.443 \widehat{z}_{i-2}, \quad (4.3.28)$$

где  $\widehat{z}_i = y_i - \widehat{t}(\tau_i)$ ,  $i=1,2,\dots,n$ . На рис. 4.3 приведены значения  $y_i$  (кривая 1) и значения  $\widehat{y}_i$  (кривая 2) для  $i=3,4,\dots,21$ .

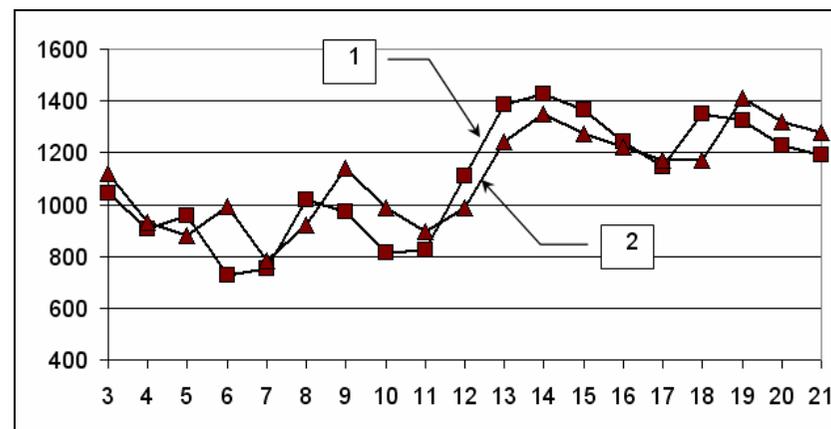


Рис. 4.3. График значений  $y_i, \widehat{y}_i$  (пример 4.3.2)

Видно, что значения  $\widehat{y}_i$  адекватные исходным  $y_i$  в большей степени, чем значения  $\widehat{y}_i$ , вычисленные по формуле (табл. 4.4):

$$\widehat{y}_i = 275.61 + 0.730 y_{i-1} + 0.022 y_{i-2}. \quad (4.3.29)$$

Это подтверждается и значениями коэффициентов детерминации, приведенными в табл. 4.4.

Таблица 4.4

Уравнение	$R^2$	$\widehat{R}^2$
Формула (4.3.29)	0.554	0.525
Формула (4.3.28)	0.713	0.695

Следовательно, можно сделать вывод, что комбинированная модель (4.3.28) более точно описывает временной ряд (выборка которого приведена в табл. 4.1) по сравнению с *чистой* моделью авторегрессии (4.3.29). ●

**Замечание 4.3.1.** Формулу (4.3.28) можно записать в другом виде, более близком к записи авторегрессионной модели временного ряда (4.1.1). Для этого в формулу (4.3.28) вместо  $\widehat{z}_i$  подставим выражение  $y_i - \widehat{t}(\tau_i)$ . Получаем следующую запись авторегрессионной модели:

$$\begin{aligned} \widehat{y}_i &= 0.783y_{i-1} - 0.443y_{i-2} + t(\tau_i) + 0.783t(\tau_{i-1}) - 0.443t(\tau_{i-2}) = \\ &= b_0(\tau_i, \tau_{i-1}, \tau_{i-2}) + b_1y_{i-1} + b_2y_{i-2}, \end{aligned}$$

где коэффициент  $b_0$  зависит от моментов времени  $\tau_i, \tau_{i-1}, \tau_{i-2}$ . Зависимость  $b_0$  от времени позволяет в определенной степени учесть нестационарность исследуемого временного ряда.

#### 4.4. Тест на наличие автокорреляции

Рассмотренная статистика Дарбина–Уотсона является наиболее важным индикатором автокорреляции (см. п. 2.4.3). Однако тест с использованием этой статистики имеет следующие недостатки:

- наличие зоны *неопределенности* значений статистики  $d$  (когда нельзя принять определенное решение);
- ограниченность применения, так как выявляется корреляция только между соседними членами ряда.

Все это приводит к необходимости использовать и другие тесты на наличие автокорреляции:  $Q$ -тест Льюинга–Бокса, который применим к стационарным временным рядам.

