
В.Н.Афанасьев, М.М.Юзбашев

АНАЛИЗ ВРЕМЕННЫХ РЯДОВ И ПРОГНОЗИРОВАНИЕ

Рекомендовано
Учебно-методическим объединением
по образованию в области статистики
и антикризисного управления,
математических методов в экономике
в качестве учебника для студентов
высших учебных заведений,
обучающихся по специальности 080601 "Статистика"
и другим экономическим специальностям



**МОСКВА
"ФИНАНСЫ И СТАТИСТИКА"
2010**

УДК 311.13 (075.8)
ББК 60.603а73
А 94

РЕЦЕНЗЕНТЫ

Кафедра социально-экономической статистики
Московского государственного университета экономики,
статистики и информатики
(заведующий кафедрой – доктор экономических наук,
профессор М.В. Карманов);
Ю.П. Лукашин,
доктор экономических наук, профессор

Афанасьев В.Н.

А94 Анализ временных рядов и прогнозирование: учебник
/В.Н. Афанасьев, М.М. Юзбашев. Изд. 2-е, перераб. и доп. –
М.: Финансы и статистика; ИНФРА-М, 2010. – 320 с.: ил.
ISBN 978-5-279-03400-0 (Финансы и статистика)

Рассматриваются показатели временного ряда, основные типы тен-
денций и методы их распознавания, методы оценки параметров колебле-
мости, измерение устойчивости уровней ряда и тенденции динамики,
моделирование и прогнозирование временных рядов. Рассчитан на лиц,
имеющих знания по общей теории статистики.

Для студентов, обучающихся по направлению и специальности
«Статистика», а также для преподавателей и аспирантов экономических
вузов.

Ф $\frac{0702000000 - 083}{010(01) - 2009}$ без объявл.

УДК 311.13 (075.8)
ББК 60.603а73

ISBN 978-5-279-03400-0

© Афанасьев В.Н.,
Юзбашев М.М., 2010
© Издательство «Финансы
и статистика», 2010

ПРЕДИСЛОВИЕ

В учебниках по общей теории статистики теме «Анализ временных рядов и прогнозирование» посвящается, как правило, одна глава «Анализ динамики» («Динамические ряды»), что явно недостаточно для глубокого изложения непрерывно развивающихся методов статистического изучения разнообразных массовых процессов. Цель предлагаемого учебника, полностью посвященного данной теме, – помочь студентам, изучающим статистику, получить более широкое и глубокое представление о всех методах статистического исследования по изменениям социально-экономических и других массовых явлений во времени. В нем подробно излагаются методы распознавания типа тренда, алгоритма расчета логарифмического, логистического, гиперболического трендов; описываются различные модели сезонных колебаний; рассматриваются моделирование функционально и корреляционно связанных систем показателей, а также система показателей устойчивости уровней и тенденций изменения, обобщенные модели эконометрического прогнозирования.

Учебник рассчитан на лиц, имеющих подготовку по общей теории статистики, хотя авторы стремились сделать изложение максимально понятным. Данное издание скорректировано с учетом опыта прочтения курса по одноименному учебнику тех же авторов¹, оно состоит из двенадцати глав, в которых раскрыты теоретические и прикладные аспекты современного статистического анализа временных рядов.

Ограниченный объем учебника не позволяет углубленно рассматривать те или иные проблемы статистики, в этих случаях авторы предлагают обратиться к работам, указанным в списке рекомендуемой литературы. Иногда авторы отсылают читателей к последующим главам, что методически не совсем правильно, но в целях экономии места и имея в виду хорошую подготовленность читателя считают оправданным.

¹ Афанасьев В.Н., Лебедева Т.В. Анализ временных рядов и прогнозирование. Учебник, изд. 1-е.- М.: Финансы и статистика, 2001.

Предлагаемый учебник, дополненный двумя учебными пособиями – «Моделирование и прогнозирование временных рядов»¹ и «Статистические методы прогнозирования в экономике»², в которых даются практические задания, тесты и вопросы по анализу временных рядов, позволит как овладеть теоретическим материалом, так и приобрести практические навыки решения задач по анализу и прогнозированию временных рядов.

Авторы выражают глубокую признательность профессорам М.В. Карманову и Ю.П. Лукашину за ценные замечания, сделанные ими при рецензировании рукописи настоящего учебника.

Большую помощь в подготовке учебника оказала кандидат экономических наук, доцент кафедры статистики и эконометрики Оренбургского государственного университета (ОГУ) Т.В. Лебедева, за что авторы выражают ей признательность. В настоящее время данный курс в ОГУ читается ею.

Авторы будут благодарны всем, кто пожелает высказать свои предложения по улучшению данного учебника. Предложения следует присылать по адресу: Afanassiev@tambler.ru

¹ *Афанасьев В.Н., Лебедева Т.В.* Моделирование и прогнозирование временных рядов: учеб. пособие. – М.: Финансы и статистика, 2009.

² *Афанасьев В.Н., Лебедева Т.В.* Статистические методы прогнозирования в экономике: учеб. пособие. – М.: Финансы и статистика, 2009.

Оглавление

Предисловие.....	8
Глава 1 Виды и построение временных рядов.....	10
1.1 Что такое «временной ряд».....	10
1.2 Классификация временных рядов.....	11
1.3 Обеспечение сопоставимости уровней временных рядов.....	16
Глава 2 Составляющие элементы временного ряда.....	20
2.1 Понятие об основной тенденции и колеблемости временных рядов.....	20
2.2 Иерархия тенденций и колебаний.....	21
2.3 Периодизация динамики.....	24
Глава 3 Показатели временного ряда и методы их исчисления...	45
3.1 Показатели, характеризующие тенденцию динамики....	45
3.2 Особенности показателей для рядов, состоящих из относительных уровней.....	52
3.3 Средние показатели временных рядов.....	57
Глава 4 Основные типы тенденций и уравнений тренда.....	68
4.1 Прямолинейный тренд и его свойства.....	68
4.2 Параболический тренд и его свойства.....	72
4.3 Экспоненциальный тренд и его свойства.....	78
4.4 Гиперболический тренд и его свойства.....	83
4.5 Логарифмический тренд и его свойства.....	86
4.6 Логистический тренд и его свойства.....	88
Глава 5 Методы распознавания типа тренда и оценки его параметров.....	92

5.1	Применение графического изображения для распознавания типа тенденции.....	92
5.2	Методика проверки статистических гипотез о типе тренда.....	95
5.3	Оценка параметров линейного, параболического и гиперболического трендов.....	104
5.3.1	Уравнение прямой линии тренда.....	105
5.3.2	Уравнение параболического (II порядка) тренда.....	109
5.3.3	Гиперболическое уравнение тренда.....	118
5.4	Оценка параметров логистического, логарифмического и экспоненциального уравнений тренда.....	120
5.4.1	Экспоненциальное уравнение тренда.....	121
5.4.2	Логарифмическое уравнение тренда.....	128
5.4.3	Логистическое уравнение тренда.....	131
5.5	Многokратное скользящее выравнивание.....	136
Глава 6	Методы распознавания типа колебаний и оценки параметров колеблемости.....	144
6.1	Графическое отображение и основные свойства разных типов колебаний.....	145
6.1.1	Пилообразная колеблемость.....	146
6.1.2	Долгопериодическая циклическая колеблемость.....	149
6.1.3	Случайно распределенная во времени колеблемость.....	152
6.2	Измерение показателей силы и интенсивности колебаний.....	154
6.2.1	Показатели абсолютной величины (силы) колебаний....	155
6.2.2	Показатели относительной интенсивности колебаний...	157
6.3	Особенности измерения сезонных колебаний.....	159

6.3.1	Статистические методы оценки уровня сезонности. Прогнозирование с помощью тренд – сезонных моделей.....	160
6.3.2	Исследование периодических колебаний методами спектрального анализа.....	175
6.3.3	Моделирование сезонных колебаний с помощью фиктивных переменных.....	183
6.4	Измерение тренда колеблемости.....	190
6.5	Автокорреляция отклонений от тренда.....	197
Глава 7	Вероятностная оценка существенности (надежности установления) параметров тренда и колеблемости.....	199
7.1	Оценка надежности параметров тренда.....	200
7.2	Доверительные границы тренда.....	211
7.3	Вероятностная оценка показателей колеблемости.....	214
Глава 8	Методы изучения и измерения устойчивости уровней ряда и тренда.....	225
8.1	Методы измерения устойчивости уровней ряда.....	226
8.2	Методы измерения устойчивости тенденции динамики.....	231
8.3	Комплексные показатели (критерии) устойчивости.....	234
Глава 9	Изучение динамики комплекса взаимосвязанных признаков.....	240
9.1	Динамика жестко связанной системы признаков (показателей).....	240
9.2	Агрегирование трендов и колебаний по совокупности объектов.....	253
9.2.1	Тренды объемных признаков.....	253
9.2.2	Тренды качественных признаков.....	256

9.2.3	Агрегирование показателей колеблемости.....	258
9.3	Корреляция по временным рядам: сущность, ограничения.....	262
9.4	Методы измерения корреляции между колебаниями признаков.....	266
9.5	Корреляция с учетом лага и циклов.....	274
9.6	Понятие о динамике комплекса статистически взаимосвязанных признаков.....	280
Глава 10	Моделирование и прогнозирование временных рядов (динамики).....	291
10.1	Сущность и условия прогноза по тренду с учетом колеблемости.....	292
10.2	Простая трендовая модель и прогноз по ней.....	297
10.3	Прогноз с учетом случайной колеблемости.....	299
10.3.1	Прогноз доверительного интервала для линии тренда...	299
10.3.2	Прогноз доверительного интервала для уровня отдельного периода (момента).....	303
10.3.3	Прогноз доверительного интервала для среднего уровня за ряд периодов.....	307
10.3.4	Расчет страхового запаса для преодоления риска отклонения от тренда.....	310
10.4	Прогнозирование комплекса жестко-взаимосвязанных признаков.....	311
10.5	Прогнозирование по смешанной трендово-факторной модели.....	314
Глава 11	Использование адаптивных методов прогнозирования в экономических исследованиях.....	316

11.1	Сущность адаптивных методов.....	316
11.2	Экспоненциальное сглаживание.....	318
11.3	Адаптивные полиномиальные модели.....	325
Глава 12	Прогнозирование с помощью ARMA и ARIMA – процессов.....	337
12.1	Стационарные временные ряды и их основные характеристики.....	337
12.2	Модели стационарных временных рядов.....	340
12.2.1	Авторегрессионные модели.....	341
12.2.2	Модели скользящего среднего.....	363
12.2.3	Авторегрессионные модели со скользящими средними в остатках.....	370
12.2.4	Прогнозирование ARMA – процессов.....	372
12.3	Методология Бокса – Дженкинса.....	378
	Литература.....	392
	Приложения.....	394

Глава 1. ВИДЫ И ПОСТРОЕНИЕ ВРЕМЕННЫХ РЯДОВ

1.1. Что такое «временной ряд»

Термин *временные ряды* в статистике России пока непривычен. В учебниках по общей теории статистики преобладают термины *ряды динамики, динамические ряды, статистическое изучение динамики*. В зарубежной англоязычной литературе принят термин *time series*, в немецкой – *zeitreihen analyse*. Оба термина ближе всего передаются по-русски как *временные ряды* или *анализ временных рядов*.

Одной из причин, препятствовавших принятию отечественной статистикой данного термина, служит особенность русского языка – сближение по звучанию и написанию совершенно разных по смыслу слова *временной*, т.е. относящийся ко времени, связанный со временем, происходящий во времени, и слова *временный*, т.е. непостоянный, преходящий, малосущественный. В европейских языках это разные слова: в немецком, например, временный – *provisorisch*, во французском – *provisoire*, в английском – *provisional*, т.е. эти слова происходят не от корня «время».

Опасение, что студенты (учащиеся) воспримут термин *временной ряд* как *временный*, заставило предпочесть новый для статистики и неточный по существу термин *динамический ряд, ряд динамики*.

Неточность последнего термина состоит в том, что не каждый ряд уровней за последовательные моменты или периоды времени содержит на самом деле (отражает) динамику какого-либо признака. Термин *динамика* правильнее относить к изменениям, направленному развитию, наличию тенденции рассматриваемых во времени показа-

телей. Про ряд уровней, содержащих лишь колебания, но не имеющих надежно установленной тенденции, говорят: «В этом ряду, в данном процессе нет никакой динамики». Так можно характеризовать экономику застойного периода, население страны или региона, находящиеся в стационарном состоянии, любую общественную или механическую систему, находящуюся в статическом состоянии. Следовательно, *динамические ряды* – понятие, относящееся к тем рядам уровней, в которых содержится тенденция изменения, а *временные ряды* – более общее понятие, включающее как динамические, так и статические последовательности уровней какого-либо показателя.

Данное существенное уточнение, а также стремление к сближению терминологии нашего учебника с принятой в зарубежных странах побудили нас принять в заглавии учебника и в тексте именно термин *временные ряды* и пренебречь «опасностью» ошибочного отождествления слов *временной* и *временный* читателями.

Итак, *временной ряд* – это последовательность упорядоченных во времени числовых показателей, характеризующих уровень состояния и изменения изучаемого явления.

1.2. Классификация временных рядов

Всякий временной ряд включает два обязательных элемента: во-первых, время и, во-вторых, конкретное значение показателя, или уровень ряда. Временные ряды различаются по следующим признакам:

1) по времени – *моментные и интервальные*. *Интервальный ряд* (табл. 1.1) – последовательность, в которой уровень явления относят к результату, накопленному или вновь произведенному за определенный интервал времени. Таковы, например, ряды показателей объема

продукции предприятия по месяцам года, количества отработанных человеко-дней по отдельным периодам (месяцам, кварталам, полугодиям, годам, пятилетиям и т.п.) и т.д. Если же уровень ряда характеризует изучаемое явление в конкретный момент времени, то совокупность уровней образует *моментный ряд*. Примерами моментных рядов могут быть последовательность показателей численности населения на начало года, поголовье скота в фермерских хозяйствах на 1 декабря или 1 июня за несколько лет, величина запаса какого-либо материала на начало периода и т.д. Важное отличие моментных рядов от интервальных состоит в том, что сумма уровней интервального ряда дает вполне реальный показатель – общий выпуск продукции за год (пятилетие, десятилетие), общие затраты рабочего времени, общий объем продаж акций и т.д., сумма же уровней моментного ряда иногда и подсчитывается, но реального содержания, как правило, не имеет;

Таблица 1.1

Запасы монетарного золота органов денежно-кредитного регулирования (на конец года; миллионов тройских унций)

Страна	1995г.	2002г.	2003г.	2004г.	2005г.	2006г.
Россия	9,41	12,46	12,54	12,44	12,4	12,9
Австралия	7,9	2,56	2,56	2,56	2,57	2,57
Австрия	11,99	10,21	10,21	9,89	9,73	9,28
Аргентина	4,37	0,01	0,01	1,77	1,76	1,76
Бельгия	20,54	8,29	8,29	8,29	7,32	7,32

Источник. Российский статистический ежегодник. 2007: Стат. сб./ Росстат.- М., 2007.-826с

2) по форме представления уровней – ряды *абсолютных* (см.

табл. 1.1), *относительных* (табл. 1.2) и *средних величин* (табл. 1.3);

Таблица 1.2

Индексы цен в секторах экономики
(декабрь к декабрю предыдущего года; в %)

Годы	2000	2001	2002	2003	2004	2005	2006
Индекс потребительских цен	120,2	118,6	115,1	112	111,7	110,9	109
Индекс цен производителей промышленных товаров	131,9	108,3	117,7	112,5	128,5	113,4	110,4
Индекс цен производителей сельскохозяйственной продукции	122,2	117,5	98,1	124,7	117,7	103	110,4
Сводный индекс цен строительной продукции	135,9	114,4	112,6	110,3	114,9	112,1	112,4
Индекс тарифов на грузовые перевозки	151,5	138,6	118,3	123,5	109,3	116,6	115,8

Источник. Российский статистический ежегодник. 2007: Стат. сб./

Росстат.- М., 2007.-826с

Таблица 1.3

**Основные показатели уровня жизни населения
в Оренбургской области, руб, (до 1998г-тыс.руб.)**

Годы	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Среднедушевые денежные доходы населения (в месяц)	674	641	1084	1451	1927	2482	3135	3873	4985	6139
Среднемесячная номинальная начисленная заработная плата работающих в экономике	788	855	1249	1849	2460	3142	3898	4735	6164	7753
Средний размер назначенных месячных пенсий	330	391	439	675	978	1307	1526	1767	2181	2510

Источник. Оренбургская область, 2007: статистический ежегодник / Территориальный орган федеральной службы гос. статистики по Оренбургской области. – Оренбург, 2007. – 428с.

3) по расстоянию между датами или интервалами времени выделяют полные и неполные временные ряды. *Полные ряды* имеют место, когда даты регистрации или окончания периодов следуют друг за другом с равными интервалами (см. табл. 1.2; табл. 1.3), *неполные* – когда принцип равных интервалов не соблюдается (см. табл. 1.1);

4) по содержанию показателей – ряды частных и агрегированных показателей. *Частные показатели* характеризуют изучаемое явление односторонне, изолированно. Например, среднесуточный объем выпуска промышленной продукции дает возможность оценить динамику промышленного производства, численность граждан, состоящих на учете в службе занятости; показывает эффективность социальной политики государства; остатки наличных денег у населения и вклады населения в банках отражают платежеспособность населения и т.д.

Агрегированные показатели (см. табл. 1.4) основаны на частных показателях и характеризуют изучаемый процесс комплексно. В таблице 1.4 приведена динамика индекса развития человеческого потенциала, предлагающегося специалистами в качестве агрегатного показателя уровня жизни. При его расчете используются следующие частные показатели:

- ожидаемая продолжительность жизни при рождении, лет;
- уровень грамотности взрослого населения, %;
- валовой показатель поступивших в учебные заведения, %;
- ВВП на душу населения по ППС, долл. США.

К сводным показателям уровня жизни можно отнести также индекс социального неблагополучия¹ (таблица 1.4.). Его расчет основан на шести показателях:

- число умерших, чел.;
- численность зарегистрированных безработных, чел.;
- число преступлений;
- число разводов;
- численность родившихся, тыс. чел.;

¹ **Афанасьев, В.Н.** Оценка эффективности реализации экономической политики государства / В.Н. Афанасьев, Е.В. Петрова, А.Б. Савельев; под ред. В.Н. Афанасьева. – М.: Финансы и статистика, 2005. – 236с.

– индекс покупательной способности.

Таблица 1.4

**Динамика интегральных показателей
оценки благосостояния населения**

Годы	1999	2000	2001	2002	2003
Индекс развития человеческого потенциала					
по Оренбургской области	0,765	0,753	0,753	0,752	0,752
по Российской Федерации	0,771	0,770	0,761	0,759	0,753
Индекс социального неблагополучия					
по Оренбургской области	32,83	32,53	33,30	33,62	33,29
по Российской Федерации	42,12	42,03	42,20	42,04	42,32

Кроме приведенных агрегатных показателей определяют агрегированный показатель экономической конъюнктуры, который дает представление о состоянии экономики в России в целом. Подобного рода показатели определяют при исследовании эффективности производства, технического уровня предприятий, качества продукции, экологического состояния. Широкое применение последних стало возможным с развитием факторного и компонентного анализа.

1.3. Обеспечение сопоставимости уровней временных рядов

Важнейшим условием правильного формирования временных рядов является сопоставимость уровней, образующих ряд. Уровни ряда, подлежащие изучению, должны быть *однородны по экономическому содержанию* и учитывать существо изучаемого явления и цель исследования.

Статистические данные, представленные в виде временных рядов, должны быть сопоставимы по территории, кругу охватываемых объектов, единицам измерения, моменту регистрации, методике расчета, ценам, достоверности.

Несопоставимость по территории возникает в результате изменений границ стран, регионов, хозяйств и т.п. Для приведения к сопоставимому виду производится пересчет прежних данных с учетом новых границ.

Полнота охвата различных частей явления – важнейшее условие сопоставимости уровней ряда. Требование одинаковой полноты охвата разных частей изучаемого объекта означает, что уровни ряда за отдельные периоды должны характеризовать размер того или иного явления по одному и тому же кругу входящих в состав его частей. Например, при характеристике динамики объемов выдачи кредитов коммерческими банками региона по годам нельзя в одни годы учитывать в качестве заемщиков только юридических лиц, а в другие – физических и юридических лиц.

При определении сравниваемых уровней ряда необходимо использовать *единую методику* их расчета. Особенно часто эта проблема возникает при международных сопоставлениях. Например, в некоторых странах (например, в США) национальный доход до последнего времени исчислялся по факторной стоимости, т.е. в рыночных ценах за вычетом косвенных налогов, но с добавлением субсидий на продукты. Такая практика представляла известное отклонение от международного стандарта в области СНС, который рекомендует оценивать все показатели в рыночных ценах.²

² Экономическая статистика: Учебник. – 3-е изд., перераб. и доп. / Под ред. проф. Ю.Н. Иванова. – М.: ИНФРА-М, 2007. – 736 с. ISBN 5-16-002083-7

Несопоставимость показателей, возникающая в силу неодинаковости применяемых *единиц измерения*, сама по себе очевидна. С различием применяемых единиц измерения приходится встречаться при изучении динамики: производственных ресурсов, когда они представляются то в стоимостном, то в трудовом исчислении; энергетических мощностей (кВт-ч, л.с.); атмосферного давления и т.д.

Трудности при сравнении данных *по моменту регистрации* возникают из-за сезонных явлений. Цены на овощи и фрукты различна зимой и летом, поэтому уровни при сравнении должны относиться к определенной дате ежегодно.

При анализе показателей в стоимостном выражении следует учитывать, что с течением времени происходит непрерывное *изменение цен*. Причин у этого процесса множество – инфляция, рост затрат, рыночные условия (спрос и предложение) и т.д. В этой связи при характеристике стоимостных показателей объема продукции во времени должно быть устранено влияние изменения цен. Для решения этой задачи количество продукции, произведенное в разные периоды, оценивают в ценах одного периода, которые называют фиксированными или в определенных статистических органах – сопоставимыми ценами.

Широкое использование в статистических исследованиях выборочного метода требует учитывать *достоверность* количественных и качественных характеристик изучаемых явлений в динамике. Различная репрезентативность выборки по периодам внесет существенные погрешности в величины уровней ряда. Так, рейтинг политических деятелей в средствах массовой информации России очень часто определяют по разному числу респондентов.

Одним из условий сопоставимости уровней интервального ряда,

кроме равенства периодов, за которые приводятся данные, является *однородность этапов*, в пределах которых показатель подчиняется одному закону развития. В этих случаях проводят периодизацию временных рядов, типологическую группировку во времени.

Все вышеназванные обстоятельства следует учитывать при подготовке информации для анализа изменений явлений во времени (динамике).

Глава 2. СОСТАВЛЯЮЩИЕ ЭЛЕМЕНТЫ ВРЕМЕННОГО РЯДА

2.1. Понятие об основной тенденции и колеблемости временных рядов

В изучении временных рядов большое место занимает вопрос о закономерностях их движения на протяжении длительного периода. Статистика должна дать характеристику изменений статистических показателей во времени. Как изменяются год за годом валовой национальный продукт и национальный доход страны? Как возрастает или снижается уровень безработицы, оплаты труда? Велики ли колебания урожайности зерновых культур и существует ли тенденция ее роста? На все поставленные вопросы ответ может дать только специальная система статистических методов, предназначенная для изучения развития, изменений во времени или, как принято в статистике говорить, для *изучения динамики*.

Познание закономерностей изменений во времени – сложная и трудоемкая процедура исследования, так как любое изучаемое явление формирует множество факторов, действующих в разных направлениях. По характеру непосредственного воздействия эти факторы могут быть разделены на две группы. К первой группе относятся факторы, определяющие основную тенденцию динамики (рост или снижение уровней). Вторая группа факторов, вызывающая случайные колебания, отклоняет уровни от тенденции то в одном, то в другом направлении. Например, тенденция динамики урожайности связана с прогрессом агротехники, с укреплением экономики данной совокупности хозяйств, совершенствованием организации и управления производством. Колеблемость урожайности вызвана чередованием благоприятных по погоде и неблагоприятных лет, циклами солнечной ак-

тивности, колебаниями в развитии вредных насекомых и болезней растений.

При статистическом изучении динамики необходимо четко разделять два основных ее элемента – тенденцию и колеблемость, чтобы дать каждому из них количественную характеристику с помощью специальных показателей.

Основной *тенденцией*, или *трендом*, называется характеристика процесса изменения явления за длительное время, освобожденная от случайных колебаний, создаваемых второй группой факторов.

В отличие от вариации явлений в пространственной совокупности, измеряемой по отклонениям уровней для отдельных единиц совокупности от их средней величины, *колеблемостью* следует называть отклонения уровней отдельных периодов времени от тенденции динамики (тренда).

2.2. Иерархия тенденций и колебаний

В предыдущих разделах были рассмотрены две основные компоненты временного ряда – его тенденция и колеблемость отдельных уровней. При решении конкретных задач статистического исследования эти компоненты следует разделять, измерять каждую из них отдельно. В то же время при рассмотрении сложных процессов на больших интервалах времени мы наблюдаем иерархию тенденций и колебаний: то, что для времени высокого порядка, например, столетия, выступает как колебания, на интервале времени низшего порядка, например трех-пяти лет, может выступать как тенденция. Например, существует 10-11-летняя циклическая колеблемость солнечной активности, одним из показателей которой служат *числа Вольфа W* (число групп солнечных пятен, умноженное на 10, плюс число отдельных пя-

тен). За 100 лет происходит в среднем 9 или 10 циклов колебаний. Но если рассматривать помесечные данные о числах Вольфа за 2-3 года фазы снижения активности Солнца, то само это снижение можно считать уже не частью колебания, а тенденцией, на фоне которой происходят хаотические, случайные колебания (табл. 2.1).

Таблица 2.1

Динамика чисел Вольфа за 1994 и 1995 гг.

Год	Месяц											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1994	58,8	35,9	31,7	16,7	18,2	28,1	35,0	22,8	26,7	43,8	18,0	26,2
1995	23,8	29,9	31,1	14,6	14,7	15,8	14,6	15,1	12,3	21,7	9,4	10,8

Мы наблюдаем явную тенденцию снижения W при наличии случайных колебаний в отдельные месяцы.

Сезонные колебания продажи ряда предметов одежды и обуви – хорошо известное явление в торговле. Оно проявляется в циклическом изменении месячных объемов продаж на протяжении ряда лет. Но если рассматривать, например, объем продаж босоножек за отдельные дни мая, то заметим на общем фоне «тенденции» роста продажи колебания в отдельные дни недели или в зависимости от погоды дня. То, что для годовых отрезков времени – колеблемость, то для суточных внутри месяца – тенденция. Следовательно, кавычки можно и снять. Последний пример сложной структуры тенденций и колебаний дает нам динамика температуры воздуха, взятая за десятки лет с разбивкой по годам, месячным, суточным и часовым данным.

Имеется «высший» уровень динамики температур – ее тенденция к повышению, в основном в результате антропогенного воздей-

ствия – роста выбросов продуктов сжигания топлива в атмосферу. Это медленная тенденция роста среднегодовых температур примерно на 0,03 градуса за год. На фоне этой тенденции среднегодовые температуры отдельных лет колеблются в среднем на 2 - 3 градуса. Внутри каждого года на средних широтах происходят колебания средних температур месяцев – циклические сезонные колебания, которые, однако, для температуры в отдельные дни выступают как тенденция снижения температуры осенью и ее роста весной.

Около этих тенденций среднесуточные температуры колеблются в основном хаотически, ввиду смены холодных и теплых воздушных масс, т.е. циклонической и антициклонической динамики атмосферы. Но если спуститься на нижележащий уровень времени и рассматривать температуру воздуха в отдельные часы суток, то мы увидим новые, мелкомасштабные циклические колебания часовых температур: с утра и до 13 - 14 ч температура имеет тенденцию роста, а к вечеру – тенденцию снижения ввиду дневного нагревания воздуха солнечным светом и охлаждения ночью. Но в отдельные часы температура колеблется около этих «тенденций» в зависимости от облачности, ветра, дождя и т.д., иногда за полчаса температура воздуха может измениться на 5 - 10 градусов.

Не менее сложны тенденции динамики и колебания потребления электроэнергии в городе, зависящие от числа зданий и предприятий, режима работы последних, от времени года, температуры воздуха, времени суток, от трансляции футбольных матчей или концерта группы «Rolling Stones»... И все эти тенденции и колебания нужно уметь измерить, учесть, прогнозировать для того, чтобы электросистема работала без сбоев и наиболее рентабельно.

В связи с этим знание статистических методов и изучение тен-

денций и колебаний для экономиста-менеджера, для статистика-аналитика имеют огромное значение.

2.3. Периодизация динамики

В экономических исследованиях, при использовании в целях прогнозирования будущих значений анализируемого явления динамических рядов, редко кто задумывается об их однородности или сопоставимости отдельных периодов.

Ярким примером в современной истории России является динамика ВВП РФ, так до кризиса 1998г. динамика показателя имела тенденцию к снижению, при этом наблюдался значительный спад показателя в 1992-1995гг. но после финансового кризиса тенденция кординальным образом поменялась. Отсюда возникает вопрос - можно ли считать данный ряд единым целым? Существует мнение, что на самом деле это разные временные ряды, так как до 1998г. на механизм генерации уровней ряда оказывали влияние одни факторы, а после данного периода уже совершенно другие (например: до 1998г. иностранный капитал, после внутренние финансовые ресурсы).

В отечественной литературе данный вопрос игнорируется, а методы, с помощью которых можно выявить однородные периоды развития явления, носят бессистемный характер. В связи с этим назрела предпосылка для разработки методики периодизации динамических рядов на основе систематизации имеющихся разработок.

Прежде чем непосредственно приступить к рассмотрению достоинств и недостатков алгоритмов периодизации введем само понятие периодизации развития.

Периодизация развития - расчленение изучаемого динамиче-

ского ряда на однородные этапы, в пределах которых показатель (или показатели) подчиняется одному закону развития.

Существующие в литературе алгоритмы можно разделить на точные методы и приближенные (рисунок 2.1), часть из которых может быть применена как к одномерным, а часть к многомерным временным рядам.

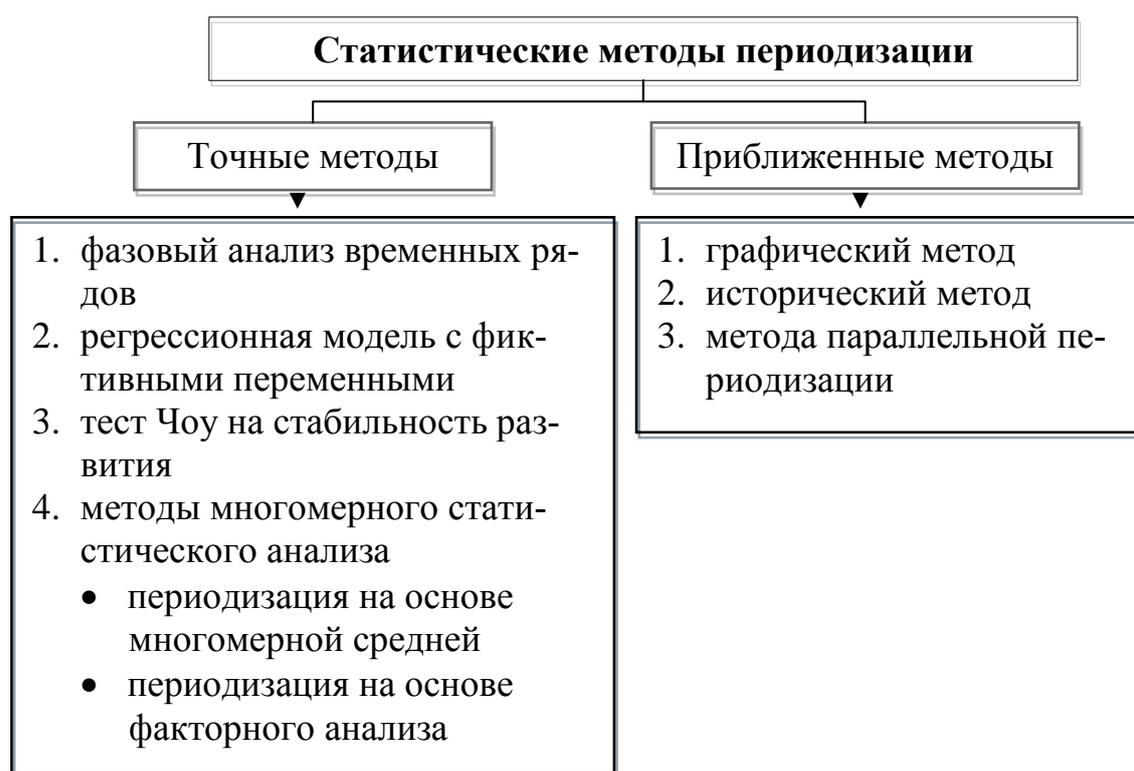


Рис. 2.1 Систематизация методов периодизации временных рядов

Графический метод наиболее простой способ обнаружения однородных периодов развития явления, но вместе с тем в отдельных случаях требует от исследователя большого мастерства.

В качестве примера выявления однородных периодов рассмотрим динамику темпов прироста (снижения) ВВП СССР и России (рисунок 2.2).

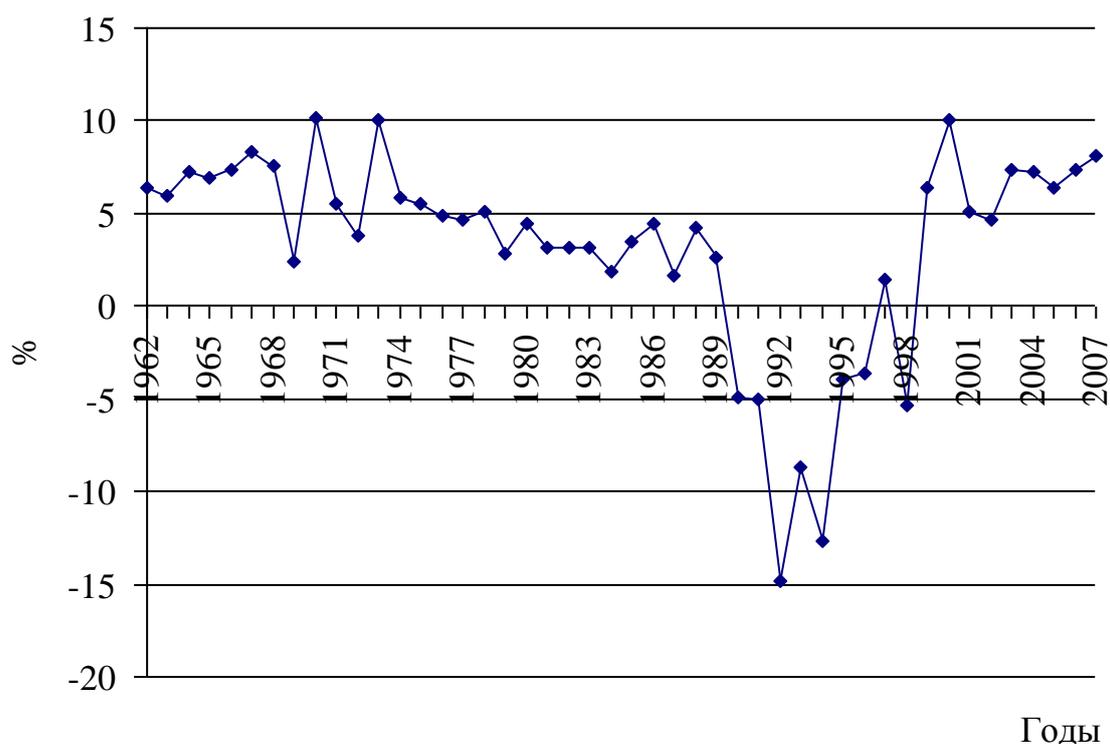


Рис. 2.2 Темп прироста ВВП СССР/России за период 1961-2007 гг., %

Источник. Данные до 1990 г. - Пономаренко А.Н. Ретроспективные национальные счета России: 1961-1990. – М.: Финансы и статистика, 2002. – 256 с.; после 1990 г. - <http://lnweb18.worldbank.org/eca/rus.nsf>

Согласно приведенной на рисунке динамике индекса ВВП, явно прослеживается период, относящийся к СССР и период новой России (до 1990г. и после данного года). При этом колеблемость уровней в первой половине ряда незначительные, а общая тенденция характеризуется снижением темпов развития. Вторая половина ряда характеризуется быстрым увеличением темпов прироста и значительной колеблемостью уровней.

Исторический метод, в данном случае периодизация осуществляется на основе «узаконенной» структуры динамики, при этом обращают внимание на значимые даты и события, а именно: время

принятия управленческих решений по данному показателю, смену хозяйственного механизма, смену руководства, войны и т. п. Недостатком этого метода является то, что точные временные границы периодов путем теоретического анализа удается получить крайне редко.

Развитием данного метода является использование для выявления и описания однородности периодов фазового анализа и теста Г. Чоу.

Применение **фазового анализа временных рядов** подробно рассмотрено в книге Ю.П. Лукашина³. Основная идея метода заключается в том, чтобы постепенно (итеративно) очищать ряд от мало-мощных колебаний, отождествляемых со случайными или второстепенными, конъюнктурными флуктуациями. В ходе такого процесса фильтрации производится сглаживание наименее мощных фаз, в результате чего соседние фазы объединяются в одну более крупную.

Вначале необходимо ввести некоторые понятия, которые понадобятся в дальнейшем.

Флуктуация - величина отклонения x_t от некоторого установленного исследователем уровня. Это может быть отклонение от среднего уровня предшествующего значения, нуля, тренда или какой-либо другой величины. Обозначим величину флуктуации в момент t через Dx_t

Мощность флуктуации - абсолютная величина флуктуации $|Dx_t|$.

Нерегулярные циклические колебания - наличие у ряда x_t разнонаправленных отклонений Dx_t с различной амплитудой без их

³ Лукашин, Ю.П. Адаптивные методы прогнозирования: учеб. пособие / Ю.П.Лукашин. – М.: Финансы и статистика, 2003. - 416 с.:ил. - . ISBN 5-279-02740-5

детерминированной повторяемости.

Фаза - период положительных или отрицательных флуктуации ряда Dx_t

Положительная и отрицательная фазы - периоды соответственно положительных и отрицательных флуктуации ряда.

Длительность фазы - временной отрезок положительных или отрицательных флуктуации ряда.

Точка смены фаз - момент смены знака флуктуации.

Мощность флуктуации ряда, или просто мощность ряда, - сумма абсолютных флуктуации всего ряда, т.е.

$$P = \sum |Dx_t| \quad (2.1)$$

Мощность фазы - сумма абсолютных флуктуации ряда внутри фазы, т.е. мощность фазы i

$$P_i = \sum_{t=t_i}^{t_{i+1}} |Dx_t| \quad (2.2)$$

где: t_i – момент начала фазы i ;

t_{i+1} - момент окончания фазы i (поворотные точки).

Уровень агрегирования фаз - номер итерации в процессе последовательного объединения фаз путем фильтрации маломощных флуктуации. Уровень теряемой мощности - процент мощности ряда, теряемой при элиминировании маломощных флуктуации.

Фазовая диаграмма - представление временного ряда в виде сменяющих друг друга фаз.

Весь процесс периодизации динамики ряда на основе фазового

анализа можно разбить на несколько этапов.

Этап 1. Задаемся критерием остановки итерационного процесса. В качестве такового могут выступать:

а) уровень теряемой мощности ν — процент совокупной мощности ряда, которым мы решаемся пожертвовать в процессе агрегирования фаз путем подавления случайных флуктуации;

б) наперед заданное число фаз, на которое желательно разбить весь выборочный период. Процесс объединения (агрегирования) фаз останавливается при достижении этого числа;

в) средняя длительность фаз. Если, например, ставится задача изучения приблизительно 4-летних циклов, то процесс целесообразно остановить, когда средняя длительность фаз будет близка 2;

г) экспертное решение, принимаемое на основе анализа графиков и числовых таблиц.

Этап 2. Находим наименее мощную фазу, например j . Возможны два варианта ее расположения в выборке: a - в начале или в конце; b - внутри. Рассмотрим второй вариант. Принимаем решение об устранении соответствующих ей флуктуации. Для этого заменяем фазы $j-1, j, j+1$ одной, мощность которой равна $p_{j-1} - p_j + p_{j+1}$. В ходе этой операции число фаз уменьшается на 2, а совокупная мощность временного ряда сокращается на $2p_j$. Таким образом, $2p_j$ - мощность, теряемая в процессе агрегирования фаз в случае b . Случай a отличается от b лишь тем, что число фаз в результате элиминирования наименее мощной краевой фазы уменьшается на одну, а не на две, так как объединению подлежат не три, а две соседние фазы (j и $j+1$, если j - первая фаза; $j-1$ и j , если j - последняя фаза).

Этап 3. Этап 2 будем повторять до тех пор, пока не достигнем заданного критерия остановки (см. этап 1). Тогда процесс агрегирова-

ния прекращается и полученные фазы подлежат содержательному или статистическому анализу.

Этап 4. В результате применения изложенной процедуры весь интервал выборки ряда будет разбит точками смены фаз на отрезки, внутри которых можно представить движение ряда упрощенно линейным образом. Для этого достаточно разделить мощность фазы на ее длительность - получим среднюю мощность флуктуации внутри данной фазы, а ее знак определит знак флуктуации. Интерес может представлять наложение графика агрегированных фаз (фазовой диаграммы) на график исходного ряда.

Рассмотрим алгоритм данного метода применительно к динамическому ряду добычи нефти.

Проведем визуализацию исследуемого динамического ряда (рисунок 2.3). Согласно приведенному рисунку наблюдается динамика показателя к росту, точка смены роста уровней ряда в августе 2005г. и с данного момента наблюдается тенденция к снижению. Тенденция рассматриваемого ряда достаточно точно аппроксимируется полиномом второй степени.

Исключим из рассматриваемых рядов детерминирующую составляющую (тренд) путем перехода к отклонениям от выбранной модели тренда. Тогда в качестве флуктуации Dx_t будут выступать остаточные величины.

В результате реализации пяти итераций фазового анализа, ряд разбит на две фазы (рисунок 2.4):

1. январь 2003г - апрель 2005г. (28 уровней) – восходящая тенденция;
2. май 2005г. – июнь 2007г. (26 уровней) – нисходящая тенденция.

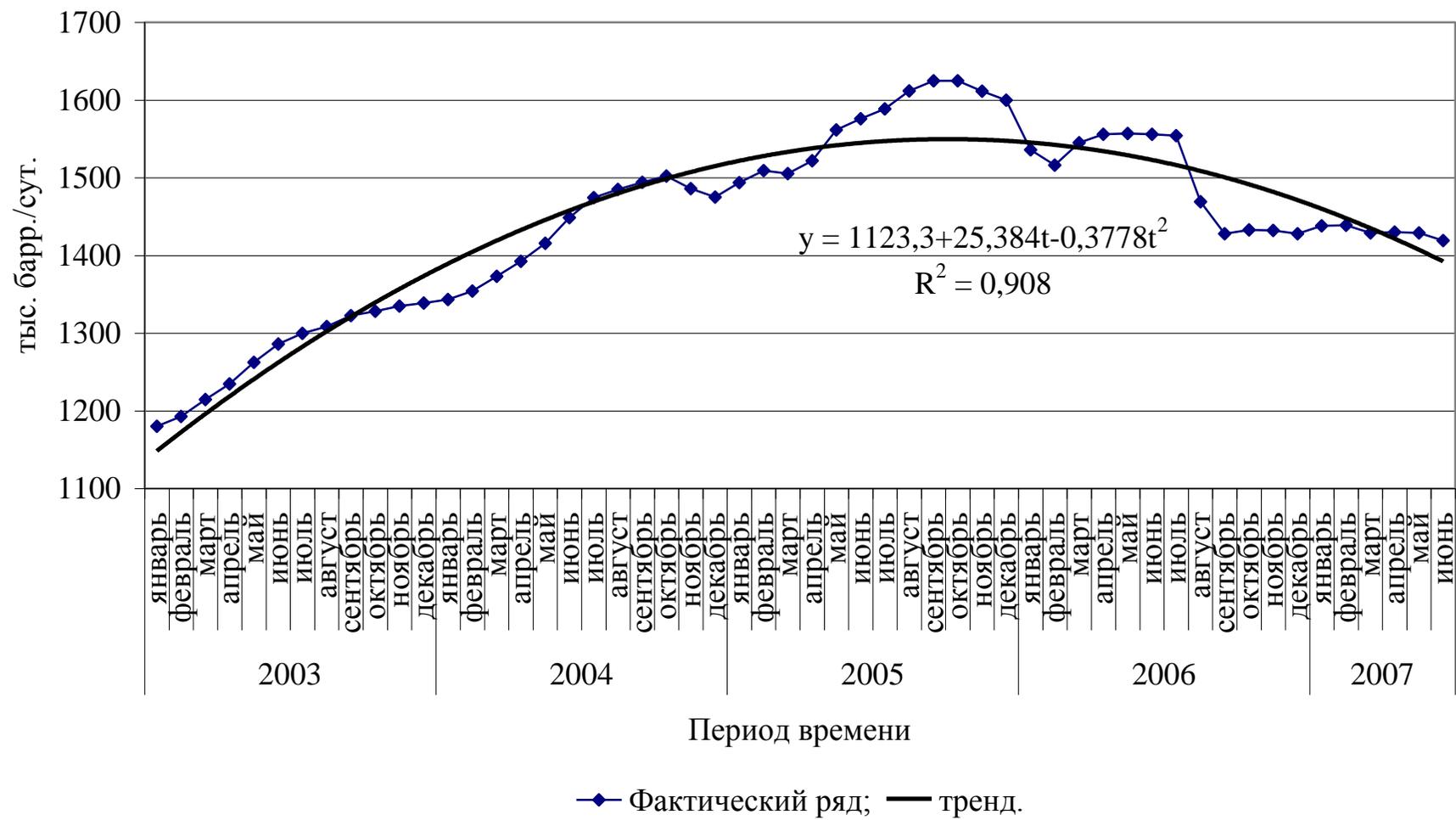


Рис. 2.3 Динамика добычи нефти

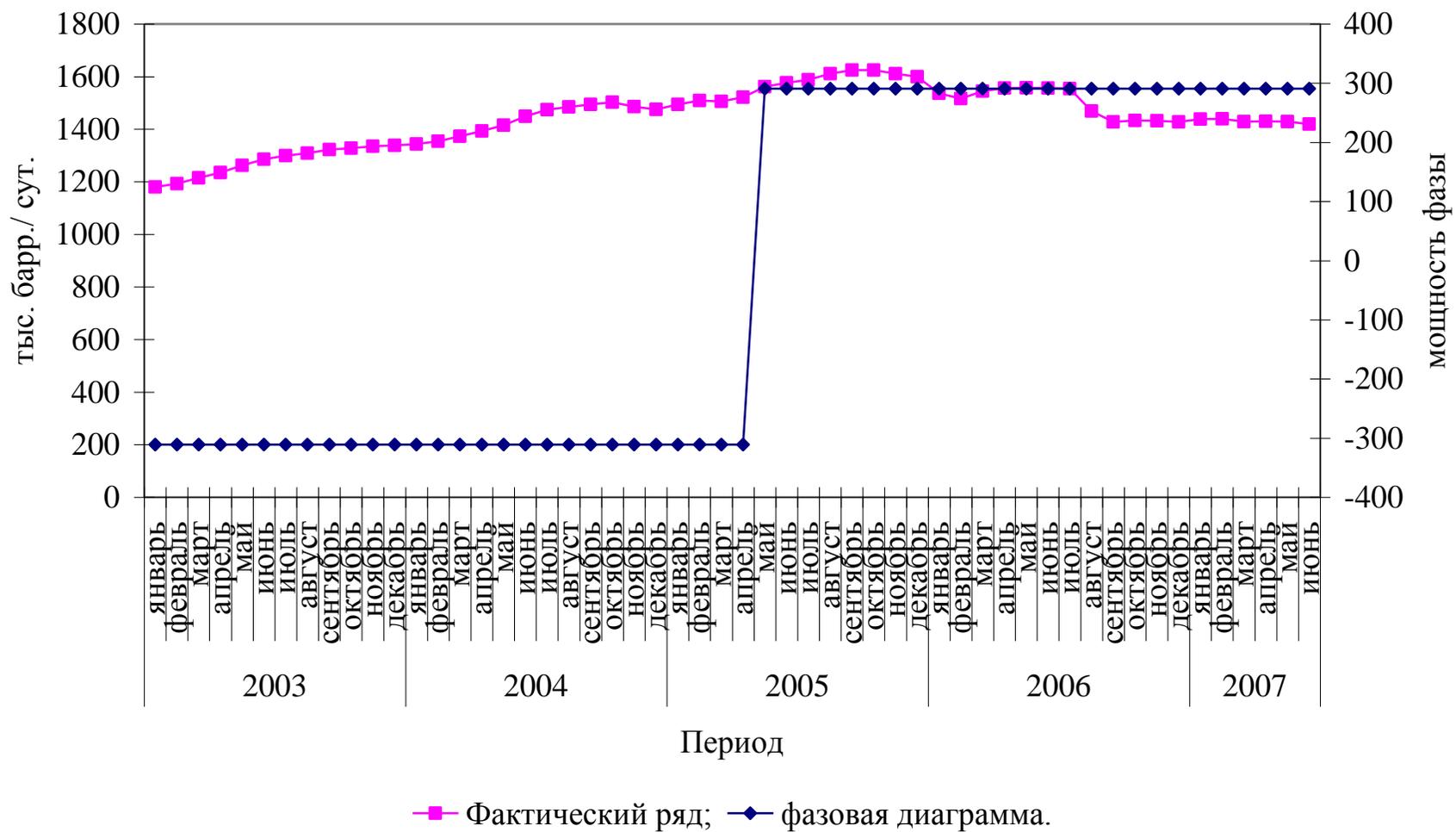


Рис. 2.4. Фазовая диаграмма временного ряда добычи нефти

В Приложении 1 представлены необходимые расчеты при агрегировании фаз.

В экономических исследованиях **тест Г. Чоу** в основном применяется для ответа на вопрос - какая регрессия наилучшим образом отражает динамику показателя, построенная на основе всего ряда или лишь на его половине. Данная постановка вопроса как нельзя лучше подходит для определения однородных отрезков развития динамического ряда.

Применим данный тест для проверки правильности разбиения ряда добычи нефти на фазы.

Основные обозначения для проведения теста Г. Чоу представлены в таблице 2.2.

Выдвинем гипотезу H_0 о структурной стабильности тенденции изучаемого временного ряда.

$$S_{K-Л} = 3765267,41 + 3382151,02 = 7147418,43$$

Тогда сокращение остаточной дисперсии при переходе от единого уравнения тренда к кусочно-линейной модели можно определить по следующей формуле:

$$\Delta S = S_0 = 60432681,18 - 7147418,43 = 53285262,76$$

Определим фактическое значение F -критерия Фишера по следующей формуле:

$$F_{\text{факт}} = \frac{\Delta S}{(k_1 + k_2) - k_0} / \frac{S_{K-Л}}{n - k_1 - k_2} = \frac{53285262,76}{3} / \frac{7147418,43}{48} = 119,28$$

Таблица 2.2

Расчет необходимых характеристик теста Чоу

Уравнение тренда для периода	Вид уравнения тренда	Число наблюдений в совокупности	Остаточная сумма квадратов	Число параметров в уравнении	Число степеней свободы остаточной дисперсии
Январь 2003г. – июнь 2007г.	$\hat{y} = 1123,3 + 25,384 \cdot t - 0,378 \cdot t^2$	$n_0=54$	$S_0= 60432681,18$	$k_0=3$	$n_0-k_0=54-3=51$
Январь 2003 г. – апрель 2005 г.	$\hat{y} = 1164 + 18,51 \cdot t - 0,199 \cdot t^2$	$n_1=28$	$S_1=3765267,405$	$k_1=3$	$n_1-k_1=28-3=25$
Май 2005 г. – июнь 2007 г.	$\hat{y} = 1617,2 - 5,155 \cdot t - 0,137 \cdot t^2$	$n_2=26$	$S_2=3382151,02$	$k_2=3$	$n_2-k_2=26-3=23$

Находим $F_{табл}$ с уровнем значимости $\alpha=0,05$ и числом степеней свободы $(k_1+k_2)-k_0=3$ и $n-(k_1+k_2)=48$, соответственно его значение равно 2,7981.

Получаем $F_{факт} > F_{табл}$, т.е. гипотеза о структурной стабильности тенденции отклоняется, т.е. подтверждается наше предположение относительно однородных отрезков развития анализируемого явления.

Идея **метода параллельной периодизации** (сопоставления параллельных рядов) заключается в следующем. Пусть y - анализируемый показатель, развернутый в динамический ряд $\{y_t\}$, где y_t - значение уровня ряда в момент (интервал) времени t . Возможно, существует показатель x , которому соответствует динамический ряд $\{x_t\}$, определяющий поведение исследуемого показателя y . Тогда в роли однородных периодов развития y нужно взять периоды x .

Рассмотрим условный пример:

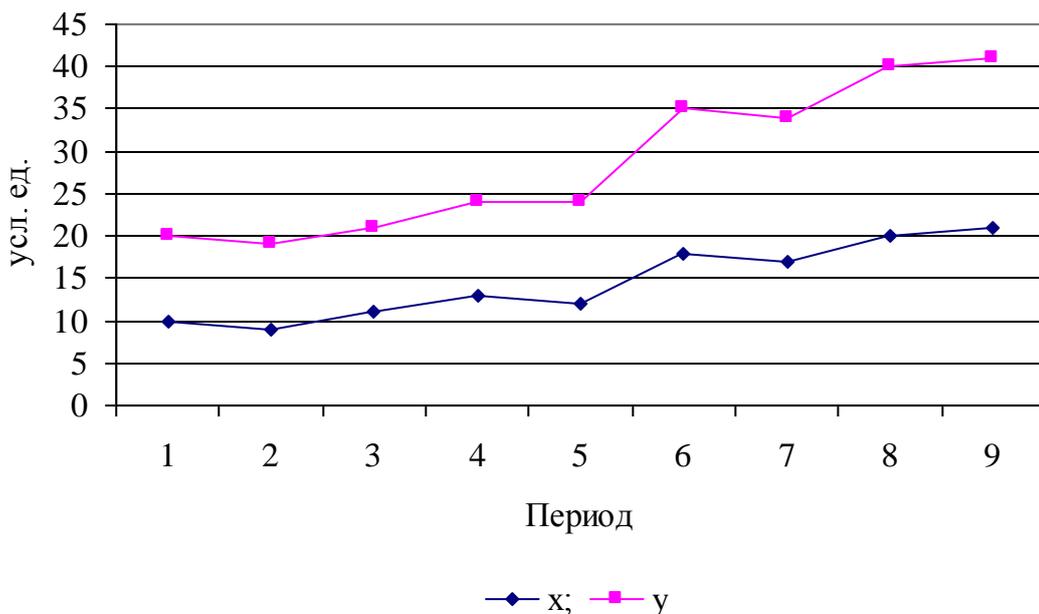


Рис. 2.5 Динамика прибыли предприятия (y) и среднегодовой стоимости основных производственных фондов (x), усл. ед.

Периоды однокачественной динамики показателей x легко выделить: это 1-4 и 5-9. Линейный коэффициент корреляции между этими рядами очень высок: $R = 0,985$. Таким образом, можно считать, что ряд x на 97,1% определяет значение уровней ряда y . Теперь, если предстоит качественный скачок показателя x , то с очень большой степенью вероятности можно ожидать аналогичных изменений показателя y . В качестве недостатка метода параллельной периодизации следует отметить сложности в нахождении x - детерминирующего показателя. Более того, во многих случаях такой параметр вообще невозможно найти, так как он должен обладать весьма редкими свойствами - связью с анализируемым показателем и, главное, неоспоримыми временными границами периодов.

Методы многомерного статистического анализа. Часто требуется выделить однородные периоды в развитии явлений или процессов, получить адекватное отображение которых с помощью одного лишь показателя трудно. К таковым относятся, в частности, здоровье населения, развитие сельскохозяйственного производства и многие другие. Очевидно, что даже такие комплексные показатели, как смертность, продолжительность жизни, заболеваемость, недостаточны для эквивалентного описания столь сложного, интегрированного явления, как здоровье. Необходима система показателей, при которой:

- учитывается многообразие аспектов явления;
- амортизируется искажающее воздействие недостоверных и неточных статистических данных;
- наличие множества показателей повышает обоснованность статистических выводов, т.е. обеспечивается надежность их экстраполяции.

Идеальным выходом является использование множества,

включающего все характеристики процесса. Однако это не всегда возможно по разным причинам, и чаще всего вследствие недоступности статистической информации. На основе комплексных динамических рядов (системы показателей) периодизация реализуется методом многомерной средней и методами факторного анализа.⁴

Периодизация на основе **факторного анализа** наиболее применима на практике, так как: во-первых, методика факторного анализа достаточно разработана, во-вторых, существует реализация метода в современных статистических пакетах программ (STATISTICA 6.0, SPSS 12.0, STADIA 6.0 и др.) что облегчает обработку информации. В качестве иллюстрации данного метода рассмотрим динамику темпов роста ВВП бывших стран «социалистического лагеря».

Анализируя рисунок 2.6 нетрудно заметить, что изображенную на нем динамику темпа прироста ВВП можно интерпретировать как цикл развития экономики. Явно прослеживаются однородные периоды развития: 1 – период спада (1989-1992гг.); 2 – период подъема (1993-1997гг.); 3 – период стабильного роста (1998-2003гг.).

В основе факторного анализа лежит механизм главных компонент. Решение задачи методом главных компонент сводится к поэтапному преобразованию матрицы исходных данных X (рисунок 2.7).

⁴ Глинский В.В., Ионин В.Г. Статистический анализ. – М.: ИНФРА-М; Новосибирск: Сибирское соглашение, 2002. – 241 с.

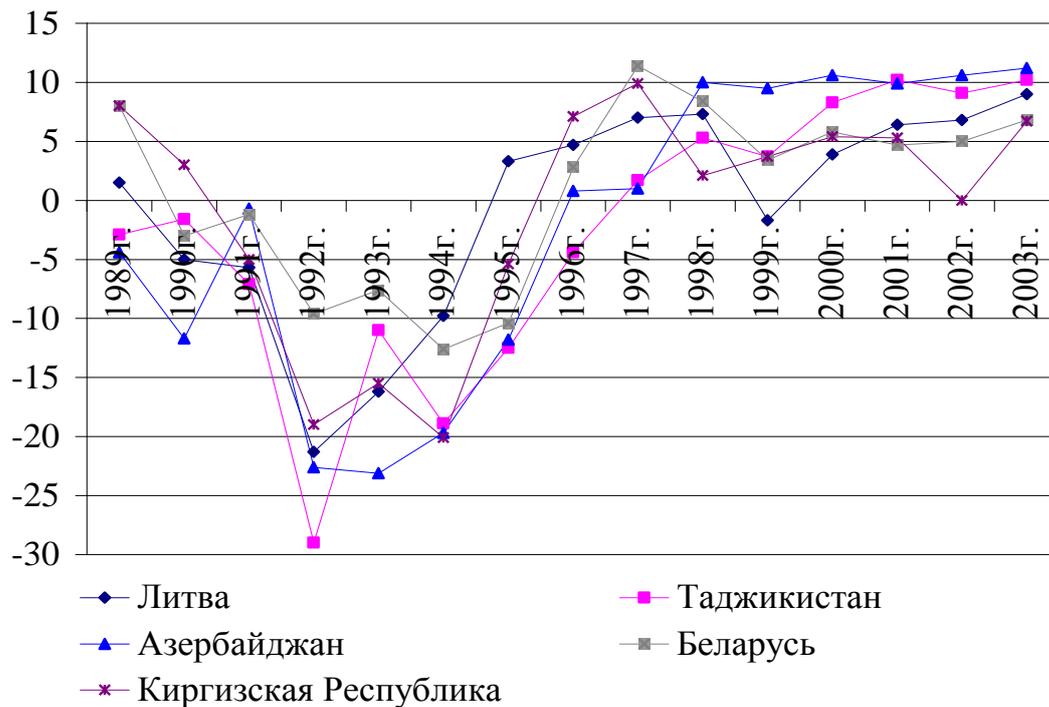


Рис.2.6 Динамика темпа прироста ВВП для бывших стран «социалистического лагеря»

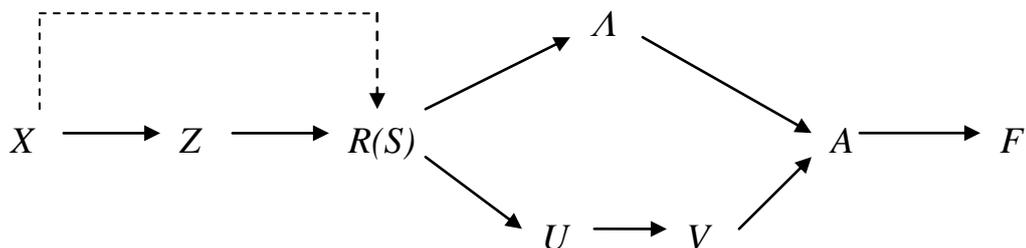


Рис. 2.7 Схема математических преобразований при проведении факторного анализа

где: X – матрица исходных данных размерностью $n \cdot m$ (n – число объектов наблюдения, m – число элементарных аналитических признаков);

Z – матрица центрированных и нормированных значений при-

Приведенная система объединяет однородные линейные уравнения, и так как число ее уравнений равно числу неизвестных, она имеет бесконечное множество решений. Конкретные значения собственных векторов при этом можно найти, задавая произвольно, по крайней мере, величину одной компоненты каждого вектора.

A – матрица факторного отображения, ее элементы a_{rj} – весовые коэффициенты. Вначале A имеет размерность $m \cdot m$ – по числу элементарных признаков X_j , затем в анализе остается r наиболее значимых компонент, $r \leq m$. Вычисляют матрицу A по известным данным матрицы собственных чисел Λ и нормированных собственных векторов V по формуле $A = V\Lambda^{1/2}$.

F – матрица значений главных компонент размерностью $r \cdot n$, $F = A^{-1}Z'$. Эта матрица в общем виде записывается:

$$\begin{array}{cccc}
 & n_1 & n_2 & \dots & n_n \\
 F_1 & f_{11} & f_{12} & \dots & f_{1n} \\
 F = F_2 & f_{21} & f_{22} & \dots & f_{2n} \\
 \dots & \dots & \dots & \dots & \dots \\
 F_r & f_{r1} & f_{r2} & \dots & f_{rn}
 \end{array}$$

В нашем случае данные представляют собой темпы роста (снижения), т.е. их можно считать стандартизованными (приведенными к общему уровню) и нет необходимости в образовании матрицы центрированных и нормированных значений признаков Z_{ij} .

Следующим этапом проведения факторного анализа является выбор оптимального числа факторов (компонент). Для этого на практике прибегают к рассмотрению объясненной дисперсии значениями главных компонент или к помощи критерия каменистой осыпи.

В нашем случае дисперсия, объясненная последовательными факторами, представлена в следующей таблице.

Таблица 2.3

**Собственные значения главных компонент
бывших стран «социалистического лагеря»**

Количество факторов	Собственные значения	% общей дисперсии	Кумулятивные собственные значения	Кумулятивный %
1	8,2025	54,683	8,203	54,683
2	2,8923	19,282	11,095	73,966
3	2,3828	15,885	13,478	89,851

Во втором столбце таблицы представлена дисперсия нового, только что выделенного фактора. В третьем столбце для каждого фактора приводится процент от общей дисперсии для каждого фактора. Четвертый столбец содержит накопленную или кумулятивную дисперсию.

Согласно приведенным данным можно утверждать, что три собственных фактора имеют значения превосходящие единицу. Следовательно, для анализа целесообразно отобрать только три фактора. Первый фактор объясняет 54,7% суммарной дисперсии, второй фактор 19,3% и третий фактор 15,9%.

Еще одним инструментом, помогающим в отборе факторов, является критерий каменистой осыпи. Это графический метод, с помощью которого собственные значения выводятся в виде простого графика.

Автор данного критерия - Кэттель предложил найти такое место на графике, где убывание собственных значений слева направо

максимально замедляется. Предполагается, что справа от этой точки находится только «факториальная осыпь» - «осыпь» является геологическим термином, обозначающим обломки горных пород, скапливающиеся в нижней части скалистого склона. В нашем случае получаем рисунок 2.8.

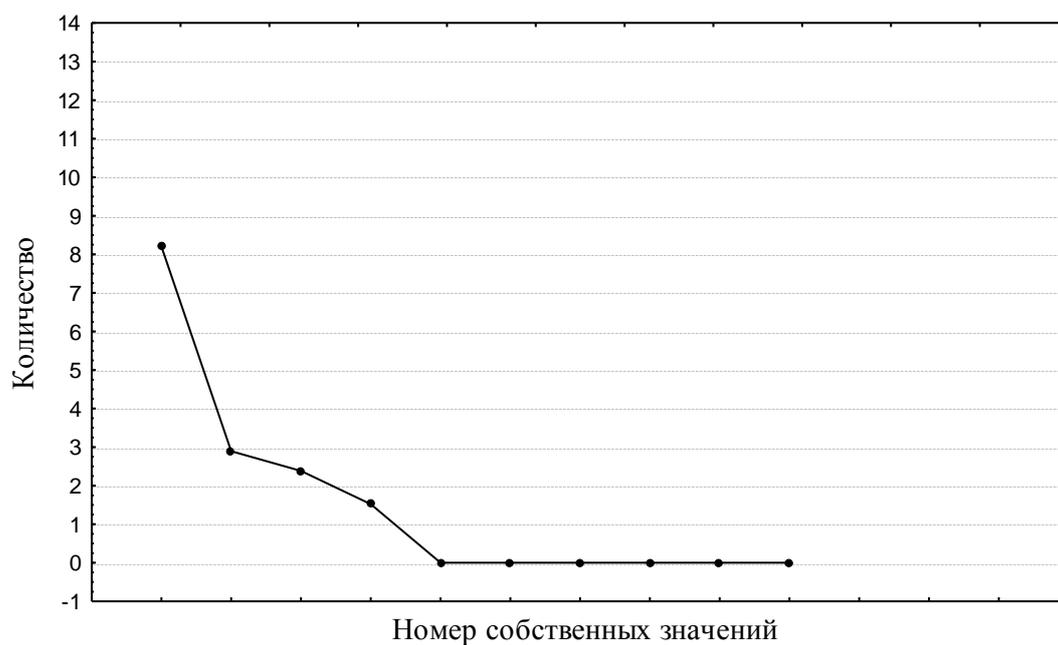


Рис. 2.8 Собственные значения факторов относительно номеров факторов для бывших стран «социалистического лагеря»

В соответствии с критерием каменистой осыпи можно оставить в нашем случае 2 или 3 фактора.

После того как было выбрано 3 фактора приступим к самой сложной части факторного анализа – объяснению отобранных факторов. В данном случае задаем в неявном виде типы динамики и состояний. В общем виде типы состояний могут быть представлены как высокий, средний и низкий, что соответствует типам динамики «рост», «стабилизация» и «снижение».

Для разграничения динамики на типы (однородные периоды развития) обратимся к факторной матрице (таблица 2.4) отмечая ту факторную нагрузку, которая имеет наибольшее абсолютное значение.

Таблица 2.4

**Матрица факторных нагрузок
для бывших стран «социалистического лагеря»**

Годы	Фактор 1	Фактор 2	Фактор 3
1989г.	0,946	-0,189	0,263
1990г.	0,718	0,757	0,402
1991г.	-0,152	-0,969	0,185
1992г.	0,652	-0,736	0,100
1993г.	0,530	0,737	0,135
1994г.	0,457	-0,243	-0,802
1995г.	0,536	0,186	-0,706
1996г.	0,742	-0,263	-0,758
1997г.	0,959	-0,259	0,791
1998г.	-0,505	-0,719	-0,729
1999г.	-0,657	-0,398	0,617
2000г.	-0,904	-0,175	0,390
2001г.	-0,955	0,294	0,000
2002г.	-0,881	-0,112	-0,378
2003г.	-0,951	0,113	-0,278

Представленные в таблице факторные нагрузки следует понимать как корреляционные коэффициенты между переменными и фак-

торами. Так 1989г. сильнее всего связан с фактором 1 (величина корреляции составляет 0,946), 1990г. сильнее всего коррелирует с фактором 2 (0,757) и т.д.

Первый фактор, по видимому, можно интерпретировать как «стабильность» в динамике темпа прироста ВВП данных странах. Второй фактор необходимо интерпретировать как «снижение» и третий как «рост».

По результатам факторного анализа в группе стран вошедших в первый кластер можно выделить следующие периоды трансформации экономики (таблица 2.5).

Таблица 2.5

**Периодизация трансформации экономики
для бывших стран «социалистического лагеря»**

Годы	Типы развития
1989	стабилизация
1990-1993	снижение / нестабильность
1994-1999	рост
1999-2003	стабилизация

Глава 3. ПОКАЗАТЕЛИ ВРЕМЕННОГО РЯДА И МЕТОДЫ ИХ ИСЧИСЛЕНИЯ

3.1. Показатели, характеризующие тенденцию динамики

Чтобы построить систему показателей, характеризующих тенденцию динамики, нужно ответить на вопрос: какие черты, свойства этой тенденции нужно измерить и выразить в статистических показателях? Очевидно, что нас интересует величина изменений уровня как в абсолютном, так и в относительном выражении (на какую долю, процент уровня, принятого за базу, произошло изменение?). Далее нас интересует, является ли изменение равномерным или неравномерным, ускоренным (замедленным). Наконец, нас интересует выражение тенденции в форме некоторого достаточно простого уравнения, наилучшим образом аппроксимирующего фактическую тенденцию динамики. Понятие об уравнении тенденции динамики было введено в статистику английским ученым Гукером в 1902 г. Он предложил называть такое уравнение *трендом* (trend).

Для того чтобы нагляднее представить показатели, характеризующие тенденцию, следует абстрагироваться от колеблемости и выявить динамический ряд в форме «чистого» тренда при отсутствии колебаний. Пример такого ряда представлен в табл. 3.1.

Абсолютное изменение уровней – в данном случае его можно назвать *абсолютным приростом* – это разность между сравниваемым уровнем и уровнем более раннего периода, принятым за базу сравнения. Если эта база – непосредственно предыдущий уровень, показатель называют *цепным*, если за базу взят, например, начальный уровень, показатель называют *базисным*. Формулы абсолютного изменения уровня:

Таблица 3.1

Абсолютные и относительные показатели тенденции

Номера периодов или моментов времени	Уровни ряда, тыс. т	Абсолютное изменение уровней, тыс. т в год	Ускорение абсолютно-го изменения, тыс. т·год ⁻²	Темп роста уровня, в % к предыдущему	Темп роста уровня, в % к начальному
0 (начальный)	100	-	-	-	-
1	112	12	-	112,0	112
2	128	16	4	114,3	128
3	148	20	4	115,6	148
4	172	24	4	116,2	172
5	200	28	4	116,3	200
6	232	32	4	116,0	232

цепное $\Delta_y = y_n - y_{n-1}$;

базисное $\Delta_0 = y_n - y_0$.

Если абсолютное изменение отрицательно, его следует называть абсолютным сокращением. Абсолютное изменение имеет ту же единицу измерения, что и уровни ряда с добавлением единицы времени, за которую определено изменение: 22 тысячи тонн в год (или 1,83 тыс. т в месяц, или 110 тыс. т в пятилетие). Без указания единицы времени, за которую произошло измерение, абсолютный прирост нельзя правильно интерпретировать.

В табл. 3.1 абсолютное изменение уровня не является константой тенденции. Оно со временем возрастает, т.е. уровни ряда изменяются с ускорением. *Ускорение – это разность между абсолютным изменением за данный период и абсолютным изменением за предыдущий период равной длительности:*

$$\Delta'_i = \Delta_i - \Delta_{i-1}.$$

Показатель абсолютного ускорения применяется только в цепном варианте, но не в базисном. Отрицательная величина ускорения говорит о замедлении роста или об ускорении снижения уровней ряда.

Как видно по данным табл. 3.1, ускорение является константой тенденции данного ряда, что свидетельствует о параболической форме этой тенденции. Ее уравнение имеет вид:

$$y_i = y_0 + at_i + bt_i^2, \quad (3.1)$$

где y_0 – уровень ряда в начальный (нулевой) период;

a – средний абсолютный прирост (по всему ряду);

b – половина ускорения;

t_i – номера периодов.

По данным табл. 3.1 имеем:

$$y_i = 100 + 10t_i + 2t_i^2$$

Показатель ускорения абсолютного изменения уровней выражается в единицах измерения уровня, деленных на квадрат длины пери-

ода. В нашем случае ускорение составило 4 тыс. т в год за год или 4 тыс. т·год². Смысл показателя следующий: объем производства (или добычи угля, руды) имел абсолютный прирост, возрастающий на 4 тыс. т в год ежегодно.

Усвоить рассмотренные показатели поможет следующая аналогия с механическим движением: уровень – это аналог пройденного пути, причем начало его отсчета не в нулевой точке. Абсолютный прирост – аналог скорости движения тела, а ускорение абсолютного прироста – аналог ускорения движения. Пройденный телом путь, считая и тот, который уже был пройден до начала отсчета времени в данной задаче, равен:

$$S = S_0 + V_0 t + \frac{at^2}{2},$$

где S_0 – путь, пройденный до начала отсчета времени;

V_0 – начальная скорость;

a – ускорение;

t – время, прошедшее от начала его отсчета в задаче. Сравнивая с формулой (3.1), видим, что S_0 – аналог свободного члена y_0 , V_0 – аналог начального абсолютного изменения a ; $a/2$ – аналог ускорения прироста b .

Система показателей должна содержать не только абсолютные, но и относительные статистические показатели. Относительные показатели динамики необходимы для сравнения развития разных объектов, особенно если их абсолютные характеристики различны. Предположим, другое предприятие увеличивало производство аналогичной продукции с тенденцией, выраженной уравнением тренда:

$y_i = 20 + 4t_i + 0,5t_i^2$. И абсолютный прирост, и ускорение роста объема продукции во втором предприятии гораздо меньше, чем в первом. Но можно ли ограничиться этими показателями и сделать вывод, что развитие второго предприятия более медленное, чем первого? Меньший уровень еще не есть меньший темп развития, и это покажет относительная характеристика тенденции динамики – темп роста.

Темп роста – это отношение сравниваемого уровня (более позднего) к уровню, принятому за базу сравнения (более раннему). Темп роста исчисляется в цепном варианте – к уровню предыдущего года и в базисном варианте – к одному и тому же, обычно начальному уровню (см. формулы (3.2)). Он говорит о том, сколько процентов составляет сравниваемый уровень по отношению к уровню, принятому за базу, или во сколько раз сравниваемый уровень больше уровня, принятого за базу. При этом если уровни снижаются со временем, то сказать, что последующий уровень "больше в 0,33 раза", или составляет 33,3% базового уровня, это, разумеется, означает, что уровень уменьшился в 3 раза. Но сказать что "уровень меньше в 0,33 раза", это неверно. Темп изменения в размах всегда говорит о том, во сколько раз сравниваемый уровень больше.

Теперь можно сказать, что относительная характеристика роста объема продукции на первом предприятии в среднем за год близка к 115% (рост приблизительно на 15% за год), и за шесть лет продукция увеличилась в 2,32 раза, а на втором предприятии, вычислив также шесть уровней параболического тренда, читатель убедится, что в среднем за год объем продукции возрастал примерно на 20% , а за шесть лет объем ее возрос в 3,1 раза. Следовательно, в относительном выражении объем продукции на втором предприятии развивался, возрастал быстрее. Только в сочетании абсолютных и относительных ха-

ра характеристик динамики можно правильно отразить процесс развития совокупности (объекта).

Рассмотрим связь абсолютных и относительных показателей динамики. Обозначим темп изменения через k . Тогда имеем:
цепной темп роста в период с номером n

$$k_n = \frac{y_n}{y_{n-1}}; \quad (3.2)$$

базисный темп роста за весь период между базой (0) и текущим годом (n)

$$k_{n/0} = \frac{y_n}{y_0}$$

Если сравниваемый уровень выразить через уровень базисного (или предыдущего) периода и абсолютное изменение, получим:

$$k_n = \frac{y_{n-1} + \Delta}{y_{n-1}} = 1 + \frac{\Delta}{y_{n-1}}, \text{ или } 100\% + \frac{\Delta 100}{y_{n-1}} \quad (3.3)$$

Величина Δ/y_{n-1} , т.е. отношение абсолютного изменения к уровню предыдущего (или базисного) года, называется *относительным приростом* (относительным сокращением, относительным изменением, процентным изменением) или *темпом прироста*. Он равен темпу изменения (роста) минус единица (минус 100%).

Темп изменения – величина всегда положительная. Если уровень ряда динамики принимает положительные и отрицательные зна-

чения, например, финансовый результат от реализации продукции предприятием может быть прибылью (+), а может быть убытком (–), тогда темп изменения и темп прироста применять нельзя. В этом случае такие показатели теряют смысл и не имеют экономической интерпретации. Сохраняют смысл только абсолютные показатели динамики.

Рассмотрим соотношения между цепными и базисными показателями на примере данных табл. 3.1:

1) сумма цепных абсолютных изменений равна базисному абсолютному изменению

$$\sum a_{i(\text{цепн})} = a_{i(\text{баз})},$$

$$12+16+20+24+28+32 = 232-100 = 132;$$

2) произведение цепных темпов изменения равно базисному темпу изменения

$$\prod_i k_{i(\text{цепн})} = k_{i(\text{баз})},$$

$$1,12 \cdot 1,143 \cdot 1,156 \cdot 1,162 \cdot 1,163 \cdot 1,16 = 2,32.$$

Неверно, будто сумма цепных темпов прироста равна базисному темпу прироста, %.

$$12 + 14,3 + 15,6 + 16,2 + 16,3 + 16 \neq 132.$$

Значения цепных темпов прироста, рассчитанных каждый к своей базе, различаются не только числом процентов, но и величиной аб-

солютного изменения, составляющей каждый процент. Поэтому складывать или вычитать цепные темпы прироста нельзя. *Абсолютное значение 1% прироста равно сотой части предыдущего уровня или базисного уровня.*

3.2. Особенности показателей для рядов, состоящих из относительных уровней

Уровнями динамического ряда могут быть не только абсолютные показатели – численности совокупностей или объемы их признаков. Ряды динамики могут отражать развитие структуры совокупности, изменение со временем вариации признака в совокупности, взаимосвязи между признаками, соотношения значений признака для разных объектов. В этих случаях уровни динамического ряда сами являются относительными показателями, нередко выражаются в процентах. Следовательно, абсолютные изменения (и ускорения) тоже окажутся относительными величинами, могут быть выражены в процентах. В процентах, разумеется, будут выражены темпы изменения и относительные приросты. Все это создает нередко путаницу в интерпретации и использовании показателей динамики в печати и даже в специальной экономической литературе.

Рассмотрим пример. В США с конца XIX в. для группы ведущих акционерных компаний исчисляется так называемый индекс Доу - Джонса – арифметическая средняя величина котировок акций на фондовых биржах. Этот показатель характеризует хозяйственную конъюнктуру: если индекс Доу - Джонса повышается, т.е. растет относительная цена акций, значит, вкладчики капитала рассчитывают получить по акциям больший дивиденд (распределяемая часть прибы-

ли). Это говорит о росте деловой активности. Падение индекса Доу - Джонса говорит о снижении деловой активности в стране. Величина этого показателя есть отношение в процентах цены акций на бирже к их номиналу (первоначальной цене при выпуске акций). Это отношение зависит не только от колебаний деловой активности, но имеет также общую тенденцию роста ввиду инфляции – падения покупательной силы доллара США. С начала века этот рост значителен, поэтому в наше время индекс Доу-Джонса составляет более 2000% (акция, когда-то выпущенная на сумму 100 долл., теперь стоит более 2000 современных долларов).

Например, «Биржевые ведомости» за 5.05.90 сообщают: индекс Доу - Джонса на 3.05.90 г. составил 2689,64% по сравнению с 2759,55% на 29.04.90 г. Если вычислить показатель абсолютного изменения индекса, т.е. $2689,64\% - 2759,55\% = -69,91\%$, и сказать, что индекс Доу - Джонса за неделю понизился почти на 70%, создается ложное впечатление о чудовищном крахе на биржах США, потому что снижение на 70% воспринимается как темп изменения – будто от прежней цены акций осталось только 30%.

На самом деле снижение показателей с 2760 до 2690% никакой катастрофой экономике США не грозит: это обычная на рынке ценных бумаг колеблемость курсов. «Биржевые ведомости» далее сообщали, что индекс Доу - Джонса на 7.06.1990 г. достиг 2911,6%, т.е. с 5.05.1990 г. возрос на 222 единицы, которые во избежание путаницы принято именовать *пунктами*. В первом рассмотренном случае индекс снизился на 70 пунктов, во втором – возрос на 222 пункта, а не процента. В процентах рост составил: $222 : 2690 = 8,25\%$ – это и есть темп прироста курса акций.

Аналогичный подход и термины должны применяться к изме-

нению структуры. Например, общее производство электроэнергии в Российской Федерации в 1980 г. составляло 805 млрд. кВт·ч, в том числе на АЭС 54 млрд. кВт·ч, т.е. их доля была равна 6,7%. В 1991 г. общее производство электроэнергии составило 1068 млрд. кВт·ч, а доля АЭС - 11,2%. Доля атомных станций за 11 лет возросла на 11,2 - 6,7 = 4,5 пункта. А темп прироста доли составил $4,5 : 6,7 = 67\%$.

Показатели динамики долей имеют еще одну особенность, вытекающую из того, что сумма всех долей в любой период времени равна единице, или 100%. Изменение, произошедшее с одной из долей, поэтому, неизбежно меняет и доли всех других частей целого, если даже по абсолютной величине эти части не изменились. Казалось бы, это положение самоочевидно, однако нередко в печати встречаются рассуждения о том, что увеличение доли пшеницы и ячменя среди зерновых культур – это хорошо, но вот плохо, что уменьшились доли ржи, овса и гречихи. Как будто все доли сразу могут увеличиться!

Если признак варьирует альтернативно, то увеличение доли одной группы равно уменьшению доли другой группы в пунктах, но темпы изменения долей в процентах при этом могут сильно различаться. Темп больше у той доли, которая в базисном периоде была меньше. Например, удельный вес жилой площади, оборудованной водопроводом, в городском государственном и общественном жилом фонде в 1970 г. составлял 78,9%, а в 1989 г. достиг 92,9%, т.е. возрос на 14 пунктов, или на $14 : 78,9 = 17,7\%$. Соответственно доля необорудованной водопроводом жилой площади снизилась за 19 лет с 21,1 до 7,1%, т.е. на те же 14 пунктов, это снижение составило уже $14 : 21,1 = 66,4\%$.

В общем виде темп роста одной из альтернативных долей зави-

сит от темпа роста другой доли и величины этой доли следующим образом:

$$k_2 = \frac{1 - k_1 x_0}{1 - x_0} \quad (3.4)$$

где x_0 – доля в базисном периоде одного из альтернативных значений признака;

k_1 – темп роста этой доли;

k_2 – темп изменения доли второго альтернативного значения признака.

Абсолютное изменение долей в пунктах зависит от величины доли и темпа роста таким образом:

$$A_1 = -A_2 = x_0 (k_i - 1) \cdot 100. \quad (3.5)$$

При наличии в совокупности не двух, а более групп абсолютное изменение каждой из долей в пунктах зависит от доли этой группы в базисный период и от соотношения темпа роста абсолютной величины объемного признака этой группы со средним темпом роста объемного признака во всей совокупности. Доля i -й группы в сравниваемый (текущий) период определяется как

$$d_{i1} = d_{i0} \frac{k_i}{k} = \frac{d_{i0} k_i}{\sum_{i=1}^m d_{i0} k_i}, \quad (3.6)$$

где d_{i0}, d_{i1} – доли i -й группы в базисный и текущий периоды;

k_i , – темп роста объемного признака в i -й группе;

\bar{k} – средний темп роста;

m – число групп.

