

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Оренбургский государственный университет»

В.Н. Афанасьев

АНАЛИЗ ВРЕМЕННЫХ РЯДОВ И ПРОГНОЗИРОВАНИЕ

Учебник

Рекомендован ученым советом федерального государственного бюджетного образовательного учреждения высшего образования «Оренбургский государственный университет» для обучающихся по образовательным программам высшего образования по направлениям подготовки 01.03.05, 01.04.05 Статистика, 38.03.01, 38.04.01 Экономика, 39.03.01, 39.04.01 Социология

Ай Пи Эр Медиа, Саратов
Оренбург, 2020

УДК 519.2
ББК 22.1
А 94

Автор:

Афанасьев В. Н. – д-р экон. наук, проф., заведующий кафедрой статистики и эконометрики Оренбургского государственного университета.

Рецензенты:

Елисеева И. И. – д-р экон. наук, проф. Санкт-Петербургского экономического университета, чл.-кор. РАН;

Мхитарян В. С. – д-р экон. наук, проф. Высшей школы экономики.

Афанасьев, В. Н.

А94 Анализ временных рядов и прогнозирование : учебник /
В. Н. Афанасьев; Ай Пи Ар Медиа – Саратов, Оренбургский гос. ун-т. –
Оренбург : 2020. – 286 с.
ISBN 978-5-4497-0269-2 (Ай Пи Эр Медиа Саратов)
ISBN (Оренбургский гос. ун-т.)

Учебник состоит из двух частей, первая из которых предназначена для подготовки бакалавров, вторая (дополнительные главы) – для магистров и аспирантов. Дополнительные главы содержат современные методы в прогнозировании различных процессов и имеют практическую направленность. Учебник является основным при изучении дисциплины «Анализ временных рядов и прогнозирование», а также может быть применен частично при изучении «Прикладной математики», «Эконометрики», «Биометрики», «Системного анализа».

Подготовлен в соответствии с требованиями Федерального государственного образовательного стандарта высшего образования.

Предназначен для студентов таких направлений подготовки, как 01.03.05, 01.04.05 «Статистика», 38.03.01, 38.04.01 «Экономика», 39.03.01, 39.04.01 «Социология», также будет полезен широкому кругу читателей, интересующихся вопросами статистических закономерностей, происходящих в жизни, и их динамики. Учебник может быть использован как для очной, так и для заочной формы обучения.

УДК 519.2
ББК 22.1

ISBN 978-5-4497-0269-2 (Ай Пи Эр Медиа Саратов)
ISBN (Оренбургский гос. ун-т.)

© Афанасьев В.Н.,
© ООО «Ай Пи Эр Медиа», 2019
© ОГУ, 2020

Содержание

ПРЕДИСЛОВИЕ.....	8
ЧАСТЬ 1.....	9
ГЛАВА 1 ВИДЫ И ПОСТРОЕНИЕ ВРЕМЕННЫХ РЯДОВ	9
1.1 Что такое "временной ряд"	9
1.2 Классификация временных рядов	10
1.3 Обеспечение сопоставимости уровней временных рядов	12
ГЛАВА 2 СОСТАВЛЯЮЩИЕ ЭЛЕМЕНТЫ ВРЕМЕННОГО РЯДА.....	14
2.1 Понятие об основной тенденции и колеблемости (волатильности) временных рядов	14
2.2 Иерархия тенденция и колебаний.....	15
2.3. Периодизация динамики.....	16
ГЛАВА 3 ПОКАЗАТЕЛИ ВРЕМЕННОГО РЯДА И МЕТОДЫ ИХ ИСЧИСЛЕНИЯ.....	34
3.1 Показатели, характеризующие тенденцию динамики.....	34
3.2 Особенности показателей для рядов, состоящих из относительных уровней	38
3.3 Средние показатели временных рядов.....	42
ГЛАВА 4 ОСНОВНЫЕ ТИПЫ ТЕНДЕНЦИЙ И УРАВНЕНИЙ ТРЕНДА. 49	
4.1. Прямолинейный тренд и его свойства	49
4.2. Параболический тренд и его свойства	52
4.3 Экспоненциальный тренд и его свойства	55
4.4 Гиперболический тренд и его свойства	59
4.5 Логарифмический тренд и его свойства	61
4.6 Логистический тренд и его свойства.....	63
ГЛАВА 5 МЕТОДЫ РАСПОЗНАВАНИЯ ТИПА ТРЕНДА И ОЦЕНКИ ЕГО ПАРАМЕТРОВ.....	66

5.1 Применение графического изображения для распознавания типа тенденции	66
5.2 Методика проверки статистических гипотез о типе тренда	68
5.3 Оценка параметров линейного, параболического и гиперболического трендов	71
5.3.1. Уравнение прямой линии тренда	72
5.3.2 Уравнение параболического (II порядка) тренда	74
5.3.3 Гиперболическое уравнение тренда	77
5.4 Оценка параметров экспоненциального, логарифмического и логистического уравнений тренда	78
5.4.1. Экспоненциальное уравнение тренда	79
5.4.2. Логарифмическое уравнение тренда	80
5.4.3. Логистическое уравнение тренда	83
5.5 Многократное скользящее выравнивание	87
ГЛАВА 6 МЕТОДЫ РАСПОЗНАВАНИЯ ТИПА КОЛЕБАНИЙ И ОЦЕНКИ ПАРАМЕТРОВ КОЛЕБЛЕМОСТИ	92
6.1 Графическое отображение и основные свойства разных типов колебаний	92
6.1.1 Пилообразная колеблемость	93
6.1.2 Долгопериодическая циклическая колеблемость	95
6.1.3 Случайно распределенная во времени колеблемость	97
6.2 Измерение показателей силы и интенсивности колебаний	99
6.2.1 Показатели абсолютной величины (силы) колебаний	99
6.2.2 Показатели относительной интенсивности колебаний	101
6.3 Особенности измерения сезонных колебаний	102
6.3.1 Плавные, синусоидальные колебания при несущественности тренда .	102
6.3.2 Сезонные колебания, не имеющие синусоидальной формы, при наличии существенной тенденции	105

6.3.3 Представление синусоидальных колебаний в форме тригонометрического уравнения Фурье	111
6.4 Измерение тренда колеблемости	112
6.5 Автокорреляция отклонений от тренда	115
ГЛАВА 7 ВЕРОЯТНОСТНАЯ ОЦЕНКА СУЩЕСТВЕННОСТИ (НАДЕЖНОСТИ УСТАНОВЛЕНИЯ) ПАРАМЕТРОВ ТРЕНДА И КОЛЕБЛЕМОСТИ	117
7.1 Оценка надежности параметров тренда.....	117
7.2 Доверительные границы тренда	123
7.3 Вероятностная оценка показателей колеблемости	125
ГЛАВА 8 МЕТОДЫ ИЗУЧЕНИЯ И ИЗМЕРЕНИЯ УСТОЙЧИВОСТИ УРОВНЕЙ РЯДА И ТРЕНДА.	133
8.1 Методы измерения устойчивости уровней ряда	135
8.2 Методы измерения устойчивости тенденции динамики.....	138
8.3 Комплексные показатели (критерии) устойчивости	140
ГЛАВА 9 ИЗУЧЕНИЕ ДИНАМИКИ КОМПЛЕКСА ВЗАИМОСВЯЗАННЫХ ПРИЗНАКОВ	143
9.1 Динамика жестко связанной системы признаков (показателей).....	144
9.2 Агрегирование трендов и колебаний по совокупности объектов	152
9.2.1 Тренды объемных признаков	152
9.2.2 Тренды качественных признаков	154
9.2.3 Агрегирование показателей колеблемости.....	155
9.3 Корреляция между временными рядами: сущность, ограничения	158
9.4 Методы измерения корреляции между колебаниями признаков.....	161
9.5 Корреляция с учетом лага и циклов	166
9.6 Понятие о динамике комплекса статистически–взаимосвязанных признаков	170
ГЛАВА 10 МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ (ДИНАМИКИ).....	177

10.1 Сущность и условия прогноза по тренду с учетом колеблемости.....	177
10.2 Простая трендовая модель и прогноз по ней	180
10.3 Прогноз с учетом случайной колеблемости	182
10.3.1 Прогноз доверительного интервала для линии тренда	182
10.3.2 Прогноз доверительного интервала для уровня отдельного периода или момента.....	184
10.3.3 Прогноз доверительного интервала для среднего уровня за ряд периодов времени.....	186
10.3.4 Расчет страхового запаса для преодоления риска отклонения от тренда.....	189
10.4 Прогнозирование по модели тренда и сезонных колебаний	190
10.4.1 Модель тренда месячных уровней и средних «индексов сезонности».....	190
10.4.2 Модель тренда готовых уровней и тригонометрического уравнения сезонности	191
10.5 Прогнозирование комплекса жестко-взаимосвязанных признаков.....	193
ЧАСТЬ 2. ДОПОЛНИТЕЛЬНЫЕ ГЛАВЫ.....	200
ВВЕДЕНИЕ.....	200
ГЛАВА 11 МЕЖДУНАРОДНЫЙ ОПЫТ В РАЗВИТИИ МЕТОДОЛОГИИ СТАТИСТИЧЕСКОГО АНАЛИЗА ВРЕМЕННЫХ РЯДОВ.....	202
11.1 Принципы практического моделирования временного ряда.....	202
11.2 Скользящие и рекурсивные оценки в системе методов эконометрического анализа одномерного временного ряда.....	208
11.3 Применение двувходового объединения и теории коинтеграции в анализе взаимосвязи временных рядов	225
ГЛАВА 12 ПОВЫШЕНИЕ КАЧЕСТВА ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ	236
12.1 Потенциал моделей с длинной памятью и стейт-спейс моделей в прогнозировании одномерных временных рядов	236

12.2 Анализ причинности взаимосвязей в многофакторном прогнозировании временных рядов	244
12.3 Сравнительная оценка альтернативных методов прогнозирования и обобщение прогнозов.....	251
ГЛАВА 13 УЧЕБНЫЙ ПРИМЕР.....	255
13.1 Наивная модель	256
13.2 Детерминированный полиномиальный тренд.....	257
13.3 Экспоненциальное сглаживание.....	258
13.4 Модели авторегрессии	260
13.5 Многофакторная модель.....	266
13.6 Сравнение методов и объединение результатов	276
РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА.....	283

Предисловие

Цель издания учебника по методам анализа временных рядов заключалась в том, чтобы студенты, изучающие статистику, и особенно специализирующиеся по Направлению «Статистика», получили более широкое и углубленное представление (знания) по всему кругу методов статистического исследования, изменениям социально-экономических, психологических, технологических, технических, биологических, климатических и других массовых явлений во времени.

По Направлению «Статистика» Министерством образования и науки Российской Федерации утверждены Федеральные государственные образовательные стандарты, которые подразумевают курс под названием «Анализ временных рядов и прогнозирование» для бакалавров и магистров. Настоящий учебник состоит из двух частей, первая для бакалавров, вторая (дополнительные главы) для магистров и аспирантов. Первый вариант учебника вышел в 2001 году в издательстве «Финансы и статистика» под редакцией М. М. Юзбашева и В.Н. Афанасьева. Второй вариант издан в том же составе авторов, издательствами «Финансы и статистика» и «Инфра-М», в 2010 году.

Мы посчитали, что материала для студентов всех ступеней (бакалавриата, магистратуры и аспирантуры) высшего образования недостаточно в существующих учебниках в условиях непрерывно развивающихся методов статистического изучения разнообразных массовых процессов. **В первой части этого учебника, подробно излагаются методы распознавания типа тренда, алгоритмы расчета логарифмического, логистического, гиперболического тренда, различные модели сезонных колебаний, комплексного изучения и моделирования функционально - и корреляционно - связанных систем показателей, система показателей устойчивости уровней и тенденций изменения.** Вторая часть (дополнительные главы), представлена в учебнике впервые.

Учебник рассчитан на лиц, уже имеющих знания по теории статистики, хотя авторы и стремились сделать изложение максимально доступным. Все методики анализа, излагаемые в учебнике, по возможности иллюстрируются расчетами по фактическим временным рядам. При этом, конечно, нужно учитывать, что цель этих примеров не в анализе конкретных процессов в той или иной области, стране, эпохе, а в подтверждении необходимости и полезности применения излагаемой методики, в лучшем ее понимании читателями. **Дополнительные главы предназначены в основном для углубленного изучения бакалаврами и обязательного – магистрантами и аспирантами, включают международный опыт в развитии методологии статистического анализа временных рядов и повышение качества прогнозирования временных рядов, а также учебные примеры со сравнением методов и объединением результатов их применения.**

В.Н. Афанасьев

Часть 1

Глава 1 Виды и построение временных рядов

1.1 Что такое "временной ряд"

Термин *временные ряды* в статистике России пока еще непривычен. В учебниках общей теории статистики преобладают термины *ряды динамики*, *динамические ряды*, *статистическое изучение динамики*. В зарубежной, англоязычной литературе принят термин - *time series*, в немецкой - *zeitreihen analyze*. 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 слов *временной*, т.е. относящейся 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. производятся эти слова не от корня "время".

Опасение, что студенты (учащиеся) воспримут термин *временной ряд* именно 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 показателей. Про ряд уровней, содержащих лишь колебания, но не имеющий надежно установленной тенденции, говорят: - "В этом ряду, в данном процессе нет никакой динамики". Так говорят об экономике периода застоя, о населении страны или региона, находящихся в стационарном состоянии, о всякой общественной или механической системе, находящейся 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 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) *п о в р е м е н и – м о м е н т н ы е и и н т е р в а л ь н ы е.* *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 последовательности показателей численности населения на начало года, поголовье скота в фермерских хозяйствах на 1 декабря или 1 июня за несколько лет, величины 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.

2) *п о ф о р м е п р е д с т а в л е н и я* уровней – ряды *абсолютных* (табл. 1.1), *относительных* (табл. 1.2), и *средних величин* (табл. 1.3).

Таблица 1.1

Поголовье сельскохозяйственных животных в хозяйствах всех категорий в РФ, тысяч голов

Годы	2010	2015	2016	2017
Крупный рогатый скот	19967,9	18992	18752,5	18643,9
Овцы и козы	21819,9	24881,1	24843,8	24532,4

Таблица 1.2

Индексы цен на реализованную сельскохозяйственную и приобретенную
сельхозпредприятиями промышленную продукцию
(в % к предыдущему периоду), РФ

	К соответствующему периоду предыдущего года					
	2011	2012	2013	2014	2015	2016
Сельскохозяйственная продукция	94,9	110,8	102,7	114,1	108,5	101,8
Промышленные товары	112,1	106,2	110,1	104,8	115,6	104,4
Минеральные удобрения	122,8	108,8	103,3	104,9	126,2	98,1
Электрoэнергия	111,9	100,3	108,6	106,5	106,3	107,5
Горючие и смазочные материалы	120,0	113,2	116,7	106,8	105,7	101,5
Топливо	114,8	107,3	111,7	108,0	103,7	103,4

Таблица 1.3

Потребление основных продуктов питания в среднем в расчете на душу
населения в РФ

Виды продукции	Годы						
	2001- 2003	2004- 2006	2007- 2009	2010- 2012	2013- 2014	2015	2016
Хлебные продукты, кг	120	120	119	119	118	118	117
Молоко, кг	226	234	242	247	246	239	236
Мясо, кг	50	56	64	71	75	73	74
Яйцо, шт	242	250	256	272	269	269	273
Картофель, кг	108	109	110	108	111	112	113
Овощи, кг	82	87	98	105	110	111	112

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 (табл. 1.2; 1.4). *Неполные*, когда принцип равных интервалов не соблюдается (табл. 1.1 1.3).

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 банки отражают платежеспособность населения и т.д.

Таблица 1.4

Прогноз агрегированного показателя экономической конъюнктуры России, 1999 г.

месяц	апрель	май	июнь	июль	август	сентябрь	октябрь	ноябрь	декабрь
уровень	104,87	105,02	105,17	105,31	105,46	105,61	105,76	105,90	106,05

Агрегированные показатели (табл. 1.4) основаны на частных показателях и характеризуют изучаемый процесс комплексно. Так, чтобы иметь представление о состоянии экономики России в целом, необходимо определять агрегированный показатель экономической конъюнктуры, включающий в себя и вышеперечисленные частные показатели. Их определяют, также, 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 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, историческому периоду, методике расчета, ценам, достоверности.

Несопоставимость по территории возникает в результате изменений границ стран, регионов, хозяйств и т.п. Для приведения данных к сравнимому 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 времени должны характеризовать размер того или иного явления по одному и тому же кругу, входящих в его состав частей. Например, при характеристике динамики урожайности овощных культур в регионе по годам нельзя в одни годы учитывать только фермерские хозяйства, а в другие - все категории хозяйств, в том числе приусадебные участки сельских жителей и сады - огороды горожан.

При определении сравниваемых уровней ряда необходимо использовать *единую методику* их расчета. Особенно часто эта проблема возникает при международных сопоставлениях. Так, например, до недавнего времени в России

урожайность сельскохозяйственных культур определяли делением валового сбора на весеннюю продуктивную площадь, в США же, на фактически убранную площадь.

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/час, л.с.), атмосферного давления (мм рт. ст., миллибарах) и т.д.

Трудности, при сравнении данных *по моменту регистрации*, возникают из-за сезонных явлений. Численность скота в домашних хозяйствах из-за экономической целесообразности различна зимой и летом. Поэтому уровни при сравнении должны относиться к определенной дате ежегодно.