**Тест Льюинга–Бокса.** Сформулируем статистические гипотезы:

$H_0$ : автокорреляция отсутствует;

$H_1$ : автокорреляция присутствует.

Тест основан на следующем рассуждении: при отсутствии автокорреляции все значения коэффициентов автокорреляции  $\rho(l)$  равны 0. Разумеется, значения выборочных коэффициентов автокорреляции  $r(l)$  отличны от нуля, но это отличие не должно быть существенным.

Критерий теста Льюинга–Бокса определяется выражением

$$Q_p = n(n+2) \sum_{l=1}^p \frac{r^2(l)}{n-l}, \quad (4.4.1)$$

где  $n$  – число наблюдений временного ряда,  $p$  – предполагаемый порядок авторегрессионной модели. Выборочные коэффициенты автокорреляции вычисляются по формуле (1.2.5).

Если верна гипотеза  $H_0$  о равенстве нулю всех коэффициентов  $\rho(l)$ ,  $l = 1, 2, \dots, p$ , то критерий  $Q_p$  имеет  $\chi_p^2$ -распределение с  $p$  степенями свободы. Если выполняется неравенство

$$Q_p > \chi_{p, 1-\alpha}^2, \quad (4.4.2)$$

то с вероятностью ошибки первого рода, равной  $\alpha$ , гипотеза  $H_0$  отвергается и принимается альтернативная гипотеза  $H_1$ , т.е. принимается предположение о наличии автокорреляции в исследуемом временном ряду.

Величина  $\chi_{p,1-\alpha}^2$  является квантилем  $\chi_p^2$ -распределения порядка  $(1-\alpha)$ . Для вычисления можно использовать функцию Excel:

$$\chi_{p,1-\alpha}^2 = \text{ХИ2ОБР}(\alpha; p). \quad (4.4.3)$$

**Пример 4.4.1.** В табл. 4.5 приведены значения коэффициентов автокорреляции  $r(l)$ , вычисленные по 100 наблюдениям стационарного временного ряда.

Таблица 4.5

$l$	$r(l)$	$p$	Значение $Q_p$	$\chi_{p,0.95}^2$
1	0.498	1	25.314	3.84
2	0.158	2	27.875	5.99
3	0.024	3	27.937	7.82

Необходимо проверить гипотезы о наличии автокорреляции для  $p = 1, 2, 3$ .

*Решение.* В табл. 4.5 приведены вычисленные значения критерия  $Q_p$ ,  $p = 1, 2, 3$  и квантили  $\chi_{p,0.95}^2$ . Видно, что для всех  $p$  выполняется неравенство (4.4.2) и, следовательно, гипотезы  $H_0$  для  $p = 1, 2, 3$  отвергаются с вероятностью ошибки первого рода, равной 0.05. ●

Рассмотренный тест Льюинга–Бокса следует использовать для стационарного временного ряда или ряда остатков  $e_i = y_i - \hat{t}(\tau_i)$ , полученных после нахождения трендовой составляющей  $\hat{t}(\tau)$ .

#### 4.5. Определение порядка авторегрессионной модели временного ряда

Ранее (при вычислении оценок для коэффициентов  $\beta_m$ ) предполагалось, что порядок  $p$  авторегрессионной модели задан. Однако в большинстве практических случаев это не так. Поэтому возникает задача оценить порядок авторегрессионной модели.

Предположим, что исследуемый временной ряд является стационарным и обратимся к соотношению (3.3.9). Видно, что

$$\beta_m = \rho_{\text{част}}(m), \quad (4.5.1)$$

где  $\rho_{\text{част}}(m)$  – частный коэффициент автокорреляции, соответствующий лагу  $m$ . Величина  $\rho_{\text{част}}(m)$  определяется как коэффициент корреляции между двумя случайными величинами  $Y(\tau_i)$  и  $Y(\tau_{i+m})$  при устранении влияния промежуточных  $Y(\tau_{i+1}), \dots, Y(\tau_{i+m-1})$  случайных величин.