Если, например, в базисном году поголовье коров в личных и частных хозяйствах составило в области 68 тыс. голов из общего поголовья 450 тыс. гол., а по прогнозу через 10 лет поголовье коров в индивидуальных хозяйствах возрастет в 5 раз при общем темпе роста поголовья в области 120%, то доля индивидуальных хозяйств в поголовья коров, по прогнозам, должна будет составить:

$$0,151 \cdot \frac{5}{1,2} = 0,6296, \text{ или почти } 63\% \text{ поголовья.}$$

Особенностью показателей динамики относительных величин интенсивности является то, что темпы роста и темпы прироста (или сокращения) прямого и обратного показателей не совпадают. Пусть, например, трудоемкость производственной операции на старом станке составляла 10 мин., а производительность труда соответственно 48 оп. за смену. После замены станка на новый трудоемкость операции снизилась в пять раз (до 2 мин.), а производительность возросла в те же пять раз – до 240 оп. за смену. Относительное изменение трудоемкости составило $(2 - 10) : 10 = -0,8$, т.е. трудоемкость снизилась на 80%. Относительное изменение производительности труда составило $(240 - 48) : 48 = 4$, или 400%, т.е. производительность труда возросла на 400%. Причина состоит в том, что пределом, к которому стремятся по мере прогресса показатели ресурсоотдачи, является бесконечность, а пределом, к которому стремятся обратные им показатели ресурсоемкости, является нуль. Понимание разного поведения показателей ди-

намики прямых и обратных мер эффективности очень важно для экономиста и статистика.

По мере приближения относительного показателя к пределу одно и то же абсолютное изменение в пунктах приобретает иное качественное содержание. Так, например, если показатель тесноты связи – коэффициент детерминации – возрос с 40 до 65% (на 25 пунктов), то система факторов в регрессионном уравнении как была, так и осталась неполной, хорошей модели не получено. Но если после изменения состава факторов коэффициент детерминации возрос с 65 до 90% – на те же 25 пунктов, это изменение имеет другое качественное содержание: получена хорошая регрессионная модель, в основном объясняющая вариацию результативного признака с достаточно полной системой факторов.

3.3. Средние показатели временных рядов

Средние показатели динамики – средний уровень ряда, средние абсолютные изменения и ускорения, средние темпы роста – характеризуют тенденцию. Они необходимы при обобщении характеристик тенденции за длительный период, по различным периодам; они незаменимы при сравнении развития за неодинаковые по длительности отрезки времени, при выборе аналитического выражения тренда. При наличии в динамическом ряду существенных колебаний уровней определение средних показателей тенденции требует применения специальных методов статистики, которые излагаются в последующих разделах. В данном разделе рассматривается только форма, математические свойства средних показателей динамики и простейшие приемы их вычисления, применимые на практике к рядам со слабой

колеблемостью.

Средний уровень интервального ряда динамики определяется как простая арифметическая средняя из уровней за равные промежутки времени:

$$\bar{y} = \frac{\sum_{(i)}^n y_i}{n} \quad (3.7)$$

или как взвешенная арифметическая средняя из уровней за неравные промежутки времени, длительность которых и является весами.

Таблица 3.2

Динамика численности населения РФ

Годы	Численность населения, тыс. человек	Годы	Численность населения, тыс. человек
1989	147400	1997	147502
1990	148041	1998	147105
1991	148543	1999	146388
1992	148704	2000	145300
1993	148673	2001	145600
1994	148366	2002	145200
1995	148306	2003	144200
1996	147976	2004	143500

Источник. **Симчера В.М.** Развитие экономики России за 100 лет: 1900-2000. Исторические ряды, вековые тренды, периодические циклы / В.М. Симчера. – М.: ЗАО «Издательство Экономика», 2007. – 683с. ISBN 5-282-02627-9

По данным табл. 3.2 определим среднегодовые уровни численности населения по восьмилетиям.

Период	Среднегодовые уровни, тыс. человек
1989 – 1996	148251
1997 – 2004	145599

Средние уровни принято относить к середине осредняемого отрезка времени, т.е. в нашем примере к средним годам каждого восьмилетия.

Если, например, с 1-го числа месяца по 18-е число на предприятии работали 45 человек, с 19-го по 27-е – 48 человек, а с 28-го по 31-е число – 54 человека, то среднее списочное число работников за месяц составит:

$$y = \frac{45 \cdot 18 + 48 \cdot 9 + 54 \cdot 4}{31} = 47,03 \text{ чел.}$$

В моментном ряду роль, смысл среднего уровня в том, что он характеризует уже не состояние объекта в отдельные моменты, а его среднее, обобщенное состояние между начальным и конечным моментом учета. Из этого следует, что роль уровней, относящихся к начальному и конечному моменту, существенно иная, чем роль уровней на моменты внутри изучаемого отрезка времени. Начальный и конечный уровни находятся на границе изучаемого интервала, они наполовину относятся к предыдущему и последующему интервалам и лишь наполовину к изучаемому. Уровни, относящиеся к моментам внутри осредняемого интервала, целиком относятся только к нему. Отсюда получаем особую форму средней арифметической величины, называемой *хронологической средней*:

$$y_{\text{хрон}} = \left(\frac{y_1}{2} + \sum_{i=2}^{n-1} y_i + \frac{y_n}{2} \right) : (n-1) \quad (3.8)$$

Проблема вычисления среднего уровня моментного ряда при неравных промежутках между моментами является спорной и здесь не рассматривается.

Если известны точные даты изменения уровней моментного ряда то средний уровень определяется как

$$\bar{y} = \frac{\sum y_i t_i}{\sum t_i} \quad (3.9)$$

где t_i – время, в течение которого сохранялся уровень.

Средний абсолютный прирост (абсолютное изменение) определяется как простая арифметическая средняя из абсолютных изменений за равные промежутки времени (цепных абсолютных изменений) или как частное от деления базисного абсолютного изменения на число осредняемых отрезков времени от базисного до сравниваемого периода:

$$\bar{\Delta} = \frac{\Delta_i}{n} = \frac{y_n - y_0}{n} \quad (3.10)$$

Например, производство телевизоров в Российской Федерации в 1980 г. составило 4013 тыс. шт., а в 1990 г. – 4717 тыс. шт. Среднегодовой абсолютный прирост производства телевизоров за 10 лет составил:

$$\bar{\Delta} = \frac{4717 - 4013}{10} = 70,4 \text{ тыс. шт. за год.}$$

Для правильной интерпретации показатель среднего абсолютно-го изменения должен сопровождаться указанием двух единиц времени: 1) времени, за которое он вычислен, к которому относится и которое он характеризует (в нашем примере это десятилетие – 1980–1990 гг.); 2) время, на которое показатель рассчитан, время, входящее в его единицу измерения, – 1 год. Можно рассчитать среднемесячный абсолютный прирост за те же 10 лет – он будет в 12 раз меньше среднегодового прироста.

Среднее ускорение абсолютного изменения применяется реже. Для его надежного расчета даже при слабых колебаниях уровней требуется применять методику аналитического выравнивания по параболе II порядка. Не рекомендуется измерять среднее ускорение без абстрагирования от колебаний уровней. Для более грубого, приближенного расчета среднего ускорения можно воспользоваться средними годовыми уровнями, сглаживающими колебания. Например, национальное богатство России объеме сопоставимых ценах, млрд. руб. составляло:

Период	1941 – 1960	1961 – 1980	1981 – 2000
Национальное богатство России объеме сопоставимых ценах, млрд. руб.	1566	7097	18434

Рассчитано по: **Симчера В.М.** Развитие экономики России за 100 лет: 1900-2000. Исторические ряды, вековые тренды, периодические циклы / В.М. Симчера. – М.: ЗАО «Издательство Экономика»,

2007. – 683с. ISBN 5-282-02627-9

Абсолютный прирост за второй период в сравнении с первым составил 5531 млрд. руб., за третий в сравнении со вторым – 11337 млрд. руб. Следовательно, ускорение в третьем пятилетии по сравнению со вторым составило $11337 - 5531 = 5806$ млрд. руб. в год за двадцать лет, а среднегодовое ускорение прироста равно: $5806 : 20 = 290,28$ млрд. руб. в год за год. Среднее ускорение требует указания трех единиц времени, хотя, как правило, две из них одинаковы: период, на который рассчитан прирост, и время, на которое рассчитано ускорение.

Средний темп изменения определяется наиболее точно при аналитическом выравнивании динамического ряда по экспоненте. Если можно пренебречь колеблемостью, то средний темп определяют как *геометрическую среднюю* из цепных темпов роста за n лет или из общего (базисного) темпа роста за n лет:

$$\bar{k} = \sqrt[n]{\prod_{i=1}^n k_i} = \sqrt[n]{\frac{y_n}{y_0}} \quad (3.11)$$

Например, стоимость потребительской корзины за год в результате инфляции возросла в шесть раз. Каков средний месячный темп инфляции?

$$\bar{k} = \sqrt[12]{6} = 1,16, \text{ или } 116\%,$$

т.е. в среднем за месяц цена увеличивалась на 16% к уровню предыдущего месяца.

Средний темп роста так же, как средний прирост, следует со-

проводить указанием двух единиц времени: периода, который им характеризуется и периода, на который рассчитан темп. Например, среднегодовой темп за последнее десятилетие; среднемесячный темп за полугодие и т.п.

Если исходной информацией служат темпы прироста и нужно вычислить их среднегодовую величину, то предварительно следует все темпы прироста превратить в темпы роста, прибавив 1, или 100%, вычислить их среднюю геометрическую и снова вычесть 1, или 100%. Интересно, что в виду асимметрии темпа прироста и темпа сокращения при равных их величинах общий темп прироста всегда отрицателен. Так, если за первый год объем производства вырос на 20%, а за второй снизился на 20% (темпы цепные), то за два года имеем:

средний темп роста $\bar{k} = \sqrt{1,2 \cdot 0,8} = 0,9798$, или 97,98% ;

средний темп прироста $\bar{k} - 1 = -0,0202$, или -2,02%.

Применяя для вычисления среднего темпа среднюю геометрическую, мы опираемся на соблюдение фактического отношения конечного уровня к начальному при замене фактических темпов на средние. В практических задачах может потребоваться вычисление среднего уровня при условии соблюдения отношения суммы уровней за период к уровню, принятому за базу. Например, если общий выпуск продукции за пятилетие должен составить 800% к базисному (среднегодовому за предыдущие 5 лет выпуску), или, что то же самое, среднегодовой уровень должен составить 160% к базовому уровню, каков должен быть среднегодовой темп роста выпуска продукции? В 1974 г. украинские статистики А. и И.Соляники предложили следующую

приближенную формулу для среднего темпа роста, удовлетворяющую этому условию:

$$\bar{k}_{nap} = 1 + \frac{-3}{2(m-1)} + \sqrt{\frac{9}{4(m-1)^2} + \frac{6}{m(m^2-1)} \left(\frac{\sum_{i=1}^m y_i}{y_0} - m \right)}, \quad (3.12)$$

где m – число суммируемых уровней;

y_0 – базисный уровень.

Темп роста данного вида называют параболическим (отсюда обозначение \bar{k}_{nap}), так как он вычисляется по уравнению параболы порядка m . При $m = 5$ имеем:

$$k_{nap} = 1 - \frac{3}{8} + \sqrt{\frac{9}{64} + \frac{1}{20} \left(\frac{\sum y_i}{y_0} - 5 \right)} = 1 - 0,375 + \sqrt{0,1406 + 0,05(8-5)} = 1,16407$$

или 116,4%

Расчет по этому среднегодовому темпу дает сумму выпуска за 5 лет в 8,069 раза больше базисной, т.е. приближение хорошее. В общем виде проблема параболических темпов исследована саратовским статистиком Л. С. Казинцом⁵. Им составлены таблицы, с помощью которых, зная отношение суммы уровней к базисному уровню и число суммируемых уровней m , можно получить \bar{k}_{nap} . Таблица Л.С. Казинца рассчитана на основе нахождения корней уравнения:

$$\bar{k} + \bar{k}^2 + \bar{k}^3 + \dots + \bar{k}^n = \left(\sum_1^n y_i : y_0 \right)$$

⁵ Казинец Л.С. Темпы роста и абсолютные приросты. – М.: Статистика, 1975.

Для нашего примера таблица Л.С. Казинца дает среднегодовой темп роста 116,1% и сумму выпуска в 8,00016 раза больше базисной.

Интересную задачу представляет определение срока, за который ряд с большим средним показателем динамики, но меньшим начальным уровнем догонит другой ряд с большим начальным уровнем, но меньшим показателем динамики. Для абсолютных приростов задача элементарна: имеем один ряд с базисным уровнем y_0^I и средним абсолютным приростом $\bar{\Delta}^I$; второй ряд с показателями соответственно y_0^{II} , $\bar{\Delta}^{II}$, причем $y_0^{II} > y_0^I$; $\bar{\Delta}^{II} < \bar{\Delta}^I$. Уровень первого ряда сравнивается с уровнем второго ряда через

$$\frac{y_0^{II} - y_0^I}{\bar{\Delta}^I - \bar{\Delta}^{II}} \text{ лет.}$$

Та же задача может быть решена на основе *ускорений*. Имеем первый ряд с базисным уровнем y_0^I , базисным абсолютным изменением a_0^I и средним ускорением \bar{b}^I ; второй ряд — с показателями y_0^{II} , a_0^{II} , b^{II} . При каком числе n периодов (лет) после базисного уровни рядов сравниваются? Тенденции рядов параболические:

$$y_n^I = y_0^I + a_0^I n + \frac{\bar{b}^I n^2}{2};$$

$$y_n^{II} = y_0^{II} + a_0^{II} n + \frac{\bar{b}^{II} n^2}{2}.$$

Приравняв правые части уравнений, получим:

$$y_0'' + a_0''n + \frac{\bar{b}''n^2}{2} = y_0' + a_0'n + \frac{\bar{b}'n^2}{2}$$

или

$$(y_0'' - y_0') + n(a_0'' - a_0') + n^2 \left(\frac{\bar{b}'' - \bar{b}'}{2} \right) = 0 \quad (3.13)$$

Искомый срок n является корнем этого квадратного уравнения.

Если, например, имеем:

$$y_0' = 500; a_0' = +40; \bar{b}' = +2; y_0'' = 300; a_0'' = +26; \bar{b}'' = +3, \text{ то}$$

$$-200 + (-14)n + 0,5n^2 = 0,$$

откуда

$$n = \frac{14 \pm \sqrt{196 + 400}}{2 \cdot 0,5} = 14 \pm 24,4.$$

Второй ряд догонит первый по уровню через 38,4 года; в прошлом уровни рядов были одинаковы 10,4 года назад. Будущие равные уровни составляют 3510, а прошлые были равны 192.

Если мы хотим найти срок n , через который уровни рядов сравняются, то эту задачу можно решить на основе *темпов изменения*.

Имеем:

$$y_0'' \cdot \bar{k}^{nII} = y_0' \bar{k}^{nI}$$

Логарифмируя это равенство, имеем:

$$n \log \bar{k}'' + \log y_0'' = n \log \bar{k}' + \log y_0',$$

откуда

$$n(\log \bar{k}'' - \log \bar{k}') = \log y_0' - \log y_0'',$$

$$n = \frac{\log y_0' - \log y_0''}{\log \bar{k}'' - \log \bar{k}'}$$

т.е. искомый срок равен частному от деления разности логарифмов уровней рядов в базисном периоде на разность логарифмов темпов изменения, только переставленных при вычитании. Обычно и в числителе, и в знаменателе от большего логарифма вычитается меньший. Например, первый ряд имеет $y_0' = 300$; $\bar{k}' = 1,09$; второй ряд — $y_0'' = 100$; $\bar{k}'' = 1,2$, тогда:

$$n = \frac{\ln 300 - \ln 100}{\ln 1,2 - \ln 1,09} = \frac{5,70382 - 4,60517}{0,18232 - 0,08618} = 11,43.$$

Через 11,43 года уровень второго ряда сравнивается с первым при сохранении экспоненциальных трендов обоих рядов.

Глава 4. ОСНОВНЫЕ ТИПЫ ТЕНДЕНЦИЙ И УРАВНЕНИЙ ТRENDA

В гл. 2 рассмотрено понятие о тенденции временного ряда, т.е. тенденции динамики развития изучаемого показателя. Задача данной главы состоит в том, чтобы рассмотреть основные типы таких тенденций, их свойства, отражаемые большей или меньшей степенью полноты уравнением линии тренда. Укажем при этом, что в отличие от простых систем механики тенденции изменения показателей сложных социальных экономических, биологических и технических систем только некоторым приближением отражаются тем или иным уравнением, линией тренда.

В данной главе рассматриваются далеко не все известные математике линии и их уравнения, а лишь набор их сравнительно простых форм, который мы считаем достаточным для отображения и анализа большинства встречающихся на практике тенденций временных рядов. При этом желательно всегда выбирать из нескольких типов линий, достаточно близко выражающих тенденцию, более простую линию. Этот «принцип простоты» обоснован тем, что чем сложнее уравнение линии тренда, чем большее число параметров оно содержит, тем при равной степени приближения труднее дать надежную оценку этих параметров по ограниченному числу уровней ряда и тем больше ошибка оценки этих параметров, ошибки прогнозируемых уровней.

4.1. Прямолинейный тренд и его свойства

Самым простым типом линии тренда является прямая линия, описываемая линейным (т.е. первой степени) уравнением тренда:

$$\hat{y}_i = a + b \cdot t_i,$$

где \hat{y}_i – выравненные, то есть лишенные колебаний, уровни тренда для лет с номером i ;

" a " – свободный член уравнения, численно равный среднему выравненному уровню для момента или периода времени, принятого за начало отсчета, т. е. для $t_i = 0$;

" b " – средняя величина изменения уровней ряда за единицу изменения времени;

t_i – номера моментов или периодов времени, к которым относятся уровни временного ряда (год, квартал, месяц, дата).

Среднее изменение уровней ряда за единицу времени – главный параметр и константа прямолинейного тренда. Следовательно, этот тип тренда подходит для отображения тенденции примерно равномерных изменений уровней: равных в среднем абсолютных приростов или абсолютных сокращений уровней за равные промежутки времени. Практика показывает, что такой характер динамики встречается достаточно часто. Причина близких к равномерному абсолютных изменений уровней ряда состоит в следующем: многие явления, как, например, урожайность сельскохозяйственных культур, численность населения региона, города, сумма дохода населения, среднее потребление какого-либо продовольственного товара и др., зависят от большого числа различных факторов. Одни из этих факторов влияют в сторону ускоренного роста изучаемого явления, другие – в сторону замедленного роста, третьи – в направлении сокращения уровней и т. д. Влияние разнонаправленных и разноускоренных (замедленных) сил факторов взаимно усредняется, частично взаимно погашается, а рав-

нодействующая их влияний приобретает характер, близкий к равномерной тенденции. Итак, равномерная тенденция динамики (или застоя) – это результат сложения влияния большого количества факторов на изменение изучаемого показателя.

Графическое изображение прямолинейного тренда – прямая линия в системе прямоугольных координат с линейным (арифметическим) масштабом на обеих осях. Пример линейного тренда дан на рис. 4.1.

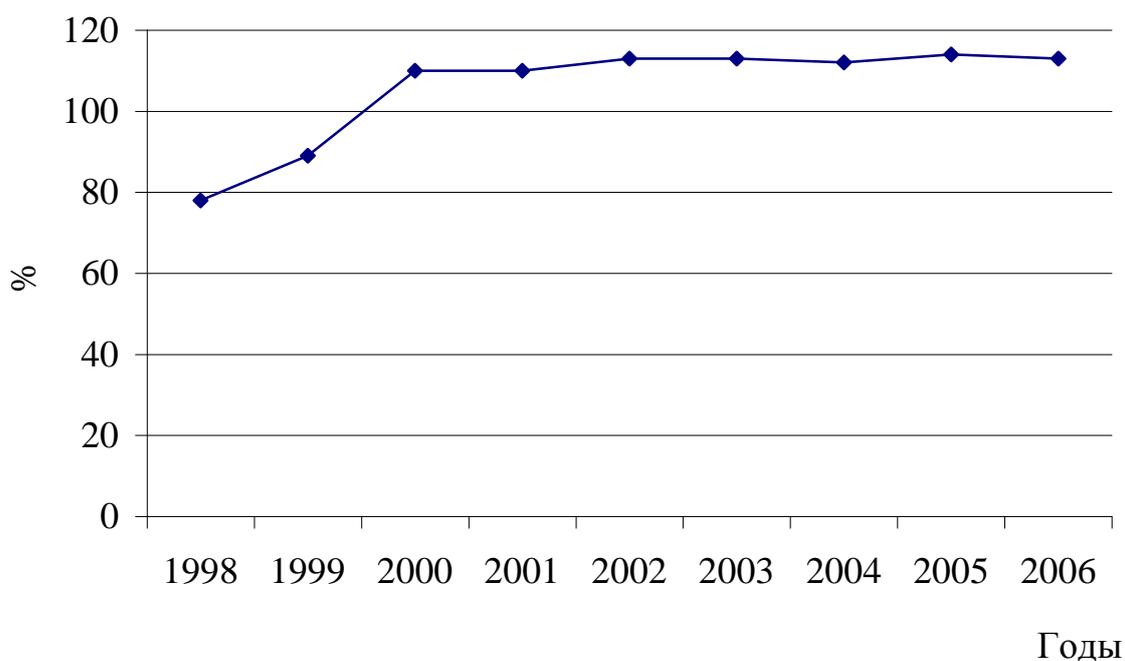


Рис. 4.1. Реальные денежные доходы населения

Оренбургской области, в % к предыдущему году

Абсолютные изменения уровней в разные годы не были точно одинаковыми, но общая тенденция роста реальных денежных доходов населения в Оренбургской области, в процентах к предыдущему году хорошо отражается прямолинейным трендом. Его параметры вычислены в гл. 5 (табл.5.5).

Основные свойства тренда в форме прямой линии таковы:

- равные изменения за равные промежутки времени;
- если средний абсолютный прирост – положительная величина, то относительные приросты или темпы прироста – постепенно уменьшаются;
- если среднее абсолютное изменение – отрицательная величина, то относительные изменения или темпы сокращения постепенно увеличиваются по абсолютной величине снижения к предыдущему уровню;
- если тенденция к сокращению уровней, а изучаемая величина является по определению положительной, то среднее изменение b не может быть больше среднего уровня a ;
- при линейном тренде ускорение, т.е. разность абсолютных изменений за последовательные периоды, равно нулю.

Свойства линейного тренда иллюстрирует табл. 4.1. Уравнение тренда: $\hat{y}_i = 100 + 20t_i$.

Таблица 4.1

**Показатели динамики при линейном тренде
к увеличению уровней: $\hat{y}_i = 100 + 20t_i$**

Номер периода t_i	Уровень \hat{y}_i	Абсолютное изменение к предыдущему периоду	Темпы (цепные), %	Ускорение
1	2	3	4	5
1	120	+20	120,0	—
2	140	+20	116,7	0
3	160	+20	114,3	0

Продолжение таблицы 4.1

1	2	3	4	5
4	180	+20	112,5	0
5	200	+20	111,1	0
6	220	+20	110,0	0

Показатели динамики при наличии тенденции сокращения уровней приведены в табл. 4.2.

Таблица 4.2

**Показатели динамики при линейном тренде
сокращения уровней: $\hat{y}_i = 100 - 20t_i$**

Номер периода t_i	Уровень \hat{y}_i	Абсолютное изменение к предыдущему периоду	Темп к предыдущему периоду, %	Ускорение
1	180	-20	90,0	—
2	160	-20	88,9	0
3	140	-20	87,5	0
4	120	-20	85,7	0
5	100	-20	83,3	0
6	80	-20	80,0	0

4.2. Параболический тренд и его свойства

Под названием параболического будем иметь в виду тренд, выраженный параболой II порядка с уравнением

$$\hat{y}_i = a + b \cdot t + c \cdot t^2$$

Параболы III порядка и более высоких порядков редко применимы для выражения тенденции динамики и слишком сложны для получения надежных оценок параметров при ограниченной длине временного ряда. Прямую линию, с точки зрения математики, можно также считать видом парабол – параболой I порядка, которая уже рассмотрена ранее.

Значения (смысл, сущность) параметров параболы II порядка таковы: свободный член a – это средний (выравненный) уровень тренда на момент или период, принятый за начало отсчета времени, т.е. $t = 0$; b – это средний за весь период среднегодовой прирост, который уже не является константой, а изменяется равномерно со средним ускорением, равным 2 с, которое и служит константой, главным параметром параболы II порядка.

Следовательно, тренд в форме параболы II порядка применяется для отображения таких тенденций динамики, которым свойственно примерно постоянное ускорение абсолютных изменений уровней. Процессы такого рода встречаются на практике гораздо реже, чем процессы с равномерным изменением, но с другой стороны, любое отклонение процесса от строго равномерного прироста (или сокращения) уровней можно интерпретировать, как наличие ускорения. Более того, существует строгое математическое правило: чем выше порядок параболы, тем ближе линия тренда к уровням исходного временного ряда. Если это правило довести до крайнего предела, то любой ряд из n уровней может быть точно отображен параболой $(n-1)$ -ого порядка! (Через любые две точки проходит одна прямая, через три точки – одна парабола II порядка и т.д.). Такое «приближение» линии тренда к эм-

пирическому ряду, содержащему как тенденцию, так и колебания, нельзя считать достижением научного анализа. Напротив, применяя параболу более высокого порядка там, где сущность процесса этого не требует, а только ради уменьшения остаточной суммы отклонений (или их квадратов) отдельных уровней от тренда, исследователь уходит от цели, смешивая тренд с колебаниями.

Парабола II порядка, как уравнение тренда, применяется к различным процессам, которые на некотором, как правило, довольно недолгом этапе развития, имеют примерно постоянное ускорение абсолютного прироста уровней. Таким бывает рост населения отдельных стран или регионов приведенный на рис. 4.2.

Расчет уравнения этой параболы представлен в гл. 5.

Основные свойства тренда в форме параболы II порядка таковы:

1) неравные, но равномерно возрастающие или равномерно убывающие абсолютные изменения за равные промежутки времени;

2) парабола, рассматриваемая относительно ее математической формы, имеет две ветви: восходящую с увеличением уровней признака, и нисходящую с их уменьшением. Но относительно статистики по содержанию изучаемого процесса изменений трендом, выражающим определенную тенденцию развития, чаще всего можно считать только одну из ветвей: либо восходящую, либо нисходящую. В особых, более конкретных ситуациях мы не отрицаем возможности объединения обеих ветвей в единый тренд;

3) так как свободный член уравнения a как значение показателя в начальный момент (период) отсчета времени, как правило, величина положительная, то характер тренда, тенденции, определяется знаками параметров b и c :

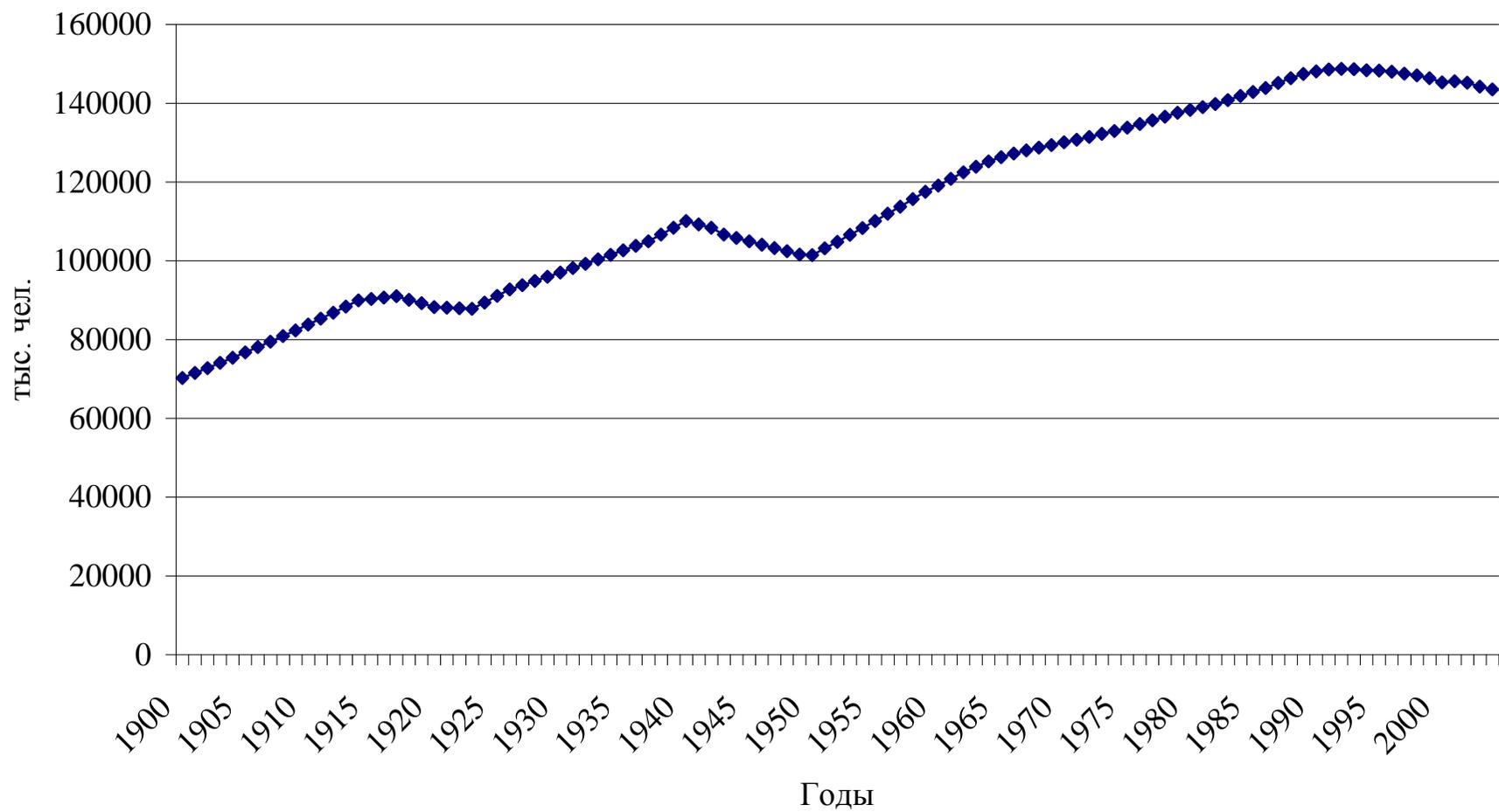


Рис. 4.2. Динамика численность населения РФ, тыс. человек

а) при $b > 0$ и $c > 0$ имеем восходящую ветвь, т.е. тенденцию к ускоренному росту уровней;

б) при $b < 0$ и $c < 0$ имеем нисходящую ветвь – тенденцию к ускоренному сокращению уровней;

в) при $b > 0$ и $c < 0$ имеем либо восходящую ветвь с замедляющимся ростом уровней, либо обе ветви параболы, восходящую и нисходящую, если их по существу можно считать единым процессом.

г) при $b < 0$ и $c > 0$ имеем либо нисходящую ветвь с замедляющимся сокращением уровней, либо обе ветви – нисходящую и восходящую, если их можно считать единой тенденцией.

4) при параболической форме тренда, в зависимости от соотношений между его параметрами, цепные темпы изменений могут либо уменьшаться, либо некоторое время возрастать, но при достаточно длительном периоде времени рано или поздно темпы роста всегда обязательно начинают уменьшаться, а темпы сокращения уровней при $b < 0$ и $c < 0$ обязательно начинают возрастать (по абсолютной величине относительного изменения).

Таблица 4.3

Показатели динамики при параболическом тренде,

когда $b > 0, c > 0$: $\hat{y}_i = 100 + 20t + 2t^2$

Номер периода t_i	Уровень \hat{y}_i	Абсолютное изменение	Цепные темпы, % к предыдущему периоду	Ускорение
1	2	3	4	5
1	122	+22	122,0	—

Продолжение таблицы 4.3

1	2	3	4	5
2	148	+26	121,3	+4
3	178	+30	120,3	+4
4	212	+34	119,1	+4
5	250	+38	117,9	+4
6	292	+42	116,8	+4

Ввиду ограниченного объема учебника рассмотрим не все четыре случая параболических трендов, а лишь два первых (табл. 4.3 и 4.4).

Таблица 4.4

Показатели динамики при параболическом тренде,

когда $b < 0, c < 0$: $\hat{y}_i = 100 - 20t - 2t^2$

Номер периода t_i	Уровень \hat{y}_i	Абсолютные изменения	Цепные темпы, % к предыдущему периоду	Ускорение
1	-22	89,0	-11,0	—
2	-26	85,4	-14,6	-4
3	-30	80,3	-19,7	-4
4	-34	72,1	-27,9	-4
5	-38	56,8	-43,2	-4
6	-42	16,0	-84,0	-4

В тех случаях, когда по существу изучаемого процесса допустимо считать единым трендом обе ветви параболы, представляет

большой интерес решение задачи о нахождении того периода или момента времени, когда уровень тренда достигает максимума (когда $b > 0$, $c < 0$) или минимума (если $b < 0$, $c > 0$). Экстремальная точка параболы $\hat{y} = a + bt + ct^2$ достигается при нулевом значении первой производной:

$$\frac{df}{dt} = (a + bt + ct^2)' = b + 2ct$$

Из равенства: $b + 2ct = 0$, имеем: $t = \frac{-b}{2c}$.

Например, если $\hat{y} = 100 + 20t - 2t^2$, то максимум параболы имеет при $t = \frac{-20}{2(-2)} = 5$.

Максимальное значение уровня тренда при $t = 5$ составит:

$$\hat{y}_{\max} = 100 + 20 \cdot 5 - 2 \cdot 5^2 = 150.$$

Если имеем параболу при $b < 0$, а $c > 0$, например: $\hat{y}_i = 200 - 20t + 2t^2$, то минимальное значение тренда достигается при $t = \frac{-b}{2c} = \frac{20}{2 \cdot 2} = 5$, и это минимальное значение составит:

$$\hat{y}_{\min} = 100 - 20 \cdot 5 + 2 \cdot 5^2 = 50.$$

4.3. Экспоненциальный тренд и его свойства

Экспоненциальным трендом называют тренд, выраженный уравнением:

$$\hat{y}_i = a \cdot k^t$$

или в форме

$$\hat{y}_i = \exp[\ln a + \ln k \cdot t_i].$$

Свободный член экспоненты a равен выравненному уровню, т.е. уровню тренда в момент или период, принятый за начало отсчета времени, т.е. при $t=0$. Основным параметром экспоненциального тренда k является постоянным темпом изменения уровней (цепным). Если $k>1$, имеем тренд с возрастающими уровнями, причем это возрастание не просто ускоренное, а с возрастающим ускорением и возрастающими производными всех более высоких порядков. Если $k<1$, то имеем тренд, выражающий тенденцию постоянного, но замедляющегося сокращения уровней, причем замедление непрерывно усиливается. Экстремума экспонента не имеет, и при $t \rightarrow \infty$ стремится либо к ∞ при $k>1$, либо к 0 при $k<1$.

Экспоненциальный тренд характерен для процессов, развивающихся в среде, не создающей никаких ограничений для роста уровня. Из этого следует, что на практике он может развиваться только на ограниченном промежутке времени, так как любая среда рано или поздно создает ограничения, любые ресурсы со временем исчерпаемы. Однако, практика показала, что, например, численность населения Земли на протяжении 1950–1985 гг. возросла примерно по экспоненте, со среднегодовым темпом роста $k \approx 1,018$ и за это время возросла вдвое – с 2,5 до 5 миллиардов человек. В настоящее время темп роста населения постепенно уменьшается.

На рисунке 4.3 представлена динамика объема национального богатства РФ в сопоставимых ценах, которая хорошо аппроксимируется экспоненциальным трендом.

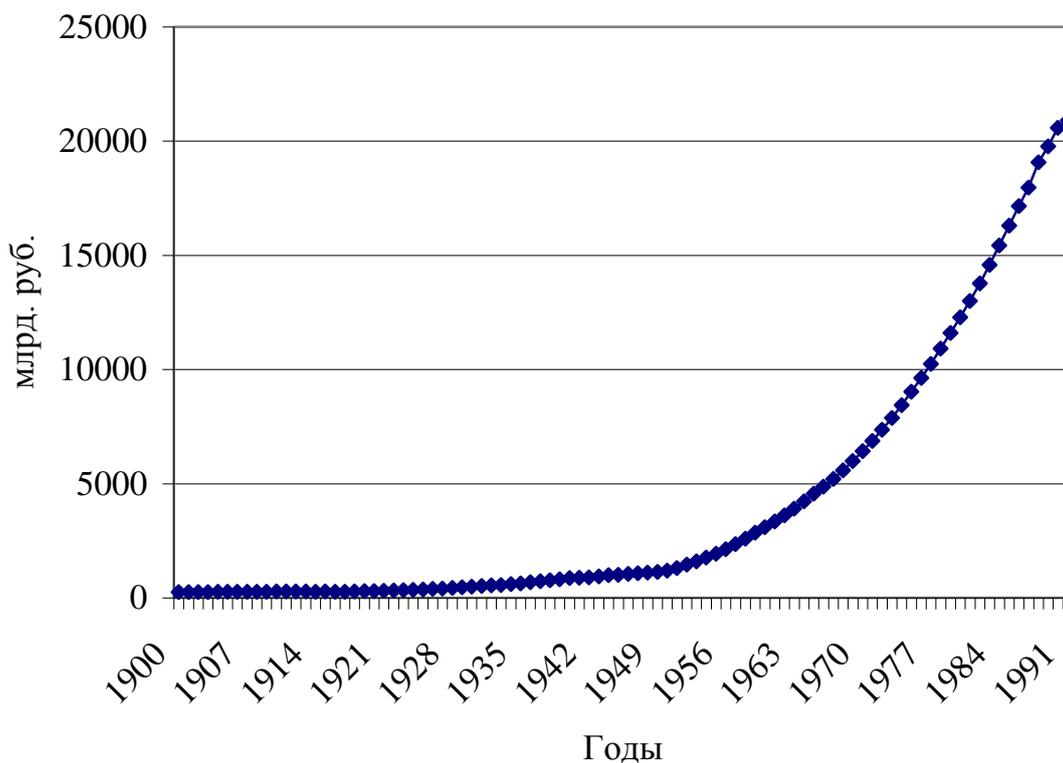


Рис. 4.3. Национальное богатство РФ в сопоставимых ценах

Экспоненциальный рост объема реализации и производства происходит при возникновении новых видов продукции и их освоения промышленностью: при появлении цветных телевизоров, видеомагнитофонов, пейджеров и т.п., но когда производство начинает наполнять рынок, приближаться к спросу, экспоненциальный рост прекращается.

Расчет экспоненциального тренда дан в гл. 5.

Основные свойства экспоненциального тренда:

1. Абсолютные изменения уровней тренда пропорциональны самим уровням.
2. Экспонента экстремумов не имеет: при $k > 1$ тренд стремится к $+\infty$, при $k < 1$ тренд стремится к нулю.
3. Уровни тренда представляют собой геометрическую прогрессию: уровень периода с номером $t = m$ есть $a \cdot k^m$.

4. При $k > 1$ тренд отражает ускоряющийся неравномерно рост уровней, при $k < 1$ тренд отражает замедляющееся неравномерно уменьшение уровней. Поведение основных показателей динамики в этих случаях рассмотрено в табл. 4.5 и 4.6.

Таблица 4.5

Экспоненциальный тренд при $k > 1$: $\hat{y}_i = 100 \cdot 1,2^i$

Номер периода t_i	Уровень \hat{y}_i	Абсолютные изменения (цепные)	Цепные темпы, % к предыдущему периоду	Ускорение	Прирост ускорения к предыдущему периоду
1	120,00	+20,00	120	-	-
2	144,00	+24,00	120	+4,00	-
3	172,80	+28,80	120	+4,80	+0,80
4	207,36	+34,56	120	+4,76	+0,96
5	248,83	+41,47	120	+6,81	+1,15
6	298,60	+49,77	120	+8,30	+ 1,39

В табл. 4.5 и 4.6 в последней графе приведены редко применяемые показатели динамики III порядка: ускорение (или прирост) ускорения и замедление ускорения. Эти абсолютные показатели даны для наглядного пояснения главного отличия экспоненциального тренда от парабол любого порядка: экспонента не имеет постоянных производных любого порядка по времени. Постоянен только цепной темп изменения.

Таблица 4.6

Экспоненциальный тренд при $k < 1$: $\hat{y}_i = 200 \cdot 0,8^i$

Номер периода t_i	Уровень \hat{y}_i	Абсолютные изменения (цепные)	Цепные темпы, % к предыдущему периоду	Ускорение	Замедление ускорения
1	160,00	-40,00	80	-	-
2	128,00	-32,00	80	+8,00	-
3	102,40	-25,60	80	+6,40	-1,60
4	81,92	-20,48	80	+5,12	-1,28
5	65,54	-16,38	80	+4,10	-1,02
6	52,43	-13,11	80	+3,27	-0,83

Читатель может заинтересоваться и таким вопросом: как назвать такую тенденцию динамики, при которой и темп изменения был бы не постоянен, а имел постоянное абсолютное или относительное изменение, например, уравнение типа: $\hat{y} = a(k + bt_i)^i$ или $\hat{y} = ak^{t^2}$ и т.д. Подобные «гиперэкспоненты» не применяются статистикой, ибо любой, сколь угодно быстрый, сколь угодно ускоряющийся рост может быть отображен обычной экспонентой, – стоит лишь уменьшить период времени, за который происходит возрастание (или сокращение) уровней в k раз. По своему существу экспоненциальное развитие процесса и есть предельно возможное, предельно благоприятное по условиям развития, так как оно осуществляется в среде, не ограничивающей развитие данного процесса. Но, следует помнить, только до поры до времени, так как каждая среда, каждый ресурс в природе ограничен. Единственный спорный в науке процесс, по которому до сих пор нет доказательства ограниченности его во времени, – это экспоненциаль-

ное замедляющееся расширение Вселенной. Ограничено ли оно и сменится ли со временем сжатием, или будет продолжаться бесконечно, зависит от значения средней плотности вещества и излучения во Вселенной, которую пока науке установить не удалось, ибо не все формы существования вещества и полей науке известны. Зато интересно знать, что самый фундаментальный процесс, охватывающий всю известную Вселенную, уже по крайней мере 12-15 миллиардов лет развивается по экспоненте.

4.4. Гиперболический тренд и его свойства

Из различных форм гипербол рассмотрим только наиболее простую: $\hat{y} = a + \frac{b}{t}$.

Если основной параметр гиперболы $b > 0$, то этот тренд выражает тенденцию замедляющегося снижения уровней и при $t \rightarrow \infty$ $\hat{y} \rightarrow a$. Таким образом, свободный член гиперболы – это предел, к которому стремится уровень тренда.

Такая тенденция наблюдается, например, при изучении процесса снижения затрат любого ресурса (труда, материалов, энергии) на единицу данного вида продукции, или ее себестоимости в целом. Затраты ресурса не могут стремиться к нулю, значит, экспонента не соответствует сущности процесса; нужно применить гиперболическую формулу тренда.

На рисунке 4.4 представлена динамика численности прибывших на территорию Оренбургской области. Как видно на графике их численность снижается в анализируемом периоде, однако данная кривая никогда не пересечет ось абсцисс. Следовательно для описания дина-

мики такого процесса целесообразно применить гиперболическую кривую.

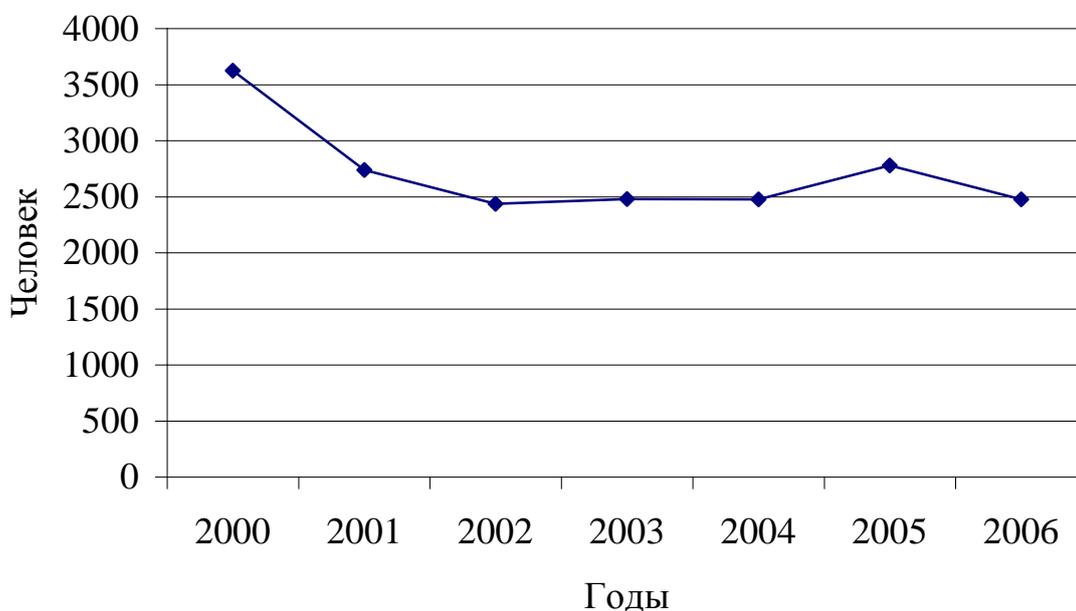


Рис. 4.4. Динамика числа прибывших в Оренбургскую область

Если параметр $b < 0$, то с возрастанием t , то есть с течением времени, уровни тренда возрастают и стремятся к величине a при $t \rightarrow \infty$.

Такой характер динамики присущ, например, показателям КПД двигателей или иных преобразователей энергии (трансформатор тока, фотоэлемент и т.п.). По мере научно-технического прогресса эти КПД постепенно повышаются, но никогда не могут превысить определенного предела для каждого типа двигателя и не могут превысить 100% в принципе для любого преобразователя энергии. При расчете гиперболического тренда нельзя нумеровать года от середины ряда, так как значения $\frac{1}{t_i}$ должны быть всегда положительными.

Основные свойства гиперболического тренда:

1. Абсолютный прирост или сокращение уровней, ускорение абсолютных изменений, темп изменения – все эти показатели не являются постоянными. При $b > 0$ уровни замедленно уменьшаются, отрицательные абсолютные изменения, а также положительные ускорения тоже уменьшаются, цепные темпы изменения растут и стремятся к 100%.

2. При $b < 0$ уровни замедленно возрастают, положительные абсолютные изменения, а также отрицательные ускорения и цепные темпы роста замедленно уменьшаются, стремясь к 100%.

Как видим, гиперболический тренд описывает в любом случае тенденцию такого процесса, показатели которого со временем затухают, т.е. происходит переход от движения к застою. Иллюстрацией этих свойств может служить табл. 4.7.

Таблица 4.7

Показатели динамики при гиперболическом тренде: $\hat{y}_i = 100 + \frac{100}{t_i}$

Номер периода t_i	Уровень \hat{y}_i	Абсолютные изменения (цепные)	Цепные темпы, % к предыдущему периоду	Ускорение
1	200,0	-	-	-
2	150,0	-50,0	75,0	-
3	133,0	-16,7	88,9	+33,3
4	125,0	-8,3	93,8	+8,4
5	120,0	-5,0	96,0	+3,3
6	116,7	-3,3	97,2	+1,7

4.5. Логарифмический тренд и его свойства

Если изучаемый процесс приводит к замедлению роста какого-то показателя, но при этом рост не прекращается, не стремится к какому-либо ограниченному пределу, то гиперболическая форма тренда уже не подходит. Тем более не подходит парабола с отрицательным ускорением, по которой замедляющийся рост перейдет со временем в снижение уровней. В указанном случае тенденция изменения лучше всего отображается логарифмической формой тренда: $\hat{y}_i = a + b \ln t_i$

Логарифмы возрастают значительно медленнее, чем сами числа (номера периодов t_i), но рост логарифмов неограничен. Подбирая начало отсчета периодов (моментов) времени, можно найти такую скорость снижения абсолютных изменений, которая наилучшим образом отвечает фактическому временному ряду.

Примером тенденций, соответствующих логарифмическому тренду, может служить динамика рекордных достижений в спорте: известно, что увеличение на 1 см рекорда прыжка в высоту или снижение на 0,1 с времени бега на 200 или 400 м требует все больших и больших затрат времени, каждый рекорд дается все большим и большим трудом. В то же время, нет и «вечных» рекордов, все спортивные достижения улучшаются, но медленнее и медленнее, то есть по логарифмическому тренду. Нередко такой же характер динамики присущ на отдельных этапах развития динамике урожайности или валового сбора какой-то культуры в данном регионе, пока новое агротехническое достижение не придаст снова тенденции ускорения. Тенденцию, близкую к логарифмической имеет динамика естественной убыли населения в Оренбургской области за 1985 – 2006гг. (рис. 4.5).

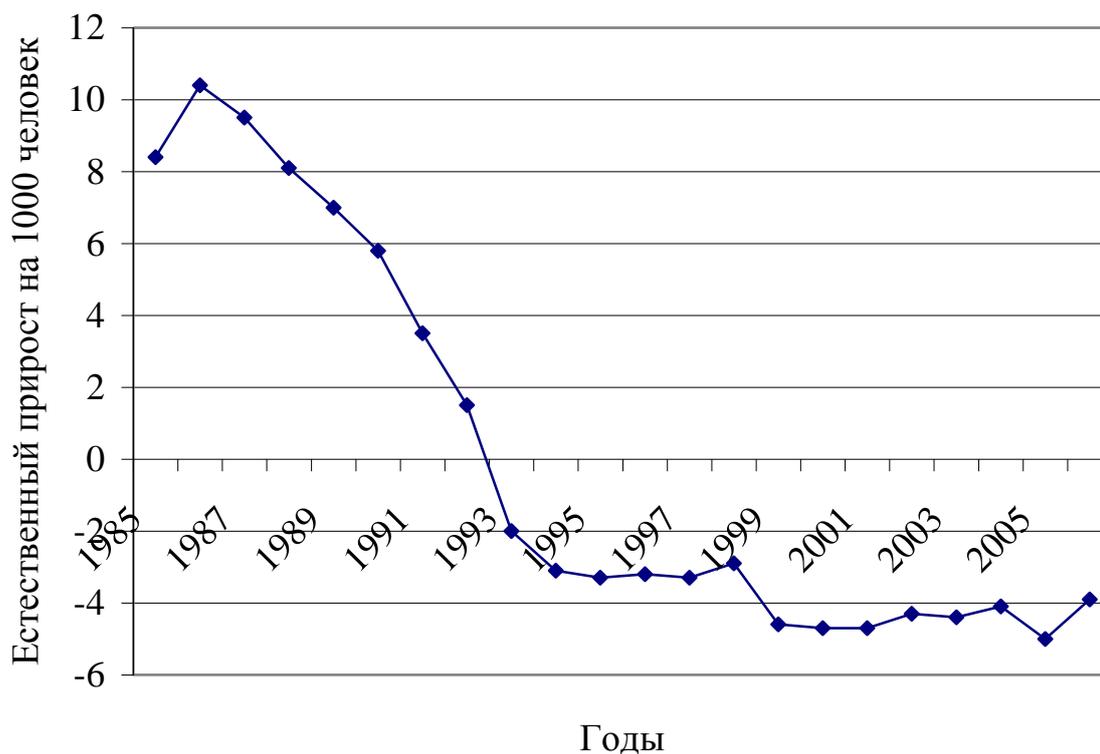


Рис. 4.5. Динамика естественного прироста (убыли (-)) в Оренбургской области

Конечно, характер тенденции маскируется колебаниями, но видно, что естественная убыль населения замедляется. Это показывают и средние уровни:

- за 1985-1991 гг. средний естественный прирост равен 7,53 ‰;
- за 1992-1998 гг. средняя естественная убыль составила 2,33 ‰ (снижение на 9,857 ‰);
- за 1999-2006 гг. средняя естественная убыль составила 4,46 ‰ (снижение на 2,329 ‰);

Основные свойства логарифмического тренда:

1. Если $b > 0$, то уровни возрастают, но с замедлением, а если $b < 0$, то уровни тренда уменьшаются, тоже с замедлением.

2. Абсолютные изменения уровней по модулю всегда уменьшаются со временем.

3. Ускорения абсолютных изменений имеют знак, противоположный самим абсолютным изменениям, а по модулю постепенно уменьшаются.

4. Темпы изменения (цепные) постепенно приближаются к 100 % при $t \rightarrow \infty$.

Можно сделать общий вывод о том, что логарифмический тренд отражает так же, как и гиперболический тренд, постепенно затухающий процесс изменений. Различие в том, что затухание по гиперболе происходит быстро при приближении к конечному пределу, а при логарифмическом тренде затухающий процесс продолжается без ограничения гораздо медленнее.

4.6. Логистический тренд и его свойства

Логистическая форма тренда подходит для описания такого процесса, при котором изучаемый показатель проходит полный цикл развития, начиная, как правило, от нулевого уровня, сначала медленно, но с ускорением возрастая, затем ускорение становится нулевым в середине цикла, т.е. рост происходит по линейному тренду, затем, в завершающей части цикла рост замедляется по гиперболе по мере приближения к предельному значению показателя.

Примером такого цикла динамики может служить изменение доли грамотных в населении страны, например России, с 1800 года до наших дней, или изменение доли семей, имеющих телевизоры, примерно с 1945 года до 2000 года в России, доли жилищ в городах, имеющих горячее водоснабжение или центральное отопление (процесс,

еще не законченный). В некоторых зарубежных программах для компьютеров логистическая кривая называется *S-образной кривой*.

Можно, конечно, логистическую тенденцию считать объединением трех разных по типу: параболической с ускоряющимся ростом на первом этапе, линейной – на втором, и гиперболической с замедляющимся ростом – на третьем. Но есть доводы и в пользу рассмотрения всего цикла развития, как особого единого типа тенденции со сложными, переменными свойствами, но постоянным направлением изменений – увеличению уровней в рассмотренных нами примерах, или уменьшению уровней, если взять противоположный процесс: сокращение доли неграмотных среди населения, доли жилищ, необорудованных газоснабжением или центральным отоплением и т.д.

Рассмотрение таких временных рядов, как проявлений единой логистической тенденции, позволяет уже на первом этапе рассчитать всю траекторию развития, определить сроки перехода от ускоренного роста к замедленному, что чрезвычайно важно при планировании производства или реализации нового вида товара, спрос на который будет проходить все этапы логистической тенденции вплоть до насыщения рынка. Так, например, обеспеченность населения России автомобилями в конце 1980-х лет находилась на начальном этапе логистической кривой, и это означало, что предстоит еще ряд лет, или даже десятилетий ускоренного роста спроса. В то же время обеспеченность фотоаппаратами уже достигла этапа замедления роста, и это означало, что расширять производство или импорт прежних типов фотоаппаратов не следует. Расширение их рынка возможно только для принципиально новых типов фотоаппаратов, насыщенность которыми еще находится в самом начале первого этапа.

В вышеописанном диапазоне изменения уровней, то есть от ну-

ля до единицы, уравнение логистического тренда имеет вид:

$$\hat{y}_i = \frac{1}{e^{a_0 + a_1 t_i} + 1}$$

При $a_0 > 0$, $a_1 < 0$ с ростом номеров периодов времени t_i получаем логистическую тенденцию роста уровней, причем если нужно начать рост почти от нулевой величины, то a_0 должно быть примерно равно 10, тогда при $t = 1$ $\frac{1}{e^9 + 1} = 0,000123$. Чем больше модуль a_1 , тем быстрее возрастание уровней. При $a_0 < 0$ и $a_1 > 0$ имеем логистический тренд со снижением уровней, причем если снижение должно начаться почти от единицы, то a_0 должно быть примерно равно -10. Чем больше a_1 , тем быстрее будут снижаться уровни, например, при $a_0 = -10$; $a_1 = 1$, уже при $t_i = 20$ уровни снизятся почти до нуля.

Если же диапазон изменения уровней ограничен не нулем и единицей, а любыми значениями, определяемыми, исходя из существа задачи, обозначаемыми y_{\max} и y_{\min} , то формула логистического тренда принимает вид:

$$\hat{y}_i = \frac{y_{\max} - y_{\min}}{e^{a_0 + a_1 t_i} + 1} + y_{\min}$$

Как видно из табл. 4.8, абсолютные изменения нарастают до середины периода, затем уменьшаются. Все они положительны. Ускорения сначала возрастают, а после середины периода снижаются, становятся отрицательными, но уменьшаются по модулю. Сумма положительных и отрицательных ускорений приблизительно равна нулю (если ряд продлить от $-\infty$ до $+\infty$, то сумма их точно равна нулю). Темпы роста возрастают до конца первой половины ряда, затем снижаются. Если ряд достаточно длинный, то темпы начинаются со 100 % и завершаются на 100 %.

Таблица 4.8

Показатели динамики при логистическом тренде:

$$\hat{y}_i = \frac{200 - 50}{e^{5-1,5t} + 1} + 50$$

Номер периода t_i	Уровень \hat{y}_i	Абсолютные изменения к предыдущему периоду	Ускорение	Темпы роста к предыду- щему перио- ду, %
0	51,0	-	-	-
1	54,4	+3,4	-	106,7
2	67,9	+13,5	+ 10,1	124,8
3	106,6	+38,7	+25,2	157,0
4	159,7	+53,1	+ 14,4	149,8
5	188,6	+28,8	-24,2	118,1
6	197,3	+8,7	-20,2	104,6
7	199,4	+2,1	-6,6	101,1

При логистическом тренде со снижающимися уровнями, показатели динамики изменяются в следующем порядке: отрицательные абсолютные изменения по модулю возрастают до середины ряда и снижаются к концу, стремясь к нулю при $t = \infty$. Ускорения в первой половине периода отрицательные и по модулю возрастающие; во второй половине периода ускорения положительные и уменьшающиеся, в пределе до нуля. Темпы изменений все меньше 100%, в конце первой половины периода наименьшие, во второй половине возрастающие с замедлением до 100% в пределе. Графическое изображение логистического тренда приведено на рис. 5.2.

Глава 5. МЕТОДЫ РАСПОЗНАВАНИЯ ТИПА ТРЕНДА И ОЦЕНКИ ЕГО ПАРАМЕТРОВ

При изучении методов распознавания типа тренда не следует забывать о существовании изучаемого процесса, который отображается временным рядом. Как правило, тип тренда должен соответствовать характерным особенностям процесса. В гл. 4 для каждого типа тренда приведены примеры выражаемых этим типом процессов. Определяя другие процессы по временным рядам, полезно по указанным примерам подобрать подходящие типы тренда.

Но жизнь, практика всегда гораздо богаче, разнообразнее любых гипотез и теорий: фактические временные ряды, особенно относящиеся к отдельным предприятиям, малоинерционным системам или к ограниченным отрезкам времени, могут и не соответствовать тем аналогам по существу процесса, которые приведены в предыдущей главе. Кроме того, характер тенденции часто маскируется значительной колеблемостью уровней ряда, поэтому требуется специальная методика распознавания типа тренда, наилучшим образом отражающего тенденцию фактического ряда уровней, чему и посвящена эта глава. После определения типа тренда необходимо вычислить оценки его параметров, как правило, по методу наименьших квадратов, а также с использованием специфических приемов для логистического типа тренда.

5.1. Применение графического изображения для распознавания типа тенденции

Графическое изображение во многих случаях позволяет прини-

женно выявить тип тенденции временного ряда. Но для этого следует соблюдать правила построения графика: точное соблюдение масштаба как по величине уровней ряда, так и по времени. Временные интервалы откладывают по оси абсцисс, величины уровней – по оси ординат. По каждой оси следует установить такой масштаб, чтобы ширина графика была примерно в 1,5 раза больше его высоты. Если уровни ряда на всем протяжении периода много больше нуля и между собой различаются не более чем на 20–30%, то следует обозначить перерыв на оси ординат, увеличить масштаб так, чтобы меньший из уровней ненамного превышал разрыв оси. Если уровни ряда различаются в десятки, сотни и тысячи раз, ось ординат следует разметить в логарифмическом масштабе, чтобы равные отрезки означали различие уровней в одинаковое число раз. Интерпретация вида графика будет другой: при линейном масштабе график, близкий к прямой линии, означает линейную тенденцию, а при логарифмическом масштабе оси ординат прямая линия показывает экспоненциальную тенденцию. Необходимо строго соблюдать равенство промежутков времени на равных отрезках оси абсцисс. Логарифмический масштаб по времени не рекомендуется, так как он крайне затруднит интерпретацию графика. Рассмотрим пример графического изображения, представленный на рис. 5.1.

Видно, что линейный тренд хорошо подходит для отражения тенденции динамики доходов бюджета Оренбургской области: прямая проходит как бы посреди колеблющихся точек – уровней месяцев.

Но не всегда график позволяет выбрать тип линии тренда. Трудно графически отличить параболу от экспоненты, логарифмическую кривую от гиперболы и т.д. Оценка типа тренда по типу графика включает субъективные моменты, что может привести к ошибке. Есть

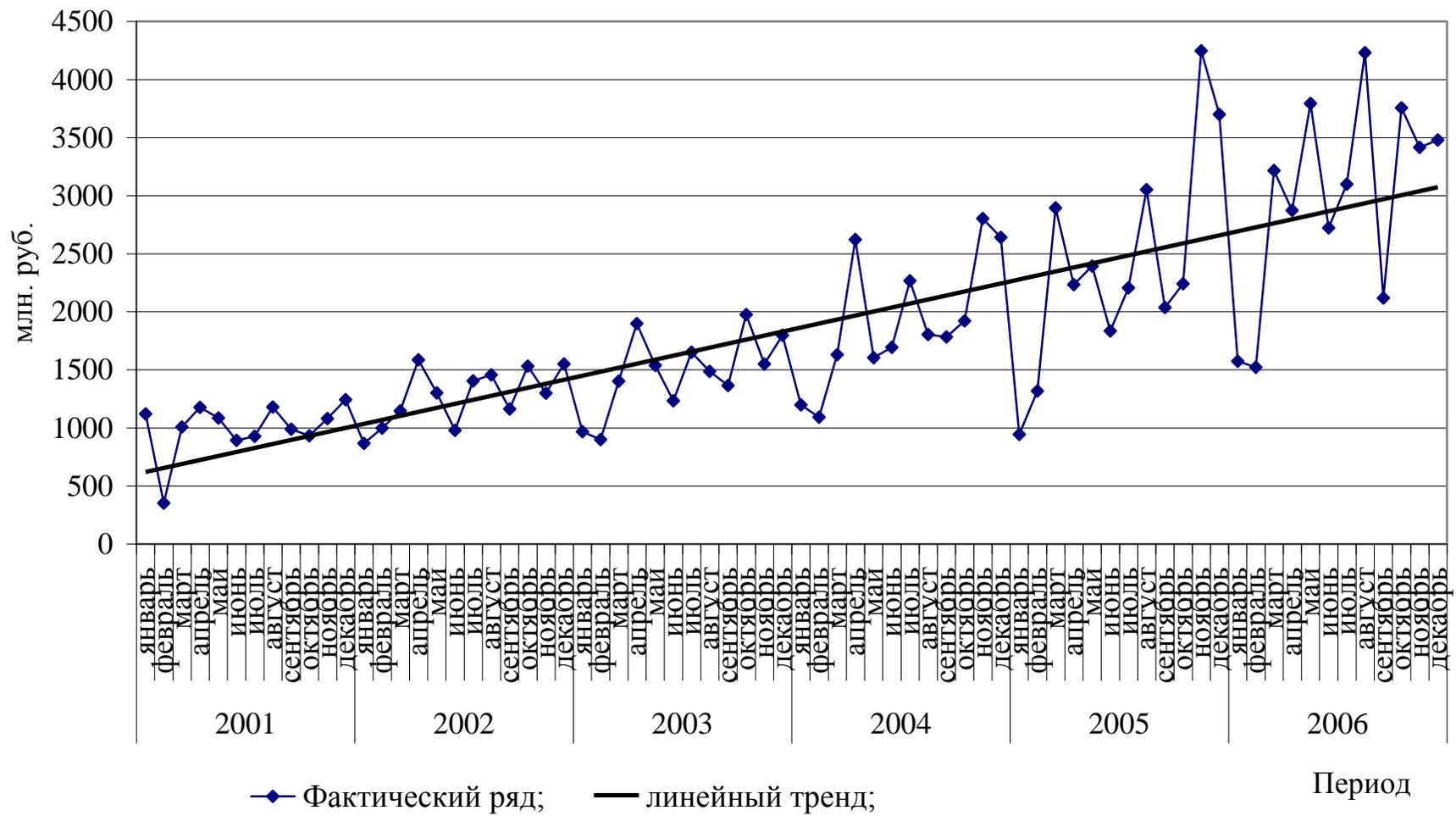


Рис. 5.1. Динамика доходов бюджета Оренбургской области

много способов объективной, статистико-математической оценки пригодности того или иного типа линии. Весьма популярен его выбор с помощью перебора на ПК всех имеющихся в пакете программ статистического анализа типов линий либо по наименьшему среднему квадратическому отклонению, либо по наименьшему модулю отклонений фактических уровней от расчетных по проверяемой линии. Недостатки данной методики заключаются в том, что, во-первых, не все пакеты программ статистического анализа содержат достаточный выбор линий тренда, но главное состоит в том, что, как уже указано в гл. 4, чем больше параметров содержит уравнение тренда, тем меньше и отклонений отдельных уровней от тренда. Парабола II порядка, а тем более III и более высоких порядков всегда при таком подходе «лучше», чем прямая или экспонента.

Но «преимущество» параболы над прямой может быть невелико. Следовательно, нужно применить опять же статистико-математические критерии существенности уменьшения среднего отклонения при переходе от прямой к параболе.

Не отрицая допустимости указанной методики с дополнительной проверкой существенности снижения среднего отклонения от тренда, рассмотрим и другие методы выбора типа тренда без вычисления последнего.

5.2. Методика проверки статистических гипотез о типе тренда

Предположим, что предварительная гипотеза о типе тренда выбрана на основе теоретических соображений об изучаемом процессе и на основе графического изображения. Для того чтобы проверить данную гипотезу, необходимо сформулировать ее математически. Так,

гипотеза о том, что тренд является прямой линией, означает, что на всем периоде временной ряд в среднем сохраняет постоянную величину абсолютного изменения уровней. Гипотеза о параболе II порядка означает, что на всем периоде (в среднем) имеется постоянная величина ускорения абсолютных изменений. Гипотеза об экспоненциальном тренде подтвердится, если можно будет доказать, что на периоде сохраняется постоянная величина (в среднем) цепного темпа изменений.

Для указанных трех типов линий предлагается следующая методика статистической проверки гипотез, разработанная М.С. Каяйкиной и А.И. Манеллей:

1) чтобы снизить искажающее тренд влияние колебаний, проводится сглаживание ряда уровней, например, по пятилетней скользящей средней;

2) по ряду сглаженных уровней вычисляются:

– для линейного тренда - цепные абсолютные изменения:

$$\Delta_i = y_{i+1} - y_i;$$

– для параболического тренда – абсолютные ускорения:

$$\Delta'_i = \Delta_i - \Delta_{i-1};$$

– для экспоненциального тренда – цепные коэффициенты роста:

$$k_n = \frac{y_n}{y_{n-1}}.$$

3) ряд разбивается на несколько равных или примерно равных подпериодов, и по каждому вычисляется средняя величина того параметра, постоянство которого подтверждает выдвинутую гипотезу о типе тренда: средний абсолютный прирост – для прямой, среднее

ускорение – для параболы, средний темп – для экспоненты (см. разд. 3.3);

4) методом дисперсионного анализа при многих средних значениях проверяемого параметра или по t -критерию при двух значениях проверяется существенность различия средних значений параметра в разных подпериодах исходного ряда. Если нельзя отклонить гипотезу о несущественности различий средних величин параметра в разных подпериодах, то принимается гипотеза о соответствующем типе тренда. Если различия средних признаются существенными, гипотеза о данном типе тренда отвергается и выдвигается следующая гипотеза в порядке усложнения: после отклонения прямой линии – об экспоненте; после отклонения экспоненты – о параболе; при отклонении параболы – о других типах линий.

Рассмотрим применение данной методики на примере динамики доходов бюджета Оренбургской области. На основании графика, представленного на рис. 5.1, предложена гипотеза о линейном тренде (табл. 5.1).

Далее проводится дисперсионный анализ различий между средними абсолютными изменениями, результаты которого представлены в табл. 5.2.

Полученное значение F -критерия значительно ниже табличного для значения 0,05, следовательно, различия между средними значениями цепных абсолютных изменений в разных подпериодах не являются существенными; вероятность нулевой гипотезы (о случайном характере этих различий) много больше 0,05, и она не может быть отклонена. Принимается исходная гипотеза о том, что средние значения абсолютных приростов доходов бюджета постоянны, тренд доходов – прямая линия.

Таблица 5.1

**Проверка гипотезы о линейном тренде доходов бюджета в
Оренбургской области**

Период		Доход, млн.руб.	Скользя- щая 5- летняя средняя	Абсо- лютные измене- ния	Средние по подпериодам абсолютные изменения
1		2	3	4	5
2001	январь	1119,3	-	-	13,14
	февраль	352,2	-	-	
	март	1006,9	948,12	-	
	апрель	1177,8	902,54	-45,58	
	май	1084,4	1017,74	115,2	
	июнь	891,4	1052,04	34,3	
	июль	928,2	1014,36	-37,68	
	август	1178,4	983,92	-30,44	
	сентябрь	989,4	1021,72	37,8	
	октябрь	932,2	1084,78	63,06	
	ноябрь	1080,4	1022,2	-62,58	
	декабрь	1243,5	1024	1,8	
2002	январь	865,5	1066,58	42,58	
	февраль	998,4	1167,62	101,04	
	март	1145,1	1179,12	11,5	
	апрель	1585,6	1202,08	22,96	
	май	1301	1283,1	81,02	
	июнь	980,3	1345,22	62,12	
	июль	1403,5	1260,8	-84,42	

Продолжение таблицы 5.1

1		2	3	4	5
2002	август	1455,7	1307	46,2	-13,89
	сентябрь	1163,5	1370,92	63,92	
	октябрь	1532	1400,04	29,12	
	ноябрь	1299,9	1302,66	-97,38	
	декабрь	1549,1	1249,96	-52,7	
2003	январь	968,8	1223,96	-26	
	февраль	900	1343,74	119,78	
	март	1402	1341,68	-2,06	
	апрель	1898,8	1394,46	52,78	
	май	1538,8	1544,48	150,02	
	июнь	1232,7	1561,46	16,98	
	июль	1650,1	1454,56	-106,9	
	август	1486,9	1541,72	87,16	
	сентябрь	1364,3	1605,4	63,68	
	октябрь	1974,6	1634,5	29,1	
	ноябрь	1551,1	1576,48	-58,02	
	декабрь	1795,6	1521,84	-54,64	
2004	январь	1196,8	1452,8	-69,04	
	февраль	1091,1	1666,62	213,82	
	март	1629,4	1628,24	-38,38	
	апрель	2620,2	1727,44	99,2	
	май	1603,7	1962,72	235,28	
	июнь	1692,8	1997,76	35,04	
	июль	2267,5	1830,28	-167,48	
	август	1804,6	1893,74	63,46	

Продолжение таблицы 5.1

1		2	3	4	5
2004	сентябрь	1782,8	2115,64	221,9	7,7
	октябрь	1921	2190,06	74,42	
	ноябрь	2802,3	2017,96	-172,1	
	декабрь	2639,6	1924,86	-93,1	
2005	январь	944,1	2119,3	194,44	
	февраль	1317,3	2005,7	-113,6	
	март	2893,2	1956,52	-49,18	
	апрель	2234,3	2134,54	178,02	
	май	2393,7	2312,16	177,62	
	июнь	1834,2	2343,86	31,7	
	июль	2205,4	2304,14	-39,72	
	август	3051,7	2273,66	-30,48	
	сентябрь	2035,7	2755,88	482,22	
	октябрь	2241,3	3054,74	298,86	
	ноябрь	4245,3	2759	-295,74	
	декабрь	3699,7	2656,16	-102,84	
2006	январь	1573	2850,94	194,78	
	февраль	1521,5	2576,38	-274,56	
	март	3215,2	2594,92	18,54	
	апрель	2872,5	2824,66	229,74	
	май	3792,4	3139,8	315,14	
	июнь	2721,7	3342,6	202,8	
	июль	3097,2	3192,02	-150,58	
	август	4229,2	3184,84	-7,18	
	сентябрь	2119,6	3323,72	138,88	

Продолжение таблицы 5.1

1		2	3	4	5
2006	октябрь	3756,5	3400,02	76,3	
	ноябрь	3416,1			Общая средняя величина
	декабрь	3478,7			
Итого		132968,7			$\bar{\Delta}_{общ} = +32,78$

Таблица 5.2

Результаты дисперсионного анализа различий между средними абсолютными изменениями

Вид вариации	Сумма квадратов отклонений	Число степеней свободы	Дисперсия	F– критерий	
				фактический	табличный для P = 0,05
Между подпериодами	3191,93	2	1595,963	0,079	3,140
Остаточная	1289613,95	64	20150,218		
Вся вариация	1197915,21	66	18150,231		

Еще один методический прием определения типа тренда – применение многократного аналитического выравнивания с последующим рассмотрением динамики изменений основного параметра тренда по скользящим интервалам. К этому методу следует обратиться после изучения многократного выравнивания, представленного в разд. 5.5.

Кроме рассмотренного подхода к выбору формы кривой можно

применить *метод характеристик прироста*.⁶ Алгоритм метода включает следующие этапы:

1 Предварительная обработка временного ряда: сглаживание ряда по простой (5.1, 5.2) или взвешенной скользящей средней (5.3).

$$\bar{y}_t = \frac{1}{m} \sum_{i=t-p}^{t+p} y_i = \frac{y_{t-p} + y_{t-p+1} + \dots + y_{t+p-1} + y_{t+p}}{2p+1} \quad (5.1)$$

где m - длина интервала сглаживания, $m = 2p + 1$;

i - порядковый номер уровня на интервале сглаживания;

y_i - фактическое значение уровня ряда на момент i .

$$\begin{aligned} \bar{y}_t &= \frac{\frac{1}{2}y_{t-p} + \sum_{i=t-p+1}^{t+p-1} y_i + \frac{1}{2}y_{t+p}}{m} = \\ &= \frac{\frac{1}{2}y_{t-p} + y_{t-p+1} + \dots + y_{t-1} + y_t + y_{t+1} + \dots + y_{t+p-1} + \frac{1}{2}y_{t+p}}{2p} \end{aligned} \quad (5.2)$$

где m - длина интервала сглаживания, $m = 2p$;

i - порядковый номер уровня на интервале сглаживания;

y_i - фактическое значение уровня ряда на момент i ;

y_t - центральный уровень активного участка.

$$\bar{y}_t = \frac{\sum_{i=t-p}^{t+p} y_i w_i}{\sum_{i=t-p}^{t+p} w_i} \quad (5.3)$$

где w_i - весовые коэффициенты (таблица 5.3).

⁶ **Бабешко Л.О.** Основы эконометрического моделирования: Учебное пособие. Изд. 2-е, испр. /Л.О. Бабешко - М.: КомКниа, 2006. – 432 с.

Таблица 5.3

Весовые коэффициенты для взвешенной скользящей средней

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

2 Определение средних приростов (5.4)

$$\bar{u}_t = \frac{\sum_{t=-p}^p t \cdot y_t}{\sum_{t=-p}^p t^2}. \quad (5.4)$$

3 Расчет характеристик изменения приростов.

В таблице 5.4 приводится сводка характерных изменений показателей для различных видов кривых, позволяющая упростить процедуру выбора формы тренда.

В качестве критерия отбора показателя, который в большей степени изменяется во времени линейно, можно воспользоваться коэффициентом детерминации линейного тренда, так как из-за различия масштабов визуальный выбор по графику в большинстве случаев за-

труднителен.⁷

Таблица 5.4

Характеристики изменения приростов

Показатель	Характер изменения во времени	Вид кривой
\bar{u}_t	Примерно одинаковые	Прямая
\bar{u}_t	Линейно изменяются	Парабола второй степени
$\bar{u}_t^{(2)} = \bar{u}_t - \bar{u}_{t-1}$	Линейно изменяются	Парабола третьей степени
$\frac{\bar{u}_t}{\bar{y}_t}$	Примерно одинаковые	Экспонента
$\frac{\bar{u}_t}{\bar{y}_t}$	Линейно изменяются	Логарифмическая парабола
$\ln \bar{u}_t$	Линейно изменяются	Модифицированная экспонента
$\ln \frac{\bar{u}_t}{\bar{y}_t}$	Линейно изменяются	Кривая Гомперца
$\ln \frac{\bar{u}_t}{\bar{y}_t^2}$	Линейно изменяются	Логистическая кривая

5.3. Оценка параметров линейного, параболического и гиперболического трендов

Данные виды трендов объединены в связи с тем, что методика оценки их параметров имеет много общего. Основой этой методики служит метод наименьших квадратов, который дает оценки параметров, отвечающие принципу максимального правдоподобия: сумма квадратов отклонений фактических уровней от тренда (от выравнен-

⁷ Бабешко Л.О. Основы эконометрического моделирования: Учебное пособие. Изд. 2-е, испр. /Л.О. Бабешко - М.: КомКниа, 2006. – 432 с.

ных по уравнению тренда уровней) должна быть минимальной для данного типа уравнения.

Эта методика близка к методике корреляционно-регрессионного анализа связей – парной регрессии. Однако между ними есть и принципиальные различия: выступающий при расчете уравнения тренда в качестве независимой переменной ряд номеров периодов или моментов времени не является случайной варьирующей переменной X регрессионного анализа. Ряд значений времени – это жестко упорядоченный ряд величин, и, следовательно, не может быть речи о корреляции между ним и значениями зависимой переменной – варьирующих уровней показателя, изменяющегося во времени. Нередко применяемые в литературе и в программах для ПК коэффициенты корреляции со временем или фактических уровней с выравненными (т.е. тоже упорядоченными) уровнями тренда таковыми на самом деле не являются и не могут измерять какой-либо «тесноты связи». Чем длиннее период, охватываемый рядом, тем автоматически становятся больше так называемые коэффициенты корреляции при той же самой скорости роста уровней и той же рамой силе колебаний. Таким образом, эти лжекоэффициенты не могут характеризовать соотношение между ролью факторов тенденции и ролью факторов колеблемости.

5.3.1. Уравнение прямой линии тренда

Уравнение имеет вид:

$$\hat{y}_i = a + bt_i,$$

где \hat{y}_i – уровень тренда для периода или момента с номером t_i ;

a – свободный член уравнения, равный среднему уровню тренда для периода (момента) с нулевым номером t_i ;

b – главный параметр линейного тренда – его константа – среднее абсолютное изменение за принятую в ряду единицу времени.

Величина параметров a и b определяется по методу наименьших квадратов путем приравнивания частных первых производных функции

$$f(a, b) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - a - bt_i)^2 \text{ к нулю.}$$

Имеем:

$$\frac{\partial f}{\partial a} = 2 \sum_{i=1}^n (y_i - a - bt_i) \cdot (-1) = 0, \quad (5.5)$$

$$\frac{\partial f}{\partial b} = 2 \sum_{i=1}^n (y_i - a - bt_i) \cdot (-t) = 0. \quad (5.6)$$

После алгебраических преобразований получаем два «нормальных уравнения» МНК для прямой:

$$na + b \sum_{i=1}^n t_i = \sum_{i=1}^n y_i \quad (5.7)$$

$$a \sum_{i=1}^n t_i + b \sum_{i=1}^n t_i^2 = \sum_{i=1}^n (y_i t_i) \quad (5.8)$$

Решая эти уравнения с двумя неизвестными по данным фактического временного ряда y_i ($i=1-n$), получаем значения a и b . Если номера периодов (моментов) времени отсчитываются от начала ряда так, что первый период (момент) обозначен номером $t=1$, то свободный член a есть уровень тренда для предыдущего периода (момента), а не

первого в ряду, как часто ошибочно полагают. Для первого периода уровень тренда \hat{y}_1 равен $a+b$, для второго $\hat{y}_2 = a+2b$ и т.д.

Однако рациональнее начало отсчета времени перенести в середину ряда, т.е. при нечетном n – на период (момент) с номером $(n+1)/2$, а при четном числе уровней ряда – на середину между периодом номером $n/2$ и $(n/2)+1$. В последнем случае все номера периодов t_i будут дробными. При нумерации периодов времени точно от середины ряда, половина номеров t_i будут отрицательными числами (аналогично годам до нашей эры), а половина – положительными, сумма их, т.е. $\sum_{i=1}^n t_i = 0$. В таком случае система нормальных уравнений МНК распадается на два уравнения с одним неизвестным в каждом:

$$na = \sum_{i=1}^n y_i \quad (5.9)$$

$$b \sum_{i=1}^n t_i^2 = \sum_{i=1}^n y_i t_i \quad (5.10)$$

Откуда имеем:

$$a = \frac{\sum_{i=1}^n y_i}{n} = \bar{y} \quad (5.11)$$

$$b = \frac{\sum_{i=1}^n (y_i t_i)}{\sum_{i=1}^n t_i^2} \quad (5.12)$$

К сожалению, многие компьютерные программы не предусматривают такого упрощения, и нумерация периодов (моментов) в них производится с начала ряда, с номера $t = 1$, причем пользователь никак об этом не предупреждается. При расчетах без компьютера, ко-

нечно, следует применять упрощенный прием. Знаменатель в формуле (5.12) при нумерации периодов от середины ряда вычисляется устно при $n \leq 10$, или по формуле:

$$\sum_{i=-\frac{n+1}{2}}^{+\frac{n+1}{2}} t_i^2 = \frac{n^3 - n}{12}.$$

Приведем расчет линейного тренда по временному ряду (см. рис. 4.1). Динамика реальных денежных доходов населения Оренбургской области с 1998 по 2006 г. представлена в табл. 5.5. В целях экономии места в той же таблице приведены и другие показатели, необходимые для измерения колеблемости, описываемые в гл. 6.

По приведенным данным параметры линейного тренда (при расчете используем формулы 5.11 и 5.12) составят:

$$a = \bar{y} = \frac{952}{9} = 105,8 \text{ \%};$$

$$b = \frac{222}{60} = 3,7 \text{ п.п. в год.}$$

Уравнение тренда:

$$\hat{y}_i = 105,8 + 3,7 \cdot t_i,$$

где $t_i = 0$ в 2002 г.

В среднем реальные денежные доходы населения увеличивались на 3,7 п.п. в год.

Сумма уровней тренда должна равняться сумме фактических уровней.

Таблица 5.5

**Динамика реальных денежных доходов населения
Оренбургской области, в % к предыдущему году**

Годы	Уровень, $y_i, \%$	Номер года, t_i	$y_i t_i$	t^2	Тренд $\hat{y}_i, \%$	Отклонение от тренда, $y_i - \hat{y}_i = u_i$	u_i^2	$u_i \cdot u_{i+1}$
1	2	3	4	5	6	7	8	9
1998	78	-4	-312	16	91	-13	168	-
1999	89	-3	-267	9	95	-6	32	73,68
2000	110	-2	-220	4	98	12	135	-65,99
2001	110	-1	-110	1	102	8	63	92,07
2002	113	0	0	0	106	7	52	57,22
2003	113	1	113	1	109	4	12	25,44
2004	112	2	224	4	113	-1	1	-4,15
2005	114	3	342	9	117	-3	8	3,39
2006	113	4	452	16	121	-8	57	21,81
Итого	952	0	222	60	952	0	530	203,47

Источник. Оренбургская область, 2007: статистический ежегодник / Территориальный орган федеральной службы гос. статистики по Оренбургской области. – Оренбург, 2007. – 428с.