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. Для сравнения социально-экономических показателей республик бывшего СССР, важно брать *исторические ряды*, то есть уровни ряда до распада союзного государства и после распада. В данном случае, изменились и форма собственности, и законодательства.

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 предложения) 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 органами - 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.

Все выше названные 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 (волатильности) 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, *изучение динамики*.

Познание закономерностей изменений во времени сложная и трудоемкая процедура исследования, так как любое изучаемое явление формируют множество факторов, действующих в разных направлениях. По характеру непосредственного их воздействия факторы эти 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 процессом агротехники, с укреплением экономики данной совокупности хозяйств сельских товаропроизводителей, совершенствованием организации и управления производства. Колеблемость (волатильность) урожайности вызвана чередованием благоприятных по погоде и неблагоприятных лет, циклами солнечной активности, колебаниями в развитии вредных насекомых и болезней растений.

При статистическом изучении динамики необходимо четко разделить ее два основных элемента – тенденцию и колеблемость (волатильность), 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 (волатильностью)* следует называть отклонения уровней отдельных периодов времени от тенденции динамики (тренда).

2.2 Иерархия тенденция и колебаний

В предыдущем параграфе были рассмотрены две основные компоненты временного ряда - его тенденция и колеблемость (волатильность) 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 пятен). 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 2016 и 2017 годы

Год	Январь	Февраль	Март	Апрель	Май	Июнь	Июль	Август	Сентябрь	Октябрь	Ноябрь	Декабрь
2016	65,8	65,3	54,5	47,3	60,2	25,0	40,9	48,7	47,8	38,0	31,0	25,7
2017	31,7	30,4	20,3	35,1	22,8	22,2	18,8	28,8	45,0	16,1	6,5	10,1

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 продажи 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 воздействия - 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 на чемпионате мира или телепередачи «Синяя птица». 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 тенденций и колебаний для бизнес-аналитика и статистика, для менеджера различных структур имеет огромное значение.

2.3 Периодизация динамики

Периодизация развития, т. е. расчленение периода развития во времени на однородные этапы, в пределах которых показатель подчиняется одному закону развития, это, по существу, *типологическая группировка во времени*.

Периодизация может осуществляться несколькими методами.

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

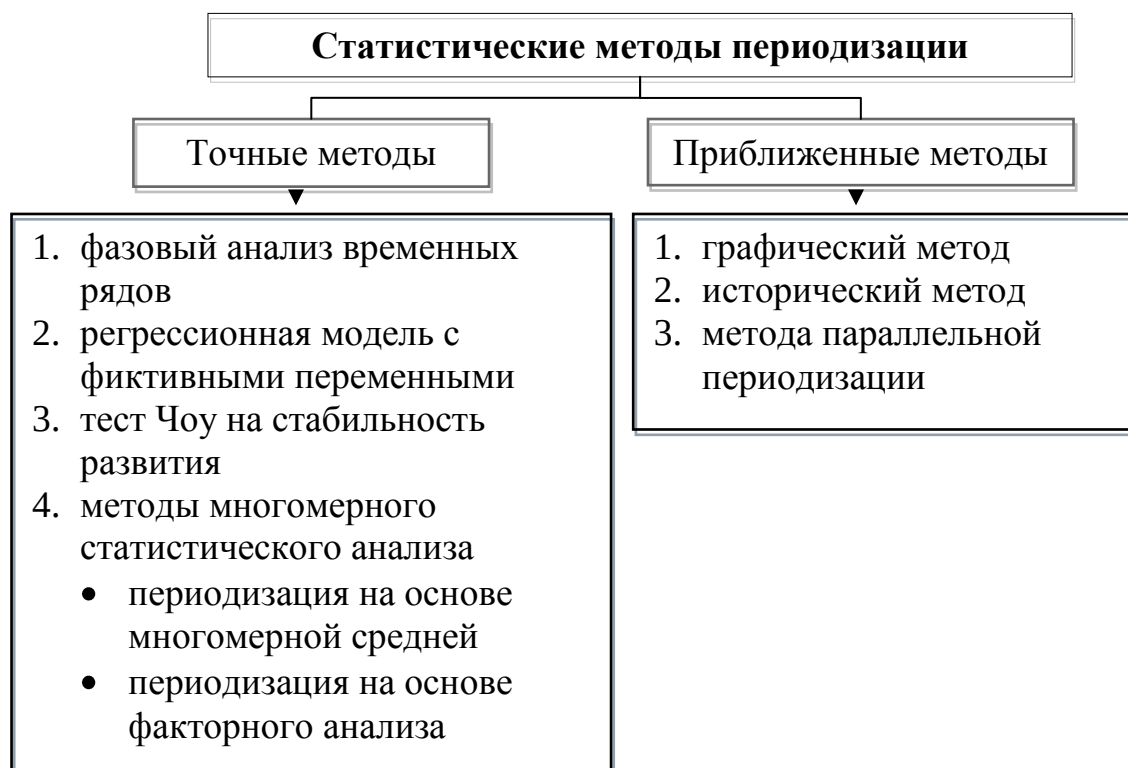
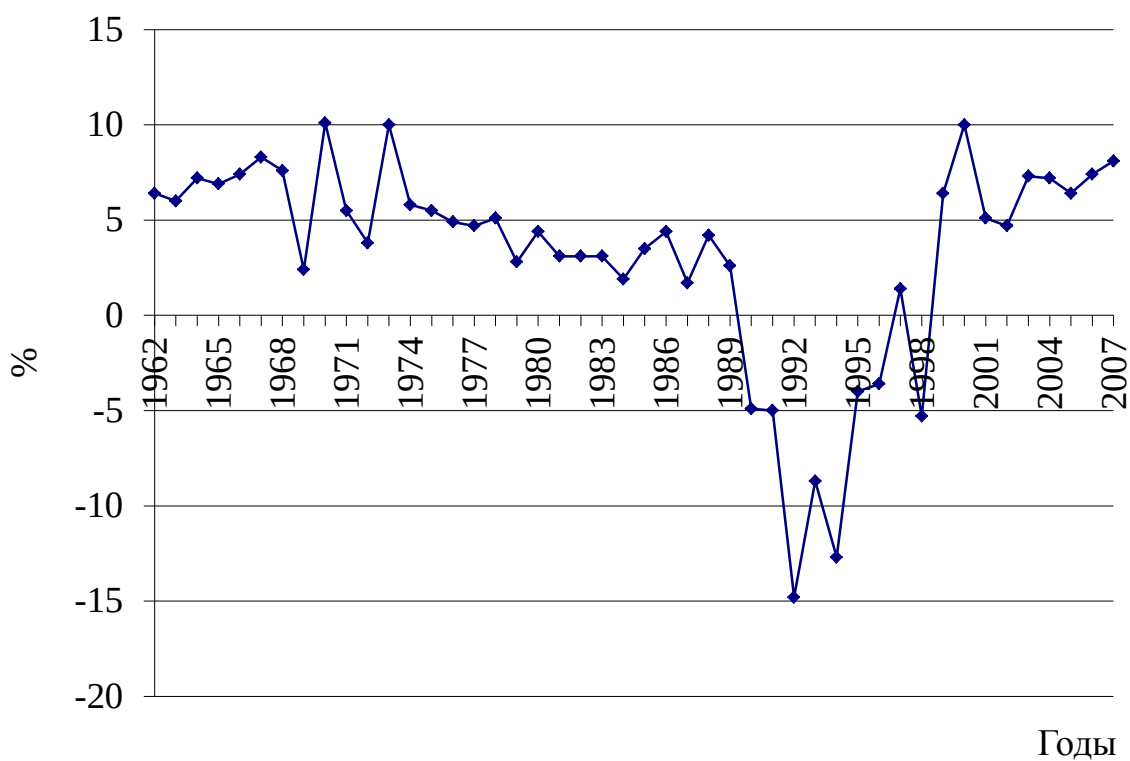


Рис. 2.1 Систематизация методов периодизации временных рядов

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).



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 проводится на основе значимых дат и событий - времени принятия управленческих решений приводящих к изменениям уровней изучаемых показателей; смена хозяйственного механизма; смена руководства; войны и т. п. Недостатком этого метода является то, что точные временные границы периодов путем теоретического анализа удастся получить крайне редко.

Помощь в выявлении и описании однородности периодов является использование *фазового анализа* и теста Г. Чоу.

Применение *фазового анализа временных рядов* подробно рассмотрено в

книге Ю.П. Лукашина¹. Основная идея метода заключается в постепенном (итеративном) очищении динамического (временного) ряда от маломощных колебаний, отождествляемых со случайными или второстепенными, конъюнктурными флуктуациями. В ходе процесса фильтрации происходит сглаживание наименее мощных фаз, в результате чего мелкие (маломощные) фазы объединяются в одну более крупную.

Рассмотрим некоторые понятия, которые будем использовать 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 (условно принятой) 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 детерминированной повторяемости.

Фаза - период положительных или отрицательных флуктуаций ряда Dx_t

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

Длительность фазы - временной отрезок положительных или отрицательных флуктуаций ряда.

Точка смены фаз - момент смены знака флуктуации.

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

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

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

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

где: t_i - момент начала фазы i ;

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

¹ Лукашин, Ю.П. *Адаптивные методы прогнозирования: учеб. пособие* / Ю.П.Лукашин. - М.: Финансы и статистика, 2003. - 416 с.:ил. - . ISBN 5-279-02740-5

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

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

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 v — процент совокупной мощности ряда, которым мы решаемся пожертвовать в процессе агрегирования фаз путем подавления случайных флуктуаций;

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

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

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

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 флуктуаций. Для этого заменяем фазы $j-1$, j , $j+1$ одной, мощность которой равна $p_{j-1} - p_j + p_{j+1}$. В ходе этой операции число фаз уменьшается на 2, а совокупная мощность временного ряда сокращается на $2p_j$. 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 агрегирования прекращается и полученные фазы подлежат содержательному или статистическому анализу.

Этап 4. В результате применения изложенной процедуры весь интервал выборки ряда будет разбит точками смены фаз на отрезки, внутри которых можно представить движение ряда упрощенно линейным образом. Для этого достаточно разделить мощность фазы на ее длительность - получим среднюю мощность 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

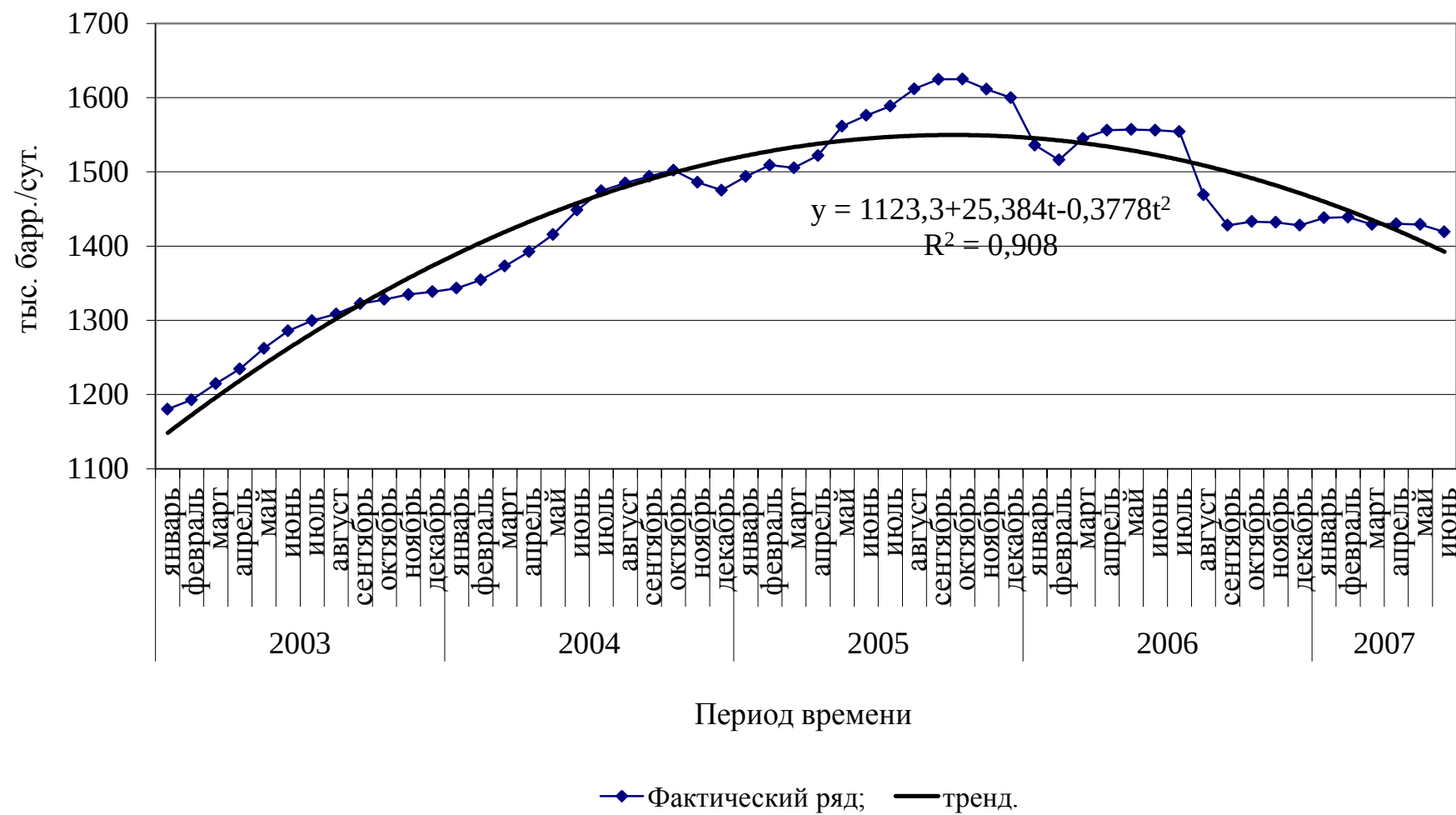


Рис. 2.3 Динамика добычи нефти

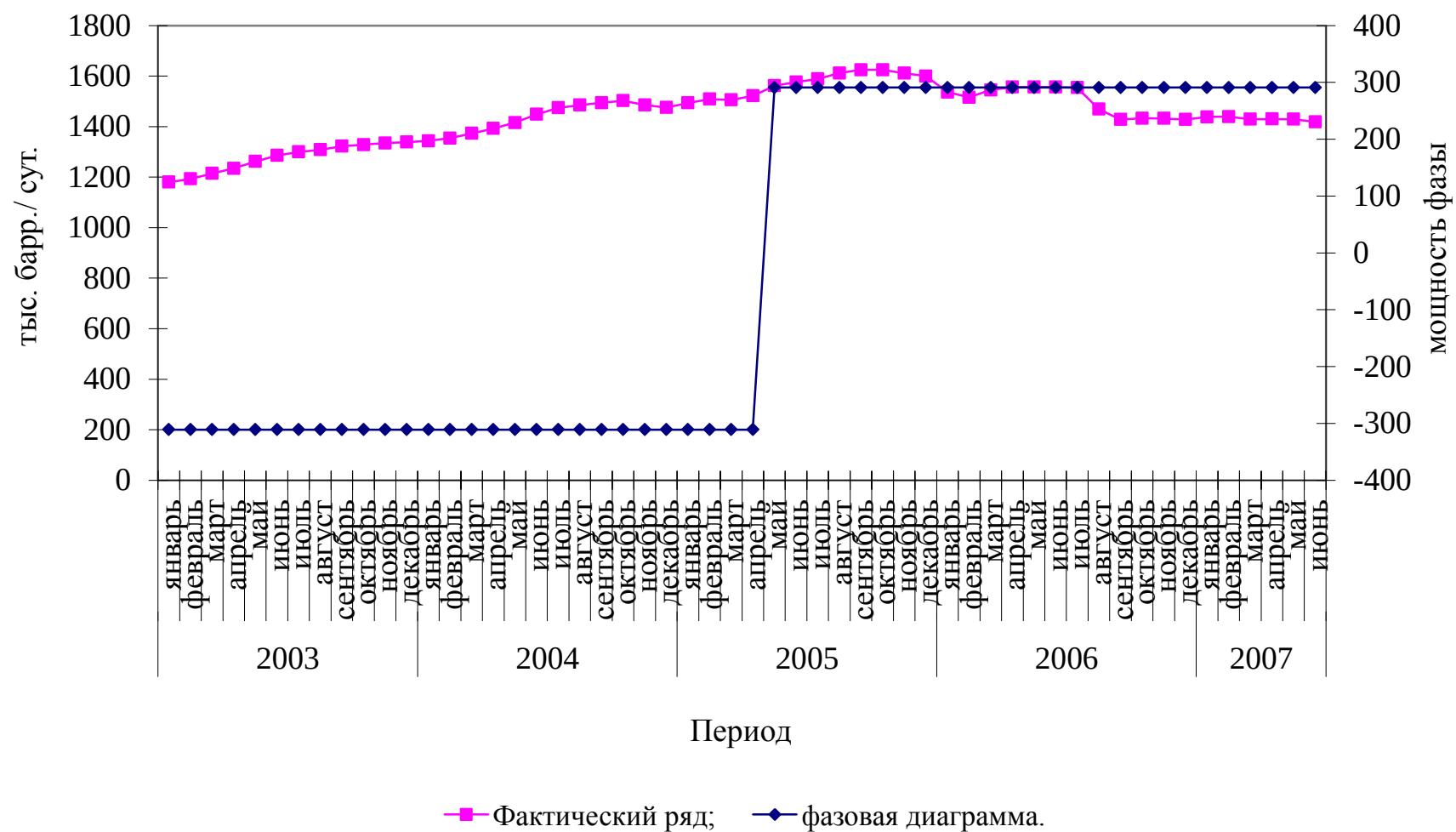


Рис. 2.4. Фазовая диаграмма временного ряда добычи нефти

Тест Г. Чоу в статистических исследованиях 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, например, для проверки правильности разбиения динамического 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:

$$F_{\text{факт}} = \frac{\Delta S}{\sum_{i=1}^m k_i} / \frac{S_{K-Л}}{n - k_1 - k_2} = \frac{53285262,76}{3} / \frac{7147418,43}{48} = 119,28$$