Оценкой для  $\rho_{\text{част}}(m)$  является выборочный частный коэффициент автокорреляции  $r_{\text{част}}(m)$  и тогда оценка  $b_m$  для коэффициента  $\beta_m$  определяется выражением

$$b_m = r_{\text{част}}(m). \quad (4.5.2)$$

Для построения формул, вычисляющих  $r_{\text{част}}(m)$ , приведем некоторые необходимые сведения о частных коэффициентах корреляции, определяемых при наличии зависимой переменной  $Y$  и  $p$  независимых объясняющих переменных  $X_1, X_2, \dots, X_p$  [5, с. 128]. Запись  $r_{YX_i(X_1, X_2, \dots, X_p)}$  означает коэффициент частной корреляции между переменными  $Y$  и  $X_i$  при «исключении» переменных  $X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_p$ . Количество исключаемых переменных (указаны в круглых скобках) определяет порядок частного коэффициента корреляции. Соответственно коэффициенты парной корреляции можно рассматривать как частные коэффициенты корреляции нулевого порядка.

Коэффициенты частной корреляции более высоких порядков определяют по следующей рекуррентной формуле:

$$r_{YX_i(X_1, \dots, X_p)} = \frac{r_{YX_i(X_1, \dots, X_{p-1})} - r_{YX_p(X_1, \dots, X_{p-1})} \cdot r_{X_i X_p(X_1, \dots, X_{p-1})}}{\sqrt{(1 - r_{YX_p(X_1, \dots, X_{p-1})}^2)(1 - r_{X_i X_p(X_1, \dots, X_{p-1})}^2)}}. \quad (4.5.3)$$

При двух независимых переменных  $X_1, X_2$  и  $i=1$  формула примет вид

$$r_{YX_1(X_2)} = \frac{r_{YX_1} - r_{YX_2} \cdot r_{X_1 X_2}}{\sqrt{(1 - r_{YX_2}^2)(1 - r_{X_1 X_2}^2)}}. \quad (4.5.4)$$

Для  $i=2$  имеем

$$r_{YX_2(X_1)} = \frac{r_{YX_2} - r_{YX_1} \cdot r_{X_1 X_2}}{\sqrt{(1 - r_{YX_1}^2)(1 - r_{X_1 X_2}^2)}}. \quad (4.5.5)$$

Для трех переменных  $X_1, X_2, X_3$  и  $i=3$  частный коэффициент корреляции второго порядка определяется выражением:

$$r_{YX_3(X_1, X_2)} = \frac{r_{YX_3(X_1)} - r_{YX_2(X_1)} \cdot r_{X_2 X_3(X_1)}}{\sqrt{(1 - r_{YX_2(X_1)}^2)(1 - r_{X_2 X_3(X_1)}^2)}}. \quad (4.5.6)$$

Для вычисления частных коэффициентов автокорреляции с использованием формулы (4.5.3) установим следующее соотношение

$$Y \leftrightarrow Y(\tau_i); X_1 \leftrightarrow Y(\tau_{i+1}); X_2 \leftrightarrow Y(\tau_{i+2}); X_3 \leftrightarrow Y(\tau_{i+3}). \quad (4.5.7)$$

Тогда с учетом формулы (4.5.5) для частного коэффициента автокорреляции первого порядка между  $Y(\tau_i)$ ,  $Y(\tau_{i+2})$  при устранении влияния  $Y(\tau_{i+1})$  получаем следующую формулу:

$$r_{част}(2) = r_{YX_2(X_1)} = \frac{r(2) - r(1) \cdot r(1,2)}{\sqrt{(1 - r^2(1))(1 - r^2(1,2))}}, \quad (4.5.8)$$

где  $r(1)$ ,  $r(2)$ ,  $r(1,2)$  – выборочные коэффициенты автокорреляции между  $Y(\tau_i)$  и  $Y(\tau_{i+1})$ ,  $Y(\tau_i)$  и  $Y(\tau_{i+2})$ ,  $Y(\tau_{i+1})$  и  $Y(\tau_{i+2})$ .