5.3.2. Уравнение параболического (II порядка) тренда

Уравнение имеет вид: $\hat{y}_i = a + bt_i + ct_i^2$ Для вычисления параметров a, b, c по методу наименьших квадратов три частных производных функции:

$f(a, b, c) = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ приравняются к нулю и после преобразований

получаем систему трех уравнений с тремя неизвестными:

$$na + b \sum_{i=1}^n t_i + c \sum_{i=1}^n t_i^2 = \sum_{i=1}^n y_i \quad (5.13)$$

$$a \sum_{i=1}^n t_i + b \sum_{i=1}^n t_i^2 + c \sum_{i=1}^n t_i^3 = \sum_{i=1}^n y_i t_i \quad (5.14)$$

$$a \sum_{i=1}^n t_i^2 + b \sum_{i=1}^n t_i^3 + c \sum_{i=1}^n t_i^4 = \sum_{i=1}^n y_i t_i^2 \quad (5.15)$$

При переносе начала отсчета периодов (моментов) времени в середину ряда суммы нечетных степеней номеров этих периодов $\sum t_i$ и $\sum t_i^3$ обращаются в нуль. При этом второе уравнение обращается в уравнение с одним неизвестным, откуда:

$$b = \frac{\sum_{i=1}^n y_i t_i}{\sum_{i=1}^n t_i^2}.$$

Уравнения (5.13) и (5.15) образуют систему двух уравнений с двумя неизвестными:

$$na + c \sum_{i=1}^n t_i^2 = \sum_{i=1}^n y_i$$

$$a \sum_{i=1}^n t_i^2 + c \sum_{i=1}^n t_i^4 = \sum_{i=1}^n y_i t_i^2$$

где, напомним,

$$\sum_{i=1}^n t_i^2 = \frac{n^3 - n}{12};$$

$$\sum_{i=1}^n t_i^4 = \frac{3n^5 - 10n^3 + 7n}{240}$$

Приведем пример расчета параболического тренда по данным рис. 4.2 и табл. 5.6, в которой присутствуют также графы, необходимые для анализа колеблемости, описываемые в гл. 6.

Вычисляем параметры параболы:

$$b = \frac{74268252}{96460} \cong 769,94;$$

$$105a + 96460c = 11943282,$$

$$96460a + 159486964c = 10971116366;$$

$$a + 918,67c = 113745,5,$$

$$a + 1653,4c = 113737,5;$$

$$734,7c = -8,07; c = -0,011$$

$$a = 113755,6$$

Уравнение тренда:

$$\hat{y}_i = 113755,6 + 769,94t_i - 0,011t_i^2,$$

где $t=0$ в 1952 г.

Интерпретация параметров тренда такова: численность населения России в 1900-2004 гг. возрастала в среднем за год на 769938 человек с замедлением роста уровней на $2 \cdot 0,011 = 0,022$ тыс. чел. Средняя численность населения на середину периода была равна 113755,6 тыс. человек.

Таблица 5.6

Расчет параболического тренда численности населения России, тыс. человек

Год	Уровень, y_i тыс. чел.	t_i	$y_i t_i$	t^2	t^4	$y_i t_i^2$	Тренд, \hat{y}_i	$y_i - \hat{y}_i = u_i$	u_i^2	$u_i \cdot u_{i+1}$
1	2	3	4	5	6	7	8	9	10	11
1900	70200	-52	-3650400	2704	7311616	189820800	73689	-3489	12172954	
1901	71476	-51	-3645276	2601	6765201	185909076	74460	-2984	8904548	10411275
1902	72748	-50	-3637400	2500	6250000	181870000	75231	-2483	6165786	7409692
1903	74043	-49	-3628107	2401	5764801	177777243	76002	-1959	3838186	4864713
1904	75361	-48	-3617328	2304	5308416	173631744	76773	-1412	1994128	2766557
1905	76702	-47	-3604994	2209	4879681	169434718	77544	-842	709168	1189189
1906	78067	-46	-3591082	2116	4477456	165189772	78315	-248	61546	208917
1907	79457	-45	-3575565	2025	4100625	160900425	79086	371	137622	-92033
1908	80871	-44	-3558324	1936	3748096	156566256	79857	1014	1028310	376189
1909	82311	-43	-3539373	1849	3418801	152193039	80628	1683	2833024	1706817
1910	83776	-42	-3518592	1764	3111696	147780864	81399	2377	5651479	4001347
1911	85267	-41	-3495947	1681	2825761	143333827	82170	3097	9594079	7363473
1912	86785	-40	-3471400	1600	2560000	138856000	82940	3845	14780949	11908383
1913	88330	-39	-3444870	1521	2313441	134349930	83711	4619	21333230	17757404

Продолжение таблицы 5.6

1	2	3	4	5	6	7	8	9	10	11
1914	89902	-38	-3416276	1444	2085136	129818488	84482	5420	29376443	25033866
1915	90262	-37	-3339694	1369	1874161	123568678	85253	5009	25092475	27150095
1916	90623	-36	-3262428	1296	1679616	117447408	86024	4599	21155363	23039975
1917	91009	-35	-3185315	1225	1500625	111486025	86794	4215	17764328	19385841
1918	90099	-34	-3063366	1156	1336336	104154444	87565	2534	6421541	10680560
1919	89198	-33	-2943534	1089	1185921	97136622	88336	862	743732	2185385
1920	88247	-32	-2823904	1024	1048576	90364928	89106	-859	738321	-741022
1921	88079	-31	-2730449	961	923521	84643919	89877	-1798	3232405	1544847
1922	87912	-30	-2637360	900	810000	79120800	90648	-2736	7482960	4918125
1923	87755	-29	-2544895	841	707281	73801955	91418	-3663	13418221	10020380
1924	89379	-28	-2502612	784	614656	70073136	92189	-2810	7894167	10292020
1925	91042	-27	-2458134	729	531441	66369618	92959	-1917	3675660	5386675
1926	92735	-26	-2411110	676	456976	62688860	93730	-995	989476	1907086
1927	93786	-25	-2344650	625	390625	58616250	94500	-714	510117	710457
1928	94848	-24	-2276352	576	331776	54632448	95271	-423	178679	301906
1929	95923	-23	-2206229	529	279841	50743267	96041	-118	13962	49947
1930	97010	-22	-2134220	484	234256	46952840	96812	198	39364	-23444
1931	98109	-21	-2060289	441	194481	43266069	97582	527	277720	104557
1932	99220	-20	-1984400	400	160000	39688000	98352	868	752730	457217
1933	100345	-19	-1906555	361	130321	36224545	99123	1222	1493849	1060408

Продолжение таблицы 5.6

1	2	3	4	5	6	7	8	9	10	11
1934	101481	-18	-1826658	324	104976	32879844	99893	1588	2521376	1940761
1935	102631	-17	-1744727	289	83521	29660359	100663	1968	3871288	3124255
1936	103794	-16	-1660704	256	65536	26571264	101434	2360	5570808	4643943
1937	104932	-15	-1573980	225	50625	23609700	102204	2728	7441848	6438719
1938	106639	-14	-1492946	196	38416	20901244	102974	3665	13430143	9997254
1939	108377	-13	-1408901	169	28561	18315713	103745	4632	21459862	16976720
1940	110098	-12	-1321176	144	20736	15854112	104515	5583	31172837	25864353
1941	109217	-11	-1201387	121	14641	13215257	105285	3932	15461182	21953790
1942	108343	-10	-1083430	100	10000	10834300	106055	2288	5234486	8996185
1943	106617	-9	-959553	81	6561	8635977	106825	-208	43368	-476453
1944	105764	-8	-846112	64	4096	6768896	107595	-1831	3353938	381382
1945	104918	-7	-734426	49	2401	5140982	108365	-3447	11885125	6313634
1946	104079	-6	-624474	36	1296	3746844	109136	-5057	25568839	17432408
1947	103246	-5	-516230	25	625	2581150	109906	-6660	44350605	33674820
1948	102420	-4	-409680	16	256	1638720	110676	-8256	68155988	54979626
1949	101601	-3	-304803	9	81	914409	111446	-9845	96917744	81274379
1950	101438	-2	-202876	4	16	405752	112216	-10778	116158300	106102782
1951	103112	-1	-103112	1	1	103112	112986	-9874	97488945	106414990
1952	104813	0	0	0	0	0	113756	-8943	79970095	88296094
1953	106542	1	106542	1	1	106542	114526	-7984	63736735	71393506

Продолжение таблицы 5.6

1	2	3	4	5	6	7	8	9	10	11
1954	108300	2	216600	4	16	433200	115295	-6995	48936125	55848266
1955	110087	3	330261	9	81	990783	116065	-5978	35740322	41820962
1956	111903	4	447612	16	256	1790448	116835	-4932	24326439	29486179
1957	113749	5	568745	25	625	2843725	117605	-3856	14868929	19018625
1958	115626	6	693756	36	1296	4162536	118375	-2749	7556143	10599611
1959	117534	7	822738	49	2401	5759166	119145	-1611	2594164	4427401
1960	119046	8	952368	64	4096	7618944	119914	-868	754146	1398706
1961	120766	9	1086894	81	6561	9782046	120684	82	6696	-71063
1962	122407	10	1224070	100	10000	12240700	121454	953	908400	77993
1963	123848	11	1362328	121	14641	14985608	122224	1624	2638646	1548207
1964	125179	12	1502148	144	20736	18025776	122993	2186	4777302	3550438
1965	126309	13	1642017	169	28561	21346221	123763	2546	6482315	5564888
1966	127189	14	1780646	196	38416	24929044	124533	2656	7056440	6763288
1967	128026	15	1920390	225	50625	28805850	125302	2724	7418950	7235425
1968	128696	16	2059136	256	65536	32946176	126072	2624	6886300	7147665
1969	129379	17	2199443	289	83521	37390531	126841	2538	6439409	6659106
1970	130079	18	2341422	324	104976	42145596	127611	2468	6091241	6262906
1971	130704	19	2483376	361	130321	47184144	128380	2324	5398703	5734527
1972	131445	20	2628900	400	160000	52578000	129150	2295	5267025	5332458
1973	132210	21	2776410	441	194481	58304610	129919	2291	5246441	5256723

Продолжение таблицы 5.6

1	2	3	4	5	6	7	8	9	10	11
1974	132941	22	2924702	484	234256	64343444	130689	2252	5071702	5158332
1975	133775	23	3076825	529	279841	70766975	131458	2317	5366631	5217083
1976	134690	24	3232560	576	331776	77581440	132228	2462	6062311	5703874
1977	135645	25	3391125	625	390625	84778125	132997	2648	7010712	6519288
1978	136596	26	3551496	676	456976	92338896	133767	2829	8005482	7491604
1979	137551	27	3713877	729	531441	100274679	134536	3015	9090460	8530739
1980	138291	28	3872148	784	614656	108420144	135305	2986	8914428	9002014
1981	139028	29	4031812	841	707281	116922548	136075	2953	8722518	8817951
1982	139816	30	4194480	900	810000	125834400	136844	2972	8833378	8777773
1983	140766	31	4363746	961	923521	135276126	137613	3153	9940343	9370529
1984	141842	32	4538944	1024	1048576	145246208	138382	3460	11968721	10907484
1985	142823	33	4713159	1089	1185921	155534247	139152	3671	13478877	12701375
1986	143835	34	4890390	1156	1336336	166273260	139921	3914	15320617	14370272
1987	145115	35	5079025	1225	1500625	177765875	140690	4425	19580404	17320042
1988	146343	36	5268348	1296	1679616	189660528	141459	4884	23851659	21610764
1989	147400	37	5453800	1369	1874161	201790600	142228	5172	26746264	25257529
1990	148041	38	5625558	1444	2085136	213771204	142997	5044	25437538	26083694
1991	148543	39	5793177	1521	2313441	225933903	143767	4776	22814675	24090437
1992	148704	40	5948160	1600	2560000	237926400	144536	4168	17375559	19910242
1993	148673	41	6095593	1681	2825761	249919313	145305	3368	11345788	14040634

Продолжение таблицы 5.6

1	2	3	4	5	6	7	8	9	10	11
1994	148366	42	6231372	1764	3111696	261717624	146074	2292	5254749	7721352
1995	148306	43	6377158	1849	3418801	274217794	146843	1463	2141302	3354401
1996	147976	44	6510944	1936	3748096	286481536	147612	364	132741	533140
1997	147502	45	6637590	2025	4100625	298691550	148381	-879	771982	-320115
1998	147105	46	6766830	2116	4477456	311274180	149150	-2045	4180242	1796405
1999	146388	47	6880236	2209	4879681	323371092	149918	-3530	12464296	7218294
2000	145300	48	6974400	2304	5308416	334771200	150687	-5387	29023820	19020029
2001	145600	49	7134400	2401	5764801	349585600	151456	-5856	34295652	31549815
2002	145200	50	7260000	2500	6250000	363000000	152225	-7025	49352030	41140735
2003	144200	51	7354200	2601	6765201	375064200	152994	-8794	77333187	61778231
2004	143500	52	7462000	2704	7311616	388024000	153763	-10263	105323750	90249772
Итого	11943282	0	74268252	96460	159486964	1097111636 6	11943277	5	1728855992	1627645986

Источник. Симчера В.М. Развитие экономики России за 100 лет: 1900-2000. Исторические ряды, вековые тренды, периодические циклы / В.М. Симчера. – М.: ЗАО «Издательство Экономика», 2007. – 683с. ISBN 5-282-02627-9

Если бы параболический тренд вычислялся на ПК по программе, предусматривающей нумерацию лет от начала с номера $t = 1$, то уравнение имело бы вид:

$$\hat{y}_i = 72918 + 771,1t_i - 0,011t_i^2,$$

где $t_i = 1$ в 1900 г.

5.3.3. Гиперболическое уравнение тренда

Уравнение имеет вид:

$$\hat{y}_i = a + \frac{b}{t_i},$$

т.е. отличается от линейного уравнения тем, что вместо t_i в первой степени включает номера периодов времени (моментов) в минус первой степени: $\frac{1}{t_i}$. Соответственно, нормальные уравнения метода наименьших квадратов получают вид:

$$na + b \sum_{i=1}^n \frac{1}{t_i} = \sum_{i=1}^n y_i \quad (5.16)$$

$$a \sum_{i=1}^n \frac{1}{t_i} + b \sum_{i=1}^n \frac{1}{t_i^2} = \sum_{i=1}^n \frac{y_i}{t_i} \quad (5.17)$$

Однако при этом нельзя, в отличие от линейного тренда, переносить начало отсчета периодов времени в середину, так как гипербола не имеет постоянного параметра изменения уровней на протяжении

всего периода, и все величины $\frac{1}{t_i}$ должны быть положительными.

Рассмотрим расчет гиперболического уравнения тренда (табл. 5.7) по данным рис. 4.4 – число прибывших в Оренбургскую область (человек).

Таблица 5.7

Расчет гиперболического уравнения тренда

Год	y_i , человек	t_i	$\frac{1}{t_i}$	$\frac{1}{t_i^2}$	$\frac{y_i}{t_i}$	Тренд, \hat{y}_i	$y_i - \hat{y}_i = u_i$	u_i^2
2000	3623,7	1,0	1,0	1,0	3623,7	3510,4	113,3	12831,1
2001	2738,7	2,0	0,5	0,3	1369,3	2878,6	-140,0	19590,9
2002	2435,7	3,0	0,3	0,1	811,9	2668,0	-232,4	54001,2
2003	2478,3	4,0	0,3	0,1	619,6	2562,8	-84,4	7127,1
2004	2475,1	5,0	0,2	0,0	495,0	2499,6	-24,5	600,1
2005	2776,8	6,0	0,2	0,0	462,8	2457,5	319,4	101997,9
2006	2476,0	7,0	0,1	0,0	353,7	2427,4	48,6	2364,0
Итого	19004,3	28,0	2,6	1,5	7736,0	19004,3	0,0	198512,3

Источник. Оренбургская область, 2007: статистический ежегодник / Территориальный орган федеральной службы гос. статистики по Оренбургской области. – Оренбург, 2007. – 428с.

Нормальные уравнения МНК:

$$7a + 2,6b = 19004,3$$

$$2,6a + 1,5b = 7736$$

Решая систему уравнений, получаем:

$$a = 2246,9; b = 1263,5.$$

Уравнение гиперболического тренда числа прибывших в Оренбургскую область имеет вид:

$$\hat{y}_i = 2246,9 + \frac{1263,5}{t_i},$$

где $t_i = 1$ в 2000г.

Величина численности прибывших 2247 человек – это предел, к которому стремится сокращение численности прибывших на территорию Оренбургской области.

5.4. Оценка параметров экспоненциального, логарифмического и логистического уравнений тренда

Данные типы трендов объединены в одну группу в связи с необходимостью при оценке их параметров прибегать к логарифмированию. При расчете логарифмического уравнения тренда логарифмируют номера периодов (моментов) времени, а при расчете параметров экспоненциального и логистического трендов – сами уровни. Поскольку отрицательные числа не имеют действительных логарифмов, если нужно логарифмировать номера периодов времени, то нельзя переносить начало их отсчета в середину ряда. Если же сами уровни могут принимать отрицательные значения, например, уровни финансового результата от реализации, уровни температуры воздуха или почвы, то необходимо перенести начало отсчета уровней на величину, ал-

гебраически меньшую наименьшего реального уровня. Например, температуру следует выразить не в градусах Цельсия, а в Кельвинах, финансовый результат при наибольшем убытке 83 млн. руб., отсчитывать от –100 млн. руб., чтобы наинизший уровень выразился как 17 млн. руб. По окончании расчета тренда нетрудно восстановить обычные единицы измерения. Так, получив тренд финансового результата при отсчете от –100 млн. руб. как

$$\hat{y}_i = 27 \cdot 1,028^t,$$

нужно по нему рассчитать все уровни тренда, а затем прибавить к ним величину –100 млн. руб. Начиная с $t = 48$, уровни тренда станут положительными числами в обычном смысле: $47 < [\ln(100:27) : \ln 1,028] < 48$.

5.4.1. Экспоненциальное уравнение тренда

Формула уравнения имеет вид:

$$\hat{y}_i = a \cdot k^t.$$

Для нахождения параметров a и k уравнение логарифмируем:

$$\ln \hat{y}_i = \ln a + t_i \ln k.$$

В такой форме, т.е. для логарифмов, уравнение соответствует линейному, а, следовательно, метод наименьших квадратов дает для

логарифмов a и k нормальные уравнения, аналогичные таковым для параметров a и b линейного тренда (см. табл. 5.2.).

$$n \ln a + \ln k \sum_{i=1}^n t_i = \sum_{i=1}^n \ln y_i \quad (5.18)$$

$$\ln a \sum_{i=1}^n t_i + \ln k \sum_{i=1}^n t_i^2 = \sum_{i=1}^n t_i \ln y_i \quad (5.19)$$

Так как номера периодов времени не логарифмируются, можно перенести начало их отсчета в середину ряда и упростить систему:

$$n \ln a = \sum_{i=1}^n \ln y_i, \text{ откуда } \ln a = \overline{\ln y_i},$$

$$\ln k \sum_{i=1}^n t_i^2 = \sum_{i=1}^n t_i \ln y_i, \text{ откуда } \ln k = \frac{\sum_{i=1}^n t_i \ln y_i}{\sum_{i=1}^n t_i^2}.$$

По данным таблицы 5.8 получим:

$$\ln a = \frac{670,18}{92} = 7,2846; \quad a = 1457,7;$$

$$\ln k = \frac{3564,6}{64883} = 0,0549; \quad k = 1,0565.$$

Уравнение тренда примет вид: $\hat{y}_i = 1457,7 \cdot 1,0565^{t_i}$, где $t = 0,5$ в 1946 г.

Итак, национальное богатство в период с 1900 по 1991г. возрастало со среднегодовым темпом роста, равным корню девяносто первой степени из среднего темпа за десятилетие, найденного по данным табл. 5.6, т.е. $\sqrt[91]{1,0565} = 1,0006$, или 0,06 % прироста в год.

Таблица 5.8

**Расчет экспоненциального тренда национальное богатство РФ
в сопоставимых ценах.**

Год	y_i млрд. руб.	t_i	$y_i t_i$	$\ln y_i$	$t_i \ln y_i$	Тренд, \hat{y}_i млрд. руб.	$y_i - \hat{y}_i = u_i$	u_i^2
1	2	3	4	5	6	7	8	9
1900	248	-45,5	5,51	-250,86	2070,25	120	128	16384
1901	248	-44,5	5,51	-245,35	1980,25	126	122	14884
1902	249	-43,5	5,52	-240,01	1892,25	134	115	13225
1903	250	-42,5	5,52	-234,66	1806,25	141	109	11881
1904	261	-41,5	5,56	-230,93	1722,25	149	112	12544
1905	262	-40,5	5,57	-225,52	1640,25	158	104	10816
1906	262	-39,5	5,57	-219,95	1560,25	166	96	9216
1907	263	-38,5	5,57	-214,53	1482,25	176	87	7569
1908	264	-37,5	5,58	-209,10	1406,25	186	78	6084
1909	265	-36,5	5,58	-203,66	1332,25	196	69	4761
1910	277	-35,5	5,62	-199,65	1260,25	207	70	4900
1911	278	-34,5	5,63	-194,15	1190,25	219	59	3481
1912	279	-33,5	5,63	-188,65	1122,25	231	48	2304
1913	280	-32,5	5,63	-183,13	1056,25	244	36	1296
1914	271	-31,5	5,60	-176,47	992,25	258	13	169
1915	273	-30,5	5,61	-171,09	930,25	273	0	0
1916	264	-29,5	5,58	-164,49	870,25	288	-24	576

Продолжение таблицы 5.8

1	2	3	4	5	6	7	8	9
1917	268	-28,5	5,59	-159,34	812,25	305	-37	1369
1918	277	-27,5	5,62	-154,66	756,25	322	-45	2025
1919	287	-26,5	5,66	-149,98	702,25	340	-53	2809
1920	296	-25,5	5,69	-145,10	650,25	359	-63	3969
1921	306	-24,5	5,72	-140,23	600,25	379	-73	5329
1922	327	-23,5	5,79	-136,06	552,25	401	-74	5476
1923	341	-22,5	5,83	-131,22	506,25	423	-82	6724
1924	355	-21,5	5,87	-126,25	462,25	447	-92	8464
1925	369	-20,5	5,91	-121,17	420,25	473	-104	10816
1926	395	-19,5	5,98	-116,59	380,25	499	-104	10816
1927	411	-18,5	6,02	-111,34	342,25	528	-117	13689
1928	438	-17,5	6,08	-106,44	306,25	557	-119	14161
1929	455	-16,5	6,12	-100,98	272,25	589	-134	17956
1930	484	-15,5	6,18	-95,82	240,25	622	-138	19044
1931	513	-14,5	6,24	-90,48	210,25	657	-144	20736
1932	545	-13,5	6,30	-85,06	182,25	694	-149	22201
1933	566	-12,5	6,34	-79,23	156,25	734	-168	28224
1934	599	-11,5	6,40	-73,55	132,25	775	-176	30976
1935	635	-10,5	6,45	-67,76	110,25	819	-184	33856
1936	683	-9,5	6,53	-62,00	90,25	865	-182	33124
1937	723	-8,5	6,58	-55,96	72,25	914	-191	36481
1938	765	-7,5	6,64	-49,80	56,25	965	-200	40000
1939	810	-6,5	6,70	-43,53	42,25	1020	-210	44100

Продолжение таблицы 5.8

1	2	3	4	5	6	7	8	9
1940	868	-5,5	6,77	-37,21	30,25	1078	-210	44100
1941	878	-4,5	6,78	-30,50	20,25	1138	-260	67600
1942	899	-3,5	6,80	-23,80	12,25	1203	-304	92416
1943	932	-2,5	6,84	-17,09	6,25	1271	-339	114921
1944	988	-1,5	6,90	-10,34	2,25	1342	-354	125316
1945	1012	-0,5	6,92	-3,46	0,25	1418	-406	164836
1946	1047	0,5	6,95	3,48	0,25	1498	-451	203401
1947	1071	1,5	6,98	10,46	2,25	1583	-512	262144
1948	1108	2,5	7,01	17,53	6,25	1672	-564	318096
1949	1134	3,5	7,03	24,62	12,25	1767	-633	400689
1950	1191	4,5	7,08	31,87	20,25	1867	-676	456976
1951	1307	5,5	7,18	39,47	30,25	1972	-665	442225
1952	1450	6,5	7,28	47,32	42,25	2083	-633	400689
1953	1597	7,5	7,38	55,32	56,25	2201	-604	364816
1954	1760	8,5	7,47	63,52	72,25	2325	-565	319225
1955	1929	9,5	7,56	71,87	90,25	2457	-528	278784
1956	2127	10,5	7,66	80,46	110,25	2595	-468	219024
1957	2353	11,5	7,76	89,28	132,25	2742	-389	151321
1958	2588	12,5	7,86	98,23	156,25	2897	-309	95481
1959	2855	13,5	7,96	107,42	182,25	3060	-205	42025
1960	3093	14,5	8,04	116,54	210,25	3233	-140	19600
1961	3341	15,5	8,11	125,77	240,25	3416	-75	5625
1962	3610	16,5	8,19	135,16	272,25	3609	1	1

Продолжение таблицы 5.8

1	2	3	4	5	6	7	8	9
1963	3905	17,5	8,27	144,73	306,25	3813	92	8464
1964	4223	18,5	8,35	154,44	342,25	4028	195	38025
1965	4561	19,5	8,43	164,29	380,25	4255	306	93636
1966	4874	20,5	8,49	174,08	420,25	4496	378	142884
1967	5209	21,5	8,56	184,00	462,25	4750	459	210681
1968	5582	22,5	8,63	194,11	506,25	5018	564	318096
1969	5988	23,5	8,70	204,39	552,25	5301	687	471969
1970	6416	24,5	8,77	214,78	600,25	5601	815	664225
1971	6866	25,5	8,83	225,28	650,25	5917	949	900601
1972	7360	26,5	8,90	235,95	702,25	6251	1109	1229881
1973	7879	27,5	8,97	246,73	756,25	6604	1275	1625625
1974	8437	28,5	9,04	257,65	812,25	6977	1460	2131600
1975	9030	29,5	9,11	268,70	870,25	7371	1659	2752281
1976	9624	30,5	9,17	279,75	930,25	7787	1837	3374569
1977	10238	31,5	9,23	290,87	992,25	8227	2011	4044121
1978	10916	32,5	9,30	302,18	1056,25	8692	2224	4946176
1979	11598	33,5	9,36	313,51	1122,25	9183	2415	5832225
1980	12283	34,5	9,42	324,85	1190,25	9701	2582	6666724
1981	13 002	35,5	9,47	336,29	1260,25	10249	2753	7579009
1982	13 764	36,5	9,53	347,84	1332,25	10828	2936	8620096
1983	14 573	37,5	9,59	359,51	1406,25	11439	3134	9821956
1984	15 433	38,5	9,64	371,30	1482,25	12085	3348	11209104
1985	16 302	39,5	9,70	383,11	1560,25	12768	3534	12489156

Продолжение таблицы 5.8

1	2	3	4	5	6	7	8	9
1986	17 152	40,5	9,75	394,87	1640,25	13489	3663	13417569
1987	17 958	41,5	9,80	406,53	1722,25	14251	3707	13741849
1988	19 065	42,5	9,86	418,86	1806,25	15056	4009	16072081
1989	19 766	43,5	9,89	430,29	1892,25	15906	3860	14899600
1990	20 579	44,5	9,93	441,98	1980,25	16804	3775	14250625
1991	20 902	45,5	9,95	452,62	2070,25	17753	3149	9916201
Итого	377965	0	670,18	3564,60	64883,00	329987	47979	172590775

Источник. Симчера В.М. Развитие экономики России за 100 лет: 1900-2000. Исторические ряды, вековые тренды, периодические циклы / В.М. Симчера. – М.: ЗАО «Издательство Экономика», 2007. – 683с. ISBN 5-282-02627-9

5.4.2. Логарифмическое уравнение тренда

Особенность этого типа тренда заключается в том, что логарифмировать необходимо номера периодов (моментов) времени: $\hat{y} = a + b \ln t$. Следовательно, все номера должны быть положительными числами. Однако это вовсе не означает, что нумерацию следует начинать с числа 1. Дело в том, что величина логарифма быстро возрастает при переходе от единицы к двум: натуральный логарифм единицы равен нулю, а логарифм двух равен 0,693, имеем рост на 0,693; в то же время логарифм четырех равен 1,386, а логарифм пяти равен 1,609, имеем прирост лишь на 0,223 и т.д. Если уровень изучаемого ряда в начале возрастает втрое быстрее, чем между четвертым и пятым периодом, тогда нумерация от единицы допустима. Если же уменьшение прироста уровней происходит значительно медленнее, нумерацию периодов (моментов) следует начинать не с единицы, а с большего числа.

Покажем методику расчета логарифмического уравнения тренда на примере динамики естественной убыли населения в Оренбургской области за 1985 – 2006гг. (см. рис. 4.5; табл. 5.9).

Таблица 5.9

**Расчет логарифмического тренда естественной убыли населения
в Оренбургской области на 1000 человек**

Годы	y_i	t_i	$\ln t_i$	$(\ln t)^2$	$y \cdot \ln t$	\hat{y}_i	$y_i - \hat{y}_i = u_i$	u_i^2	$u_i u_{i+1}$
1	2	3	4	5	6	7	8	9	10
1985	8,4	1	0	0	0	13,7	-5,344	28,558	-
1986	10,4	2	0,693	0,480	7,209	9,4	0,970	0,941	26,860
1987	9,5	3	1,099	1,207	10,437	6,9	2,593	6,725	6,325
1988	8,1	4	1,386	1,922	11,229	5,1	2,984	8,902	59,863
1989	7	5	1,609	2,590	11,266	3,7	3,272	10,708	95,323
1990	5,8	6	1,792	3,210	10,392	2,6	3,207	10,285	110,133
1991	3,5	7	1,946	3,787	6,811	1,6	1,866	3,483	35,826
1992	1,5	8	2,079	4,324	3,119	0,8	0,697	0,486	1,694
1993	-2	9	2,197	4,828	-4,394	0,1	-2,070	4,283	2,083
1994	-3,1	10	2,303	5,302	-7,138	-0,6	-2,514	6,320	27,067
1995	-3,3	11	2,398	5,750	-7,913	-1,2	-2,121	4,497	28,421
1996	-3,2	12	2,485	6,175	-7,952	-1,7	-1,479	2,188	9,840
1997	-3,3	13	2,565	6,579	-8,464	-2,2	-1,081	1,169	2,557
1998	-2,9	14	2,639	6,965	-7,653	-2,7	-0,220	0,048	0,056

Продолжение таблицы 5.9

1	2	3	4	5	6	7	8	9	10
1999	-4,6	15	2,708	7,334	-12,457	-3,1	-1,490	2,221	0,107
2000	-4,7	16	2,773	7,687	-13,031	-3,5	-1,189	1,413	3,139
2001	-4,7	17	2,833	8,027	-13,316	-3,9	-0,811	0,659	0,931
2002	-4,3	18	2,890	8,354	-12,429	-4,2	-0,056	0,003	0,002
2003	-4,4	19	2,944	8,670	-12,956	-4,6	0,181	0,033	0,000
2004	-4,1	20	2,996	8,974	-12,283	-4,9	0,800	0,640	0,021
2005	-5	21	3,045	9,269	-15,223	-5,2	0,204	0,041	0,027
2006	-3,9	22	3,091	9,555	-12,055	-5,5	1,593	2,538	0,105
Итого	0,7	-	48,471	120,988	-86,801	0,7	5,336	67,583	410,381

Источник. Оренбургская область, 2007: статистический ежегодник / Территориальный орган федеральной службы гос. статистики по Оренбургской области. – Оренбург, 2007. – 428с.

Система нормальных уравнений для оценки параметров тренда имеет вид:

$$\begin{cases} a \cdot n + b \cdot \sum \ln t = \sum y \\ a \cdot \sum \ln t + b \cdot \sum (\ln t)^2 = \sum y \cdot \ln t \end{cases}$$

По данным таблицы 5.9 получим:

$$\begin{cases} 22 \cdot a + 48,471 \cdot b = 0,7 \\ 48,471 \cdot a + 120,99 \cdot b = -86,8 \end{cases}$$

Решив систему получим: $a = 13,744$, $b = -6,224$

Итак, уравнение логарифмического тренда имеет вид:

$$\hat{y} = 13,744 - 6,224 \cdot \ln t,$$

где $t = 1$ в 1985 году.

По этому уравнению рассчитаны уровни тренда \hat{y}_i в табл. 5.9. Суммы фактических и теоретических уровней полностью совпали. Кривая хорошо отражает тенденцию.

5.4.3. Логистическое уравнение тренда

Уравнение имеет наиболее общий вид:

$$\hat{y}_i = \frac{y_{\max} - y_{\min}}{e^{a_0 + a_1 t_i} + 1} + y_{\min}$$

При расчете этого уравнения логарифмируют величину, произ-

водную от уровней ряда, но не номера периодов (моментов) времени. Поэтому рационально производить эту нумерацию от середины ряда. Особенностью логистического тренда является этап обоснования значений максимального и минимального уровней временного ряда. Это обоснование производится на основе, во-первых, уровней фактического ряда, а во-вторых, теоретических соображений, т.е. внешних по отношению к статистике, относящихся к содержанию изучаемого процесса.

Уравнение логистического тренда в общем виде непосредственно логарифмировать невозможно. Преобразуем его в форму:

$$\frac{\hat{y}_{\max} - \hat{y}_{\min}}{\hat{y}_i - \hat{y}_{\min}} - 1 = e^{a_0 + a_1 t_i}$$

и обозначим его левую часть, т.е.

$$\frac{\hat{y}_{\max} - \hat{y}_{\min}}{\hat{y}_i - \hat{y}_{\min}} - 1 = \hat{\zeta}_i, \text{ т.е. } \hat{\zeta}_i = e^{a_0 + a_1 t_i}; \ln \hat{\zeta}_i = a_0 + a_1 t_i$$

Условие метода наименьших квадратов:

$$\sum_{i=1}^n (\ln \zeta_i - \ln \hat{\zeta}_i)^2 \rightarrow \min ,$$

подставляя значение $\ln \hat{\zeta}_i$, имеем:

$$\sum_{i=1}^n (\ln \zeta_i - a_0 - a_1 t_i)^2 \rightarrow \min$$

После вычисления частных производных по a_0 и по a_1 , получаем нормальные уравнения МНК для логистической кривой, аналогичные таковым для прямой линии, т.к. заменой на ζ фактически проведена линеаризация функции логистической кривой:

$$na_0 + a_1 \sum_{i=1}^n t_i = \sum_{i=1}^n \ln \zeta_i \quad (5.25)$$

$$a_0 \sum_{i=1}^n t_i + a_1 \sum_{i=1}^n t_i^2 = \sum_{i=1}^n t_i \ln \zeta_i \quad (5.26)$$

При переносе начала отсчета периодов (моментов) времени в середину ряда система упрощается до двух уравнений с одним неизвестным в каждом из них:

$$na_0 = \sum_{i=1}^n \ln \zeta_i, \text{ откуда } a_0 = \overline{\ln \zeta_i},$$

$$a_1 \sum_{i=1}^n t_i^2 = \sum_{i=1}^n t_i \ln \zeta_i, \text{ откуда } a_1 = \frac{\sum_{i=1}^n t_i \ln \zeta_i}{\sum_{i=1}^n t_i^2}.$$

Итак, алгоритм расчета логистической кривой состоит из десяти этапов:

- 1) обоснование величин \hat{y}_{\max} и \hat{y}_{\min} ;
- 2) вычисление по фактическому временному ряду значений

$$\zeta_i = \frac{\hat{y}_{\max} - \hat{y}_{\min}}{y_i - \hat{y}_{\min}} - 1;$$

- 3) вычисление $\ln \zeta_i$;

- 4) нумерация периодов или моментов времени от середины ряда;
- 5) умножение $\ln \zeta_i$ на t_i ;
- 6) подсчет итоговых сумм $\sum_{i=1}^n \ln \zeta_i$; $\sum_{i=1}^n t_i \ln \zeta_i$.
- 7) вычисление a_0 и a_1 ;
- 8) вычисление $\ln \hat{\zeta}_i = a_0 + a_1 t_i$;
- 9) вычисление $\hat{\zeta}_i = \exp(a_0 + a_1 t_i)$ для всех периодов;
- 10) вычисление уровней тренда $\hat{y}_i = \frac{\hat{y}_{\max} - \hat{y}_{\min}}{\hat{\zeta}_i + 1} + \hat{y}_{\min}$.

Проведем расчет логистического тренда по данным рис. 5.2.

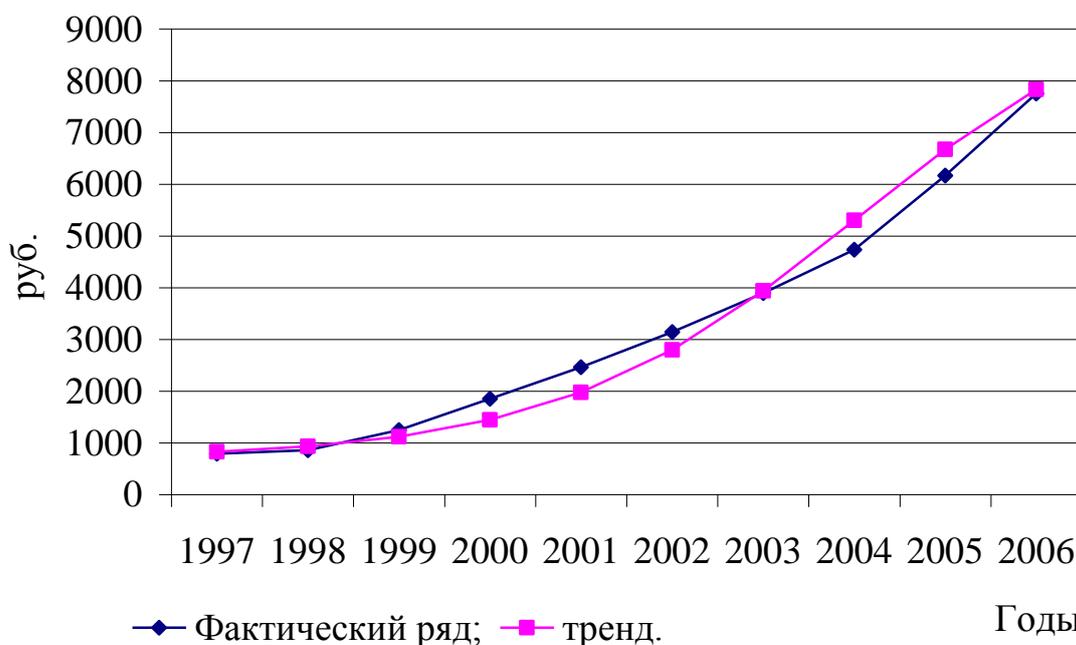


Рис. 5.2. Логистическая тенденция динамики среднемесячной номинальной начисленной заработной платы работающих в экономике, руб. (до 1998г-тыс.руб.)

Таблица 5.10

**Расчет логистического тренда среднемесячной номинальной начисленной заработной платы
работающих в экономике, руб. (до 1998г-тыс.руб.)**

Годы	y_i	$\zeta_i = \frac{10000 - 700}{y_i - 700} - 1$	$\ln \zeta_i$	t_i	$t_i \ln \zeta_i$	t^2	$\hat{\zeta}_i$	Тренд $\hat{y}_i = \frac{10000 - 700}{\hat{\zeta}_i + 1} + 700$	u_i	u_i^2
1997	788	104,68	4,7	-4,5	-20,9	20,3	71,1	829	-41	1676
1998	855	59,00	4,1	-3,5	-14,3	12,3	38,8	934	-79	6201
1999	1249	15,94	2,8	-2,5	-6,9	6,3	21,2	1120	129	16686
2000	1849	7,09	2,0	-1,5	-2,9	2,3	11,5	1442	407	165711
2001	2460	4,28	1,5	-0,5	-0,7	0,3	6,3	1976	484	234605
2002	3142	2,81	1,0	0,5	0,5	0,3	3,4	2799	343	117567
2003	3898	1,91	0,6	1,5	1,0	2,3	1,9	3940	-42	1729
2004	4735	1,30	0,3	2,5	0,7	6,3	1,0	5304	-569	323227
2005	6164	0,70	-0,4	3,5	-1,2	12,3	0,6	6676	-512	261661
2006	7753	0,32	-1,1	4,5	-5,1	20,3	0,3	7835	-82	6756
Итого	32893	-	15,4	-	-50,0	82,5	-	32853	40	1597

Источник. Оренбургская область, 2007: статистический ежегодник / Территориальный орган федеральной службы гос. статистики по Оренбургской области. – Оренбург, 2007. – 428с.

Исходя из границ периода времени и фактических уровней ряда, получаем:

$$\hat{y}_{\min} = 700; \hat{y}_{\max} = 10000;$$

$$a_0 = \frac{15,4}{10} = 1,154; a_1 = \frac{-50,0}{82,5} = -0,606$$

$$\hat{\zeta}_i = \exp[1,154 + t_i(-0,606)].$$

Уравнение логистического тренда среднемесячной номинальной начисленной заработной платы работающих в экономике имеет вид:

$$\hat{y}_i = \frac{10000 - 700}{e^{1,154 - 0,606t_i} + 1} + 700.$$

Рисунок 5.2. показывает достаточно близкое приближение логистической кривой к исходным данным. Напомним, что, в отличие от прямой и параболы, алгоритм расчета других кривых не предусматривает автоматического равенства сумм выравненных и фактических уровней, они совпадают только при идеальном выражении тенденции ряда данным уравнением тренда.

5.5. Многократное скользящее выравнивание

Как видно из табл. 5.5–5.7, при расчете параметров тренда разные уровни имеют неодинаковые веса, так как умножаются на разные величины t_i . Наибольшие веса имеют уровни, стоящие в начале и кон-

це временного ряда, что особенно явно видно при нумерации лет от середины ряда. То же самое происходит и при нумерации периодов (моментов) от начала ряда, так как можно легко доказать, что в этом случае в расчет входят не сами номера лет t_i , а их отклонения от среднего номера, т.е. $t_i - \bar{t}$ а это то же самое, что и номера периодов от середины ряда.

Ввиду этого если в начале ряда находятся уровни с отрицательными отклонениями от нормы, от тренда (например, естественная убыль населения), а в конце ряда располагаются уровни с положительными отклонениями от тренда (естественный прирост населения), то среднегодовой прирост в линейном тренде, или ускорение прироста в параболе, будет завышен за счет случайной колеблемости. Если же в начале ряда будут находиться уровни с положительными отклонениями от тренда, а в конце его – с отрицательными, то параметры трендов будут занижены.

Следующий шаг в освобождении параметров тренда от влияния случайного распределения положительных и отрицательных колебаний уровней на протяжении временного ряда можно сделать, применяя методику многократного скользящего выравнивания⁸.

Сущность данного метода довольно проста: чтобы избежать преимущественного влияния уровней, стоящих на концах временного ряда, следует сделать так, чтобы «на концах» побывали все уровни. Для этого следует достаточно длинный временной ряд выравнивать не в один прием, а скользящим способом по более дробным отрезкам. При этом каждый раз вычисляется среднегодовой прирост, например b линейного тренда, а на концах будут года, то благоприятные, то не-

⁸ Юзбашев М.М., Манелля А.И. Статистический анализ тенденций и колеблемости. – М.: Финансы и статистика, 1983.

благоприятные. Обозначив буквой n длину каждой базы, т.е. число уровней, по которым производится расчет параметра, а число таких баз расчета, укладываемых в ряд длиной N уровней, – буквой l , составим равенство:

$$n + l = N + 1.$$

В ряду из 26 уровней уложатся 12 баз по 15 уровней в каждой. Получим 12 значений среднегодового прироста, часть из них – заниженные, часть – завышенные, часть – неискаженные. Теперь разумно усреднить полученные значения параметра: ведь в средней величине случайные отклонения взаимно погашаются. Получим значение среднегодового прироста, максимально освобожденное от влияния случайного распределения неблагоприятных или благоприятных лет по длине исходного временного ряда.

Методика многократного скользящего выравнивания имеет, как, впрочем, и всякая иная, свои ограничения.

Во-первых, для ее применения необходимо иметь достаточно длинный временной ряд при наличии в нем единой качественной тенденции. Если для однократного расчета параметра достаточным (минимальным) можно считать ряд из 9 – 11 уровней, а для достаточной степени взаимопогашения в средней величине следует иметь не менее 6–8 заниженных и завышенных значений параметра, т.е. минимальное значение будет $l = 6 - 8$, минимальная длина исходного временного ряда, т.е. N , должна составлять $m + l - 1 = (9 + 6 - 1) - (11 + 8 - 1)$, или от 14 до 18 уровней. При более коротких рядах применение многократно скользящего выравнивания нецелесообразно.

Во-вторых, многократное выравнивание следует применять, ес-

ли колеблемость исходных уровней достаточно существенная, коэффициент колеблемости (см. о нем в гл. 6) должен быть не ниже 5%. При более слабой колеблемости искажения параметра a невелики и при однократном выравнивании, поэтому нет необходимости «стрелять из пушки по воробьям», применяя сложную методику многократного выравнивания.

В-третьих, при наличии долгопериодических (циклических) колебаний, чтобы многократное выравнивание не привело к искажению значения параметра тренда, необходимо соблюдать хотя бы одно из двух условий:

1) длина базы выравнивания, т.е. n , должна быть равна или кратна длине цикла;

2) число баз скользящего выравнивания, т.е. l , должно быть равно длине цикла.

При соблюдении одного из указанных условий или обоих будут перебраны на началах и концах базы выравнивания все фазы цикла в равном количестве, и тогда циклическая колеблемость, равно как и случайная, в основном будет исключена из усредненного значения параметра тренда.

Наконец, следует помнить, что искажающее влияние распределения случайных отклонений по длине временного ряда относится только к параметру динамики – среднегодовому (месячному и т.д.) приросту, ускорению и т.д., но не к среднему уровню ряда, не к свободному члену уравнения тренда. В связи с этим не нужно усреднять значения свободного члена по скользящим базам, а в качестве свободного члена для прямой следует взять общую среднюю величину уровней исходного ряда, т.е. $a = \sum_1^N y_i : N$. Для параболы, экспоненты и т.д.

свободный член определяется расчетом на основе этой же средней величины. Так, для параболы:

$$a = \bar{y} - \frac{c \sum_{i=1}^n t_i^2}{n} .$$

Рассмотрим пример многократного скользящего выравнивания по данным табл. 5.1. Тренд, как показано ранее, линейный, но колеблемость существенная. Сделаем 30 скользящих баз расчета среднегодового прироста по 43 уровня в каждой.

Таблица 5.11

Результаты расчетов по многократному выравниванию

Скользящие базы расчета, период	Среднегодовой прирост, b_i , млн.руб. в год
1	2
01.2001-07.2004	22,9
02.2001-08.2004	23,7
03.2001-09.2004	21,8
04.2001-10.2004	22,3
05.2001-11.2004	26,0
06.2001-12.2004	28,6
07.2001-01.2005	25,0
08.2001-02.2005	22,7
09.2001-03.2005	26,2
10.2001-04.2005	26,7
11.2001-05.2005	27,3

Продолжение таблицы 5.11

1	2
12.2001-06.2005	26,4
01.2002-07.2005	27,1
02.2002-08.2005	29,1
03.2002-09.2005	28,0
04.2002-10.2005	27,8
05.2002-11.2005	35,3
06.2002-12.2005	39,8
07.2002-01.2006	36,1
08.2002-02.2006	33,5
09.2002-03.2006	36,4
10.2002-04.2006	37,1
11.2002-05.2006	41,6
12.2002-06.2006	41,6
01.2003-07.2006	43,4
02.2003-08.2006	46,6
03.2003-09.2006	42,4
04.2003-10.2006	44,9
05.2003-11.2006	47,7
06.2003-12.2006	49,1
Итого	986,9
В среднем	32,9

Средний уровень: $a = \bar{y} = \frac{132968,7}{72} = 1846,8$ млн.руб.

Среднее среднегодовое изменение (прирост):

$$\bar{b} = \frac{\sum_1^{30} b_i}{30} = \frac{986,9}{30} = 32,9 \text{ млн.руб. в год.}$$

Уравнение тренда:

$$\hat{y}_i = 1846,8 + 32,9 \cdot t_i,$$

где $t_i = 0,5$ в январе 2004 г.

По этому уравнению в приложении 2 вычислены уровни тренда и отклонения от него.

Как видно из табл. 5.11, среднегодовой прирост по скользящим базам расчета то возрастает, то снижается. Поскольку нет определенного направления тенденции изменения величины b_i , можно считать, что их различие – следствие колебаний уровней и небольших колебаний скорости роста доходов бюджета, однако, в пределах единой линейной тенденции. В связи с этим допустимо усреднение значений среднегодового прироста.

Если же в результате многократного скользящего выравнивания обнаружится систематическое и существенное возрастание или убывание среднегодового прироста, это означает, что тенденция на самом деле не линейная, а параболическая, экспоненциальная, гиперболическая или логарифмическая. Таким образом, по результатам многократного выравнивания можно исправить допущенную на предыдущих этапах (если они выполнялись) ошибку в определении типа тренда или в периодизации динамики.

В заключение данной главы на примере последнего тренда покажем, как рассчитать описанные в гл. 3 показатели динамики.

Абсолютное изменение: если тренд линейный, то оно – главный параметр, т.е. b или \bar{b} при многократном выравнивании. Если тренд криволинейный, то абсолютное изменение – непостоянная величина. За любой период его можно вычислить, вычитая из текущего выравненного уровня базисный выравненный, т.е. абсолютное изменение равно $\hat{y}_m - \hat{y}_0$.

Абсолютное ускорение: для параболического тренда II порядка оно – главный параметр, но не забывайте, что оно равно 2 с, т.е. удвоенному квадратическому члену (ускорение – вторая производная по времени; вторая производная от ct^2 по t будет равна 2 с).

При других типах тренда ускорение за период m можно найти из уравнений:

$$(\hat{y}_m - \hat{y}_{m-1}) - (\hat{y}_{m-1} - \hat{y}_{m-2}) = \hat{y}_m - 2\hat{y}_{m-1} + \hat{y}_{m-2}.$$

Например, ускорение естественной прироста населения в Оренбургской области на 1000 человек в 1987 году (см. табл. 5.9) составило:

$$\hat{y}_{87} - 2\hat{y}_{86} + \hat{y}_{85} = 6,9 - 2 \cdot 9,4 + 13,7 = 1,8 \text{ в год за год.}$$

Темп роста – основной параметр экспоненциального тренда k . Для всех других типов тренда для вычисления темпа роста за некоторый период m следует разделить \hat{y}_m на \hat{y}_0 , т.е. на выравненный уровень базисного года. Так, для линейного тренда (см. табл. 5.5) темп изменения реальных денежных доходов населения в Оренбургской области (в % к предыдущему году) в 2006 г. по сравнению с 1998 г. составил:

$$121: 91 = 0,7521, \text{ или } 75,21\%.$$

Глава 6. МЕТОДЫ РАСПОЗНАВАНИЯ ТИПА КОЛЕБАНИЙ И ОЦЕНКИ ПАРАМЕТРОВ КОЛЕБЛЕМОСТИ

В гл. 2 было показано, что временной ряд, как правило, содержит два основных элемента: тенденцию динамики и колеблемость. Эти составляющие в разных реальных временных рядах находятся в неодинаковом соотношении, а в крайних случаях остается один элемент: ряд без колеблемости уровней представляет собой тренд в чистом виде, а ряд без тенденции динамики, но с колебаниями уровней около постоянной средней величины – это стационарный временной ряд. Оба крайних случая крайне редки на практике. Обычно тенденция и колеблемость сочетаются в исходном ряду, и методы статистического анализа, изложенные в гл. 4 и 5, призваны «очистить» тенденцию от колебаний, измерить ее параметры. Колеблемость в этом случае выступала как помеха, «шум», мешающий выделить и интерпретировать «сигнал», т.е. параметры тренда. Нередко в учебной литературе взгляд на колеблемость как на помеху в изучении тенденции преобладает или является единственным.

Однако сама колеблемость также представляет собой важный предмет статистического исследования временных рядов. Значение колеблемости многогранно:

1. она позволяет выдвинуть гипотезы о причинах колебаний, о путях влияния на них;
2. на основе параметров колеблемости можно ее прогнозировать или учитывать как фактор ошибки прогноза (гл. 10), т.е. сделать прогноз наиболее надежным и (или) точным;
3. на основе параметров и прогнозов колебаний можно рассчитать резервы, страховой запас, необходимый для преодоления вред-

ных последствий колебания уровней, например, валовых сборов зерна.

Колебания уровней временного ряда могут иметь разную форму, разное распределение по времени, разную частоту и амплитуду. В данной главе рассматриваются методы исследования этих свойств колеблемости, их отображения в системе показателей, характеризующих колеблемость тех или иных явлений. Что же касается дальнейшего изучения причин, механизма колебаний, то эта задача уже выходит за пределы статистического исследования и должна выполняться наукой, изучающей по существу те явления и процессы, динамика которых отражена временным рядом.

6.1. Графическое отображение и основные свойства разных типов колебаний.

Так же, как изучение тенденции, исследование колебаний целесообразно начать с графического изображения – обобщающего, целостного впечатления о временном ряде.

Все многообразие встречающихся колебаний во временных рядах можно представить как «смесь» в разных пропорциях трех основных типов:

- пилообразной или маятниковой колеблемости;
- долгопериодических циклов колебаний;
- случайно распределенной во времени колеблемости.

Графическое изображение каждого из этих типов и описание основных свойств каждого типа колеблемости, во-первых, помогают по виду фактического ряда определить, каков преобладающий в нем тип колебаний; во-вторых, помогают экономисту, менеджеру, другому

специалисту понять, какие последствия могут иметь колебания для его сферы деятельности и как с этими колебаниями (если нужно) бороться.

6.1.1. пилообразная колеблемость

Характерной чертой этого типа колеблемости является правильное, регулярное чередование отклонений от тренда «вверх» и «вниз», т.е. положительных по знаку и отрицательных через одно. Поскольку это похоже на колебание маятника часов вправо-влево, данный тип колеблемости называют также «маятниковой колеблемостью». Название же «пилообразная» происходит от вида графика (рис. 6.1), похожего на зубья пилы (хотя величина «зубьев», разумеется, не должна быть, как у хорошей пилы, одинаковой).

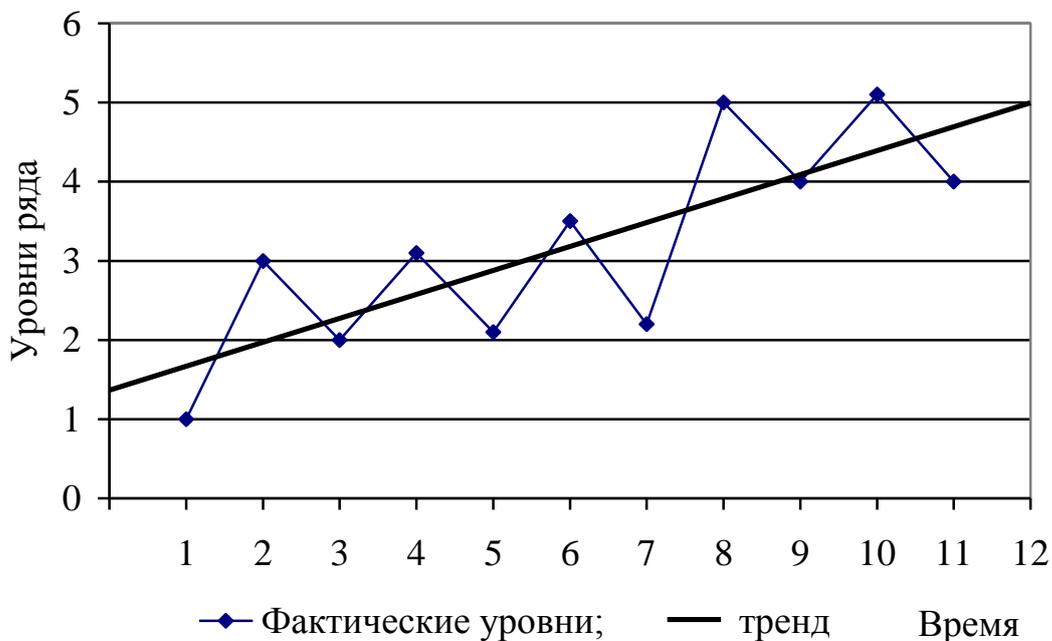


Рис. 6.1. Пилообразная колеблемость

Свойства пилообразной колеблемости таковы: из-за частой смены знака отклонения от тренда не происходит аккумуляции ни положительных, ни отрицательных отклонений. Следовательно, нет необходимости создавать для их компенсации значительный страховой запас. Регулярность чередования отклонений дает возможность надежного их прогнозирования: если в данный период отклонение отрицательное, то в периоде 5 вперед оно будет положительным (данный период считать номером нулевым). Число положительных отклонений при достаточно большой длине ряда равно (точнее стремится к равенству) числу отрицательных отклонений, а общее количество локальных экстремумов (отклонений от тренда, которые либо меньше, либо больше двух соседних по алгебраической величине) равно числу уровней.

Причины пилообразной колеблемости зависят как от внутренних факторов системы, так и от внешних. Внутренние для агротехнической системы причины пилообразной колеблемости урожайности – это колебания содержания питательных веществ в почве. Если, по какой-либо внешней причине, в данном году получен особо высокий урожай, то он выносит из почвы больше питательных веществ, чем в среднем в ней образуется за счет деятельности микроорганизмов, вносится с осадками и ветром. Следовательно, в следующем году ввиду более низкого содержания питательных веществ в почве урожай будет ниже нормы (средней, тренда), и вынесет меньше питательных веществ из почвы, чем в ней образуется за год, а, следовательно, следующий урожай (второго от базы выравнивания периода) опять будет выше среднего и т.д. Конечно, рациональная агротехника подавит пилообразные колебания, например, увеличив после высокого урожая внесение в почву удобрений, и компенсирует повышенные этим за-

траты, сократив (а не повысив, как обычно думают) внесение удобрений после низкого урожая, например, после засухи. В чистом виде пилообразные колебания урожаев не наблюдаются в нашу эпоху, но как составляющая часть колеблемости, особенно на коротких отрезках времени, они существенны.

Распознать наличие пилообразных колебаний как элемента во временном ряду можно, во-первых, по виду графика, во-вторых, подсчетом числа локальных экстремумов в ряду отклонений от тренда: чем это число ближе к числу уровней ряда, тем большую роль играют пилообразные колебания в их общем комплексе. Третий способ распознавания – по знаку и величине коэффициента автокорреляции отклонений от тренда I порядка, т.е. со сдвигом (лагом) на 1 год.

Коэффициент автокорреляции отклонений имеет формулу:

$$r_u^{a'} = \frac{\sum_{i=1}^{n-1} U_i U_{i+1}}{\frac{U_1^2}{2} + \sum_{i=2}^{n-1} U_i^2 + \frac{U_n^2}{2}} \quad (6.1)$$

Числитель коэффициента – сумма произведений каждого отклонения на следующее, кроме последнего в ряду отклонений. В этих произведениях первое отклонение и последнее, т.е. U_1 и U_n участвуют только по одному разу, а отклонения от U_2 до U_{n-1} – по два раза. Соответственно, в знаменателе в сумму квадратов отклонений от U_2^2 до U_{n-1}^2 входят квадраты с единичным весом, а квадраты первого и последнего отклонений U_1^2 и U_n^2 – с половинным весом.

Чем ближе коэффициент автокорреляции к -1 , тем большую роль играет пилообразная составляющая в общей колеблемости изу-

чаемого временного ряда. При коэффициенте, по алгебраической величине превышающем $-0,3$, можно считать пилообразную составляющую несущественной или отсутствующей вовсе, если длина ряда не больше 20 уровней.

6.1.2. Долгопериодическая циклическая колеблемость

Характерной чертой этого типа колебаний является наличие нескольких (многих) подряд отклонений одного знака, затем сменяющихся примерно таким же количеством отклонений противоположного знака подряд. Затем весь цикл вновь повторяется, причем, как правило, длина всех циклов одинакова или хотя бы примерно равная. Если равенство отдельных циклов существенно нарушается, говорят о *квазициклической колеблемости*, т.е. как бы циклической.

Свойства циклической колеблемости (рис. 6.2) таковы: отклонения одного и того же знака следуют подряд в течение примерно половины длины цикла. Следовательно, эти отклонения аккумулируются, и для их компенсации (если таковая требуется), нужен большой страховой запас. Например, надой молока от коров находится ниже тренда в течение 6 месяцев года (с октября до марта включительно) в большинстве сельхозпредприятий Ленинградской области и других регионов России. Следовательно, для удовлетворения спроса на молоко в осенне-зимний период нужен запас в форме сухого молока, масла и других хранящихся молочных продуктов.

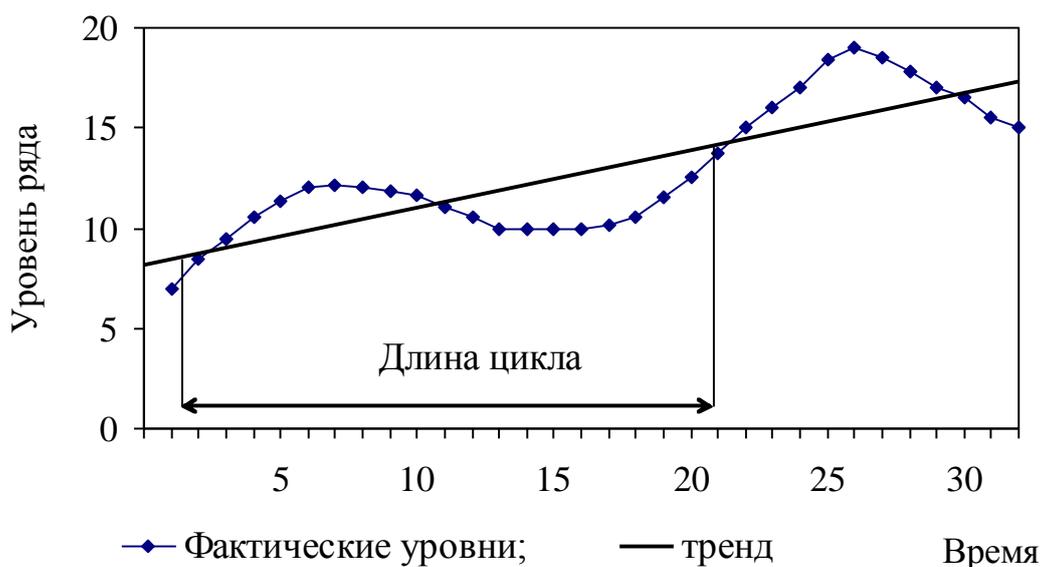


Рис. 6.2. Циклическая долгопериодическая колеблемость

Для прогнозирования циклическая колеблемость благоприятна, особенно, если длина цикла строго постоянна. Прогноз на любой будущий период состоит из прогноза тренда и циклического отклонения от него, соответствующего фазе цикла в прогнозируемый период. Например, зная, что солнечная активность имеет 10–11-летнюю периодичность, и что предыдущий цикл имел максимум в 1990–1991 гг., можно уверенно прогнозировать следующий максимум на 2000–2001 гг.

Как правило, за цикл наблюдаются два экстремума отклонений от тренда – один максимум и один минимум. Следовательно, за период, состоящий из N уровней, насчитывается экстремумов:

$$K = 2 \frac{N}{l}, \quad (6.2)$$

где l – длина цикла.

Причиной циклической колеблемости является какая-либо основная сила, влияющая на уровень изучаемого явления. Иначе говоря, есть главный фактор, вызывающий колебания. Сезонные колебания температуры, осадков, а, следовательно, также и производства, и потребления многих видов продукции зависят от одного фактора – наклона Земной оси к плоскости орбиты Земли. Причина циклической колеблемости солнечной активности пока науке неизвестна.

Распознать циклическую долгопериодическую колеблемость можно по виду графика, подсчетом числа экстремумов в ряду отклонений от тренда и по коэффициенту автокорреляции отклонений I порядка. Если число локальных экстремумов в ряду отклонений мало, то можно предположить наличие циклической колеблемости. Поскольку отклонения одного и того же знака следуют подряд, их произведения являются положительными числами, а отрицательные произведения встречаются лишь дважды за цикл – при пресечении графиком фактического ряда уровней тренда вниз и вверх. Следовательно, коэффициент автокорреляции при долгопериодической колеблемости – величина положительная, стремящаяся к +1 при $l \rightarrow \infty$. При наличии фактического коэффициента больше чем +0,3 можно считать, что в общей колеблемости временного ряда есть существенная циклическая составляющая, а при $\tau_U^{al} > 0,7 - 0,6$ циклическая составляющая является главной.

Для нахождения длины цикла, особенно, если цикличность не строгая, а «квази», нужно последовательно вычислить коэффициенты автокорреляции отклонений от тренда разных порядков, т.е. с лагом 1, 2, 3 и т.д. периодов времени. Наибольший по алгебраической величине коэффициент автокорреляции отметит длину цикла.

6.1.3. Случайно распределенная во времени колеблемость

Характерной чертой данного типа колебаний является хаотичность последовательности отклонений: после отрицательного отклонения от тренда может следовать снова отрицательное или даже два-три, а может и положительное (два-три). Это как бы мелкие «куски» пилообразной и циклической колеблемости разных длин цикла, перемешанные друг с другом. Иногда случайно распределенную колеблемость и называют «интерференцией колебаний» (термин, заимствованный из физики).

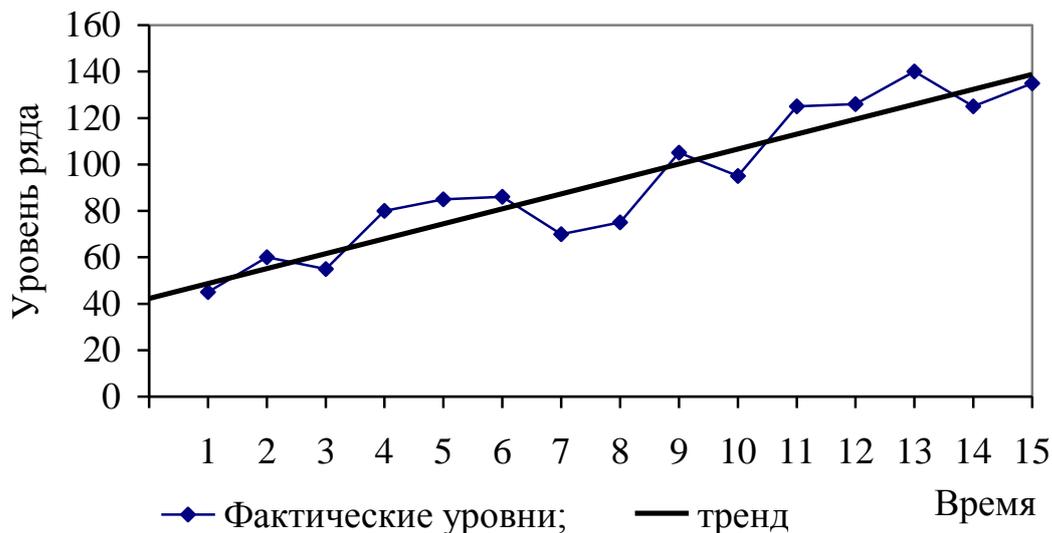


Рис. 6.3. Случайно распределенная во времени колеблемость.

Для колеблемости, изображенной на рис. 6.3, характерны два свойства:

- из-за хаотического чередования знаков отклонений от тренда, их взаимопогашение наступает только на достаточно длительном периоде, а на коротких отрезках отклонения могут аккумулироваться,

например, могут быть три неурожайных года подряд, или два-три высокоурожайных. Значит, необходимы довольно значительные резервы, страховые запасы для гарантии от колебаний.

– случайно распределенная во времени колеблемость неблагоприятна для прогнозирования, ибо в любом прогнозируемом периоде может осуществиться с равной вероятностью как положительное, так и отрицательное отклонение от тренда. (Как увидим в гл. 10, прогнозировать можно лишь интервал, в котором с заданной вероятностью может оказаться уровень.)

Причиной случайно распределенных колебаний служит наличие большого комплекса независимых или слабо связанных между собой факторов, влияющих на уровни изучаемого явления. Так, колебания урожайности зависят от осадков в разные периоды роста культур, от температуры воздуха и почвы, от силы ветра, от развития вредных насекомых, болезнетворных микроорганизмов, от соблюдения агротехники, от качества семян и еще от многих других факторов. Практика статистических исследований колеблемости урожаев показала, что преобладают именно случайно распределенные колебания. Наличие множества примерно равноправных и независимых факторов означает также, что нельзя существенно уменьшить колеблемость, воздействуя только на какой-либо отдельный фактор. Необходимо, если это возможно, регулировать все основные факторы, как, например, и делается в защищенном грунте (теплицах).

Распознать случайно распределенную во времени колеблемость по виду графика труднее, чем два других типа колебаний. Число локальных экстремумов может также колебаться. В среднем, как доказал

английский статистик М. Кендэл⁹, их число составляет $2/3(n-2)$, при среднем квадратическом отклонении, равном $\sqrt{\frac{16n-29}{90}}$. Ряд, изображенный на рис. 6.3. имеет 10 локальных экстремумов (точек перегиба ломаной линии), при $2/3(15-2)=8,7$, и среднем квадратичном отклонении, равном $\sqrt{\frac{16 \cdot 15 - 29}{90}} = 1,53$. Как видим, фактическое число экстремумов попадает в интервал $\bar{x} \pm \sigma$, т.е. вероятность того, что распределение отклонений от тренда является случайным, довольно велика, следовательно, эта гипотеза не может быть отклонена.

Коэффициент автокорреляции отклонений от тренда при случайно распределенной колеблемости стремится к нулю при $n \rightarrow \infty$. Если ряд состоит менее чем из 19–22 уровней, коэффициенты автокорреляции I порядка, не превышающие 0,3 по абсолютной величине, свидетельствуют о преобладании случайной компоненты в общем комплексе колебаний. В случае, изображенном на рис. 6.3, $r_U^a = -0,025$.

6.2. Измерение показателей силы и интенсивности колебаний

Показатели силы и интенсивности колебаний аналогичны таковым же по построению, по форме показателям силы и интенсивности вариации признака в пространственной совокупности. По существу они отличаются тем, что показатели вариации вычисляются на основе отклонений от постоянной средней величины, а показатели, характеризующие колеблемость уровней временного ряда – по отклонениям отдельных уровней от тренда, который можно считать «подвижной

⁹ Кендэл М. Временные ряды: Пер. с англ. – М.: Финансы и статистика, 1981.

средней величиной».

6.2.1. Показатели абсолютной величины (силы) колебаний

Первый показатель – амплитуда (размах) колебаний – разность между наибольшим и наименьшим по алгебраической величине отклонениями от тренда. Например, размах колебаний численности населения России в 1900 – 2004 г. (см. табл. 5.6) составил: $148704 - 70200 = 78504$ тыс. человек.

Показатель амплитуды колебаний характеризует лишь крайние пределы, но не среднюю силу колеблемости. Чем длиннее ряд, тем больше вероятность того, что в нем встретится особенно большое отклонение от тренда. Поэтому, с увеличением длины изучаемого периода возрастает в среднем и амплитуда колебаний, в отличие от всех других показателей колеблемости, которые не зависят от длины ряда.

Вторым показателем колеблемости по абсолютной величине (силе) является среднее по модулю отклонение от тренда, которое мы обозначим как $a(t)$:

$$a(t) = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (6.3)$$

Знак t отличает указанный и все последующие показатели от аналогичного среднего по модулю отклонения от постоянной средней величины, меры силы вариации в пространственной совокупности. Средний модуль отклонений измеряется в тех же единицах, что и уровни ряда. Например, согласно данным табл. 5.6, среднее по моду-

лю отклонение от тренда численности населения России в 1900–2004 гг. может составить примерно 3297,3 тыс. человек.

Хотя средний модуль отклонений тренда вполне пригоден как обобщающий показатель силы колебаний за изучаемый период времени, но, как известно, модули имеют и существенные недостатки, в частности, с ними невозможно связать вероятностные законы распределения. Поэтому модули непригодны для прогнозирования достоверных границ возможных колебаний в будущем с заданной вероятностью (см. гл.10).

Чаще всего в качестве *третьего показателя* силы колебаний используется среднее квадратическое отклонение уровней ряда от тренда, обозначаемое $\sigma(t)$ или $S(t)$.

Если речь идет только об измерении колеблемости во временном ряду и не ставится задача оценки силы колебаний вообще в прогнозе на будущее, тогда следует вычислять и использовать обычное среднее квадратическое отклонение:

$$\sigma(t) = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}. \quad (6.4)$$

Если же речь идет о вычислении *оценки генерального показателя колеблемости*, а исходный временной ряд рассматривается как *выборка* из генерального ряда, продолжаемого и в прошлое и в будущее, то следует учитывать потерю степеней свободы колеблемости и применять показатель:

$$S(t) = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - p}}, \quad (6.5)$$

где p – число параметров в уравнении тренда.

Причину учета числа параметров тренда можно проиллюстрировать следующими примерами.

Линейный тренд имеет два параметра – a и b . Если из ряда уровней взять только уровни двух любых периодов, то, как известно из геометрии, прямая точно пройдет через две любые точки, мы увидим только тренд и не увидим никаких колебаний. Аналогично, если оставить от ряда три любых уровня, тренд в форме параболы II порядка, имеющий три параметра, точно пройдет через три точки графика, в результате колеблемость останется «за кадром», так как у нее нет ни одной степени свободы. Поэтому, оценивая генеральное среднее квадратическое отклонение уровней от тренда, нужно учесть потерю степеней свободы колебаний на величину, равную количеству параметров уравнения тренда. Именно такая несмещенная оценка генерального параметра может быть распространена на будущие периоды, т.е. она необходима в прогнозировании (см. гл. 10). Среднее квадратическое отклонение, как известно, входит в формулу нормального закона распределения вероятностей, на его основе можно рассчитывать вероятности ошибок прогнозов и их доверительные границы.

6.2.2. Показатели относительной интенсивности колебаний

Показатели относительной интенсивности вариации рассчитываются как отношение ее абсолютных показателей к постоянной сред-

ней величине, относительной интенсивности колебаний – как отношение индивидуальных отклонений отдельного периода к уровню тренда за этот же период, а обобщающие показатели – как отношения обобщающих показателей силы колебаний за весь ряд к обобщающему показателю уровней ряда – среднему уровню.

Например, мы хотим оценить интенсивность отклонения доходов бюджета Оренбургской области от ее тренда в январе 2002 года. Абсолютное отклонение составило – 208,15 млн. руб., а уровень тренда (см. приложение 2) = 1073,7 млн. руб. Интенсивность отклонения (колебания) равна – $208,15 : 1073,7 = -0,194$ или –19,4 %. В феврале 2006 года отклонение доходов бюджета от тренда по абсолютной величине тоже было значительным: –1164,3 млн. руб. Но в этом году уровень тренда поднялся уже до 2685,8 млн. руб., поэтому интенсивность отклонения составила – $1164,3 : 2685,8 = -0,433$, или – 43,3%. Следовательно, можно сделать вывод, что интенсивность отклонений от тренда усилилась.

Обобщающим показателем интенсивности колебаний служит *коэффициентом колеблемости*, который рассчитывается аналогично коэффициенту пространственной вариации как отношение среднего квадратического отклонения от тренда к среднему уровню временного ряда, который мы обозначаем, для отличия от коэффициента пространственной вариации, как $V(t)$. Его формула такова:

$$V(t) = \frac{S(t)}{y} \quad (6.6)$$

– для оценки генеральной величины и прогнозов или

$$V(t) = \frac{\sigma(t)}{y} \quad (6.7)$$

–для измерения интенсивности колебаний за данный период как изолированный отрезок, без распространения на прошлые и будущие периоды времени.

Величина коэффициента колеблемости также играет важную роль при анализе устойчивости в динамике (см. гл. 8).

В заключение необходимо подчеркнуть, что любая погрешность в определении типа тренда или в расчете его параметров приводит к преувеличению показателей силы и интенсивности колебаний. Так как реальные временные ряды всегда отклоняются от строго линейной, параболической, экспоненциальной или иной любой абстрактно-математической линии, то колеблемость всегда несколько преувеличивается за счет неполного соответствия истинной тенденции динамики какому-либо принятому типу линии тренда.

6.3. Особенности измерения сезонных колебаний.

Сезонными называют колебания, связанные со сменой времен года и повторяющиеся поэтому ежегодно. Связь может быть непосредственной, как, например, связь сезонной смены температур воздуха с объемом товарооборота разных видов одежды и обуви или мороженого. В других случаях связь колебаний изучаемого показателя с временами года опосредована социальными, юридическими и экономическими факторами, как, например, сезонное увеличение средней зарплаты и среднедушевого дохода в декабре (13-ая зарплата, премии по итогам годовой деятельности, распределение доходов к Новому году и Рождеству и т.п.). Таковы же сезонные колебания числа

браков, приурочиваемых традицией к тем или иным праздникам.

6.3.1 Статистические методы оценки уровня сезонности.

Прогнозирование с помощью тренд – сезонных моделей

По временным рядам за лет в помесечном или поквартальном разрезе могут наблюдаться сезонные колебания.

Сезонные колебания – это разновидность периодических колебаний. Для них характерны внутригодовые, повторяющиеся устойчиво из месяца в месяц (из квартала в квартал) изменения в уровнях, т.е. это регулярно повторяющиеся подъемы и снижение уровней временного ряда внутри года на протяжении ряда лет.

Существует 2 модели сезонности: аддитивная и мультипликативная.

В аддитивной модели сезонность выражается в виде абсолютной величины, которая добавляется или вычитается из среднего значения ряда, чтобы выделить показатель сезонности.

В мультипликативной модели сезонность выражена как процент от среднего уровня, который должен быть учтен при прогнозировании путем умножения на него среднего значения ряда.

Методика построения аддитивной и мультипликативной модели различается в зависимости от того, есть или нет тенденций в ряду динамики¹⁰.

Если во временном ряду отсутствует тенденция, то уровень ряда рассматривается как функция сезонности и случайности (рисунок 6.4):

¹⁰ Статистика: учебник / И.И. Елисева [и др.]; под ред. проф. И.И. Елисеевой. – М.: КНОРУС, 2006.- 552 с.

$$y_i = f(s, \varepsilon), \quad (6.8)$$

где y_i - фактические уровни временного ряда;

S – сезонная составляющая;

ε - случайная компонента.

При аддитивной модели уровень такого ряда можно представить как:

$$y_i = \bar{y} + S + \varepsilon \quad (6.9)$$

тогда: $(y_i - \bar{y}) = (\bar{y}_s - \bar{y}) + (y_i - \bar{y}_s)$

где \bar{y}_s - средний уровень ряда соответствующего периода внутри года (месяца, квартала) за ряд лет.

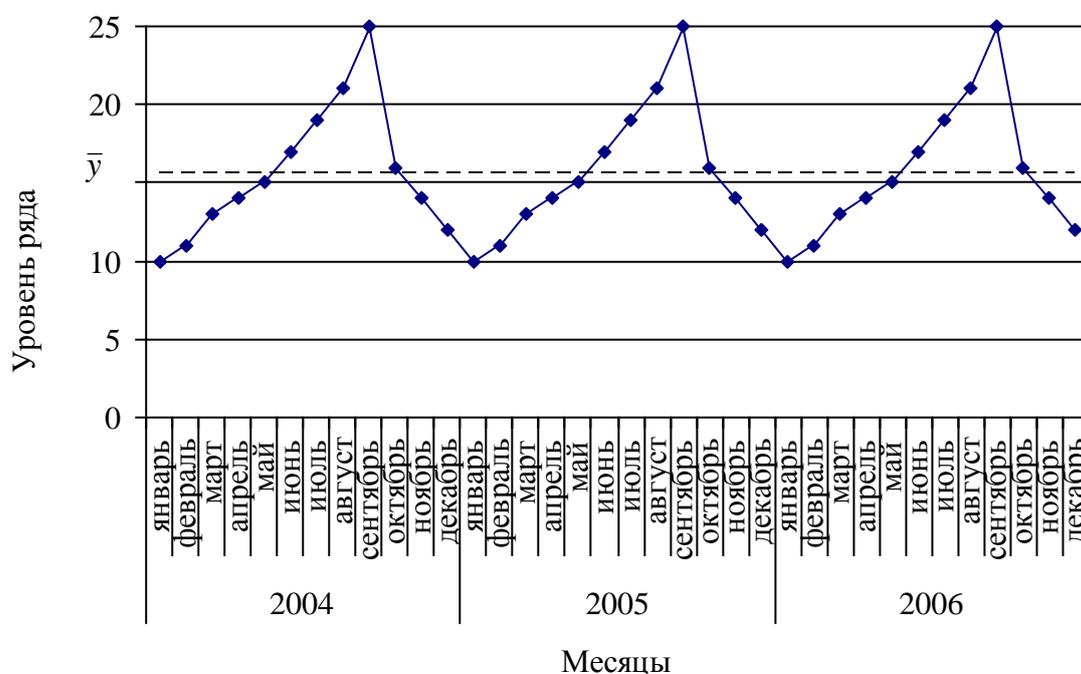


Рис. 6.4 Стационарный временной ряд с сезонными колебаниями

Величина $(\bar{y}_s - \bar{y})$ отражает влияние сезонности (сезонная составляющая S), а величина $(y_i - \bar{y}_s)$ характеризует влияние случайной компоненты.

При мультипликативной модели уровень динамического ряда можно представить как произведение его составляющих:

$$y_i = \bar{y} \cdot \frac{\bar{y}_s}{\bar{y}} \cdot \frac{y_i}{\bar{y}_s} \quad (6.10),$$

где $\frac{\bar{y}_s}{\bar{y}}$ - коэффициент сезонности (K_s),

$\frac{y_i}{\bar{y}_s}$ - отражает влияние случайного фактора.

Чем больше коэффициент сезонности, тем больше амплитуда колебаний уровней ряда относительно его среднего уровня, тем существеннее влияние сезонности. Чем меньше влияние случайной составляющей, тем в большей мере рассматриваемая модель адекватно описывает исходный временной ряд¹¹.

Прогнозирование динамического ряда с сезонными колебаниями при отсутствии в нем тенденции сводится к прогнозированию среднего уровня (\bar{y}_p) с последующей корректировкой его на сезонную компоненту:

$$y_p = \bar{y}_p \pm S \quad (6.11)$$

- аддитивная модель;

¹¹ Статистика: учебник / И.И. Елисеева [и др.]; под ред. проф. И.И. Елисеевой. – М.: КНОРУС, 2006.- 552 с.

$$y_p = \bar{y}_p \cdot K_s \quad (6.12)$$

- мультипликативная модель.

Значительно распространена ситуация, когда динамический ряд имеет тенденцию.

В этом случае уровень временного ряда рассматривается как функция тенденции (t), сезонности (S), и случайности (ε). Тогда аддитивная модель уровня динамического ряда примет вид:

$$y_i = \hat{y}_t + S + \varepsilon, \quad (6.13)$$

где \hat{y}_t - теоретическое значение уровня ряда согласно тенденции;

S – сезонная составляющая;

ε - случайная компонента.

Общая колеблемость уровней временного ряда раскладывается на 3 составляющие:

$$(y_i - \bar{y}) = (\hat{y}_t - \bar{y}) + (y_s - \hat{y}_t) + (y_i - y_s) \quad (6.14)$$

где $(y_i - \bar{y})$ - общая вариация;

$(\hat{y}_t - \bar{y})$ - влияние тенденции;

$(y_s - \hat{y}_t)$ - влияние сезонности;

$(y_i - y_s)$ - влияние случайности;

y_s - тренд с учетом сезонности.