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:

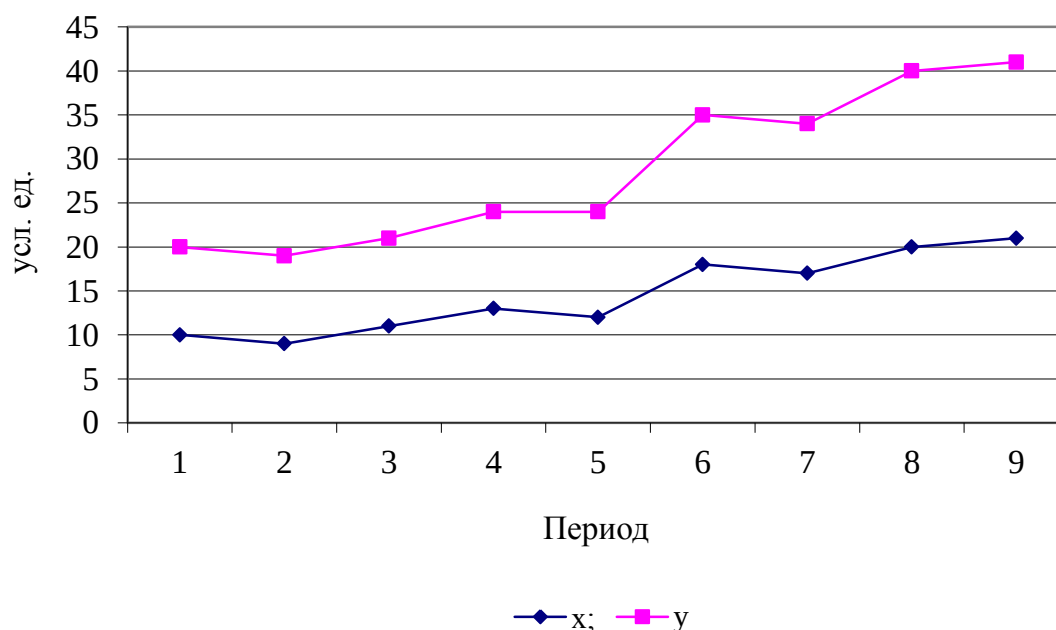


Рис. 2.5 Динамика прибыли предприятия (y) и среднегодовой стоимости основных производственных фондов (x), усл. ед.

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 . Недостатком метода параллельной

периодизации является сложность в нахождении x - детерминирующего показателя. Более того, во многих случаях такой параметр вообще невозможно найти, так как он должен обладать 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, чаще всего из-за отсутствия статистической информации.

На основе комплексных динамических рядов (системы показателей) периодизация проводится методом многомерной средней и методами факторного анализа.²

Периодизация на основе *факторного анализа* наиболее применима на практике. Во-первых, потому, что методика факторного анализа сравнительно достаточно разработана; во-вторых, реализация метода существует в современных статистических пакетах программ (STATISTICA, SPSS, STADIA 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

² Глинский В.В., Ионин В.Г. *Статистический анализ.* – М.: ИНФРА-М; Новосибирск: Сибирское соглашение, 2002. – 241 с.

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).

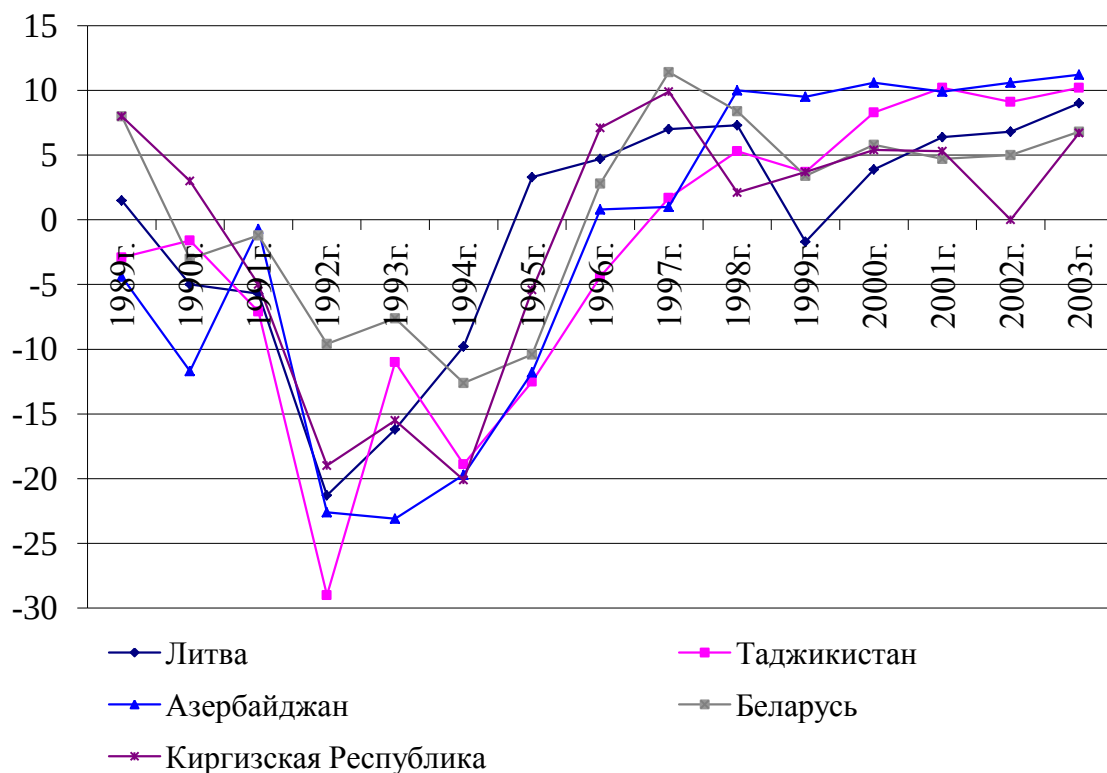


Рис.2.6 Динамика темпа прироста ВВП для бывших стран «социалистического лагеря»

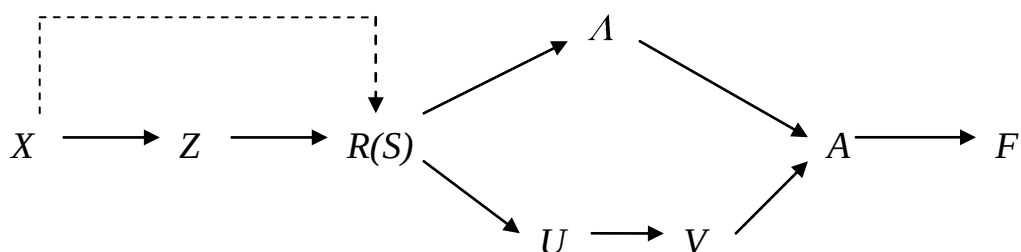


Рис. 2.7 Схема математических преобразований при проведении факторного анализа

Z – матрица центрированных и нормированных значений признаков, элементы матрицы вычисляют по одной из формул:

$$Z_{ij} = \frac{X_{ij}}{X_{\max j}}; \quad Z_{ij} = \frac{X_{ij}}{X_{\min j}}; \quad Z_{ij} = \frac{X_{ij}}{\bar{X}_j}; \quad Z_{ij} = \frac{X_{ij} - \bar{X}_j}{\sigma_j}$$

Если предварительная стандартизация данных не проводилась, то на данном шаге получают матрицу $S = (1/n) \cdot X' \cdot X$, элементы матрицы X для расчета будут центрированными величинами.

Множество решений λ_j находят решением характеристического уравнения $|R - \lambda E| = 0$, λ_j – это характеристики вариации, точнее, показатели дисперсии каждой главной компоненты. Суммарное значение $\sum \lambda_j$ равно сумме дисперсий элементарных признаков X_j . При условии стандартизации исходных данных, эта сумма равна числу элементарных признаков m .

$$(1-\lambda_j) \cdot u_{1j} + r_{12} \cdot u_{2j} + r_{13} \cdot u_{3j} + \dots + r_{1m} \cdot u_{mj} = 0,$$

$$r_{21} \cdot u_{1j} + (1 - \lambda_j) \cdot u_{2j} + r_{23} \cdot u_{3j} + \dots + r_{2m} \cdot u_{mj} = 0,$$

.....

$$r_{m1} \cdot u_{1j} + r_{m2} \cdot u_{2j} + r_{m3} \cdot u_{3j} + \dots + (1 - \lambda_j) \cdot u_{mj} = 0.$$

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:

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

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

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 методом, 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.

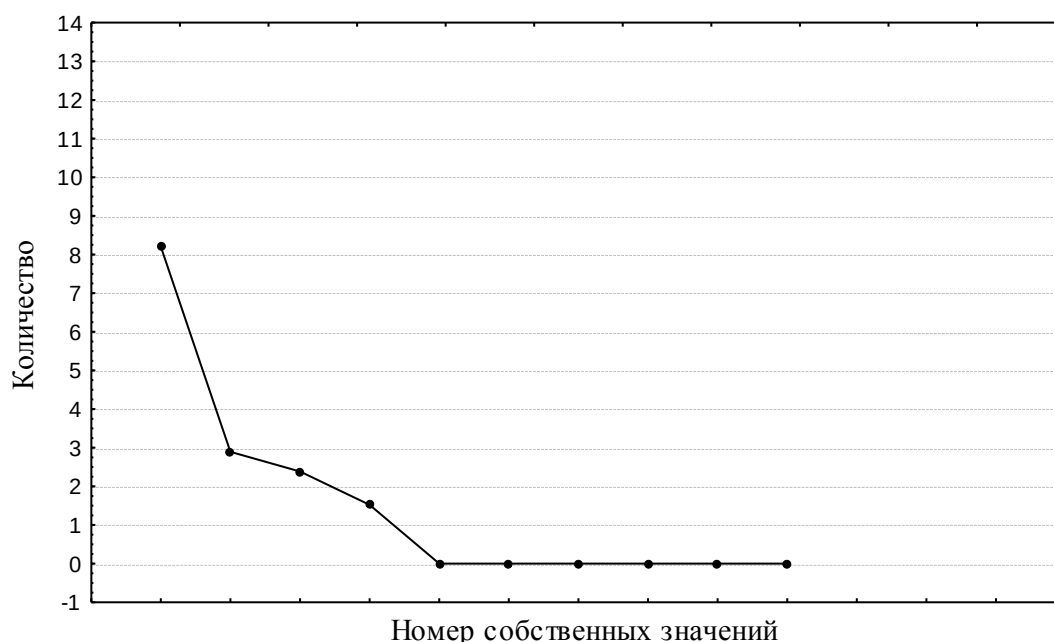


Рис. 2.8 Собственные значения факторов относительно номеров факторов для бывших стран «социалистического лагеря»

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 – объяснении 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

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 факторами. 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 динамики 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

Подробный анализ применения различных методов периодизации динамики, подчеркивает важность исследования качественного содержания уровней рядов динамики (их типологизацию). Факторы, определяющие периоды в динамических рядах играют основополагающую роль в прогнозировании.

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 (волатильности) 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:

$$\text{цепное, } \Delta_{\text{ц}} = y_n - y_{n-1};$$

$$\text{базисное, } \Delta_0 = y_n - y_0.$$

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.

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 абсолютного изменения, тыс. т. в год	Темп роста уровня, в % к предыдущему	Темп роста – уровня, в % к начальному
0 (начальный)	100	-	-	-	100
1	112	12	-	112	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

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 периода. В нашем случае ускорение составило 4 тыс. т в год за год или 4 тыс. т·год². Смысл показателя следующий: объем произ-

водства (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: $y_i = 20 + 4t + 0,5t^2$. 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 медленными темпами, 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, а в базисном — к одному и тому же, обычно начальному уровню, что иллюстрируется формулой (3.2). Он свидетельствует 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 будет неверно, если сказать, что «уровень меньше в 0,33 раза». Темп изменения в размах всегда говорит о том, во сколько раз сравниваемый уровень больше.

Теперь можно утверждать, 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. На втором предприятии, вычислив также шесть уровней параболического тренда, студент 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 характеристик динамики можно правильно отразить процесс развития совокупности (объекта).

Рассмотрим связь абсолютных и относительных показателей динамики. Обозначим темп изменения через k , тогда имеем:

$$\text{цепной темп роста в период с номером } n \quad k_n = \frac{y_n}{y_{n-1}} \quad (3.2)$$

$$\begin{array}{l} \text{базисный темп роста за весь период} \\ \text{между базой (0) и текущим годом (n)} \end{array} \quad k_{n/0} = \frac{y_n}{y_0}$$

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 значения, 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

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

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

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

$$12 + 14,3 + 15,6 + 16,2 + 16,3 + 16 \neq 132 \text{ (в процентах)}.$$

Значения цепных темпов прироста, рассчитанных каждый к своей базе, различаются не только числом процентов, но и величиной абсолютного изменения, составляющей каждый процент. Поэтому складывать или вычитать цепные темпы прироста нельзя. *Абсолютное значение 1%-ного 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 — численность 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 прибыли). 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 дол., теперь стоит более 2000 современных долларов).

Например, «Биржевые ведомости» за 5.05.90 сообщают: индекс Доу Джонса на 3.05.90 составил 2689,64% в сравнении с 2759,55% на 29.04.90. 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 (волатильность) курсов. «Биржевые ведомости» далее сообщали, что индекс Доу Джонса на 7.06.90 достиг 2911,6%, т.е. с 5.05.90 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 изменению 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 очевидно, 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 зависит 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:

$$A_1 = -A_2 = x_0 (k_1 - 1) \cdot 100. \quad (3.5)$$

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

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

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 фермерских 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 поголовье 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 динамики 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, т.е. хорошая модель не получилась. 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 главах. В данной главе рассматриваются только форма, математические свойства средних показателей динамики и простейшие приемы их вычисления, применимые на практике к рядам со слабой колеблемостью (волатильностью).

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 урожайности картофеля N-й области

Годы	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Урожайность, ц/га	149	145	168	146	177	176	190	186	176	211	170

По данным табл. 3.2 определим среднегодовые уровни урожайности картофеля по пятилетиям.

Пятилетия	Среднегодовые уровни, ц/га
2007 – 2011	157,0
2013 – 2017	186,6

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

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:

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

В моментном ряду смысл среднего уровня в том, что он характеризует уже не состояние объекта в отдельные моменты, а его среднее, обобщенное состояние между начальным и конечным моментами учета. Из этого следует, что уровни, относящиеся к начальному и конечному моментам, играют не ту роль, что уровни, относящиеся к моментам 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:

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

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:

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

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:

$$\Delta = \frac{4717 - 4016}{10} = 70,4 \text{ тыс. шт. за год.}$$

Для правильной интерпретации показатель среднего абсолютного 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); 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, среднегодовое производство мяса в Российской Федерации составляло:

Годы	2002 - 2006	2007 - 2011	2012 – 2016
Млн. т.	5,00	6,69	9,03

Абсолютный прирост за второе пятилетие в сравнении с первым составил 1,69 млн. т, за третье в сравнении со вторым — 2,34 млн. т. Следовательно, ускорение в третьем пятилетии по сравнению со вторым составило $2,34 - 1,69 = 0,65$ млн. т в год за пять лет, а среднегодовое ускорение прироста равно: $0,65 : 5 = 0,13$ млн. т 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 (волатильностью), 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 сопровождать 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:

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

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

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? Существует приближенная формула для среднего темпа роста, удовлетворяющая этому условию:

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

где m — число суммируемых уровней;
 y_0 — базисный уровень.