Из соответствий (4.5.7) следует очевидное тождество

$$r_{част}(3) = r_{YX_3(X_1, X_2)}. \quad (4.5.9)$$

Тогда, используя формулу (4.5.6), можно получить выражение для вычисления выборочного частного коэффициента автокорреляции второго порядка между  $Y(\tau_i)$  и  $Y(\tau_{i+3})$  при устранении влияния  $Y(\tau_{i+1})$ ,  $Y(\tau_{i+2})$ . При этом частные корреляции первого порядка  $r_{YX_3(X_1)}$ ,  $r_{YX_2(X_1)}$ ,  $r_{X_2 X_3(X_1)}$  определяются соотношениями, аналогичными (4.5.4), (4.5.5):

$$r_{YX_3(X_1)} = \frac{r(3) - r(1) \cdot r(1,3)}{\sqrt{(1 - r^2(1))(1 - r^2(1,3))}}; \quad (4.5.10)$$

$$r_{YX_2(X_1)} = r_{част}(2); \quad (4.5.11)$$

$$r_{X_2 X_3(X_1)} = \frac{r(2,3) - r(1,3) \cdot r(1,2)}{\sqrt{(1 - r^2(1,3))(1 - r^2(1,2))}}, \quad (4.5.12)$$

где  $r(2,3)$ ,  $r(1,3)$  – выборочные коэффициенты автокорреляции между величинами  $Y(\tau_{i+2})$  и  $Y(\tau_{i+3})$ ,  $Y(\tau_{i+1})$  и  $Y(\tau_{i+2})$ .

**Замечание 4.5.1.** Вычисление выборочного частного коэффициента автокорреляции третьего порядка  $r_{част}(4)$  требует большого объема вычислений, и некоторые специализированные статистические пакеты включают функцию, позволяющую вычислить  $r_{част}(l)$ ,  $l=1,2,\dots,p$ . ♦

Предположим, что все вычисленные частные коэффициенты автокорреляции значимы до порядка  $p$  включительно. Тогда можно принять *порядок автокорреляции модели (4.1.1), равным величине  $p$* .

Для проверки значимости коэффициентов  $r_{част}(l)$  сформулируем статистические гипотезы:

$H_0: \rho_{част}(l) = 0$  (т.е. коэффициент  $r_{част}(l)$  не значим);

$H_1: \rho_{част}(l) \neq 0$  (т.е. коэффициент  $r_{част}(l)$  значим).

Для проверки этих гипотез введем критерии

$$T_r(l) = \frac{r_{част}(l)}{\sqrt{1-r_{част}^2(l)}} \cdot \sqrt{n-l-1} \quad (4.5.13)$$

Если выполняется неравенство

$$|T_r(l)| > t(1-\alpha, n-l-1), \quad (4.5.14)$$

то с вероятностью ошибки первого рода, равной  $\alpha$ , отвергается гипотеза  $H_0$  и принимается альтернативная  $H_1$  (т.е. коэффициент  $r_{част}(l)$  является значимым).

Для вычисления значения  $t(1-\alpha, n-l-1)$  используется функция Excel:

$$t(1-\alpha, n-l-1) = \text{СТЮДРАСПОБР}(\alpha, n-l-1). \quad (4.5.15)$$

**Пример 4.5.1.** В табл. 2.1 приведены данные, отражающие спрос (в условных единицах) на некоторый товар за восьмилетний период (см. пример 2.1.1).

По данным табл. 2.1 необходимо:

- вычислить коэффициенты автокорреляции (для лагов  $l=1, 2, 3$ );
- вычислить частный коэффициент автокорреляции первого порядка  $r_{част}(2)$ .

*Решение.* На рис. 4.4. приведен фрагмент документа Excel, в котором показано вычисление коэффициентов автокорреляции  $r(1)$ ,  $r(2)$ ,  $r(3)$  с помощью функции Excel КОРРЕЛ (см. пример 1.2.2). Получены следующие значения:

$$r(1) = 0.725, \quad r(2) = 0.842, \quad r(3) = 0.909, \quad r(1,2) = 0.825.$$

Подставляя необходимые значения в формулу (4.5.8), получаем (ячейка E9)  $r_{част}(2) = 0.627$ . ●