Алгоритм построения тренд – сезонной аддитивной модели

1) сглаживание временного ряда с помощью простой скользящей средней. Период скользящего окна должен быть равен 1 году (если период четный, то проводится центрирование скользящей средней);

2) рассчитывают абсолютные показатели сезонности:

$$S_i = y_i - \tilde{y} \quad (6.15)$$

где \tilde{y} - выровненные скользящие средние);

3) рассчитываются средние показатели сезонности для одноименных кварталов (месяцев):

$$\bar{S}_i = \frac{1}{n} \sum S_i ; \quad (6.16)$$

4) если $\sum \bar{S}_i \neq 0$, проводится корректировка сезонной компоненты:

$$\hat{S}_i = \bar{S}_i - \frac{1}{n} \sum_{i=1}^n \bar{S}_i ; \quad (6.17)$$

5) проводим десеонализацию временного ряда: из исходных уровней вычитаем скорректированную сезонную компоненту:

$$y_i - \hat{S}_i ; \quad (6.18)$$

6) по десеонализованному временному ряду проводим анали-

тическое выравнивание;

7) рассчитываем тренд с учетом сезонности:

$$y_s = \hat{y}_t + \hat{S}_i \quad (6.19).$$

При мультипликативной модели уровень временного ряда можно представить в виде сомножителей:

$$y_i = \hat{y}_t \cdot K_s \cdot E, \quad (6.20)$$

где K_s - коэффициент сезонности;

E – коэффициент влияния случайности $\left(\frac{y_i}{y_s} \right)$.

Рассмотрим пример построения аддитивной тренд – сезонной модели для временного ряда объема продаж.

Графический анализ исходного временного ряда (рисунок 6.5) свидетельствует о наличии трендовой компоненты, характер которой близок к линейному развитию: имеется устойчивая, ярко выраженная тенденция роста объемов продаж.

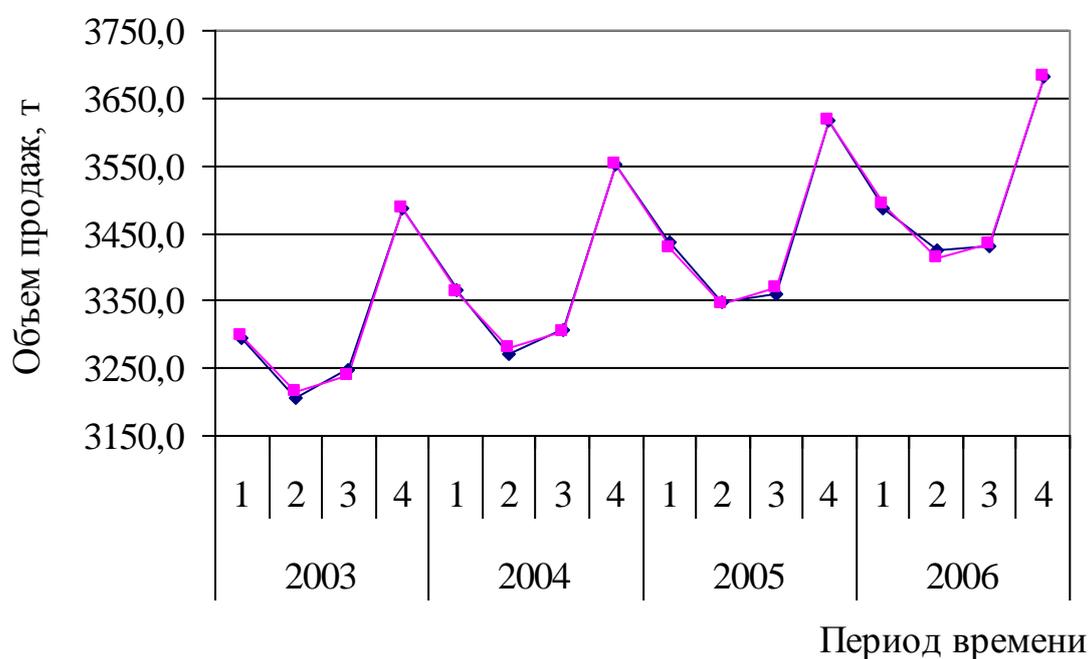
Также отчетливо видны сезонные колебания (период которых равен 1 году). Наиболее существенные «всплески» в динамике показателя просматриваются в 4 квартале. Так как амплитуда сезонных колебаний остается примерно постоянной, то для описания и прогнозирования динамики временного ряда можно использовать аддитивную модель.

Проведем сглаживание временного ряда с помощью центрированной скользящей средней по формуле (период скольжения равен 1

году, т.е. для нашего примера он равен 4):

$$\tilde{y}_i = \frac{1/2 \cdot y_{i-2} + y_{i-1} + y_i + y_{i+1} + 1/2 \cdot y_{i+2}}{4}.$$

$$\tilde{y}_3 = \frac{1/2 \cdot 3294,1 + 3205,9 + 3247,1 + 3488,2 + 1/2 \cdot 3364,7}{4} = 3317,6 \text{ и т.д.}$$



- ◆— фактический ряд объема продаж;
- расчетный объем продаж по тренд - сезонной модели.

Рис.6.5 Квартальная динамика объема продаж

Рассчитаем абсолютные показатели сезонности по формуле (6.15). Результаты расчетов скользящей средней и показателя сезонности представлены в таблице 6.1.

Таблица 6.1

Динамика объема продаж продукции

Годы	№ квартала	y_i , Т	\tilde{y}_i , Т	S_i
2003	1	3294,1	-	-
	2	3205,9	-	-
	3	3247,1	3317,6	-70,6
	4	3488,2	3334,6	153,7
2004	1	3364,7	3350,0	14,7
	2	3270,6	3365,4	-94,9
	3	3305,9	3382,4	-76,5
	4	3552,9	3400,7	152,2
2005	1	3435,3	3416,9	18,4
	2	3347,1	3431,6	-84,6
	3	3358,8	3446,3	-87,5
	4	3617,6	3462,5	155,1
2006	1	3488,2	3480,9	7,4
	2	3423,5	3497,8	-74,3
	3	3429,4	-	-
	4	3682,4	-	-

Определим средние показатели сезонности по формуле (6.16), т.е. для 1 квартала средний показатель сезонности составит:

$$\bar{S}_1 = \frac{14,7 + 18,4 + 7,4}{3} = 13,4804.$$

Аналогично рассчитывают для других кварталов.

Так как сумма средних показателей сезонности не равна нулю,

проведем их корректировку по формуле (6.17).

Скорректированный показатель сезонности для 1 квартала составит:

$$\hat{S}_1 = 13,4804 - \frac{1}{4} \cdot 4,4118 = 12,3775 \text{ и т.д.}$$

Результаты расчетов средних и скорректированных показателей сезонности заносим в таблицу 6.2.

Таблица 6.2

Оценивание сезонной компоненты в аддитивной модели

№ квартала	\bar{S}_j	\hat{S}_j
1	13,4804	12,3775
2	-84,5588	-85,6618
3	-78,1863	-79,2892
4	153,6765	152,5735
Итого	4,4118	0

На следующем этапе определим десеоналированный ряд объема продаж: из исходных уровней вычитаем скорректированную сезонную компоненту: $y_i - \hat{S}_i$. По десеоналированному временному ряду проводим аналитическое выравнивание по линейному тренду и рассчитываем тренд с учетом сезонности: $y_s = \hat{y}_t + \hat{S}_i$.

Уравнение тренда имеет вид: $\hat{y}_t = 3267,2 + 16,442t$ ($R^2 = 0,993$).

Результаты расчетов представлены в таблице 6.3.

Ожидаемый объем продаж в первом полугодии составит:

$$\hat{y}_{sp} = 3559,4 + 3477,8 = 7037,2 \text{ тонн.}$$

Таблица 6.3

**Прогнозирование объема продаж с помощью
аддитивной тренд – сезонной модели**

№ года	№ квартала	t	y_i, T	\hat{S}_j	$y_i - \hat{S}_j$	\hat{y}_i, T	y_s, T
2003	1	1	3294,1	12,3775	3281,7	3283,9	3296,3
	2	2	3205,9	-85,662	3291,5	3300,4	3214,7
	3	3	3247,1	-79,289	3326,3	3316,8	3237,5
	4	4	3488,2	152,574	3335,7	3333,3	3485,8
2004	1	5	3364,7	12,3775	3352,3	3349,7	3362,1
	2	6	3270,6	-85,662	3356,3	3366,2	3280,5
	3	7	3305,9	-79,289	3385,2	3382,6	3303,3
	4	8	3552,9	152,574	3400,4	3399,0	3551,6
2005	1	9	3435,3	12,3775	3422,9	3415,5	3427,9
	2	10	3347,1	-85,662	3432,7	3431,9	3346,3
	3	11	3358,8	-79,289	3438,1	3448,4	3369,1
	4	12	3617,6	152,574	3465,1	3464,8	3617,4
2006	1	13	3488,2	12,3775	3475,9	3481,2	3493,6
	2	14	3423,5	-85,662	3509,2	3497,7	3412,0
	3	15	3429,4	-79,289	3508,7	3514,1	3434,8
	4	16	3682,4	152,574	3529,8	3530,6	3683,1
2007*	1	17*	-	12,3775	-	3547,0	3559,4
	2	18*	-	-85,662	-	3563,5	3477,8
* - прогнозируемый уровень							

**Алгоритм построения тренд – сезонной
мультипликативной модели:**

- 1) сглаживание временного ряда с помощью скользящей средней;
- 2) рассчитываем коэффициент сезонности

$$K_s = \frac{y_i}{\tilde{y}_i} \quad ; \quad (6.21)$$

- 3) определяем средние показатели сезонности для одноименных кварталов (месяцев):

$$\bar{K}_j = \frac{1}{n} \sum K_{si} \quad ; \quad (6.22)$$

- 4) если при поквартальном наблюдении $\sum \bar{K} \neq 4$, а при помесячном $\sum \bar{K} \neq 12$, то выполняется корректировка коэффициента сезонности:

$$\hat{K}_j = \bar{K}_j \cdot \frac{4(12)}{\sum \bar{K}_j} \quad ; \quad (6.23)$$

- 5) исключаем сезонность из уровней ряда:

$$\frac{y_i}{\hat{K}_j} \quad ; \quad (6.24)$$

б) проводится аналитическое выравнивание десеоналированного ряда;

7) рассчитываются уровни временного ряда, обусловленные влиянием тенденции и сезонности:

$$y_s = \hat{y}_t \cdot \hat{K}_j. \quad (6.25)$$

Рассмотрим пример построения мультипликативной тренд – сезонной модели временного ряда объема производства.

Графический анализ исходного временного ряда (рисунок 6.6) свидетельствует о наличии трендовой компоненты, характер которой близок к линейному развитию: имеется устойчивая, ярко выраженная тенденция снижения объема производства. Также отчетливо видны сезонные колебания (период которых равен 1 году). Наиболее существенные «всплески» в динамике показателя просматриваются в 1 квартале. Так как амплитуда сезонных колебаний постепенно уменьшается, то для описания и прогнозирования динамики временного ряда можно использовать мультипликативную модель.

Проведем сглаживание временного ряда с помощью централизованной скользящей средней по формуле (период скольжения равен 1 году, т.е. для нашего примера он равен 4):

$$\tilde{y}_i = \frac{1/2 \cdot y_{i-2} + y_{i-1} + y_i + y_{i+1} + 1/2 \cdot y_{i+2}}{4}.$$

$$\tilde{y}_3 = \frac{1/2 \cdot 92,0 + 73,6 + 58,9 + 80,9 + 1/2 \cdot 88,3}{4} = 75,9 \text{ и т.д.}$$



Рис.6.6. Квартальная динамика объема производства

Рассчитаем коэффициенты сезонности по формуле (6.21). Результаты расчетов скользящей средней и коэффициента сезонности представлены в таблице 6.4.

Определяем средние показатели сезонности для одноименных кварталов (месяцев) по формуле (6.22), т.е. для 1 квартала средний коэффициент сезонности составит:

$$\bar{K}_1 = \frac{1,193 + 1,197 + 1,212}{3} = 1,201.$$

Аналогично рассчитывают для других кварталов.

Таблица 5.4

Разложение уровней ряда по мультипликативной модели

Годы	№ квар- тала	t	y_i , тыс. т	\tilde{y}_i , тыс. т	K_s	\hat{K}_j	$\frac{y_i}{\hat{K}_j}$	\hat{y}_i , тыс. т	y_s , тыс. т
2003	1	1	92,0	-	-	1,200	76,6	79,3	95,2
	2	2	73,6	-	-	0,948	77,7	77,8	73,7
	3	3	58,9	75,9	0,776	0,785	75,0	76,2	59,8
	4	4	80,9	75,0	1,080	1,068	75,8	74,7	79,8
2004	1	5	88,3	74,0	1,193	1,200	73,6	73,1	87,8
	2	6	69,9	72,7	0,962	0,948	73,8	71,6	67,8
	3	7	55,2	70,8	0,779	0,785	70,3	70,1	55,0
	4	8	73,6	69,0	1,067	1,068	68,9	68,5	73,2
2005	1	9	80,9	67,6	1,197	1,200	67,4	67,0	80,4
	2	10	62,5	66,2	0,944	0,948	66,0	65,4	62,0
	3	11	51,5	64,4	0,800	0,785	65,6	63,9	50,1
	4	12	66,2	62,5	1,059	1,068	62,0	62,3	66,6
2006	1	13	73,6	60,7	1,212	1,200	61,3	60,8	72,9
	2	14	55,2	58,9	0,938	0,948	58,2	59,2	56,1
	3	15	44,1	-	-	0,785	56,3	57,7	45,3
	4	16	58,9	-	-	1,068	55,1	56,1	59,9
2007*	1	17	-	-	-	-	-	54,6	65,5
	2	18	-	-	-	-	-	53,0	50,3
*- прогнозируемый уровень									

Так как сумма средних коэффициентов сезонности не равна 4, проведем их корректировку по формуле (6.23). Так скорректирован-

ный коэффициент сезонности для 1 квартала составит:

$$\hat{K}_j = 1,201 \cdot \frac{4}{4,002} = 1,200 \text{ и т.д.}$$

Результаты расчетов средних и скорректированных показателей сезонности заносим в таблицу 6.5.

Таблица 6.5

Оценивание сезонной компоненты в мультипликативной модели

№ квартала	\bar{K}_j	\hat{K}_j
1	1,201	1,200
2	0,948	0,948
3	0,785	0,785
4	1,068	1,067
Итого	4,002	4,000

На следующем этапе определим десезонализованный ряд объема производства по формуле (6.24). По десезонализованному временному ряду проводим аналитическое выравнивание по линейному тренду и рассчитываем тренд с учетом сезонности (6.25).

Уравнение тренда имеет вид:

$$\hat{y}_t = 80,881 - 1,5467t \quad (R^2 = 0,9723).$$

Результаты расчетов представлены в таблице 6.4.

Ожидаемый объем производства в первом полугодии составит:

$$\hat{y}_{sp} = 65,5 + 50,3 = 115,8 \text{ тыс. тонн.}$$

Аддитивная модель целесообразна, если размах сезонных колебаний изменяется слабо.

6.3.2. Исследование периодических колебаний методами спектрального анализа

При наличии периодических колебаний в ряду динамики модель прогноза должна учитывать эти колебания. С этой целью может быть использован ряд Фурье.

Рассмотрим график временного ряда с периодическими колебаниями (рисунок 6.7).

Уровни временного ряда варьируют вокруг среднего значения (\bar{y}), при этом эти колебания (волны) повторяются, т.е. перед нами периодический временной ряд. Интервал времени, необходимый для того, чтобы временной ряд начал повторяться, называется периодом (P).

Величина, обратная периоду, называется частотой (f).

$$f = 1/p \tag{6.26}$$

Она указывает число повторений цикла в единицу времени.

Отклонения от среднего уровня до пика (или впадины) называется амплитудой временного ряда (A).

Расстояние между началом отсчета времени (точкой в которой $t=0$) и ближайшим пиковым значением называется фазой (Ф).

Стационарный периодический временной ряд, можно задать

четырьмя параметрами: периодом или частотой, амплитудой, фазой и средним значением:

$$y_t = \bar{y} + A \cdot \cos(\omega t - \Phi), \quad (6.27)$$

которая называется гармоническим представлением.

ω (омега) – угловая частота, измеряемая в радианах в единицу времени и равна:

$$\omega = 2\pi f; \quad 0 \leq \omega \leq 2\pi. \quad (6.28)$$

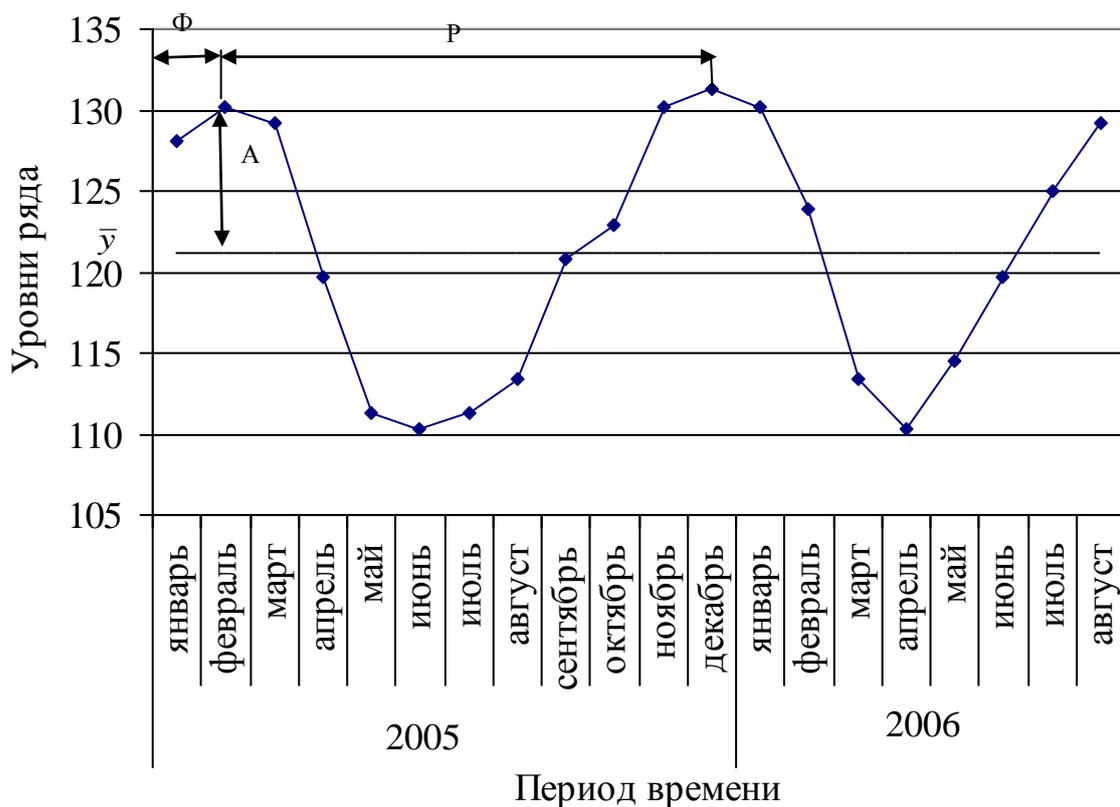


Рис.6.7 Периодический временной ряд

Данное выражение часто записывают через синусы и косинусы без упоминания о фазе:

$$y_t = \bar{y} + a \cos \omega t + b \sin \omega t, \quad (6.29)$$

где $a = A \cos \Phi$, $b = A \sin \Phi$,

т.к. $\cos^2 x + \sin^2 x = 1$ то $a^2 + b^2 = A^2$,

кроме того, т.к. $\operatorname{tg} x = \frac{\sin x}{\cos x}$, то $\operatorname{tg} \Phi = \frac{\sin \Phi}{\cos \Phi} = \frac{b}{a}$ или $\operatorname{arctg} \frac{b}{a} = \Phi$.

Теоретически любой стационарный временной ряд может быть представлен как сумма среднего значения и ряда синусоид и косинусоид, что и называется рядом Фурье:

$$y_t = \bar{y} + \sum_{i=1}^{\infty} a_i \cos \omega_i t + \sum_{i=1}^{\infty} b_i \sin \omega_i t \quad (6.30)$$

Так как реальный ряд динамики имеет конечную длину N , то ряд Фурье приобретает вид:

$$y_t = \bar{y} + \sum_{i=1}^n a_i \cos \omega_i t + \sum_{i=1}^n b_i \sin \omega_i t \quad (6.31)$$

где $n=N/2$ (N - длина временного ряда).

При этом \bar{y} заменяется параметром a_0 , т.е. в окончательном виде имеем:

$$y_t = a_0 + \sum_{i=1}^n a_i \cos \omega_i t + \sum_{i=1}^n b_i \sin \omega_i t \quad (6.32)$$

Оценка параметров данного уравнения при компьютерной об-

работке обычно дается традиционным МНК. Рассмотрим систему нормальных уравнений для случая одной гармоники:

$$y_t = a_0 + a_1 \cos t + b_1 \sin t, \quad (6.33)$$

где t принимает значения от 0 с постоянным увеличением на $2\pi/N$.

Система нормальных уравнений:

$$\begin{cases} Na_0 + a_1 \sum \cos t + b_1 \sum \sin t = \sum y_t \\ a_0 \sum \cos t + a_1 \sum \cos^2 t + b_1 \sum \sin t \cos t = \sum y \cos t \\ a_0 \sum \sin t + a_1 \sum \cos t \sin t + b_1 \sum \sin^2 t = \sum y \sin t \end{cases} \quad (6.34)$$

Т.к. $\sum \sin t \cos t = \sum \cos t = \sum \sin t = 0$, то из 1-го уравнения получаем, что

$$a_0 = \sum y_t / N = \bar{y} \quad (6.35)$$

из второго уравнения:

$$a_1 = \frac{\sum y \cos t}{\sum \cos^2 t} \quad (6.36)$$

из третьего уравнения:

$$b_1 = \frac{\sum y \cdot \sin t}{\sum \sin^2 t}, \quad (6.37)$$

т.к. $\cos^2 t = \frac{1 + \cos 2t}{2}$, то $\sum \cos^2 t = \frac{N}{2}$ соответственно и $\sum \sin^2 t = N/2$.

Следовательно, $a_1 = \frac{2\sum y \cdot \cos t}{N}$ и $b_1 = \frac{2\sum y \cdot \sin t}{N}$. При определении второй гармоники рассчитываются a_2 и b_2 :

$$a_2 = \frac{2}{N} \sum y \cos 2t; \quad (6.38)$$

$$b_2 = \frac{2}{N} \sum y \sin 2t. \quad (6.39)$$

Иными словами, параметры ряда Фурье определяется следующим образом:

$$a_i = \frac{2}{N} \sum y \cos \omega_i t \quad (6.40)$$

и

$$b_i = \frac{2}{N} \sum y \sin \omega_i t \quad (6.41)$$

Часто хорошее описание фактического временного ряда достигается с использованием двух гармоник¹²:

$$y_t = a_0 + a_1 \cos t + b_1 \sin t + a_2 \cos 2t + b_2 \sin 2t \quad (6.42)$$

¹² Статистика: учебник / И.И. Елисева [и др.]; под ред. проф. И.И. Елисеевой. – М.: КНОРУС, 2006.- 552 с.

Применим спектральный анализ временного ряда производства продукции (тыс. тонн). Так как разложение по ряду Фурье применяется только для стационарных временных рядов необходимо проверить гипотезу о наличии тенденции либо применить графический анализ.

По виду графика анализируемого временного ряда можно сделать вывод о его стационарности (рисунок 6.8).

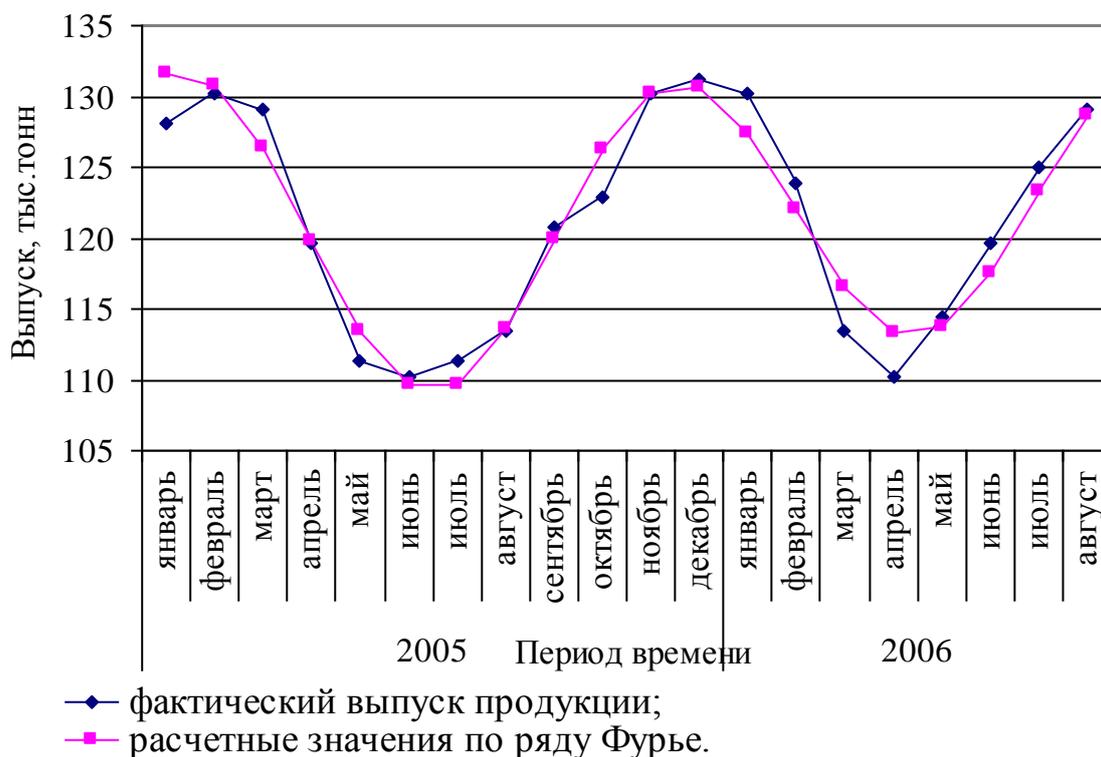


Рис. 6.8 Динамика объема производства продукции

Построим ряд Фурье с двумя гармониками.

Отсчет t ведется с нуля, с шагом $\frac{2\pi}{N}$, т.е. в нашем случае $0,1\pi$.

Параметры ряда Фурье определяются по формулам:

$$a_i = \frac{2}{N} \sum y \cos \omega_i t, \quad b_i = \frac{2}{N} \sum y \sin \omega_i t$$

Таблица 6.6

Расчет параметров по ряду Фурье

Годы	месяц	y_t	t	$\cos t$	$\sin t$	$\cos 2t$	$\sin 2t$	\tilde{y}_t
1	2	3	4	5	6	7	8	13
2005	январь	128,1	0	1,000	0,000	1,000	0,000	131,7
	февраль	130,2	0,314	0,951	0,309	0,809	0,588	130,8
	март	129,2	0,628	0,809	0,588	0,309	0,951	126,3
	апрель	119,7	0,942	0,588	0,809	-0,309	0,951	119,7
	май	111,3	1,257	0,309	0,951	-0,809	0,588	113,4
	июнь	110,3	1,571	0,000	1,000	-1,000	0,000	109,6
	июль	111,3	1,885	-0,309	0,951	-0,809	-0,588	109,7
	август	113,4	2,199	-0,588	0,809	-0,309	-0,951	113,6
	сентябрь	120,8	2,513	-0,809	0,588	0,309	-0,951	119,9
	октябрь	122,9	2,827	-0,951	0,309	0,809	-0,588	126,2
	ноябрь	130,2	3,142	-1,000	0,000	1,000	0,000	130,3
	декабрь	131,3	3,456	-0,951	-0,309	0,809	0,588	130,7

Продолжение таблицы 6.6

1	2	3	4	5	6	7	8	13
2006	январь	130,2	3,769	-0,809	-0,588	0,309	0,951	127,4
	февраль	123,9	4,084	-0,588	-0,809	-0,309	0,951	122,0
	март	113,4	4,398	-0,309	-0,951	-0,809	0,588	116,5
	апрель	110,3	4,712	0,000	-1,000	-1,000	0,000	113,4
	май	114,5	5,027	0,309	-0,951	-0,809	-0,588	113,7
	июнь	119,7	5,340	0,588	-0,809	-0,309	-0,951	117,5
	июль	125,0	5,655	0,809	-0,588	0,309	-0,951	123,3
	август	129,2	5,969	0,951	-0,309	0,809	-0,588	128,7
Итого		2424,5	59,690	0,000	0,000	0,000	0,000	2424,5
Прогноз на сентябрь 2006 г.		-	6,283	1,000	0,000	1,000	0	132,7

Так как мы строим ряд с двумя гармониками, нам необходимо рассчитать параметры: $a_1 = \frac{2 \sum y \cdot \cos t}{N}$; $a_2 = \frac{2}{N} \sum y \cos 2t$; $b_1 = \frac{2 \sum y \cdot \sin t}{N}$; $b_2 = \frac{2}{N} \sum y \sin 2t$. Необходимые расчеты занесем в таблицу 6.6.

Подставим соответствующие значения в формулы и получим:

$$a_0 = 121,223,$$

$$a_1 = 0,700,$$

$$a_2 = 9,753,$$

$$b_1 = -1,884,$$

$$b_2 = 2,791.$$

Тогда ряд Фурье с двумя гармониками примет вид:

$$\tilde{y}_t = 121,222 + 0,700 \cdot \cos t - 1,885 \cdot \sin t + 9,753 \cdot \cos 2t + 2,791 \cdot \sin 2t.$$

Чтобы получить прогноз на следующий месяц (сентябрь 2005г.) необходимо подставить в уравнение следующее значение $t = 2\pi$. Таким образом, объем производства в сентябре составит 132,7 тыс.т.

6.3.3. Моделирование сезонных колебаний с помощью фиктивных переменных

Рассмотрим ещё один метод моделирования временного ряда, содержащего сезонные колебания, - построение модели регрессии с

включением фактора времени и фиктивных переменных¹³.

Количество фиктивных переменных в такой модели должно быть на единицу меньше числа моментов (периодов) времени внутри одного цикла колебаний. Каждая фиктивная переменная отражает сезонную (циклическую) компоненту временного ряда, для какого – либо одного периода. Она равна 1 для данного периода и нулю для всех остальных.

Пусть имеется временной ряд, содержащий циклические колебания периодичностью K . Модель регрессии с фиктивными переменными для этого ряда:

$$y_t = a + bt + c_1x_1 + \dots + c_jx_j + \dots + c_{k-1}x_{k-1} + \varepsilon_t, \quad (6.43)$$

$$\text{где } x_j = \begin{cases} 1 & \text{для каждого } j \text{ внутри каждого цикла} \\ 0 & \text{во всех остальных случаях} \end{cases}$$

Например, при моделировании сезонных колебаний на основе поквартальных данных за несколько лет число кварталов внутри одного года $K=4$, а общий вид модели:

$$y_t = a + bt + c_1x_1 + c_2x_2 + c_3x_3 + \varepsilon_t, \quad (6.44)$$

$$\text{где } x_1 = \begin{cases} 1 & \text{для 1 квартала} \\ 0 & \text{во всех остальных случаях} \end{cases} \quad (6.45)$$

¹³ Эконометрика: учебник / И.И. Елисеева [и др.]. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2006. – 576 с.

$$x_2 = \begin{cases} 1 & \text{для 2 квартала} \\ 0 & \text{во всех остальных случаях} \end{cases} \quad (6.46)$$

$$x_3 = \begin{cases} 1 & \text{для 3 квартала} \\ 0 & \text{во всех остальных случаях} \end{cases} \quad (6.47)$$

Уравнение тренда для каждого квартала будет иметь следующий вид:

$$\text{– для 1 квартала: } y_t = a + bt + c_1 + \varepsilon_t; \quad (6.48)$$

$$\text{– для 2 квартала: } y_t = a + bt + c_2 + \varepsilon_t; \quad (6.49)$$

$$\text{– для 3 квартала: } y_t = a + bt + c_3 + \varepsilon_t; \quad (6.50)$$

$$\text{– для 4 квартала: } y_t = a + bt + \varepsilon_t. \quad (6.51)$$

Таким образом, фиктивные переменные позволяют дифференцировать величину свободного члена уравнения регрессии для каждого квартала. Она составит:

$$\text{– для 1 квартала } (a + c_1);$$

$$\text{– для 2 квартала } (a + c_2);$$

$$\text{– для 3 квартала } (a + c_3);$$

$$\text{– для 4 квартала } a.$$

Параметр b в этой модели характеризует среднее абсолютное изменение уровней ряда под воздействием тенденции¹⁴.

¹⁴ Эконометрика: учебник / И.И. Елисеева [и др.]. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2006. – 576 с.

Таблица 6.7

**Исходные данные для расчета параметров уравнения регрессии с
фиктивными переменными
во временном ряду объема производства, (т)**

Годы	№ квартала	y	t	x ₁	x ₂	x ₃	\hat{y}_t			
							1 кв.	2 кв.	3 кв.	4 кв.
2003	1	92,0	1	1	0	0	92,9			
	2	73,6	2	0	1	0		74,5		
	3	58,9	3	0	0	1			61,7	
	4	80,9	4	0	0	0				79,1
2004	1	88,3	5	1	0	0	86,8			
	2	69,9	6	0	1	0		68,4		
	3	55,2	7	0	0	1			55,5	
	4	73,6	8	0	0	0				73,0
2005	1	80,9	9	1	0	0	80,6			
	2	62,5	10	0	1	0		62,2		
	3	51,5	11	0	0	1			49,3	
	4	66,2	12	0	0	0				66,8
2006	1	73,6	13	1	0	0	74,5			
	2	55,2	14	0	1	0		56,1		
	3	44,1	15	0	0	1			43,2	
	4	58,9	16	0	0	0				60,7
2007*	1		17				59,1			
	2		18					57,6		
* - прогноз										

Спрогнозируем объем производства с помощью модели регрессии с включением фактора времени и фиктивных переменных. Модель для квартальной динамики имеет вид:

$$y_t = a + bt + c_1x_1 + c_2x_2 + c_3x_3 + \varepsilon_t,$$

$$\text{где } x_1 = \begin{cases} 1 & \text{для 1 квартала} \\ 0 & \text{во всех остальных случаях} \end{cases}$$

$$x_2 = \begin{cases} 1 & \text{для 2 квартала} \\ 0 & \text{во всех остальных случаях} \end{cases}$$

$$x_3 = \begin{cases} 1 & \text{для 3 квартала} \\ 0 & \text{во всех остальных случаях} \end{cases}$$

Занесем значение фиктивных переменных и фактора времени в таблицу (таблица 6.7).

Оценим параметры уравнения традиционным МНК с помощью ППП Excel (таблица 6.8).

Уравнение регрессии примет вид:

$$\hat{y}_t = 85,3 - 1,54 \cdot t + 9,17 \cdot x_1 - 7,68 \cdot x_2 - 19,02 \cdot x_3.$$

Параметры c_1, c_2, c_3 характеризуют отклонения уровней временного ряда от уровней, учитывающих сезонные воздействия в 4 квартале. Величина параметра $b = -1,54$ говорит о том, что в среднем

за квартал происходит снижение объема производства на 1,54 тонны (рисунок 6.9).

Таблица 6.8

Результаты оценивания регрессионной модели с фиктивными переменными

Регрессионная статистика

Множественный R	0,99
R-квадрат	0,99
Нормированный R-квадрат	0,99
Стандартная ошибка	1,62
Наблюдения	16

Дисперсионный анализ

	df	SS	MS	F	Значимость F
Регрессия	4	2758,4	689,6	262,2	0,000
Остаток	11	28,93	2,6		
Итого	15	2787,2			

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
a	85,30	1,22	70,14	0,00	82,63	87,98
b	-1,54	0,09	-16,99	0,00	-1,74	-1,34
c1	9,17	1,18	7,78	0,00	6,58	11,77
c2	-7,68	1,16	-6,61	0,00	-10,23	-5,12
c3	-19,02	1,15	-16,53	0,00	-21,55	-16,48

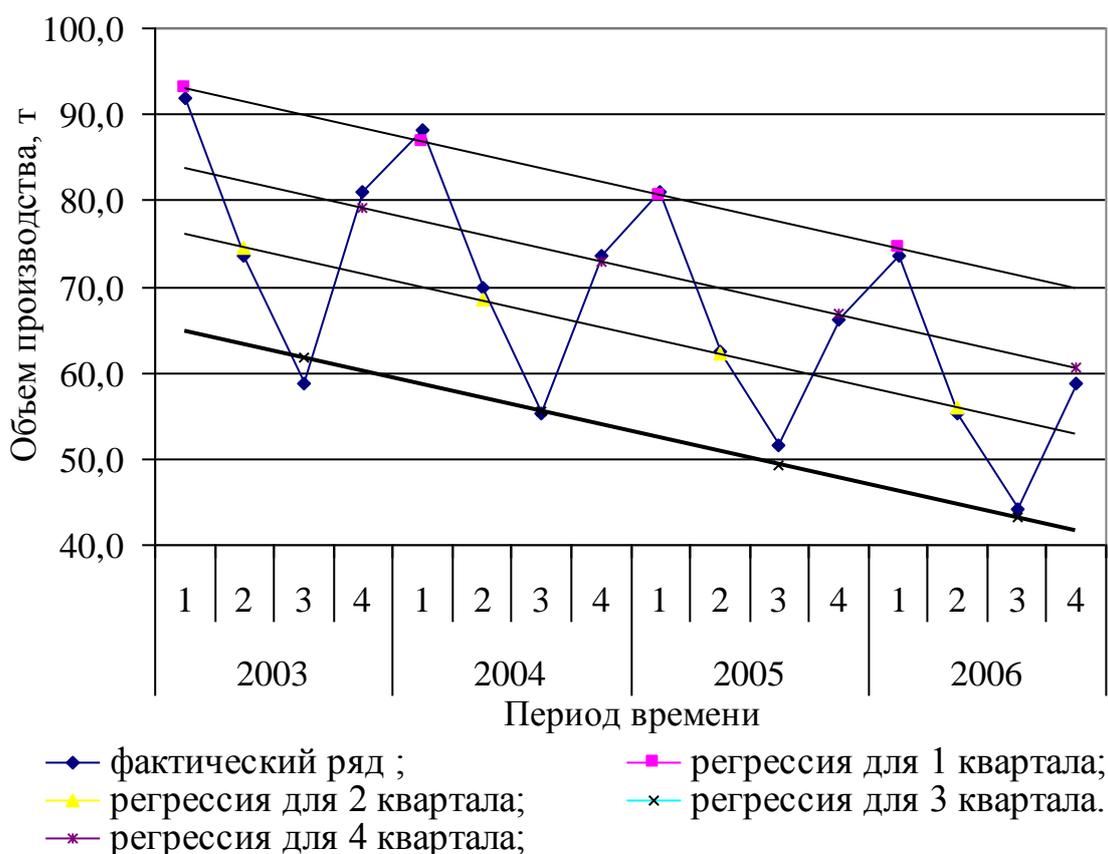


Рис.6.9 Моделирование сезонных колебаний объема производства с помощью фиктивных переменных

Уравнение тренда для каждого квартала будет иметь следующий вид:

– для 1 квартала: $\hat{y}_t = 94,48 - 1,54t$;

– для 2 квартала: $\hat{y}_t = 77,62 - 1,54t$;

– для 3 квартала: $\hat{y}_t = 66,29 - 1,54t$;

– для 4 квартала: $\hat{y}_t = 85,3 - 1,54t$.

Чтобы получить прогнозные значения объема производства на 1 и 2 кварталы 2007 года необходимо в соответствующее уравнение регрессии подставить следующие значения фактора времени t .

Так прогноз на 1 квартал составит: $\hat{y}_t = 94,48 - 1,54 \cdot 17 = 59,1$ т,
на второй: $\hat{y}_t = 77,62 - 1,54 \cdot 18 = 57,6$ т.

6.4. Измерение тренда колеблемости

Неоднократно указывалось на большое значение мониторинга колебаний. Как правило, производство, экономика заинтересованы в уменьшении колеблемости. Чтобы измерить изменение абсолютного показателя силы колебаний $S(t)$, проще всего рассчитать эту величину за последовательные отрезки времени, а затем по полученным значениям $S(t)_1, S(t)_2$ и т.д. до $S(t)_n$ провести аналитическое выравнивание, т.е. вычислить тренд того или другого типа. Однако, для более или менее надежного вычисления меры колеблемости необходимо, минимально, 7–9 уровней первичного временного ряда, а для вычисления тренда по этим мерам колеблемости, опять же 7–9 таких частных мер $S(t)$. А для этого первичный ряд должен содержать примерно $8 \cdot 8 = 64$ уровня. Такие ряды анализируются нечасто, а значит, нет и условий для расчета тренда мер колеблемости.

Положение отчасти спасает то, что для вычисления тренда колеблемости вовсе необязательно, чтобы за весь изучаемый период существовал единый тренд уровней показателя. Вполне допустимо для расчета тренда колеблемости объединить отрезки времени с разными по типу трендами или с *кусочно-линейным* трендом. От изменения скорости роста, или даже типа роста, или направления тенденции динамики, колеблемость мало или совсем не зависит. Но и с учетом этой ее особенности измерить тренд колеблемости по ряду отдельных отрезков времени сложно. При длине первичного ряда в 15–20 уровней

получается всего 2 значения $S(t)$, чего явно не хватает для расчета тренда.

Не вполне корректным с математической точки зрения является расчет скользящих показателей колеблемости со сдвигом в один период времени и последующее их аналитическое выравнивание. Конечно, скользящие показатели уже не являются независимыми друг от друга, но выявить общую тенденцию изменения силы колебаний и приближенно измерить тренд $S(t)$ все же возможно. Покажем применение этого метода на примере временного ряда доходов бюджета Оренбургской области (см. разд. 5.1). В таблице 6.9 вычислены отклонения уровней от тренда, с которых и начинается измерение тренда среднего квадратического отклонения.

Скользящие показатели колеблемости $S(t)_i$ будем рассчитывать по 11-летним подпериодам, т.е. первый за январь 2001 – ноябрь 2001 гг., второй за февраль 2001 – декабрь 2001 гг. и т.д. Первая величина $S(t)$ будет относиться к середине подпериода, т.е. июлю 2001 года и т.д., последняя скользящая средняя за февраль 2006 г. – декабрь 2006 г. относится к июлю 2006 года. Итого получаем 62 скользящих значений показателя колеблемости, которые и выравниваем по уравнению прямой.

Тренд среднего квадратического отклонения уровней доходов бюджета Оренбургской области от их тренда имеет вид:

$$\hat{S}(t) = 471,19 + 13,03 \cdot t_i;$$

где $t_i = 0,5$ в январе 2004 г.

Таким образом, имеется тенденция повышения силы колебаний

Таблица 6.9

Вычисление тренда показателя колеблемости

Годы	Месяцы	U_i	U_i^2	Скользятые $S(t)_i$	t_i	t_i^2	$S(t)_i t_i$	$\hat{S}(t)_i$	$U_{S(t)_i}$	$U_{S(t)_i}^2$
1	2	3	4	5	6	7	8	9	10	11
2001	январь	440,45	193996,2	-	-	-	-	-	-	-
	февраль	-359,55	129276,2	-	-	-	-	-	-	-
	март	262,25	68775,1	-	-	-	-	-	-	-
	апрель	400,25	160200,1	-	-	-	-	-	-	-
	май	273,95	75048,6	-	-	-	-	-	-	-
	июнь	48,05	2308,8	254,80	-30,5	930,25	-7771,3	73,79	181,01	32764,02
	июль	51,95	2698,8	225,88	-29,5	870,25	-6663,5	86,82	139,06	19338,58
	август	269,25	72495,6	207,87	-28,5	812,25	-5924,2	99,85	108,02	11668,17
	сентябрь	47,35	2242,0	194,99	-27,5	756,25	-5362,1	112,88	82,11	6741,89
	октябрь	-42,75	1827,6	153,16	-26,5	702,25	-4058,8	125,91	27,26	742,89
	ноябрь	72,55	5263,5	179,34	-25,5	650,25	-4573,1	138,94	40,40	1632,09
	декабрь	202,75	41107,6	181,07	-24,5	600,25	-4436,1	151,97	29,10	846,79
2002	январь	-208,15	43326,4	196,43	-23,5	552,25	-4616	165,00	31,43	987,94
	февраль	-108,15	11696,4	183,27	-22,5	506,25	-4123,6	178,03	5,24	27,51
	март	5,65	31,9	188,36	-21,5	462,25	-4049,6	191,06	-2,70	7,29

Продолжение таблицы 6.9

1	2	3	4	5	6	7	8	9	10	11
	апрель	413,25	170775,6	195,05	-20,5	420,25	-3998,5	204,09	-9,04	81,68
	май	95,75	9168,1	199,90	-19,5	380,25	-3898	217,11	-17,22	296,52
	июнь	-257,85	66486,6	192,82	-18,5	342,25	-3567,2	230,14	-37,32	1392,88
	июль	132,45	17543,0	185,51	-17,5	306,25	-3246,4	243,17	-57,66	3325,09
	август	151,75	23028,1	236,74	-16,5	272,25	-3906,2	256,20	-19,46	378,80
	сентябрь	-173,35	30050,2	298,19	-15,5	240,25	-4622	269,23	28,96	838,55
	октябрь	162,25	26325,1	273,83	-14,5	210,25	-3970,5	282,26	-8,43	71,11
	ноябрь	-102,75	10557,6	290,08	-13,5	182,25	-3916,1	295,29	-5,21	27,13
	декабрь	113,55	12893,6	280,08	-12,5	156,25	-3501	308,32	-28,24	797,57
2003	январь	-499,65	249650,1	302,35	-11,5	132,25	-3477	321,35	-19,00	361,16
	февраль	-601,35	361621,8	298,90	-10,5	110,25	-3138,5	334,38	-35,48	1258,71
	март	-132,25	17490,1	301,15	-9,5	90,25	-2860,9	347,41	-46,26	2140,10
	апрель	331,65	109991,7	317,12	-8,5	72,25	-2695,5	360,44	-43,32	1876,59
	май	-61,25	3751,6	321,90	-7,5	56,25	-2414,2	373,47	-51,57	2659,92
	июнь	-400,25	160200,1	328,58	-6,5	42,25	-2135,7	386,50	-57,92	3355,24
	июль	-15,75	248,1	292,19	-5,5	30,25	-1607,1	399,53	-107,34	11521,34
	август	-211,85	44880,4	304,76	-4,5	20,25	-1371,4	412,56	-107,80	11620,48
	сентябрь	-367,35	134946,0	387,57	-3,5	12,25	-1356,5	425,59	-38,02	1445,72
	октябрь	210,05	44121,0	385,19	-2,5	6,25	-962,98	438,62	-53,43	2854,65

Продолжение таблицы 6.9

1	2	3	4	5	6	7	8	9	10	11
	ноябрь	-246,35	60688,3	432,92	-1,5	2,25	-649,38	451,65	-18,73	350,73
	декабрь	-34,75	1207,6	432,17	-0,5	0,25	-216,08	464,68	-32,51	1057,18
2004	январь	-666,45	444155,6	443,78	0,5	0,25	221,892	477,71	-33,93	1150,98
	февраль	-805,05	648105,5	443,57	1,5	2,25	665,353	490,74	-47,17	2225,04
	март	-299,65	89790,1	438,26	2,5	6,25	1095,66	503,77	-65,50	4290,84
	апрель	658,25	433293,1	445,87	3,5	12,25	1560,55	516,80	-70,93	5030,78
	май	-391,15	152998,3	445,47	4,5	20,25	2004,64	529,83	-84,35	7115,46
	июнь	-334,95	112191,5	481,84	5,5	30,25	2650,13	542,86	-61,02	3723,03
	июль	206,85	42786,9	455,42	6,5	42,25	2960,24	555,89	-100,47	10093,47
	август	-288,95	83492,1	552,67	7,5	56,25	4145	568,92	-16,25	264,06
	сентябрь	-343,65	118095,3	619,24	8,5	72,25	5263,55	581,95	37,29	1390,86
	октябрь	-238,35	56810,7	611,18	9,5	90,25	5806,19	594,98	16,20	262,48
	ноябрь	610,05	372161,0	600,83	10,5	110,25	6308,68	608,01	-7,18	51,55
	декабрь	414,45	171768,8	592,28	11,5	132,25	6811,21	621,04	-28,76	826,95
2005	январь	-1314	1726464,6	615,12	12,5	156,25	7689,01	634,07	-18,94	358,91
	февраль	-973,65	947994,3	613,57	13,5	182,25	8283,19	647,10	-33,53	1123,98
	март	569,35	324159,4	628,16	14,5	210,25	9108,3	660,13	-31,97	1021,88
	апрель	-122,45	14994,0	640,98	15,5	240,25	9935,14	673,15	-32,18	1035,45
	май	4,05	16,4	621,22	16,5	272,25	10250,2	686,18	-64,96	4220,24

Продолжение таблицы 6.9

1	2	3	4	5	6	7	8	9	10	11
	июнь	-588,35	346155,7	787,58	17,5	306,25	13782,6	699,21	88,37	7808,42
	июль	-250,05	62525,0	754,53	18,5	342,25	13958,8	712,24	42,29	1788,22
	август	563,35	317363,2	767,56	19,5	380,25	14967,4	725,27	42,28	1787,84
	сентябрь	-485,55	235758,8	826,38	20,5	420,25	16940,7	738,30	88,07	7757,00
	октябрь	-312,85	97875,1	839,02	21,5	462,25	18038,9	751,33	87,68	7688,61
	ноябрь	1658,25	2749793,1	839,81	22,5	506,25	18895,7	764,36	75,45	5692,14
	декабрь	1079,75	1165860,1	875,31	23,5	552,25	20569,8	777,39	97,92	9588,19
2006	январь	-1079,9	1166076,0	872,54	24,5	600,25	21377,1	790,42	82,11	6742,61
	февраль	-1164,3	1355478,1	859,08	25,5	650,25	21906,4	803,45	55,62	3094,02
	март	496,55	246561,9	938,77	26,5	702,25	24877,5	816,48	122,29	14954,85
	апрель	120,95	14628,9	964,40	27,5	756,25	26520,9	829,51	134,89	18194,07
	май	1007,95	1015963,2	859,86	28,5	812,25	24506,1	842,54	17,32	300,06
	июнь	-95,65	9148,9	806,55	29,5	870,25	23793,2	855,57	-49,02	2403,16
	июль	246,95	60984,3	751,05	30,5	930,25	22907,2	868,60	-117,55	13817,17
	август	1346,05	1811850,6	-	-	-	-	-	-	-
	сентябрь	-796,45	634332,6	-	-	-	-	-	-	-
	октябрь	807,55	652137,0	-	-	-	-	-	-	-
	ноябрь	434,25	188573,1	-	-	-	-	-	-	-
	декабрь	463,95	215249,6	-	-	-	-	-	-	-
Итого		-0,9	20451512,2	29214	0	19855,5	258712	29214	0,00	268271

доходов бюджета за рассмотренный период времени. Остается проверить надежность расчета среднегодового снижения величины $S(t)$, т.е. сравнить $b_{S(t)}$ со средней ошибкой репрезентативности. Это необходимо для применения полученного тренда силы колебаний в прогнозировании доходов бюджета, т.е. для распространения выборочной оценки на генеральную совокупность периодов времени.

Для указанной цели придется использовать излагаемую только в главе 7 методику вероятностных оценок параметров:

Средняя ошибка репрезентативности среднегодового изменения – $b_{S(t)}$, т.е.

$$m_{b_{S(t)}} = \frac{S(t)_{S(t)}}{\sqrt{\sum_1^{16} t_i^2}}.$$

Здесь в числителе стоит величина среднего квадратического отклонения скользящих значений $S(t)_i$ от их трендовых значений $\hat{S}(t)_i$ (вторая графа справа в табл. 6.9).

$$\overline{S(t)} = \frac{29214,05}{62} = 471,19; b_{S(t)} = \frac{258711,6}{19855,5} = 13,03 \text{ млн. руб. в год.}$$

Имеем:

$$S(t)_{S(t)} = \sqrt{\frac{\sum_1^{62} U_{S(t)}^2}{62-1}} = \sqrt{\frac{268270,7}{61}} = 66,32$$

$$m_{b_{S(t)}} = \frac{S(t)_{S(t)}}{\sqrt{\sum_1^{62} t_i^2}} = \frac{66,32}{\sqrt{19855,5}} = 5,09$$

Критерий Стьюдента, равен отношению $\frac{b_{S(t)}}{m_{b_{S(t)}}} = \frac{13,03}{5,09} = 2,56$. Таб-

личное значение критерия Стьюдента при 61 степенях свободы вариации и значимости 0,05 составляет 1,999. Фактическое значение критерия больше табличного, можно считать достаточно надежно установленным увеличение колебаний доходов бюджета в Оренбургской области за январь 2001г – декабрь 2006г. (см. также разд. 8.3).

6.5. Автокорреляция отклонений от тренда

Автокорреляция – это корреляция уровней ряда друг с другом, либо отклонений от тренда друг с другом, т.е. корреляция внутри одного и того же временного ряда, но с разными сдвигами во времени. Автокорреляция уровней ряда, если она существенна, говорит о наличии тренда, т.е. служит одним из методов обнаружения тренда. В данном разделе рассматривается автокорреляция отклонений от тренда, как один из способов исследования колеблемости.

Методика состоит из последовательного вычисления коэффициентов автокорреляции отклонений с разными сдвигами во времени. Коэффициент автокорреляции со сдвигом на 1 интервал времени уже был рассмотрен в разд. 6.1. Аналогично строятся и формулы коэффициентов автокорреляции со сдвигом в два, три и т.д. периодов времени. В общем виде коэффициент автокорреляции порядка m , т.е. со сдвигом на m периодов времени, вычисляется по формуле:

$$r_{U_i U_{i+m}} = \frac{\sum_1^{n-m} U_i \cdot U_{i+m}}{\frac{\sum_1^{m-1} U_i^2}{2} + \sum_m^{n-m} U_i^2 + \frac{\sum_{n-m+1}^n U_i^2}{2}}$$

Первые $(m - 1)$ отклонений от тренда и последние $(m - 1)$ отклонений участвуют в произведениях (в числителе) по одному разу, остальные – дважды. Соответственно, в знаменателе первые $(m - 1)$ квадратов и последние $(m - 1)$ квадратов входят с половинным весом в сравнении со средними отклонениями. Рассмотрим пример расчета коэффициентов автокорреляции отклонений от тренда и их значения (табл. 6.10).

Таблица 6.10

**Коэффициенты автокорреляции отклонений от тренда
доходов бюджета Оренбургской области**

Лаг, месяц	1	2	3	4	5	6	7	8	9	10	11	12	13
Коэффициент автокорреляции	0,040	-0,312	0,021	0,203	0,185	-0,119	0,084	0,360	-0,046	-0,485	0,116	0,615	0,083

По данным таблицы можно сделать вывод, что во временном ряду доходов бюджета Оренбургской области наблюдается циклическая составляющая равная 12 месяцам, т.е. сезонность.

Глава 7. ВЕРОЯТНОСТНАЯ ОЦЕНКА СУЩЕСТВЕННОСТИ (НАДЕЖНОСТИ УСТАНОВЛЕНИЯ) ПАРАМЕТРОВ ТРЕНДА И КОЛЕБЛЕМОСТИ

Статистика лишь в виде редкого исключения может вести анализ какого-то процесса от начала до конца. Обычно исходный временной ряд – это лишь выборка во времени, отражающая некоторый этап или просто отрезок развития данного процесса и его показателей. Однако задача исследования может заключаться не только в получении характеристик процесса на ограниченном отрезке времени (показателей выборки), но и в оценке генеральных параметров процесса (показателей гипотетической генеральной совокупности). Например, проведен анализ динамики среднегодовой температуры воздуха в Ленинграде – Санкт-Петербурге за последние 40 лет и измерен линейный тренд. Но нас интересует среднегодовой прирост не только как факт, относящийся к 1957 –1997 гг., но и как характеристика процесса потепления климата города вообще для распространения ее на будущее, например, столетие. В этом случае параметры полученного тренда – лишь выборочные оценки генеральных параметров с некоторой вероятной ошибкой.

Наличие случайных колебаний уровней в отдельные периоды или моменты времени вносят неизбежный элемент случайности во все показатели динамики, если их хотят распространить на генеральную совокупность.

Само наличие тренда или его отсутствие на изучаемом отрезке времени может быть доказано лишь с некоторой вероятностью, для чего используются специальные критерии. При изучении случайной колеблемости очень важно определить вероятность крайних, макси-

мальных отклонений от тренда: сильных неурожаев, морозов, наводнений и т.п.

По указанным причинам в данной главе рассматриваются методы вероятностной оценки параметров тренда и колеблемости, которые приводились в предыдущих главах без таковой, но на самом деле обязательно должны сопровождаться указанием степени надежности и доверительным интервалом для оценки генеральной величины показателя.

7.1. Оценка надежности параметров тренда

Вероятностная оценка любого выборочного показателя осуществляется с помощью сравнения его величины с величиной средней квадратической ошибки (среднего квадратического отклонения выборочных показателей при данном типе и объеме выборки от генерального показателя). Подробнее об этом следует читать в учебных пособиях, посвященных выборочному методу.

Надежность следует проверять для основного параметра тренда: среднегодового абсолютного изменения при линейном тренде, ускорения при параболе II порядка, коэффициента роста при экспоненте. Свободный член, если он ненадежно отличен от нуля, нужно оценить с точки зрения экономики, технологии или другой науки по существу процесса, и если такое положение допустимо, то тренд надежен, если надежен его главный параметр. Если же по существу свободный член, то есть уровень тренда в период, принятый за начало отсчета времени, не может быть равен нулю, то тренд ненадежен, несмотря на надежность главного параметра.

Рассмотрим проверку надежности тренда реальных денежных

доходов населения, (см. рис. 4.1 и табл. 5.5.).

Тренд имеет вид:

$$\hat{y}_i = 105,8 + 3,7 \cdot t_i,$$

где $t_i = 0$ в 2002 г., среднее квадратическое отклонение уровней от тренда $S(t) = 8,703 \%$.

Средняя ошибка репрезентативности выборочного коэффициента линейного тренда определяется по формуле:

$$m_b = \frac{S(t)}{\sqrt{\sum_{i=1}^n t_i^2}},$$

где $S(t)$ – оценка среднего квадратического отклонения уровней от тренда;

$\sum_{i=1}^n t_i^2$ – рассчитывается при отсчете t_i от середины ряда или

$\sum_{i=1}^n (t_i - \bar{t})^2$ – при отсчете t_i от начала ряда;

n – число уровней ряда.

$$m_b = \frac{8,703}{\sqrt{60}} = 1,12.$$

Отношение среднегодового изменения к его средней ошибке – это *t-критерий Стьюдента*:

$$t = \left| \frac{b}{m_b} \right| = \left| \frac{3,7}{1,12} \right| = 3,304.$$

Величину критерия сравниваем с табличной величиной критерия Стьюдента для $9 - 2 = 7$ степеней свободы, которая для значимости (вероятности нулевой гипотезы) 0,05 равна 2,3646. Фактическая величина критерия больше табличного, следовательно, вероятность нулевой гипотезы (о равенстве параметра b нулю) чрезвычайно мала. Достоверно известно, что тренд существовал и что реальные доходы населения в РФ в анализируемом периоде увеличивались не случайно.

Если исходный ряд достаточно велик и применялось многократное скользящее определение среднего изменения уровней, формула средней ошибки параметра тренда видоизменяется. Рассмотрим актуальную научную задачу: насколько надежно можно установить наличие тренда среднегодовой температуры воздуха, например, по данным ряда температур в Ленинграде – Санкт-Петербурге за 1957–1997 гг. (табл. 7.1).

Таблица 7.1.

**Среднегодовая температура воздуха
в Ленинграде – Санкт-Петербурге, °С**

Год	Температура, y_i						
1	2	3	4	5	6	7	8
1957	6,2	1967	5,6	1977	4,7	1987	3,2
1958	4,6	1968	4,3	1978	3,8	1988	5,9
1959	5,6	1969	3,8	1979	4,8	1989	7,4

1960	4,6	1970	4,9	1980	4,5	1990	6,4
1961	6,3	1971	4,8	1981	5,3	1991	6,2
1962	5,0	1972	5,8	1982	5,2	1992	6,3

Продолжение таблицы 7.1

1	2	3	4	5	6	7	8
1963	4,3	1973	5,1	1983	6,1	1993	5,3
1964	5,2	1974	6,7	1984	5,8	1994	5,5
1965	4,3	1975	6,4	1985	3,6	1995	6,5
1966	3,5	1976	2,9	1986	4,8	1996	5,2
						1997	5,7

Проведено многократное выравнивание: 21 раз по 21 уровню в каждой базе. Тренд имеет вид:

$$\hat{y}_i = 5,183 + 0,02554t_i; t = 0 \text{ в } 1977 \text{ г.}$$

Колеблемость характеризуется величиной $S(t) = 1,121$ градуса.

Величина среднегодового прироста температуры очень мала – сотые доли градуса за год, что вызывает подозрение в его несущественном, ненадежном отличии от нуля. Необходимо проверить вероятность нулевой гипотезы.

Каждое из 21 значений параметра тренда – это одна выборка. Можно для каждой такой выборки определять величину $S(t)$ и ошибки оценки среднегодового изменения, а затем вычислить ошибку среднего значения параметра всех 21 выборок, которая будет в $\sqrt{21}$ раз меньше. Однако, по нашему мнению, можно упростить расчет ошиб-

ки, применив формулу:

$$m_{\bar{b}} = \frac{S(t)}{\sqrt{l \cdot \sum_{i=1}^{21} t_i^2}}$$

Здесь l – число баз расчета среднего параметра;

$\sum_{i=1}^{21} t_i^2$ – сумма квадратов номеров периода при отсчете от середины ряда в 21 уровень.

Имеем:

$$m_{\bar{b}} = \frac{1,121}{\sqrt{21 \cdot 707}} = 0,00920 \text{ градуса.}$$

При этом t -критерий Стьюдента равен:

$$\frac{\bar{b}}{m_{\bar{b}}} = \frac{0,02554}{0,00920} = 2,78.$$

Табличное значение критерия для значимости 0,05 (вероятность нулевой гипотезы) при $41 - 2 = 39$ степенях свободы вариации составляет 2,02. Следовательно, вероятность нулевого значения среднегодового прироста температуры менее 0,05, а надежность того, что среднегодовая температура воздуха в городе повышается, больше 0,95. Нужно, конечно, оговорить причины потепления: не только общее изменение температуры по всему Земному шару, но и рост энергопотребления в самом городе. Для того чтобы установить, происходит ли общее потепление, нужно вести анализ не по городам, а по территориям, не имеющим местных источников возможного потепления, и на

большом числе таких территорий.

Для основного параметра параболы II порядка c средняя ошибка репрезентативности выборочной оценки параметра вычисляется по формуле:

$$m_c = \frac{S(t)}{\sqrt{t_i^4 - (\overline{t_i^2})^2}}.$$

Под корнем, при условии отсчета номеров периодов или моментов времени от середины ряда, стоят выражения: средняя величина четвертых степеней t_i минус квадрат среднего квадрата t_i ; по существу, это дисперсия, но не линейная, а квадратическая аргумента параболы. Если же отсчет периодов времени идет не от середины ряда, а от начала, то подкоренное выражение принимает вид:

$$\overline{(t_i - \bar{t})^4} - \left[\overline{(t_i - \bar{t})^2} \right]^2.$$

Здесь черта над скобками – знак средних величин.

Рассмотрим пример по данным, представленным рис. 4.2, – динамика численности населения РФ в 1900 – 2004 гг., имеющая параболический тренд. Его уравнение имеет вид:

$$\hat{y}_i = 113755,6 + 769,94t_i - 0,011t_i^2,$$

где $t=0$ в 1952 г.

Проверим, надежно ли отличие от нуля параметра c , половины ускорения. Колеблемость уровней численности населения измеряется величиной

$$S(t) = \sqrt{\frac{1728888461}{105 - 3}} = 4117,024.$$

Имеем:

$$m_c = \frac{4117,024}{\sqrt{1518923,467 - 843948,44}} = 5,011$$

Критерий Стьюдента равен отношению $\frac{c}{m_c} = \frac{0,0111}{5,011} = 0,0022$. Таб-

личное значение критерия при ста двух степенях свободы составляет 1,9835. Таким образом, замедленное повышение численности населения РФ за 1900–2004 гг. статистически незначимо отлично от нуля.

Для оценки основного параметра экспоненциального тренда – среднего коэффициента изменения уровней k – целесообразнее всего применить предложенную Е.М.Четыркиным¹⁵ методику: проверяется отличие от нуля логарифма среднего коэффициента изменения с учетом среднего квадратического отклонения логарифмов фактических уровней от логарифмов уровней тренда. Иначе говоря – методика та же, как для прямой линии, но только не для абсолютных величин, а для их логарифмов.

Формула средней ошибки логарифма коэффициента изменения k имеет вид:

$$m_{\ln k} = \frac{S_{(t)\ln y_i}}{\sqrt{\sum t_i^2}}$$

Рассмотрим эту методику на примере экспоненциального роста

¹⁵ Четыркин Е.М. Статистические методы прогнозирования. – М.: Статистика, 1977. с. 173-174

национального богатства РФ за 1900–1991 гг. (см. рис. 4.3. и табл. 5.8). Тренд имеет вид:

$$\hat{y}_i = 1457,7 \cdot 1,0565^{t_i}$$

В логарифмическом виде:

$$\ln \hat{y}_i = 7,2846 + 0,0549t_i$$

Дополнительно вычисляем отклонения логарифмов уровней от логарифмов тренда (табл. 7.2).

Таблица 7.2.

**Определение отклонений логарифмов уровней
от логарифмов тренда**

Годы	$\ln y_i$	$\ln \hat{y}_i$	Отклонения $U_{\ln y_i}$	$U^2_{\ln y_i}$	t_i	t_i^2
1	2	3	4	5	6	7
1900	5,51	4,79	0,72	0,52	-45,5	2070,25
1901	5,51	4,84	0,67	0,45	-44,5	1980,25
1902	5,52	4,90	0,62	0,39	-43,5	1892,25
1903	5,52	4,95	0,57	0,32	-42,5	1806,25
1904	5,56	5,01	0,55	0,31	-41,5	1722,25
1905	5,57	5,06	0,51	0,26	-40,5	1640,25
1906	5,57	5,12	0,45	0,21	-39,5	1560,25
1907	5,57	5,17	0,40	0,16	-38,5	1482,25
1908	5,58	5,23	0,35	0,13	-37,5	1406,25
1909	5,58	5,28	0,30	0,09	-36,5	1332,25
1910	5,62	5,34	0,28	0,08	-35,5	1260,25
1911	5,63	5,39	0,24	0,06	-34,5	1190,25

1912	5,63	5,45	0,18	0,03	-33,5	1122,25
1913	5,63	5,50	0,13	0,02	-32,5	1056,25
1914	5,6	5,56	0,04	0,00	-31,5	992,25
1915	5,61	5,61	0,00	0,00	-30,5	930,25
1916	5,58	5,67	-0,09	0,01	-29,5	870,25
1917	5,59	5,72	-0,13	0,02	-28,5	812,25
1918	5,62	5,77	-0,15	0,02	-27,5	756,25
1919	5,66	5,83	-0,17	0,03	-26,5	702,25

Продолжение таблицы 7.2

1	2	3	4	5	6	7
1920	5,69	5,88	-0,19	0,04	-25,5	650,25
1921	5,72	5,94	-0,22	0,05	-24,5	600,25
1922	5,79	5,99	-0,20	0,04	-23,5	552,25
1923	5,83	6,05	-0,22	0,05	-22,5	506,25
1924	5,87	6,10	-0,23	0,05	-21,5	462,25
1925	5,91	6,16	-0,25	0,06	-20,5	420,25
1926	5,98	6,21	-0,23	0,05	-19,5	380,25
1927	6,02	6,27	-0,25	0,06	-18,5	342,25
1928	6,08	6,32	-0,24	0,06	-17,5	306,25
1929	6,12	6,38	-0,26	0,07	-16,5	272,25
1930	6,18	6,43	-0,25	0,06	-15,5	240,25
1931	6,24	6,49	-0,25	0,06	-14,5	210,25
1932	6,3	6,54	-0,24	0,06	-13,5	182,25
1933	6,34	6,60	-0,26	0,07	-12,5	156,25
1934	6,4	6,65	-0,25	0,06	-11,5	132,25
1935	6,45	6,71	-0,26	0,07	-10,5	110,25
1936	6,53	6,76	-0,23	0,05	-9,5	90,25
1937	6,58	6,82	-0,24	0,06	-8,5	72,25
1938	6,64	6,87	-0,23	0,05	-7,5	56,25
1939	6,7	6,93	-0,23	0,05	-6,5	42,25
1940	6,77	6,98	-0,21	0,05	-5,5	30,25
1941	6,78	7,04	-0,26	0,07	-4,5	20,25
1942	6,8	7,09	-0,29	0,09	-3,5	12,25
1943	6,84	7,15	-0,31	0,09	-2,5	6,25

1944	6,9	7,20	-0,30	0,09	-1,5	2,25
1945	6,92	7,26	-0,34	0,11	-0,5	0,25
1946	6,95	7,31	-0,36	0,13	0,5	0,25
1947	6,98	7,37	-0,39	0,15	1,5	2,25
1948	7,01	7,42	-0,41	0,17	2,5	6,25
1949	7,03	7,48	-0,45	0,20	3,5	12,25
1950	7,08	7,53	-0,45	0,20	4,5	20,25
1951	7,18	7,59	-0,41	0,17	5,5	30,25

Продолжение таблицы 7.2

1	2	3	4	5	6	7
1952	7,28	7,64	-0,36	0,13	6,5	42,25
1953	7,38	7,70	-0,32	0,10	7,5	56,25
1954	7,47	7,75	-0,28	0,08	8,5	72,25
1955	7,56	7,81	-0,25	0,06	9,5	90,25
1956	7,66	7,86	-0,20	0,04	10,5	110,25
1957	7,76	7,92	-0,16	0,02	11,5	132,25
1958	7,86	7,97	-0,11	0,01	12,5	156,25
1959	7,96	8,03	-0,07	0,00	13,5	182,25
1960	8,04	8,08	-0,04	0,00	14,5	210,25
1961	8,11	8,14	-0,03	0,00	15,5	240,25
1962	8,19	8,19	0,00	0,00	16,5	272,25
1963	8,27	8,25	0,02	0,00	17,5	306,25
1964	8,35	8,30	0,05	0,00	18,5	342,25
1965	8,43	8,36	0,07	0,01	19,5	380,25
1966	8,49	8,41	0,08	0,01	20,5	420,25
1967	8,56	8,46	0,10	0,01	21,5	462,25
1968	8,63	8,52	0,11	0,01	22,5	506,25
1969	8,7	8,57	0,13	0,02	23,5	552,25
1970	8,77	8,63	0,14	0,02	24,5	600,25
1971	8,83	8,68	0,15	0,02	25,5	650,25
1972	8,9	8,74	0,16	0,03	26,5	702,25
1973	8,97	8,79	0,18	0,03	27,5	756,25
1974	9,04	8,85	0,19	0,04	28,5	812,25
1975	9,11	8,90	0,21	0,04	29,5	870,25

1976	9,17	8,96	0,21	0,04	30,5	930,25
1977	9,23	9,01	0,22	0,05	31,5	992,25
1978	9,3	9,07	0,23	0,05	32,5	1056,25
1979	9,36	9,12	0,24	0,06	33,5	1122,25
1980	9,42	9,18	0,24	0,06	34,5	1190,25
1981	9,47	9,23	0,24	0,06	35,5	1260,25
1982	9,53	9,29	0,24	0,06	36,5	1332,25
1983	9,59	9,34	0,25	0,06	37,5	1406,25

Продолжение таблицы 7.2

1	2	3	4	5	6	7
1984	9,64	9,40	0,24	0,06	38,5	1482,25
1985	9,7	9,45	0,25	0,06	39,5	1560,25
1986	9,75	9,51	0,24	0,06	40,5	1640,25
1987	9,8	9,56	0,24	0,06	41,5	1722,25
1988	9,86	9,62	0,24	0,06	42,5	1806,25
1989	9,89	9,67	0,22	0,05	43,5	1892,25
1990	9,93	9,73	0,20	0,04	44,5	1980,25
1991	9,95	9,78	0,17	0,03	45,5	2070,25
Итого	670,18	670,18	0,00	7,27	0	64883,00

Среднее квадратическое отклонение логарифмов:

$$S_{(t) \ln y_i} = \sqrt{\frac{7,27}{90 - 2}} = 0,2843$$

Средняя ошибка логарифма коэффициента изменения:

$$m_{\ln k} = \frac{0,2843}{\sqrt{64883,00}} = 0,0011$$

Критерий Стьюдента:

$$\frac{\ln k}{m_{\ln k}} = \frac{0,0549}{0,0011} = 49,193$$

Табличный критерий Стьюдента при девятиста степенях свободы и значимости 0,01 равен 2,63. Полученное значение критерия много больше табличного, так что вероятность нулевой гипотезы можно считать равной нулю, а рост национального богатства РФ – достоверным.

Для кривых, не имеющих постоянного основного параметра, вышеизложенный метод проверки надежности неприменим. В таких случаях можно, во-первых, проверять сам факт наличия какого-либо тренда путем сравнения средних уровней за первую и за вторую половину периода, во-вторых, обычной методикой проверки надежности различия двух средних величин в теории выборочного метода. Если различие средних уровней в более ранний период и в более поздний период надежно (нулевая гипотеза отвергается), значит, тренд существует. А о форме уравнения тренда судим по тем методикам и показателям, которые изложены в гл. 5.

7.2. Доверительные границы тренда

Если уравнение тренда рассматривается как выборочное, имеющее ошибки репрезентативности своих параметров, то можно рассчитать доверительные границы, внутри которых с заданной, достаточно большой вероятностью, проходит линия тренда в генеральной совокупности. Рассмотрим эту проблему на примере простейшего, линейного тренда. Оба его параметра, свободный член a и среднее изменение за единицу времени b имеют ошибки репрезентативности

выборочных оценок. Свободный член уравнения тренда – это выборочная средняя величина уровней временного ряда, средняя ошибка репрезентативности которой $m_a = \frac{S(t)}{\sqrt{n}}$. Средняя ошибка репрезентативности параметра b , как упоминалось выше, равна:

$$m_b = \frac{S(t)}{\sqrt{\sum_{i=1}^n t_i^2}}.$$

Свободный член уравнения линейного тренда и среднее изменение за единицу времени – величины независимые, а, следовательно, согласно теореме сложения дисперсий независимых величин, дисперсия их суммы равна сумме дисперсий слагаемых, а среднее квадратическое отклонение (средняя ошибка) – корню квадратному из суммы дисперсий, т.е. из суммы квадратов ошибок m_a^2 и m_b^2 . Однако, мы рассматриваем ошибку не в статике, а в динамике. Средняя ошибка положения линии тренда за счет ошибки свободного члена – это константа для любой точки линии тренда, а средняя ошибка изменения уровня тренда за счет ошибки параметра b – это величина переменная, ибо в разных точках линии тренда его уровень равен $a + bt_i$, и ошибка параметра b возрастет в t_i раз по сравнению с ошибкой в точке, где $t_i = 1$. Следовательно, ошибка линии тренда минимальна в середине базы его расчета – в середине временного ряда. В этой точке, где $t = 0$, средняя ошибка положения линии тренда равна ошибке его свободно-го члена, т.е. $\frac{S(t)}{\sqrt{n}}$, а в любой иной точке тренда его средняя ошибка вычисляется по формуле:

$$m_{\hat{y}_m} = \sqrt{m_a^2 + m_b^2 \cdot t_m^2} = \sqrt{\frac{S^2(t)}{n} + \frac{S^2(t) \cdot t_m^2}{\sum_{i=1}^n t_i^2}} = S(t) \sqrt{\frac{1}{n} + \frac{t_m^2}{\sum_{i=1}^n t_i^2}}$$

– для однократного выравнивания и при $t_i = 0$ в середине ряда. При нумерации периодов времени от начала ряда вместо t_i в формулу следует подставить величину $(t_i - \bar{t})^2; (t_m - \bar{t})^2$.

При многократном скользящем определении параметра b второе слагаемое подкоренного выражения примет вид:

$$\frac{t_m^2}{l \sum_{i=1}^n t_i^2},$$

где n – длина одной базы расчета тренда;

l – число баз.

Рассчитаем среднюю ошибку тренда среднегодовой температуры воздуха в Ленинграде – Санкт-Петербурге по данным таблицы 7.1:

$$m_{\hat{y}_n} = 1,121 \cdot \sqrt{\frac{1}{41} + \frac{t_m^2}{21 \cdot 707}}.$$

Для середины ряда – 1977 г. – средняя ошибка тренда составила:

$$1,121 \cdot \sqrt{\frac{1}{41}} = 0,175^\circ. \text{ А для крайних уровней 1957 г. и 1997 г. средняя}$$

$$\text{ошибка тренда составляет } 1,121 \cdot \sqrt{\frac{1}{41} + \frac{20^2}{21 \cdot 707}} = 0,254^\circ.$$

Таким образом, ошибка тренда возрастает от середины базы его расчета (середина ряда) к его краям, образуя конусообразную зону вероятных значений генерального тренда.

Если эту зону мы хотим определить с достаточно большой вероятностью, то среднюю ошибку следует умножить на величину t -

критерия Стьюдента для соответствующей вероятности. Границы доверительной зоны тренда среднегодовой температуры с вероятностью 0,95 изображены на рис. 7.1.

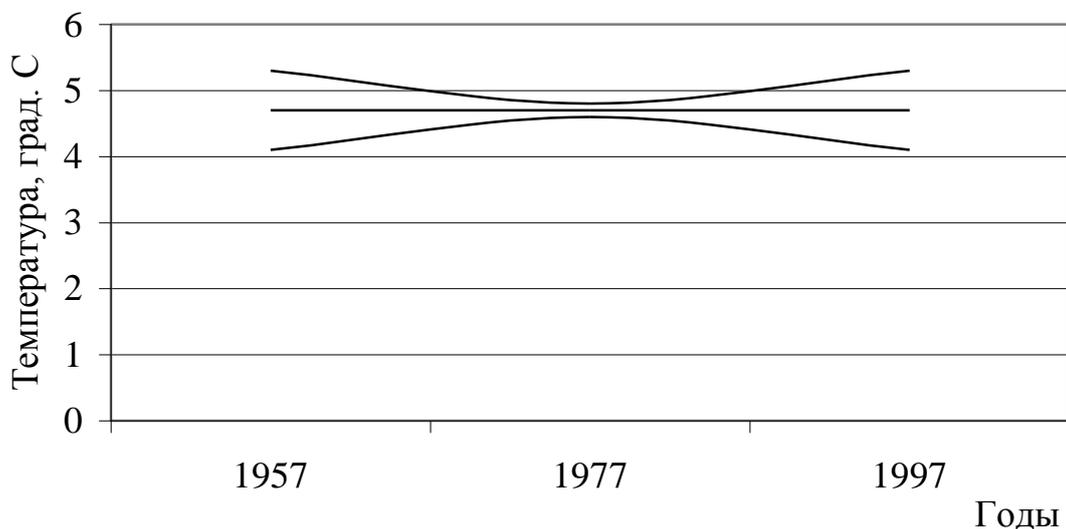


Рис. 7.1. Доверительные границы генерального тренда среднегодовой температуры воздуха в Ленинграде – Санкт-Петербурге.

Чем сильнее колеблемость уровней и чем меньше база расчета тренда, тем шире доверительная зона генерального тренда, и тем быстрее она расширяется от середины ряда к его концам. Зона для параболического тренда расширятся при этом гораздо сильнее, чем для линейного тренда.

7.3. Вероятностная оценка показателей колеблемости.

Для сравнения показателей колеблемости разных временных рядов необходимо использовать известные в математической статистике методы вероятностной оценки среднего квадратического откло-

нения или коэффициента вариации. Их можно применять для вероятностных оценок среднего квадратического отклонения уровней ряда от тренда и коэффициента колеблемости.

Средняя ошибка репрезентативности выборочной оценки генерального среднего квадратического отклонения от тренда при их нормальном распределении имеет вид¹⁶:

$$m_{S(t)} = \frac{S(t)}{\sqrt{2n}},$$

где $S(t)$ – среднее квадратическое отклонение уровней от тренда;
 n – число уровней.

Критерий Стьюдента – отношение среднего квадратического отклонения уровней от тренда к его средней ошибке – примет вид: $S(t) : m_{S(t)} = \sqrt{2n}$. Так как эту величину, как и табличное значение критерия Стьюдента для вероятностей 0,95 и 0,99, можно свести в одну таблицу, получаем готовую таблицу для оценки надежности отличия генерального среднеквадратического отклонения уровней от нуля (табл. 7.3).

Таким образом, если обнаружена колеблемость уровней ряда, число уровней которого более 5, то можно считать достаточно надежно установленным, что отличие $S(t)$ от нуля не случайно.

Доверительная граница среднего квадратического отклонения уровней от тренда с заданной вероятностью равна $S(t) \pm t_{\text{Стюд.}} \cdot m_{S(t)}$. Например, доверительный интервал средней силы колебаний средне-

¹⁶ Юзбашев М.М. Экспериментальная проверка закономерности распределения отклонений урожайности отдельных лет от тренда по их величине // Науч. труды ЛСХИ. – Т. 322. –Л. – Пушкин. – 1981.

годовой температуры воздуха в Ленинграде – Санкт-Петербурге за 1957–1997 гг. с вероятностью 0,95 составил:

$$1,121 \pm 2,03 \cdot \frac{1,121}{\sqrt{2 \cdot 41}} = 1,121 \pm 0,251^\circ.$$

Таблица 7.3.

Вероятность отличия $S(t)$ от нуля

n	$\sqrt{2n}$	Табличный критерий t для вероятности			Вывод о надежности от- личия
		0,90	0,95	0,99	
2	2,2	2,92	4,30	9,92	Значительно ниже 0,9
3	2,45	2,35	3,18	5,84	Выше 0,9, но ниже 0,95
5	3,16	2,02	2,57	4,03	Выше 0,95, но ниже 0,99
8	4,0	1,86	2,30	3,35	Выше 0,99
9	4,24	1,83	2,26	3,25	
10	4,47	1,81	2,23	3,17	
12	4,90	1,78	2,18	3,06	
15	5,48	1,75	2,13	2,95	Практически достоверно
18	6,00	1,73	2,10	2,90	
20	6,32	1,72	2,09	2,84	
25	7,07	1,71	2,08	2,79	
30	7,75	1,70	2,04	2,75	
40	8,94	1,64	2,03	2,72	
50	10,00	1,64	1,98	2,70	
100	14,14	1,64	1,96	2,62	

Доверительный интервал среднего квадратического отклонения доходов бюджета Оренбургской области (см. табл. 6.5.) с вероятностью

стью 0,95 составляет: $540,52 \pm 1,9944 \cdot \frac{540,52}{\sqrt{2 \cdot 72}} = 540,52 \pm 89,8365$ млн.руб.

Ввиду довольно значительной силы колебаний, доверительный интервал оценки генерального среднего квадратического колебания довольно широк. Ошибка возрастает прямо пропорционально силе колеблемости и росту надежности оценки, а уменьшается обратно пропорционально корню квадратному из числа уровней ряда.

Средняя ошибка репрезентативности выборочной оценки генерального коэффициента колеблемости имеет вид¹⁷:

$$m_{V(t)} = \frac{V(t)}{\sqrt{2n}} \cdot \sqrt{1 + 2 \left(\frac{V(t)}{100} \right)^2},$$

где $V(t)$ – коэффициент колеблемости, %.

Например, коэффициент колеблемости доходов бюджета Оренбургской области за январь 2001г. по декабрь 2006 г. составил 29,27%. Если рассматривать этот показатель как выборочный для Оренбургской области на больший период, то средняя ошибка коэффициента как оценки генерального равна:

$$\frac{29,27}{\sqrt{2 \cdot 72}} \cdot \sqrt{1 + 2 \cdot \left(\frac{29,27}{100} \right)^2} = 2,64\% .$$

С вероятностью 0,95 при 71 степенях свободы колеблемости доверительные границы генерального коэффициента вариации составят $29,27\% \pm 1,99 \cdot 2,64\%$, или от 24,02% до 34,52%.

Не менее, а может и более важная задача, чем вероятностная

¹⁷ Юзбашев М.М., Манелля А.И. Статистический анализ тенденций и колеблемости. – М.: Финансы и статистика, 1983.