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

$$k_{\text{пар}} = 1 - \frac{3}{8} + \sqrt{\frac{9}{64} + \frac{1}{20} \left(\frac{\sum y_i}{y_0} - 5 \right)} = 1 - 0.375 + \sqrt{0.1406 + 0.05 \left(-5 \right)} = 1.16407$$

≈ 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. Казинцом [8]. Им составлены таблицы, с помощью которых, зная отношение суммы уровней к базисному уровню и число суммируемых уровней m , можно получить $\bar{k}_{\text{пар}}$. Таблица Л.С. Казинца рассчитана на основе нахождения корней уравнения:

$$\bar{k} + k^{-2} + k^{-3} + \dots + k^{-n} = \left(\sum_{i=1}^n y_i : y_0 \right)$$

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

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

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 $\bar{b}^{\text{I}}$; второй ряд — с показателями y_0^{II} , a_0^{II} , $\bar{b}^{\text{II}}$. При каком числе n периодов (лет) после базисного уровни рядов сравниваются? Тенденции рядов параболические:

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

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

Приравняв правые части уравнений, получим:

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

или

$$(y_0^{\text{II}} - y_0^{\text{I}}) + n(a_0^{\text{II}} - a_0^{\text{I}}) + n^2 \left(\frac{\bar{b}^{\text{II}} - \bar{b}^{\text{I}}}{2} \right) = 0$$

(3.13)

Искомый срок n является корнем этого квадратного уравнения. Если, например, имеем:

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

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

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 сравниваются, то эту задачу можно решить на основе *темпов изменения*. Имеем:

$$y_0^{\text{II}} \cdot \bar{k}^{n\text{II}} = y_0^{\text{I}} \bar{k}^{n\text{I}}$$

Логарифмируя это равенство, имеем:

$$n \log \bar{k}^{\text{II}} + \log y_0^{\text{II}} = n \log \bar{k}^{\text{I}} + \log y_0^{\text{I}},$$

откуда

$$n(\log \bar{k}^{\text{II}} - \log \bar{k}^{\text{I}}) = \log y_0^{\text{I}} - \log y_0^{\text{II}}, \quad (3.14)$$

$$n = \frac{\log y_0^{\text{I}} - \log y_0^{\text{II}}}{\log \bar{k}^{\text{II}} - \log \bar{k}^{\text{I}}}$$

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 $y_0^{\text{I}} = 300$; $\bar{k}^{\text{I}} = 1,09$; второй ряд имеет $y_0^{\text{II}} = 100$; $\bar{k}^{\text{II}} = 1,2$. 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 тренда

В гл. 2 было 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 с 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 с некоторым приближением отражается тем или иным уравнением, линией тренда.

В данной главе далеко не все известные в математике линии и их уравнения, а лишь набор сравнительно простых их 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:

$$\hat{f}_i = a + b * t_i,$$

где $\hat{f}_i$ - выровненные, то есть лишенные колебаний уровни тренда для лет с номером i ,

a – свободный член уравнения, численно равный среднему выровненному уровню для момента или периода времени, принятого за начало отсчета, то есть, для $t_i = 0$;

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

t_i - номера моментов или периодов времени, к которым относятся уровни временного ряда (года, кварталы, месяцы, даты).

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 другие 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 так далее. Влияние разнонаправленных и разноускоренных (замедленных) сил факторов взаимно усредняется, частично взаимно погашается, а равнодействующая 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.

Абсолютные изменения уровней в разные годы не были точно одинаковыми, но общая тенденция увеличения среднедушевых денежных доходов населения РФ очень хорошо отражается прямолинейным трендом. Его параметры вычислены в гл. 5 табл. 5.3.

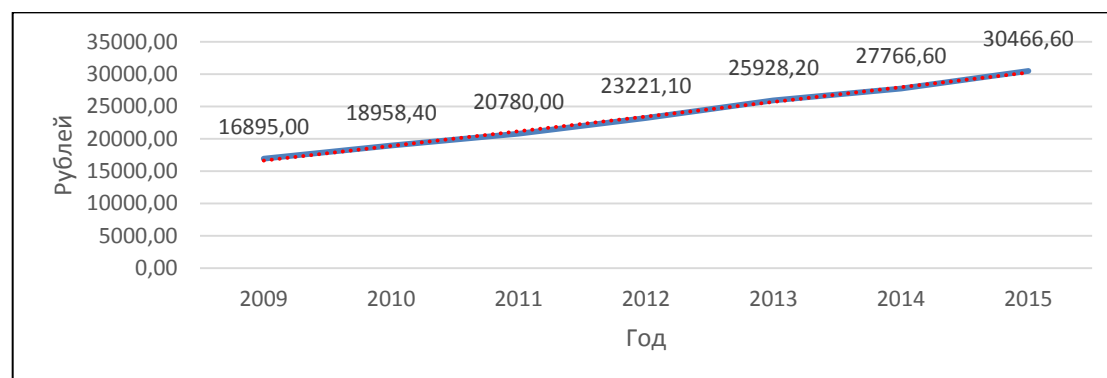


Рис. 4.1. Динамика среднедушевых денежных доходов населения РФ, рублей

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 таблица 4.1.
Уравнение тренда: $\bar{Y}_i = 100 + 20t_i$.

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

Таблица 4.1.

Показатели динамики при линейном тренде к увеличению уровней.

$$\bar{Y}_i = 100 + 20t_i$$

Номер периода, t_i	Уровни, $\bar{Y}_i$	Абсолютное изменение (цепное)	Темп изменения (цепной), %	Ускорение
1	120	+20	120	...
2	140	+20	116,7	0
3	160	+20	114,3	0
4	180	+20	112,5	0
5	200	+20	111,1	0
6	220	+20	110,0	0

Таблица 4.2.

Показатели динамики при линейном тренде сокращения

$$\text{уровней: } \mathcal{Y}_i = 200 - 20 \cdot t_i$$

Номер периода, t_i	Уровни, $\mathcal{Y}_i$	Абсолютное изменение (цепное)	Темп изменения (цепной), %	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:

$$\mathcal{Y}_i = a + b \cdot t + c \cdot t^2.$$

Парабола III 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 (выровненный) 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 $2c$, 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 $(n-1)$ -го порядка! (Через любые две точки проходит одна прямая, через три точки – одна парабола II порядка и т.д.) Такое «приближение» линии тренда к эмпирическому 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 городов и регионов, ускоренное увеличение объема продукции в фазе циклического подъема, как, например, динамика торгового баланса Китая в 2008-2015 годах, приведенная на рис. 4.2.

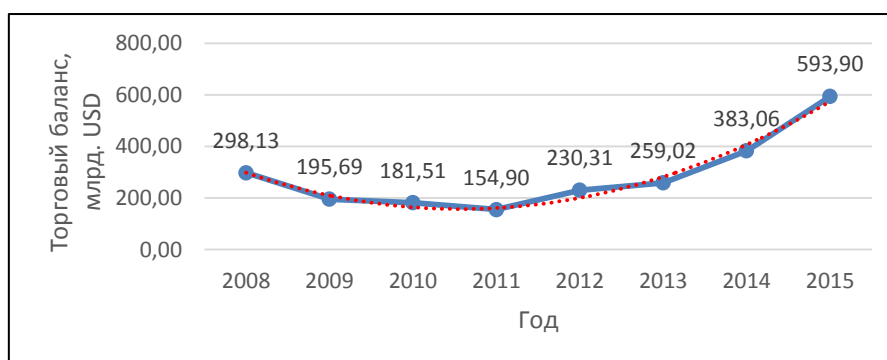


Рис. 4.2. Торговый баланс Китая, млрд. USD

Расчет уравнения этой параболы приведен далее 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 с точки зрения ее математической формы, имеет две ветви: восходящую с увеличением уровня 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 редких 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 :

а) при $b > 0$ и $c > 0$ имеем восходящую ветвь, то есть 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. 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).

Ввиду ограниченного объема учебника рассмотрим не все 4 случая параболических трендов, а лишь два первых (табл. 4.3 и 4.4).

Таблица 4.3.

Показатели динамики при параболическом тренде,

когда $b > 0, c > 0$: $y_i = 100 + 20t + 2t^2$

№ периода t_i	Уровни y_i	Абсолютные изменения	Цепные темпы, % к предыдущ.	Ускорение
1	122	+22	122	...
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

Таблица 4.4.

**Показатели динамики при параболическом тренде,
когда $b < 0, c < 0$: $\mathcal{F} = 200 - 20t - 2t^2$**

№ периода t_i	Уровни $\mathcal{F}_i$	Абсолютны е изменения	Цепные темпы, % к предыдущ.	Ускорение	Цепное относитель ное изменение в % к предыдущ.
1	178	-22	89,0	...	-11
2	152	-26	85,4	-4	-14,6
3	122	-30	80,3	-4	-19,7
4	88	-34	72,1	-4	-27,9
5	50	-38	56,8	-4	-43,2
6	8	-42	16,0	-4	-84,0

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 $\mathcal{F} = a + bt + ct^2$ 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 $\mathcal{F} = 100 + 20t - 2t^2$, то максимум параболы имеет при:

$$t = \frac{-20}{2(-2)} = 5.$$

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

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

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

$$\mathcal{F}_{min} = 200 - 20 \cdot 5 + 2 \cdot 5^2 = 150.$$

4.3 Экспоненциальный тренд и его свойства

Экспоненциальным трендом называют тренд, выраженный уравнением: $y_i = a \cdot k^{t_i}$ или в форме $y_i = \exp[a + \ln k \cdot t_i]$. Свободный член экспоненты a равен выровненному уровню, т.е. уровню тренда в момент или период, принятый за начало отсчета времени, то есть 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 Земного шара на протяжении 1950 – 1985 годов возрастала 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.

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.

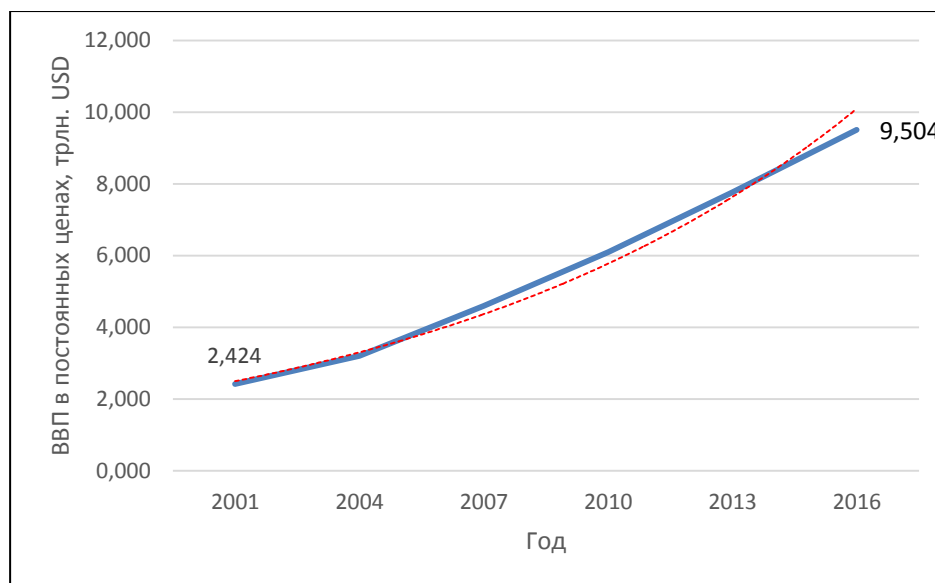


Рис. 4.3. Рост ВВП Китая в постоянных ценах, трлн. USD

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 таблицах 4.5. и 4.6.

В таблицах 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 приведены 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 4.5.

Экспоненциальный тренд при $k>1$: $\mathcal{F}_i = 100 * 1,2^i$

№ периода t_i	Уровни $\mathcal{F}_i$	Абсолютные изменения (цепные)	Цепные темпы роста, %	Ускорение	Прирост ускорения (цепной)
1	120	+20	120	...	
2	144	+24	120	+4	...
3	172,80	+28,80	120	+4,80	+0,80
4	207,36	+34,56	120	+5,76	+0,96
5	248,83	+41,47	120	+6,81	+1,15
6	298,60	+49,77	120	+8,30	+1,39

Таблица 4.6.

Экспоненциальный тренд при $k<1$: $\mathcal{F}_i = 200 * 0,8^i$

№ периода t_i	Уровни $\mathcal{F}_i$	Абсолютные изменения (цепные)	Цепные темпы роста, %.	Ускорение	Замедление ускорения
1	160	-40	80	...	...
2	128	-32	80	+8	...
3	102,40	-25,6	80	+6,4	-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 поинтересоваться 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: $\mathcal{F} = a \mathcal{E} + bt_i$ или $\mathcal{F} = ak^{t^2}$ 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, 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, что это происходит только до определенного 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 экспоненциальное 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: $y = a + \frac{b}{t}$.

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

Такая тенденция наблюдается, например (рис. 4.4), при изучении процесса снижения затрат любого ресурса (труда, материалов, энергии) на единицу данного вида продукции или ее себестоимости в целом. Затрата ресурса не может стремиться к нулю, значит, экспонента не соответствует сущности процесса; нужно применить гиперболическую формулу тренда.

Если параметр $b < 0$, то с возрастанием t , т.е. 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.

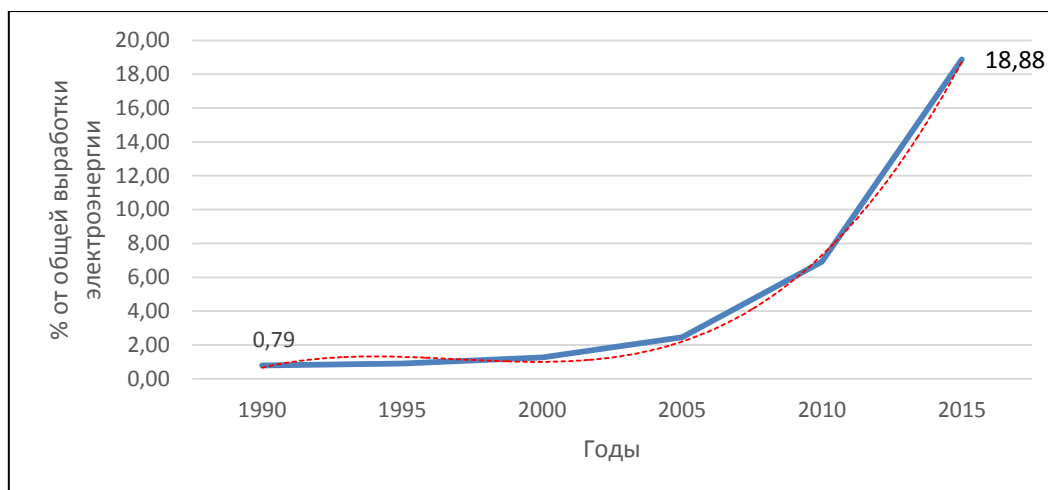


Рис. 4.4. Выработка возобновляемой электроэнергии в Бельгии, % от общей выработки электроэнергии

Основные свойства гиперболического тренда таковы:

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

2. При $b < 0$ уровни замедленно возрастают, положительные абсолютные изменения уменьшаются, отрицательные ускорения также уменьшаются 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, переход от движения к застою. Иллюстрацией этих свойств может служить таблица 4.7.

Таблица 4.7

Показатели динамики при гиперболическом тренде:

$$\mathcal{F}_i = 100 + \frac{100}{t_i}$$

№ периода t_i	Уровни $\mathcal{F}_i$	Абсолютные изменения (цепные)	Ускорение	Цепные темпы
1	200	...	...	...
2	150	-50	...	75,0
3	133,3	-16,7	+33,3	88,9
4	125,0	-8,3	+8,4	93,8
5	120,0	-5,0	+3,3	96,0
6	116,7	-3,3	+1,7	97,2

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: $\mathcal{Y}_i = a + b \ln t_i$

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 0,1с 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, что иллюстрирует рис. 4.5.

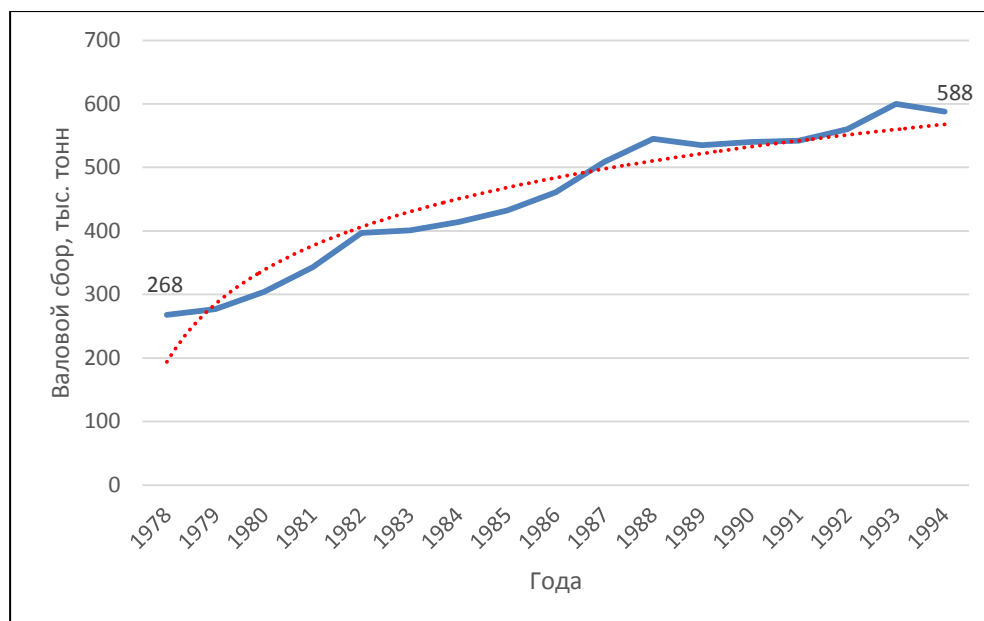


Рис. 4.5. Динамика валового сбора чая в Китае

Конечно, характер тенденции маскируется колебаниями, но видно, что рост валового сбора замедляется. Это показывают и средние уровни сбора чая:

за 1978-1983 гг. средний сбор равен 333 тыс. т;

за 1984-1989 гг. средний сбор равен 483 тыс. т, рост на 150 тыс. т;

за 1990-1994 гг. средний сбор равен 566 тыс. т, рост на 83 тыс. т.

На рисунке, для убедительности нанесен и логарифмический тренд, расчет которого дан в гл. 5. Заметны также 5-6-летние циклические колебания валового сбора чая.

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 состоит 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. 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 в 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 или переводных программах для ПЭВМ логистическая кривая называется *S-образной кривой*.