	A	B	C	D	E	F
1	$\tau_i$	$y_i$	=КОРРЕЛ(B2:B8;B3:B9)			
2	1	213				
3	2	171		$r(1) =$	0,725	
4	3	291		$r(2) =$	0,842	
5	4	309		$r(3) =$	0,909	
6	5	317		$r(1,2) =$	0,825	
7	6	362				
8	7	351		=КОРРЕЛ(B3:B8;B4:B9)		
9	8	361	$r_{част}(2) =$		0,627	
10						
11			=(E4-E3*E6)/КОРЕНЬ((1-E3^2)*(1-E6^2))			

Рис. 4.4. Вычисление частного коэффициента автокорреляции (пример 4.5.1)

## ЛАБОРАТОРНАЯ РАБОТА

### Вычисление коэффициентов автокорреляции

*Исходные данные.* В табл. 4.1 представлены данные, отражающие динамику курса акций некоторой компании (в условных единицах).

*Необходимо:*

1. Вычислить выборочные коэффициенты автокорреляции  $r(l)$  для лагов  $l = 1, 2, 3, 4$ .
2. Вычислить выборочный частный коэффициент автокорреляции первого порядка  $r_{\text{част}}(2)$ .
3. Проверить значимость коэффициента  $r_{\text{част}}(2)$ .
4. Сделать обоснованный вывод о порядке авторегрессионной модели для исследуемого временного ряда.

*Рекомендации.* При выполнении лабораторной работы используйте пример 4.5.1.

### КОНТРОЛЬНАЯ РАБОТА

*Исходные данные.* В таблице, представленной ниже, находятся данные, отражающие динамику курса акций некоторой компании (в условных единицах), где  $N$  – последняя цифра меняющегося номера зачетной книжки.

$i$	1	2	3	4	5	6
$y_i$	971+N	1166-N	1044	907+N	957	727+N
$i$	7	8	9	10	11	12
$y_i$	1019	972+N	815	823+N	1112+N	1386-N
$i$	13	14	15	16	17	18
$y_i$	1364-N	1241+N	1145	1351+N	1325-N	1226+N

*Необходимо:*

1. Вычислить выборочные коэффициенты автокорреляции  $r(l)$  для лагов  $l = 1, 2, 3, 4$ .
2. Вычислить выборочный частный коэффициент автокорреляции первого порядка  $r_{\text{част}}(2)$ ,  $r_{\text{част}}(3)$ .
3. Проверить значимость коэффициента  $r_{\text{част}}(2)$ ,  $r_{\text{част}}(3)$ .
4. Сделать обоснованный вывод о порядке авторегрессионной модели для исследуемого временного ряда.
5. Построить авторегрессионную модель временного ряда с выбранным порядком авторегрессии.
6. Построить графики значений  $y_i, \hat{y}_i, i = 1, \dots, 18$ .
7. Вычислить величину коэффициента детерминации  $R^2$ .
8. По построенной модели осуществить прогноз значений временного ряда для  $\tau_i = 19, 20, 22$ .

*Рекомендации.* При выполнении контрольной работы используйте пример 4.5.1.

### ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Что такое лаговые переменные?
2. Какой вид имеет авторегрессионная модель второго и третьего порядков?

3. Как оцениваются коэффициенты авторегрессионной модели на основе метода наименьших квадратов?

4. Запишите матрицу  $X$  в матричном представлении

$$y = X\beta + \varepsilon$$

авторегрессионной модели второго и третьего порядков.

5. Какова сущность оценивания коэффициентов авторегрессионной модели из решения уравнений Юла–Уокера?

6. Что такое частный коэффициент автокорреляции?

7. Как определить порядок авторегрессионной модели?

8. Приведите примерные значения частных коэффициентов автокорреляции, при которых делается вывод о втором порядке авторегрессии.

## ЗАКЛЮЧЕНИЕ

Проводимые на различных этапах эконометрического моделирования при анализе временных рядов вычисления являются весьма трудоемкими, что в ряде случаев сдерживает использование эконометрических методов на практике или делает их недостаточно эффективными.

Преодолеть эти трудности можно применяя:

– универсальные статистические пакеты: Statgraphics, EViews, Statistica;

– табличный процессор Excel.