оценка генеральных параметров колеблемости, является вероятностная оценка крайних отклонений от тренда, например, сильных неурожаяев, повышения цен, себестоимости затрат и т.п. Эти экстремальные отклонения определяют производственные риски, а оценка вероятности рисков – одна из главных задач менеджмента в любой отрасли народного хозяйства.

Вероятностная оценка отклонений от тренда возможна в том случае, если известен закон вероятностей их распределения по величине отклонений. Хотя ни в одном реальном временном ряду отклонения не подчиняются *абсолютно точно* какому-то теоретическому распределению вероятностей, во многих процессах распределение вероятностей отклонения от тренда близко к нормальному закону. В нашем примере распределение отклонений от тренда доходов бюджета Оренбургской области близко к нормальному (рисунок 7.2).

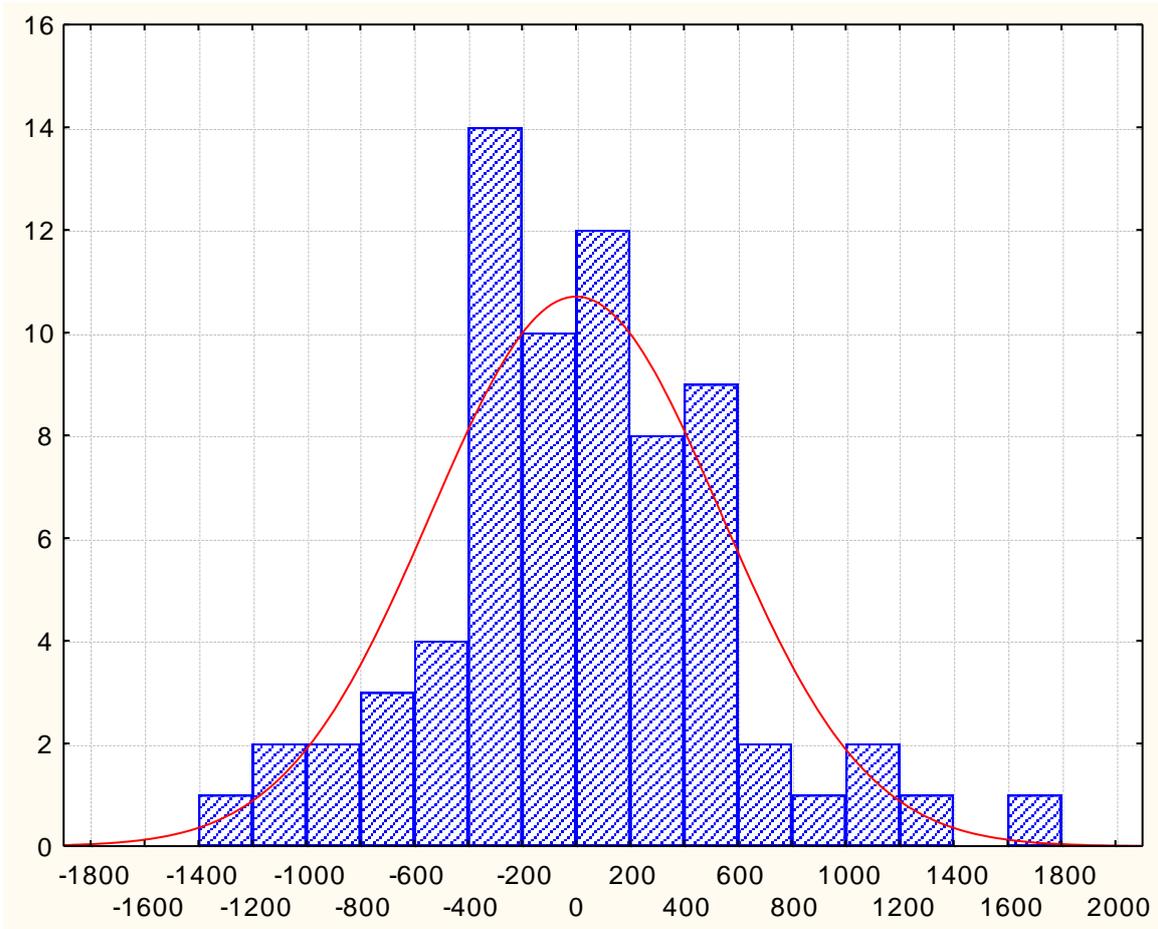


Рис. 7.2 Гистограмма отклонений от тренда временного ряда доходов бюджета Оренбургской области

Среднее квадратическое отклонение, согласно данным табл. 6.5, равно 540,52 млн. руб. Находим *вероятности рисков*, т.е. что отклонение от тренда вниз превышает уровни –5 млн. руб.; –10 млн. руб.; –15 млн. руб. (табл. 7.4).

Таблица 7.4

**Расчет вероятностей рисков доходов бюджета
в Оренбургской области**

Отклонения вниз от тренда, U_i ,	Нормированное отклонение	Вероятность отклоне- ния,
---------------------------------------	-----------------------------	------------------------------

млн. руб.	$t = U_i: S(t)$	P
–5 и более	0,01	0,496
–10 и более	0,02	0,492
–15 и более	0,03	0,488

Вероятность P равна половине разности между единицей и $F(t)$, то есть применяется *односторонний критерий* (иногда в литературе приводится готовая таблица вероятностей именно этого критерия).

Поясним определение этой вероятности с помощью графика (рис.7.3), из которого ясно и то, что у нас обозначено, как $F(t)$.

Таким образом, вероятность отклонения на 5 млн. руб. или больше почти равна 50 %, или в среднем случится один раз в месяц.

Логически ясно (это видно из графика, рис. 7.3), что точно такая же, как вероятность отклонения от тренда больше, чем на $2S(t)$ вниз, и вероятность вверх от тренда.

Если же распределение колебаний по их величине далеко от нормального, закон распределения вообще неизвестен, приближенную оценку вероятностей риска возникновения больших отклонений от тренда можно получить на основе эмпирических частот таких отклонений. Для этого, конечно, необходим достаточно длинный временной ряд. Нельзя на основе данных за 5–6 лет предсказывать вероятность отклонения, случающегося в среднем раз в 20–25 лет. Методику эмпирической оценки возможности крупных отклонений покажем на условном примере, приведенном в табл. 7.5.

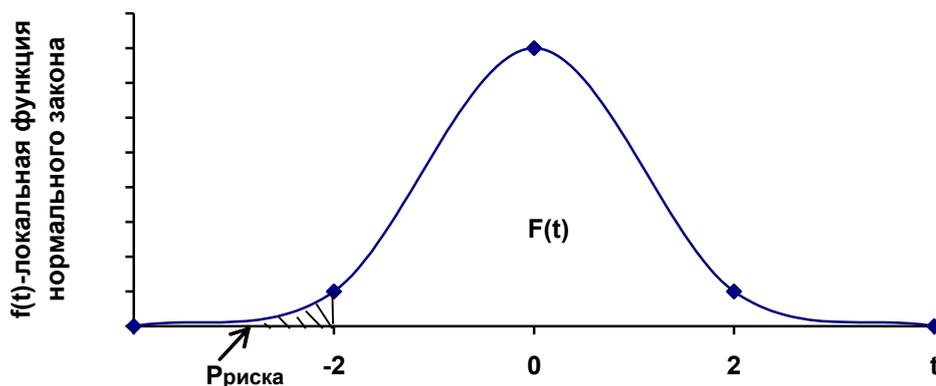


Рис. 7.3 Вероятность отрицательного отклонения, большего по величине, чем заданная граница.

Средняя ошибка репрезентативности выборочной доли (частоты), как известно, равна:

$$m_d = \frac{\sqrt{d \cdot (1-d)}}{\sqrt{n}}$$

Вычислив средние ошибки всех частот, умножаем их на 2, получая вероятные ошибки приблизительно с вероятностью 0,95, или на 3, приблизительно с вероятностью 0,995. Так как распределение не является нормальным, лучше для гарантии взять трехкратную среднюю ошибку частоты и сделать вывод о возможной частоте отклонения от тренда на указанный процент по величине этой частоты плюс трехкратная средняя ошибка.

Таблица 7.5.

**Оценка вероятности отклонений от тренда
при неизвестном законе их распределения**

Отклонения от тренда, %	Частота f_i	Частость $f_i : \sum f_i$	Трехкратная ошибка	Предельная частость
Ниже				

более чем на 20	3	0,06	0,10	0,16
от 20 до 10	7	0,14	0,15	0,29
от 0 до 10	12	0,24	0,18	0,42
Выше				
от 0 до 10	5	0,10	0,13	0,23
от 10 до 20	14	0,28	0,19	0,47
от 20 до 30	7	0,14	0,15	0,29
более чем на 30	2	0,04	0,08	0,12
Итого	50	1,00	—	—

Таким образом, крайне маловероятно, что отклонение вниз от тренда более чем на 20% встретится чаще, чем 16 раз за 100 рассматриваемых периодов (это могут быть и годы, и месяцы, и другие отрезки времени в исходном ряду). Вероятность отклонения от тренда вверх более чем на 30%, наверняка, не превысит 0,12 или 12 раз за 100 интервалов времени. Напомним, что расчет этот сделан с большим запасом осторожности ввиду неизвестности закона распределения и не очень большого объема выборки (числа уровней в исходном ряду).

В заключение рассмотрим задачу о сравнении двух значений показателей колеблемости, которая тоже требует вероятностной оценки. Задача связана с мониторингом колебаний; при этом весьма важно следить за тем, чтобы управленческие решения приводил к уменьшению величины колебаний. Для того чтобы определить, надежно ли изменение величины $S(t)$ в сравнении с прошлым периодом (например, десятилетием), нужно проверить нулевую гипотезу о случайном

различии величин $S(t)_0$ – базисного периода и $S(t)_1$ – текущего периода. Для решения задачи о различии двух или более дисперсий (т.е. $S(t)^2$) применяется *критерий Бартлетта M*. Он основан на том, что если сравниваемые величины равны, то их арифметическая средняя (взвешенная или простая) равна их геометрической средней, а если величины различаются, то чем больше они различаются, тем больше и различие между арифметической и геометрической средними.

Взвешенная арифметическая средняя дисперсия равна:

$$\overline{S(t)^2}_{ap} = \frac{\sum_{i=1}^k S(t)_i^2 \cdot n_i}{\sum_{i=1}^k n_i},$$

где k – число дисперсий;

n_i – их веса, число уровней в подпериодах.

Взвешенная геометрическая средняя:

$$\overline{S(t)^2}_{geom.} = \sum_{i=1}^k n_i \sqrt[k]{\prod_{i=1}^k [S(t)^2]^{n_i}}.$$

Критерий Бартлетта имеет вид:

$$M = \ln \frac{\overline{S(t)^2}_{ap.}}{\overline{S(t)^2}_{geom.}} \cdot \sum_{i=1}^k n_i,$$

его средняя ошибка:

$$C = 1 + \frac{\sum_{i=1}^k \frac{1}{n_i} - \frac{1}{\sum_{i=1}^k n_i}}{3(k-1)}.$$

Отношение M/C имеет распределение χ^2 ("хи-квадрат") с числом степеней свободы $k - 1$.

При сравнении двух дисперсий и равном числе уровней в каждом подпериоде (средние будут невзвешенные), формулы упрощаются:

$$M = 2 \ln \frac{S(t)_{ap.}^2}{S(t)_{geom.}^2} \cdot \sum_{i=1}^k n_i;$$

$$C = 1 + \frac{\frac{1}{n} - \frac{1}{2n}}{3}.$$

Например, сравним силу колебаний доходов бюджета Оренбургской области за первые три года (январь 2001г. – декабрь 2003 г.) и за последние 3 года (январь 2004г. – декабрь 2006 г.):

Соответственно, дисперсии равны:

$$S^2(t)_1 = 67597,7; \quad S^2(t)_2 = 516731,2;$$

их арифметическая средняя равна:

$$S^2_{арифм} = 292164,5,$$

а их геометрическая средняя:

$$S^2_{geom} = \sqrt{67597,7 \cdot 516731,2} = 186895,3;$$

$$M = 2 \ln \frac{292164,5}{186895,3} \cdot 72 = 64,33;$$

$$C = 1 + \frac{\frac{1}{36} - \frac{1}{72}}{3} = 1,014.$$

$M/C = 63,45$. Табличное значение критерия χ^2 при 1 степени свободы и значимости 0,05 составляет 3,84. Фактическое значение 63,45 больше табличного, следовательно, можно считать, что колеблемость в последние 3 года выше, чем в первые 3 года изучавшегося периода, т.е. колеблемость доходов бюджета в Оренбургской области усилилась.

Глава 8. МЕТОДЫ ИЗУЧЕНИЯ И ИЗМЕРЕНИЯ УСТОЙЧИВОСТИ УРОВНЕЙ РЯДА И ТРЕНДА

Устойчивость временного ряда – понятие многоплановое. Его следует рассматривать с двух позиций:

- устойчивости уровней временного ряда;
- устойчивости тенденции (тренда).

Вопрос определения понятия *устойчивость* невозможно решить без статистической теории динамического ряда, разработанной известными статистиками В.М. Обуховым, Н.С. Четвериковым, Альб. Л. Вайнштейном, С.П. Бобровым, Б.С. Ястремским. Согласно этой теории статистический показатель содержит в себе элементы необходимого и случайного. Необходимость проявляется в форме тенденции динамического ряда, случайность – в форме колебаний уровней относительно кривой, выражающей тенденцию. Тенденцией характеризуется процесс эволюции. В явном виде невозможно видеть все причины, порождающие тенденцию (тренд). Полное разделение элементов случайного и необходимого существует только в виде научной абстракции.

Расчленение динамического ряда на составляющие элементы – условный описательный прием. Тем не менее, несмотря на взаимозависимость тенденции и колеблемости, решающим фактором, обуславливающим тенденцию, является целенаправленная деятельность человека, а главной причиной колеблемости – изменение условий жизнедеятельности. Исходя из вышеизложенного, можно отметить следующее. Устойчивость не означает обязательное повторение одинакового уровня из года в год; такое понимание устойчивости приравнивало бы ее к застою состоянию изучаемого явления. Слишком узким и

жестким было бы понятие устойчивости ряда как полное отсутствие в динамическом ряду всяких колебаний, так как полностью устранить влияние случайных факторов на показатель невозможно. Сокращение колебаний уровней ряда – одна из главных задач при повышении устойчивости, но этим она не исчерпывается, необходимо развитие явления. Отсюда и следует, что устойчивость временного ряда – понятие не простое, а многоплановое.

Устойчивость временного ряда – это наличие необходимой тенденции изучаемого статистического показателя с минимальным влиянием на него неблагоприятных условий.

Отсюда вытекают основные требования устойчивости:

- минимизация колебаний уровней временного ряда;
- наличие определенной, необходимой для общества тенденции изменения.

Оценивать устойчивость временного ряда можно на различных явлениях. При этом в зависимости от явления будут меняться показатели, которые используются в качестве форм выражения существа исследуемого процесса. Содержание понятия *устойчивость* при этом остается неизменным.

8.1. Методы измерения устойчивости уровней ряда

Наиболее простым, аналогичным размаху вариации при измерении устойчивости уровней временного ряда является размах колеблемости средних уровней за благоприятные и неблагоприятные, в отношении к изучаемому явлению, периоды времени:

$$R_{\hat{y}} = \bar{y}_{\text{благ}} - \bar{y}_{\text{неблаг}} \cdot \quad (8.1)$$

Причем к благоприятным периодам времени относятся все периоды с уровнями выше тренда, к неблагоприятным – ниже тренда (однако, например, при изучении динамики производительности труда, если это трудоемкость, то все должно быть наоборот).

Отношение средних уровней за благоприятные периоды времени к средним уровням за неблагоприятные $\bar{y}_{\text{благ}} / \bar{y}_{\text{неблаг}}$ также может служить показателем устойчивости уровней. Чем ближе к единице отношение, тем меньше колеблемость, а соответственно выше устойчивость. Назовем его индексом устойчивости уровней динамических рядов и обозначим:

$$i_{\bar{y}} = \frac{\bar{y}_{\text{благ}}}{\bar{y}_{\text{неблаг}}} \quad \text{или} \quad i_{\bar{y}} = \frac{\bar{y}_в}{\bar{y}_н} \quad (8.2.)$$

– отношение средней уровней выше тренда к средней уровней ниже тренда (при тенденции роста).

При измерении колеблемости уровней исчисляются обобщающие показатели отклонений уровней от тренда за исследуемый период.

Основными абсолютными показателями являются среднее линейное и среднее квадратическое отклонения (см. гл. 6., формулы 6.4; 6.5):

среднее линейное отклонение

$$a(t) = \frac{\sum_{i=1}^n |y_i - \tilde{y}_i|}{n - p}, \quad (8.3)$$

среднее квадратическое отклонение

$$S_y(t) = \sqrt{\frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{n - p}}, \quad (8.4)$$

где y_i – фактический уровень;

\tilde{y}_i – выравненный уровень;

n – число уровней;

p – число параметров тренда;

t – номера лет (знак отклонения от тренда).

Эти показатели выражаются в единицах измерения анализируемых уровней и не могут служить для сравнения колебаний различных динамических рядов. Сравнение средних линейных и квадратических отклонений по базам скользящего при многократном аналитическом выравнении дает информацию о снижении или повышении устойчивости уровней за период исследования. Аналитическое выравнение $a(t)$ и $S_y(t)$ и расчет параметров уравнения их трендов позволяют определить количественные характеристики изменения абсолютной колеблемости во времени: среднегодовое изменение, темп изменения. Снижение колеблемости во времени будет равнозначно повышению устойчивости уровней (см. разд. 6.4.).

Для характеристики устойчивости (неустойчивости) Д. Бланфорд и С. Оффат рекомендуют следующие показатели¹⁸:

¹⁸ *Blanford D. Offut S. A Review of Empirical Techniques for the Analysis of Commodity Instability. – USsl. – 1983.*

1. *Процентный размах* (Percentage Range) – PR:

$$PR = W_M - W_m, \quad (8.5)$$

где $W_M = \max(W_2, \dots, W_{t-1})$;

$W_m = \min(W_2, \dots, W_{t-1})$;

$$W_t = \frac{|X_t - X_{t-1}|}{X_{t-1}} \cdot 100, \quad t = 1, \dots, n.$$

PR оценивает разность между максимальным и минимальным относительными приростами в процентах.

2. *Показатель скользящие средние* (Moving Average) – МА, оценивает величину среднего отклонения от уровня скользящих средних:

$$MA = \frac{\sum_{i=t-r+1}^{t+r} \left[\frac{x_i - x_t}{x_t} \right]}{n+1-m}, \quad (8.6)$$

$$X_t = \frac{\sum_{i=t-r}^{t+r} x_i}{m},$$

где $r = (m - 1)/2$;

m – период скользящей средней.

3. *Среднее процентное изменение* (Average Percentage Change) – АРС, которое оценивает среднее значение абсолютных величин относительных приростов и квадратов относительных приростов:

$$APC = \frac{\sum_{t=2}^n \left[\frac{X_t - X_{t-1}}{\max(X_t - X_{t-1})} \right]^2}{n-1} \cdot 100 \quad (8.7)$$

Бланфорд и Оффат, анализируя вышеперечисленные коэффициенты, отмечают их хорошую согласованность относительно коэффициента Спирмена.

Относительные показатели колеблемости, чаще всего используемые в статистике, вычисляются делением абсолютных показателей на средний уровень за весь изучаемый период (см. разд. 6.2.2):

$$\text{коэффициент линейной колеблемости } V_y^d(t) = \frac{d_y(t)}{\bar{y}}, \quad (8.8)$$

$$\text{коэффициент колеблемости } V_y(t) = \frac{S_y(t)}{\bar{y}}, \quad (8.9)$$

где \bar{y} – средний уровень ряда.

Эти показатели отражают величину колеблемости в сравнении со средним уровнем ряда. Они необходимы для сравнения колеблемости двух различных явлений и чаще всего выражаются в %. Если $V_y(t)$ – коэффициент колеблемости, то величину

$$K_y = 100 - V_y(t), \quad (8.10)$$

называют *коэффициентом устойчивости*. Такое определение коэффициента устойчивости интерпретируется как обеспечение устойчивости уровней ряда относительно тренда лишь в $(100 - V_y(t))$ случаях. Если K_y составил 0,9, это означает, что среднее колебание составляет

10% среднего уровня. Однако вероятность того, что отдельное колебание (т.е. отклонение от тренда в отдельном периоде времени) не превзойдет средней величины колебаний $S_y(t)$ составляет лишь 0,68, если распределение колебаний по их величине близко к нормальному.

Например, (см. гл. 6, разд. 6.2.2) коэффициент колеблемости доходов бюджета Оренбургской области за январь 2001г. – декабрь 2006г. составил 29,27%, следовательно, коэффициент устойчивости уровней равен 70,73%.

8.2. Методы измерения устойчивости тенденции динамики

Наиболее простым показателем устойчивости тенденции временного ряда является коэффициент Спирмэна Kp ¹⁹:

$$Kp = 1 - \frac{6 \sum_{i=1}^n d^2}{n^3 - n} \quad (8.11)$$

где d – разность рангов уровней изучаемого ряда (P_y) и рангов номеров периодов или моментов времени в ряду (P_t);

n – число таких периодов или моментов.

Для определения коэффициента Спирмена величины уровней изучаемого явления y_i нумеруются в порядке возрастания, а при наличии одинаковых уровней им присваивается определенный ранг, равный частному от деления суммы рангов, приходящихся на эти значения, на число этих равных значений. При наличии дробных рангов

¹⁹ Афанасьев В.Н., Юзбашев М.М. Коэффициент корреляции рангов как показатель устойчивости динамики // Вестник статистики. – 1983. – № 11, 3с.

необходима поправка к формуле Спирмена:

$$Kp = 1 - \frac{6 \sum_{i=1}^n d^2 - A}{n^3 - n - 12A}, \quad (8.12)$$

где $A = \frac{1}{12} \sum_{j=1}^m (A_j^3 - A_j)$;

j – номера связок по порядку,

A_j – число одинаковых рангов в j -й связке (число одинаковых уровней).

При малой вероятности совпадения уровней и достаточном их числе эта поправка несущественна.

Коэффициент рангов периодов времени и уровней динамического ряда, может принимать значения в пределах от 0 до ± 1 .

Интерпретация этого коэффициента такова: если каждый уровень ряда исследуемого периода выше, чем предыдущего, то ранги уровней ряда и номера лет совпадают, $Kp = +1$. Это означает полную устойчивость самого факта роста уровней ряда, непрерывность роста.

Чем ближе Kp к $+1$, тем ближе рост уровней к непрерывному, выше устойчивость роста. При $Kp = 0$ рост совершенно неустойчив. При отрицательных значениях Kp чем ближе к -1 , тем устойчивее снижение изучаемого показателя. В рассмотренном ранее ряду динамики доходов бюджета Оренбургской области коэффициент Спирмена составил 83,72%.

Коэффициент устойчивости роста (Kp) можно получить и по другой формуле:

$$Kp = \frac{12 \sum_{i=1}^n P_{ii} \cdot P_{yi}}{n^3 - n} - \frac{3(n+1)}{n-1}. \quad (8.13)$$

Этот вариант расчета несколько сокращает вычисления. Коэффициент Спирмена здесь применен в совершенно новой функции, и его нельзя трактовать, как *меру связи* изучаемого явления со временем. Преимуществом коэффициента корреляции рангов как показателя устойчивости является то, что для его вычисления не требуется аналитическое выравнивание динамического ряда – сложная и чреватая ошибками стадия анализа динамики.

Следует иметь в виду, что даже при полной (100%) устойчивости роста (снижения) в ряду динамики может быть колеблемость уровней, и коэффициент их устойчивости будет ниже 100%. При слабой колеблемости, но еще более слабой тенденции, напротив возможен высокий коэффициент устойчивости уровней, но близкий к нулю коэффициент устойчивости изменения.

Обычно эти показатели изменяются совместно, большая устойчивость уровней наблюдается при большей устойчивости изменения.

Недостатком коэффициента устойчивости роста Kp является его слабая чувствительность к изменениям скорости роста уровней ряда, он может показать устойчивый рост при незначительно отличающихся от нуля приростах уровней.

В качестве характеристики устойчивости изменения можно применить индекс корреляции:

$$J_r = \sqrt{1 - \frac{\sum (y_i - \tilde{y}_i)^2}{\sum (y_i - \bar{y})^2}}, \quad (8.14)$$

где y_i – уровни динамического ряда;
 \bar{y} – средний уровень ряда;
 \tilde{y}_i – теоретические уровни ряда.

Индекс корреляции показывает степень сопряженности колебаний исследуемых показателей с совокупностью факторов, изменяющих их во времени. Приближение индекса корреляции к 1 означает большую устойчивость изменения уровней динамического ряда.

Сравнение индексов корреляции по разным показателям возможно лишь при условии равенства числа уровней, так с ростом длины периода при том же среднем приросте (b_y), той же абсолютной ($S_y(t)$) и относительной колеблемости ($V_y(t)$), он автоматически увеличивается из-за накопления изменений за счет тренда.

8.3. Комплексные показатели (критерии) устойчивости.

Сущность комплексных показателей заключается в определении их не через уровни динамического ряда, а через показатели их динамики. Так М.С. Каяйкиной²⁰ был предложен один из таких показателей (K). Он определяется как отношение среднего прироста линейного тренда $\tilde{y}_i = a + bt_i$, т.е. параметра b к среднему квадратическому отклонению уровней от тренда $S_y(t)$:

$$K = \frac{b}{S_y(t)} \quad (8.15)$$

²⁰ Калинина М.С. Статистические методы изучения динамики урожайности. – Л.: ЛСХИ, 1969.

Чем больше величина K , тем менее вероятно, что уровень ряда в следующем периоде будет меньше предыдущего. Например, если считать, как и ранее, что распределение колебаний близко к нормальному, то при $K = 1$ вероятность того, что отклонение от тренда будет не больше прироста (по модулю), составляет $F(1) \approx 0,68$. Поскольку отклонения от тренда разных знаков одинаково вероятны, можно сказать, что вероятность того, что уровень следующего года (месяца, дня) будет ниже, чем предыдущего, составит $0,5 - F(1):2 = 0,5 - 0,34 = 0,16$. Если же показатель K составляет только $0,25$, то вероятность снижения уровня следующего периода по сравнению с предыдущим составит $0,5 - F(0,25) = 0,5 - 0,1974:2 = 0,4013$. При отрицательном b вероятность снижения уровня становится больше $0,5$: так, если $b = -0,4S_y(t)$, т.е. $K = -0,4$, вероятность снижения следующего уровня такова:

$$0,5 - F(-0,4):2 = 0,5 + F(0,4):2 = 0,5 + 0,3108:2 = 0,6554.$$

Как видим, при $K = -0,4$ тенденция снижения уровней еще довольно неустойчива.

Рассмотрим показатели такого же рода для экспоненциального и параболического трендов. Основным параметром, характеризующим динамику по экспоненте, служит средний темп роста (коэффициент роста уровней в размах) k уравнения экспоненты: $\tilde{y} = ak^t$. k – величина отвлеченная, при том всегда положительная (знакопеременные уровни здесь не рассматриваются).

Недопустимо сопоставлять темпы с абсолютным показателем колеблемости $S_y(t)$, логично сравнить темпы роста уровней по экспоненциальному тренду с темпами изменения колеблемости. Для этого необходимо построить динамический ряд величин $S'_y(t)$, хотя бы

скользящим способом, и выравнивать его тоже по экспоненте, чтобы определить величину среднегодового темпа (в разгах) величины колебаний, то есть показатель $K_{S_y(t)}$. Так как для одноразового надежного вычисления показателя колеблемости уже необходимо иметь не менее 11–15 уровней, то для получения динамического ряда $S_y(t)$ и его среднегодового темпа изменения необходим динамический ряд исходных уровней значительной длины (не менее 11–15 плюс еще 9–11), т.е. более 20 уровней, а лучше около или более 30. Далеко не всегда можно получить такой длинный ряд достаточно однокачественных уровней с единым трендом.

Сопоставляя темпы роста уровней ряда с темпами изменения колеблемости, получим *показатель опережения*:

$$O_{kэ} = \frac{\bar{k}}{\bar{k}_{S_y(t)}} \quad (8.16)$$

Если $O_{kэ} > 1$, это свидетельствует, что уровни ряда в среднем растут быстрее колебаний (или снижаются медленнее колебаний). В таком случае, как понятно без доказательства, коэффициент колеблемости уровней будет снижаться, а коэффициент устойчивости уровней повышаться. Если $O_{kэ} < 1$, наоборот, колебания растут быстрее уровней тренда и коэффициент колеблемости растет, а коэффициент устойчивости уровней снижается. Таким образом, величина $O_{kэ}$ определяет направление динамики коэффициента устойчивости уровней.

Параболический тренд: $\tilde{y}_i = a + bt_i + ct_i^2$ имеет два динамических параметра: среднегодовой прирост b и половину ускорения прироста c . Величина b в параболе не является константой, и для построения

показателей комплексной устойчивости W нужно взять среднюю за весь ряд величину \bar{b} . В остальном интерпретация та же, что и для прямой. Второй показатель – половину ускорения c или ускорение прироста $2c$ – логично сопоставлять уже не с самой величиной колеблемости $S_y(t)$, а с ее среднегодовым приростом $b_{S_y(t)}$, полученным по достаточно длинному ряду путем выравнивания показателей $S_y(t)$, скользящих или следующих друг за другом. Имеем показатель

$$O_c = \frac{2c}{b_{S_y(t)}} \quad (8.17)$$

Интерпретация показателя O_c такова: если $O_c > 1$, значит, положительное ускорение (прирост абсолютного прироста уровней) больше, чем прирост среднего квадратического отклонения от тренда. Значит, отношение прироста уровней к среднему отклонению от тренда станет увеличиваться, т.е. показатель K будет возрастать, что говорит о повышении устойчивости динамики тренда. Если $O_c < 1$, значит, колебания растут сильнее, чем прирост уровней, показатель устойчивости K будет снижаться.

Это общее положение, однако, требует конкретизации, так как числитель и знаменатель показателя O_c могут принимать как положительные, так и отрицательные значения. Следовательно, может иметь место восемь возможных сочетаний: четыре – по знакам и два – по величине. Рассмотрим интерпретацию каждого из восьми возможных сочетаний:

1. $c > 0$; $b_{S_y(t)} > 0$; $2c > b_{S_y(t)}$.

Прирост уровней ряда растет, колебания тоже растут, но мед-

леннее, в результате K увеличивается, т.е. устойчивость тенденции возрастает. Уточним, что при этом не обязательно растут и уровни ряда, так как параметр b_y может быть и отрицательным, так что часть периода уровни ряда могут снижаться.

$$2. c > 0; b_{S_y(t)} > 0; 2c < b_{S_y(t)}.$$

Хотя прирост уровней возрастает (ускоряется), но колеблемость растет еще быстрее, а, значит, показатель устойчивости тенденции K снижается. Это менее благоприятный тип динамики, чем случай 1.

$$3. c > 0; b_{S_y(t)} < 0; 2c > b_{S_y(t)} - \text{очевидная ситуация.}$$

Эта комбинация означает, что прирост уровней растет, а колеблемость снижается. Ясно, что при этом показатель устойчивости тенденции K возрастает.

4. $c > 0; b_{S_y(t)} < 0; 2c < b_{S_y(t)}$ – нереальная комбинация, третье неравенство противоречит двум первым.

5. $c < 0; b_{S_y(t)} > 0; 2c > b_{S_y(t)}$ – также нереальное сочетание по той же причине.

$$6. c < 0; b_{S_y(t)} > 0; 2c < b_{S_y(t)} - \text{очевидная ситуация.}$$

Это означает, что прирост уровней снижается, а колебания возрастают. Естественно, показатель устойчивости тенденции K уменьшается и за счет знаменателя, устойчивость падает, это самый неблагоприятный тип динамики производства с точки зрения его устойчивости.

$$7. c < 0; b_{S_y(t)} < 0; 2c > b_{S_y(t)}.$$

Отсюда следует, что прирост уровней сокращается, но медленнее, чем колеблемость, так как неравенство $2c > b_{S_y(t)}$ понимается по алгебраической величине, а не по модулю, т.е., например, $c = -0,05$, а

$2c > b_{S_y(t)} = -0,13$, имеем: $2c = -0,1$, что больше, чем $-0,13$. В таком случае показатель устойчивости тенденции K будет возрастать, хотя уровни ряда либо тоже снижаются, либо растут с замедлением, так что для производства это не самый благоприятный тип динамики.

8. $c < 0$; $b_{S_y(t)} < 0$; $2c < b_{S_y(t)}$ – также понимается по алгебраической величине.

Прирост уровней снижется быстрее, чем колебания, показатель устойчивости K снижается, тип динамики неблагоприятный, хотя и не столь сильно, как тип 6.

Итак, исключив два нереальных сочетания из восьми, получим при параболическом тренде шесть типов динамики устойчивости, из них типы 1 и 3 благоприятные для производства, 2 и 7 благоприятны в одном отношении, но неблагоприятны в другом, а типы 6 и 8 явно неблагоприятны с точки зрения устойчивости.

Еще раз оговорим, что для надежного определения всей предлагаемой системы показателей устойчивости при параболическом тренде необходим достаточно длинный динамический ряд – не менее 20 уровней при едином типе тенденции. При более коротких рядах следует ограничиться показателями, не требующими оценки тенденции динамики колебаний $b_{S_y(t)}$.

Глава 9. ИЗУЧЕНИЕ ДИНАМИКИ КОМПЛЕКСА ВЗАИМОСВЯЗАННЫХ ПРИЗНАКОВ

В предыдущих главах рассматривалась динамика одного признака, выраженного тем или иным показателем, но фактически наука и практика всегда имеют дело не с изолированными признаками, а с их системами, жестко связанными функциональной либо корреляционной связью. В данной главе будут последовательно рассмотрены методики анализа таких систем признаков, а также свойства трендов и колеблемости при агрегировании объектов по совокупности, описаны связи, особенно корреляционные, в динамике. Все эти проблемы на порядок сложнее ранее изложенных и ввиду ограниченности объема данного учебника могут быть изложены только очень сжато и неполно. Желаящим глубже изучить проблемы анализа и прогнозирования систем взаимосвязанных признаков рекомендуется обратиться к специальной литературе.²¹

9.1. Динамика жестко связанной системы признаков (показателей)

Насколько нам известно, в полном объеме эта проблема в нашей литературе впервые описана Л.Н. Кривенковой в диссертации, защищенной в Санкт-Петербургском университете экономики и финан-

²¹ *Андерсон Т.* Статистический анализ временных рядов: Пер. с англ. – М.: Мир, 1976.

Вайну Я.Я.Ф. Корреляция рядов динамики. – М.: Финансы и статистика, 1982.

Гренджер К., Хатанака М. Спектральный анализ временных рядов в экономике. – М.: Статистика, 1972.

Кендэл М. Временные ряды: Пер. с англ. – М.: Финансы и статистика, 1981.

Статистический анализ экономических временных рядов и прогнозирование: ЦЭМИ АН СССР. – М.: Наука, 1973.

Хенан Э. Многомерные временные ряды: Пер. с англ. – М.: Мир, 1974.

Четыркин Е.М. Статистические методы прогнозирования. – М.: Статистика, 1977.

Юл Д.Э., Кендэл М.Д. Теория статистики: Пер. с англ. – М.: Госстатиздат, 1960.

сов²². Изложение материала начнем с конкретной задачи: необходимо рассмотреть тенденции и колеблемость трех взаимосвязанных функционально признаков: площади посева зерновой культуры, ее урожайности и валовой сбора зерна (табл. 9.1). Если площади в разные годы обозначим как P_i , урожайности – y_i , валовой сбор – b_i , то имеем функциональную связь: $b_i = P_i \cdot y_i$, справедливую для каждого года (ошибки регистрации не принимаем во внимание). Для соблюдения жесткости связи численные значения округлим до целых (табл. 9.1). Тренды площади и урожайности берем линейные.

Таблица 9.1

Динамика площади, урожайности, валового сбора

Номер периода	Фактические уровни			Уровни трендов			Отклонения от трендов		
	P_i , га	y_i , ц/га	b_i , ц	\hat{P}_i , га	\hat{y}_i , ц/га	\hat{b}_i , ц	U_{P_i} , га	U_{y_i} , ц/га	U_{b_i} , ц
-4	94	27	2538	100	25	2494	-6	+2	+44
-3	113	23	2599	105	26	2724	+8	-3	-125
-2	116	26	3016	110	27	2963	+6	-1	+53
-1	103	28	2884	115	28	3213	-12	0	-329
0	125	35	4375	120	29	3472	+5	+6	+903
1	124	25	3100	125	30	3741	-1	-5	-641
2	126	35	4410	130	31	4021	-4	+4	+389
3	142	26	3692	135	32	4310	+7	-6	-618
4	137	36	4932	140	33	4608	-3	+3	+324
$\Sigma 0$	1080	261	31546	1080	261	31546	0	0	0

²² См.: Кривенкова Л.Н. Статистические методы анализа и моделирования свиноводства (на уровнях отдельного предприятия и региона): Дисс. ... канд. экон. наук. – СПб., 1993. – С. 95–112.

Тренд площади: $\hat{\Pi}_i = 120 + 5t_i$, $t=0$ в пятом периоде от начала ряда.

Тренд урожайности: $\hat{y}_i = 29 + t_i$, $t=0$ в пятом периоде от начала ряда.

Тренд валового сбора: $\hat{b}_i = 3472,2 + 264,3t_i + 4,94t_i^2$, $t=0$ в пятом периоде от начала ряда.

Сначала рассмотрим взаимосвязь трендов в случае, когда колеблемость отсутствует. Тогда валовой сбор каждого года является произведением уровней трендов площади и урожайности, которые совпадают с фактическими уровнями площади и урожайности, т.е. имеет место равенство: $\hat{b}_i = \hat{\Pi}_i \cdot \hat{y}_i = b_i$, а вектор валового сбора представлен в табл. 9.2.

Таблица 9.2

Вектор валового сбора

Номера периодов от середины ряда	-4	-3	-2	-1	0	1	2	3	4
Валовой сбор, т	250 0	273 0	297 0	322 0	348 0	375 0	403 0	432 0	462 0
Абсолютный прирост к предыдущему, т	...	230	240	250	260	270	280	290	300
Ускорение	+10	+10	+10	+10	+10	+10	+10

Как видим, тренд валового сбора при отсутствии колебаний

площади и урожайности был бы параболой II порядка с параметрами:

$$\hat{B}_i = 3480 + 265t + 5t^2.$$

(Напомним, что параметр c – это половина ускорения; параметр b – средняя по всем периодам величина среднего абсолютного прироста; параметр a – уровень тренда в период с нулевым значением t_i .)

Уравнение тренда валового сбора с уравнениями трендов площади и урожайности при условии отсутствия колебаний связано также, как сам показатель валового сбора с показателями площади и урожайности.

Тренд признака – произведения есть произведение трендов признаков – сомножителей, если колеблемость равна нулю:

$$\hat{b} = \hat{\Pi} \cdot \hat{y} = (120 + 5t) \cdot (29 + t) = 120 \cdot 29 + 5t \cdot 29 + 120 \cdot t + 5t \cdot t = 3480 + (145 + 120)t + 5t^2,$$

что точно совпадает с ранее полученным по ряду уровней самого валового сбора уравнением его тренда. Полученный результат полностью соответствует логике взаимосвязи показателей и кажется тривиальным. Однако, фактический тренд валового сбора по данным табл. 9.1. вовсе не соответствует этой логике, т.е. тренд валового сбора при наличии колеблемости площади и (или) урожайности уже не равен произведению трендов площади и урожайности. Парабола II-го порядка, вычисленная по данным ряда валового сбора таблицы 9.1., имеет вид:

$$\hat{B}_i = 3472,2 + 264,3t_i + 4,9t_i^2.$$

И если в данном примере расхождения параметров невелики, то при более сильной колеблемости они могут оказаться уже значительно большими. Главный результат наших исследований состоит в том, что установлен факт несовпадения тренда произведения с произведе-

нием трендов сомножителей.

Следующая наша задача – теоретическое объяснение этого факта. Введем обозначения: x_i и z_i – фактические значения уровней временных рядов признаков-сомножителей; \hat{x}_i, \hat{z}_i – их трендовые значения; \hat{y}_i – трендовые значения признака-произведения, y_i – его фактические уровни. При этом имеется точное равенство $y_i = x_i z_i$. Тренды \hat{x}_i, \hat{z}_i полагаем линейными, следовательно, тренд \hat{y}_i – парабола II порядка. Будем также для упрощения записи вести отсчет номеров периодов (моментов) времени t_i от середины временных рядов. Фактические уровни признаков можно представить как сумму уровня тренда и отклонения от него, обозначаемого соответственно $U_{x_i}, U_{z_i}, U_{y_i}$, так что:

$$x_i = \hat{x}_i + U_{x_i}; z_i = \hat{z}_i + U_{z_i}; y_i = \hat{y}_i + U_{y_i}.$$

Так как

$$y_i = x_i z_i, \text{ то } y_i = (\hat{x}_i + U_{x_i}) \cdot (\hat{z}_i + U_{z_i}). \quad (9.1)$$

Рассмотрим произведение трендов сомножителей:

$$\begin{aligned} \hat{x}_i \cdot \hat{z}_i &= (\bar{x} + b_x t_i) \cdot (\bar{z} + b_z t_i) = \bar{x} \cdot \bar{z} + \bar{x} b_z t_i + \bar{z} b_x t_i + b_x t_i \cdot b_z t_i = \\ &= \bar{x} \cdot \bar{z} + (\bar{x} b_z + \bar{z} b_x) t_i + b_x b_z t_i^2 \end{aligned} \quad (9.2)$$

Уравнение (9.2) есть уравнение параболы II порядка, в котором свободный член равен произведению средних величин признаков-

сомножителей, он же – средняя величина признака произведения \bar{y} . Второй член – это средний абсолютный прирост признака-произведения за период, а третий член – половина ускорения признака-произведения. Эти результаты не новы, но следует твердо усвоить, что при равномерном росте (изменении) признаков x и z , их произведение y изменяется не равномерно, а с ускорением. Если изменения признаков-сомножителей имеют одинаковые знаки, то это ускорение – положительная величина; если изменение признаков имеют разные знаки, ускорения их произведения – отрицательная величина. При наличии более двух сомножителей, тренд их произведения будет параболой более высокого порядка со значительно более сложным поведением, в данном учебнике подробно не рассматривается.

Упомянем все же, что если оба признака-сомножителя изменяются по параболе II порядка, то тренд их произведения будет уже параболой IV порядка. Если тренды сомножителей – экспоненты, то и тренд их произведения – тоже экспонента, но вот каков ее параметр, об этом часто судят неверно. Многие руководители предприятий полагают, что если число работников будет возрастать на 10%, а производительность их труда на 8% в год, то выпуск продукции будет возрастать на $10 + 8 = 18\%$ или даже на $10 \cdot 8 = 80\%$ в год! Оба эти ответа неправильны. Тренд произведения будет иметь среднегодовой темп роста, равный произведению темпов сомножителей, т.е. $1,08 \cdot 1,10 = 1,188$ или 118,8%; следовательно, прирост продукции составит 18,8% в год к предыдущему уровню.

Далее рассмотрим свойства тренда признака-произведения при наличии колебаний каждого из признаков-сомножителей, опишем структуру каждого из параметров его параболического тренда

$\hat{y}_i = a_y + b_y t_i + c_y t_i^2$, начиная со среднего абсолютного прироста b_y , который и вычисляется первым из уравнения МНК:

$$b_y = \frac{\sum_{i=1}^n y_i t_i}{\sum_{i=1}^n t_i^2}.$$

Далее не будем указывать границ суммирования, они всегда проходят по всем уровням ряда (по всем периодам). При этом, т.к. $y_i = x_i z_i$, имеем:

$$\begin{aligned} b_y &= \frac{\sum x_i z_i t_i}{\sum t_i^2} = \frac{\sum (\hat{x}_i + U_{xi}) \cdot (\hat{z}_i + U_{zi}) t_i}{\sum t_i^2} = \frac{\sum (\bar{x} + b_x t_i + U_{xi}) \cdot (\bar{z} + b_z t_i + U_{zi}) t_i}{\sum t_i^2} = \\ &= \frac{\sum (\bar{x} \cdot \bar{z} t_i + \bar{x} b_z t_i^2 + \bar{z} b_x t_i^2 + \bar{x} U_{zi} t_i + \bar{z} U_{xi} t_i + b_x U_{zi} t_i^2 + b_z U_{xi} t_i^2 + b_x b_z t_i^3 + U_{xi} U_{zi} t_i)}{\sum t_i^2} \end{aligned} \quad (9.3)$$

Рассмотрим суммы каждого из слагаемых в числителе (9.3):

1. $\frac{\sum \bar{x} \cdot \bar{z} t_i}{\sum t_i^2} = \frac{\bar{x} \cdot \bar{z} \sum t_i}{\sum t_i^2} = 0$, так как $\sum t_i = 0$.

2 и 3. $\frac{\sum \bar{x} b_z t_i^2}{\sum t_i^2} + \frac{\sum \bar{z} b_x t_i^2}{\sum t_i^2} = \frac{\bar{x} \sum b_z t_i^2}{\sum t_i^2} + \frac{\bar{z} \sum b_x t_i^2}{\sum t_i^2} = \frac{\bar{x} b_z \sum t_i^2}{\sum t_i^2} + \frac{\bar{z} b_x \sum t_i^2}{\sum t_i^2} = \bar{x} b_z + \bar{z} b_x$,

что равно второму члену произведения трендов сомножителей, см. (9.2).

4 и 5. $\frac{\sum \bar{x} U_{zi} t_i}{\sum t_i^2} + \frac{\sum \bar{z} U_{xi} t_i}{\sum t_i^2} = \frac{\bar{x} \sum U_{zi} t_i}{\sum t_i^2} + \frac{\bar{z} \sum U_{xi} t_i}{\sum t_i^2} = 0$

Основание равенства равно нулю: сумма или математическое ожида-

ние произведений величин, математические ожидания (или суммы) каждого из которых равны нулю, тоже равна нулю:

$$\sum U_{zi}t_i = 0; \quad \sum U_{xi}t_i = 0.$$

$$6 \text{ и } 7. \quad \frac{\sum b_x U_{zi}t_i^2}{\sum t_i^2} + \frac{\sum b_z U_{xi}t_i^2}{\sum t_i^2} = \frac{b_x \sum U_{zi}t_i^2}{\sum t_i^2} + \frac{b_z \sum U_{xi}t_i^2}{\sum t_i^2}.$$

Эти члены разложения (9.3) в общем случае не равны нулю, так как $U_{zi}t_i^2$, $U_{xi}t_i^2$ – случайные величины, зависящие от распределения отклонений от тренда по периодам времени.

$$8. \quad \frac{\sum b_x b_z t_i^3}{\sum t_i^2} = \frac{b_x b_z \sum t_i^3}{\sum t_i^2} = 0, \text{ так как } \sum t_i^3 = 0.$$

$$9. \quad \frac{\sum U_{xi} U_{zi} t_i}{\sum t_i^2}. \text{ Этот член произведения (9.3) в общем случае не равен}$$

нулю, если имеет место корреляция отклонений от трендов признаков x и z .

Итак, кроме членов, равных аналогичным параметрам произведения трендов сомножителей, средний прирост в тренде произведения b_y содержит еще три члена, в общем случае не равных нулю. Следовательно, в общем случае $b_y \neq b_x b_z$, что мы и наблюдаем на примере табл. 9.1.

Рассмотрим далее квадратический параметр тренда признака– произведения, т.е. c_y . Из расчета по методу наименьших квадратов (см. главу 6) для параболы II порядка имеем:

$$c_y = \left(\frac{\sum y_i}{n} - \frac{\sum y_i t_i^2}{\sum t_i^2} \right) : \left(\frac{\sum t_i^2}{n} - \frac{\sum t_i^4}{\sum t_i^2} \right) \quad (9.4)$$

Выражение (9.4) во второй скобке не содержит величин признаков и не нуждается в анализе. В первую скобку подставляем значения:

$$y_i = \frac{\sum (\bar{x} + b_x t_i + U_{xi}) \cdot (\bar{z} + b_z t_i + U_{zi})}{n} - \frac{\sum (\bar{x} + b_x t_i + U_{xi}) \cdot (\bar{z} + b_z t_i + U_{zi}) \cdot t_i^2}{\sum t_i^2} =$$

$$= \frac{\sum \bar{x} \cdot \bar{z} + \sum \bar{x} b_z t_i + \sum \bar{z} b_x t_i + \sum b_x b_z t_i^2 + \sum \bar{x} U_{zi} + \sum \bar{z} U_{xi} + \sum b_x U_{zi} t_i + \sum b_z U_{xi} t_i + \sum U_x U_z}{n} -$$

$$\frac{\sum \bar{x} \cdot \bar{z} t_i^2 + \sum \bar{x} b_z t_i^3 + \sum \bar{z} b_x t_i^3 + \sum b_x b_z t_i^4 + \sum \bar{x} U_{zi} t_i^2 + \sum \bar{z} U_{xi} t_i^2 + \sum b_x U_{zi} t_i^3 + \sum b_z U_{xi} t_i^3 + \sum U_{xi} U_{zi} t_i^2}{\sum t_i^2}$$

Рассмотрим каждый из 18 членов разложения, используя уже известные из предыдущего анализа равенства.

1. $\sum \frac{\bar{x} \cdot \bar{z}}{n} = \bar{x} \cdot \bar{z}$, т.е. равен первому (свободному) члену произведения трендов сомножителей.

2 и 3. $\frac{\sum \bar{x} b_z t_i}{n} + \frac{\sum \bar{z} b_x t_i}{n} = \frac{\bar{x} b_z \sum t_i}{n} + \frac{\bar{z} b_x \sum t_i}{n} = 0$, так как $\sum t_i = 0$.

4. $\frac{\sum b_x b_z t_i^2}{n} = \frac{b_x b_z \sum t_i^2}{n}$.

5. и 6. $\frac{\sum \bar{x} U_{zi}}{n} + \frac{\sum \bar{z} U_{xi}}{n} = \frac{\bar{x} \sum U_{zi}}{n} + \frac{\bar{z} \sum U_{xi}}{n} = 0$, так как

$$\sum U_{xi} = \sum U_{zi} = 0.$$

7 и 8. $\frac{\sum b_x U_{zi} t_i}{n} + \frac{\sum b_z U_{xi} t_i}{n} = \frac{b_x \sum U_{zi} t_i}{n} + \frac{b_z \sum U_{xi} t_i}{n} = 0$, как уже показана

но ранее.

9. $\frac{\sum U_{xi} U_{zi}}{n}$. Это член произведения в общем случае не равен нулю при наличии корреляции между отклонениями от тренда.

$$10. -\frac{\sum \bar{x} \cdot \bar{z} t_i^2}{\sum t_i^2} = -\frac{\bar{x} \cdot \bar{z} \sum t_i^2}{\sum t_i^2} = -\bar{x} \bar{z}.$$

$$11 \text{ и } 12. -\frac{\sum \bar{x} b_z t_i^3}{\sum t_i^2} - \frac{\sum \bar{z} b_x t_i^3}{\sum t_i^2} = -\frac{\bar{x} b_z \sum t_i^3}{\sum t_i^2} - \frac{\bar{z} b_x \sum t_i^3}{\sum t_i^2} = 0, \text{ так как } \sum t_i^3 = 0.$$

$$13. -\frac{\sum b_x b_z t_i^4}{\sum t_i^2} = -b_x b_z \frac{\sum t_i^4}{\sum t_i^2}.$$

$$14 \text{ и } 15. -\frac{\sum \bar{x} U_{zi} t_i^2}{\sum t_i^2} - \frac{\sum \bar{z} U_{xi} t_i^2}{\sum t_i^2} = -\frac{\bar{x} \sum U_{zi} t_i^2}{\sum t_i^2} - \frac{\bar{z} \sum U_{xi} t_i^2}{\sum t_i^2} \quad \text{— в общем}$$

случае, как ранее показано, не равны нулю, так как зависят от распределения отклонений от трендов по времени.

$$16 \text{ и } 17. -\frac{\sum b_x U_{zi} t_i^3}{\sum t_i^2} - \frac{\sum b_z U_{xi} t_i^3}{\sum t_i^2} = -\frac{b_x \sum U_{zi} t_i^3}{\sum t_i^2} - \frac{b_z \sum U_{xi} t_i^3}{\sum t_i^2}.$$

В общем случае эти члены не равны нулю при асимметричном распределении отклонений от тренда по длине периода, особенно при ограниченной длине ряда.

$$18. -\frac{\sum U_{xi} U_{zi} t_i^2}{\sum t_i^2}, \text{ в общем случае нулю и этот член не равен при}$$

наличии корреляции между отклонениями.

Суммируя члены разложения 1, 4, 10, 13, получаем:

$$\bar{x} \cdot \bar{z} + \frac{b_x b_z \sum t_i^2}{n} - \bar{x} \cdot \bar{z} - \frac{b_x b_z \sum t_i^4}{\sum t_i^2} = b_x b_z \left(\frac{\sum t_i^2}{n} - \frac{\sum t_i^4}{\sum t_i^2} \right).$$

После деления этого элемента на правую часть формулы (9.4) имеем: $b_x b_z$, т.е. точные значения квадратического члена произведения трендов сомножителей.

Но в общем виде из-за наличия дополнительных членов разложения, не равны нулю члены разложения 9, 14, 15, 16, 17 и 18, квадратический член параболы – тренда признака-произведения – не равен аналогичному члену произведения трендов сомножителей, что и видим по данным табл. 9.1.

Что касается свободного члена тренда признака-произведения, то он вычисляется системно вместе с квадратическим членом, а, значит, расхождение последнего с таковым в произведении трендов сомножителей означает, что и свободные члены расходятся. Следовательно, в общем случае $a_y \neq \bar{x} \cdot \bar{z}$, свободный член уравнения параболы тренда при неравенстве нулю квадратического параметра вообще никогда не равен средней арифметической величине признака:

$$a_y \neq \bar{y} = \bar{x} \cdot \bar{z}.$$

Итак, на вопрос о причинах отличия параметров тренда признака-произведения от произведения соответствующих параметров трендов сомножителей можно дать ответ: *параметры тренда признака-произведения при наличии колебаний уровней признаков-сомножителей относительно их трендов содержат дополнительные случайные члены, зависящие от распределения отклонений признаков-сомножителей от тренда по длине ряда и от наличия корреляции между этими отклонениями.*

Можно сказать, что тренд произведения больше зависит от случайностей, чем зависело бы произведение трендов сомножителей. Это положение необходимо учитывать при обсуждении методики прогнозирования системы жестко связанных признаков.

Теперь кратко рассмотрим связи между колебаниями признаков.

Из табл. 9.1. видно, что лишь четыре раза из девяти позиций знак отклонения от тренда валового сбора соответствует знаку произведения отклонений от тренда площади и урожайности. Представляется на первый взгляд, что колебания признаков вообще никак не связаны.

Более точный анализ связи показал, что коэффициенты корреляции между отклонениями от трендов составили:

$$r_{U_n U_y} = -0,326;$$

$$r_{U_n U_b} = -0,047;$$

$$r_{U_y U_b} = 0,725.$$

Следовательно, колебания валового сбора в основном были вызваны колебаниями урожайности, а колебания размеров площади слабо связаны и с колебаниями урожайности и с колебаниями валового сбора.

Что касается интенсивности и силы колебаний, то имеем следующие показатели:

$$S(t)n = 7,4 \text{ га}; V(t)n = 6,2\%; \text{ колеблемость площади слабая};$$

$$S(t)y = 4,4 \text{ ц/га}; V(t)y = 15,2\%; \text{ колеблемость умеренная};$$

$$S(t)b = 633,6 \text{ ц}; V(t)b = 15,3\% \text{ колеблемость умеренная}.$$

Величина каждого отклонения валового сбора от тренда, ввиду несовпадения тренда последнего с произведением трендов площади и урожайности, не равна сумме произведения отклонения площади на

трендовый уровень урожайности плюс произведение отклонения урожайности на трендовую величину площади, как «должно было бы быть». Между отклонениями от тренда нет жесткой функциональной связи: множественный коэффициент детерминации колебаний валового сбора колебаниями площади и урожайности равен лишь 0,566 или 56,6%. Жесткая связь колебаний была бы только при такой же жесткой связи колебаний площади и урожайности. Но такой связи не может быть на практике, ибо причины колебаний размера посевной площади в основном имеют экономическую или организационно-хозяйственную основу, а колебания урожайности в основном имеют причины природного характера.

Итак, можно сделать лишь качественные выводы о связи и силе колебаний жестко-взаимосвязанных признаков:

1) при существенной и прямой связи колебаний факторов-сомножителей колебания признака-произведения будут в среднем сильнее, чем каждого из сомножителей, а при обратной и существенной связи колебаний сомножителей колеблемость признака-произведения будет, в среднем, слабее, чем колеблемость сомножителей;

2) при слабой связи между колебаниями сомножителей, колебания признака-произведения приблизительно такие же, как колебания сомножителя с наибольшей колеблемостью по величине коэффициента $V(t)$;

3) ввиду случайного распределения колебаний сомножителей во времени для изучения их связи необходимо рассмотреть достаточно длинные ряды, не менее 13–15 уровней в каждом.

9.2. Агрегирование трендов и колебаний по совокупности объектов

9.2.1. Тренды объемных признаков

Рассмотрим проблему соотношения тренда и колеблемости по совокупности объектов (например, тренда и колеблемости валового сбора по району в целом) и соотношения трендов колебаний того же показателя в каждой единице совокупности (по каждому хозяйству). Иначе говоря, в отличие от мультипликативной системы, представленной в разд. 9.1., рассмотрим аддитивную систему.

Эта проблема в нашей статистической литературе рассматривалась очень кратко для частного случая И.Поповой²³ и в общем случае В.Н.Афанасьевым²⁴.

Сначала обсудим проблему агрегирования трендов объемных признаков, например, валового сбора. Очевидно, что каждый уровень признака по совокупности хозяйств равен сумме валовых сборов всех единиц этой совокупности:

$$X_i = \sum_{j=1}^k x_j .$$

Средний уровень за ряд лет по совокупности – свободный член линейного тренда – равен, следовательно, сумме свободных членов линейных трендов валового сбора по всем единицам совокупности.

Далее покажем, из чего складывается среднегодовой прирост валового сбора по совокупности:

²³ Попова И. Агрегирование трендов объемных показателей и их колеблемости по совокупности предприятий // Экономико-математические и статистические методы в прогнозировании и управлении сельскохозяйственного производства. – Сб. трудов. – Спб.: Спб. ГАУ, 1993. с. 57-61

²⁴ Афанасьев В.Н. Статистическое обеспечение проблемы устойчивости сельскохозяйственного производства. – М.: Финансы и статистика, 1996.

$$B_x = \frac{\sum_{i=1}^n x_i t_i}{\sum_{i=1}^n t_i^2} = \frac{\sum_{i=1}^n \sum_{j=1}^k x_{ij} t_i}{\sum_{i=1}^n t_i^2} = \sum_{j=1}^k \frac{\sum_{i=1}^n x_{ij} t_i}{\sum_{i=1}^n t_i^2} = \sum_{j=1}^k b_j ,$$

где j – номера единиц совокупности.

Следовательно, средний абсолютный прирост тренда по совокупности в целом равен сумме средних абсолютных приростов по всем единицам совокупности. Таким образом, теорема агрегирования для линейных трендов доказана.

Для параболических трендов средний абсолютный прирост совпадает с таковым для прямой, доказательство уже имеется. Система уравнений МНК для других параметров параболы по совокупности в целом имеет вид:

$$\begin{aligned} nA + C \sum_{i=1}^n t_i^2 &= \sum_{i=1}^n X_i \\ A \sum_{i=1}^n t_i^2 + C \sum_{i=1}^n t_i^4 &= \sum_{i=1}^n X_i t_i^2. \end{aligned}$$

Подставляя в правые части $X_i = \sum_{j=1}^k x_{ij}$, имеем:

$$\begin{aligned} nA + C \sum_{i=1}^n t_i^2 &= \sum_{i=1}^n \sum_{j=1}^k x_{ij} \\ A \sum_{i=1}^n t_i^2 + C \sum_{i=1}^n t_i^4 &= \sum_{i=1}^n \sum_{j=1}^k x_{ij} t_i^2. \end{aligned}$$

Решая эту систему уравнений, получаем:

$$C = \left(\frac{\sum_{i=1}^n \sum_{j=1}^k x_{ij}}{n} - \frac{\sum_{i=1}^n \sum_{j=1}^k x_{ij} t_i^2}{\sum_{i=1}^n t_i^2} \right) : \left(\frac{\sum_{i=1}^n t_i^2}{n} - \frac{\sum_{i=1}^n t_i^4}{\sum_{i=1}^n t_i^2} \right)$$

Вторая скобка не содержит величины признака x_{ij} и в рассмотрении не нуждается. Первая скобка преобразуется в следующее выражение:

$$\sum_{j=1}^k \left(\frac{\sum_{i=1}^n x_{ij}}{n} - \frac{\sum_{i=1}^n x_{ij} t_i^2}{\sum_{i=1}^n t_i^2} \right),$$

что после деления каждого из j слагаемых на вторую скобку дает $\sum_{j=1}^k C_j$, т.е. квадратический параметр параболы по совокупности в целом равен сумме квадратических параметров по всем единицам совокупности. Свободный член параболического тренда по совокупности A вычисляем после нахождения C по формуле:

$$A = \bar{X} - \frac{C \sum_{i=1}^n t_i^2}{n} = \sum_{j=1}^k \bar{x}_j - \frac{\sum_{j=1}^k C_j \sum_{i=1}^n t_i^2}{n} = \sum_{j=1}^k \left(\bar{x}_j - \frac{C_j \sum_{i=1}^n t_i^2}{n} \right) = \sum_{j=1}^k a_j.$$

Таким образом, свободный член параболы по совокупности в целом равен сумме свободных членов уравнений трендов по всем единицам совокупности. Доказана и теорема сложения для параболических трендов. Разумеется, если по части единиц совокупности трен-

ды линейные, а по другим единицам – параболические, то и в этом случае соблюдается правило суммирования трендов. Прямую можно считать частным случаем параболы при нулевом ускорении.

В случае экспоненциальных трендов по каждой единице совокупности, тренд по совокупности в целом также является экспонентой, коэффициент роста которой k является не постоянной, а переменной величиной, в каждом периоде равной средней арифметической взвешенной из индивидуальных темпов k_{ij} по величине уровней предыдущего периода. С течением времени общий темп роста по совокупности асимптотически приближается к величине темпа роста, являющегося наибольшим из всех индивидуальных темпов, так как уровень признака у единицы совокупности с наибольшим темпом роста со временем становится преобладающим в совокупности, его доля стремится к единице. Разумеется, теорема сложения трендов к экспонентам неприменима. Она заменяется теоремой усреднения трендов, которую здесь излагать не будем.

9.2.2. Тренды качественных признаков

Более сложная проблема – агрегирование трендов качественных признаков, таких, как урожайность, производительность труда, коэффициент рентабельности и т.д. Очевидно, что величина каждого уровня качественного признака по совокупности в целом есть средняя взвешенная арифметическая величина из значений данного признака по единицам совокупности, весами являются значения объемного признака – знаменателя изучаемого качественного показателя, для урожайности это площадь посева.

Кратко изложим результат исследования, начиная с простейше-

го случая: при постоянстве весов, т.е. постоянном распределении площади (весового признака) между единицами совокупности, параметры тренда урожайности по совокупности в целом (для всех парабол, включая прямую линию) есть средние взвешенные на доли единиц совокупности в общей площади параметры из всех трендов по каждой единице: $A = \bar{a}; B = \bar{b}$. Таким образом, тренд урожайности по совокупности хозяйств есть средняя величина из трендов по отдельным хозяйствам. При малой колеблемости долей хозяйств в общей площади культуры по совокупности тренд урожайности в совокупности будет приблизительно равен среднему взвешенному тренду отдельных хозяйств. При существенных изменениях в распределении площадей между хозяйствами с разными трендами общий тренд урожайности по совокупности уже не будет равен среднему из трендов по хозяйствам.

Если бы число единиц совокупности было достаточно большим, а изменения их долей в общем объеме признака - веса были случайными, не связанными или слабо связанными с уровнями урожайности и со скоростями ее изменения в отдельных хозяйствах, то, в силу закона больших чисел, параметры тренда урожайности по совокупности в целом в вероятностном смысле приближались бы к их математическому ожиданию, т.е. к среднему из всех индивидуальных трендов. Насколько реальное изменение площадей в совокупности хозяйств отвечает этим условиям, необходимо конкретно исследовать в каждой отдельной задаче.

9.2.3. Агрегирование показателей колеблемости

Ранее доказано, что каждый фактический уровень объемного признака X_i по совокупности в целом равен сумме уровней этого признака для всех единиц совокупности:

$$X_i = \sum_{j=1}^k x_{ji}.$$

Точно так же каждый уровень тренда \hat{X}_i по совокупности есть сумма уровней трендов по единицам совокупности:

$$\hat{X}_i = \sum_{j=1}^k \hat{x}_{ji}.$$

Тогда и каждое отклонение от тренда по совокупности в целом:

$$U_i = X_i - \hat{X}_i = \sum_{j=1}^k x_{ij} - \hat{x}_{ij} = \sum_{j=1}^k u_{ij}.$$

Квадрат отклонения в i -м году от тренда по совокупности в целом равен: $U_i^2 = \left(\sum_{j=1}^k u_{ij} \right)^2$; сумма квадратов отклонений по совокупности

в целом:

$$\sum_{i=1}^n U_i^2 = \sum_{i=1}^n \left(\sum_{j=1}^k u_{ij} \right)^2 = \sum_{i=1}^n \left(\sum_{j=1}^k u_{ij}^2 + 2 \sum_{\substack{m, p \in j \\ m \neq p}}^k u_{im} u_{ip} \right), \quad (9.5)$$

$$m, p \in j; m \neq p$$

Формула (9.5) означает, что сумма квадратов отклонений уровней признака по совокупности от их тренда равна сумме по годам сумм по единицам совокупности квадратов их отклонений от своих

трендов плюс удвоенная сумма произведений отклонений за тот же год отклонений уровней для разных единиц совокупности от своих трендов. Эта последняя удвоенная сумма парных отклонений по всем C_k^2 (сочетание из k по 2) есть удвоенная сумма ковариаций колебаний по всем возможным парам единиц совокупности. Так как коэффициент каждой парной корреляции колебаний – величина

$$r_{u_m u_p} = \frac{\sum_{i=1}^n U_{mi} U_{pi}}{\sqrt{\sum_{i=1}^n U_{mi}^2 \sum_{i=1}^n U_{pi}^2}},$$

то

$$\sum_{i=1}^n U_{mi} U_{pi} = r_{u_m u_p} \sqrt{\sum_{i=1}^n U_m^2 \cdot \sum_{i=1}^n U_p^2} = r_{u_m u_p} \sqrt{\gamma S_m^2(t) \gamma S_p^2(t)} = r_{u_m u_p} \gamma S_m S_p, \quad (9.6)$$

где γ – число степеней свободы (для прямой $\gamma = n - 2$, для параболы $\gamma = n - 3$).

В свою очередь, $\sum_{i=1}^n U_i^2$ по совокупности в целом можно выразить как $\gamma S^2(t)_{сов.}$ по совокупности в целом. Учитывая это и результат (9.6), можно записать вместо (9.5):

$$\gamma S^2(t)_{сов.} = \sum_{j=1}^k \gamma \cdot S_j^2(t) + 2 \sum_1^{C_k^2} r_{u_m u_p} \cdot \gamma S_m(t) \cdot S_p(t)$$

Сократив обе части равенства на число степеней свободы γ , имеем окончательный результат для объемных признаков:

$$S^2(t)_{сов.} = \sum_{j=1}^k S^2(t)_j + 2 \sum_1^{C_k^2} S_m(t) \cdot S_p(t) \cdot r_{u_m u_p} \quad (9.7.)$$

Итак, можно сделать вывод: *дисперсия колебаний признака в целом по совокупности, объемом k единиц, равна сумме дисперсий по всем k единицам плюс удвоенная сумма произведений средних квадратических отклонений по всем сочетаниям единиц совокупности C_k^2 на парные коэффициенты корреляции колебаний.*

Из этого важного вывода вытекает следствие: если бы колебания признака у всех единиц совокупности были независимы друг от друга (все $r_{u_m u_p} = 0$), дисперсия признака по совокупности в целом была бы равна сумме дисперсий признака для всех единиц совокупности.

Например, если в каждом из 20 предприятий района валовой сбор имел бы дисперсию колебаний, равную 9000 ц², то дисперсия валового сбора по району была бы равна 180000 ц². В таком случае имели бы: $S(t)_{сов.} = \sqrt{180000} = 424,26$ ц, в то время как по каждому предприятию $S(t)_j = \sqrt{9000} = 94,87$ ц, и их сумма по 20 предприятиям составила бы $94,87 \cdot 20 = 1897,49$. Отсутствие связи колебаний у разных единиц совокупности, независимость их распределения во времени более чем вчетверо снизили бы величину колебаний признака по совокупности в целом. К сожалению, в границах не только административного района, но даже и области, края, небольшого государства, многие факторы колебаний валового сбора сельскохозяйственных культур являются общими, действующими на всей территории более или менее согласованно. Это означает, что коэффициенты корреляции $r_{u_m u_p}$ в преобладающей части – положительные величины. Если предположить, что в среднем общие факторы объясняют половину колебаний, т.е. $\bar{r}^2 = 0,5$, $\bar{r} \approx 0,7$ то получим следующий результат по (9.7):

$$S(t)_{cov.} = \sqrt{180000 + C_{20}^2 \cdot 94,87^2} = \sqrt{180000 + 190 \cdot 94,87^2} = 1374,8 \text{ ц.}$$

Как видим, и эта величина все еще существенно меньше, чем сумма колебаний по 20 единицам. Так как на практике невозможно, чтобы все факторы колеблемости для всех единиц совокупности были только общими, всегда есть и часть специфических факторов колеблемости для отдельных предприятий, то коэффициенты корреляции отклонений от трендов всегда в среднем меньше единицы, а тогда правая часть выражения (9.7) меньше, чем квадрат суммы колебаний. Итак, имеем *общий закон агрегирования колебаний объемного признака для совокупности хозяйств или любых иных объектов: абсолютная колеблемость объемного признака в совокупности всегда меньше, чем сумма абсолютных мер колеблемости по всем единицам совокупности, и коэффициент колеблемости по совокупности меньше средней величины коэффициентов колеблемости в единицах совокупности:*

$$S(t)_{cov.} < \sum_{j=1}^k S(t)_j,$$

$$V(t)_{cov.} < \bar{V}(t)_j.$$

Если же имеет место обратная корреляция колебаний между единицами совокупности, например, между колебаниями валового сбора в разных регионах большой страны или всего мира, то компенсирующие друг друга колебания могут еще резче снизить общую колеблемость по совокупности и даже свести ее к нулю²⁵.

Данный закон справедлив и для вторичных признаков, как уро-

²⁵ Проблема взаимопогашения колебаний валового сбора и урожайности для России была исследована в ряде работ А.И. Манелли, например, в монографии [20, гл. 2].

жайность. Если бы колебания урожайности у всех единиц совокупности были жестко связаны (т.е. все $r_{u_m u_p}$ были равны единице), то колебания урожайности по совокупности были равны средней из показателей $S(t)_j$ каждой единицы совокупности. Но так как в разных предприятиях, хозяйствах, есть не только общие для совокупности факторы колеблемости, но и специфические, все $r_{u_m u_p} < 1$, а, значит, колебания средней урожайности по совокупности хозяйств, даже если взять простую среднюю, будут меньше, чем среднее квадратическое отклонение по всем единицам. А если еще среднюю урожайность по совокупности вычислить как взвешенную по площадям, то еще и их колебания, тоже, конечно, не строго согласованные по всем единицам совокупности, будут снижать колеблемость средней урожайности по совокупности.

Знающие векторную алгебру легко усвоят закономерности уменьшения колеблемости при агрегировании объектов, если примут во внимание, что колебания – не скалярная величина, а векторная, направление которой – ее распределение во времени. Векторная сумма всегда меньше скалярной суммы векторов, не учитывающей их направленности.

9.3. Корреляция по временным рядам: сущность, ограничения

Предполагается, что читатель знаком с теорией корреляции в пространственных совокупностях и ее показателями, которые здесь используются. Корреляция временных рядов применяется:

– взамен пространственной корреляции, ввиду отсутствия однородной совокупности или данных о таковой. Например, при изучении

связи между средним душевым доходом в стране и душевым потреблением картофеля. Совокупность стран явно неоднородна, не везде потребляется картофель, единственная возможность измерить связь – по данным той же страны за ряд лет;

– при изучении взаимодействующих процессов, например, при изучении связи между урожайностью и колебаниями солнечной активности. Изучать эту связь по пространственной совокупности вообще невозможно: для всех регионов на Земле показатели солнечной активности одинаковы;

– там, где следует применять пространственную корреляцию. Например, дипломник проходил практику в отдельном колхозе, на предприятии, а не в районе. У него нет данных по совокупности хозяйств о внесении удобрений и об урожайности, он берет данные колхоза за 7–11 лет и по ним измеряет связь урожайности с дозой удобрений, получая, как правило, низкий коэффициент корреляции, или даже отрицательный, потому что урожайность разных лет колеблется вовсе не из-за различия доз удобрения, а совсем из-за других причин. Это просто суррогат настоящей пространственной корреляции, к которому прибегать не рекомендуется.

Корреляция между двумя (для простоты возьмем два) признаками означает, что если величина одного из них больше средней по совокупности, то и величина другого в основном тоже больше его средней (прямая связь) или же в основном меньше его средней (обратная связь). Но если оба признака имеют одинаково направленные тренды, то уровни лет после середины периода, как правило, больше средних величин или, при трендах к снижению, оба признака имеют уровни меньше средних. Выходит, что в динамике между любыми признаками, имеющими тенденцию изменения, всегда есть связь: либо прямая

(оба тренда в одном направлении), либо обратная (тренды в разных направлениях). Результат абсурдный. В любой развитой стране в 1970–1990 гг. рос уровень производства компьютеров. Одновременно росло число инфицированных ВИЧ-инфекцией и больных СПИД. Но при очень высокой корреляции уровней обоих рядов, никакой реальной связи процессов нет. Это один из видов ложной корреляции. Как же отличить ложную корреляцию от истинной? Конечно, прежде всего, как и при изучении связей в пространственной совокупности, нужно обосновать связь по существу, объяснить ее причинный механизм. Эта задача не статистическая, в данном учебнике не рассматривается. Она решается специалистом в той сфере знаний, которая изучает объект, процесс, – агрономом, инженером, экономистом, социологом, биохимиком, астрономом и т.д. Без причинного обоснования лучше не начинать измерение связи в динамике.

Но даже и после такого обоснования остается открытым вопрос: при наличии одинаково направленных трендов двух причинно-связанных признаков не преувеличится ли теснота связи за счет трендов? Если, например, в стране растет производство и применение минеральных удобрений, растет и урожайность сельскохозяйственных культур, но ведь она растет не только по причине роста применения удобрений, а также и за счет других факторов – селекции новых сортов, мелиорации, орошения, механизации производства, роста экономической заинтересованности фермеров и еще ряда факторов. А при коррелировании уровней урожайности и доз удобрений за 20–25 лет прогресс всех факторов урожайности будет отнесен на дозу удобрений. Получится коэффициент детерминации, превышающий 50 или даже 70%, и где гарантия, что к истинной корреляции и здесь не примешана ложная? Такой гарантии нет.

Могут возразить: «А разве не может так случиться, что и в пространственной совокупности предприятий, у тех из них, которые вносят больше дозы минеральных удобрений, одновременно и семена лучше, и сельскохозяйственные машины, и кадры более подготовлены, и экономика сильнее?» Да, это возможно, но именно лишь возможно, как возможно и несовпадение факторов, влияющих на урожайность. А параллельная тенденция динамики факторов во времени – это не просто возможность, а в 90% стран и регионов – достоверный факт. Так что примесь ложной корреляции в пространственных совокупностях намного меньше, чем при коррелировании временных рядов. И, следовательно, если есть возможность изучить, измерить, моделировать связь результативного признака с его факторами не по рядам динамики, а в пространственной совокупности, это обязательно следует делать.

Проблема ложной корреляции почти целиком снимается, если причинная связь обоснована не столько между тенденциями динамики, сколько между колебаниями факторного и результативного признаков. Например, колебания урожайности во влагонедостаточных регионах, таких, как Оренбургская область, причинно связана не с какой-либо тенденцией изменения суммы осадков, а с её колебаниями в отдельные годы. К тенденции же роста урожайности осадки никакого отношения (причинной связи) не имеют. Снимается ложная корреляция тем, что колебания других факторов, влияющих на урожайность – экономических, организационных – не связаны или слабо связаны с колебаниями осадков. Тенденции факторов связаны часто, колебания – почти никогда. Поэтому связь между колебаниями одного фактора с результативным показателем (его колебаниями) почти всегда свободна от ложной корреляции, наведенной другими факторами.

В последующих разделах данной главы в основном будут рассматриваться корреляция между колебаниями признаков, а также методики ее измерения и моделирования. Что же касается проблемы измерения связи между тенденциями, между самими уровнями временных рядов, включающих тенденцию, а не только колебания, то эта проблема не может считаться решенной. Некоторые указания читатели учебника могут найти далее в разделе о смешанных прогностических моделях (гл. 10). Излагаемые здесь методики решают только ограниченный класс задач – измерение связи между колебаниями факторного (факторных) признака и колебаниями результативного признака.

Строго говоря, это жесткое ограничение относится и к пространственной корреляции в том смысле, что и в ней измеряется связь вариации результативного признака с вариацией фактора. Например, за счет вариации дозы минеральных удобрений объясняется 38% вариации урожайности пшеницы между хозяйствами области ($r^2 = 0,38$), а не 38% уровня урожайности, как иногда неверно считают.

9.4. Методы измерения корреляции между колебаниями признаков

Итак, в предыдущем разделе было установлено, что единственная «чистая» задача об измерении корреляции временных рядов – это измерение связи между колебаниями их уровней. *Колебания* – это, как правило, случайная составляющая, в отличие от тренда. Если же и колебания не случайны, а строго упорядочены, как, например, сезонные, то и задача о связи таких колебаний не является «чистой», содержит риск ложной связи. Поэтому в дальнейшем рассматривается лишь

случайно-распределенные во времени колебания, например, колебания урожайности.

Классический пример, иллюстрирующий отличие корреляции отклонений от тренда и корреляции уровней ряда – связь, наблюдавшаяся в 1970–1989 гг. в СССР между урожайностью сельскохозяйственных культур и себестоимостью единицы их продукции. Урожайность большинства культур в подавляющей части регионов в 70–80% хозяйств имела тенденцию роста, хотя и медленного, а в отдельных хозяйствах – довольно быстрого. Согласно законам экономики, как рыночной, так и плановой, рост урожайности должен приводить к снижению себестоимости единицы продукции. Однако на самом деле в большинстве, если не во всех хозяйствах и регионах, наоборот, себестоимость имела тенденцию роста. Скрытой причиной этого явления была не признаваемая официально инфляция – рост цен на все элементы затрат на производство: сельхозмашины, энергоносители, удобрения. Рассмотрим пример, представленный в табл. 9.3.

Средние: $\bar{x} = 119,92 \approx 120$; $\bar{y} = 19,0$

Уравнения трендов:

урожайности: $\hat{x} = 119,9 + 3,81t$

себестоимости: $\hat{y} = 19,0 + 1,22t$, где $t=0$ в 1983 году.

Таблица 9.3.

Корреляция урожайности картофеля с его себестоимостью

Период времени	Урожай- ность, ц/га x_i	Себестои- мость, руб./ц y_i	Тренды		Отклонения от трендов		Квадраты от- клонений		Произве- дения $U_{x_i}^2 U_{y_i}^2$	$\hat{y}_i + (-0,124U_{x_i})$, руб./ц
			\hat{x}_i	\hat{y}_i	U_{x_i}	U_{y_i}	$U_{x_i}^2$	$U_{y_i}^2$		
1	108	11,8	97	11,7	+11	+0,1	121	0,01	+1,1	10,3
2	81	15,4	101	12,9	-20	+2,5	400	6,25	-50,0	15,4
3	106	13,0	105	14,1	+1	-1,1	1	1,21	-1,1	14,0
4	124	13,9	108	15,3	+16	-1,4	256	1,96	-22,4	13,3
5	103	15,1	112	16,6	-9	-1,5	81	2,25	+13,5	17,7
6	106	19,6	116	17,8	-10	+1,8	100	3,24	-18,0	19,0
7	149	16,2	120	19,0	+29	-2,8	841	7,84	-81,2	15,4
8	148	17,2	124	20,2	+24	-3,0	576	9,00	-72,0	17,2
9	102	24,0	128	21,4	-26	+2,6	676	6,76	-67,6	24,6
10	130	22,4	131	22,7	-1	-0,3	1	0,09	+0,3	22,8
11	80	32,3	135	23,9	-55	+8,4	3025	70,56	-462,0	30,7
12	139	24,7	139	25,1	0	-0,4	0	0,16	0	25,1
13	183	21,1	143	26,3	+40	-4,9	1600	24	-196,0	21,3
Σ	1559	247,0	1559	247,0	0	0	7678	133,3	-952,7	246,8

Если рассчитывать коэффициент корреляции между уровнями рядов по обычной формуле

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}},$$

то получаем величину $-0,06$, незначимо отличную от нуля. Параллельность трендов урожайности и себестоимости погасила обратную связь их колебаний, что и привело к результату, противоречащему законам экономики.

Рассмотрим теперь другую методику: измерение корреляции между отклонениями уровней от трендов. Подставляя отклонения от трендов в обычную формулу коэффициента корреляции, имеем:

$$r_{U_x U_y} = \frac{\sum_{i=1}^n (U_{x_i} - \bar{U}_x) \cdot (U_{y_i} - \bar{U}_y)}{\sqrt{\sum_{i=1}^n (U_{x_i} - \bar{U}_x)^2 \cdot \sum_{i=1}^n (U_{y_i} - \bar{U}_y)^2}}.$$

Однако так как средние величины отклонений от линейных и параболических трендов всегда равны нулю, а от других форм тренда близки к нулю, если эти формы трендов правильно выбраны, то

$$\bar{U}_x = \bar{U}_y = 0,$$

и формула приобретает вид:

$$r_{U_x U_y} = \frac{\sum_{i=1}^n U_{x_i} U_{y_i}}{\sqrt{\sum_{i=1}^n U_{x_i}^2 \cdot \sum_{i=1}^n U_{y_i}^2}} \quad (9.8)$$

Соответственно, формула коэффициента регрессии также меняется:

$$b = \frac{\sum_{i=1}^n U_{x_i} U_{y_i}}{\sum_{i=1}^n U_{x_i}^2} \quad (9.9)$$

Свободный член уравнения регрессии определяем по обычной формуле: $a = \bar{y} - b\bar{x}$, т.е. для отклонений от трендов $a = \bar{U}_y - b\bar{U}_x = 0$.

Уравнение регрессии имеет вид:

$$U_{y_i} = bU_{x_i} \quad (9.10)$$

Подставляя данные из табл. 9.3, получаем:

$$r_{U_x U_y} = \frac{-952,7}{\sqrt{7678 \cdot 133,3}} = -0,941; r^2 = 88\%$$

Таким образом, колебания себестоимости картофеля в совхозе почти целиком были связаны с колебаниями урожайности, связь обратная, как и требуют законы экономики. И вся она была подавлена тем, что оба тренда имели одно и то же направление по совершенно разным причинам: прогресс агротехники – не причина инфляции и роста цен. Равно как и наоборот: инфляция скорее тормозила прогресс урожайности.

Коэффициент регрессии $b = \frac{-952,7}{7678} = -0,124$, уравнение регрессии:

$U_{y_i} = -0,124U_{x_i}$. Смысл этого уравнения таков: в среднем, отклонение себестоимости от ее тренда в i -м году составляет 0,124 величины отклонения урожайности от своего тренда с обратным знаком. Значения себестоимости, рассчитанные по модели с учетом тренда себестоимости и колебаний урожайности, приведены в последней графе табл. 9.3:

$$\hat{y}(x)_i = (19,0 + 1,22t_i) + (-0,124U_{x_i}).$$

Как видим, полученные по этой модели уровни себестоимости довольно близки к фактическим.