Можно, конечно, логистическую тенденцию считать объединением трех разных по типу тенденций: параболической с ускоряющимся ростом на первом этапе, линейной - на втором, и гиперболической с замедляющимся ростом - на третьем этапе. Но есть доводы и в пользу рассмотрения всего цикла развития как особого единого типа тенденции со сложными, переменными свойствами, но постоянным направлением изменений в сторону увеличения уровней в рассмотренных нами примерах или уменьшения 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 проявление 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 годов 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 (смартфонов), насыщенность которыми еще находится в начале первого этапа.

В вышеописанном диапазоне изменения уровней, то есть от нуля до единицы, уравнение логистического тренда имеет вид:

$$y_i = \frac{1}{e^{a_0 + a_1 t_i} + 1}$$

При $a_0 > 0$, $a_1 < 0$ с ростом номеров периодов времени 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 $a_0 < 0$; $a_1 > 0$ 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:

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

Таблица 4.8.

Показатели динамики при логистическом тренде:

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

№ периода t_i	Уровни y_i	Абсолютные изменения к предыдущ. периоду	Ускорение	Темп роста к предыдущ. Периоду, %
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

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, темпы начинаются со 100 % и завершаются на 100 %.

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. 4.7.

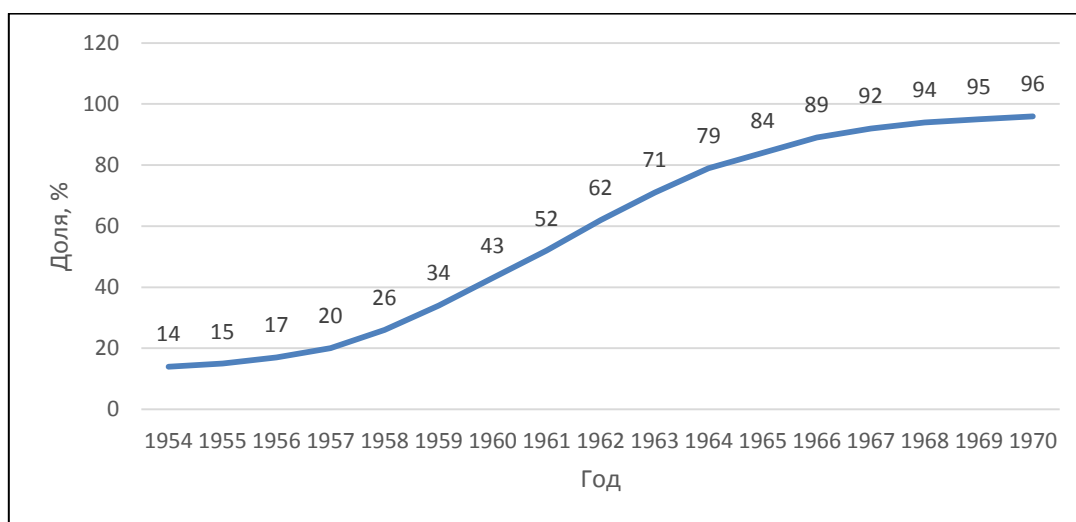


Рис. 4.7. Логистическая тенденция динамики доли тепловозной и электровозной тяги в грузообороте железных дорог СССР

Период охватывает почти весь процесс замены паровозов тепловозами и электровозами. Наиболее быстро этот процесс происходит с 1960 по 1964 годы.

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. Если, например, изучается динамика продуктивности коров или валового надоя молока, то они аналогичны указанной в гл. 4 динамике урожайности, и скорее всего, отобразятся линейным трендом. Если изучается динамика расхода бензина на 100 км пробега автомобиля по мере развития и совершенствования двигателей, то этот процесс аналогичен динамике снижения трудоемкости при освоении технологии производства изделий, и, скорее всего, отображается гиперболическим трендом.

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 главе 4. Кроме того, характер тенденции часто маскируется значительной колеблемостью уровней ряда. Поэтому и является необходимой 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 логарифмического или логистического типа тренда.

5.1 Применение графического изображения для распознавания типа тенденции

Графическое изображение во многих случаях позволяет приближенно 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, иначе говоря, равные отрезки будут означать 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 по ординате прямая линия означает экспоненциальную тенденцию. Необходимо строго соблюдать равенство промежутков времени на равных отрезках оси абсцисс. Логарифмический масштаб по времени не рекомендуется, он крайне затруднит интерпретацию графика. Рассмотрим пример графического изображения, представленный на рис. 5.1.

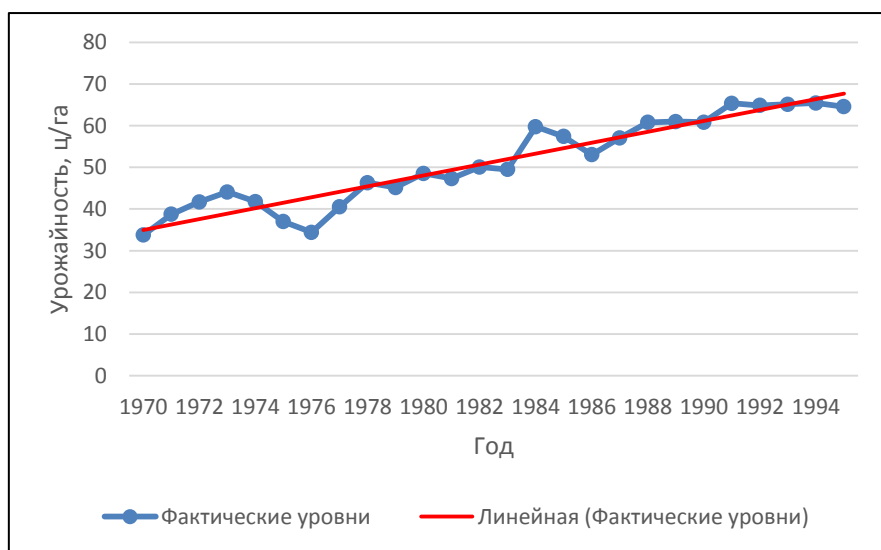


Рис. 5.1. Динамика урожайности зерновых во Франции

Видно, что линейный тренд хорошо подходит для отражения тенденции динамики урожайности зерновых культур во Франции: прямая проходит как бы посреди колеблющихся точек–уровней лет.

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 много способов объективной, статистико-математической оценки пригодности того или иного типа линии. Весьма популярен выбор его с помощью перебора на электронных вычислительных

машинах (ЭВМ) всех имеющихся в пакете 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, 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 математико-статистические 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, а также средних отклонений.

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. Теперь необходимо 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 параболы II порядка означает, что на всем периоде (в среднем) имеется постоянная величина ускорения абсолютных изменений. Гипотеза об экспоненциальном тренде подтвердится, если можно 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 цепные абсолютные изменения $\Delta_i = y_{i+1} - y_i$ (для параболы - ускорения, для экспоненты - темпы).

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.

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 урожайности зерновых культур во Франции, для которой на основании графика рис. 5.1. предложена гипотеза о линейном тренде (табл. 5.1).

Таблица 5.1.

Проверка гипотезы о линейном тренде

Годы	Урожайность, центнеров с 1 гектара	Скользящая 5-летняя средняя, центнеров с 1 гектара	Абсолютные изменения, центнеров с 1 гектара	Средние по подпериодам абсолютные изменения, центнеров с 1 гектара
1970	33,7			
1971	38,8			
1972	41,7	40,02		
1973	44,1	40,68	+0,66	} $\bar{\Delta}_1 = +0,79$
1974	41,8	39,80	-0,88	
1975	37,0	39,56	-0,24	
1976	34,4	40,00	+0,44	

1977	40,5	40,68	+0,68	
1978	46,3	42,98	+2,30	
1979	45,2	45,56	+2,58	
1980	48,5	47,48	+1,92	
1981	47,3	48,12	+0,64	} $\bar{\Delta}_2 = +1,71$
1982	50,1	51,02	+2,90	
1983	49,5	52,78	+1,76	
1984	59,7	53,92	+1,14	
1985	57,3	55,30	+1,38	
1986	53,0	57,56	+2,26	
1987	57,0	57,82	+0,26	
1988	60,8	58,52	+0,70	} $\bar{\Delta}_2 = +1,04$
1989	61,0	61,00	+2,48	
1990	60,8	62,58	+1,58	
1991	65,4	63,44	+0,86	
1992	64,9	64,34	+0,90	
1993	65,1	64,82	+0,48	
1994	65,5	-	-	Общая средняя величина $\bar{\Delta}_{\text{общ}} = +1,18$
1995	63,2	-	-	
Итого	1332,6	-	-	

Источник данных "Развитие рынка зерна в России", изд. ЦЭК при Правительстве РФ. М.: 1997 год, стр. 53.

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

Таблица 5.2.

**Результаты дисперсионного анализа различий между средними
абсолютными изменениями**

Вид вариации	Сумма квадратов отклонений	Число степеней свободы	Дисперсия	F-критерий	
				фактиче ский	табличный для P=0,05
1. Между подпериодами	3,1682	2	1,5841	1,7365	3,55
2. Остаточная	16,4198	18	0,9122	—	—
Вся вариация	19,588	20	0,9794	—	—

Полученное значение F-критерия значительно ниже табличного для значимости 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 урожайности постоянны, тренд урожайности – прямая линия.

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 в разд. 5.5.

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, дающий оценки параметров, отвечающие принципу максимального правдоподобия: сумма квадратов отклонений фактических уровней от тренда (от выравненных 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 ЭВМ 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 самой силе колебаний. Таким образом, эти лжекоэффициенты не могут характеризовать соотношение между ролью факторов тенденции и ролью факторов колеблемости (волатильностью).

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 № 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 $f(a, b) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - a - bt_i)^2$ к нулю.

Имеем:

$$\frac{\partial f}{\partial a} = 2 \sum_{i=1}^n (y_i - a - bt_i) = 0 \quad (5.1)$$

$$\frac{\partial f}{\partial b} = 2 \sum_{i=1}^n (y_i - a - bt_i) t_i = 0 \quad (5.2)$$

После алгебраических преобразований получаем два "нормальных уравнения" МНК для прямой:

$$na + b \sum_{i=1}^n t_i = \sum_{i=1}^n y_i \quad (5.3)$$

$$a \sum_{i=1}^n t_i + b \sum_{i=1}^n t_i^2 = \sum_{i=1}^n y_i t_i \quad (5.4)$$

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 ϵ_1 равен $a+b$, для второго $\epsilon_2 = a+2b$ 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 № $n/2$ 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, 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:

$$na = \sum_{i=1}^n y_i \quad (5.5)$$

$$b \sum_{i=1}^n t_i^2 = \sum_{i=1}^n y_i t_i \quad (5.6)$$

Откуда имеем:

$$a = \frac{\sum_{i=1}^n y_i}{n} = \bar{y} \quad (5.7)$$

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

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 об этом не предупреждается. При расчетах без компьютера, конечно, следует применить упрощенный прием. Знаменатель в формуле (5.8) 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 среднедушевых денежных доходов населения в РФ, с 2009 по 2015 г. представлена в табл. 5.3. В целях экономии, в той же таблице вычислены и другие показатели, необходимые для измерения колеблемости, описываемые в гл. 6.

Таблица 5.3.

Расчет линейного тренда

Годы	Уровни, y_i млн. чел.	номера лет t_i	$y_i t_i$	тренд, $\hat{y}_i$, руб.	отклоне ние от тренда $y_i - \hat{y}_i = u_i$	u_i^2	$u_i * u_{i+1}$
2009	16895,00	-3	-50685	16629,48	265,52	70500,87	16409,14
2010	18958,40	-2	-37916,8	18896,6	61,80	3819,24	-23713,9
2011	20780,00	-1	-20780	21163,72	-383,72	147241,04	80481,43
2012	23221,10	0	0	23430,84	-209,74	43990,87	-48290,5
2013	25928,20	1	25928,2	25697,96	230,24	53010,46	-45698
2014	27766,60	2	55533,2	27965,08	-198,48	39394,31	-46523,7
2015	30466,60	3	91399,8	30232,2	234,40	54943,36	
Σ	164015,90	0	63479,40	164015,88	0,02	412900,14	-67335,61

$$a = \bar{y} = \frac{164015,90}{7} = 23430,84 \text{ рублей} \quad b = \frac{63479,40}{28} = 2267,12 \text{ рублей}$$

Уравнение тренда: $\hat{y}_i = 23430,84 + 2267,12t_i$, $t_i = 0$ в 2012 году. В среднем, среднедушевые денежные доходы населения увеличивались на 2267,12 рублей в год. Сумма уровней тренда должна равняться сумме фактических уровней, различие в четвертой значащей цифре связано с округлением значений параметров.

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 производные 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:

$$na + b \sum_{i=1}^n t_i + c \sum_{i=1}^n t_i^2 = \sum_{i=1}^n y_i \quad (5.9)$$

$$a \sum_{i=1}^n t_i + b \sum_{i=1}^n t_i^2 + c \sum_{i=1}^n t_i^3 = \sum_{i=1}^n y_i t_i \quad (5.10)$$

$$a \sum_{i=1}^n t_i^2 + b \sum_{i=1}^n t_i^3 + c \sum_{i=1}^n t_i^4 = \sum_{i=1}^n y_i t_i^2 \quad (5.11)$$

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:

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

Уравнение (5.9) и (5.11) образуют систему двух уравнений с двумя неизвестными:

$$na + c \sum_{i=1}^n t_i^2 = \sum_{i=1}^n y_i \quad (5.12)$$

$$a \sum_{i=1}^n t_i^2 + c \sum_{i=1}^n t_i^4 = \sum_{i=1}^n y_i t_i^2 \quad (5.13)$$

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. 5.4, в которой присутствуют также графы, необходимые для анализа колеблемости (волатильности), описываемые в гл. 6.

Таблица 5.4.

Расчет параболического тренда торгового баланса Китая

Годы	Уровни, y_i , млрд. USD	t_i	$y_i t_i$	$y_i t_i^2$ *	Тренд $\hat{y}_i$ *	$y_i - \hat{y}_i = u_i$	u_i^2	$u_i u_{i+1}$
2008	298,13	-3,5	-1043,46	3 652,09	297,89	0,25	0,06	-3,43
2009	195,69	-2,5	-489,225	1 223,06	209,68	-13,99	195,58	-244,53
2010	181,51	-1,5	-272,265	408,40	164,03	17,49	305,73	-105,52
2011	154,90	-0,5	-77,45	38,73	160,94	-6,04	36,42	-180,48
2012	230,31	0,5	115,155	57,58	200,41	29,91	894,31	-700,23
2013	259,02	1,5	388,53	582,80	282,44	-23,42	548,26	561,14
2014	383,06	2,5	957,65	2 394,13	407,03	-23,97	574,32	-472,71
2015	593,90	3,5	2078,65	7 275,28	574,18	19,73	389,08	-0,79
Σ	2296,52	0	1657,59	15 632,05	2 296,56	-0,04	2943,75	-1 146,54

* округлены до целых.

$$\sum_1^8 t_i^2 = \frac{8^3 - 8}{12} = 42; \sum_1^8 t_i^4 = \frac{3 * 8^5 - 10 * 8^3 + 7 * 8}{240} = 388,5$$

Вычисляем параметры параболы:

$$b = \frac{1657,59}{42} \cong 39,47;$$

$$8a + 42c = 2296,52$$

$$42a + 388,5c = 15632,05$$

$$a + 5,25c = 287,07$$

$$a + 9,25c = 372,19$$

$$4c = 85,12; c = 21,28$$

$$a = 175,35$$

Уравнение тренда:

$$\hat{y}_i = 175,35 + 39,47t_i + 21,28t_i^2$$

где $t=0,5$ в 2012 году.

Интерпретация параметров тренда такова: торговый баланс Китая в 2008-2015 годах возрастал в номинальной оценке ускоренно, со средним ускорением $2 * 21,28 = 42,56$ млрд. USD в год за год, средний за весь период прирост торгового баланса составил 39,47 млрд. USD в год, средний уровень торгового баланса на середину периода был равен 175,35 млрд. USD.

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

$$\hat{y}_i = 428,69 - 152,07t_i + 21,28t_i^2,$$

где $t_i = 0$ в 2008 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:

$$na + b \sum_{i=1}^n \frac{1}{t_i} = \sum_{i=1}^n y_i \quad (5.14)$$

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

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. 5.5) по данным рис. 4.4. – динамика изменения доли выработки возобновляемой энергии от общей выработки электроэнергии, в %.

Таблица 5.5.

Расчет гиперболического уравнения тренда

Годы	y_i , %	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
1990	0,79	1	1,000	1,000	0,79	-1,97	2,76	7,62
1995	0,91	2	0,500	0,250	0,45	4,10	-3,19	10,15
2000	1,26	3	0,333	0,111	0,42	6,12	-4,86	23,57
2005	2,46	4	0,250	0,063	0,61	7,13	-4,67	21,81
2010	6,92	5	0,200	0,040	1,38	7,73	-0,81	0,66
2015	18,88	6	0,167	0,028	3,15	8,14	10,74	115,38
Σ	31,22	-	2,45	1,49	6,81	31,24	-0,02	179,20

Нормальные уравнения МНК:

$$\begin{aligned}6a + 2,45b &= 31,22 \\ 2,45a + 1,49b &= 6,81\end{aligned}$$

Решая систему уравнений, получаем:

$$a = 10,16; b = -12,13$$

Уравнение гиперболического тренда удельного расхода топлива имеет вид:

$$\bar{x}_t = 10,16 - \frac{12,13}{t_t}$$

Величина доли 10,16% – это предел, к которому стремится выработка возобновляемой электроэнергии от общей выработки при существующем уровне развития технологий.