В учебном пособии выбран второй путь по следующим причинам:

– табличный процессор Excel является доступной русифицированной лицензионной программой, в то время как названные статистические пакеты труднодоступны и в основном являются контрафактными;

– использование табличного процессора Excel подразумевает программирование расчетных выражений, что способствует лучшему усвоению расчетных соотношений и методов эконометрического моделирования.

Наличие большого числа примеров, лабораторных работ, контрольных вопросов и заданий делает учебное пособие полезным не только при изучении курса «Эконометрика», но и при выполнении курсовых и дипломных работ, а также при использовании различных дистанционных форм обучения в сети Интернет.

Полезным дополнением к данному пособию является учебный материал, изложенный в пособиях [5, 12].

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. **Тимошенко Е. И.** Теория вероятностей : учеб. пособие / Е. И. Тимошенко, Ю. Е. Воскобойников ; Новосиб. гос. архитектур.-строит. ун-т. – Новосибирск : НГАСУ, 2003. – 88 с. (электр. версия: <http://www.sibstrin.ru/prikl/terver.html>).
2. **Воскобойников Ю. Е.** Математическая статистика : учеб. пособие / Ю. Е. Воскобойников, Е. И. Тимошенко ; Новосиб. гос. архитектур.-строит. ун-т. – Новосибирск : НГАСУ, 2000. (электр. версия: <http://www.sibstrin.ru/prikl/stat.html>).
3. **Воскобойников Ю. Е.** Математическая статистика (с примерами в Excel) : учеб. пособие / Ю. Е. Воскобойников, Е. И. Тимошенко ; Новосиб. гос. архитектур.-строит. ун-т. – Новосибирск : НГАСУ (Сибстрин), 2006. – 152 с. (электр. версия: [http://www.sibstrin.ru/prikl/stat\\_excel.html](http://www.sibstrin.ru/prikl/stat_excel.html)).
4. **Гмурман В. Е.** Теория вероятностей и математическая статистика : учеб. для вузов / В. Е. Гмурман. – М. : Высшая школа, 1998.
5. **Воскобойников Ю. Е.** Эконометрика в Excel. Часть 1. Парный и множественный регрессионный анализ : учеб. пособие / Ю. Е. Воскобойников ; Новосиб. гос. архитектур.-строит. ун-т. – Новосибирск : НГАСУ (Сибстрин), 2005. – 152 с.
6. **Кремер Н. Ш.** Эконометрика / Н. Ш. Кремер, Б. А. Путко. – М. : ЮНИТИ, 2002.
7. **Айвазян С. А.** Прикладная статистика и основы эконометрики / С. А. Айвазян, В. С. Мхитарян. – М. : ЮНИТИ, 1998.
8. **Минус Я. Р.** Эконометрика. Начальный курс / Я. Р. Минус, Л. К. Катышев, А. А. Пересецкий. – М. : Дело, 2000.
9. **Эконометрика** / под ред. Н. И. Елисеевой. – М. : Финансы и статистика, 2001.
10. **Арженовский С. В.** Эконометрика : учеб. пособие / С. В. Арженовский, О. Н. Федосова. – Ростов н/Д, 2002.
11. **Макарова Н. В.** Статистика в EXCEL : учеб. пособие / Н. В. Макарова, В. Я. Трофимец. – М. : Финансы и статистика, 2002.
12. **Воскобойников Ю. Е.** Эконометрика : метод. указания к лабораторным и контрольным работам / Ю. Е. Воскобойников, Т. Н. Воскобойникова. – Новосибирск : Изд-во Новосибирского филиала Санкт-Петербургской академии управления и экономики, 2006. (электр. версия: <http://www.sibstrin.ru/prikl/econometr.html>).
13. **Воскобойников Ю. Е.** Устойчивые методы и алгоритмы параметрической идентификации : монография / Ю. Е. Воскобойников ; Новосиб. гос. архитектур.-строит. ун-т. – Новосибирск : НГАСУ (Сибстрин), 2006. – 180 с. (электр. версия: <http://www.sibstrin.ru/prikl/monogr07.html>).