Другим методом измерения корреляции между временными рядами служит метод корреляции цепных показателей динамики, которые являются константами трендов. Для линейных трендов это – абсолютные цепные изменения. Метод предпочтительно применять для таких рядов, в которых среднее изменение (параметр b) существенно меньше, чем среднее колебание $S(t)$, иначе говоря, показатель K значительно меньше единицы.

Логика применения метода заключается в том, что если колеблемость намного больше изменения тренда за единицу времени, то цепные абсолютные изменения, т.е. разности соседних уровней, в основном состоят из колебаний. В связи с этим корреляция абсолютных изменений будет мало отличаться от корреляции отклонений от тренда. Метод имеет и преимущество: не нужно вычислять тренд, ошибка в выборе типа тренда не влияет на конечный результат. Расчет идет непосредственно по исходным временным рядам. По данным табл. 9.4, имеем:

$$\bar{\Delta}_x = +5,57 \approx +5,8; \bar{\Delta}_y = +0,738 \approx +0,74.$$

Таблица 9.4

Корреляция отклонений от средних отклонений

Период времени	x_i	y_i	$\Delta x_i = x_{i+1} - x_i$	$\Delta y_i = y_{i+1} - y_i$	$\Delta_{x_i} - \bar{\Delta}_x$	$\Delta_{y_i} - \bar{\Delta}_y$	$(\Delta_{x_i} - \bar{\Delta}_x) \times (\Delta_{y_i} - \bar{\Delta}_y)$	$(\Delta_{x_i} - \bar{\Delta}_x)^2$	$(\Delta_{y_i} - \bar{\Delta}_y)^2$
1	108	11,8
2	81	15,4	-27	+3,6	-32,8	+2,86	-93,8	1075,8	8,2
3	106	13,0	+25	-2,4	+19,2	-3,14	-60,3	368,6	9,9
4	124	13,9	+18	+0,9	+12,2	+0,16	+2,0	148,8	0,0
5	103	15,1	-21	+1,2	-26,8	+0,46	-12,3	718,2	0,2
6	106	19,6	+3	+4,5	-2,8	+3,76	-10,5	7,8	14,1
7	149	16,2	+43	-3,4	+37,2	-4,14	-154,0	1383,8	17,1
8	148	17,2	-1	+1,0	-6,8	+0,26	-1,8	46,2	0,1
9	102	24,0	-46	+6,8	-51,8	+6,06	-313,9	2683,2	36,7
10	130	22,4	+28	-1,6	+22,2	-2,34	-51,9	492,8	5,5
11	80	32,3	-50	+9,9	-55,8	+9,16	-511,1	3113,6	83,9
12	139	24,7	+59	-7,6	+53,2	-8,34	-443,7	2830,2	69,6
13	183	21,4	+44	-3,3	+38,2	-4,04	-154,3	1459,2	16,3
Σ	1559	247,0	+75	+9,6	—	—	-1805,6	14328,2	261,6

В отличие от отклонений от тренда, средняя величина цепных абсолютных изменений не равна нулю. В связи с этим для расчета параметров корреляции необходимо пользоваться полными формулами, а не сокращенной формулой (9.8). Соответствующие суммы квадратов и произведения отклонений от средних приростов вычислены в табл. 9.4. Исходя из них, имеем:

$$r_{\Delta_x \Delta_y} = \frac{\sum_{i=1}^n (\Delta_{x_i} - \bar{\Delta}_x) \cdot (\Delta_{y_i} - \bar{\Delta}_y)}{\sqrt{\sum_{i=1}^n (\Delta_{x_i} - \bar{\Delta}_x)^2 \cdot \sum_{i=1}^n (\Delta_{y_i} - \bar{\Delta}_y)^2}} = \frac{-1805,6}{\sqrt{14328,2 \cdot 261,6}} = -0,933,$$

что почти совпадает с ранее полученной величиной коэффициента корреляции отклонений от трендов.

Если тренды признаков являются экспонентами, то вместо корреляции отклонений от трендов можно применить метод корреляции цепных темпов роста уровней, поскольку именно темпы роста – основной параметр экспоненциальных трендов.

Остаются недостаточно проработанные вопросы: насколько допустима корреляция абсолютных изменений, если тренды имеют другой вид (гиперболический, логистический, логарифмический и т.д.)?; если тренд факторного признака одного типа, а результативного – другого типа? Достаточного практического опыта для убедительного ответа на эти вопросы у авторов нет, они будут благодарны читателям, если кто-то обладает подобным практическим опытом, или изучал эти вопросы. Еще раз, и не последний, авторы подчеркивают, что наука – открытая система, продолжающийся процесс познания, открытие новых «материков» (реже) и «островов» (чаще) в бесконечном

океане неведомого.

В заключение напомним, что метод корреляции отклонений от трендов – более чистый, основной, он работает независимо от того, одинаковы типы трендов коррелируемых показателей или нет. Прочие методы – суррогаты, имеющие чаще всего, ограничения по типам трендов. Эти методы лучше применять только при явном преобладании колеблемости над тенденцией изменения за единицу времени, т.е. при малом показателе K для линейных трендов или малых аналогичных показателях для других типов трендов (см. разд. 8.3).

9.5. Корреляция с учетом лага и циклов

Среди природных и общественных явлений нередко встречаются такие, которые связаны между собой не в одном и том же периоде времени, а с некоторым запозданием – по-английски – lag, откуда пошел термин *лаг*. Например, капиталовложения в создание машиностроительного, автомобильного завода отразятся в росте объема производства не в том году, когда они произведены, а через два-три и более лет, капиталовложения в строительство крупной гидроэлектростанции – через 6–8 лет. При наличии лага в реальной связи изучаемых явлений измерять корреляцию факторного признака с результативным нужно, конечно, не по одновременным уровням, а с учетом лага. Например, отклонение от тренда капиталовложений скажется на отклонении от тренда выпуска продукции через k лет. Значит, измерять корреляцию нужно через произведения $U_{x(t_i)}, U_{y(t_i+k)}$.

Методика корреляции с учетом лага делится на два подвида.

А. Случай, когда величина лага известна заранее.

Б. Случай, когда саму величину лага следует определить на основе измерения корреляции.

Рассмотрим вначале случай А. Например, на сельскохозяйственном предприятии принят и длительное время действует следующий севооборот: после трех лет многолетних трав участок занимает пропашная культура: картофель, бобовые, овощи, под которые вносятся высокая доза органических удобрений, а в следующем году на участке высевают зерновые культуры. Необходимо измерить связь между дозой органических удобрений, внесенных под пропашные культуры, и урожайностью зерновых. В данном случае $k = 1$ году и расчет корреляции приведен в табл. 9.5. При этом будем считать, что тренд дозы внесенных органических удобрений отсутствует или незначителен.

Средняя доза удобрений: $\bar{x} = 451:11 = 41$ т/га.

Тренд урожайности: $\hat{y}_i = 18,0 + 0,6 \cdot t_i$; $t = 0$ в 1992 г.

Коэффициент корреляции с учетом лага в 1 год имеет вид:

$$r_{\Delta x_i U_{y(i+1)}} = \frac{\sum_{i=1}^{n-1} \Delta x_i U_{y_{i+1}}}{\sqrt{\sum_{i=1}^{n-1} \Delta x_i^2 \cdot \sum_{i=2}^n U_{y_i}^2}} = \frac{70,4}{\sqrt{305 \cdot 46,44}} = +0,5915, r^2 = 0,35 \text{ или } 35\% .$$

Связь колебаний дозы удобрений под предшественника зерновых с колебаниями их урожайности на следующий год оказалась средней силы: за счет этой связи объясняется 35% всей колеблемости урожайности.

Коэффициент регрессии $b(x) = \frac{70,4}{305} = 0,2308$, т.е. 1 т удобрений под пропашные культуры в среднем давала прибавку урожайности зерновых на следующий год на 0,23 ц/га.

Таблица 9.5.

Корреляция с лагом в 1 год

Период времени	Доза, т/га x_i	Урожайность, ц/га y_i	Тренд \hat{y}_i	Отклонение U_{y_i}	$x_i - \bar{x} = \Delta_{x_i}$	$\Delta x_i U_{y_{i+1}}$	$\Delta_{x_i}^2$	$U_{y_i}^2$	\hat{y}'
1	45	13,0	15,0	-2,0	+4	+9,6	16	4,00	15,0
2	36	18,0	15,6	+2,4	-5	-1,5	25	5,76	16,5
3	47	16,5	16,2	+0,3	+6	-10,2	36	0,09	15,0
4	33	15,1	16,8	-1,7	-8	+23,2	64	2,89	18,2
5	42	14,5	17,4	-2,9	+1	0	1	8,41	15,6
6	51	18,0	18,0	0	+10	+38,0	100	0,00	18,2
7	36	22,4	18,6	+3,8	-5	+2,5	25	14,44	20,9
8	42	18,7	19,2	-0,5	+1	+3,2	1	0,25	18,0
9	35	23,0	19,8	+3,2	-6	+3,6	36	10,24	20,0
10	40	19,8	20,4	-0,6	-1	+2,0	1	0,36	19,0
11	44	19,0	21,0	-2,0	+3	...	9	4,00	20,8
Σ	451	198,0	198,0	0	0	+70,4	314	50,44	197,2

Уравнение регрессии имеет вид: $U_{y_{i+1}(x)} = 0,2308\Delta x_i$, свободного члена это уравнение не имеет, так как средние отклонения от тренда и от средней дозы равны нулю. Рассчитанные по этой формуле значения урожайности, то есть трендовые значения $\hat{y}_i + U_{y(x)_{i+1}}$ даны в последней графе табл. 9.5.

Обратите внимание на особенности сумм произведений и сумм квадратов в формулах коэффициента корреляции и коэффициента регрессии: в сравнении с суммами при корреляции отклонений без лага, число слагаемых на единицу меньше: в одной из сумм – от конца, в других – от начала. Если же лаг велик, то число слагаемых сильно сократится, а значит, корреляция станет менее надежной: ведь оценка надежности коэффициентов должна рассчитываться в этом случае не по общему числу членов первичного ряда, а исходя из числа реально участвующих в работе коэффициентов. При лаге в 5 лет это число составит $n - 5$, а затем еще надо исключить 2 степени свободы при парной корреляции. Откуда следует еще один вывод: при коротком исходном ряде (рядах) и большом лаге показатели связи колебаний признаков будут заведомо ненадежны.

Теперь рассмотрим случай Б, когда величина лага заранее неизвестна и должна быть определена с помощью корреляционного анализа. Имея в данном случае дело с недостаточно изученными явлениями, назовем коррелируемые признаки "икс" и "игрек". Если их временные ряды достаточно велики, находим тренды \hat{x} и \hat{y} , отклонения отдельных уровней от трендов U_{x_i} , U_{y_i} и начинаем вычислять корреляцию между ними: сначала без лага, затем с лагом один период, с лагом два периода и т.д. Получается серия (или вектор) коэффициентов корреляции между колебаниями признаков x и y с возрастающим лагом.

Графическое изображение этого вектора принято называть *коррелограммой*.

Коррелограмма может иметь два вида:

- коэффициенты до какого-то сдвига растут, а затем убывают до незначимо отличных от нуля величин, тогда лаг считается равным тому сдвигу отклонений, при котором коэффициент корреляции по модулю максимален;
- коэффициенты поочередно растут и убывают, образуя циклы или квазициклы, т.е. локальные максимумы наблюдаются, скажем, то через три года, то через четыре года. Лагом в этом случае считается средний промежуток времени между локальными максимумами коэффициентов корреляции, между отклонениями от трендов.

Рассчитываем коэффициенты корреляции отклонений от тренда, начиная с нулевого лага (таблица 9.6.):

$$r_{u_x u_{y_i}} = \frac{\sum_{i=1}^n U_{x_i} U_{y_i}}{\sqrt{\sum_{i=1}^n U_{x_i}^2 \cdot \sum_{i=1}^n U_{y_i}^2}} = \frac{109,5}{\sqrt{1058 \cdot 100,72}} = 0,335;$$

$$r_{u_x u_{y_{i+1}}} = \frac{\sum_{i=1}^{n-1} U_{x_i} U_{y_{i+1}}}{\sqrt{\sum_{i=1}^{n-1} U_{x_i}^2 \cdot \sum_{i=2}^n U_{y_i}^2}} = \frac{-173,2}{\sqrt{1057 \cdot 84,72}} = -0,579;$$

$$r_{u_x u_{y_{i+2}}} = \frac{\sum_{i=1}^{n-2} U_{x_i} U_{y_{i+2}}}{\sqrt{\sum_{i=1}^{n-2} U_{x_i}^2 \cdot \sum_{i=3}^n U_{y_i}^2}} = \frac{-49,8}{\sqrt{1057 \cdot 73,83}} = -0,178;$$

$$r_{u_x u_{y_{i+3}}} = \frac{\sum_{i=1}^{n-3} U_{x_i} U_{y_{i+3}}}{\sqrt{\sum_{i=1}^{n-3} U_{x_i}^2 \cdot \sum_{i=4}^n U_{y_i}^2}} = \frac{184}{\sqrt{1048 \cdot 63,59}} = 0,713;$$

$$r_{u_{x_i} u_{y_{i+4}}} = \frac{\sum_{i=1}^{n-4} U_{x_i} U_{y_{i+4}}}{\sqrt{\sum_{i=1}^{n-4} U_{x_i}^2 \cdot \sum_{i=5}^n U_{y_i}^2}} = \frac{-57,4}{\sqrt{823 \cdot 51,34}} = -0,279;$$

$$r_{u_{x_i} u_{y_{i+5}}} = \frac{\sum_{i=1}^{n-5} U_{x_i} U_{y_{i+5}}}{\sqrt{\sum_{i=1}^{n-5} U_{x_i}^2 \cdot \sum_{i=6}^n U_{y_i}^2}} = \frac{-104,7}{\sqrt{787 \cdot 51,09}} = -0,522.$$

Таблица 9.6

Корреляция отклонений от тренда с неизвестным заранее лагом

Период	t_i	x_i	\hat{x}_i	U_{x_i}	y_i	\hat{y}_i	U_{y_i}	$U_{x_i}^2$	$U_{y_i}^2$
1	-7	67	80	-13	18,2	22,2	-4,0	169	16,00
2	-6	84	82	+2	26,7	23,4	+3,3	4	10,89
3	-5	95	84	+11	27,8	24,6	+3,2	121	10,24
4	-4	77	86	-9	22,3	25,8	-3,5	81	12,25
5	-3	100	88	+12	27,5	27,0	+0,5	144	0,25
6	-2	98	90	+8	30,6	28,2	+2,4	64	5,76
7	-1	93	92	+1	26,8	29,4	-2,6	1	6,76
8	0	91	94	-3	31,9	30,6	+1,3	9	1,69
9	1	91	96	-5	32,3	31,8	+0,5	25	0,25
10	2	111	98	+13	31,9	33,0	-1,1	169	1,21
11	3	94	100	-6	30,7	34,2	-3,5	36	12,25
12	4	87	102	-15	38,4	35,4	+3,0	225	9,00
13	5	107	104	+3	39,5	36,6	+2,9	9	8,41
14	6	106	106	0	37,8	37,8	0	0	0
15	7	109	108	+1	36,6	39,0	-2,4	1	5,76
Σ	0	1410	1410	0	459	459	0	1058	100,72

Нет смысла продолжать корреляцию, так как остается все меньше и меньше слагаемых в суммах, и коэффициент становится все более случайным. Можно сделать достаточно уверенно вывод о том, что лаг равен трем годам, так как коэффициент с лагом три года довольно резко выделяется. Такой вывод будет справедлив, если по существу известно, что связь должна быть прямой, например, x – капиталовложения, млрд. руб., y – ввод в эксплуатацию жилой площади, млн. м². Если же неизвестен априори не только лаг, но даже и направление связи, то следует проверить и альтернативную гипотезу: обратную связь при лаге в два года.

9.6. Понятие о динамике комплекса статистически взаимосвязанных признаков.

В разд. 9.1. была рассмотрена методика изучения динамики комплекса жестко взаимосвязанных признаков. Здесь мы представим лишь пути исследования динамики статистически взаимосвязанных, в основном корреляционно-связанных признаков²⁶.

Очень упрощенно, проблема состоит в следующем: имеем пространственную многофакторную регрессионную модель:

$$\hat{y}_i = a + \sum_{j=1}^k b_j x_{ji} ,$$

выражающую корреляционную связь результативного признака y от

²⁶ Это сложнейшую проблему в отечественной статистической литературе разрабатывали А.А. Френкель, О.П. Крастинь, С.П. Бобров, Н.С. Четвериков, Б.С. Ястремский, Н.К. Дружинин, Р.П. Рудакова, авторы данного учебника.

ряда факторных признаков x_1, x_2, \dots, x_k с коэффициентом детерминации $R^2_{yx_1 \dots x_k}$. С развитием процесса во времени могут изменяться значения факторных признаков, теснота их связи с результативным признаком y , а также относительная роль данного комплекса факторов в общей вариации результативного показателя.

Самая простая задача – разложение изменения результативного признака в текущем периоде в сравнении с базисным за счет:

- изменения величин факторных признаков;
- изменения коэффициентов регрессии при факторах;
- прочих причин.

Чтобы разложить общее изменение средней величины результативного признака на элементы, используем обычную последовательность индексирования: сначала изменяем значения объемных показателей, которыми в данной задаче являются средние величины факторных признаков при сохранении базисных значений качественных показателей. В данной задаче это коэффициенты регрессии, измеряющие эффект каждого фактора. Далее индексируем коэффициенты регрессии при неизменных на уровне отчетного периода средних значениях. Рассмотрим пример по изменению среднего надоя молока на корову в отчетном периоде по сравнению с базисным и системы факторных признаков:

y – средний удой молока на корову, кг/год;

x_1 – затраты труда на 1 среднегодовую корову, ч;

x_2 – доля концентратов в рационе коров, %;

x_3 – доля корнеплодов в рационе коров, %;

x_4 – средняя длительность сухостойного периода, дни;

x_5 – средняя длительность сервис-периода, дни;

x_6 – доля коров классов «элита», «элита-рекорд» в стаде, %;

x_7 – общий объем рациона коров, ц корм. ед./год;

x_8 – оценка сельхозугодий, баллы.

За базисный год регрессионная модель имела вид:

$$\hat{y}_{82} = 3,36x_1 + 11,02x_2 + 16,69x_3 - 6,07x_4 - 2,92x_5 + 8,08x_6 + 62,51x_7 + 3,48x_8 - 64,50$$

За текущий год регрессионная модель приняла следующий вид:

$$\hat{y}_{83} = 2,95x_1 + 18,4x_2 + 19,55x_3 - 5,70x_4 - 2,95x_5 + 8,54x_6 + 52,74x_7 + 4,37x_8 + 124,1$$

Средние значения признаков за оба периода приведены в табл.

9.7.

Таблица 9.7

Значения факторных признаков

Год	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	y	R^2
Базисный	145,3	34,80	9,00	74,2	105,6	41,7	42,1	29,6	3270,2	0,926
Отчетный	146,4	31,68	9,18	73,6	102,3	46,1	44,0	29,6	3440,7	0,909

Изменения средних значений факторов за год невелики. Существенно возросла доля элитных коров в стаде, зато снизилась доля концентратов в рационе. Изменения коэффициентов регрессии более значительны, возросли коэффициенты при x_2 и x_3 , снизился коэффициент при x_7 , что может указывать на ухудшение качества кормов.

Общий индекс среднего надоя молока от коровы составил:

$$\frac{3440,7}{3270,2} \cdot 100\% = 105,2\% , \text{ абсолютное изменение: } +170,5 \text{ кг/гол.}$$

Изменение среднего надоя за счет изменения средних величин признаков определяется по формуле:

$$I_{\bar{y}(\bar{x})} = \frac{\sum_{j=1}^k b_{0j} \bar{x}_{1j} + a_0}{\sum_{j=1}^k b_{0j} \bar{x}_{0j} + a_0} = \frac{3410,2}{3270,2} = 1,0428 \text{ или } 104,28\% .$$

Абсолютное изменение составило +140 кг от коровы. Таким образом, изменение всего факторного комплекса в целом было благоприятным для повышения продуктивности коров. Данный индекс можно разложить на восемь субиндексов, последовательно меняя в числителе по одному фактору, начиная от его базисного значения и до отчетного. Спорной будет последовательность изменения факторов, которую желательно, если это возможно, согласовать с реальной последовательностью в процессе производства.

Изменение среднего надоя за счет параметров модели, которые можно условно трактовать как изменение эффекта факторного комплекса, определяется по формуле

$$I_{\bar{y}(b_j, a)} = \frac{\sum_{j=1}^k b_{1j} x_{1j} + a_1}{\sum_{j=1}^k b_{0j} x_{1j} + a_0} = \frac{3440,7}{3410,2} = 1,0089 \text{ или } 100,89\% .$$

Абсолютное изменение составило +30,5 кг от коровы. Следовательно, несмотря на разнонаправленные изменения коэффициентов регрессии, преобладающим (или равнодействующей) оказалось повышение эффекта факторов.

Другая задача, решаемая с помощью анализа динамики корреляционно-связанного комплекса, это получение более надежных параметров связи, абстрагированных от колебаний под влиянием усло-

вий отдельных лет. Одногодичная модель связи может оказаться нетипичной. Например, в засушливый год влияние дозы удобрений на урожайность окажется очень слабым, так как растения не в состоянии их использовать. Как же получить типичную регрессионную модель? Эту задачу подробно рассмотрел О.П. Крастинь²⁷. Им показано, что лучше применять усреднение коэффициентов годичных моделей за ряд лет, чем предварительное усреднение информации за ряд лет, а затем вычисление по ней показателей связи. В последнем приеме возможно возникновение неинтерпретируемых «парадоксальных» коэффициентов регрессии, либо больших, чем все годовые, либо меньших. Также нарушает условия применения МНК так называемый *метод за-водо-лет* (или *хозяйство-лет*) – простое суммирование (без усреднения) первичной информации за ряд лет, особенно, если совокупность невелика.

Условиям задачи наиболее полно отвечает ковариационный анализ, при котором различия, обусловленные особенностями периодов времени, полностью отделяются от вариаций результативного показателя за счет факторных признаков. (В данном учебнике достаточно сложные методы ковариационного анализа, введение дополнительных переменных для периодов времени не рассматриваются.)

К наиболее трудным задачам исследования динамики комплекса корреляционно-связанных признаков – это изучение динамики регрессионной модели и формирование ее прогнозируемой формы на будущий период. Эта проблема в нашей статистике подробно рассмотрена в работах А.А.Френкеля²⁸. В кратком изложении разрабо-

²⁷ Крастинь О.П. Изучение статистических зависимостей по многолетним данным. – М.: Финансы и статистика, 1981.

²⁸ Френкель А.А. Математические методы анализа динамики и прогнозирования производитель-

танный им алгоритм построения прогнозируемой модели связи состоит из следующих этапов.

1. Строятся одногодичные модели связи за 9–12 последовательных лет.

2. Изучается динамика каждого коэффициента условно-чистой регрессии: тип его тренда, надежность уравнения тренда. При этом по разным коэффициентам могут применяться разные типы тренда, включая модель экспоненциального сглаживания.

3. Вычисляются прогнозируемые значения коэффициентов условно-чистой регрессии на заданный период: b_{jk} – для каждого фактора, а также β_{jk} .

4. Анализируются тенденции средних значений факторов и сопоставляются с тенденциями коэффициентов регрессии при них. Отмечено, что часто при увеличении среднего значения фактора наблюдается сокращение коэффициента регрессии. (Это замечено и авторами данного учебника: снижение эффекта фактора может быть результатом несистемного изменения значений факторов; эффективно только увязанное технологически и экономически изменение всей системы, а не отдельного только ее элемента.)

5. Оцениваются ошибки прогнозов для коэффициентов регрессии и для ожидаемых значений факторных признаков.

6. Вычисляется точечный прогноз результативного признака и его доверительные границы.

Применение данного метода может быть ограничено следующими обстоятельствами. По пункту 1: в модель следует включать за все годы одни и те же факторы, однако в отдельные годы тот или иной

ности труда. – М.: Экономика, 1972.

фактор может оказаться статистически несущественным. По пункту 2: как поступить с факторами, коэффициенты при которых неустойчивы и тем более, если они в разные годы меняют знак? По пункту 4: нельзя изменять значения факторов так, что их системный характер разрушится. При разных типах трендов факторов это вполне может случиться. При разных типах трендов факторов это вполне может случиться. По пункту 5: сам А.А.Френкель отметил, что «... было не ясно, как определять ошибки прогноза по факторам x_3 и x_4 , доверительные интервалы для выработки по моделям (9.4.3) и (9.4.4) построены не были»²⁹.

Указанные трудности следует иметь в виду при построении ожидаемого на будущий период уравнения многофакторной или парной регрессии.

Динамика системы корреляционно-связанных признаков может изучаться не только по рядам пространственных регрессионных моделей, но и по динамическим рядам уровней взаимосвязанных признаков. Методика такого изучения впервые была разработана русскими статистиками: С.П. Бобровым и Б.С. Ястремским, Н.С. Четвериковым и получила у них название метода *переменной корреляции*.

Несколько ранее американский статистик-экономист У. Персонс предложил, чтобы убедиться в надежности коэффициента корреляции между динамическими рядами, лучше не ограничиваться вычислением его за весь период в целом, и, кроме этого разделить период на две части и по каждой половине тоже рассчитать коэффициент, и если коэффициенты по двум полупериодам и периоду в целом не близки друг к другу, то доверять им не следует. Однако У. Персонс не ставил вопроса об изучении динамики силы связи.

²⁹ Френкель А.А. Математические методы анализа динамики и прогнозирования производительности труда. – М.: Экономика, 1972. с.173

С.П. Бобров и Б.С. Ястремский в 1922–1923 гг. пошли по другому пути. Они впервые поставили вопрос о важном значении, которое может иметь в экономике, биологии и других науках факт изменения тесноты связи между факторным и результативным признаком, и разработали свои методики решения этой задачи.

Б.С. Ястремский³⁰ предложил для изучения изменения тесноты связи во времени построить ряд нарастающих числителей коэффициента, т.е. сумм $\sum_{i=1}^n x_i \cdot y_i$, если эти суммы с ростом n возрастают равномерно, по прямой, то теснота связи признаков постоянна, а если по кривой, то теснота связи эволюционирует. Сам коэффициент

Б.С.Ястремский вычислял по формуле $r_{xy} = \frac{\sum_{i=1}^n x_i y_i}{n \sigma_x \sigma_y}$. При этом σ_x и σ_y он

считал неизменными на уровне всего динамического ряда.

С.П. Бобров справедливо упрекает Ястремского за принятие гипотезы о постоянстве «сигм», так как они также могут эволюционировать, как и числитель. Кроме этого, у Б.С. Ястремского нет ясности о том, какие величины он принял за x_i и y_i – уровни рядов динамики или же отклонения уровней от трендов.

Б.С. Бобров предложил вычислять *скользящий коэффициент корреляций* между временными рядами аналогично скользящей средней³¹. При этом он считает, что не может быть общего, единообразного правила выбора шага сдвига и длины базы расчета скользящего коэффициента корреляции. Эти вопросы должны, по мнению исследователя, решаться конкретно для каждой взаимосвязи с учетом экономи-

³⁰ Ястремский Б.С. Переменная корреляция // Вестник статистики. – 1924. – Кн. XVII.

³¹ Бобров С.П. Конъюнктурные кривые 1922– 1923 гг. //Экономическое обозрение. – 1924. – № 5 – 7.

ческих факторов. Заметим, что и у Боброва также нет определенного указания на то, что речь должна идти о корреляции между отклонениями уровней от трендов, а не самих уровней двух рядов.

Идея же скользящего расчета коэффициента – несомненная заслуга С.П. Боброва, она вписывается в целое семейство аналогичных методов статистики: скользящую среднюю, как приближенное выражение тренда, скользящее многократное выравнивание ряда для более надежного измерения параметров тренда, и скользящего расчета коэффициента корреляции временных рядов.

Наиболее совершенную методику изучения изменений тесноты связи между временными рядами предложил Н.С. Четвериков (1885–1973), ученик А.А. Чупрова. В его работе, написанной совместно с Е.В. Лунеевой³² в 1924 г., не только изложена методика анализа, но и проведено ее приложение к связи урожайности основных зерновых культур с ценами на зерно данной культуры в России за 1890–1913 гг.

Методика Н.С. Четверикова, который сам написал работу о «ложной корреляции» уровней временных рядов, прежде всего предусматривает вычисление трендов факторного и результативного признаков. Корреляция изучается только по рядам отклонений от трендов. Строятся три кумулятивных ряда:

– нарастающие итоги произведений отклонений от трендов – в наших обозначениях, принятых в данной главе, это ряды сумм

$$\sum_{i=1}^m U_{xi} U_{yi}, \text{ где } m \text{ от } 1 \text{ до } n;$$

– нарастающие итоги квадратов отклонений каждого признака от своего тренда:

³² Четвериков И.С. Статистические и стохастические исследования. – М.: Госстатиздат, 1963.

$$\sum_{i=1}^m U_{xi}^2 \text{ и } \sum_{i=1}^m U_{yi}^2 ;$$

– все три нарастающие ряда выравняются методом наименьших квадратов по наиболее подходящей линии, причем Н.С. Четвериков не советует брать сложные кривые, лучше ограничиваться прямой или параболой II порядка.

Коэффициенты корреляции получаются для любого периода t_m делением ординаты первой выровненной линии, т.е. $\sum_{i=1}^m U_{xi}U_{yi}$ для года номер t на корень квадратный из произведения ординат нарастающих сумм квадратов отклонений, т.е. на $\sqrt{\sum_{i=1}^m U_{xi}^2 \cdot \sum_{i=1}^m U_{yi}^2}$. Это построение полностью отвечает изложенной ранее в данной главе методике измерения корреляции временных рядов. Но Четвериков идет дальше, вычисляя r_{u_x, u_y} для многих периодов скользящим способом: конкретно для периодов от t_m до t_n , где $n > m$, а m достаточно велико для получения надежной меры тесноты связи. В результате исследователь получает ряд коэффициентов корреляции для периодов от номера t_m до t_n и может судить об эволюции тесноты связи факторного признака с результативным.

Можно даже предложить еще одну стадию анализа, так как полученный ряд скользящих коэффициентов корреляции, хотя и был по возможности абстрагирован от случайностей при выравнивании первичных рядов и при суммировании за m лет нарастающих сумм произведений отклонений и квадратов отклонений, но и после этого, скорее всего, коэффициенты корреляции могут иметь колебания. А значит, ряд скользящих коэффициентов корреляции можно снова выравнять методом наименьших квадратов по той или иной линии, получив

уравнение тренда коэффициента корреляции между отклонениями от своих трендов уровней факторного и результативного признаков. Однако сам Н.С. Четвериков воздерживается от такого предложения и указывает на ограничения и недостатки предложенной им методики: 1) методика, по мнению ученого, пригодна при плавных изменениях уровней первичных рядов и силы связи признаков; 2) параметры выравнивающих линий – трендов – имеют ошибки, особенно в начале и в конце рядов, и из-за этого скользящие коэффициенты корреляции иногда выходят за пределы допустимых значений коэффициента от +1 до -1; 3) выбор типа тренда тоже может содержать ошибку.

Н.С. Четвериков³³ рассчитывал коэффициенты корреляции урожайности с ценами на зерно, учитывая лаг в основном: цена августа следующего года с урожайностью предыдущего, но и с другими лагами от 1 месяца до целого года до окончания сбора следующего урожая. Лаг зависит от конкретной задачи, сущности изучаемого процесса, о чем уже сказано ранее. Метод Н.С. Четверикова и сейчас сохраняет свое значение, является важным вкладом отечественной статистики в методологию изучения связей и динамики.

³³ Четвериков И.С. Статистические и стохастические исследования. – М.: Госстатиздат, 1963.

Глава 10. МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ (ДИНАМИКИ).

В данной главе рассмотрим следующий за анализом этап – построение модели развития изучаемого показателя и прогнозирование его возможных значений на будущее. Собственно, уравнение тренда (см. гл. 5) уже есть модель временного ряда. В гл. 6, в частности, в разделе о сезонных (и иных циклических) колебаниях, получены и некоторые модели колеблемости. Остается свести их в общую модель изменения изучаемого показателя с течением времени и оценить возможность прогнозирования его будущих значений.

Прогноз (в переводе с греческого языка – предвидение, предсказание, предзнание) – неотъемлемая составляющая всей человеческой деятельности, в том числе экономической. Это промежуточное звено между познанием объективной реальности и деятельностью людей по ее преобразованию. Один из основоположников позитивизма Огюст Конт (1798–1857) говорил: «*Savoir pour prévoir; prévoir pour agir*» (знать, чтобы предвидеть; предвидеть, чтобы действовать).

Самые разные прогнозы – от прогноза погоды на завтра до прогноза результатов президентских выборов – составляют значительную часть информации, циркулирующей в обществе. Разработкой прогнозов рынка сбыта, финансовых потоков, курса валют и других важнейших показателей деятельности заняты тысячи, если не миллионы работников банков, фирм, государственных органов, частных компаний.

Создание методов прогнозирования – одна из главных проблем науки, и, может быть, труднейшая из них. Не случайно ученый-геолог, писатель-фантаст и один из самых глубоких мыслителей России XX в. Ив. Ант. Ефремов (1907–1972) предусмотрел в далеком бу-

душем человечества наличие специальной «Академии Стохастики и Прогнозирования» для изучения возможных рисков при осуществлении проектов изучения других звездных систем и крупных проектов на Земле. Увы, сейчас нет ни такой академии, ни методики предсказания землетрясений, ни погоды, хотя бы на полгода вперед... Излагаемые в данной главе методы, как будет показано, имеют серьезные ограничения, которые нужно хорошо знать пользователям. Но задача настолько важна, что любой, пусть и несовершенный, ограниченный метод прогнозирования заслуживает внимательного изучения и проверки в практической деятельности.

10.1. Сущность и условия прогноза по тренду с учетом колеблемости

Рассказывают, что однажды к древнегреческому философу Диогену Синопскому (ок. 400 – ок. 325 до н.э.), проживавшему в бочке на берегу залива, обратился неизвестный путник с посохом и мешком за плечами: «Скажи, мудрый человек, дойду ли я отсюда к закату до Афин?» Диоген посмотрел на стоящего путника, и сказал ему: «Иди!» – «Но я же тебя спрашиваю, дойду ли я до Афин засветло?» – повторил странник. – «Иди!!» – еще громче, сердито закричал Диоген. Путник пожал плечами и пошел вдоль берега. Диоген смотрел вслед ему некоторое время и закричал: «Вернись!» Путник вернулся. «Вот теперь я могу тебе сказать, что до заката солнца ты до Афин не дойдешь. Лучше оставайся до завтра у меня». – «А что же ты мне сразу не сказал, зачем прогнал меня?» – «А как же я скажу, дойдешь ли ты к закату до Афин, если я не видел, как быстро ты идешь?»

В этом предании выражена, можно сказать, суть прогнозирова-

ния по тренду: чтобы знать, какого уровня достигнет тот или иной «идущий» процесс, скажем, через пять лет, нужно знать среднюю скорость изменения уровня за год, т.е. знать параметры тренда.

Более того, притча о Диогене содержит и ограничения прогноза по тренду. Представим себе, что путник, спросивший «прогноз» у Диогена, был бы хорошим атлетом и, услышав неудовлетворительный прогноз, взял бы да побежал в Афины бегом, таким образом, опровергнув прогноз Диогена! Ведь и в экономике предприятие или другой объект прогноза, могут принять меры к ускорению движения в сравнении с прежним трендом, и прогноз по нему не оправдывается. Однако, и в этом случае прогноз вовсе не бесполезен, наоборот, он сыграл роль «предупреждения» о необходимости изменить скорость процесса. Роль предупреждающего прогноза не в том, чтоб он исполнился, наоборот, его роль именно в том, чтобы менеджер фирмы, агроном, банкир, правительство страны приняли меры, не допускающие исполнения прогноза.

Тренд производственных показателей не всегда может быть изменен даже в отдельном предприятии. Для этого необходимы средства – капитал, знания (ноу-хау), воля менеджера, квалифицированные и заинтересованные в прогрессе предприятия работники. Если эти условия налицо – прогноз по-прежнему тренду сохраняет только значение предупреждающего. Если же указанные условия изменения тренда отсутствуют, то прогноз по тренду осуществится на деле. Как говорил В. Черномырдин: «Хотели как лучше, а получилось, как всегда!», т.е. хотели изменить тренд, но не сумели.

Если же объектом прогнозирования является крупная система, например, сельское хозяйство региона, страны, то изменить тренд в короткие сроки, как правило, невозможно: для этого потребовались

бы нереально большие средства. Невозможно за пять-шесть лет существенно изменить плодородие почв области, чтобы резко увеличить урожайность. Тем более, за десяток лет не изменится тренд численности народонаселения Земли. Не остановится и не замедлится существенно тенденция роста потребления человечеством топлива и других источников энергии, а значит, и тенденция роста средней температуры воздуха. Отсюда вывод:

- для крупных систем и объектов, обладающих большой инерционностью развития, прогноз по тренду за предыдущее время, как правило, возможен и реален;
- второе условие возможности прогноза по тренду связано с надежностью его параметров, рассмотренной в гл. 7. Если эти параметры ненадежны, ненадежен и прогноз;
- период прогнозирования, т.е. срок удаления прогнозируемого уровня во времени от конца базы расчета тренда, должен быть не более трети, в крайнем случае, половины длительности базы (так рекомендуют, как правило, пособия по статистическому прогнозированию). Чем дальше удален прогнозный уровень от базы расчета тренда, тем больше ошибка прогноза, как будет показано в дальнейшем.

Прогноз по тренду – лишь один из статистических методов прогнозирования. Полезно сравнить его свойства, положительные и негативные, со свойствами прогнозирования на основе многофакторных регрессионных моделей. Начнем с положительных свойств прогноза по тренду. Коэффициент при номере периода в уравнении тренда (b – в линейном уравнении) – это комплексный *коэффициент регрессии* при всех реальных факторах, влияющих на уровень изменяющегося показателя, которые сами изменяются во времени. Подчеркнем: при всех факторах! Ни в одну факторную регрессионную модель мы не

можем включить все факторы, влияющие на изучаемый показатель, например, на урожайность. Во-первых, часть факторов вообще неизвестна, так как наши знания, наука не имеют статуса абсолютной, полной истины. Во-вторых, часть факторов теоретически известна, но на практике по ним нет достаточно надежной или даже вообще никакой информации. В-третьих, если число известных факторов велико, то всех их явно невозможно включить в уравнение регрессии по математическим ограничениям: мультиколлинеарность, гетероскедастичность, превышение числа факторов над численностью выборки и т.п. Таким образом, уравнение тренда имеет преимущество в охвате (хотя и в неявной форме) всех факторов изменения уровней прогнозируемого показателя.

Второе преимущество состоит в том, что уравнение тренда есть модель динамики процесса, и на основании него мы прогнозируем динамику, т.е. логическая основа соответствует задаче. Напротив, уравнение многофакторной регрессии – это модель вариации уровня показателя в статической совокупности. Эта модель объясняет не изменение, например, урожайности во времени, а ее различия в совокупности хозяйств в данный период. Логическая база прогноза по многофакторной регрессии в статике неадекватна задаче прогнозирования. Конкретный пример: один из главных факторов вариации урожайности в регрессионной модели – тип почвы, почвенная разность, но почвы области не будут в динамике за несколько лет меняться, и на динамику этот фактор не влияет. Зато в регрессионной модели за данный год по всем хозяйствам области средняя температура месяца почти одинакова и из регрессионной модели этот фактор исключается. Однако в динамике температура месяца может сильно колебаться, и в прогнозе следовало бы его учитывать.

Последнее, хотя и не очень существенное преимущество прогноза по тренду заключается в том, что для него не требуется большого объема исходной информации о факторах. Достаточно однородного по характеру тенденции периода за 20–25 лет, т.е. всего два десятка уровней, например, урожайности.

Но у прогнозирования по тренду есть, конечно, и свои недостатки. Неявность факторов динамики, скрытых за «номером периода», лишает исследователя возможности учесть ожидаемый или планируемый перелом, скачок в развитии того или иного фактора. Нет возможности проигрывать разные варианты прогноза при разных сочетаниях значений факторов, что обычно делается при прогнозе по регрессионной модели с управляемыми факторами.

Прогноз по тренду несет в себе как бы черты фатализма: будет то-то, изменить ничего нельзя. Ведь мы не можем изменить или отменить ход времени, а аргумент уравнения тренда – это время. Конечно, на самом деле тренд образовался как под влиянием природных факторов, так и деятельности человека. Но слитность этих факторов все равно оставляет впечатление, что человек устранен из процесса, так что психологически данный метод нередко отторгается именно по причине своего фаталистического имиджа. Особенно это чувствовалось в планово-командной экономике, когда считалось, что в будущем будет то и столько, сколько мы запланируем. Прогнозирование в этой системе управления было подавлено «прямым директивным планированием».

Теперь ясно, что прогнозирование – неотъемлемый элемент менеджмента, оно составляет этап в разработке стратегии развития, и плана деятельности предприятия, фирмы, правительства.

10.2. Простая трендовая модель и прогноз по ней.

Простая трендовая модель динамики – это уравнение тренда с указанием начала отсчета единиц времени. Прогноз по этой модели заключается в подстановке в уравнение тренда номера периода, который прогнозируется. Например, тренд цепных темпов роста реальных денежных доходов населения в Оренбургской области, рассчитанный в гл. 5 (таблица 5.5), имеет вид:

$$\hat{y}_i = 105,8 + 3,7 \cdot t_i, t_i = 0 \text{ в } 2002 \text{ г.}$$

Прогноз по этому тренду на 2010 г., номер которого от 2002 года равен 8, составит:

$$\hat{y}_i = 105,8 + 3,7 \cdot 8 = 135,4 \%$$

Интерпретация этого прогноза должна быть следующей: если цепной темп роста реальных денежных доходов населения в Оренбургской области будет возрастать до 2010 г. с той же средней скоростью (среднегодовым приростом), с каким он рос в период с 1998 по 2006 г., то тренд темпов роста в среднем пройдет в 2010 году через точку 135,4 %. Такой прогноз и называется *точечным прогнозом*. Разумеется, точечный прогноз – это скорее абстракция, чем реальность. Если уровни темпов роста и параметры тренда можно было бы определять с бесконечной степенью точности, то и вероятность точного осуществления точечного прогноза темпов роста реальных денежных доходов, составляющего 135,377777777778 ... %, была бы равна нулю. Поскольку мы дали прогноз с одним знаком за запятой, то реально это уже не строго точечный прогноз, а прогноз попадания тренда в интер-

вал от 135,277777777778 до 135,477777777778 %, т.е. в интервал шириной 0,1 %.

От строго математических дефиниций перейдем к более практическим свойствам точечного прогноза. Он означает, что при нормальном законе распределения отклонений от тренда вероятность того, что темп роста реальных денежных доходов населения Оренбургской области окажется ниже точечного прогноза, и что она окажется выше него, равны между собой (каждая равна 0,5). Точечный прогноз в то же время указывает наивероятнейшее из всех возможных значений прогнозируемого показателя. Он, таким образом, является и средней величиной, и медианой, и модой возможных значений прогнозируемого показателя.

При расчете точечного прогноза не обращалось внимания на колеблемость уровней признака. Если бы колеблемость полностью отсутствовала, точечный прогноз явился бы уже не только средним ожидаемым значением, но и единственно возможным значением признака (при соблюдении, естественно, условий реальности прогноза по тренду вообще). Также и автомобиль с отъехавшим от нас товарищем, двигаясь по шоссе пять часов со строго постоянной скоростью 90 км/ч, оказался бы на расстоянии 450 км от точки отъезда. Но ни автомобиль не может пять часов ехать с точно неизменной скоростью, ни, тем более, реальные денежные доходы за четыре года не могут возрасти без малейших колебаний точно на 3,7 п.п. В гл. 7 было показано, что, распространяя уравнение тренда на будущее, мы обязаны считать его лишь выборочной оценкой генеральных параметров, точно нам неизвестных. Наличие случайной колеблемости уровней порождает ошибку репрезентативности выборочных оценок тренда, которую следует принять во внимание при прогнозировании.

Есть, однако, такие процессы, при которых колеблемость несущественна. Таковы, например, процессы распада радиоактивных элементов. Зная точную скорость протекания ядерных реакций, персонал атомных электростанций может рассчитать долю прореагировавшего урана 235 в топливных элементах на любой срок вперед, а, значит, и планировать их замену. Итак, при несущественности колебаний процесса точечный прогноз оказывается самодостаточным и не требует каких-либо дополнений. В экономике, увы, «бесколебательные» процессы не встречаются.

10.3. Прогноз с учетом случайной колеблемости

При таком прогнозе учитывается как вызванная колеблемостью ошибка репрезентативности выборочной оценки тренда, так и колебания уровней в отдельные периоды (моменты) относительно тренда. Следует при этом строго различать три вида прогнозов:

- доверительного интервала для линии тренда;
- доверительного интервала для уровня отдельного периода (момента);
- доверительного интервала среднего уровня за ряд периодов (моментов).

10.3.1. Прогноз доверительного интервала для линии тренда

Напомним (подробнее см. гл. 7), что средняя ошибка свободного члена линейного тренда составляет:

$$m_a = \frac{S(t)}{\sqrt{n}},$$

а средняя ошибка среднегодового прироста:

$$m_b = \frac{S(t)}{\sqrt{\sum_1^n t_i^2}}, \quad (10.1)$$

где n – число уровней базы тренда;

$S(t)$ – среднее квадратическое отклонение уровней ряда от тренда.

Объединяя эти ошибки как независимые по правилу для дисперсий независимых переменных и учитывая, что ошибка среднегодового прироста за t_k лет (или иных отрезков времени) возрастет в t_k раз, получаем формулу средней ошибки прогноза для линии тренда на период с удалением t_k от середины базы прогноза:

$$m_{\hat{y}_k} = \sqrt{m_a^2 + m_b^2 \cdot t_k^2} = \sqrt{\frac{S(t)^2}{n} + \frac{S(t)^2 \cdot t_k^2}{\sum t_i^2}} = S(t) \sqrt{\frac{1}{n} + \frac{t_k^2}{\sum t_i^2}} \quad (10.2)$$

при однократном выравнивании ряда за n периодов, где $\sum t_i^2$ – это сумма квадратов номеров лет t_i от $-\frac{n-1}{2}$ до $+\frac{n-1}{2}$.

По рассмотренному нами тренду цепных темпов роста реальных денежных доходов населения в Оренбургской области точечный прогноз на 2010 г. составил: $\hat{y}_i = 105,8 + 3,7 \cdot 8 = 135,4 \%$, а $S(t) = 8,7 \%$

Средняя ошибка прогноза тренда на 2010 г. составит:

$$m_{\hat{y}_{2010}} = 8,7 \cdot \sqrt{\frac{1}{9} + \frac{13^2}{60}} = 14,89 \%$$

Вероятность того, что фактическая ошибка не превысит одного среднего квадратического отклонения, т.е. $m_{\hat{y}}$, равна при нормальном распределении 0,68. Чтобы получить доверительный интервал прогноза линии тренда с большей надежностью, например, с вероятностью 0,95, среднюю ошибку нужно умножить на величину критерия t -критерия Стьюдента для вероятности 0,95 и семи степеней свободы вариации ($9-2$ параметра линейного тренда). Получаем вероятную ошибку $\alpha_{\hat{y}_{2010}} = t_{Cm} \cdot m_{\hat{y}_{2010}} = 2,3646 \cdot 14,89 = 35,21$.

Итак, с вероятностью 0,95 тренд темпов роста реальных денежных доходов населения в Оренбургской области в 2010 г. пройдет в границах: $135,4 \pm 35,2$ или от 100,2 до 170,6 %.

При линейном тренде и многократном выравнении средняя ошибка прогноза для линии тренда на период с номером t_k от середины базы примет вид:

$$m_{\hat{y}_n} = S(t) \sqrt{\frac{1}{N} + \frac{t_k^2}{\ell \cdot \sum_1^n t_i^2}}, \quad (10.3)$$

где ℓ – число сдвигов базы расчета среднегодового прироста b ;

n – длина каждой базы расчета;

N – общая длина временного ряда.

Например, тренд доходов бюджета Оренбургской области рассчитанный в гл. 5, имеет вид: $\hat{y}_i = 1846,8 + 32,9 \cdot t_i$, где $t_i = 0,5$ в январе 2004 г.

$N = 72$, $\ell = 30$, $n = 43$, $S(t) = 540,52$ млн. руб.

Средняя ошибка прогноза для линии тренда на июль 2007 г. с номером 42,5 составит:

$$m_{\hat{y}_{07.2007}} = 540,52 \sqrt{\frac{1}{72} + \frac{42,5^2}{30 \cdot 31098}} = 67,996 \text{ млн.руб.}$$

Критерий Стьюдента при 70 степенях свободы вариации для вероятности 0,95 составляет 1,9944. Таким образом, с вероятностью 0,95 тренд при сохранении прежней скорости роста доходов бюджета в июле 2007 г. пройдет в интервале: $3245,05 \pm 1,9944 \cdot 67,996$ или от 3109,44 до 3380,66 млн.руб.

При тренде в форме параболы II порядка параметры a и c не являются независимыми друг от друга и их совокупная ошибка определяется сложнее. Независим от них параметр b , ошибка которого аналогична таковой же для линейного тренда. После соединения ошибок всех параметров общая формула средней ошибки прогноза положения параболического тренда на период с номером t_k от середины базы расчета тренда приобретает вид³⁴:

$$m_{\hat{y}_k} = S(t) \sqrt{\frac{t_k^2}{\sum t_i^2} + \frac{\sum t_i^4 - (2 \sum t_i^2) \cdot t_k^2 + n t_k^4}{n \sum t_i^4 - (\sum t_i^2)^2}}. \quad (10.4)$$

При использовании многократного скользящего выравнивания для расчета параметров параболы II порядка знаменатели обеих дробей подкоренного выражения умножаются на число сдвигов базы расчета ℓ , а суммы, стоящие в формулах, исчисляются за одну базу. В числителе же последней дроби будет $N \cdot t_k^4$.

Для экспоненциального тренда рассчитывается ошибка прогноза логарифма линии тренда как для прямой, а затем доверительный интервал логарифма линии тренда. Его границы потенцируются, получаем несимметрично удаленные от точечного прогноза границы самого

³⁴ Четыркин Е.М. Статистические методы прогнозирования. – М.: Статистика, 1977. с. 171

прогноза тренда. Для других форм тренда методика расчета средних ошибок и доверительных границ развита недостаточно, многие вопросы остаются спорными и здесь рассматриваться не будут.

10.3.2. Прогноз доверительного интервала для уровня отдельного периода (момента)

Неопределенность прогноза уровня отдельного периода складывается из двух элементов: ошибки линии тренда для прогнозируемого периода и колебаний уровня около тренда. Первый элемент рассмотрен в предыдущем разделе. Колеблемость отдельных уровней относительно линии тренда измеряется средним квадратическим отклонением $S(t)$. Однако необходимо ответить на вопрос: допустимо ли переносить значение этой величины, полученное за период-базу на прогнозируемый период? Теоретически могут иметь место изменения величины колебаний и в сторону их роста при тенденции роста уровней и постоянстве факторов колеблемости (постоянном коэффициенте колеблемости), и в сторону сокращения абсолютной величины колебаний при сознательном их подавлении: например, колебаний доходов населения при стабилизации экономической ситуации в стране.

Таким образом, в расчет ошибки прогноза и тренда, но особенно ожидаемого отдельного уровня в прогнозируемом периоде, следует взять ожидаемое значение показателя колеблемости для этого же прогнозируемого периода $\hat{S}(t)_k$, расчет которого приведен в разд. 6.4. Читатель, склонный к математическому образу мышления, тут же может заметить, что тренд колеблемости и ее прогноз на будущее – опять же не абсолютная истина, он имеет свою ошибку, которую опять же

нужно учесть, и в прогнозе уровня использовать не $\hat{S}(t)_k$, а доверительный интервал $\hat{S}(t)_k \pm t_{Cm} \cdot m_{\hat{S}(t)_k}$ и т.д. На эти теоретически верные рассуждения следует ответить тем, что ограниченная точность и надежность исходных уровней временного ряда приводит к тому, что дальнейшие итеративные шаги по расчету ошибки ошибок прогноза и т.д. становятся всего лишь математическим упражнением, не улучшающим точность прогноза на практике. В связи с этим при расчете средней ошибки прогноза уровней ряда мы рекомендуем использовать только точечный прогноз силы колебаний $\hat{S}(t)_k$, если тренд колеблемости надежно установлен (см. разд. 7.3.), или величину $S(t)$ по базе прогноза, если изменение этого показателя со временем не установлено надежно.

Кроме этого необходимо быть осторожным с линейными трендами уменьшения колеблемости. Пока это сокращение незначительно и $S(t)$ далеко от нуля, использовать прогноз по линейному тренду можно, но ведь при дальнейшем снижении по прямой показатель колеблемости когда-то станет равным нулю, чего на самом деле быть не может. Колеблемость – такой же незыблемый закон природы, как и вариация признаков в пространстве: она не может быть уничтожена. Для прогнозов показателей колеблемости на более далекую перспективу при тенденции ее сокращения следует применять гиперболическую форму тренда.

Определив величину показателя колеблемости для прогнозируемого периода, подставляем этот показатель вместе с ошибкой прогноза тренда в единую формулу средней ошибки прогноза конкретного отдельного уровня, опираясь на правило сложения независимых дисперсий:

$$m_{y_k} = \sqrt{m_{\hat{y}_k}^2 + S^2(t)_k}. \quad (10.5)$$

Эта формула является общей для любых типов линии тренда. Для каждого типа различны первые слагаемые – ошибки тренда на период t_k . Для линейного тренда при однократном его расчете, используя формулы в (10.1) и (10.3), имеем:

$$m_{y_k} = \sqrt{\frac{S(t)^2}{n} + \frac{S(t)^2 \cdot t_k^2}{\sum t_i^2} + S(t)_k^2}. \quad (10.6)$$

Эта формула должна применяться, если приведен расчет величины $S(t)_k$ на прогнозируемый период, так как в этом случае в первых двух дробях в числитель входит величина $S(t)$ за период-базу, а третье слагаемое подкоренного выражения – это прогнозируемая величина колеблемости на прогнозный период. Таким образом, «вынести за знак корня величину» $S(t)$ нельзя, так как они под корнем различные.

Если же на период прогноза принята та же величина показателя колеблемости, как и за период-базу расчета тренда, то эта величина выносится из-под корня, тогда имеем:

$$m_{y_k} = S(t) \sqrt{1 + \frac{1}{n} + \frac{t_k^2}{\sum_1^n t_i^2}} \quad (10.7)$$

Именно данная формула приводится обычно в учебниках.

Соответственно, при многократном расчете среднегодового прироста \bar{b} , получаем формулу:

$$m_{yk} = S(t) \sqrt{1 + \frac{1}{N} + \frac{t_k^2}{\ell \sum_1^n t_i^2}} \quad (10.8)$$

Значения обозначений те же, что в формулах (10.1) и (10.3).

Для всех других типов тренда средняя ошибка отдельного уровня вычисляется по общей формуле.

По ранее рассмотренным примерам имеем.

1. Прогноз цепных темпов роста реальных денежных доходов населения в Оренбургской области на 2010 г. без учета тренда колеблемости:

$$m_{\hat{y}_{2010}} = 8,7 \cdot \sqrt{1 + \frac{1}{9} + \frac{13^2}{60}} = 17,25 \text{ \%}.$$

Итак, с вероятностью 0,95 тренд темпов роста реальных денежных доходов населения в оренбургской области в 2010 г. пройдет в границах:

С вероятностью 0,95 реальные денежные доходы населения Оренбургской области, в % к предыдущему году в 2010 году должны составить $135,4 \pm 2,3646 \cdot 17,25$ или от 94,59 до 176,16 %.

В этом случае ввиду слабой колеблемости и малого срока прогноза ошибка прогноза отдельного уровня лишь на 15% больше ошибки положения тренда на 2010 г.

2. Прогноз доходов бюджета Оренбургской области на июль 2008 г. при многократном выравнивании с учетом тренда колеблемости, рассчитанного в разд. 6.4. Прогнозное значение среднего квадратического отклонения урожайности отдельных лет от ее тренда на

июль 2007 г. составляет $S(t)_k = \bar{S}(t) + b_{S(t)} \cdot t_k = 471,19 + 13,03 \cdot 42,5 = 1024,97$ млн.руб. Подставляя его в расчет по формуле для многократного выравнивания, имеем:

$$m_{y_{2000}} = \sqrt{\frac{471,19^2}{72} + \frac{471,19^2 \cdot 42,5^2}{30 \cdot 31098} + 1024,97^2} = 1026,68 \text{ млн.руб.}$$

Заметим, что из трех слагаемых подкоренного выражения наибольшим является последнее, т.е. $S(t)_k^2$. Ошибка прогноза уровня доходов бюджета, в основном, вызвана колебаниями, возможными в июле 2007 г., а не ошибкой прогноза тренда, существенно сниженной применением многократного выравнивания и длиной базой.

С вероятностью 0,95 доверительные границы прогноза доходов бюджета Оренбургской области, при условии сохранения до июля 2007 г. прежней скорости их увеличения, составят: $3245,05 \pm 1,9944 \cdot 1026,68$ или от 1197,44 до 5292,66 млн.руб.

10.3.3. Прогноз доверительного интервала для среднего уровня за ряд периодов

Предметом прогнозирования может быть не только уровень отдельного года или месяца, но и средний уровень за ряд периодов (моментов). Рассмотрим, что же в этом случае можно считать *точечным прогнозом*, и какова формула средней ошибки прогноза среднего уровня.

При линейной форме тренда приросты уровней или их сокращения – постоянная величина. В этом случае средняя величина прогнозируемых уровней равна уровню на середину прогнозируемого периода, например, при прогнозе среднегодового уровня на 2004–2008 гг.

за точечный прогноз следует взять прогнозируемый на 2006 г. уровень. Он равен среднему уровню за 2004–2008 гг., так как

$$\frac{\hat{y}_{2004} + (\hat{y}_{2004} + b) + (\hat{y}_{2004} + 2b) + (\hat{y}_{2004} + 3b) + (\hat{y}_{2004} + 4b)}{5} =$$

$$\frac{5\hat{y}_{2004} + 10b}{5} = \hat{y}_{2004} + 2b = \hat{y}_{2006}$$

Если же тренд нелинейный, то среднюю величину прогнозируемых уровней за m лет следует вычислять как простую арифметическую величину после вычисления точечных прогнозов на каждый из m лет. При экспоненциальной форме тренда для расчета среднего уровня за m лет можно приближенно использовать ту же формулу, как для линейного тренда, только, если средний темп роста (снижения) близок к единице, отличается, например, не более, чем на $\pm 0,05$ от 1. В таких случаях геометрическая средняя (уровень середины периода m) мало отличается от арифметической средней.

Средняя ошибка прогноза тренда изменяется даже для линейного тренда неравномерно. Поэтому в общем случае она не равна ошибке прогноза тренда для среднего периода прогнозируемого отрезка времени. Но для линейного тренда различие невелико, так что допустимо пользоваться расчетом ошибки прогноза тренда на середину прогнозируемого периода (первая составляющая), т.е. для $t_{k+\frac{m}{2}}$, для которого при линейном тренде вычисляется и точечный прогноз уровня. Вторая составляющая ошибки прогноза среднего уровня на m периодов времени, т.е. среднее квадратическое отклонение от тренда для m периодов, согласно теории выборочного метода, уменьшается в \sqrt{m} раз. Итак, получаем общую формулу средней ошибки прогноза сред-

него уровня для отрезка m единичных периодов после периода с номером t_k от середины базы расчета тренда:

$$m_{\bar{y}_{k-k+m}} = \sqrt{\frac{S^2(t)}{n} + \frac{S^2(t)t_{k+0,5m}^2}{\sum_l^n t_i^2} + \frac{S^2(t)_{k+0,5m}}{m}} \quad (10.9)$$

Если тренд $S(t)$ и его прогноз не вычислялся или несущественен, то имеем:

$$m_{\bar{y}_{k-k+m}} = S(t) \sqrt{\frac{1}{n} + \frac{t_{k+0,5m}^2}{\sum_l^n t_i^2} + \frac{1}{m}} \quad (10.10)$$

для однократного выравнивания и соответственно измененную формулу – для многократного расчета параметра линейного тренда, которую легко выведут внимательные читатели сами.

Для нелинейных трендов ошибку тренда для прогнозируемого отрезка в m периодов следует вычислять как среднюю арифметическую величину из всех индивидуальных ошибок прогнозов тренда для каждого из m периодов с номерами от t_{k+1} до t_{k+m} от середины базы расчета тренда. Затем к квадрату ошибки тренда добавляется $\frac{S^2(t)}{m}$ и вычисляется квадратный корень из суммы подкоренных дробей.

Например, по данным за 1900–1991 гг. тренд национального богатства России имел вид: $\hat{y}_i = 1457,7 \cdot 1,0565^{t_i}$ (см. п. 5.4.1), $t=0$ в 1946 г.

Рассчитаем по этому тренду прогноз национального богатства на 1992–1994 гг. и его доверительные границы без учета тренда ко-

леблемости при $S(t) = 1384,8$ млрд.руб. За точечный берем прогноз тренда на 1993 г.:

$$t = 47, \hat{y} = 1457,7 \cdot 1,0565^{47,5} = 19836,7 \text{ млрд.руб.};$$

$$m_{\bar{y}_{k-k+m}} = 1384,8 \sqrt{\frac{1}{92} + \frac{47,5^2}{64883} + \frac{1}{3}} = 852,5 \text{ млрд.руб.}$$

С вероятностью 0,95 доверительный интервал прогноза национального богатства России при условии сохранения до 1994 года прежнего тренда, составит: $19836,7 + 1,9867 \cdot 852,5$ или от 18143,1 до 21530,3 млрд.руб. Широкий интервал вызван значительной силой колеблемости и требованием высокой надежности прогноза.

10.3.4. Расчет страхового запаса для преодоления риска отклонения от тренда

Доверительные интервалы прогноза могут служить основанием для расчета рисков больших отклонений от тренда и страхового запаса для преодоления этого риска. В предыдущем разделе вычислены границы доверительного интервала прогноза национального богатства России на 1992-1994 гг.: среднегодовое национальное богатство ожидалось с вероятностью 0,95 в границах от 18143,1 до 21530,3 млрд.руб.

Это означает, что при близком к нормальному закону распределении отклонений от тренда вероятность того, что национальное богатство окажется в среднем за год ниже 18143,1 млрд. руб., равна: $(1 - 0,95) : 2 = 0,025$ или риск 2,5%.

Однако, если для поддержания национальной экономики в нормальном состоянии необходимо, иметь ежегодно размер национального богатства не ниже 21000 млрд. руб., то вероятность риска будет

больше. Для ее вычисления следует вычислить кратность отклонения 21000 млрд. руб. от точечного прогноза тренда 19836,7 млрд. руб. в единицах $S(t)$, т.е. $1350,0 : 1384,8 \approx 0,98$. Вероятность того, что отклонение от тренда превысит $0,98 S(t)$ есть $(1 - F(0,98)) : 2 = (1 - 0,6729) : 2 = 0,164$ или 16,4%. Такой высокий риск неприемлем, и для его перекрытия и доведения до приемлемого уровня в 2,5% необходимо иметь страховой запас $21000 - 18143,1 = 2856,9$ млрд. руб. на год. По верхней границе доверительного интервала можно рассчитать вероятность иного, не столь драматического риска: риск перепроизводства, снижение цены на рынке.

10.4. Прогнозирование комплекса жестко-взаимосвязанных признаков.

В данном разделе на основе доказанных в гл. 9 свойств трендов и колебаний в системе жестко-взаимосвязанных признаков рассматриваются проблемы прогнозирования такой системы, как например, площадь посева, урожайность и валовой сбор сельскохозяйственных культур, или численность работников предприятия, производительность их труда, выпуск продукции.

В разд. 9.1. было доказано, что при наличии колеблемости признаков-сомножителей тренд признака-произведения содержит дополнительные случайные элементы, зависящие от сочетания разнораспределенных по времени колебаний сомножителей и корреляции между ними. Из этого положения вытекает, что предпочтительнее прогнозировать уровни признаков-сомножителей, а их произведение даст прогноз признака-произведения. Такой прогноз будет меньше зависеть от случайного распределения отклонений уровней в отдельные

годы, чем прогноз по тренду признака-произведения.

Конечно, не следует забывать, что мы имеем дело со статистической закономерностью: не в каждой отдельной задаче прогнозирования, а лишь как математическое ожидание или средняя величина ошибок прогнозов по множеству задач проявится теоретическое преимущество прогнозов по трендам сомножителей. В отдельном случае ошибка прогноза по тренду произведения может быть и меньше, ошибка может быть случайно вообще равна нулю. Но решать проблему выбора методики необходимо, исходя из доказанного преимущества прогноза по произведению трендов сомножителей.

По данным табл. 9.1 вычислим прогноз валового сбора зерна на период времени с номером 6 от середины базы расчета трендов. Алгоритм расчета точечного прогноза валового сбора может иметь два варианта:

а) прогнозируем площадь, прогнозируем урожайность по их трендам, после чего перемножаем прогнозы:

$$\hat{n}_6 = 120 + 5 \cdot 6 = 150 \text{ га}; \quad \hat{y}_6 = 29 + 6 = 35 \text{ ц/га};$$

$$\hat{B}_6 = \hat{P}_6 \cdot \hat{y}_6 = 150 \cdot 35 = 5250 \text{ ц};$$

б) перемножаем тренды площади и урожайности и на основе полученной параболы вычисляем прогноз валового сбора:

$$\hat{B}_t = [(120 + 5t) \cdot (29 + t)] = 3480 + 265t + 5t^2;$$

$$\hat{B}_6 = 3480 + 265 \cdot 6 + 5 \cdot 6^2 = 5250 \text{ ц}.$$

Средняя ошибка выборочной оценки или прогноза произведения двух переменных вычисляется по формулам:

$$а) m_{\Pi_k \cdot y_k} = \sqrt{(m_{\Pi_k} \cdot \hat{y}_k)^2 + (m_{y_k} \cdot \hat{\Pi}_k)^2}, \quad (10.11)$$

если признаки Π и y независимы друг от друга;

$$б) m_{\Pi_k \cdot y_k} = \sqrt{(m_{\Pi_k} \cdot \hat{y}_k)^2 + (m_{y_k} \cdot \hat{\Pi}_k)^2 + 2(m_{\Pi_k} \cdot \hat{y}_k)(m_{y_k} \cdot \hat{\Pi}_k) \cdot r'_{\Pi y}}, \quad (10.12)$$

если признаки Π и y в динамике коррелированы.

Здесь m_{Π_k} и m_{y_k} – средняя ошибка прогнозов признаков-сомножителей на период t_k ; $r'_{\Pi y}$ – коэффициент корреляции между признаками-сомножителями в динамике, вычислении которого описано в разд. 9.4., т.е. коэффициент корреляции отклонений от трендов, ранее уже рассчитанный выше: $r'_{\Pi y} = -0,326$. Средние ошибки прогнозов отдельного признака рассмотрены в разд. 10.3.2. Имеем:

$$m_{\Pi_6} = S(t)_{\Pi} \cdot \sqrt{1 + \frac{1}{9} + \frac{6^2}{60}} = 7,4 \sqrt{1 + \frac{1}{9} + \frac{6^2}{60}} = 13,16 \text{ га};$$

$$m_{y_6} = S(t)_y \cdot \sqrt{1 + \frac{1}{9} + \frac{6^2}{60}} = 4,4 \sqrt{1 + \frac{1}{9} + \frac{6^2}{60}} = 5,76 \text{ ц/га}.$$

Подставляя в формулу средней ошибки прогноза валового сбора с учетом корреляции колебаний площади и урожайностью, имеем:

$$\begin{aligned} m_{B_6} &= \sqrt{(13,16 \cdot 35)^2 + (5,76 \cdot 150)^2 - 2(13,16 \cdot 35)(5,76 \cdot 150) \cdot 0,326} = \\ &= \sqrt{699179,5} = 836,2 \end{aligned}$$

С вероятностью 0,9 (t -критерий Стьюдента при семи степенях свободы равен 1,95) доверительный интервал валового сбора составит:

$$5250 \pm 1,95 \cdot 836,2 \text{ или от } 3619 \text{ до } 6881 \text{ ц.}$$

Широкий интервал прогноза, во-первых, связан с короткой базой расчета трендов, во-вторых, с довольно значительной колеблемо-

стью урожайности.

10.5 Прогнозирование по смешанной трендово-факторной модели

Общий вид трендово-факторной прогностической модели:

$$\tilde{b}_{y_k} = f(b_{x_1, k-\ell}; b_{x_2, k-\ell}; \dots; b_{x_i, k-\ell}; \dots; b_{x_n, k-\ell}), \quad (10.13)$$

где b_y – средневременной прирост результативного показателя;

$b_{x_1}, b_{x_2}, b_{x_i}, b_{x_n}$ – приросты факторов;

k – номер прогнозируемого периода времени;

ℓ – период упреждения.

Трендово-факторные модели позволяют выявить силу влияния приростов основных факторов и величину прибавки результативного показателя. Смешанное прогнозирование устраняет часть недостатков факторного и авторегрессионного. Оно позволяет включить скачкообразное изменение управляемых факторов.

Рассмотрим трендово-факторную модель прироста урожайности зерновых культур в хозяйствах Оренбургской области:

$$\tilde{b} = 0,0115b_{x_2} + 1,028b_{x_3} + 0,4544b_{x_4} - 0,0794,$$

где b_{x_2} – среднегодовой абсолютный прирост материально-денежных затрат без заработной платы и затрат на удобрения;

b_{x_3} – среднегодовой абсолютный прирост энергообеспеченности;

b_{x_4} – среднегодовой абсолютный прирост доз удобрений;

Подставляем ожидаемые значения приростов факторов в мо-

дель:

$$\tilde{b}_{y_{2000}} = 0,0115 \cdot 4,0 + 1,028 \cdot 0,055 + 0,4544 \cdot 0,50 - 0,0794 = 0,26 \text{ ц/га.}$$

Среднее квадратическое отклонение среднегодового абсолютно-го прироста урожайности зерновых культур в хозяйствах Оренбургской области S_b равно 0,132 ц/га, а средняя ошибка многофакторной трендово-факторной модели – 0,086 ц/га, тогда средняя ошибка прогноза:

$$m_{y_{2000}} = \sqrt{m^2_{\tilde{b}_{yk}} + S^2_{\tilde{b}_{y(t)k}}} = 0,15 \text{ ц/га.}$$

Прогнозируемый среднегодовой абсолютный прирост урожайности зерновых культур в Оренбургской области с вероятностью 0,95 заключен в границах:

$$\tilde{b}_{y_{2000}} \pm 1,96 \cdot 0,15 = 0,26 \pm 0,29 \text{ ц/га.}$$

Из-за большой колеблемости ($V_{b(t)} = 62,8\%$) среднегодового абсолютно-го прироста вероятная ошибка его прогноза превышает величину самого прироста.

Данную методику можно использовать при прогнозировании других результативных показателей, например, рентабельности, производительности труда, показателей устойчивости производства и финансов.

Глава 11. ИСПОЛЬЗОВАНИЕ АДАПТИВНЫХ МЕТОДОВ ПРОГНОЗИРОВАНИЯ В ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЯХ

11.1 Сущность адаптивных методов

В настоящее время одним из наиболее перспективных направлений исследования и прогнозирования одномерных временных рядов считаются адаптивные методы.

Адаптивными называются методы прогнозирования, позволяющие строить самокорректирующиеся (самонастраивающиеся) экономико-математические модели, которые способны оперативно реагировать на изменение условий путем учета результата прогноза, сделанного на предыдущем шаге, и учета различной информационной ценности уровней ряда.

При обработке временных рядов, как правило, наиболее ценной бывает информация последнего периода, т.к. необходимо знать, как будет развиваться тенденция, существующая в данный момент, а не тенденция, сложившаяся в среднем не всем рассматриваемом периоде. Адаптивные методы позволяют учесть различную информационную ценность уровней временного ряда, степень «устаревания» данных.

Прогнозирование методом экстраполяции на основе кривых роста в какой-то мере тоже содержат элемент адаптации, поскольку с получением «свежих» фактических данных параметры кривых пересчитываются заново. Поступление новых данных может привести и к замене выбранной ранее кривой на другую модель. Однако степень адаптации в данном случае весьма незначительна, кроме того, она па-

дает с ростом длины временного ряда, т.к. при этом уменьшается «весомость» каждой новой точки. В адаптивных методах различную ценность уровней в зависимости от их «возраста» можно учесть с помощью системы весов, придаваемых этим уровням.

Важнейшее достоинство адаптивных методов – построение самокорректирующихся моделей, способных учитывать результат прогноза, сделанного на предыдущем шаге.

Пусть модель находится в некотором состоянии, для которого определены текущие значения ее коэффициентов. На основе этой модели делается прогноз. При поступлении фактического значения оценивается ошибка прогнозного значения (разница между этим значением и полученным по модели). Ошибка прогнозирования через обратную связь поступает в модель и учитывается в ней в соответствии с принятой процедурой перехода от одного состояния в другое. В результате вырабатываются «компенсирующие» изменения, состоящие в корректировании параметров в целях большего согласования поведения модели с динамикой ряда. Затем рассчитывается прогнозная оценка на следующий момент времени, и весь процесс повторяется вновь.

Таким образом, адаптация осуществляется итеративно с получением каждой новой фактической точки ряда. Модель постоянно «впитывает» новую информацию, приспосабливается к ней и поэтому отражает тенденцию развития, существующую в данный момент.

Быстроту реакции модели на изменения в динамике процесса характеризует *параметр адаптации*. Он должен быть выбран таким образом, чтобы обеспечивалось адекватное отображение тенденции при одновременной фильтрации случайных отклонений. Значение параметра адаптации может быть определено:

- на основе эмпирических данных,
- выведено аналитическим способом,
- получено на основе метода проб.

В качестве критерия оптимальности при выборе параметра адаптации обычно принимают критерий минимума среднего квадрата ошибок прогнозирования³⁵.

11.2 Экспоненциальное сглаживание

Для экспоненциального сглаживания ряда используется рекуррентная формула:

$$S_t = \alpha \cdot y_t + \beta \cdot S_{t-1}, \quad (11.1)$$

где S_t - значение экспоненциальной средней в момент t ;

α - параметр сглаживания, $\alpha = const$, $0 < \alpha < 1$;

$\beta = 1 - \alpha$.

Если последовательно использовать соотношение (11.1), то экспоненциальную среднюю S_t можно выразить через предшествующие значения уровней временного ряда:

$$\begin{aligned} S_t &= \alpha \cdot y_t + \beta \cdot S_{t-1} = \alpha \cdot y_t + \beta \cdot (\alpha \cdot y_{t-1} + \beta \cdot S_{t-2}) = \alpha \cdot y_t + \alpha \cdot \beta \cdot y_{t-1} + \\ &+ \beta^2 S_{t-2} = \dots = \alpha \cdot y_t + \alpha \cdot \beta \cdot y_{t-1} + \alpha \cdot \beta^2 \cdot y_{t-2} + \dots + \alpha \cdot \beta^i \cdot y_{t-i} + \dots + \beta^n \cdot S_0 \end{aligned}$$

Таким образом,

³⁵ Дуброва, Т.А. Статистические методы прогнозирования: учеб. пособие для вузов / Т.А. Дуброва - М.: ЮНИТИ-ДАНА, 2003. -206 с.

$$S_t = \alpha \sum_{i=0}^{n-1} \beta^i \cdot y_{t-i} + \beta^n S_0, \quad (11.2)$$

где n – длина ряда.

При $n \rightarrow \infty$ $\beta^n \rightarrow 0$, следовательно,

$$S_t = \alpha \sum_{i=0}^{\infty} \beta^i \cdot y_{t-i} \quad (11.3)$$

Таким образом, величина S_t оказывается взвешенной суммой всех членов ряда. Причем веса отдельных уровней ряда убывают по мере их удаления в прошлое соответственно экспоненциальной функции (в зависимости от «возраста» наблюдений). Именно поэтому величина S_t называется экспоненциальной средней.

Автор модели английский математик Р. Браун показал, что математическое ожидание временного ряда и экспоненциальной средней совпадут, но в то же время дисперсия экспоненциальной средней $D[S_t]$ меньше дисперсии временного ряда σ^2 :

$$D[S_t] = \frac{\alpha}{2-\alpha} \sigma^2. \quad (11.4)$$

Так как $0 < \alpha < 1$, $D[S_t]$ меньше дисперсии временного ряда, равной σ^2 .

При высоком значении α дисперсия экспоненциальной средней незначительно отличается от дисперсии ряда. С уменьшением α

дисперсия экспоненциальной средней сокращается, возрастает ее отличие от дисперсии ряда. Тем самым экспоненциальная средняя начинает играть роль «фильтра», поглощающего колебания временного ряда.

Таким образом, с одной стороны, следует увеличивать вес более свежих наблюдений, что может быть достигнуто повышением α , с другой стороны, для сглаживания случайных отклонений значение α нужно уменьшить. Эти два требования находятся в противоречии. Поиск компромиссного значения параметра сглаживания α составляет задачу оптимизации модели. При краткосрочном прогнозировании желательно как можно быстрее отразить изменения ряда и в то же время очистить ряд, отфильтровав случайные колебания. Для этого величине α следует присваивать одно из промежуточных значений от 0 до 1. При этом для оперативных, конъюнктурных прогнозов в большей степени должна учитываться свежая информация, поэтому α следует брать большим. При увеличении срока прогнозирования α следует уменьшить. В некоторых работах приводится формула для расчета α :

$$\alpha = \frac{2}{n+1}, \quad (11.5)$$

где n – длина ряда.

При расчете экспоненциальной средней в момент времени t всегда требуется значение экспоненциальной средней в предыдущий момент времени. Часто на практике в качестве начального значения S_0 используется среднее арифметическое значение из всех имеющихся

уровней временного ряда или из какой-то их части.

При использовании экспоненциальной средней для краткосрочного прогнозирования предполагается, что модель ряда имеет вид:

$$y_t = a_{1,t} + \varepsilon_t, \quad (11.6)$$

где $a_{1,t}$ - варьирующий во времени средний уровень временного ряда;

ε_t - случайные неавтокоррелированные отклонения с нулевым мат. ожиданием и дисперсией σ^2 .

Прогнозная модель определяется равенством:

$$\hat{y}_\tau(t) = \hat{a}_{1,t}, \quad (11.7)$$

где $\hat{y}_\tau(t)$ - прогноз, сделанный в момент t на τ единиц времени вперед;

$\hat{a}_{1,t}$ - оценка $a_{1,t}$.

Единственный параметр модели $\hat{a}_{1,t}$ определяется экспоненциальной средней:

$$\hat{a}_{1,t} = S_t;$$

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

Выражение (11.1) можно представить по-другому, перегруппировав члены:

$$S_t = S_{t-1} + \alpha(y_{t-1} - S_{t-1}). \quad (11.8)$$

Величину $(y_{t-1} - S_{t-1})$ можно рассматривать как погрешность прогноза. Тогда новый прогноз S_t получается в результате корректировки предыдущего прогноза с учетом его ошибки. В этом и состоит адаптация модели³⁶.

Рассмотрим пример. Для данных о доходах консолидированного бюджета Оренбургской области (таблица 11.1) рассчитаем экспоненциальную среднюю. В качестве начального значения экспоненциальной средней возьмем среднее значение из пяти первых уровней ряда параметр адаптации примем $\alpha = 0,5$.

$$\text{Определим } S_0 = \frac{1}{5} \sum_{t=1}^5 y_t = 948,12$$

Найдем значение экспоненциальной средней при $\alpha = 0,5$.

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

$$S_1 = 0,5 \cdot 1119,3 + 0,5 \cdot 948,12 = 1033,7$$

$$S_2 = 0,5 \cdot 352,2 + 0,5 \cdot 1033,71 = 693,0$$

...

$$S_{75} = 0,5 \cdot 3478,7 + 0,5 \cdot 3366,8 = 3422,8$$

В таблице 11.1 и на рисунке 11.1 представлены результаты расчетов и динамика исходного временного ряда.

Как видно на графике расчетные значения близки к исходному временному ряду, ошибка аппроксимации составила 14,57%. Это свидетельствует о хорошей точности модели.

³⁶ Дуброва, Т.А. Статистические методы прогнозирования: учеб. пособие для вузов / Т.А. Дуброва - М.: ЮНИТИ-ДАНА, 2003. -206 с.

Таблица 11.1

**Расчетные значения по экспоненциальной средней доходов консолидированного бюджета
Оренбургской области, млн.руб.**

Период	2001 г.	2002 г.	2003 г.	2004 г.	2005 г.	2006 г.
январь	1033,7	1001,1	1208,9	1454,2	1717,6	2526,9
февраль	693,0	999,8	1054,5	1272,7	1517,5	2024,2
март	849,9	1072,4	1228,2	1451,0	2205,3	2619,7
апрель	1013,9	1329,0	1563,5	2035,6	2219,8	2746,1
май	1049,1	1315,0	1551,2	1819,7	2306,8	3269,2
июнь	970,3	1147,7	1391,9	1756,2	2070,5	2995,5
июль	949,2	1275,6	1521,0	2011,9	2137,9	3046,3
август	1063,8	1365,6	1504,0	1908,2	2594,8	3637,8
сентябрь	1026,6	1264,6	1434,1	1845,5	2315,3	2878,7
октябрь	979,4	1398,3	1704,4	1883,3	2278,3	3317,6
ноябрь	1029,9	1349,1	1627,7	2342,8	3261,8	3366,8
декабрь	1136,7	1449,1	1711,7	2491,2	3480,7	3422,8

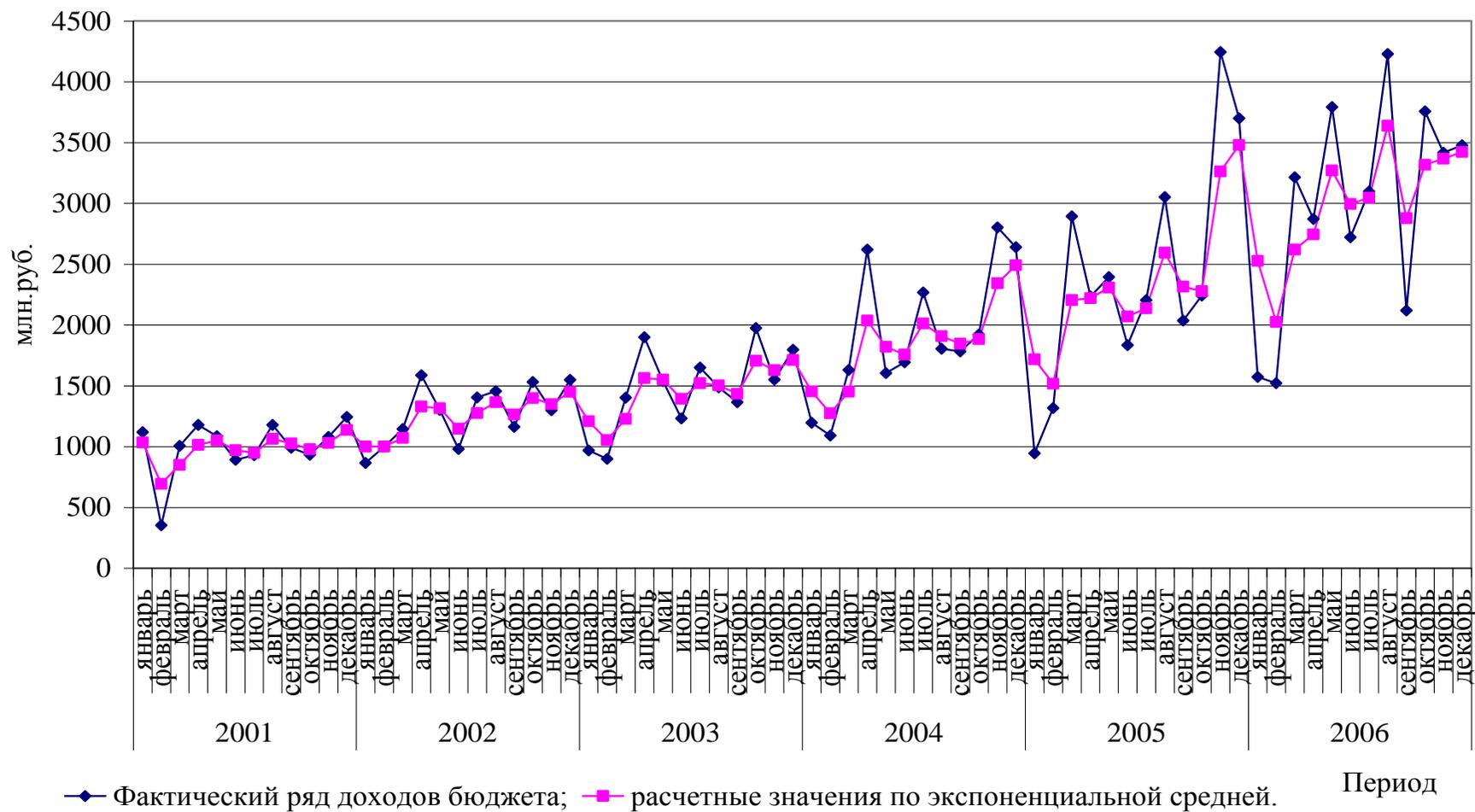


Рис. 11.1 Экспоненциальное сглаживание временного ряда доходов бюджета

Спрогнозируем доходы бюджета на январь 2007 года по формуле:

$$S_t = S_{t-1} + \alpha \cdot (y_{t-1} - S_{t-1}) = 3422,8 + 0,5 \cdot (3478,7 - 3422,8) = 3450,7 \text{ млн.р.}$$

11.3 Адаптивные полиномиальные модели

Если для прогнозирования временного ряда, имеющего ярко выраженную линейную тенденцию, использовать равенство (11.7), опирающееся на модель экспоненциального сглаживания, то модель, как правило, будет давать смещенные прогнозы, т.е. систематическую ошибку. Для таких временных рядов целесообразно использовать модели линейного роста, также применяющие процедуру экспоненциального сглаживания.