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 – 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, алгебраически меньшую наименьшего реального уровня. Например, температуру выразить не в градусах Цельсия, а в Кельвинах, финансовый результат, при наибольшем убытке 83 млн. руб. отсчитывать от минус 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:

$$\mathcal{E}_i = 27 * 1,028^{t_i}$$

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:
 $47 < \left[\ln 100 : 27 : \ln 1,028 \right] < 48$.

5.4.1. Экспоненциальное уравнение тренда

Формула уравнения имеет вид: $\mathcal{E}_i = a * k^{t_i}$

Для нахождения параметров a и k уравнение логарифмируем:

$$\ln \mathcal{E}_i = \ln a + t_i \ln k$$

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.).

$$n \ln a + \ln k \sum_{i=1}^n t_i = \sum_{i=1}^n \ln y_i \quad (5.16)$$

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

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

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

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

Приведем пример расчета экспоненциального тренда по данным рис. 4.3 (табл. 5.6).

Таблица 5.6.

**Расчет экспоненциального тренда ВВП Китая в постоянных ценах в
2001–2016 годах**

Годы	y_i , трлн. USD	t_i	$y_i t_i$	$\ln y_i$	$t_i \ln y_i$	Тренд $\hat{y}_i$, трлн. USD	$y_i - \hat{y}_i = u_i$	u_i^2
2001	2,424	-2,5	-6,059	0,885	-2,213	134,442	-132,019	17428,927
2004	3,205	-1,5	-4,807	1,165	-1,747	177,733	-174,528	30460,051
2007	4,597	-0,5	-2,298	1,525	-0,763	234,963	-230,366	53068,541
2010	6,101	0,5	3,050	1,808	0,904	310,621	-304,520	92732,457
2013	7,767	1,5	11,650	2,050	3,075	410,641	-402,874	162307,465
2016	9,504	2,5	23,761	2,252	5,629	542,867	-533,363	284475,618
Σ	33,596	0,0	25,296	9,685	4,886	270,156	35740,867	314999,061

Источники данных: Всемирный Банк

$$\ln a = \frac{33,596}{6} = 5,599; a = 270,156;$$

$$\ln k = \frac{4,886}{17,5} = 0,279; k = 1,322$$

Уравнение тренда:

$$\hat{y}_i = 270,156 * 1,322^{t_i}$$

Итак, ВВП Китая в постоянных ценах в период с 2001 по 2016 года возрастал со среднегодовым темпом роста, равным корню четвертой степени из среднего темпа за десятилетие, найденного по данным табл. 5.6., т.е. $\sqrt[4]{1,322} = 1,072$ или 7,2% прироста в год.

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 и 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 валового сбора чая в Китае (см. рис. 4.5; табл. 5.7).

Таблица 5.7.

Расчет логарифмического тренда валового сбора чая в Китае

Годы	y_i тыс.т.	t_i	$\ln t_i$	$\mathcal{F}_i$	$y_i - \mathcal{F}_i = u_i$	u_i^2	$u_i u_{i+1}$
1978	268	4	1,386	221	47	2209	235
1979	277	5	1,609	272	5	25	-50
1980	304	6	1,792	314	-10 •	100	60
1981	343	7	1,946	349	-6	36	-102
1982	397	8	2,079	380	17 •	289	-102
1983	401	9	2,197	407	-6	36	102
1984	414	10	2,303	431	-17	289	357
1985	432	11	2,398	453	-21 •	441	252
1986	461	12	2,485	473	-12	144	-216
1987	509	13	2,565	491	18	324	648
1988	545	14	2,639	509	36 •	1296	396
1989	535	15	2,708	524	11	121	11
1990	540	16	2,773	539	1	1	-11
1991	542	17	2,833	553	-11 •	121	66
1992	560	18	2,890	566	-6	36	-132
1993	600	19	2,944	578	22 •	484	-44
1994	588	20	2,996	590	-2	4	-
Итого	7716	-	-	7650	k=6	652,6	1470

Временной ряд, прежде всего нужно разделить на несколько частей, например на три части, и в каждой части вычислить средний уровень, тыс. т:

период 1978–1983 гг. — 331,7;

период 1984–1989 гг. — 482,7;

период 1990–1994 гг. — 566,0.

Эти усредненные уровни относятся соответственно к середине между 1980 и 1981 гг., к середине между 1986 и 1987 гг. и к 1992 г. Если первую дату обозначить годом номер x , то вторая будет годом

номер $x+6$, а третья – годом номер $x+11,5$. Исходя из уравнения логарифмического тренда, имеем уравнения:

$$a + b \ln x = 331,7 \quad (5.18)$$

$$a + b \ln (x + 6) = 482,7 \quad (5.19)$$

$$a + b \ln (x + 11,5) = 566 \quad (5.20)$$

Вычитая (5.18) из (5.19), имеем: $b[\ln (x + 6) - \ln (x)] = 151$
(5.21)

Вычитая (5.19) из (5.20), имеем: $b[\ln (x + 11,5) - \ln (x + 6)] = 83,3$
(5.22)

Делим второй результат на первый:

$$\frac{\ln (x + 11,5) - \ln (x + 6)}{\ln (x + 6) - \ln x} = \frac{83,3}{151} = 0,5517.$$

Это число говорит о степени замедления роста средних уровней между подпериодами ряда. Теперь необходимо подобрать такое значение x , при котором получаем наибольшее приближение к рассчитанному показателю замедления роста уровней.

При $x = 2$ получим $\frac{\ln (2 + 11,5) - \ln (2 + 6)}{\ln (2 + 6) - \ln 2} = \frac{0,5323}{1,3863} = 0,384$, что слишком мало.

Увеличим x до 6:

$$\frac{\ln (6 + 11,5) - \ln (6 + 6)}{\ln (6 + 6) - \ln (6)} = \frac{0,3773}{0,6931} = 0,5443$$
 - все еще ниже наблюдаемой величины.

Примем $x=8$.

$$\frac{\ln (8 + 11,5) - \ln (8 + 6)}{\ln (8 + 6) - \ln 8} = \frac{0,3314}{0,5596} = 0,5922$$
 - что уже больше наблюдаемого значения.

При $x=7$, имеем:

$$\frac{\ln (7 + 11,5) - \ln (7 + 6)}{\ln (7 + 6) - \ln 7} = \frac{0,3528}{0,6190} = 0,5699$$
 - немного больше необходимого.

Примем $x=6,5$:

$$\frac{\ln (6,5 + 11,5) - \ln (6,5 + 6)}{\ln (6,5 + 6) - \ln 6,5} = \frac{0,3646}{0,6539} = 0,5576$$

Можно, принимая дробные значения x подойти еще ближе к фактическому значению, однако, вряд ли целесообразно применять мелкодробные номера периодов времени, да и сам процесс усреднения уровней по подпериодам ряда включает субъективные моменты, так что лучше ограничиться приближением $x \approx 6,5$ лет, а,

следовательно: середина между 1980 и 1981 гг. — это номер 6,5 от начала отсчета номеров лет, тогда 1978 год – это номер $t=4$. Исходя из этого нумеруем все годы в таблице 5.7, начиная с $t=4$ до $t=20$.

Зная величину $x=6,5$, подставляем ее в уравнения (5.21) и (5.22), чтобы вычислить по ним величину параметра b . Из (5.21):

$$b(\ln 12,5 - \ln 6,5) = 151,$$

откуда $b=230,9$

Из уравнения (5.22):

$$b(\ln 18 - \ln 12,5) = 83,3,$$

откуда $b=228,4$.

Принимаем среднее из двух независимых оценок параметра " b ", равное 229,6.

Теперь, подставляя значения x и b в уравнения (5.18), (5.19), (5.20) получим три независимых оценки параметра a :

Из (5.18): $a + 229,6 \ln 6,5 = 331,7$; откуда $a = -98,1$;

Из (5.19): $a + 229,6 \ln 12,5 = 482,7$; откуда $a = -97,2$;

Из (5.20): $a + 229,6 \ln 18 = 566$; откуда $a = -97,6$

Средняя оценка параметра a равна $-97,6$

Итак, уравнение логарифмического тренда имеет вид:

$$\hat{y} = -97,6 + 229,6 \ln t,$$

где $t = 0$ в 1974 г.

По этому уравнению рассчитаны уровни тренда $\hat{y}$, в табл. 5.7. Хотя суммы уровней немного разошлись, кривая, как видно на рис. 4.5., 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, производную 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, во-вторых, теоретических т.е.

внешних по отношению к статистике, соображений, относящихся к содержанию изучаемого процесса.

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

$$\frac{\mathcal{F}_{max} - \mathcal{F}_{min}}{\mathcal{F}_i - \mathcal{F}_{min}} - 1 = e^{a_0 + a_1 t_i}.$$

и обозначим его левую часть, т.е.

$$\dots\dots\dots \frac{\mathcal{F}_{max} - \mathcal{F}_{min}}{\mathcal{F}_i - \mathcal{F}_{min}} - 1 = \zeta_i, \text{ т.е. } \zeta_i = e^{a_0 + a_1 t_i}; \ln \zeta_i = a_0 + a_1 t_i$$

Условие метода наименьших квадратов:

$$\sum_{i=1}^n (\ln \zeta_i - \ln \xi_i) \rightarrow \min,$$

подставляя значение $\ln \xi_i$, имеем:

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

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:

$$n a_0 + a_1 \sum_{i=1}^n t_i = \sum_{i=1}^n \ln \zeta_i \quad (5.23)$$

$$a_0 \sum_{i=1}^n t_i + a_1 \sum_{i=1}^n t_i^2 = \sum_{i=1}^n t_i \ln \zeta_i \quad (5.24)$$

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

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

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

Итак, алгоритм расчета логистической кривой состоит из этапов:

1. обоснование величин $\mathcal{F}_{max}$ и $\mathcal{F}_{min}$

2. вычисление по фактическому временному ряду значений

$$\zeta_i = \frac{f_{\max} - f_{\min}}{y_i - f_{\min}} - 1$$

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 $\ln \xi_i = a_0 + a_1 t_i$.

9. вычисление $\xi_i = \exp [a_0 + a_1 t_i]$ для всех периодов.

10. вычисление уровней тренда $f_i = \frac{f_{\max} - f_{\min}}{\xi_i + 1} + f_{\min}$.

Проведем расчет логистического тренда по данным рис. 4.7.

Период охватывает почти весь процесс замены паровозов тепловозами и электровозами. Наиболее быстро этот процесс происходил с 1960 по 1964 г.

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

$$\begin{aligned} f_{\min} &= 10\%; f_{\max} = 100\% \\ a_0 &= \frac{-0,228}{14} = -0,016286; a_1 = \frac{-99,325}{227,5} = -0,436593 \\ \xi_i &= \exp [0,016286 + t_i (-0,436593)] \end{aligned}$$

Уравнение логистического тренда доли прогрессивных видов тяги в грузообороте железных дорог за 1955–1968 гг. имеет вид:

$$f_i = \frac{100 - 10}{e^{-0,016286 - 0,436593 t_i} + 1} + 10$$

Табл. 5.8 показывает достаточно близкое приближение логистической кривой, судя по тому, что сумма уровней тренда различается от суммы фактических уровней менее, чем на 1%. 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.

Таблица 5.8.

Расчет логистического тренда

Годы	$y_i, \%$	$\zeta_i = \frac{100 - 10}{y_i - 10} - 1$	$\ln \zeta_i$	t_i	$t_i \ln \zeta_i$	ξ_i	$\epsilon_i = \frac{100 - 10}{\xi_i + 1} + 10, \%$	u_i	u_i^2
1955	15,0	17,00	2,833	-6,5	-18,416	16,803	15,1	-0,1	0,01
1956	17,0	11,86	2,473	-5,5	-13,601	10,854	17,6	-0,6	0,36
1957	20,0	8,00	2,079	-4,5	-9,357	7,017	21,2	-1,2●	1,44
1958	26,4	4,49	1,501	-3,5	-5,255	4,535	26,3	+0,1	0,01
1959	33,5	2,83	1,040	-2,5	-2,601	2,931	32,9	+0,6	0,36
1960	43,2	1,71	0,537	-1,5	-0,805	1,814	41,1	+2,16●	4,41
1961	51,8	1,15	0,142	-0,5	-0,071	1,224	50,5	+1,3●	1,69
1962	61,8	0,74	-0,305	0,5	-0,152	0,791	60,3	+1,5●	2,25
1963	70,7	0,48	-0,728	1,5	-1,093	0,510	69,6	+1,1●	1,21
1964	78,9	0,31	-1,183	2,5	-2,958	0,330	77,7	+1,2●	1,44
1965	84,5	0,21	-1,570	3,5	-5,495	0,213	84,2	+0,3	0,09
1966	88,8	0,14	-1,951	4,5	-8,779	0,138	89,1	-0,3●	0,09
1967	92,4	0,09	-2,383	5,5	-13,109	0,089	92,6	-0,2●	0,04
1968	94,4	0,07	-2,713	6,5	-17,633	0,058	95,1	-0,7	0,49
Σ	778,4	-	-0,228	0	-99,325	-	773,3	-	13,89

5.5 Многократное скользящее выравнивание

Как видно из таблиц 5.3, 5.4, 5.5, 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 конце 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 и есть то же самое, что и номера периодов от середины ряда.

Ввиду этого, если в начале ряда находится уровни с отрицательными отклонениями от нормы, от тренда (например, неурожайные годы), а в конце ряда уровни с положительными отклонениями от тренда (высокоурожайные годы), то среднегодовой прирост урожайности в линейном тренде или ускорение прироста в параболе будут завышены 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.

Чтобы сделать следующий шаг в освобождении параметров тренда от влияния случайного распределения положительных и отрицательных колебаний уровней на протяжении временного ряда, предложено (см. Юзбашев М.М., Манелля А.И. "Статистический анализ тенденций и колеблемости" М.: Финансы и статистика, 1983 год) применять методику многократного скользящего выравнивания.

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. Например, ряд динамики урожайности зерновых культур во Франции (табл. 5.1.), состоящий из 26 уровней ($N = 26$), выравнивать по 15 уровням, сначала - с 1970 по 1984 год, затем с 1971 по 1985 год и так далее, скользя по ряду на 1 год, вплоть до последних 15 уровней с 1981 года по 1995 год. Каждый раз вычисляется среднегодовой прирост, например "b" линейного тренда, при этом на концах будут то года, благоприятные для урожая зерновых, то неблагоприятные и по метеорологическим, и по экономическим условиям. Всего получим 12 разных "баз" по 15 лет; обозначив длину каждой базы, т.е. число уровней, по которым производится расчет параметра, "n", а число таких баз расчета, укладываемых в ряд длиной N уровней, как "l", имеем равенство:

$$n + l = N + 1$$

В ряду из 26 уровней уложатся 12 баз по 15 уровней в каждой. Получим 12 значений среднегодового прироста урожайности, 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 неурожайных 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 качественно 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, то есть минимальное значение $l \approx 6 - 8$, то минимальная длина исходного временного ряда, то есть N должно составлять $m+l-1 = (9+6-1) - (11+8-1)$ или от 14 до 18 уровней. При более коротких рядах применение многократного скользящего выравнивания нецелесообразно.

Во-вторых, многократное выравнивание следует применять, если колеблемость исходных уровней достаточно существенная, скажем, коэффициент колеблемости (см. о нем в следующей главе), хотя бы не ниже 5%. При более слабой колеблемости искажения параметра невелики и при однократном выравнивании, и 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, или в-четвертых, 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. Поэтому не нужно усреднять значения свободного члена по скользящим базам, следует в качестве свободного члена для прямой взять общую среднюю величину уровней исходного ряда, т.е. $a = \sum_{i=1}^N y_i : N$. Для параболы, экспоненты и т.д. свободный член находится расчетом от 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 таблицы 5.1. Тренд, как показано ранее, линейный, но колеблемость существенная. Сделаем 12 скользящих баз расчета среднегодового прироста по 15 уровней в каждой. Вид таблицы для расчета без помощи ЭВМ приведен в приложении 1.

Таблица 5.9

Результаты расчетов по многократному выравниванию

Скользящие базы расчета, годы	Среднегодовой прирост, ц с 1 га в год, b_i
1970 – 1984	1,25
1971 – 1985	1,29
1972 – 1986	1,29
1973 – 1987	1,42
1974 – 1988	1,66
1975 – 1989	1,77
1976 – 1990	1,68
1977 - 1991	1,54
1978 - 1992	1,46
1979 - 1993	1,46
1980 - 1994	1,36
1981 - 1995	1,24
Итого, Σ	17,42

Средний уровень: $a = \bar{y} = \frac{1332,4}{26} = 51,25$ ц с 1 га

Среднее среднегодовое изменение (прирост):

$$\bar{b} = \frac{\sum_{i=1}^{12} b_i}{12} = \frac{17,42}{12} = 1,452 \text{ ц с 1 га в год.}$$

Уравнение тренда:

$$\hat{y}_i = 51,25 + 1,452 \cdot t_i,$$

где $t_i = 0,5$ в 1983 году.

По этому уравнению в приложении 1 вычислены уровни тренда и отклонений от него.

В табл. 5.9 по скользящим базам расчета сначала среднегодовой прирост несколько возрастает, а затем уменьшается. Поскольку нет определенного направления тенденции изменения величины b_i , можно считать, что их различие - следствие колебаний уровней и небольших колебаний скорости роста урожайности, однако, в пределах единой линейной тенденции. Поэтому-то 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.