В этих моделях прогноз может быть получен с помощью следующего выражения:

$$\hat{y}_\tau(t) = \hat{a}_{1,t} + \hat{a}_{2,t} \tau \quad (11.9)$$

где $\hat{a}_{1,t}$ и $\hat{a}_{2,t}$ — текущие оценки коэффициентов;

τ — время упреждения прогноза.

В таблице 11.2 представлены три модели данного типа: двухпараметрическая модель Ч. Хольта, однопараметрическая модель Р. Брауна и трехпараметрическая модель Дж. Бокса и Г. Дженкинса, отличающиеся рекуррентными выражениями для пересчета текущих оценок коэффициентов (параметры адаптации или параметры экспоненциального сглаживания $0 < \alpha_1, \alpha_2, \alpha_3, \beta < 1$).

Таблица 11.2

Модели линейного роста для расчетов прогноза $\hat{y}_t(t) = \hat{a}_{1,t} + \hat{a}_{2,t}\tau$

Название модели	Оценка коэффициентов
Модель Ч. Хольта	$\hat{a}_{1,t} = \alpha_1 y_t + (1 - \alpha_1)(\hat{a}_{1,t-1} + \hat{a}_{2,t-1})$ $\hat{a}_{2,t} = \alpha_2 (\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \alpha_2)\hat{a}_{2,t-1}$
Модель Р. Брауна	$\hat{a}_{1,t} = \hat{a}_{1,t-1} + \hat{a}_{2,t-1} + (1 - \beta^2)e_t$ $\hat{a}_{2,t} = \hat{a}_{2,t-1} + (1 - \beta^2)e_t$
Модель Дж. Бокса и Г. Дженкинса	$\hat{a}_{1,t} = \alpha_1 y_t + (1 - \alpha_1)(\hat{a}_{1,t-1} + \hat{a}_{2,t-1}) + \alpha_3(e_t - e_{t-1})$ $\hat{a}_{2,t} = \alpha_2 (\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \alpha_2)\hat{a}_{2,t-1}$

В ППП чаще представлена модель Ч. Хольта с возможностью выбора оптимальных параметров по критерию минимума среднеквадратической ошибки путем перебора на сетке возможных значений. Рекуррентные формулы для оценки коэффициентов по этой модели могут быть преобразованы к следующему виду, явно показывающему зависимость «корректирующего воздействия» от величины ошибки:

$$\begin{aligned} \hat{a}_{1,t} &= \hat{a}_{1,t-1} + \hat{a}_{2,t-1} + \alpha_1 e_t, \\ \hat{a}_{2,t} &= \hat{a}_{2,t-1} + \alpha_1 \alpha_2 e_t, \end{aligned} \quad (11.10)$$

где $e_t = y_t - \hat{y}_1(t-1)$ - ошибка прогноза

Из последних выражений видно, что частным случаем модели Ч. Хольта можно считать модель Р. Брауна, представленную в таблице 11.2. При этом единственный параметр β играет роль коэффициента дисконтирования наблюдений.

Трехпараметрическая модель Дж. Бокса и Г. Дженкинса пред-

ставляет собой развитие модели Ч. Хольта, причем усовершенствование осуществляется путем включения в выражение для оценивания $\hat{a}_{1,t}$ разности ошибок.

В ряде работ проводилось теоретическое и экспериментальное сравнение различных моделей. В результате при обработке экономических временных рядов значения коэффициента α оказывались близкими к нулю, и авторы пришли к выводу о нецелесообразности включения в модель разности ошибок.

Понятие экспоненциальной средней можно обобщить в случае экспоненциальных средних более высоких порядков.

Выравнивание p -го порядка

$$S_t^{(p)} = \alpha S_t^{(p-1)} + \beta S_{t-1}^{(p)} \quad (11.11)$$

является простым экспоненциальным сглаживанием, примененным к результатам сглаживания $(p - 1)$ -го порядка.

Если предполагается, что тренд некоторого процесса может быть описан полиномом степени n , то коэффициенты предсказывающего полинома могут быть вычислены через экспоненциальные средние соответствующих порядков.

В случае, когда исследуемый процесс, состоящий из детерминированной и случайной компоненты, описывается полиномом n -го порядка, прогноз на τ шагов вперед осуществляется по формуле:

$$\hat{y}_\tau(t) = \hat{a}_1 + \hat{a}_2 \tau + \frac{1}{2} \hat{a}_3 \tau^2 + \dots + \frac{1}{n!} \hat{a}_{n+1} \tau^n \quad (11.12)$$

где $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_{n+1}$ - оценки параметров

На практике обычно используются полиномы не выше второго порядка.

В таблице 11.3 приведены формулы, необходимые для расчета по моделям нулевого, первого и второго порядков.

Процедура прогнозирования временных рядов на основе адаптивных полиномиальных моделей состоит из следующих этапов³⁷:

1 Выбирается вид модели экспоненциального сглаживания задается значение параметра сглаживания α . При выборе порядка адаптивной полиномиальной модели могут использоваться различные подходы, например графический анализ, применение метода последовательных разностей и др.

2 Определяются начальные условия. Например, для полиномиальной модели первого порядка необходимо определить $\hat{a}_{1,0}, \hat{a}_{2,0}$.

Чаще всего в качестве этих оценок берут коэффициенты соответствующих полиномов, полученные с помощью МНК. Начальные условия для модели нулевого порядка обычно получают усреднением нескольких первых уровней ряда. Зная эти оценки, с помощью указанных в таблице 11.3 формул находят начальные значения экспоненциальных средних (гр. 2).

3 Проводится расчет значений соответствующих экспоненциальных средних (гр. 3).

³⁷ Дуброва, Т.А. Статистические методы прогнозирования: учеб. пособие для вузов / Т.А. Дуброва - М.: ЮНИТИ-ДАНА, 2003. -206 с.

Таблица 11.3

Основные формулы для прогнозирования по адаптивным полиномиальным моделям

Сте- пень модели n	Начальные условия	Экспоненциаль- ные средние	Оценка коэффициентов	Модель прогноза
1	2	3	4	5
0	$S_0^{(1)} = \hat{a}_{1,0}$	$S_t^{(1)} = \alpha y_t + \beta S_{t-1}^{(1)}$	$\hat{a}_{1,t} = S_t^{(1)}$	$\hat{y}_\tau(t) = \hat{a}_{1,t}$
1	$S_0^{(1)} = \hat{a}_{1,0} - \frac{\beta}{\alpha} \hat{a}_{2,0}$ $S_0^{(2)} = \hat{a}_{1,0} - \frac{2\beta}{\alpha} \hat{a}_{2,0}$	$S_t^{(1)} = \alpha y_t + \beta S_{t-1}^{(1)}$ $S_t^{(2)} = \alpha S_t^{(1)} + \beta S_{t-1}^{(2)}$	$\hat{a}_{1,t} = 2S_t^{(1)} - S_t^{(2)}$ $\hat{a}_{2,t} = \frac{\alpha}{\beta} [S_t^{(1)} - S_t^{(2)}]$	$\hat{y}_\tau(t) = \hat{a}_{1,t} + \tau \hat{a}_{2,t}$
2	$S_0^{(1)} = \hat{a}_{1,0} - \frac{\beta}{\alpha} \hat{a}_{2,0} + \frac{\beta \cdot (2-\alpha)}{2\alpha^2} \hat{a}_{3,0}$ $S_0^{(2)} = \hat{a}_{1,0} - \frac{2\beta}{\alpha} \hat{a}_{2,0} + \frac{\beta(3-2\alpha)}{\alpha^2} \hat{a}_{3,0}$ $S_0^{(3)} = \hat{a}_{1,0} - \frac{3\beta}{\alpha} \hat{a}_{2,0} + \frac{3\beta(4-3\alpha)}{2\alpha^2} \hat{a}_{3,0}$	$S_t^{(1)} = \alpha y_t + \beta S_{t-1}^{(1)}$ $S_t^{(2)} = \alpha S_t^{(1)} + \beta S_{t-1}^{(2)}$ $S_t^{(3)} = \alpha S_t^{(2)} + \beta S_{t-1}^{(3)}$	$\hat{a}_{1,t} = 3(S_t^{(1)} - S_t^{(2)}) + S_t^{(3)}$ $\hat{a}_{2,t} = \frac{\alpha}{2\beta^2} [(6-5\alpha)S_t^{(1)} - 2(5-4\alpha)S_t^{(2)} + (4-3\alpha)S_t^{(3)}]$ $\hat{a}_{3,t} = \frac{\alpha^2}{\beta^2} (S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)})$	$\hat{y}_\tau(t) = \hat{a}_{1,t} + \tau \hat{a}_{2,t} + \frac{1}{2} \tau^2 \hat{a}_{3,t}$

4 Находятся оценки коэффициентов модели (гр. 4).

5 Осуществляется прогноз на одну точку вперед (гр. 5 при $\tau = 1$); находится отклонение фактического значения временного ряда от прогнозируемого. Шаги с 3-го по 5-й данной процедуры повторяются для всех $t \leq n$, где n — длина ряда.

Окончательная прогнозная модель формируется на последнем шаге в момент $t = n$. Прогноз получается на базе выражения (11.12) путем подстановки в него последних значений коэффициентов и времени упреждения τ .

К положительным особенностям рассмотренных моделей следует отнести то, что при поступлении новой, «свежей» информации расчеты повторять не придется. Достаточно принять в качестве начальных условий последние значения функций сглаживания $S_t^{(i)}$ и продолжить вычисления³⁸.

Рассмотрим пример. По данным таблицы 5.1 рассчитаем прогноз по адаптивным полиномиальным моделям.

Увеличим значение параметра адаптации с целью придания большего веса последним наблюдениям: $\alpha = 0,7$. Применим аналитическое выравнивание в виде полиномиального тренда 1-го порядка. Уравнение тренда примет вид: $\hat{y}_t = 586,26 + 34,535 \cdot t$. Тогда

$$\hat{a}_1 = 586,26; \hat{a}_2 = 34,535.$$

Находим начальные значения экспоненциальных средних:

$$S_0^1 = \hat{a}_{1,0} - \frac{\beta}{\alpha} \cdot \hat{a}_{2,0} = 586,26 - \frac{1-0,7}{0,7} \cdot 34,535 = 571,46$$

$$S_0^2 = \hat{a}_{1,0} - \frac{2\beta}{\alpha} \cdot \hat{a}_{2,0} = 586,26 - \frac{2 \cdot (1-0,7)}{0,7} \cdot 34,535 = 556,66$$

³⁸ Дуброва, Т.А. Статистические методы прогнозирования: учеб. пособие для вузов / Т.А. Дуброва - М.: ЮНИТИ-ДАНА, 2003. -206 с.

Экспоненциальные средние составят:

$$S_t^1 = \alpha y_t + \beta S_{t-1}^1,$$

$$\text{где } y_t = \hat{y}_{t=n} = 586,26 + 34,535 \cdot 72 = 3072,8$$

$$S_t^1 = 0,7 \cdot 3072,8 + (1 - 0,7) \cdot 571,46 = 1979,51$$

$$S_t^2 = \alpha S_t^1 + \beta S_{t-1}^2 = 0,7 \cdot 1979,51 + (1 - 0,7) \cdot 556,66 = 1218,66$$

Находим оценки коэффициентов модели (скорректированные параметры линейного тренда):

$$\hat{a}_{1,t} = 2S_t^1 - S_t^2 = 2 \cdot 1979,51 - 1218,66 = 2740,36$$

$$\hat{a}_{2,t} = \frac{\alpha}{\beta} [S_t^1 - S_t^2] = \frac{0,7}{1 - 0,7} \cdot [1979,51 - 1218,66] = 1775,31$$

Осуществляем прогноз на одну точку вперед:

$$\tilde{y}_\tau(t) = \hat{a}_{1,t} + \tau \hat{a}_{2,t} = 2740,36 + 1 \cdot 1775,31 = 4515,675 \text{ млн.р.}$$

Экспоненциальное сглаживание в ППП STATISTICA

Выберем значение параметра адаптации α используя перебор по сетке значений. Стандартным образом запустите модуль «Временные ряды и прогнозирование» (рисунок 11.2). В появившемся окне выбираем «Экспоненциальное сглаживание и прогноз» и задаем переменную в окне «Переменные» (рисунок 11.3).

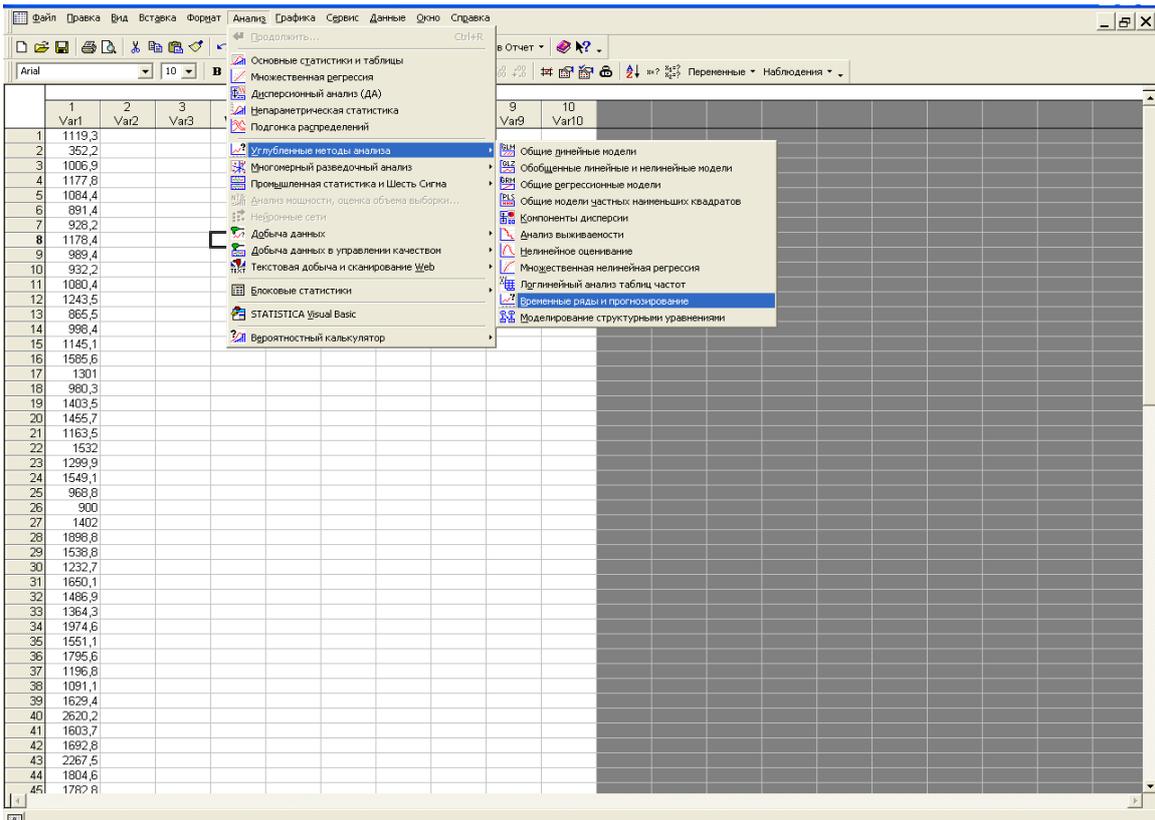


Рис. 11.2 Запуск модуля «Временные ряды и прогнозирование»

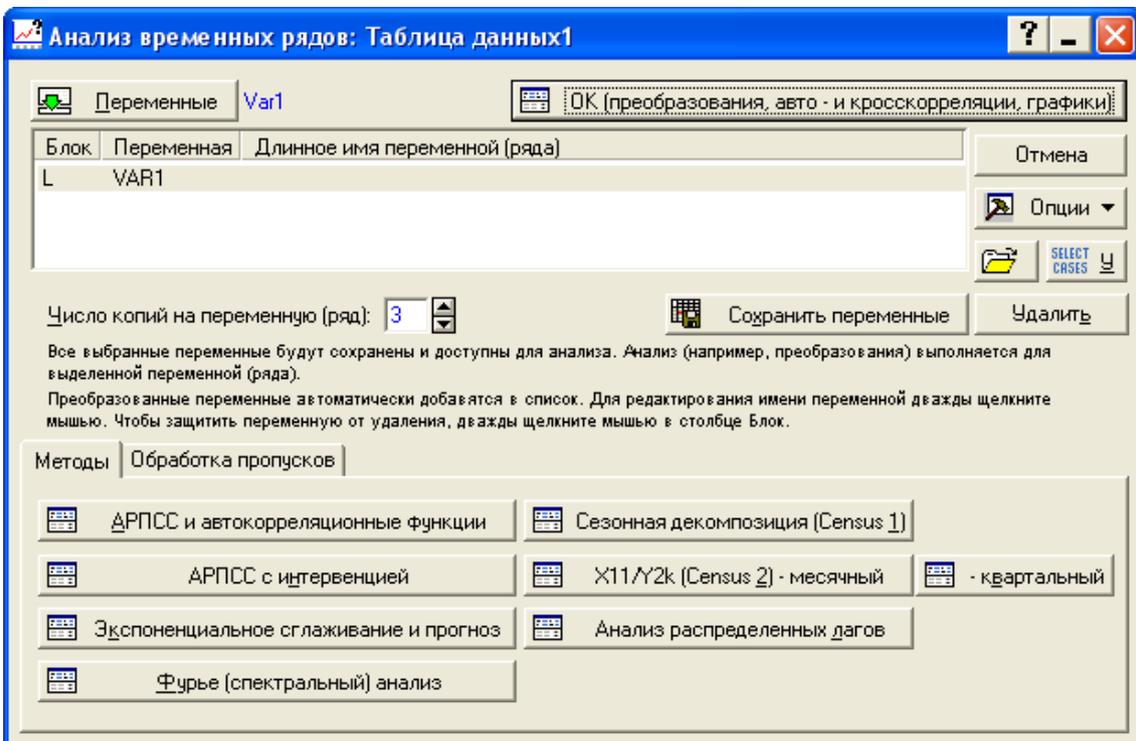


Рис. 11.3 Диалоговое окно «Временные ряды и прогнозирование»

В появившемся диалоговом окне выбираем вкладку «Поиск на сетке» (рисунок 11.4).

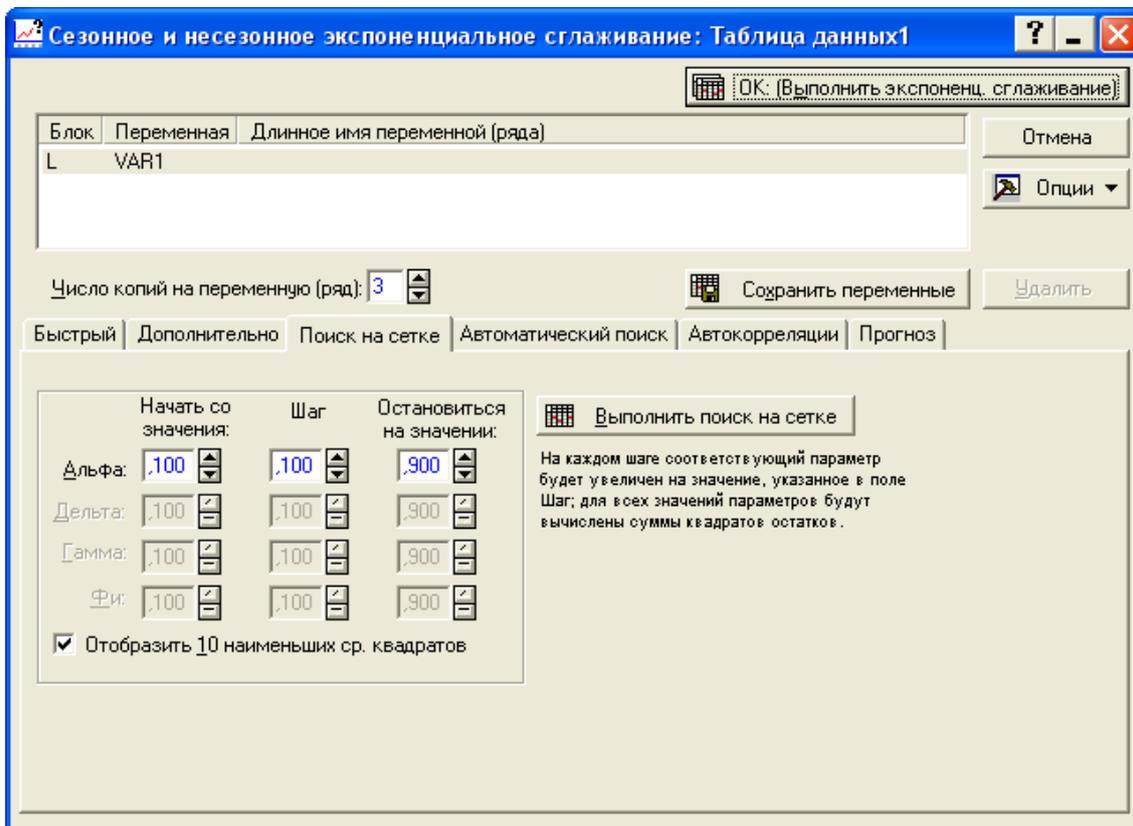


Рис. 11.4 Диалоговое окно экспоненциального сглаживания

По умолчанию перебор осуществляется с начального значения $\alpha = 0,100$ до $\alpha = 0,900$ с шагом $0,100$ (данные значения можно менять). В результате выводится 10 наилучших значений α . Если необходимо просмотреть все значения α , то необходимо убрать метку в строке «Отобразить 10 наименьших ср.квдратов».

Для просмотра сетки значений необходимо нажать кнопку «Выполнить поиск на сетке». По анализируемым данным наилучшим значением для параметра адаптации является $\alpha = 0,2$, при нем наблюдаются наименьшие значения суммы квадратов отклонений (рисунок 11.5).

В Г Э	Поиск параметров на сетке (наименьшие абс.ошибки выделяются) (Таблица данных1)								
	Модель: Без тр., нет сезон. ; SO=1847, VAR1								
Модель Номер	Альфа	Средняя ошибка	Ср. абс. ошибка	Сумма квадрат.	Средние квадраты	Средн. % ошибка	Ср. абс. % ошиб.		
2	0,200000	96,0306	435,9648	26526858	368428,6	-7,50635	28,00154		
3	0,300000	68,9371	430,7110	26685294	370629,1	-7,99902	27,57308		
4	0,400000	53,5734	436,0078	27897413	387464,1	-8,27288	27,72252		
5	0,500000	43,7774	445,7961	29500724	409732,3	-8,37729	28,20425		
1	0,100000	151,9501	493,4418	30844185	428391,5	-6,98307	31,19912		
6	0,600000	37,0035	458,7238	31291131	434599,0	-8,34582	28,89259		
7	0,700000	32,0200	469,3140	33192141	461002,0	-8,19962	29,37873		
8	0,800000	28,1756	485,0209	35179881	488609,5	-7,95394	29,93973		
9	0,900000	25,1176	509,6560	37263138	517543,6	-7,62160	30,85988		

Рис. 11.5 Сетка значений для выбора α

В качестве начального значения экспоненциальной средней берется средняя арифметическая простая из всех уровней временного ряда, которая для анализируемого ряда составила $S_0 = 1847$. Адаптивная модель примет вид: $S_t = 0,2 \cdot y_t + 0,8 \cdot S_{t-1}$. Чтобы рассчитать ряд экспоненциальных средних и прогноз на следующий период, в диалоговом окне экспоненциального сглаживания необходимо выбрать вкладку «Дополнительно» сделать установки как показано на рисунке 11.6 и нажать на кнопку «ОК: (Выполнить экспоненц. сглаживание)». В результате появится окно, содержащее значения экспоненциальной средней и прогноз на 10 шагов (рисунок 11.7).

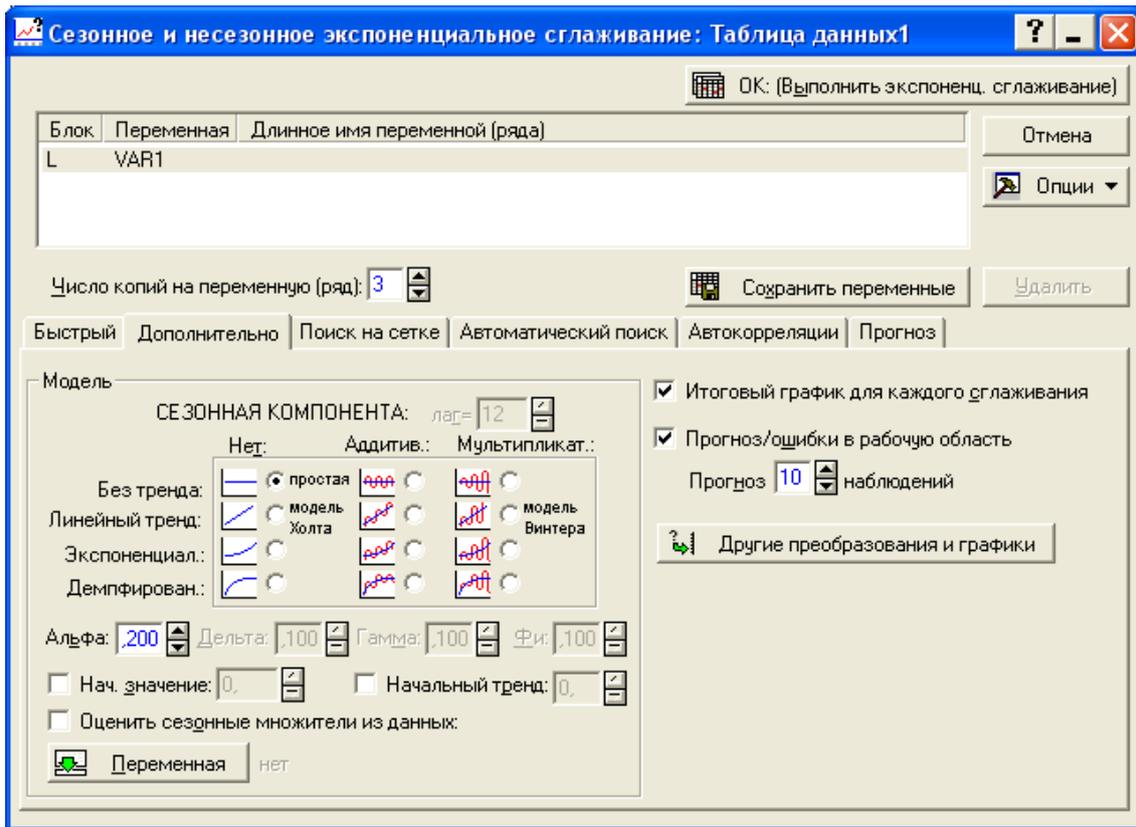


Рис. 11.6 Задание параметров для экспоненциального сглаживания

Экспон. сглаживание: $S_0=1847$, (Таблица данных1) Без тр., нет сезон.; Альфа= ,200 VAR1			
Набл.	VAR1	Сглажен. ряд	Остатки
1	1119,300	1846,787	-727,49
2	352,200	1701,290	-1349,09
3	1006,900	1431,472	-424,57
4	1177,800	1346,558	-168,76
5	1084,400	1312,806	-228,41
6	891,400	1267,125	-375,72
7	928,200	1191,980	-263,78
8	1178,400	1139,224	39,18
9	989,400	1147,059	-157,66
10	932,200	1115,527	-183,33
11	1080,400	1078,862	1,54
12	1243,500	1079,169	164,33
13	865,500	1112,036	-246,54
14	998,400	1062,728	-64,33
15	1145,100	1049,863	95,24
16	1585,600	1068,910	516,69
17	1301,000	1172,248	128,75
18	980,300	1197,999	-217,70
19	1403,500	1154,459	249,04
20	1455,700	1204,267	251,43
21	1163,500	1254,554	-91,05
22	1532,000	1236,343	295,66
23	1299,900	1295,474	4,43
24	1549,100	1296,359	252,74
25	968,800	1346,908	-378,11
26	900,000	1271,286	-371,29
27	1402,000	1197,029	204,97
28	1898,800	1238,023	660,78
29	1538,800	1370,178	168,62
30	1232,700	1403,903	-171,20
31	1650,100	1369,662	280,44
32	1486,900	1425,750	61,15
33	1364,300	1437,980	-73,68
34	1974,600	1423,244	551,36
35	1551,100	1533,515	17,58
36	1795,600	1537,032	258,57
37	1196,800	1588,746	-391,95
38	1091,100	1510,357	-419,26
39	1629,400	1426,505	202,89
40	2620,200	1467,084	1153,12
41	1603,700	1697,707	-94,01

Рис. 11.7 Результаты экспоненциального сглаживания
и прогнозирования

Глава 12. ПРОГНОЗИРОВАНИЕ С ПОМОЩЬЮ ARMA И ARIMA –ПРОЦЕССОВ

12.1 Стационарные временные ряды и их основные характеристики

Рассмотрим формальное определение стационарности.

Стохастический процесс Y_t называется *стационарным в сильном смысле (строго стационарным или стационарным в узком смысле)*, если совместное распределение вероятностей всех переменных $y_{t1}, y_{t2}, \dots, y_{tm}$ точно то же самое, что и для переменных $y_{t1+\tau}, y_{t2+\tau}, \dots, y_{tm+\tau}$.

Под стационарным процессом в слабом смысле (в широком смысле) понимается стохастический процесс, для которого среднее и дисперсия независимо от рассматриваемого периода времени имеют постоянное значение, а автоковариация зависит только от длины лага между рассматриваемыми переменными:

$$\begin{aligned}\mu(y_t) &= \mu(y_{t+\tau}) = \mu; \\ D(y_t) &= \mu(y_t - \mu)^2 = \mu(y_{t+\tau} - \mu)^2 = \gamma(0) = const; \\ cov(y_t, y_{t+\tau}) &= \mu[(y_t - \mu)(y_{t+\tau} - \mu)] = \gamma(\tau).\end{aligned}\tag{12.1}$$

Из этого следует, что автокорреляция будет зависеть только от сдвига по времени τ и не будет зависеть от t .

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

С понятием автоковариационной функции тесно связано понятие автокорреляционной функции (АКФ):

$$\rho(\tau) = \frac{\text{cov}(y_t, y_{t+\tau})}{D(y_t)} = \frac{\gamma(\tau)}{\gamma(0)} \quad (12.2)$$

Значения АКФ также характеризуют тесноту статистической связи между уровнями временного ряда, разделенными τ временными тактами. Однако, в отличие от значений автоковариационной функции они не зависят от масштаба измерения уровней временного ряда и подчиняются ограничению: $|\rho(\tau)| \leq 1$.

Выборочная оценка коэффициента автокорреляции $r(\tau)$ может быть определена следующим образом:

$$r(\tau) = \frac{\frac{1}{n-\tau} \sum_{t=1}^{n-\tau} (y_t - \bar{y})(y_{t+\tau} - \bar{y})}{\frac{1}{n} \sum_{t=1}^n (y_t - \bar{y})^2} \quad (12.3)$$

где n - длина временного ряда;

τ - временной сдвиг (лаг);

\bar{y} - среднее значение временного ряда.

Числитель выражения представляет выборочную оценку коэффициента автоковариации.

В практических руководствах рекомендуется поддерживать соотношение $\tau \leq \frac{n}{4}$.

Для стационарного временного ряда с увеличением τ АКФ должна демонстрировать свойство монотонного убывания по абсолютной величине, т.к. взаимосвязь между уровнями ряда с ростом τ

ослабевает. Однако условие может нарушаться для выборочных АКФ³⁹.

Идея перенесения частной корреляции на временные ряды находит свое выражение в частной АКФ – ЧАКФ. С помощью ЧАКФ измеряется корреляция между уровнями ряда y_t и $y_{t+\tau}$, разделенными τ временными тактами, при исключении влияния на эту взаимосвязь всех промежуточных уровней ряда $y_{t+1}, y_{t+2}, \dots, y_{t+\tau-1}$.

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

$$\rho_\tau(2) = \rho(y_t, y_{t+2} | y_{t+1}) = \mu. \quad (12.4)$$

Формула для расчета выборочной оценки частного коэффициента автокорреляции:

$$r_{ij.k} = \frac{r_{ij} - r_{ik} \cdot r_{jk}}{\sqrt{(1 - r_{ik}^2)(1 - r_{jk}^2)}}. \quad (12.5)$$

Например, r_τ 1-го порядка между y_t и y_{t+2} при устранении влияния y_{t+1} :

$$r_\tau(2) = r_{02.1} = \frac{r(2) - r(1) \cdot r(1,2)}{\sqrt{(1 - r^2(1))(1 - r^2(1,2))}}, \quad (12.6)$$

³⁹ Дуброва, Т.А. Статистические методы прогнозирования: учеб. пособие для вузов / Т.А. Дуброва - М.: ЮНИТИ-ДАНА, 2003. –206 с.

где $r(2)$ - коэффициент корреляции между y_t и y_{t+2} ;

$r(1)$ - коэффициент корреляции между y_t и y_{t+1} ;

$r(1,2)$ - коэффициент корреляции между y_{t+1} и y_{t+2} .

В практической аналитической работе стационарность временного ряда означает отсутствие⁴⁰:

- тренда;
- систематических изменений дисперсии;
- строго периодичных флуктуаций;
- систематически изменяющихся взаимосвязей между элементами временного ряда.

12.2 Модели стационарных временных рядов

Часто экономические показатели, представленные временными рядами, имеют настолько сложную структуру, что моделирование таких рядов путем построения моделей тренда, сезонности и применения традиционных подходов не приводит к удовлетворительным результатам. Во временных рядах остатков прослеживаются статистические зависимости, которые можно моделировать.

В последнее время большое внимание уделяется моделированию стационарных временных рядов, т.к. многие временные ряды могут быть приведены к стационарному виду после операции выделения тренда, фильтрации сезонной компоненты или взятия разности. Как правило, ряд остатков – это стационарный ряд. Наиболее распространенные модели стационарных рядов – модели авторегрессии и модели

⁴⁰ Эконометрика: учебник / И.И. Елисеева [и др.]. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2006. – 576 с.

скользящего среднего.

12.2.1 Авторегрессионные модели

В авторегрессии каждое значение ряда находится в линейной зависимости от предыдущих значений. Если анализируемый динамический процесс зависит от значений, отстоящих на p временных лагов назад, то авторегрессионный процесс порядка p , т.е. AR (p):

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t, \quad (12.7)$$

где ε_t - «белый шум» с $\mu_\varepsilon = 0$.

α_0 - свободный член (часто приравнивается к нулю (опускается)).

Используя функцию оператора лага

$$(1 - \alpha_0 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p) y_t = \Phi(B) y_t = \varepsilon_t, \quad (12.8)$$

где B - оператор сдвига, т.е. преобразование ряда, смещающего его на один временной такт;

$\Phi(B)$ – оператор авторегрессии.

Для выполнения условия стационарности все корни многочлена $\Phi(B)$ должны лежать вне единичного круга, т.е. все корни характеристического уравнения $1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_p z^p = 0$ должны быть по модулю больше 1 и различны, т.е. $|z| > 1$. Если $|z| = 1$, процесс называется

ся процессом единичного корня и является нестационарным.

Рассмотрим простейший вариант линейного авторегрессионного процесса – модель авторегрессии 1-го порядка – AR(1), или марковский процесс⁴¹.

Эта модель может быть представлена в виде:

$$y_t = \alpha y_{t-1} + \varepsilon_t, \quad (12.9)$$

где α - числовой коэффициент, $|\alpha| < 1$,

ε_t - последовательность случайных величин, образующих белый шум.

Основные свойства Марковского процесса:

$$\begin{aligned} \mu_{y_t} &= 0 \\ D(y_t) &= \frac{\sigma_0^2}{1 - \alpha^2} \\ \text{cov}(y_t, y_{t \pm k}) &= \alpha^k D(y_t) \\ \rho(y_t, y_{t \pm k}) &= \alpha^k \end{aligned} \quad (12.10)$$

Значения частной автокорреляционной функции равны нулю для всех лагов $k \geq 2$, что может быть использовано при подборе модели. Этот результат для теоретической ЧАКФ и может не выполняться для выборочной АКФ. Однако если выборочные частные корреляции статистически незначимо отличаются от нуля при $k \geq 2$, то использование модели AR(1) не противоречит исходным данным.

Условие стационарности ряда для AR(1) определяется требованием к коэффициенту α : $|\alpha| < 1$.

⁴¹ Эконометрика: учебник / И.И. Елисеева [и др.]. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2006. – 576 с.

Из авторегрессионных процессов выше первого порядка в экономической практике часто встречаются так называемые процессы Юла. Они описываются с помощью модели AR(2):

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t \quad (12.11)$$

Выражение для вычисления любого значения АКФ $\rho(k)$:

$$\rho(k) = \alpha_1 \rho(k-1) + \alpha_2 \rho(k-2). \quad (12.12)$$

Подставим в данное выражение значение $k=1,2$. С учетом того, что $\rho(0)=1$, а $\rho(-1)=\rho(1)$, получим:

$$\begin{cases} \rho(1) = \alpha_1 + \alpha_2 \rho(1) \\ \rho(2) = \alpha_1 \rho(1) + \alpha_2 \end{cases} \quad (12.13)$$

Эта система называется системой Юла - Уокера для AR(2). Из нее можно получить выражения для определения параметров α_1 и α_2 :

$$\alpha_1 = \frac{\rho(1)[1 - \rho(2)]}{1 - \rho^2(1)} \quad (12.14)$$

$$\alpha_2 = \frac{\rho(2) - \rho^2(1)}{1 - \rho^2(1)}. \quad (12.15)$$

Условия стационарности процесса AR(2):

$$\begin{aligned}
|\alpha_2| &< 1; \\
\alpha_1 + \alpha_2 &< 1; \\
\alpha_2 - \alpha_1 &< 1.
\end{aligned}
\tag{12.16}$$

ЧАКФ для процесса AR(p) будет иметь ненулевые значения лишь при $k \leq p$, а начиная с лага $k = p + 1$ теоретическая ЧАКФ равна нулю. Это свойство становится ключевым при подборе порядка p авторегрессионной модели для конкретных экономических временных рядов⁴².

Рассмотрим пример. Построим авторегрессионную модель первого порядка доходов бюджета Оренбургской области (таблица 5.1). Для оценивания параметров модели AR (p) можно применить один из следующих вариантов:

1) регрессия обычным МНК y_t на $y_{t-1}, y_{t-2}, \dots, y_{t-p}$, хотя некоторые свойства тестовых статистик будут искажены ввиду присутствия среди регрессоров лагов зависимой переменной;

2) максимизация логарифмической функции правдоподобия (МП – оценивание). Для длинных временных рядов различие с результатами МНК незначительно;

3) разрешение уравнений Юла-Уокера.

Рассмотрим первый способ: МНК – оценивание параметров уравнения AR (p).

Так как данный вид модели применяется только для стационарных временных рядов необходимо проверить гипотезу о наличии тенденции либо применить графический анализ.

⁴² Эконометрика: учебник / И.И. Елисеева [и др.]. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2006. – 576 с.

Так как временной ряд доходов бюджета Оренбургской области нестационарный то для построения по нему авторегрессионной модели его необходимо привести к стационарному виду. Для этого необходимо исключить тенденцию. Так как она близка к линейной, найдем отклонения от прямолинейного тренда и построим авторегрессионную модель для ряда остатков. Уравнение тренда имеет вид:

$$\hat{y}_t = 586,3 + 34,54 \cdot t .$$

Для выбора порядка авторегрессионной модели необходимо изучить поведение автокорреляционной (АКФ) и частной автокорреляционной (ЧАКФ) функций. Воспользуемся ППП STATISTICA.

Стандартным образом запустите модуль «Временные ряды и прогнозирование». В этом модуле выбираем «АРПСС и автокорреляционные функции».

В появившемся окне выбираем вкладку «Автокорреляции». Для получения АКФ нажимаем «Автокорреляции», а для ЧАКФ – «Частные автокорреляции» (рисунок 12.1). По умолчанию расчет производится до 15 лага с уровнем надежности 0,05 (данные значения исследователь может задать самостоятельно, так в нашем случае зададим максимальное число лагов 18).

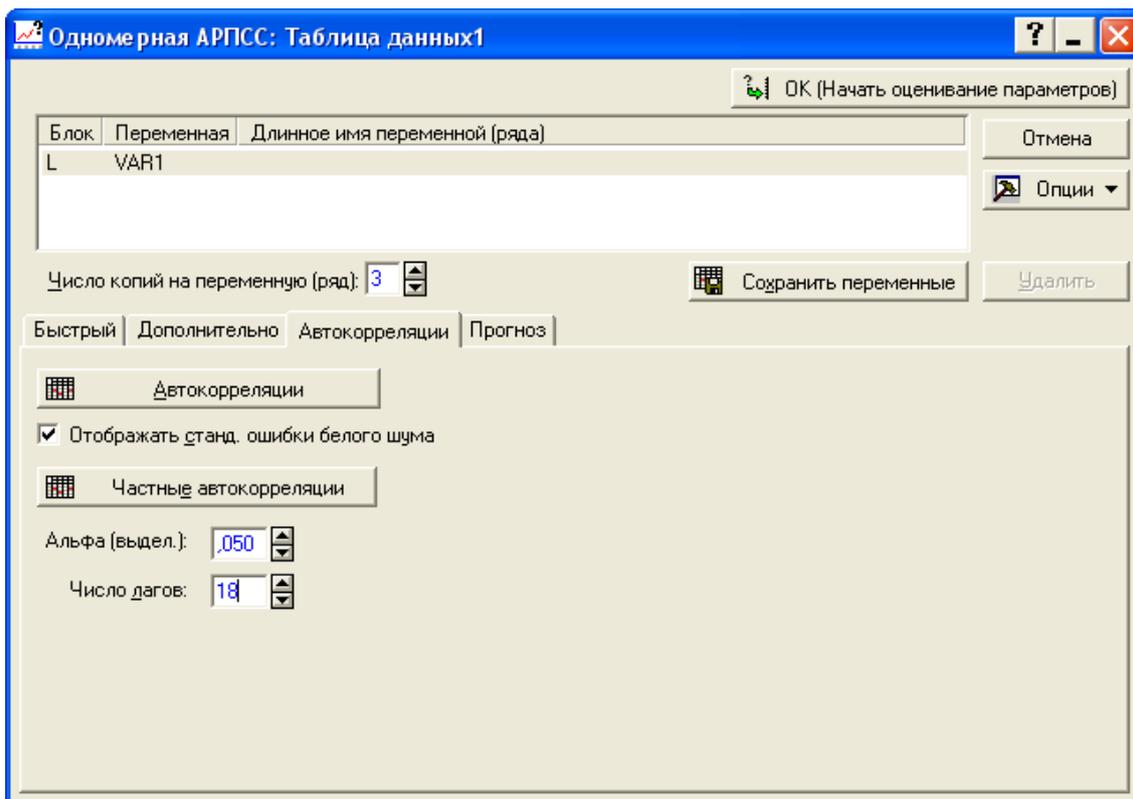


Рис. 12.1 Построение АКФ и ЧАКФ

Как видно на рисунках 12.2 и 12.3, АКФ и ЧАКФ экспоненциально затухает, меняя знак. Следовательно, можно предположить, что для описания временного ряда целесообразно применить модель авторегрессии со скользящими средними в остатках 1 порядка ARMA (1,1). Вместе с тем рассмотрим построение различных моделей и выберем наилучшую из них.

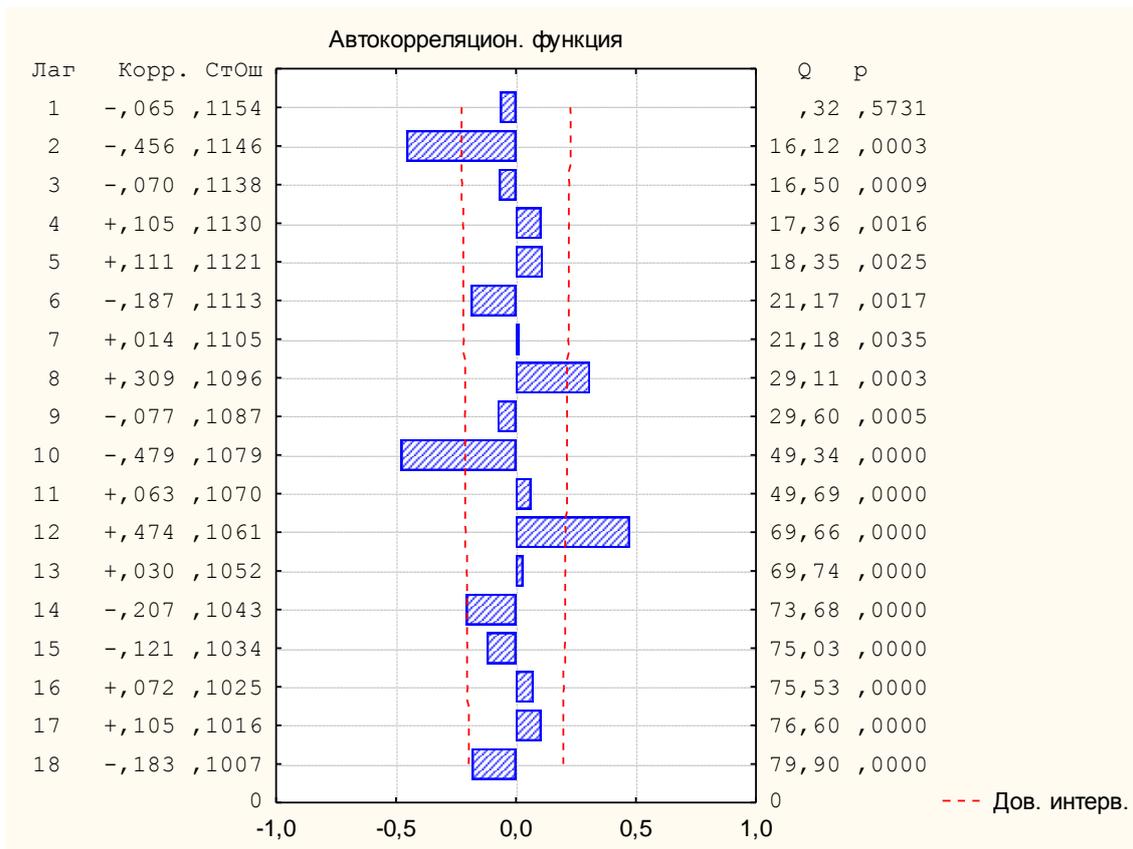


Рис. 12.2 Автокорреляционная функция временного ряда доходов бюджета

Построение модели AR (1): $\tilde{y}_t = \alpha_1 y_{t-1} + \varepsilon_t$.

Параметры данной модели определим с помощью МНК, используя функцию ППП Excel Сервис – Анализ данных - Регрессия. «Входной интервал Y» - выделяем столбец ε_t , «Входной интервал X» - выделяем столбец ε_{t-1} (исходные данные для построения модели представлены в таблице 12.1). Результаты регрессионного анализа представлены в таблице 12.2.

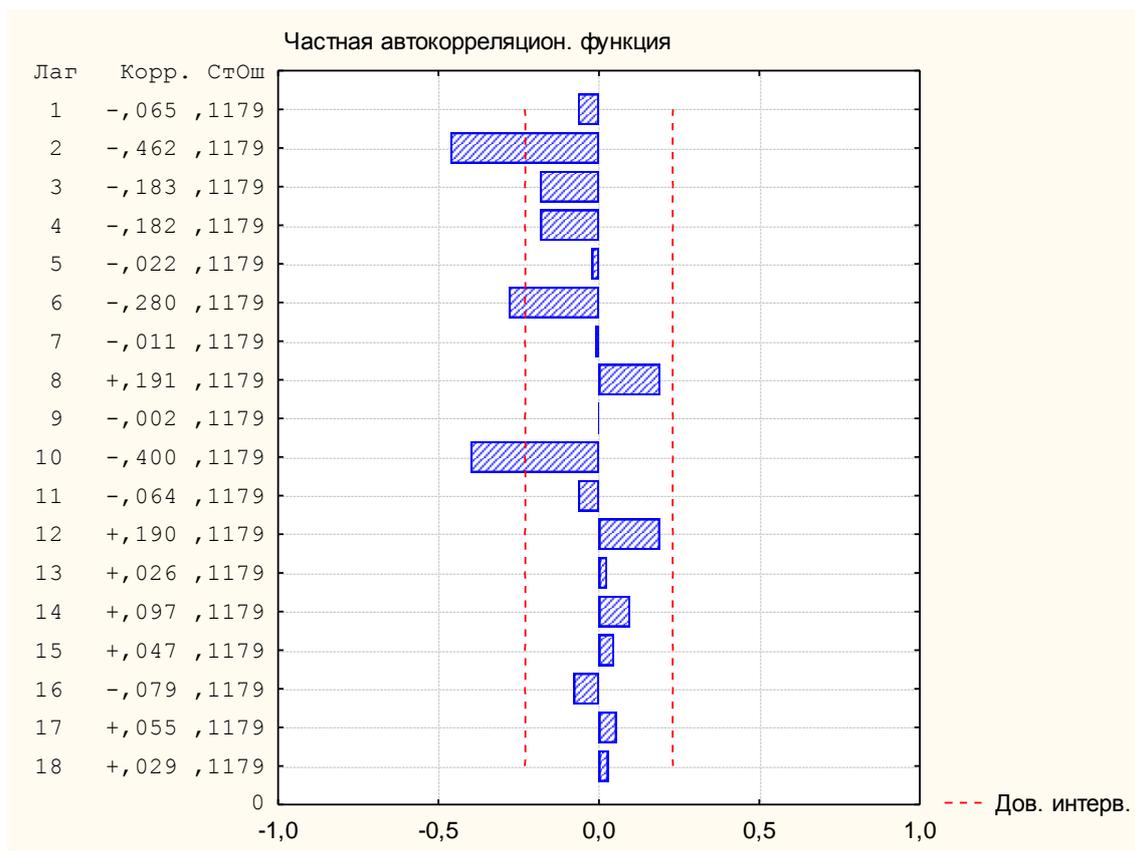


Рис. 12.3 Частная автокорреляционная функция доходов бюджета

Таким образом, нами получена модель AR(1): $\varepsilon_t = 0,03588 \cdot \varepsilon_{t-1}$.

Так как случайный компонент может быть выражен как $\varepsilon_t = y_t - \hat{y}_t = y_t - 586,26 - 34,535 \cdot t$, то подставив это выражение в модель AR (1) получим:

$$y_t - 586,26 - 34,535 \cdot t = 0,03588 \cdot (y_{t-1} - 586,26 - 34,535 \cdot (t-1))$$

В результате соответствующих преобразований получим следующую модель доходов бюджета:

$$\tilde{y}_t = 0,03588 \cdot y_{t-1} + 33,296 \cdot t + 516,464$$

Расчетные значения по модели представлены в таблице 12.1 и на рисунке 12.4. Средняя относительная ошибка аппроксимации полученной модели составила 23,77 %. Что свидетельствует об удовлетворительном качестве модели.

Таблица 12.1

**Исходные и расчетные данные для построения AR(1) – модели
доходов бюджета Оренбургской области (млн. руб.)**

Период	Доход, млн.руб.	t	\hat{y}_t	$\varepsilon_t = y_t - \hat{y}_t$	Исходные данные для оценивания параметров		\tilde{y}_t	$\left \frac{y - \tilde{y}}{y} \right $	
					ε_t	ε_{t-1}			
1	2	3	4	5	6	7	8	9	
2001	январь	1119,3	1	620,8	498,5	-303,13	498,5	-	-
	февраль	352,2	2	655,3	-303,1	317,035	-303,1	549,76	0,56093
	март	1006,9	3	689,9	317,0	453,4	317,0	602,781	0,40135
	апрель	1177,8	4	724,4	453,4	325,465	453,4	637,98	0,45833
	май	1084,4	5	758,9	325,5	97,93	325,5	672,538	0,37981
	июнь	891,4	6	793,5	97,9	100,195	97,9	707,074	0,20678
	июль	928,2	7	828,0	100,2	315,86	100,2	741,609	0,20102
	август	1178,4	8	862,5	315,9	92,325	315,9	776,144	0,34136
	сентябрь	989,4	9	897,1	92,3	0,59	92,3	810,68	0,18064
	октябрь	932,2	10	931,6	0,6	114,255	0,6	845,215	0,09331

Продолжение таблицы 12.1

1		2	3	4	5	6	7	8	9
	ноябрь	1080,4	11	966,1	114,3	242,82	114,3	879,75	0,18572
	декабрь	1243,5	12	1000,7	242,8	-169,72	242,8	914,285	0,26475
2002	январь	865,5	13	1035,2	-169,7	-71,35	-169,7	948,82	0,09627
	февраль	998,4	14	1069,8	-71,4	40,815	-71,4	983,355	0,01507
	март	1145,1	15	1104,3	40,8	446,78	40,8	1017,89	0,11109
	апрель	1585,6	16	1138,8	446,8	127,645	446,8	1052,42	0,33626
	май	1301	17	1173,4	127,6	-227,59	127,6	1086,96	0,16452
	июнь	980,3	18	1207,9	-227,6	161,075	-227,6	1121,49	0,14403
	июль	1403,5	19	1242,4	161,1	178,74	161,1	1156,03	0,17632
	август	1455,7	20	1277,0	178,7	-148	178,7	1190,56	0,18214
	сентябрь	1163,5	21	1311,5	-148,0	185,97	-148,0	1225,1	0,05294
	октябрь	1532	22	1346,0	186,0	-80,665	186,0	1259,64	0,17778
	ноябрь	1299,9	23	1380,6	-80,7	134	-80,7	1294,17	0,00441
	декабрь	1549,1	24	1415,1	134,0	-480,84	134,0	1328,71	0,14227

Продолжение таблицы 12.1

	1	2	3	4	5	6	7	8	9
2003	январь	968,8	25	1449,6	-480,8	-584,17	-480,8	1363,24	0,40714
	февраль	900	26	1484,2	-584,2	-116,71	-584,2	1397,78	0,55308
	март	1402	27	1518,7	-116,7	345,56	-116,7	1432,31	0,02162
	апрель	1898,8	28	1553,2	345,6	-48,975	345,6	1466,85	0,22749
	май	1538,8	29	1587,8	-49,0	-389,61	-49,0	1501,38	0,02432
	июнь	1232,7	30	1622,3	-389,6	-6,745	-389,6	1535,92	0,24598
	июль	1650,1	31	1656,8	-6,7	-204,48	-6,7	1570,45	0,04827
	август	1486,9	32	1691,4	-204,5	-361,62	-204,5	1604,99	0,07942
	сентябрь	1364,3	33	1725,9	-361,6	214,15	-361,6	1639,52	0,20173
	октябрь	1974,6	34	1760,5	214,2	-243,89	214,2	1674,06	0,15221
	ноябрь	1551,1	35	1795,0	-243,9	-33,92	-243,9	1708,59	0,10153
	декабрь	1795,6	36	1829,5	-33,9	-667,26	-33,9	1743,13	0,02922
2004	январь	1196,8	37	1864,1	-667,3	-807,49	-667,3	1777,66	0,48534
	февраль	1091,1	38	1898,6	-807,5	-303,73	-807,5	1812,2	0,66089
	март	1629,4	39	1933,1	-303,7	652,54	-303,7	1846,73	0,13338

Продолжение таблицы 12.1

	1	2	3	4	5	6	7	8	9
	апрель	2620,2	40	1967,7	652,5	-398,5	652,5	1881,27	0,28201
	май	1603,7	41	2002,2	-398,5	-343,93	-398,5	1915,8	0,19461
	июнь	1692,8	42	2036,7	-343,9	196,235	-343,9	1950,34	0,15214
	июль	2267,5	43	2071,3	196,2	-301,2	196,2	1984,87	0,12464
	август	1804,6	44	2105,8	-301,2	-357,54	-301,2	2019,41	0,11903
	сентябрь	1782,8	45	2140,3	-357,5	-253,87	-357,5	2053,94	0,15209
	октябрь	1921	46	2174,9	-253,9	592,895	-253,9	2088,48	0,08718
	ноябрь	2802,3	47	2209,4	592,9	395,66	592,9	2123,01	0,2424
	декабрь	2639,6	48	2243,9	395,7	-1334,4	395,7	2157,55	0,18262
2005	январь	944,1	49	2278,5	-1334,4	-995,71	-1334,4	2192,08	1,32187
	февраль	1317,3	50	2313,0	-995,7	545,655	-995,7	2226,62	0,69029
	март	2893,2	51	2347,5	545,7	-147,78	545,7	2261,15	0,21846
	апрель	2234,3	52	2382,1	-147,8	-22,915	-147,8	2295,69	0,02747
	май	2393,7	53	2416,6	-22,9	-616,95	-22,9	2330,22	0,02652
	июнь	1834,2	54	2451,2	-617,0	-280,28	-617,0	2364,76	0,28926

Продолжение таблицы 12.1

	1	2	3	4	5	6	7	8	9
	июль	2205,4	55	2485,7	-280,3	531,48	-280,3	2399,29	0,08792
	август	3051,7	56	2520,2	531,5	-519,06	531,5	2433,83	0,20247
	сентябрь	2035,7	57	2554,8	-519,1	-347,99	-519,1	2468,36	0,21254
	октябрь	2241,3	58	2589,3	-348,0	1621,48	-348,0	2502,9	0,11672
	ноябрь	4245,3	59	2623,8	1621,5	1041,34	1621,5	2537,43	0,4023
	декабрь	3699,7	60	2658,4	1041,3	-1119,9	1041,3	2571,97	0,30482
2006	январь	1573	61	2692,9	-1119,9	-1205,9	-1119,9	2606,5	0,65703
	февраль	1521,5	62	2727,4	-1205,9	453,235	-1205,9	2641,04	0,73581
	март	3215,2	63	2762,0	453,2	76	453,2	2675,57	0,16784
	апрель	2872,5	64	2796,5	76,0	961,365	76,0	2710,11	0,05653
	май	3792,4	65	2831,0	961,4	-143,87	961,4	2744,64	0,27628
	июнь	2721,7	66	2865,6	-143,9	197,095	-143,9	2779,18	0,02112
	июль	3097,2	67	2900,1	197,1	1294,56	197,1	2813,71	0,09153
	август	4229,2	68	2934,6	1294,6	-849,58	1294,6	2848,25	0,32653
	сентябрь	2119,6	69	2969,2	-849,6	752,79	-849,6	2882,78	0,36006

Продолжение таблицы 12.1

1	2	3	4	5	6	7	8	9
октябрь	3756,5	70	3003,7	752,8	377,855	752,8	2917,32	0,22339
ноябрь	3416,1	71	3038,2	377,9	405,92	377,9	2951,85	0,1359
декабрь	3478,7	72	3072,8	405,9	-	-	2986,39	0,14152
Итого	132968,7	-	-	-	-	-	-	16,8796
В среднем	1846,7875	-	-	-	-	-	-	23,7741

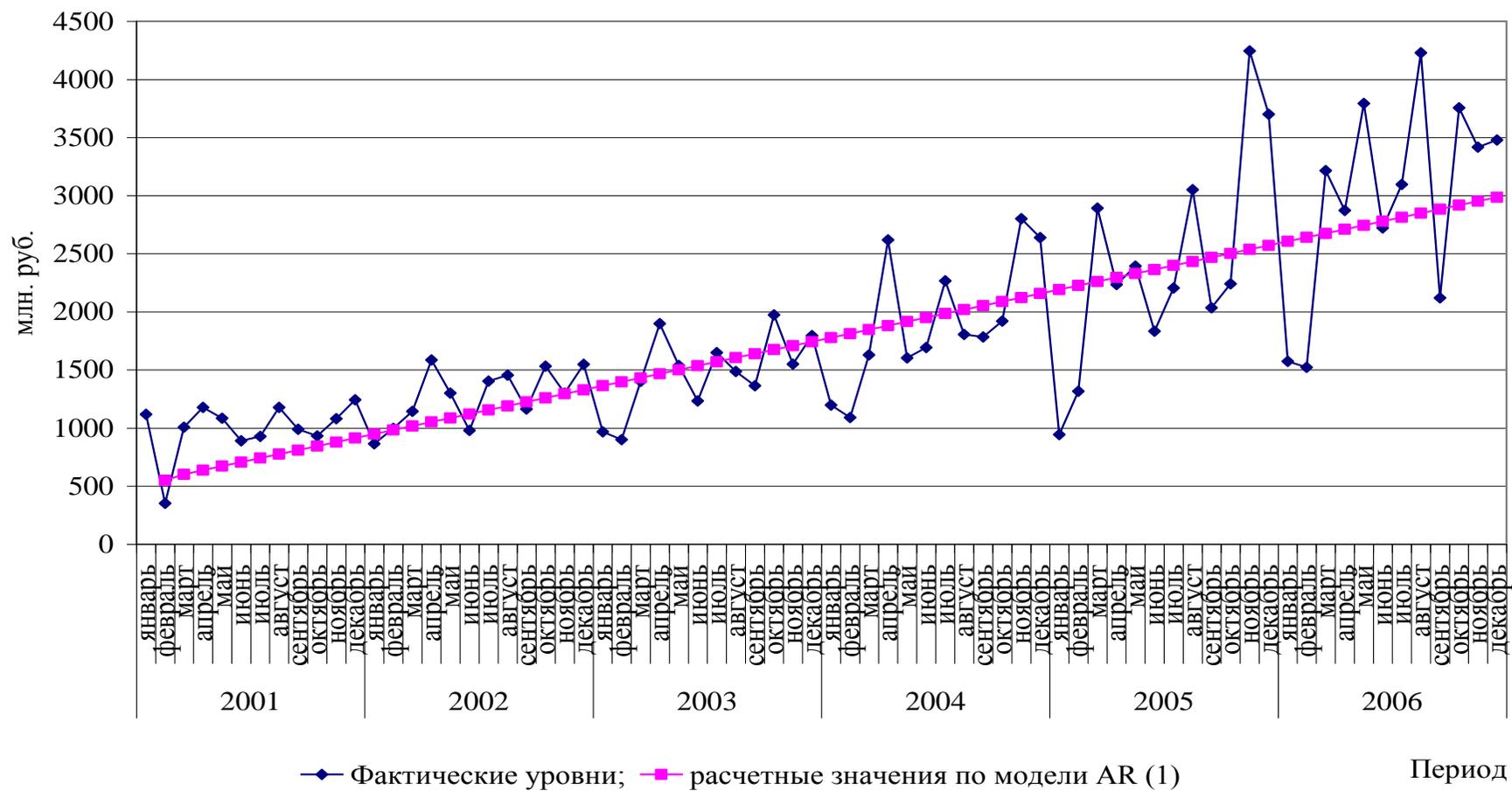


Рис. 12.4 Теоретические значения по авторегрессионной модели

Таблица 12.2

Результаты оценивания модели AR (1) МНК

Регрессионная статистика

Множественный R	0,0359
R-квадрат	0,0013
Нормированный R-квадрат	-0,01299
Стандартная ошибка	535,77
Наблюдения	71

Дисперсионный анализ

	df	SS	MS	F	Значимость F
Регрессия	1	26008,11	26008,11	0,0906	0,7643
Остаток	70	20093865	287055,2		
Итого	71	20119874			

	Коэффициенты	Стандартная ошибка	t-статистика	P-значение	Нижние 95%	Верхние 95%
α_0	0	-	-	-	-	-
α_1	0,03588	0,1192	0,3010	0,7643	-0,2018	0,2736

Прогнозирование значений y_t на период $(t+l)$ по авторегрессионной модели производят следующим образом.

Сначала вычисляют значение \dot{y}_{t+1} по формуле

$$\dot{y}_{t+1} = \alpha_1 y_t + \alpha_2 y_{t-1} + \dots + \alpha_p y_{t-p}$$

Затем в модель $\dot{y}_{t+2} = \alpha_1 \dot{y}_{t+1} + \alpha_2 y_t + \dots + \alpha_p y_{t-p+1}$ подставляют вычисленное значение \dot{y}_{t+1} и определяют величину \dot{y}_{t+2} и т.д.

Спрогнозируем доходы бюджета на январь и февраль 2007г.

Точечный прогноз на январь составит:

$$\dot{y}_{t+1} = 0,03588 \cdot 2986,39 + 33,296 \cdot 73 + 516,464 = 3054,22 \text{ млн.р.},$$

на февраль 2007 г.:

$$\dot{y}_{t+2} = 0,03588 \cdot 3054,22 + 33,296 \cdot 74 + 516,464 = 3089,95 \text{ млн.р.}$$

Интервальный прогноз определяется по формуле:

$$\dot{y}_{t+l} \pm S_y \cdot t(\alpha; k)$$

Среднее квадратическое отклонение:

$$S_y = \sqrt{\frac{\sum (y_t - \tilde{y}_t)^2}{n}} = \sqrt{\frac{20556319,18}{72}} = 534,326.$$

Табличное значение t- критерия Стьюдента: $t(0,05; 69) = 1,995$

Тогда доверительные границы прогноза на январь составят: $3054,22 \pm 1065,95$, т.е. с вероятностью 95 % в январе 2007 года доходы бюджета будут находиться в пределах от 1988,27 до 4120,18 млн.р. В феврале 2007 г. доходы бюджета с заданной вероятностью могут составить от 2024 млн. р. до 4155,9 млн.р.

Построение модели AR (p) в STATISTICA

В модуле «Временные ряды и прогнозирование» выбираем «АРПСС и автокорреляционные функции». В появившемся окне вы-

бираем вкладку «Дополнительно». Так как ряд нестационарный его необходимо преобразовать. Используем для этого 2 способа:

- 1) отклонение от линейного тренда;
- 2) первые разности.

Рассмотрим первый способ

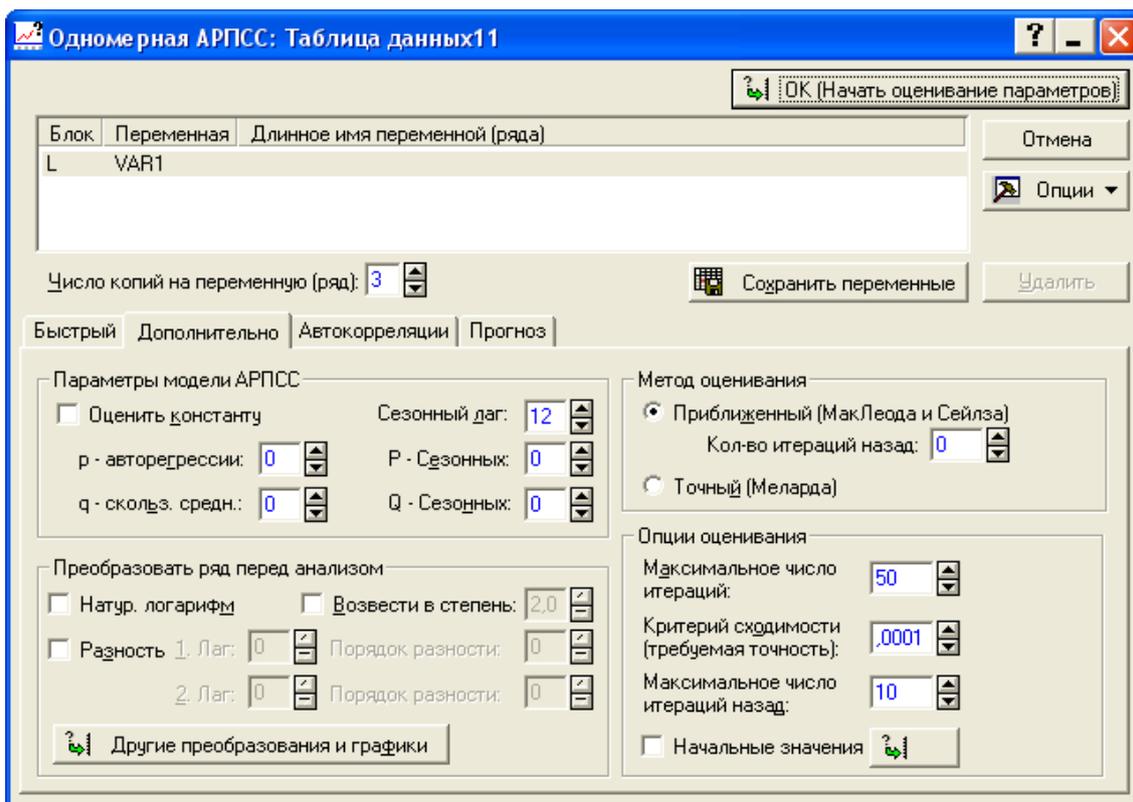


Рис. 12.5 Построение авторегрессионной модели

Для нахождения отклонений от линейного тренда необходимо нажать на кнопку «Другие преобразования и графики» (рисунок 12.5). В появившемся окне необходимо выбрать «Вычесть тренд» ($X = X - (a + b * t)$) и нажать на кнопку «ОК (Преобразовать выделенную переменную)» (рисунок 12.6). На экране появится график отклонений от тренда (рисунок 12.7).

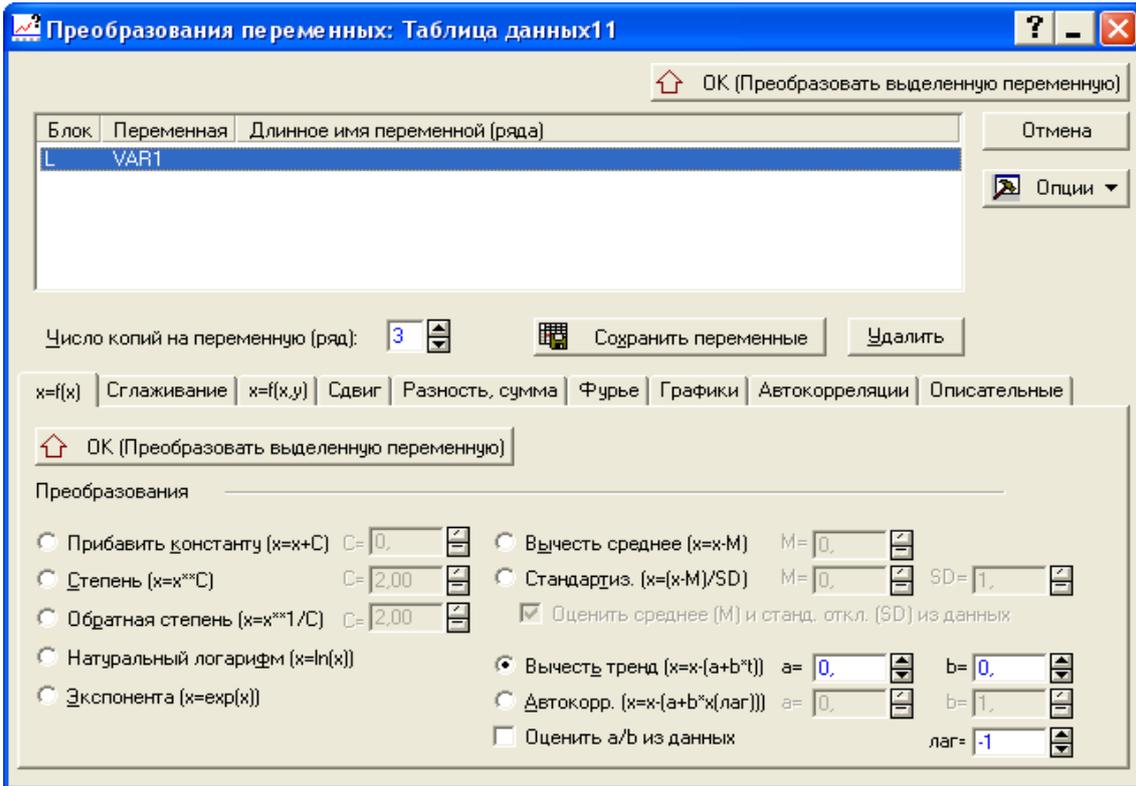


Рис. 12.6 Преобразование временного ряда



Рис. 12.7 Отклонения от линейного тренда

Возвращаемся в диалоговое окно «АРПСС и автокорреляционные функции», нажав на расположенную в левом нижнем углу кнопку «Другие преобразования и графики» и в появившемся окне на кнопку «Отмена».

В результате преобразования ряда появилась новая переменная VAR 1 X-586,3-34,5t (рисунок 12.8), которую мы используем в дальнейших расчетах (параметры уравнения тренда совпадает с оцененными нами ранее).

Для построения модели AR(1) в ячейке «p – авторегрессии» задаем значение порядка авторегрессии. В нашем примере $p = 1$ (рисунок 12.8) и нажимаем на кнопку «ОК (Начать оценивание параметров)».

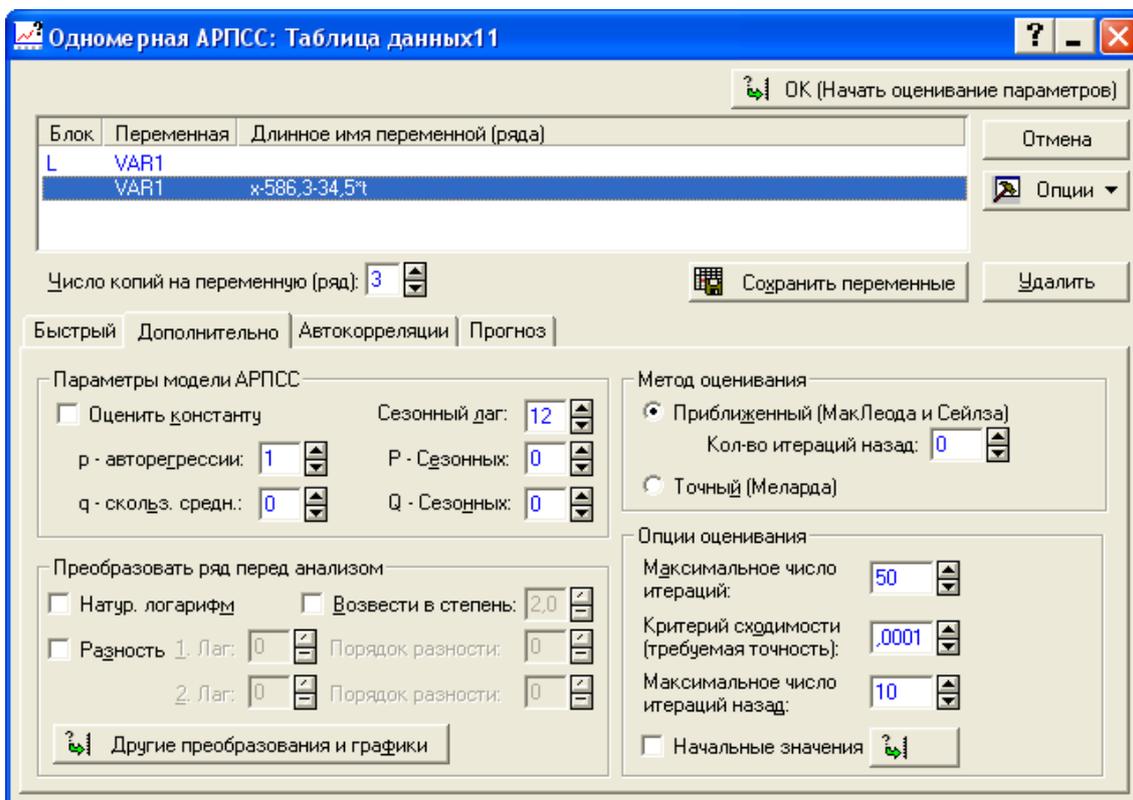


Рис. 12.8 Выбор порядка модели авторегрессии

На экране появится окно, содержащее значение параметра α (рисунок 12.9). Его значение составило 0,03588 (статистически значимые параметры выделяются красным). Значение совпадает с рассчитанным нами ранее.

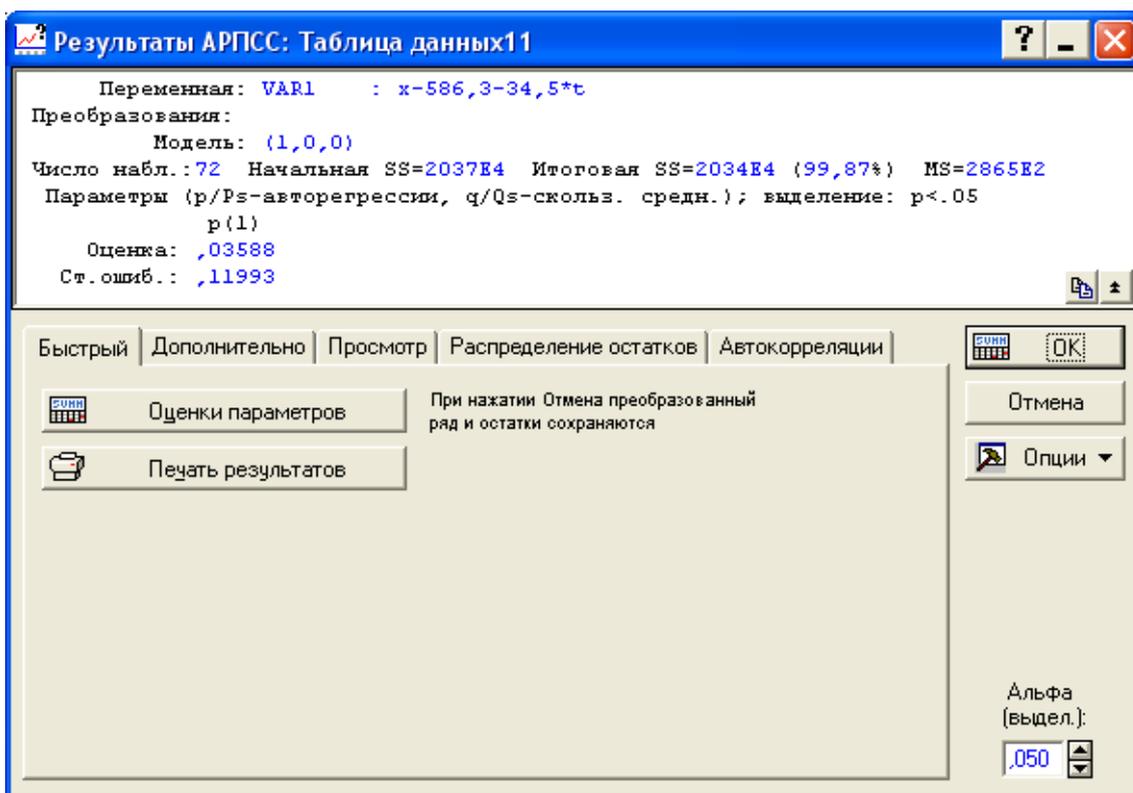


Рис. 12.9 Вывод итогов

Второй способ

Модель авторегрессии первого порядка по первым разностям имеет вид:

$$\Delta_{y_t}^1 = \alpha \cdot \Delta_{y_{t-1}}^1,$$

где $\Delta_{y_t}^1 = y_t - y_{t-1}$, $\Delta_{y_{t-1}}^1 = y_{t-1} - y_{t-2}$, ..., $\Delta_{y_{n-1}}^1 = y_n - y_{n-1}$ - первые разности.

Для расчета ее параметров в модуле «АРПСС и автокорреля-

ционные функции» необходимо задать порядок разности. В ячейке «Разность» устанавливаем его значение как показано на рисунке 12.10 (берем первые разности), в ячейке «р – авторегрессии» задаем значение порядка авторегрессии (р = 1) и нажимаем на кнопку «ОК (Начать оценивание параметров)».