Особенно сложно оценить параметры тренда при несинусоидальных и сезонных колебаниях (см. раздел 6.3). Для каждого типа тренда необходима специальная методика, иначе параметры тренда будут искажены, а значит, и сами колебания преувеличены или наоборот. Такие методики не излагаются, насколько нам известно, ни в каких учебниках или монографиях, их нет и в программах (пакетах статистических программ) для ЭВМ. Данное пособие не позволяет по своему объему включить много таких методик, поэтому изложена одна - для линейного тренда в разделе 6.3.

В заключение данной главы на примере последнего тренда покажем, как рассчитать изложенные в главе 3 показатели динамики:

1) Абсолютное изменение.

Если тренд является линейным, то оно - главный параметр, т.е. " b " или " $\bar{b}$ " при многократном выравнивании. Если тренд

криволинейный, то абсолютное изменение - не постоянная величина. За любой период его можно вычислить, вычитая из текущего выравненного уровня базисный выравненный: Абсолютное изменение = $\bar{y}_m - \bar{y}_0$.

2) Абсолютное ускорение.

Для параболического II порядка тренда оно - главный параметр, но не забывайте, что оно равно 2 с, удвоенному квадратическому члену (ускорение - вторая производная по времени, вторая производная от ct^2 по t будет равна $2c$).

При других типах тренда ускорение за период "m" можно найти как:

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

Например, ускорение валового сбора чая в КНР в 1980 году (см. табл. 5.7) составило:

$$\bar{y}_{80} - 2\bar{y}_{79} + \bar{y}_{78} = 314 - 2 \cdot 272 + 221 = -9 \text{ тыс. тонн в год за год} \\ \left(-9 \frac{\text{тыс. т}}{\text{год}^2} \right).$$

3) Темп роста - основной параметр экспоненциального тренда "k". Для всех других типов тренда следует для вычисления темпа роста за некоторый период "m" разделить $\bar{y}_m$ на $\bar{y}_0$, то есть на выравненный уровень базисного года. Для линейного тренда (табл. 5.3) темп изменения числа занятых в народном хозяйстве России в 1996 г. по сравнению с 1990 годом составил:

$$65,7:75,3 = 0,8725 \text{ или } 87,25\%.$$

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 учесть, 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 вредных 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

Также, 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).

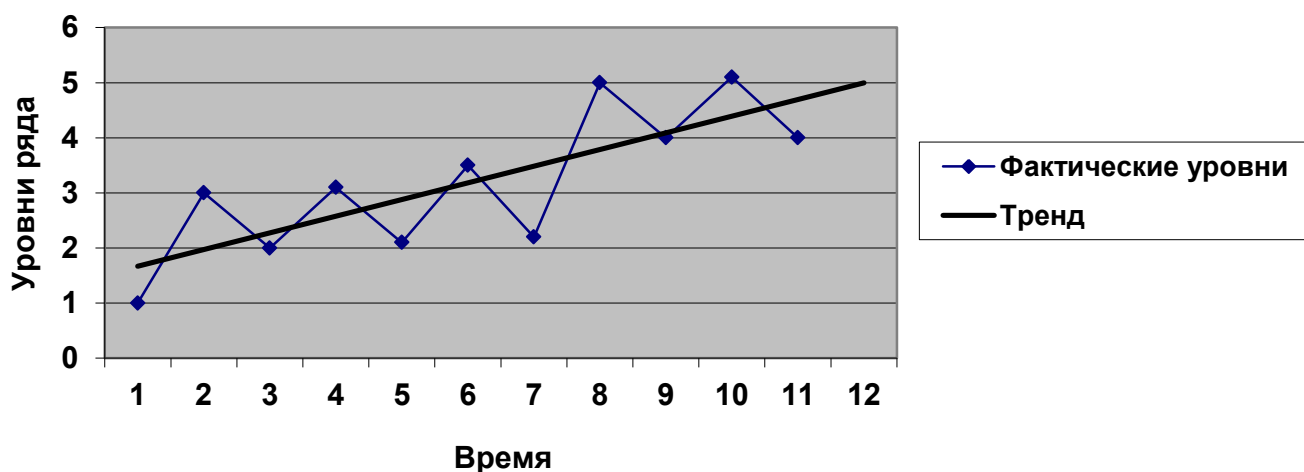


Рис. 6.1. Пилообразная колеблемость.

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 обеспечивает их надёжное прогнозирование: 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) с числом 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 могут быть как внутренними факторами системы, так и внешними. 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), в результате будет вынесено меньше питательных веществ из почвы, чем в ней образуется за год, а, следовательно, следующий урожай (второго от базы 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 затраты, 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, как элемента, пилообразных колебаний 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:

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

Числитель коэффициента – сумма произведений каждого отклонения на следующее, кроме последнего в ряду отклонения. 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 $k-1$, тем большую роль играет пилообразная составляющая в общей колеблемости изучаемого 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.

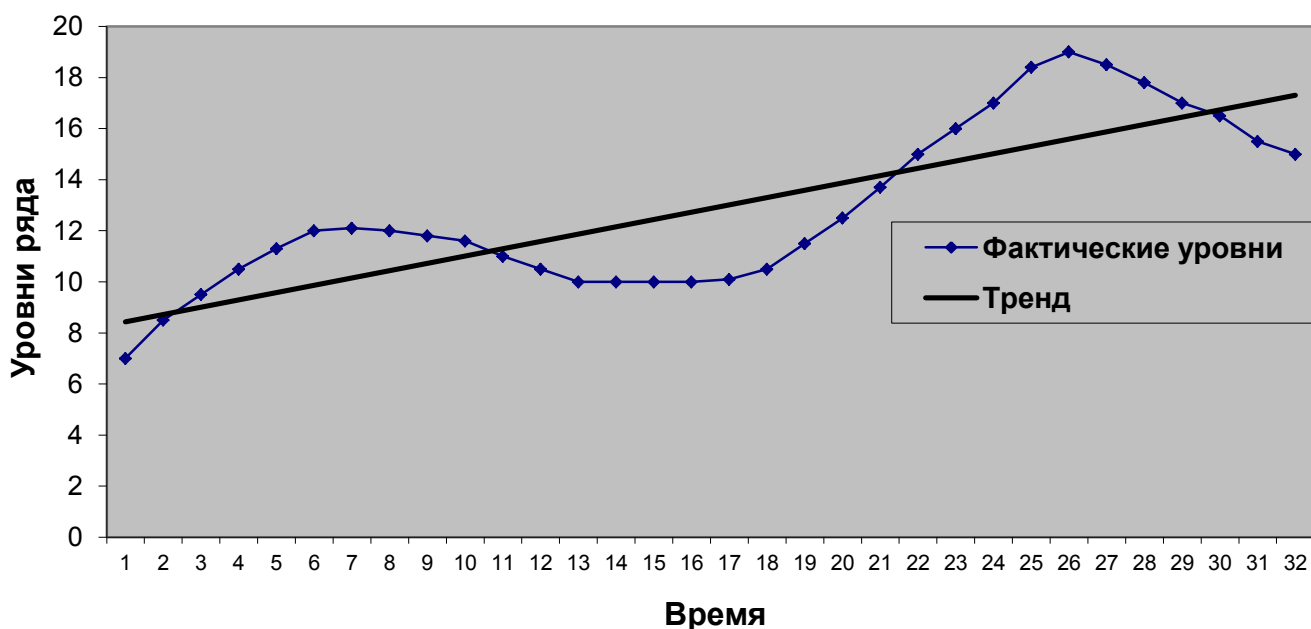


Рисунок 6.2. Циклическая, долгопериодическая колеблемость.

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 1990-91 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

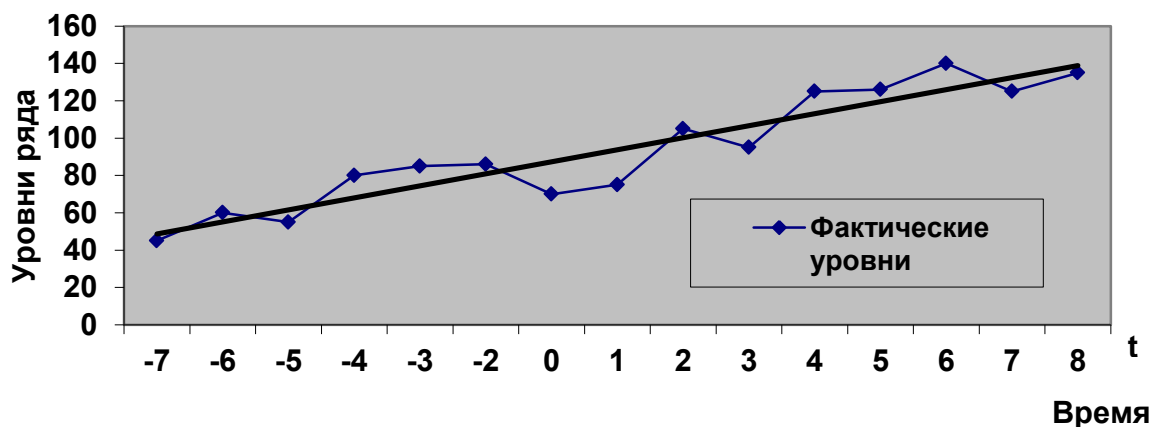
по коэффициенту автокорреляции отклонений I порядка. Если число локальных экстремумов в ряду отклонений мало, то можно предположить наличие циклической колеблемости. Поскольку отклонения одного и того же знака следуют подряд, их произведения являются положительными числами, а отрицательные произведения 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). Представим случайно распределенную во времени колеблемость на рис. 6.3.

Рисунок 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 времени, 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 слабосвязанных 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 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. Кендэл [10], 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, 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, $\tau_U^a = -0,025$.

6.2 Измерение показателей силы и интенсивности колебаний

Показатели силы и интенсивности колебаний аналогичны по построению, по форме показателям силы и интенсивности вариации признана 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 средней величиной».

6.2.1 Показатели абсолютной величины (силы) колебаний

Первый показатель - амплитуда (размах) колебаний – разность между наибольшим и наименьшим по абсолютной величине отклонениями от тренда. Например, размах колебаний объема экспорта из Японии за 1988-1995 гг. (см. табл. 5.4) составил: $5 - (-4) = 9$ млрд дол. Размах колебаний затрат условного топлива на 1 кВт·ч электроэнергии (см. табл. 5.5) составил: $14 - (-8) = 22$ г топлива на кВт·ч.

Размах колебаний урожайности зерновых культур во Франции (см. приложение 1) составил $6,6 - (-7,4) = 14$ ц/га. Показатель амплитуды колебаний характеризует лишь крайние пределы, но не среднюю силу колеблемости. Чем длиннее ряд, тем больше вероятность, 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 $a(t)$.

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

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 уровни ряда. Например, согласно данным табл. 5.6 среднее по модулю отклонение от

тренда численности населения Земли в 1950-2000 гг. может составить примерно 43,3 млн чел. Средний модуль отклонений урожайности зерновых культур от тренда во Франции по данным приложения 1 составил 2,68 ц/га.

Хотя средний модуль отклонений тренда вполне пригоден как обобщающий показатель силы колебаний за изучаемый период, но, как известно, модули имеют и существенные недостатки, в частности, с ними невозможно связать вероятностные законы распределения. Поэтому модули непригодны для прогнозирования доверительных 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 нет задачи оценки силы колебаний «вообще», о прогнозе на будущее, тогда следует вычислять и использовать обычное среднее квадратическое отклонение:

$$\sigma(t) = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (6.4)$$

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:

$$S(t) = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - p}} \quad (6.5)$$

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 в виде отношения ее абсолютных показателей к постоянной средней величине, относительной интенсивности колебаний - как отношения 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 урожайности зерновых во Франции от ее тренда в 1976 г. Абсолютное отклонение составило – 7,4 ц/га, а уровень тренда (см. приложение 1) =41,8 ц/га. Интенсивность отклонения (колебания) равна – $7,4:41,8 = -0,177$ или –17,7%. Это очень серьезный неурожай. В 1995 г. отклонение урожайности зерновых от тренда по абсолютной величине тоже было значительным: -6,2 ц/га. Но в этом году уровень тренда поднялся уже до 69,4 ц/га, поэтому интенсивность отклонения составила $-6,2:69,4 = -0,0896$ или – 8,96%, что можно уже считать не сильным, а умеренным неурожаем.

Обобщающим показателем интенсивности колебаний урожайности зерновых культур во Франции служит отношение оценки генерального среднего квадратичного отклонения уровней от тренда $S(t)$ и средней величине урожайности за весь период 1970-1995 гг., что, согласно приложению 1, составляет: $3,54 \text{ ц/га} : 51,25 \text{ ц/га} = 0,069$ или 6,9%.

Напомним, что при криволинейном тренде средний уровень не равен свободному члену уравнения тренда, так же как и при прямолинейном тренде, но при отчете периодов времени от начала, а не от середины ряда. В этих случаях делить обобщающий показатель силы колебаний $S(t)$ нужно не на свободный член уравнения, а на средний уровень изучаемого показателя. Например, интенсивность колебаний расхода условного топлива на выработку 1 кВт·ч электроэнергии (см. табл. 5.5) составляет $\sqrt{\frac{382}{7-2}} : \frac{2555}{7} = 8,74 : 365 = 0,0239$ или 2,39%. Колеблемость очень слабая. Аналогично коэффициенту пространственной вариации, отношение среднего квадратического отклонения от тренда к среднему уровню временного ряда называют *коэффициентом колеблемости*, который мы обозначаем, для отличия от коэффициента пространственной вариации, как $V(t)$. Его формула

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

-для оценки генеральной величины и прогнозов или

$$v(t) = \frac{\sigma(t)}{y}$$

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

Величина коэффициента колеблемости играет также 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. Например, наверняка часть колеблемости численности населения Земли (см. табл. 5.6) на самом деле объясняется тем, что «истинная» тенденция роста населения не являлась за 1950-2000 гг. строго экспоненциальной.

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.

Непосредственно связанные со сменой температуры колебания имеют характер плавных циклов, без скачкообразных изменений уровней, т.е. так, как меняется в течение года сама температура воздуха. Опосредованные же сезонные колебания могут иметь резкие скачки уровней, несколько максимумов и несколько минимумов за год. Это различие существенно для выбора статистической модели сезонной колеблемости.

Для правильного измерения сезонных колебаний очень важно, чтобы тренд был рассчитан правильно, что, в свою очередь, требует учета сезонных колебаний (см. разд. 5.5).

6.3.1 Плавные, синусоидальные колебания при несущественности тренда

Поскольку колебания такого рода связаны с сезонным ходом температуры воздуха, целесообразно рассмотреть колебания самой этой температуры (табл. 6.1).

Таблица 6.1.

Динамика средних месячных температур

Месяц, №i	1995г., y_{i95}	1996г. y_{i96}	1997г. y_{i97}	В среднем за 1988- 1997гг. $\bar{y}_i$	В % к среднегодов ым $\bar{y}_i : \bar{y}$	$(\bar{y}_i - \bar{y})^2$
Январь	-4,0	-6,7	-5,2	-3,9	-65,7	96
Февраль	0,0	-9,5	-3,4	-4,3	-72,5	104
Март	+1,1	-2,4	-0,8	0,0	0	35
Апрель	+5,0	+4,0	+3,1	+5,2	87,7	0,5
Май	+11,3	+10,7	+9,7	+11,2	188,9	28
Июнь	+19,2	+15,4	+17,3	+16,2	273,2	106
Июль	+16,8	+16,6	+19,8	+18,4	310,3	156
Август	+17,1	+18,2	+19,1	+16,8	283,3	119
Сентябрь	+12,5	+9,8	+10,7	+11,2	188,9	28
Октябрь	+8,4	+7,0	+4,3	+5,0	84,3	1
Ноябрь	-1,6	+4,2	-0,4	-0,7	-11,8	44
Декабрь	-8,3	-5,0	-5,2	-3,9	-65,8	96
Средняя за год	+6,46	+5,19	+5,75	+5,93	100	813

в Ленинграде - Санкт-Петербурге

Данные табл. 6.1 позволяет сделать ряд важных выводов для методики изучения сезонных колебаний:

1) температура воздуха в одноименные месяцы разных лет неодинакова. Самым холодным является то январь, то февраль, то декабрь; самым теплым бывает июнь, июль или август. Вывод: в уровнях отдельного года отражены не только закономерные сезонные колебания для климата

данного города, но и случайные отклонения погоды в отдельные годы от климатической нормы. А значит, случайные колебания будут (были!) присущи и всем экономическим показателям этих лет, связанным с изменением температуры воздуха.

2) средняя температура воздуха за 1995-1997 гг. совпадает со средней за 1988-1997 гг., что означает отсутствие существенной общей тенденции на протяжении 10 лет (более подробные исследования динамики температуры воздуха в Ленинграде – Санкт-Петербурге за 40 лет показало, что тенденция существует, но слабая: среднегодовой абсолютный прирост температуры составил $0,0255^0$ в год, что на протяжении до 10 лет, конечно, несущественно.)

3) по данным одного только года нельзя точно измерить сезонные колебания, так как они будут смешаны со случайными колебаниями. Чтобы измерить сезонные колебания, необходимо усреднить уровни каждого месяца за достаточное число смежных лет, чтобы случайные колебания уровней, в основном, взаимопогасились. В данном примере усреднены месячные температуры за 10 лет. Часто в учебниках по статистике для экономии места приводят при анализе сезонных колебаний среднемесячные уровни за 2-3 года, что, конечно, совершенно недостаточно для взаимопогашения случайных колебаний, особенностей отдельных лет.