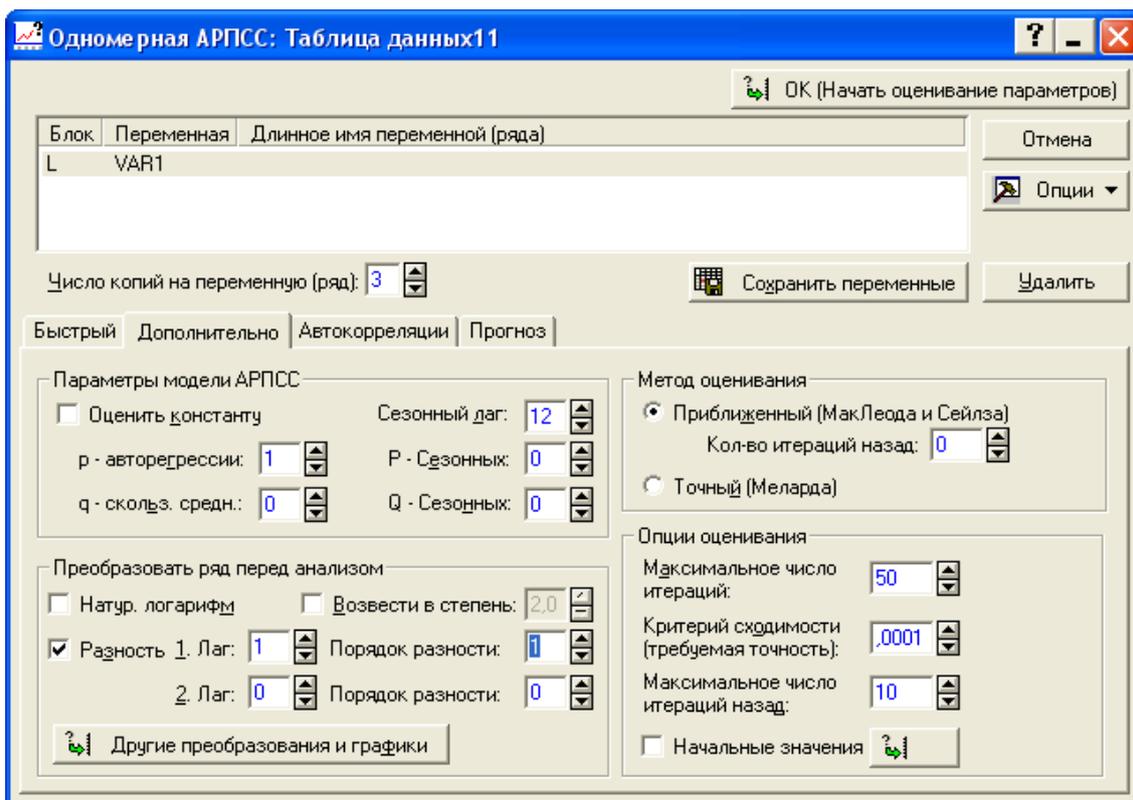


Рис. 12.10 Выбор параметров для построения авторегрессионной модели

Результаты построения авторегрессионной модели по первым разностям представлены на рисунке 12.11. Параметр $\alpha = -0,3105$ статистически значим. Следовательно, модель примет вид:

$$\Delta_{y_t}^1 = -0,3105 \cdot \Delta_{y_{t-1}}^1.$$

В результате соответствующих преобразований получим следующую модель:

$$\tilde{y}_t = 0,6895y_{t-1} + 0,3105y_{t-2}.$$

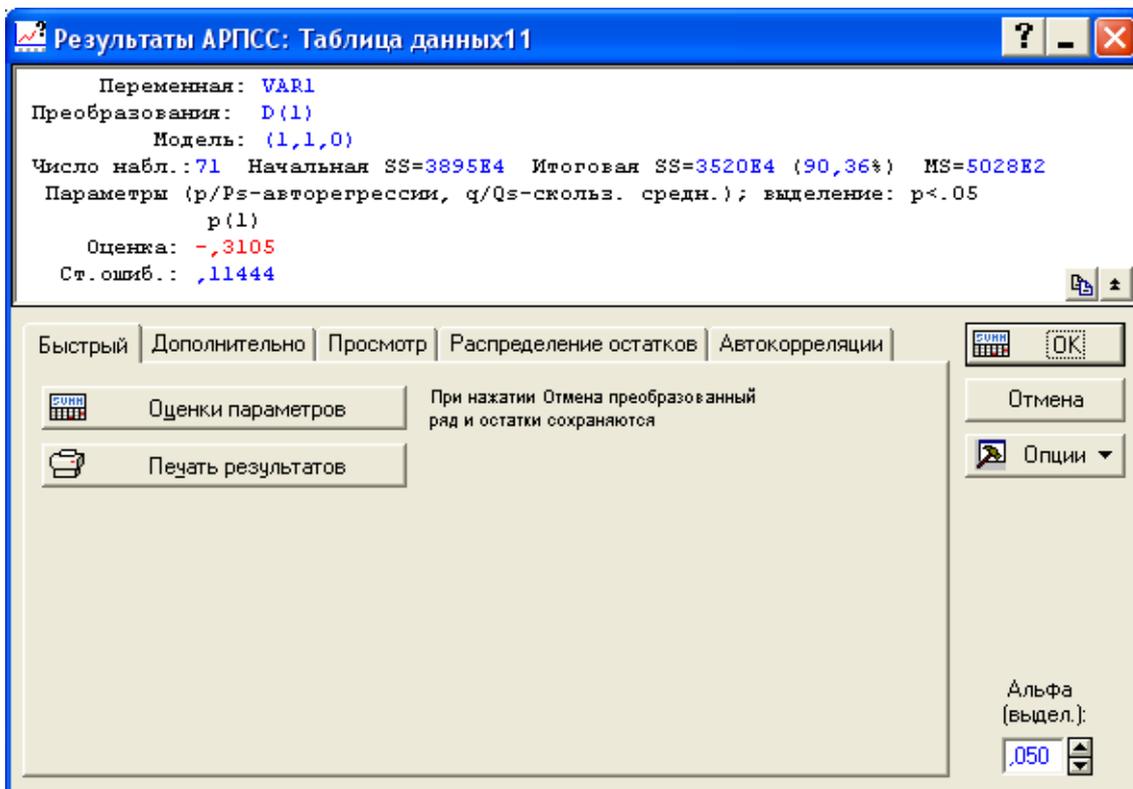


Рис. 12.11 Вывод итогов авторегрессионного моделирования

12.2.2 Модели скользящего среднего

Модель скользящего среднего предполагает, что в ошибках модели в предшествующие периоды сосредоточена информация обо всей предыстории ряда. В этой модели каждое новое значение – среднее между текущей флуктуацией и несколькими (в частности, одной) предыдущими ошибками.

Процесс скользящего среднего порядка q , обозначаемые

МА(q) имеют вид:

$$y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \dots - \beta_q \varepsilon_{t-q}, \quad (12.17)$$

где ε_t - «белый шум» (импульс, шок) с $\mu = 0$.

В моделях скользящего среднего для обеспечения стационарности ряда не требуется накладывать никаких ограничений на параметры $\beta_1, \beta_2, \dots, \beta_q$. Однако, если в модели МА(1) параметр $|\beta| \geq 1$, то текущее значение y_t будет зависеть от своих прошлых значений бесконечно растущими по мере удаления в прошлое:

$$\begin{aligned} \varepsilon_t &= y_t + \beta \varepsilon_{t-1} = y_t + \beta(y_{t-1} + \beta \varepsilon_{t-2}) = y_t + \beta y_{t-1} + \beta^2(y_{t-2} + \beta \varepsilon_{t-3}) = \dots \\ &= y_t + \beta y_{t-1} + \beta^2 y_{t-2} + \beta^3 y_{t-3} + \dots \quad \Rightarrow y_t = \varepsilon_t - \sum_{k=1}^{\infty} \beta^k y_{t-k} \end{aligned} \quad (12.18)$$

Чтобы избежать этого, надо, чтобы веса образовывали сходящийся ряд, т.е. $|\beta| < 1$.

Широко распространены в статистической практике модели скользящего среднего 1-го и 2-го порядка:

$$\text{МА(1): } y_t = \varepsilon_t - \beta \varepsilon_{t-1} \quad (12.19)$$

$$\text{МА(2): } y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2}. \quad (12.20)$$

Рассмотрим пример построения модели МА (q). В случае чистого МА (q) – процесса обычно используются нелинейные методы

наименьших квадратов для оценивания параметров. Наиболее распространен метод условной суммы квадратов (CSS), в котором отсутствующие данные значений «белого шума» ε_t генерируются как ex-post ошибки прогноза при условии минимума суммы квадратов ошибок.

Так для процесса МА (1): $y_t = \varepsilon_t - \beta \cdot \varepsilon_{t-1}$ генерируются значения $\hat{\varepsilon}_t = y_t + \hat{\beta} \cdot \hat{\varepsilon}_{t-1}$ с некоторым коэффициентом $\hat{\beta}$, который оценивается при условии $s(\hat{\beta}) = \sum_1^T \hat{\varepsilon}_t^2 = \min$. Поскольку это нелинейная функция относительно параметра β , минимизация происходит в результате итеративного процесса.

Оценим параметры модели МА (1): $y_t = \varepsilon_t - \beta \cdot \varepsilon_{t-1}$ по данным таблицы 12.2 в ППП STATISTICA.

Так как ряд нестационарный, рассмотрим два способа приведения к стационарному виду:

Взятие первых разностей

Как и при построении AR (p)- модели задаем необходимые значения параметров (рисунок 12.12).

Итог оценивания параметров представлен на рисунке 12.13.

Таким образом, нами получена модель вида:

$$\Delta_{y_t}^1 = \varepsilon_t + 0,79188 \cdot \varepsilon_{t-1}$$

От нее можно перейти к модели для уровней ряда:

$$y_t = y_{t-1} + \varepsilon_t + 0,79188 \cdot \varepsilon_{t-1}.$$

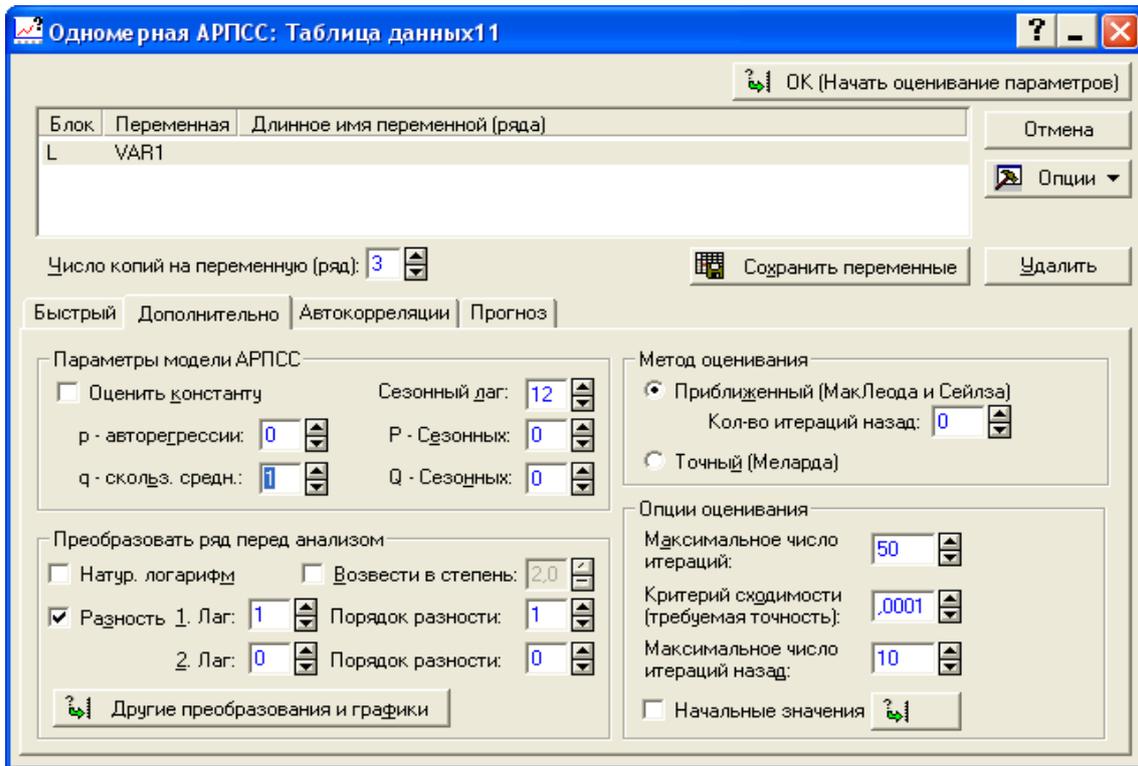


Рис. 12.12 Построение модели скользящего среднего по первым разностям

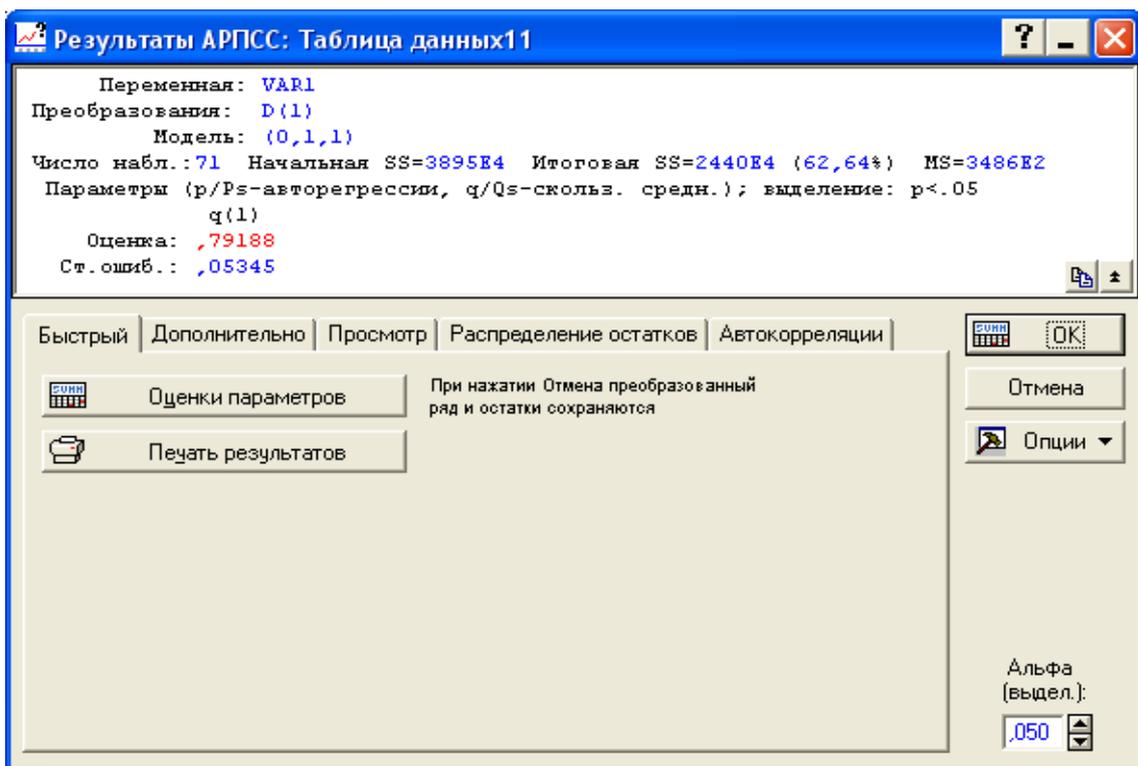


Рис. 12.13 Результаты оценивания MA(1) по первым разностям
Отклонение от линейного тренда

На первом этапе строим линейный тренд и находим от него отклонения «Другие преобразования и графики». В появившемся окне выбираем «Вычесть тренд» ($X=X-(a+b*t)$) и нажимаем на кнопку «ОК (Преобразовать выделенную переменную)». Затем задаем параметры как показано на рисунке 12.14.

Результаты оценивания параметров представлены на рисунке 12.15.

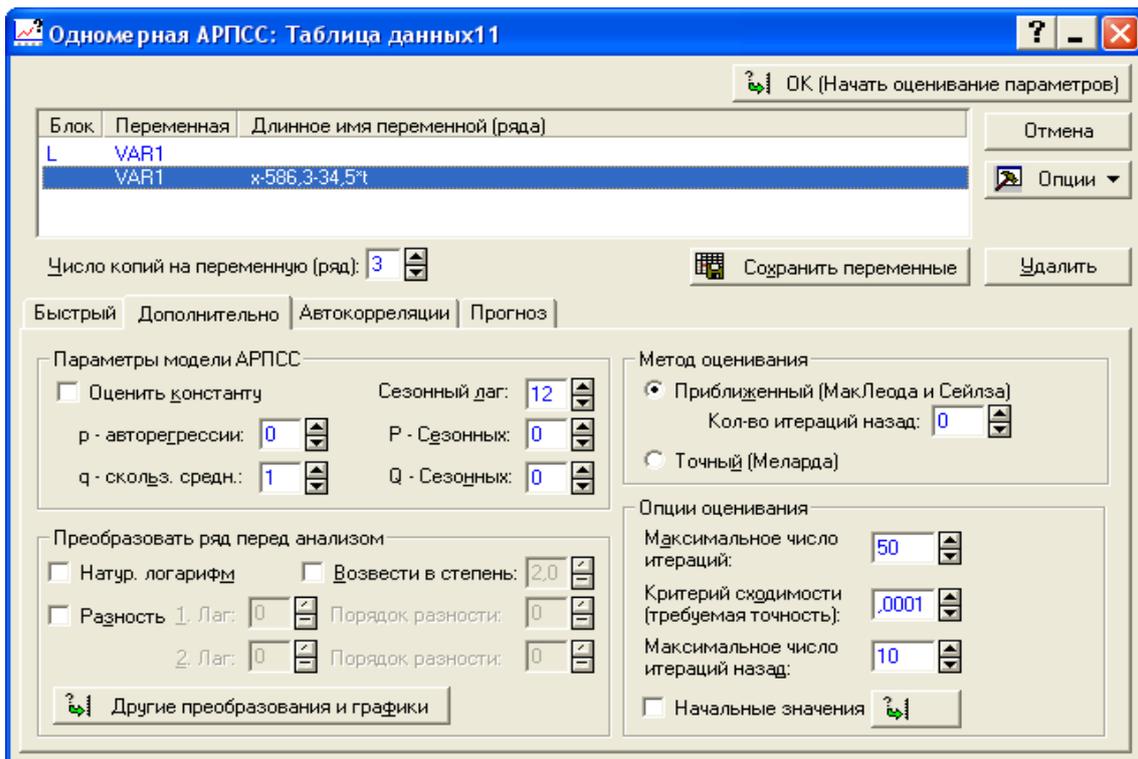


Рис. 12.14 Построение модели скользящего среднего по отклонениям от тренда

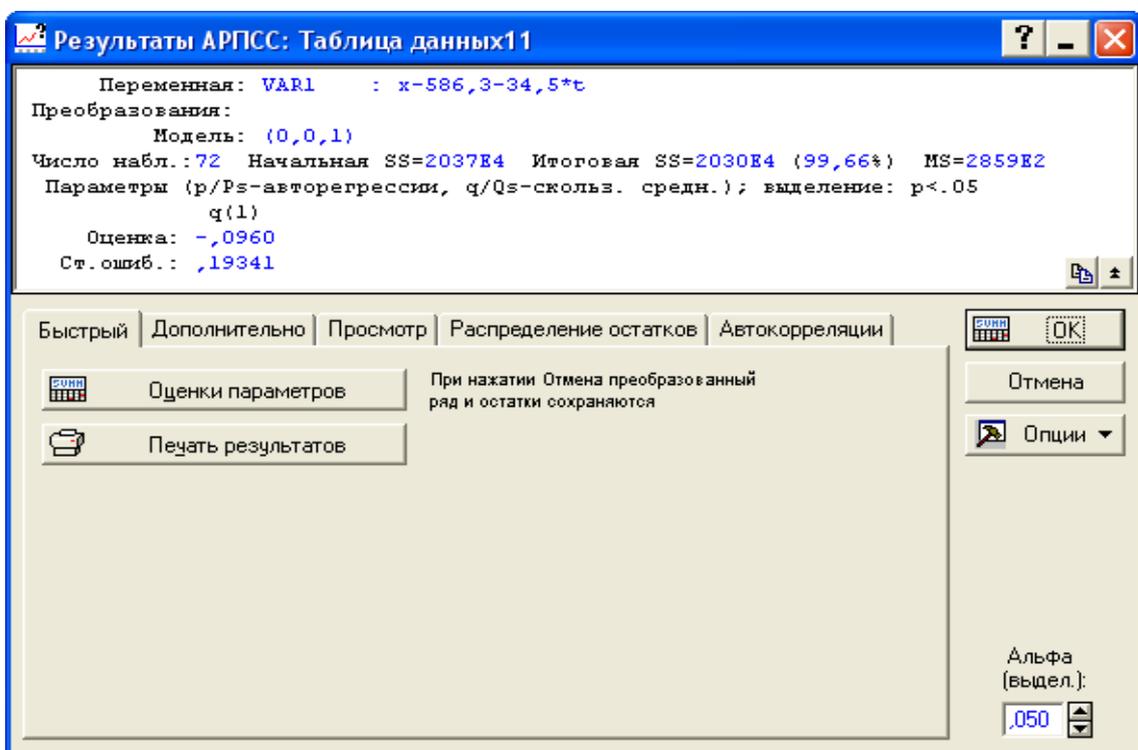


Рис. 12.15 Результаты оценивания MA(1) по отклонениям от тренда

Полученная модель имеет вид:

$$y_t - \tilde{y}_t = \varepsilon_t - 0,0960\varepsilon_{t-1}.$$

Откуда

$$y_t = \tilde{y}_t + \varepsilon_t - 0,0960 \cdot \varepsilon_{t-1} = 586,3 + 34,5 \cdot t + \varepsilon_t - 0,0960 \cdot \varepsilon_{t-1}$$

где \tilde{y}_t - расчетные значения по линейному тренду.

При прогнозировании на практике реальные параметры β_j заменяются своими оценками $\hat{\beta}_j$, а случайные шоки ε_t - на остатки $\hat{\varepsilon}_t$, полученные при оценивании модели, или на ошибки предыдущих прогнозов.

Для модели МА (1) формулы для прогнозирования имеют вид:

$$\begin{aligned} \hat{Y}_{t+1} &= \beta \cdot \varepsilon_t \\ \hat{Y}_{t+h} &= 0 \text{ для } h \geq 2 \end{aligned}$$

В ППП STATISTICA прогнозные значения по полученным моделям можно получить в табличной форме (например, для МА(1) по первым разностям в окне результатов оценивания (рисунок 12.13) на вкладке «Дополнительно» нажав на кнопку «Прогноз») и графической (в том же окне нажав на кнопку «График ряда и прогнозов») соответственно рисунки 12.16 и 12.17.

Прогнозы; Модель:(0,0,1) Сезонный лаг: 12 (Таблица данных11)				
Исход.:VAR1 : $x-586,3-34,5^t$				
Начало исходных: 1 Конец исходн.: 72				
Набл. N	Прогноз	Нижний 90,0000%	Верхний 90,0000%	Ст.ошиб.
73	36,23661	-854,879	927,3527	534,6911
74	0,00000	-895,213	895,2134	537,1496
75	0,00000	-895,213	895,2134	537,1496
76	0,00000	-895,213	895,2134	537,1496
77	0,00000	-895,213	895,2134	537,1496
78	0,00000	-895,213	895,2134	537,1496
79	0,00000	-895,213	895,2134	537,1496
80	0,00000	-895,213	895,2134	537,1496
81	0,00000	-895,213	895,2134	537,1496
82	0,00000	-895,213	895,2134	537,1496

Рис. 12.16 Прогноз по МА(1) для первых разностей

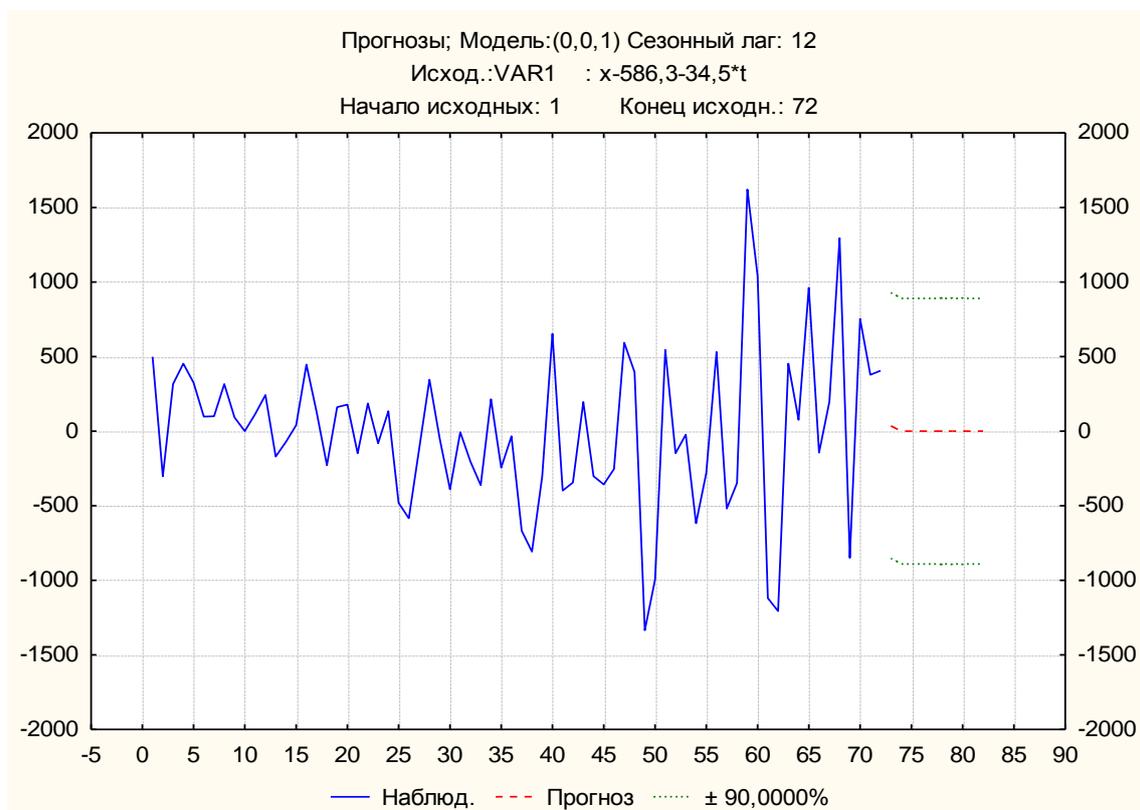


Рис. 12.17 Фактические, прогнозные значения и доверительные границы прогноза доходов бюджета Оренбургской области

12.2.3 Авторегрессионные модели со скользящими средними в остатках

На практике в целях экономичного описания анализируемого процесса в модель могут быть включены как члены, описывающие авторегрессионные составляющие, так и члены, моделирующие остаток в виде процесса скользящих средних. Такой процесс называется процессом авторегрессии скользящего среднего – ARMA (p,q)⁴³:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \dots - \beta_q \varepsilon_{t-q} \quad (12.21)$$

или

$$y_t - \alpha_0 - \alpha_1 y_{t-1} - \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \dots - \beta_q \varepsilon_{t-q} \quad (12.22)$$

Здесь единственное слагаемое ошибки ε_t , AR – процесса заменяется на процесс MA (q).

Такая модель может интерпретироваться как линейная модель множественной регрессии, в которой в качестве объясняющих переменных выступают прошлые значения самой зависимой переменной, а в качестве регрессионного остатка – скользящие средние из элементов белого шума.

Стационарность процесса ARMA обеспечивается условием $|\alpha| < 1$, а обратимость, в свою очередь гарантируется выполнением условия $|\beta| < 1$.

Одним из наиболее важных этапов построения моделей стац-

⁴³ Дуброва, Т.А. Статистические методы прогнозирования: учеб. пособие для вузов / Т.А. Дуброва - М.: ЮНИТИ-ДАНА, 2003. –206 с.

онарных временных рядов является определение ее порядка. Предварительная оценка производится на основе экономического анализа. Чрезмерное повышение порядка модели может и не повысить ее точность. Одновременно расчет большего числа коэффициентов модели при неизменной выборке снижает достоверность оценки каждого из коэффициентов. В то же время недостаточное число коэффициентов модели не позволяет отразить в должной мере динамику процесса и оценить его дальнейшие изменения.

Для определения порядка процесса модели исследуются такие характеристики, как автокорреляционная функция и частная автокорреляционная функция.

На практике, как правило, используют следующие виды моделей, идентифицировать которые можно с помощью анализа АКФ и ЧАКФ (таблица 12.3)⁴⁴.

ARMA – процессы имеют более сложную структуру по сравнению со схожими по поведению AR – или MA – процессами в чистом виде, но при этом ARMA – процессы характеризуются меньшим количеством параметров, что является одним из их преимуществ.

12.2.4 Прогнозирование ARMA - процессов

При прогнозировании на практике реальные параметры ARMA - процесса α_k и β_j заменяются своими оценками $\hat{\alpha}_k$ и $\hat{\beta}_j$, а случайные шоки ε_t - на остатки $\hat{\varepsilon}_t$, полученные при оценивании модели, или на ошибки предыдущих прогнозов.

Прогнозирование значения y_t на период $(t + h)$ по авторегрес-

⁴⁴ Дуброва, Т.А. Статистические методы прогнозирования: учеб. пособие для вузов / Т.А. Дуброва - М.: ЮНИТИ-ДАНА, 2003. –206 с.

сионной модели производят следующим образом.

Сначала вычисляют значения

$$\hat{y}_{t+1} = \alpha_0 + \alpha_1 y_t + \alpha_2 y_{t-1} + \dots + \alpha_p y_{t-p} + \varepsilon_t. \quad (12.23)$$

Затем в модель

$$\hat{y}_{t+2} = \alpha_0 + \alpha_1 \hat{y}_{t+1} + \alpha_2 y_t + \dots + \alpha_p y_{t-p+1} + \varepsilon_t \quad (12.24)$$

подставляют вычисленное значение \hat{y}_{t+1} и определяют величину \hat{y}_{t+2} и т.д.

Таблица 12.3

Свойства АКФ и ЧАКФ

Функция	ARMA (1,0)	ARMA (2,0)	ARMA (0,1)	ARMA (0,2)	ARMA (1,1)
АКФ	Экспоненциально затухает (монотонно или знакопеременно)	Экспоненциально затухает или имеет форму синусоидальной волны	Выброс (пик) на лаге 1	Выбросы (пики) на лагах 1,2	Экспоненциально затухает от значения $\rho(1)$ (монотонно или знакопеременно)
ЧАКФ	Выброс (пик) на лаге 1	Выбросы (пики) на лагах 1,2	Экспоненциально затухает (монотонно или знакопеременно)	Экспоненциально затухает или имеет форму синусоидальной волны	Экспоненциально затухает от значения $\rho_c(1)$ (монотонно или знакопеременно)

Для модели МА (1) формулы для прогнозирования имеют вид:

$$\begin{aligned} \hat{Y}_{t+1} &= \beta \cdot \varepsilon_t \\ \hat{Y}_{t+h} &= 0 \text{ для } h \geq 2 \end{aligned} \quad (12.25)$$

Для процесса МА (2) формулы для прогнозирования:

$$\begin{aligned} \hat{Y}_{t+1} &= -\beta_1 \cdot \varepsilon_t - \beta_2 \varepsilon_{t-1} \\ \hat{Y}_{t+2} &= -\beta_2 \varepsilon_t \\ \hat{Y}_{t+h} &= 0 \text{ для } h \geq 3 \end{aligned} \quad (12.26)$$

Для модели ARMA (1,1) формулы для прогнозирования имеют вид:

$$\begin{aligned} \hat{Y}_{t+1} &= \alpha_0 + \alpha_1 y_t - \beta \cdot \varepsilon_t \\ \hat{Y}_{t+h} &= \alpha_0 + \alpha_1 y_{t+h-1} \text{ для } h \geq 2 \end{aligned} \quad (12.27)$$

Доверительный интервал прогноза в предположении, что ε_t имеет характеристики белого шума, вычисляется по формуле⁴⁵:

$$\hat{y}_t - t_\alpha \hat{\sigma}_{\varepsilon_t} \leq y_t \leq \hat{y}_t + t_\alpha \hat{\sigma}_{\varepsilon_t}, \quad (12.28)$$

где y_t - истинное значение исследуемого параметра;

\hat{y}_t - предсказываемое значение исследуемого параметра;

$$\hat{\sigma}_{\varepsilon_t} = \frac{\sum_{t=1}^n \varepsilon_t^2}{n-p} - \text{оценка дисперсии случайной величины } \varepsilon_t;$$

⁴⁵ Эконометрика: учебник / И.И. Елисеева [и др.]. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2006. – 576 с.

ε_t - остатки в уравнении авторегрессии;

n- число наблюдений;

p – порядок авторегрессии;

t_α - табличное значение t- статистики Стьюдента.

Рассмотрим пример построения модели ARMA (p,q). Если модель ARMA содержит скользящие средние, то МНК – оценивание, как и в случае с МА- процессами, уже не является возможным. В связи с этим оценивание параметров моделей ARMA в основном проводится по тем же принципам, что и оценивание параметров для МА – процессов, но становится намного сложнее. Например, появляется проблема выбора первоначальных значений y_t из-за наличия регрессоров – лагов зависимой переменной. Наиболее распространенными методами оценивания параметров являются: нелинейный МНК и метод максимального правдоподобия.

Оценим параметры модели ARMA по отклонениям от линейного тренда, используя ППП STATISTICA. Зададим значения параметров как показано на рисунке 12.18.

Таким образом, нами получена модель (рисунок 12.19):

$$y_t - \tilde{y}_t = -0,7775 \cdot (y_{t-1} - \tilde{y}_{t-1}) + \varepsilon_t - 0,9618 \varepsilon_{t-1}.$$

Подставим в полученную модель вместо \tilde{y} уравнение тренда и после раскрытия скобок и приведения подобных слагаемых перейдем к модели вида:

$$y_t = 1046,15 - 61,324 \cdot t - 0,7775 \cdot y_{t-1} + \varepsilon_t - 0,9618 \cdot \varepsilon_{t-1}.$$

Прогнозные значения отклонений от тренда представлены на рисунке 12.20, их графическое изображение на рисунке 12.21.

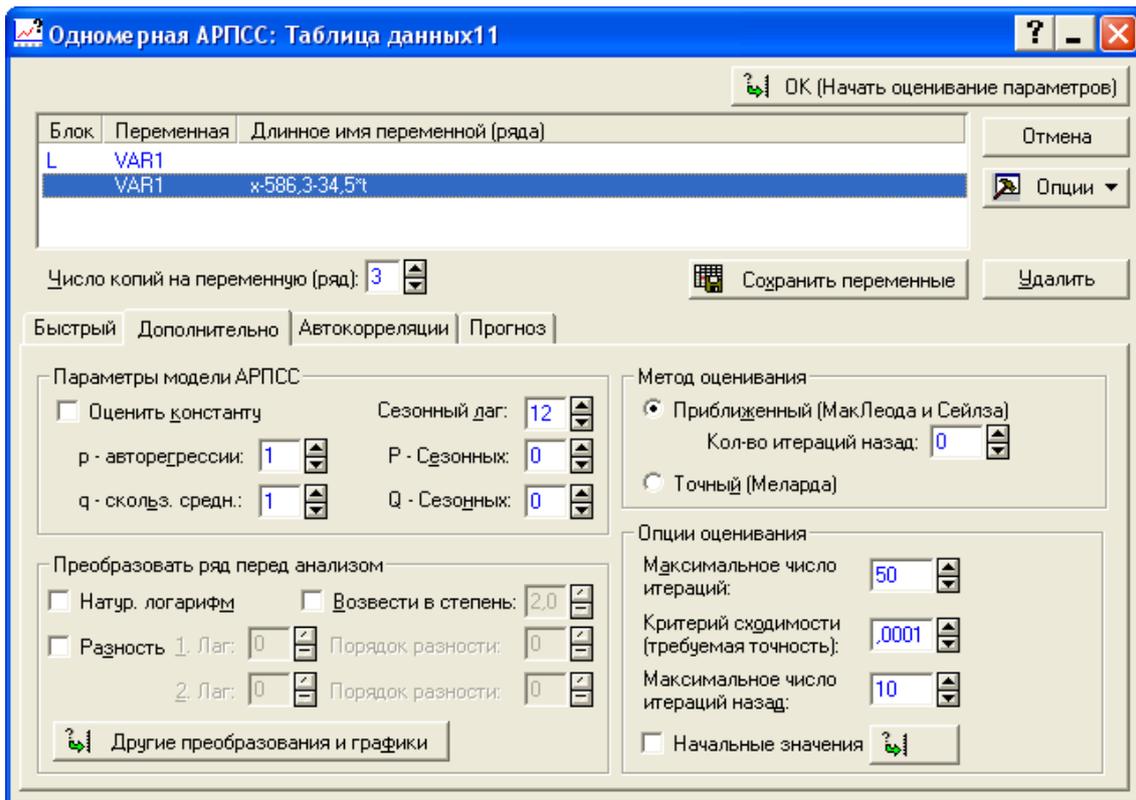


Рис. 12.18 Построение ARMA – модели по отклонениям от тренда

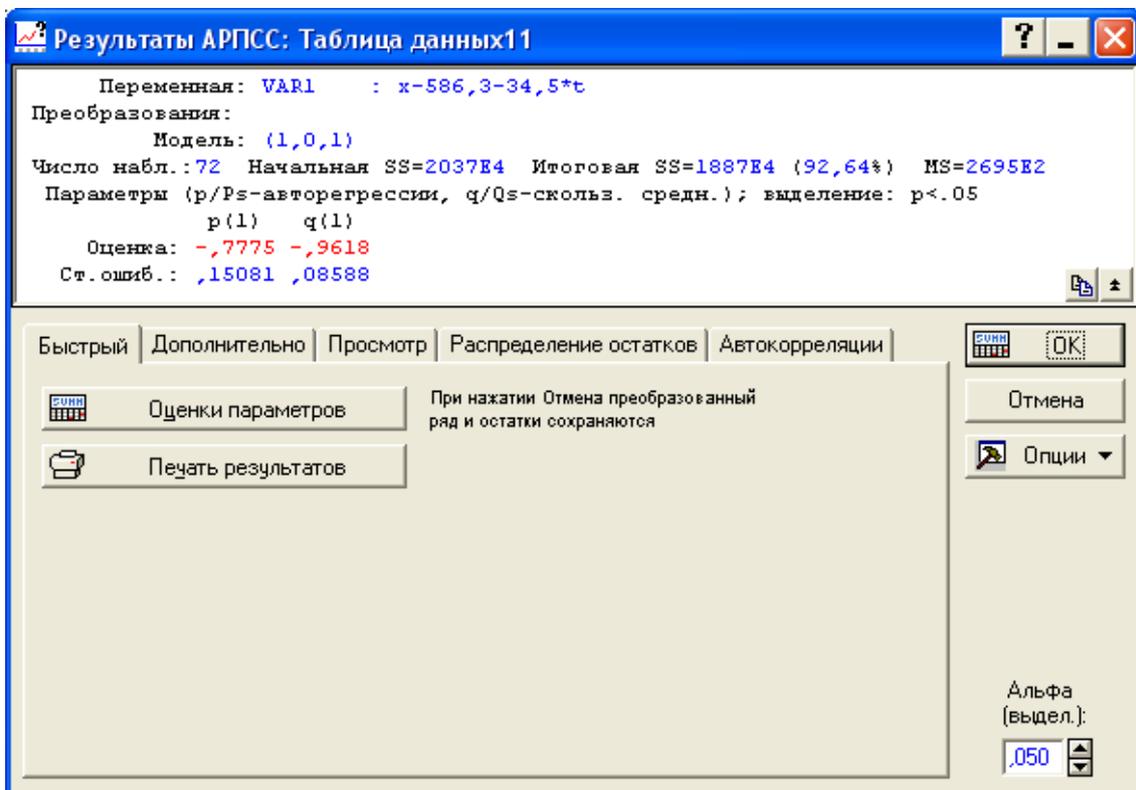


Рис. 12.19 Результаты оценивания модели ARMA

Прогнозы; Модель:(1,0,1) Сезонный лаг: 12 (Таблица данных11)				
Исход.:VAR1 : $x-586,3-34,5^*t$				
Начало исходных: 1 Конец исходн.: 72				
Набл. N	Прогноз	Нижний 90,0000%	Верхний 90,0000%	Ст.ошиб.
73	218,670	-646,76	1084,102	519,1820
74	-170,015	-1050,02	709,995	527,9275
75	132,186	-756,52	1020,892	533,1445
76	-102,774	-996,70	791,149	536,2736
77	79,906	-817,15	976,967	538,1563
78	-62,127	-961,08	836,826	539,2913
79	48,303	-851,79	948,397	539,9761
80	-37,555	-938,34	863,228	540,3897
81	29,199	-872,00	930,399	540,6396
82	-22,702	-924,15	878,750	540,7906

Рис. 12.20 Прогноз по модели ARMA(1,1)

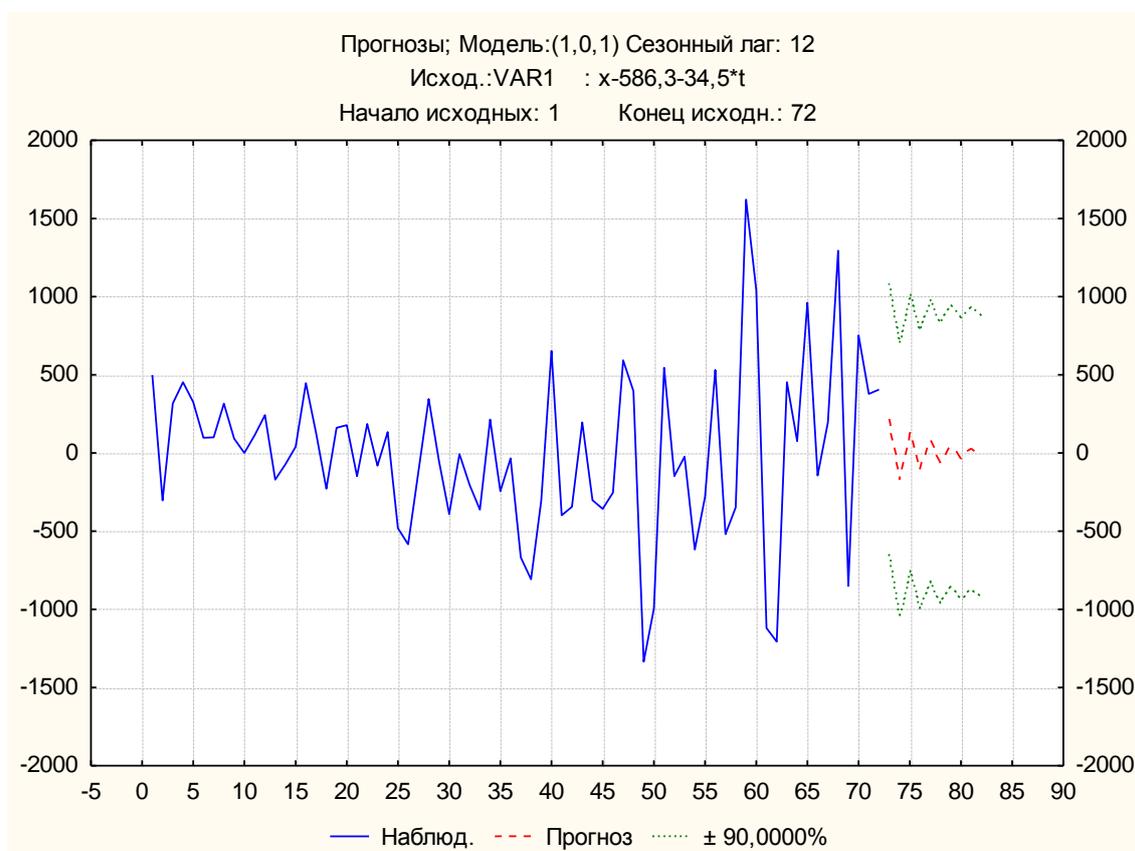


Рис. 12.21 Графическое изображение прогноза по модели ARMA(1,1)

Модель ARMA(1,1) по первым разностям представляет собой процесс ARIMA(1,1,1). К построению которой мы и переходим.

12.3 Методология Бокса – Дженкинса

Экономические временные ряды за редким исключением нестационарны. Нестационарность чаще всего проявляется в наличии зависящей от времени неслучайной составляющей $f(t)$. Для описания таких рядов используется модель авторегрессии – проинтегрированного скользящего среднего ARIMA (p,d,q) (модель Бокса-Дженкинса).

Модель ARIMA используется для описания временных рядов, обладающих свойствами:

1) ряд включает аддитивно составляющую $f(t)$, имеющую вид алгебраического полинома;

2) ряд, получившийся после применения к нему процедур последовательных разностей, может быть описан моделью ARMA (p,q).

Пусть X_t - нестационарный процесс со стационарными разностями d - го порядка, т.е. $Y_t = \Delta^d X_t$ - стационарный процесс, а $\Delta^{d-1} X_t$ - нестационарный. Это означает, что X_t интегрируем d - го порядка.

Если Y_t - процесс ARMA (p,q), т.е.

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \dots - \beta_q \varepsilon_{t-q}, \quad (12.29)$$

тогда X_t называется процессом ARIMA (p,d,q). Часто среднее или свободный член приравнивается к нулю (опускается)⁴⁶.

⁴⁶ Эконометрика: учебник / И.И. Елисеева [и др.]. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2006. – 576 с.

Построение модели ARIMA по реализации случайного процесса Бокс и Дженкинс предложили разбить на несколько этапов:

1) установить порядок интеграции d , т.е. добиться стационарности ряда, взяв достаточное количество последовательных разностей. Для определения значения d может быть применен эвристический критерий. Использование данного критерия основано на определении оценки

$$\hat{\sigma}^2(k) = \frac{1}{n-k} \frac{\sum_{t=1}^{n-k} (\Delta^k y_t)^2}{C_{2k}^k}, \quad (12.30)$$

где $\Delta^k y_t$ - последовательные разности исходного ряда y_1, y_2, \dots, y_n ,

k – порядок разностей, $k = 1, 2, \dots$

Начиная с некоторого значения $k = k_0$ величина $\hat{\sigma}^2(k)$ стабилизируется, оставаясь примерно на одном и том же уровне при росте k . Тогда порядок разности (d) следует принять равным k_0 .

Также о том, что необходимая для стационарности ряда степень разности достигнут, будет свидетельствовать быстрое затухание АКФ.

2) для полученного стационарного временного ряда строятся АКФ и ЧАКФ. Исследуя характер их поведения, выдвигаются гипотезы о значениях параметров p и q , т.е. подбирается модель ARMA(p, q). На данном этапе формируется базовый набор моделей, включающий 1, 2 или даже больше количество моделей.

3) оценивание коэффициентов $\alpha_0, \alpha_1, \dots, \alpha_p, \beta_1, \beta_2, \dots, \beta_q$ для всех отобранных на 2 этапе моделей. Для этого можно использовать сле-

дующие методы:

- традиционный МНК;
- метод максимального правдоподобия;
- нелинейный МНК;
- алгоритм Марквардта.

Все эти оценки при больших объемах выборок асимптотически эквивалентны.

4) выбор наиболее подходящей модели среди оцененных:

4.1 проверка адекватности модели на основе анализа остатков (у адекватной модели остатки должны быть похожи на белый шум). Для этого проводится проверка значимости коэффициентов автокорреляции используя следующие подходы:

– если выборочный коэффициент автокорреляции r_k выходит за интервал $\pm \frac{t_\alpha}{\sqrt{n}}$, то гипотеза H_0 о равенстве нулю коэффициента автокорреляции ρ_k отвергается;

– проверка равенства нулю сразу τ первых значений АКФ на основе Q – статистики Бокса – Пирса:

$$Q = n \sum_{k=1}^{\tau} r_k^2 \quad (12.31)$$

или тест Бокса – Льюнга:

$$\tilde{Q} = n(n+2) \sum_{k=1}^{\tau} \frac{r_k^2}{n-k}. \quad (12.32)$$

Если Q (или \tilde{Q}) $> \chi_{табл}^2$ с $\nu = \tau - p - q$ степенями свободы, то как

группа первые τ коэффициентов автокорреляции значимы (рекомендуется рассматривать $\tau \approx n/4$);

4.2 отбор оптимальной модели по наивысшему качеству с меньшим числом параметров с использованием информационного критерия Акайка и Шварца:

$$- \text{информационный критерий Акайка: } AIC = \frac{p+q}{n} + \ln \left(\frac{\sum_{t=1}^n e_t^2}{n} \right); \quad (12.33)$$

$$- \text{критерий Шварца: } SIK = \frac{(p+q) \ln n}{n} + \ln \left(\frac{\sum_{t=1}^n e_t^2}{n} \right). \quad (12.34)$$

Предпочтение следует отдать модели с меньшим значением критерия.

Прогнозирование ARIMA – процессов Y_t может быть представлено в виде двухшаговой процедуры⁴⁷:

1) экстраполируется стационарный ARMA – процесс;

2) вместо взятия разностей провести обратную операцию интегрируемости, т.е. суммирования спрогнозированных на шаге 1 приращений $\hat{Y}_{T(+h)} = \Delta^d \hat{X}_{T(+h)}$, чтобы получить сначала $\Delta^{d-1} \hat{X}_{T(+h)}$, а затем по аналогии $\Delta^{d-2} \hat{X}_{T(+h)}$ и, наконец, $\hat{X}_{T(+h)}$. Оценка дисперсии ошибки прогноза, а следовательно, и ширины доверительного интервала прогноза проводится аналогичным образом – повторным суммированием дисперсий ошибок прогноза ARMA – процесса X_t .

⁴⁷ Эконометрика: учебник / И.И. Елисеева [и др.]. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2006. – 576 с.

Другим возможным вариантом является построение индивидуальной одношаговой формулы для получения прогноза.

С этой целью в уравнение вместо Y_t подставляют разности

$$\Delta^d X_T = (1 - L)^d X_T. \quad (12.35)$$

Решив полученное уравнение относительно X_t , получим формулу, которая может быть экстраполирована для $t = T + h$ и таким образом преобразована в формулу для прогнозирования на h шагов вперед величин $\hat{X}_{T(+h)}$ с началом отсчета в момент времени T .

Рассмотрим пример построения модели ARIMA (p,d,q). Оценим параметры модели в ППП STATISTICA, задав значение параметров, как показано на рисунке 12.22.

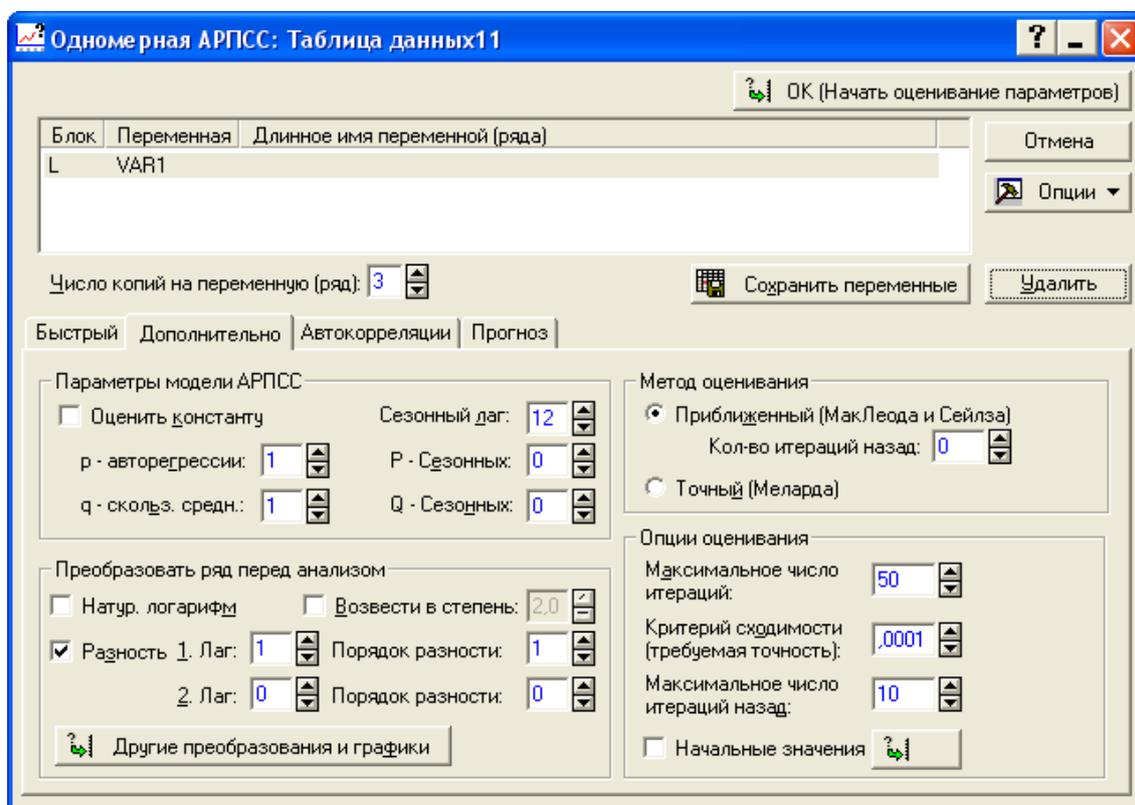


Рис. 12.22 Построение ARIMA– модели

Результаты оценивания представлены на рисунке 12.23.

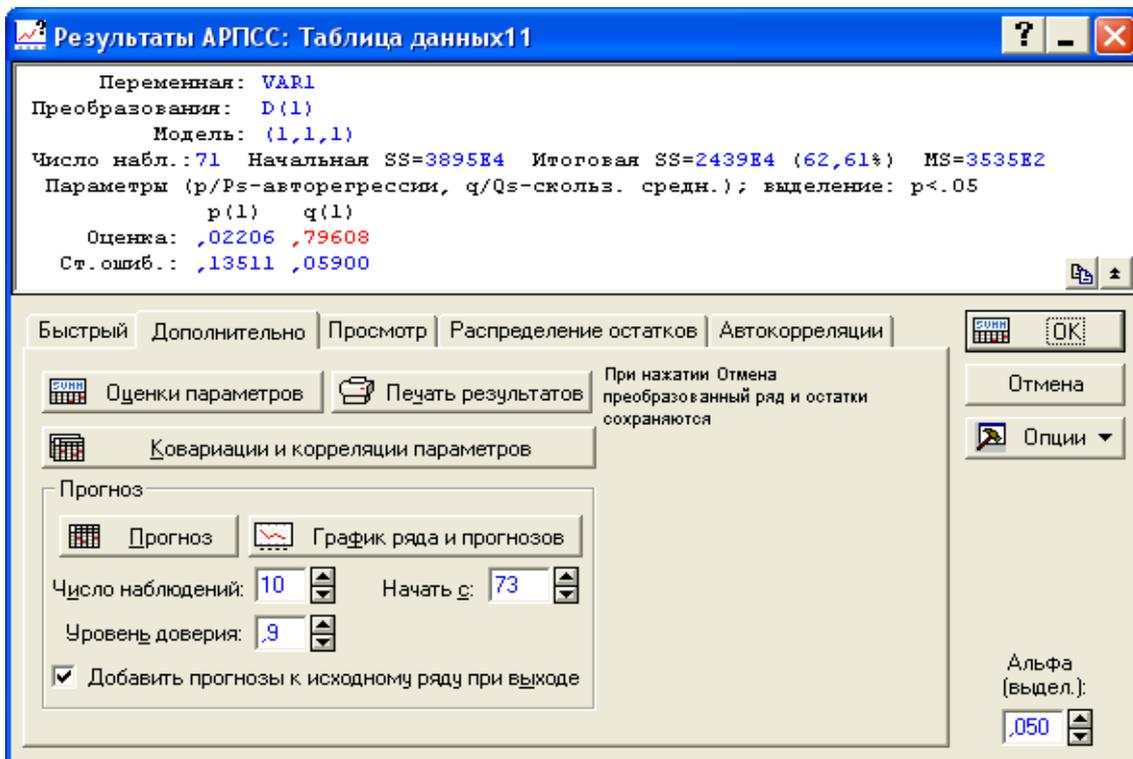


Рис. 12.23 Результаты оценивания модели ARIMA

Таким образом, нами получена модель: $\Delta_t^1 = 0,02206 \cdot \Delta_{t-1}^1 + \varepsilon_t - 0,79608 \cdot \varepsilon_{t-1}$. Подставив в модель вместо $\Delta_t^1 = y_t - y_{t-1}$ и $\Delta_{t-1}^1 = y_{t-1} - y_{t-2}$, раскрыв скобки и приведя подобные слагаемые получим модель: $\tilde{y}_t = 1,02206 y_{t-1} - 0,02206 y_{t-2} + \varepsilon_t - 0,79608 \varepsilon_{t-1}$. Рассчитаем по ней теоретические значения доходов бюджета (таблица 12.4).

Как видно на рисунке 12.24, расчетные значения «запаздывают» на 1 месяц по сравнению с фактическими данными. При этом средняя относительная ошибка аппроксимации составила 29,46 %, что свидетельствует об удовлетворительной точности модели.

Прогноз для модели ARIMA(1,1,1) на 1 шаг вперед определяется по формуле:

$$\hat{y}_{t+1} = (1 + \alpha) \cdot y_t - \alpha \cdot y_{t-1} - \beta \varepsilon_t,$$

Таблица 12.4

Расчет прогноза по модели ARIMA(1,1,1)

Период	Доход, млн.р.	t	$\Delta^{(1)} = y_t - y_{t-1}$	$0,02206 \cdot \Delta^{(1)}$	$\varepsilon_t = \Delta^{(1)} - 0,02206 \cdot \Delta^{(1)}$	\tilde{y}_t	
1	2	3	4	5	6	7	
2001	январь	1119,3	1				
	февраль	352,2	2	-767,1	-16,9	14,8	
	март	1006,9	3	654,7	14,4	3,5	326,9
	апрель	1177,8	4	170,9	3,8	-2,1	1016,5
	май	1084,4	5	-93,4	-2,1	-4,2	1179,1
	июнь	891,4	6	-193	-4,3	0,9	1086,6
	июль	928,2	7	36,8	0,8	5,5	891,9
	август	1178,4	8	250,2	5,5	-4,3	920,3
	сентябрь	989,4	9	-189	-4,2	-1,2	1186,2
	октябрь	932,2	10	-57,2	-1,3	3,3	989,5
	ноябрь	1080,4	11	148,2	3,3	3,5	931,8
	декабрь	1243,5	12	163,1	3,6	-8,4	1072,4

Продолжение таблицы 12.4

	1	2	3	4	5	6	7
2002	январь	865,5	13	-378	-8,3	3,1	1256,9
	февраль	998,4	14	132,9	2,9	3,2	857,9
	март	1145,1	15	146,7	3,2	9,6	1008,5
	апрель	1585,6	16	440,5	9,7	-6,5	1134,2
	май	1301	17	-284,6	-6,3	-6,9	1593,5
	июнь	980,3	18	-320,7	-7,1	9,5	1309,7
	июль	1403,5	19	423,2	9,3	0,9	966,6
	август	1455,7	20	52,2	1,2	-6,5	1405,6
	сентябрь	1163,5	21	-292,2	-6,4	8,3	1470,3
	октябрь	1532	22	368,5	8,1	-5,3	1145,2
	ноябрь	1299,9	23	-232,1	-5,1	5,6	1550,0
	декабрь	1549,1	24	249,2	5,5	-12,9	1277,4
2003	январь	968,8	25	-580,3	-12,8	-1,2	1563,6
	февраль	900	26	-68,8	-1,5	11,1	968,1
	март	1402	27	502	11,1	10,7	900,4

Продолжение таблицы 12.4

1	2	3	4	5	6	7	
	апрель	1898,8	28	496,8	11,0	-8,2	1396,4
	май	1538,8	29	-360	-7,9	-6,6	1909,7
	июнь	1232,7	30	-306,1	-6,8	9,4	1545,5
	июль	1650,1	31	417,4	9,2	-3,8	1214,7
	август	1486,9	32	-163,2	-3,6	-2,6	1659,7
	сентябрь	1364,3	33	-122,6	-2,7	13,5	1498,9
	октябрь	1974,6	34	610,3	13,5	-9,6	1341,2
	ноябрь	1551,1	35	-423,5	-9,3	5,6	2001,3
	декабрь	1795,6	36	244,5	5,4	-13,3	1524,0
2004	январь	1196,8	37	-598,8	-13,2	-2,0	1809,6
	февраль	1091,1	38	-105,7	-2,3	11,9	1197,1
	март	1629,4	39	538,3	11,9	21,6	1100,9
	апрель	2620,2	40	990,8	21,9	-22,9	1601,2
	май	1603,7	41	-1016,5	-22,4	2,5	2662,8
	июнь	1692,8	42	89,1	2,0	12,6	1592,0

Продолжение таблицы 12.4

1	2	3	4	5	6	7	
	июль	2267,5	43	574,7	12,7	-10,5	1674,2
	август	1804,6	44	-462,9	-10,2	-0,3	2288,3
	сентябрь	1782,8	45	-21,8	-0,5	3,1	1797,7
	октябрь	1921	46	138,2	3,0	19,4	1799,3
	ноябрь	2802,3	47	881,3	19,4	-4,0	1904,6
	декабрь	2639,6	48	-162,7	-3,6	-37,3	2787,6
2005	январь	944,1	49	-1695,5	-37,4	9,1	2674,8
	февраль	1317,3	50	373,2	8,2	34,6	934,1
	март	2893,2	51	1575,9	34,8	-15,3	1282,7
	апрель	2234,3	52	-658,9	-14,5	3,8	2944,0
	май	2393,7	53	159,4	3,5	-12,4	2204,3
	июнь	1834,2	54	-559,5	-12,3	8,5	2415,6
	июль	2205,4	55	371,2	8,2	18,5	1833,6
	август	3051,7	56	846,3	18,7	-22,8	2176,0
	сентябрь	2035,7	57	-1016	-22,4	5,0	3093,6

Продолжение таблицы 12.4

	1	2	3	4	5	6	7
	октябрь	2241,3	58	205,6	4,5	44,1	2053,4
	ноябрь	4245,3	59	2004	44,2	-13,0	2197,7
	декабрь	3699,7	60	-545,6	-12,0	-46,6	4253,2
2006	январь	1573	61	-2126,7	-46,9	-0,1	3724,7
	февраль	1521,5	62	-51,5	-1,1	37,4	1563,6
	март	3215,2	63	1693,7	37,4	-8,4	1482,2
	апрель	2872,5	64	-342,7	-7,6	20,5	3279,7
	май	3792,4	65	919,9	20,3	-24,1	2824,6
	июнь	2721,7	66	-1070,7	-23,6	8,8	3840,7
	июль	3097,2	67	375,5	8,3	24,8	2715,9
	август	4229,2	68	1132	25,0	-47,1	3038,7
	сентябрь	2119,6	69	-2109,6	-46,5	37,1	4328,8
	октябрь	3756,5	70	1636,9	36,1	-8,3	2035,2
	ноябрь	3416,1	71	-340,4	-7,5	1,5	3800,8
	декабрь	3478,7	72	62,6	1,4	52,0	3459,4

Продолжение таблицы 12.4

1	2	3	4	5	6	7
Итого	132968,7	-	2359,4	52,0	119,9	128472,4
В среднем	1846,7875	-	32,769444	0,7	1,7	1835,3

т.е. прогноз на январь 2007 года составит:

$$(1+0,02206) \cdot 3478,7 - 0,02206 \cdot 3416,1 - 0,79608 \cdot (52,0) = 3438,7 \text{ млн. р.}$$

На 2 шага вперед: $\hat{y}_{t+2} = (1 + \alpha) \cdot y_{t+1} - \alpha \cdot y_t$, тогда прогноз на февраль 2007 года составит:

$$(1+0,02206) \cdot 3438,7 - 0,02206 \cdot 3478,7 = 3437,8 \text{ млн.р.}$$

Интервальный прогноз при среднем квадратическом отклонении 763,95 и статистике Стьюдента на 5% уровне значимости для 68 степеней свободы составившей 1,995 для января 2007 года будет находится в границах от 1914,7 млн.р. до 4202,7 млн.р., а в феврале 2007 году доходы бюджета с вероятностью 95 % составит от 1913,8 млн.р. до 4961,8 млн.р.

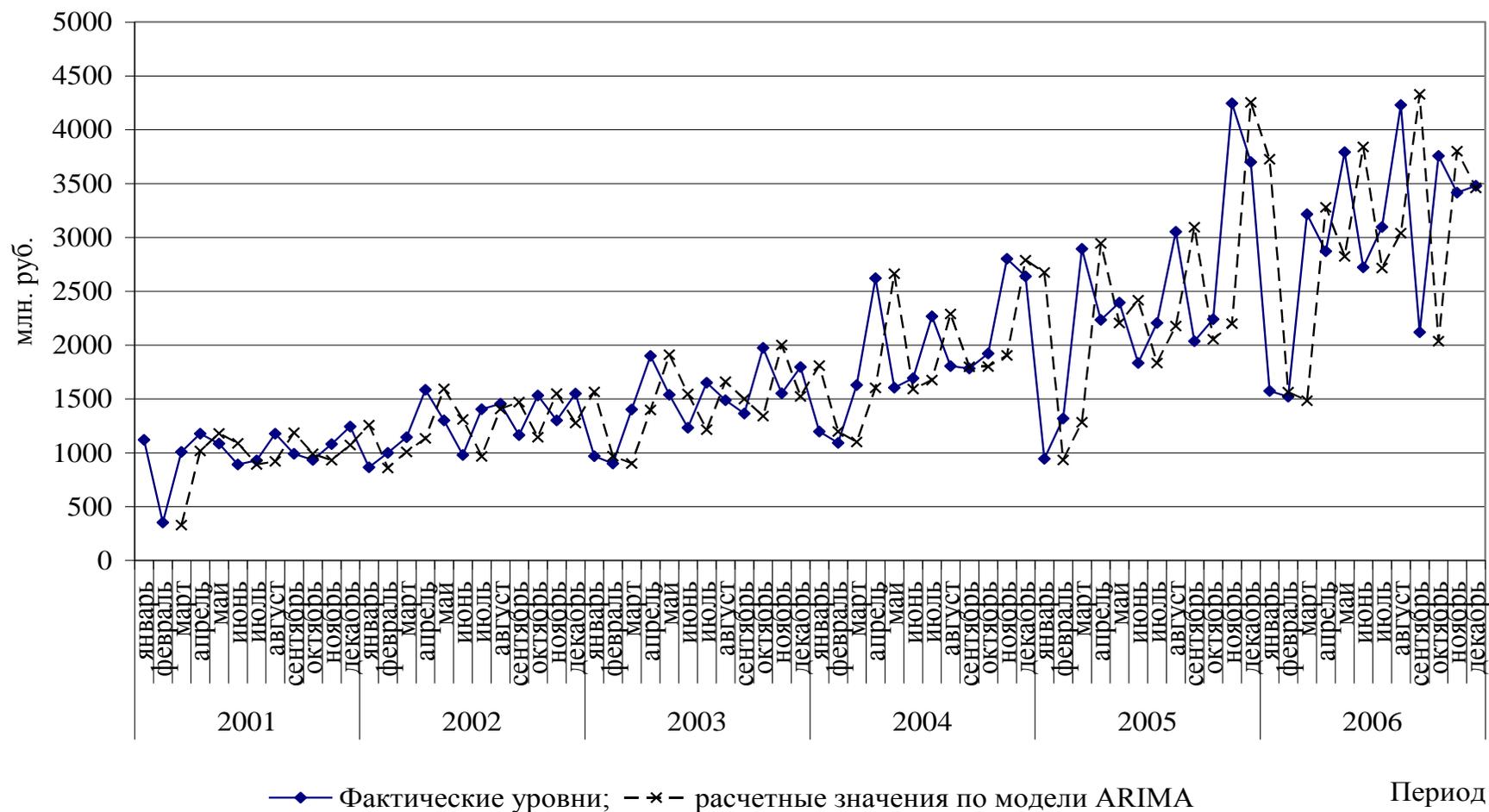


Рис. 12.24 Динамика доходов бюджета Оренбургской области

ЛИТЕРАТУРА

1. *Андерсон Т.* Статистический анализ временных рядов: Пер. с англ. – М.: Мир, 1976.
2. *Афанасьев В.Н.* Статистическое обеспечение проблемы устойчивости сельскохозяйственного производства. – М.: Финансы и статистика, 1996.
3. *Афанасьев В.Н., Юзбашев М.М.* Коэффициент корреляции рангов как показатель устойчивости динамики // Вестник статистики. – 1983. – № 11.
4. *Бобров С.П.* Конъюнктурные кривые 1922– 1923 гг. //Экономическое обозрение. – 1924. – № 5 – 7.
5. *Вайну Я.Я.Ф.* Корреляция рядов динамики. – М.: Финансы и статистика, 1982.
6. *Дубров А.М., Мхитарян В.С, Трошин Л.И.* Многомерные статистические методы. – М.: Финансы и статистика, 1999.
7. *Казинец Л.С.* Темпы роста и абсолютные приросты. – М.: Статистика, 1975.
8. *Калинина М.С.* Статистические методы изучения динамики урожайности. – Л.: ЛСХИ, 1969.
9. *Кендэл М.* Временные ряды: Пер. с англ. – М.: Финансы и статистика, 1981.
10. *Крастинь О.П.* Изучение статистических зависимостей по многолетним данным. – М.: Финансы и статистика, 1981.
11. *Лапин Г.Ф.* Биометрия. – М.: Высшая школа, 1968.
12. *Лукашин Ю.П.* Адаптивные методы прогнозирования: учеб. пособие / Ю.П.Лукашин. – М.: Финансы и статистика, 2003. - 416 с.:ил. - . ISBN 5-279-02740-5

13. *Попова И.* Агрегирование трендов объемных показателей и их колеблемости по совокупности предприятий // Экономико-математические и статистические методы в прогнозировании и управлении сельскохозяйственного производства. – Сб. трудов. – Спб.: Спб. ГАУ, 1993.
14. Статистический анализ экономических временных рядов и прогнозирование: ЦЭМИ АН СССР. – М.: Наука, 1973.
15. *Френкель А.А.* Математические методы анализа динамики и прогнозирования производительности труда. – М.: Экономика, 1972.
16. *Хенан Э.* Многомерные временные ряды: Пер. с англ. – М.: Мир, 1974.
17. *Четвериков И.С.* Статистические и стохастические исследования. – М.: Госстатиздат, 1963.
18. *Четыркин Е.М.* Статистические методы прогнозирования. – М.: Статистика, 1977.
19. Эконометрика: учебник / И.И. Елисеева, [и др.]; – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2006. – 576 с. -ISBN 5-279-02786-3
20. *Юзбашев М.М.* Экспериментальная проверка закономерности распределения отклонений урожайности отдельных лет от тренда по их величине // Науч. труды ЛСХИ. – Т. 322. –Л. –Пушкин. – 1981.
21. *Юзбашев М.М., Манелля А.И.* Статистический анализ тенденций и колеблемости. – М.: Финансы и статистика, 1983.
22. *Юл Д.Э., Кендэл М.Д.* Теория статистики: Пер. с англ. – М.: Госстатиздат, 1960.
23. *Ястремский Б.С.* Переменная корреляция // Вестник статистики. – 1924. – Кн. XVII.
24. *Blanford D. Offut S.* A Review of Empirical Techniques for the Analysis of Commodity Instability. – USsl. – 1983.

ПРИЛОЖЕНИЯ

Приложение 1

Агрегирование фаз – фильтрация маломощных флуктуаций временного ряда добычи нефти

Флуктуации					Знаки и мощность фаз на итерации					
ГО- ДЫ	МЕСЯЦЫ	y_t , ТЫС. барр./сут.	\hat{y}_t , ТЫС. барр./сут.	$D_y = y_t - \hat{y}_t$	перво- началь- но	1	2	3	4	5
1	2	3	4	5	6	7	8	9	10	11
2003	январь	1180,311	1148,323	31,988	155,714 (9)	155,714 (9)	155,714 (9)	155,714 (9)	-310,864 (28)	-310,86 (28)
	фев- раль	1192,928	1172,592	20,336	-294,053 (9)	-466,578 (19)	-466,578 (19)	-466,58 (19)	565,309 (15)	290,82 6 (26)
	март	1214,628	1196,107	18,521	14,942 (4)	409,255 (8)	565,309 (15)	565,309 (15)	-274,483 (11)	
	апрель	1234,588	1218,868	15,720	-157,583 (6)	-39,099 (2)	-321,716 (8)	-274,48 (11)		
	май	1262,243	1240,875	21,368	409,255 (8)	116,955 (5)	47,233 (3)			
	июнь	1285,765	1262,128	23,637	-39,099 (2)	-321,716 (8)				
	июль	1299,602	1282,627	16,975	116,955 (5)	47,233 (3)				
	август	1308,539	1302,372	6,167	-321,716 (8)					

Продолжение приложения 1

1	2	3	4	5	6	7	8	9	10	11
	сен- тябрь	1322,365	1321,363	1,002	47,233 (3)					
	ок- тябрь	1328,043	1339,600	-11,557						
	ноябрь	1334,828	1357,083	-22,255						
	де- кабрь	1338,495	1373,812	-35,317						
2004	январь	1343,207	1389,787	-46,580						
	фев- раль	1354,295	1405,008	-50,713						
	март	1373,036	1419,475	-46,439						
	апрель	1392,472	1433,188	-40,716						
	май	1415,498	1446,147	-30,649						
	июнь	1448,525	1458,352	-9,827						
	июль	1474,336	1469,803	4,533						
	август	1484,841	1480,500	4,341						
	сен- тябрь	1493,97	1490,443	3,527						
	ок- тябрь	1502,173	1499,632	2,541						
ноябрь	1485,787	1508,067	-22,280							

Продолжение приложения 1

1	2	3	4	5	6	7	8	9	10	11
	де-кабрь	1475,196	1515,748	-40,552						
2005	январь	1493,724	1522,675	-28,951						
	фев-раль	1509,043	1528,848	-19,805						
	март	1505,344	1534,267	-28,923						
	апрель	1521,86	1538,932	-17,072						
	май	1561,392	1542,843	18,549						
	июнь	1576,003	1546,000	30,003						
	июль	1588,495	1548,403	40,092						
	август	1611,704	1550,052	61,652						
	сен-тябрь	1624,743	1550,947	73,796						
	ок-тябрь	1624,775	1551,088	73,687						
	ноябрь	1611,273	1550,475	60,798						
	де-кабрь	1599,786	1549,108	50,678						
2006	январь	1536	1546,987	-10,987						
	фев-раль	1516	1544,112	-28,112						
	март	1545	1540,483	4,517						

Продолжение приложения 1

1	2	3	4	5	6	7	8	9	10	11
	апрель	1556	1536,100	19,900						
	май	1557	1530,963	26,037						
	июнь	1556	1525,072	30,928						
	июль	1554	1518,427	35,573						
	август	1469	1511,028	-42,028						
	сентябрь	1428	1502,875	-74,875						
	октябрь	1433	1493,968	-60,968						
	ноябрь	1432	1484,307	-52,307						
	декабрь	1428	1473,892	-45,892						
2007	январь	1438	1462,723	-24,723						
	февраль	1439	1450,800	-11,800						
	март	1429	1438,123	-9,123						
	апрель	1430	1424,692	5,308						
	май	1429	1410,507	18,493						
	июнь	1419	1395,568	23,432						

**Многokратное скользящее выравнивание доходов бюджета
Оренбургской области**

Период	Доход, млн.руб.	t	Тренд, Y_i	$Y_i - y$	$(Y_i - y)^2$	
1	2	3	4	5	6	
2001	январь	1119,3	-35,5	678,9	440,45	193996,2
	февраль	352,2	-34,5	711,8	-359,55	129276,2
	март	1006,9	-33,5	744,7	262,25	68775,1
	апрель	1177,8	-32,5	777,6	400,25	160200,1
	май	1084,4	-31,5	810,5	273,95	75048,6
	июнь	891,4	-30,5	843,4	48,05	2308,8
	июль	928,2	-29,5	876,3	51,95	2698,8
	август	1178,4	-28,5	909,2	269,25	72495,6
	сентябрь	989,4	-27,5	942,1	47,35	2242,0
	октябрь	932,2	-26,5	975,0	-42,75	1827,6
	ноябрь	1080,4	-25,5	1007,9	72,55	5263,5
	декабрь	1243,5	-24,5	1040,8	202,75	41107,6
2002	январь	865,5	-23,5	1073,7	-208,15	43326,4
	февраль	998,4	-22,5	1106,6	-108,15	11696,4
	март	1145,1	-21,5	1139,5	5,65	31,9
	апрель	1585,6	-20,5	1172,4	413,25	170775,6
	май	1301	-19,5	1205,3	95,75	9168,1
	июнь	980,3	-18,5	1238,2	-257,85	66486,6
	июль	1403,5	-17,5	1271,1	132,45	17543,0
	август	1455,7	-16,5	1304,0	151,75	23028,1
	сентябрь	1163,5	-15,5	1336,9	-173,35	30050,2
	октябрь	1532	-14,5	1369,8	162,25	26325,1
	ноябрь	1299,9	-13,5	1402,7	-102,75	10557,6
	декабрь	1549,1	-12,5	1435,6	113,55	12893,6
2003	январь	968,8	-11,5	1468,5	-499,65	249650,1
	февраль	900	-10,5	1501,4	-601,35	361621,8
	март	1402	-9,5	1534,3	-132,25	17490,1
	апрель	1898,8	-8,5	1567,2	331,65	109991,7
	май	1538,8	-7,5	1600,1	-61,25	3751,6

Продолжение приложения 2

	1	2	3	4	5	6
	июнь	1232,7	-6,5	1633,0	-400,25	160200,1
	июль	1650,1	-5,5	1665,9	-15,75	248,1
	август	1486,9	-4,5	1698,8	-211,85	44880,4
	сентябрь	1364,3	-3,5	1731,7	-367,35	134946,0
	октябрь	1974,6	-2,5	1764,6	210,05	44121,0
	ноябрь	1551,1	-1,5	1797,5	-246,35	60688,3
	декабрь	1795,6	-0,5	1830,4	-34,75	1207,6
2004	январь	1196,8	0,5	1863,3	-666,45	444155,6
	февраль	1091,1	1,5	1896,2	-805,05	648105,5
	март	1629,4	2,5	1929,1	-299,65	89790,1
	апрель	2620,2	3,5	1962,0	658,25	433293,1
	май	1603,7	4,5	1994,9	-391,15	152998,3
	июнь	1692,8	5,5	2027,8	-334,95	112191,5
	июль	2267,5	6,5	2060,7	206,85	42786,9
	август	1804,6	7,5	2093,6	-288,95	83492,1
	сентябрь	1782,8	8,5	2126,5	-343,65	118095,3
	октябрь	1921	9,5	2159,4	-238,35	56810,7
	ноябрь	2802,3	10,5	2192,3	610,05	372161,0
	декабрь	2639,6	11,5	2225,2	414,45	171768,8
2005	январь	944,1	12,5	2258,1	-1314	1726464,6
	февраль	1317,3	13,5	2291,0	-973,65	947994,3
	март	2893,2	14,5	2323,9	569,35	324159,4
	апрель	2234,3	15,5	2356,8	-122,45	14994,0
	май	2393,7	16,5	2389,7	4,05	16,4
	июнь	1834,2	17,5	2422,6	-588,35	346155,7
	июль	2205,4	18,5	2455,5	-250,05	62525,0
	август	3051,7	19,5	2488,4	563,35	317363,2
	сентябрь	2035,7	20,5	2521,3	-485,55	235758,8
	октябрь	2241,3	21,5	2554,2	-312,85	97875,1
	ноябрь	4245,3	22,5	2587,1	1658,25	2749793,1
	декабрь	3699,7	23,5	2620,0	1079,75	1165860,1

Продолжение приложения 2

	1	2	3	4	5	6
2006	январь	1573	24,5	2652,9	-1079,9	1166076,0
	февраль	1521,5	25,5	2685,8	-1164,3	1355478,1
	март	3215,2	26,5	2718,7	496,55	246561,9
	апрель	2872,5	27,5	2751,6	120,95	14628,9
	май	3792,4	28,5	2784,5	1007,95	1015963,2
	июнь	2721,7	29,5	2817,4	-95,65	9148,9
	июль	3097,2	30,5	2850,3	246,95	60984,3
	август	4229,2	31,5	2883,2	1346,05	1811850,6
	сентябрь	2119,6	32,5	2916,1	-796,45	634332,6
	октябрь	3756,5	33,5	2949,0	807,55	652137,0
	ноябрь	3416,1	34,5	2981,9	434,25	188573,1
	декабрь	3478,7	35,5	3014,8	463,95	215249,6
		132968,7		132970	-0,9	20451512,2

**Значение t -критерия Стьюдента при уровне значимости 0,10;
0,05; 0,01**

Число степеней свободы $d.f.$	P			$d.f.$	P		
	0,10	0,05	0,01		0,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7633
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

Значение F–критерия Фишера при уровне значимости 0,05

$d.f_2$	$d.f_1$																
	1	2	3	4	5	6	7	8	9	10	11	12	14	15	20	30	∞
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248	250	254
2	18,5	19,0	19,1	19,2	19,3	19,3	19,3	19,3	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,5
3	10,1	9,55	9,28	9,19	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74	8,71	8,69	8,66	8,62	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91	5,87	5,84	5,80	5,74	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68	4,64	4,60	4,56	4,50	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,96	3,92	3,87	3,81	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57	3,52	3,49	3,44	3,38	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28	3,23	3,20	3,15	3,08	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07	3,02	2,98	2,93	2,86	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91	2,86	2,82	2,77	2,70	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79	2,74	2,70	2,65	2,57	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69	2,64	2,60	2,54	2,46	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60	2,55	2,51	2,46	2,38	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53	2,48	2,44	2,39	2,31	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48	2,43	2,39	2,33	2,25	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42	2,37	2,33	2,28	2,20	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38	2,33	2,29	2,23	2,15	1,96

Продолжение Приложения 4

$d.f_2$	$d.f_1$																
	1	2	3	4	5	6	7	8	9	10	11	12	14	15	20	30	∞
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,29	2,25	2,19	2,11	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	2,34	2,31	2,26	2,21	2,15	2,07	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,31	2,28	2,23	2,18	2,12	2,04	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	2,20	2,15	2,09	2,00	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30	2,26	2,23	2,18	2,13	2,07	1,98	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32	2,28	2,24	2,20	2,14	2,10	2,04	1,96	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,22	2,18	2,13	2,09	2,02	1,94	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,41	2,34	2,26	2,24	2,20	2,16	2,11	2,06	2,00	1,92	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,10	2,05	1,99	1,90	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,30	2,25	2,20	2,16	2,13	2,08	2,03	1,97	1,88	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,19	2,15	2,12	2,06	2,02	1,96	1,87	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,35	2,28	2,22	2,18	2,14	2,10	2,05	2,00	1,94	1,85	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,16	2,12	2,09	2,04	1,99	1,93	1,84	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	2,04	2,00	1,95	1,90	1,84	1,74	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,98	1,95	1,90	1,85	1,78	1,69	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,86	1,81	1,75	1,65	1,39
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,88	1,85	1,79	1,75	1,68	1,57	1,28
∞	3,84	2,99	2,60	2,37	2,21	2,09	2,01	1,94	1,88	1,83	1,79	1,75	1,69	1,64	1,57	1,46	1,00

$d.f_1$ – число степеней свободы для большей дисперсии; $d.f_2$ – число степеней свободы для меньшей дисперсии

Значение χ^2 – критерия Пирсона при уровне значимости 0,10;
0,05; 0,01

<i>d.f.</i>	<i>P</i>			<i>d.f.</i>	<i>P</i>		
	0,10	0,05	0,01		0,10	0,05	0,01
1	2,71	3,84	6,63	20	28,41	31,41	37,57
2	4,61	5,99	9,21	21	29,62	32,67	38,93
3	6,25	7,81	11,34	22	30,81	33,92	40,29
4	7,78	9,49	13,28	23	32,01	35,17	41,64
5	9,24	11,07	15,09	24	33,20	36,42	42,98
6	10,64	12,59	16,81	25	34,38	37,65	44,31
7	12,02	14,07	18,48	26	35,56	38,89	45,64
8	13,36	15,51	20,09	27	36,74	40,11	46,96
9	14,68	16,92	21,67	28	37,92	41,34	48,28
10	16,91	18,31	23,21	29	39,09	42,56	49,59
11	17,28	19,68	24,72	30	40,26	43,77	50,89
12	18,55	21,03	26,22	40	51,80	55,76	63,69
13	19,81	22,36	27,69	50	63,17	67,50	76,15
14	21,06	23,68	29,14	60	74,40	79,08	88,38
15	22,31	25,00	30,58	70	85,53	90,53	100,42
16	23,54	26,30	32,00	80	96,58	101,88	112,33
17	24,77	27,59	33,41	90	107,56	113,14	124,12
18	25,99	28,87	34,81	100	118,50	124,34	135,81
19	27,20	30,14	36,19				

Учебное издание

**Афанасьев Владимир Николаевич
Юзбашев Михаил Михайлович**

**АНАЛИЗ ВРЕМЕННЫХ РЯДОВ
И ПРОГНОЗИРОВАНИЕ**

Заведующая редакцией *Л.А. Табакова*
Ведущий редактор *Е.А. Рыжова*
Младший редактор *О.О. Салтыкова*
Художественный редактор *Г.С. Семенова*
Технический редактор *Т.С. Маринина*
Корректор *Н.Б. Вторушина*
Компьютерная верстка *И.В. Рощиной*
Оформление художника *Н.М. Биксентеева*

ИБ № 5332

Подписано в печать 01.02.2010. Формат 60 × 90/16.
Гарнитура «Таймс». Печать офсетная.
Усл. п. л. 20,0. Уч. изд. л. 18,8.
Тираж 1500 экз. Заказ № 3120 «С» 008.

Издательство «Финансы и статистика»
101000, Москва, ул. Покровка, 7
Телефоны: (495) 625-35-02, 625-47-08. Факс (495) 625-09-57
E-mail: mail@finstat.ru <http://www.finstat.ru>

Издательский Дом «ИНФРА-М»
127282, Москва, ул. Полярная, д. 31в
Телефон (495) 380-05-40, 380-05-43. Факс (495) 363-92-12
E-mail: books@infra-m.ru <http://www.infra-m.ru>

Отпечатано с готовых диапозитивов в ОАО ордена «Знак Почета»
«Смоленская областная типография им. В. И. Смирнова».
214000, г. Смоленск, проспект им. Ю. Гагарина, 2.