В чем же состоит измерение сезонных колебаний по усредненным за ряд лет данным? Традиционным показателем служат так называемые *индексы сезонности*, под именем которых понимают *отношения уровней каждого месяца к среднемесячному уровню за весь год*. Обычно их выражают в процентах. Например, средняя температура июля составляет в Ленинграде-Санкт-Петербурге 310% к средней температуре за год. Отрицательные индексы в данном примере неинтерпретируемы, так как температура исчислялась от условного нуля, а не от абсолютного нуля (в шкале Кельвина).

Обобщающим абсолютным показателем силы сезонных колебаний служит среднее квадратическое отклонение средних температур месяцев от среднегодовой температуры:

$$\sigma_{из} = \sqrt{\frac{\sum_{i=1}^{12} (x_i - \bar{y})^2}{12}} = \sqrt{\frac{813}{12}} = 8,23^0$$

Эта величина – один из основных показателей климата данной территории. Например, в регионах с так называемым морским климатом, на островах, побережье океанов сезонные колебания температур намного слабее, чем в глубине материков, в регионах с континентальным климатом, где колебания гораздо сильнее. Например, на северо-западе Великобритании $\sigma \approx 3^0$, а в Узбекистане (г. Бухара) $\sigma \approx 12^0$.

Относительный показатель интенсивности колебаний для температур в Петербурге не пригоден по уже указанной причине, как для всех рядов, имеющих положительные и отрицательные уровни.

Сезонные колебания можно изобразить графически двумя способами: в прямоугольных координатах и полярных координатах. На рис 6.4 хорошо видно, что в разные года продолжительность лета и зимы разная.

Выше 15^0 – дни считаются летними, ниже 0^0 - зимними.

Графическое изображение сезонных колебаний в полярных координатах покажем на примере другого вида колебаний (рис. 6.4.).

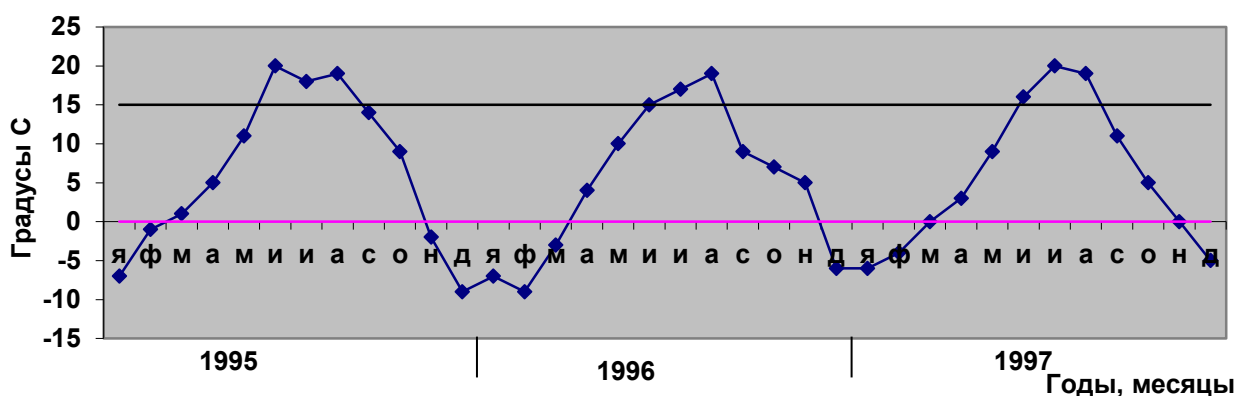


Рис. 6.4. Колебания месячной температуры воздуха в Петербурге за 1995-1997 гг

6.3.2Сезонные колебания, не имеющие синусоидальной формы, при наличии существенной тенденции

В качестве примера такого вида сезонных колебаний рассмотрим динамику затрат труда (в тыс. часов) ежемесячно в сельхозпредприятии при его развитии с ростом числа работников на протяжении трех лет. (Больше, чем за три года число месячных уровней трудно разместить в таблице). Сезонных «пиков» затрат труда два: связанные с весенними полевыми работами в мае-начале июня и связанный с уборкой урожая в августе-сентябре, часто и в октябре.

Таблица 6.2.

Расчет параметров тренда при асимметричных сезонных колебаниях

Год	квартал	Уровни, y_i	t_i	$y_i t_i$	Первич. тренд, $\hat{y}_i$	Коррект. тренд $\hat{y}_{i,кор}$	$i_{сез}$
I	1	20	-4	-80	22	29	0,690
	2	26	-3	-78	27	32	0,812
	3	28	-2	-56	32	36	0,778
	4	60	-1	-60	38	39	1,538
II	1	30	0	0	43	43	0,698
	2	38	1	38	48	46	0,826
	3	42	2	84	53	50	0,840
	4	100	3	300	58	53	1,887
III	1	40	4	160	63	57	0,702
Итого		384	0	308	384	385	

$$a=42,67; b = \frac{308}{60} = 5,13$$

Первичный тренд: $\hat{y} = 42,67 + 5,13t_i$, где $t=0$ в 1 квартале II года.

При наличии сезонных колебаний, не имеющих синусоидального характера, особенно для рядов, имеющих резкие пик в первые или в последние месяцы года, методики расчета параметров тренда, описанные ранее (см. гл. 5), оказываются недостаточно пригодными, особенно если ряд не очень длинный и нельзя применить многократное выравнивание. Рассмотрим, например, ряд квартальных уровней за два года и один квартал, так как необходимо, как уже подчеркивалось в гл. 5, чтобы начало и конец ряда (база выравнивания) приходились на одну и ту же фазу цикла, или часть года (квартал, месяц). Далее, положим, что резкий «пик» уровней приходится ежегодно на 4-ый квартал.

Резко выделяющийся пик уровней приходится на периоды со значениями t_i , равными -1 и 3 , в среднем, положительными. Наоборот, минимальные уровни первых кварталов приходятся на значение t_i , равные соответственно -4 ; 0 ; 4 , в среднем нулевые веса; низкие значения уровней вторых кварталов приходятся на значение t_i , равные соответственно -3 и 1 , в среднем отрицательные. Значения уровней третьих кварталов, также более низкие, чем в среднем за год, приходятся на значения t_i , равные: -2 и 2 , в среднем нулевые. Итак, в целом высокие значения уровней входят в расчет

параметра b с положительными весами, а остальные, низкие, уровни – с нулевыми или отрицательными весами. Следовательно, параметр b (средний годовой прирост) завышается за счет асимметричного расположения пика уровней в году. Не помогло даже соблюдение правила об окончании ряда (базы расчета параметров) на той же фазе (квартале), как и начало ряда.

Чтобы скорректировать расчет, необходимо «снять» из числителя параметра b указанное неравенство, т.е. превышение положительных произведений отклонений от тренда на веса по четырем кварталам над отрицательными произведениями отклонений от тренда по остальным кварталам.

Средний вес «пиковых» уровней равен 1, следовательно, положительное превышение за счет асимметрии весов, равно $(100+60) \cdot 1 = 160$. Нулевые произведения не дают искажений, а отрицательные произведения дают уровни 2 кварталов, их средний вес равен -1 , произведение равно $(26+38) \cdot (-1) = -64$. Избыток положительного искажения над отрицательным составил: $160 - 64 = 96$. Эту величину следует исключить из числителя при расчете параметра b . В результате имеем:

$$b_{\text{кор}} = (308-96)/60 = 3,533$$

Итак, скорректированное уравнение тренда имеет вид:

$$\hat{y}_{\text{корр.}} = 42,67 + 3,53 t_i, t = 0 \text{ в 1 квартале года II.}$$

Таким образом, преувеличение среднего прироста уровней за квартал за счет несимметричного распределения сезонных пиков уровней составляло $5,13 - 3,53 : 3,53 = 0,45$ или 45%. *Индекс сезонности* для 1-го квартала I года при первичном тренде составил бы $20 : 22 = 0,909$, а при скорректированном тренде 0,690, т.е. величина сезонного снижения уровня составила не 9,1% а 31%, т.е. втрое больше. Следовательно, без корректировки тренда вся картина динамики была бы сильно искажена.

К сожалению, еще более сложные методики корректировки для других типов тренда не могут быть здесь изложены, тем более что многие из них еще предстоит разработать и ввести в пакеты статистических программ ЭВМ.

При длительном временном ряде и расположении пика сезонных колебаний в середине года либо примерно на равном расстоянии от середины года, достаточно выполнить многократное скользящее выравнивание. Рассмотрим подробно измерение сезонных колебаний затрат труда на прогрессивно развивающемся сельскохозяйственном предприятии за три года (табл. 6.3).

Таблица 6.3.

Сезонные колебания затрат труда.

Год	месяц, i	Уровни, тыс. час, y_{ij}	Тренд $\bar{y}_i$	$i_{сез\ ij}$	$\bar{y}_i \cdot \bar{i}_{сезi}$	$y_i - \bar{y}_i \cdot \bar{i}_{сезi}$ $= u_{сл}$	$u_{сл}^2$	$\bar{y}_i - \bar{y} \cdot \bar{i}_{сезi}$ $= u_{сез}$	$u_{сез}^2$
I	Я	17	40	0,425	17	0		23	529
	Ф	16	42	0,381	16	0		26	676
	М	22	43	0,512	22	0		21	441
	А	35	44	0,795	34	1		10	100
	М	76	45	1,689	72	4		-27	729
	И	67	46	1,457	67	0		-21	441
	И	57	47	1,213	57	0		-10	100
	А	78	48	1,625	80	-2		-32	1024
	С	84	49	1,714	78	6		-29	841
	О	64	50	1,280	60	4		-10	100
	Н	35	51	0,686	37	-2		14	196
	Д	24	52	0,462	23	1		29	841
II	Я	20	53	0,377	23	-3		30	900
	Ф	19	55	0,345	21	-2		34	1156
	М	29	56	0,518	29	0		27	729
	А	42	57	0,737	44	-2		13	169
	М	96	58	1,655	93	3		-35	1225
	И	91	59	1,542	86	5		-27	729
	И	71	60	1,183	72	-1		-12	144
	А	102	61	1,672	101	1		-40	1600
	С	102	62	1,645	99	3		-37	1369
	О	76	63	1,206	75	1		-12	144
	Н	48	64	0,750	47	1		17	289
	Д	29	65	0,446	30	-1		35	1225
III	Я	33	67	0,493	29	4		38	1444
	Ф	31	68	0,456	27	4		41	1681
	М	36	64	0,522	35	1		34	1156
	А	56	70	0,800	54	2		16	256
	М	107	71	1,507	114	-7		-43	1849
	И	104	72	1,444	106	-2		-34	1156
	И	92	73	1,260	88	4		-15	225
	А	129	74	1,743	123	6		-49	2401
	С	112	75	1,493	120	-8		-45	2025
	О	86	70	1,132	91	-5		-15	225
	Н	59	77	0,766	56	3		21	441
	Д	40	79	0,506	37	3		42	1764
IV	Я	35	80	0,438	34	1		46	2116
Σ		2220	2221	-	2197	23	393	+24	32436

После вычисления тренда и его уровней за все месяцы вычисляются отношения фактических уровней к уровням тренда, т.е. *индексы сезонности*.

Однако в них включены и случайные колебания. Чтобы очистить индексы сезонных колебаний от случайности, нужно их усреднить за несколько (лучше 10 и более) лет. В учебном примере у нас только три года (для января – четыре), что на самом деле недостаточно для отделения сезонных, типичных колебаний от случайных особенностей процесса в разные годы. Вычисляем средние индексы сезонных колебаний:

$i_{\text{сез.корр.}}$	Месяц	$i_{\text{сез}}$
0,429	январь	$(0,425 + 0,377 + 0,493 + 0,438) : 4 = 0,433$ и так
далее		
0,390	февраль	0,394
0,512	март	0,517
0,769	апрель	0,777
1,601	май	1,617
1,466	июнь	1,481
1,207	июль	1,219
1,663	август	1,680
1,601	сентябрь	1,617
1,194	октябрь	1,206
0,727	ноябрь	0,734
0,466	декабрь	0,471

$$\sum_{i=1}^{12} = 12,026$$

Сумма индексов составила 12,146, хотя средний индекс должен быть равен единице. Следует корректировать индексы на пропорциональную величину, т.е. от больших отнять больше, от меньших – меньше, примерно на 0,01 от общей величины. Корректированные индексы запишем слева от названий месяцев.

Далее, умножая уровень тренда на корректированные средние индексы, находим уровни с учетом тренда и сезонных колебаний, но, исключая случайные колебания: $\bar{f}_i i_{\text{сез}}$ округлены в табл. 6.3 до целых. То, что $\sum_{i=1}^{37} \bar{f}_i i_{\text{сез}}$

меньше $\sum_{i=1}^{37} \bar{f}_i$ не является недостатком расчета: дело в «лишнем» январе, уровень которого с учетом сезонного колебания в среднем за 3 года ниже тренда на 30, в результате даже с учетом этого остается небольшой избыток $\sum_{i=1}^{37} \bar{f}_i i_{\text{сез}}$, объясняемый округлением. Избыток на 6 при сумме уровней 2220, разумеется, несущественен.

Далее вычисляем отклонения фактических уровней от $\bar{f}_i i_{\text{сез}}$, т.е. случайные колебания и их квадраты, с целью вычисления среднего квадратического отклонения уровней затрат труда от «модели», учитывающей тренд и средние сезонные колебания:

$$S(t)_{случ.} = \sqrt{\frac{393}{37-2-11}} = 4,05 \text{ тыс. часов}$$

В знаменателе стоит число степеней свободы случайной колеблемости: вычитается из числа уровней 37 две степени свободы линейного тренда и 11 степеней свободы месячных колебаний (двенадцатый индекс сезонности – величина несвободная, т.к. задана их сумма за год, равная 12 целым). Коэффициент случайной колеблемости составил $4,05 : 60 = 0,0675$ или 6,75%. Колеблемость слабая. Силу самих же сезонных колебаний можно оценить по их среднему квадратическому колебанию:

$$S(t)_{сез} = \sqrt{\frac{\sum_{i=1}^{37} (\bar{y}_i - \bar{y}_{сез i})^2}{11 \cdot 3}} = \sqrt{\frac{32436}{11 \cdot 3}} = 31,35 \text{ тыс. час.}$$

Сезонные колебания за год имели 11 степеней свободы вариации, но в ряду отклонений $\bar{y}_i - \bar{y}_{сез i}$, повторяются три раза, так что правильно будет считать всего 33 квадрата сезонных колебаний и делить сумму квадратов на 33, иначе получится нереально большая величина. Вопрос о степенях свободы вариации при сезонных колебаниях требует дальнейшего исследования. Коэффициент сезонной колеблемости $V(t)_{сез} = 31,35/60 = 0,522$ или 52,2%. Сезонная колеблемость сильная.

Графическое изображение сезонных колебаний затрат труда в сельскохозяйственном предприятии построим в полярных координатах (рис. 6.5), т.е. каждый месяц в окружности занимает $360^0 : 12 = 30^0$. Радиус равен 1, а точки откладываются от центра на величину $\rho = \bar{y}_{сез i}$.

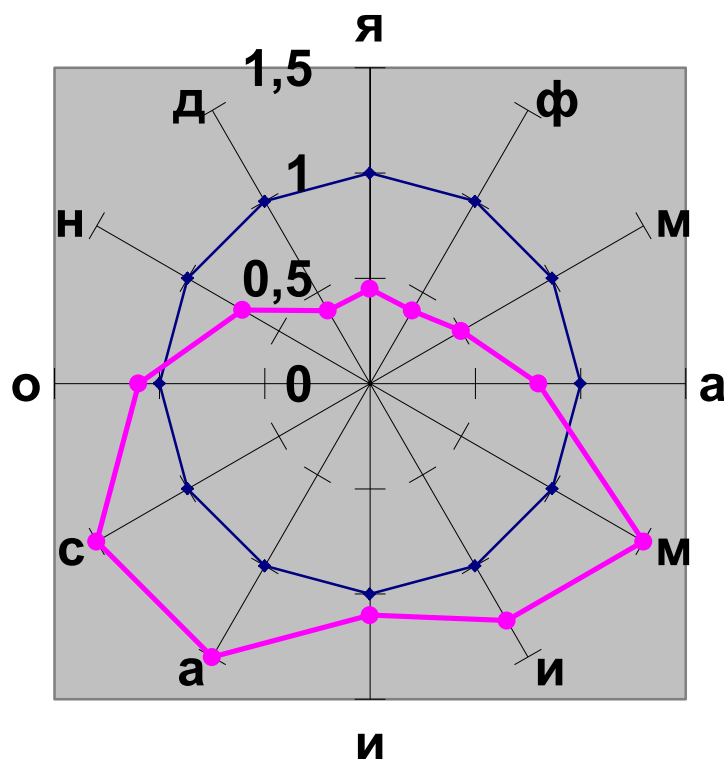


Рис. 6.5. Сезонные колебания затрат труда в сельскохозяйственном предприятии.

При отсутствии сезонности фигура I (см. рис. 6.5) лежала бы точно по окружности.

6.3.3 Представление синусоидальных колебаний в форме тригонометрического уравнения Фурье

Выдающийся французский математик Жан Батист Жозеф Фурье (1768-1830) предложил метод преобразования периодических функций в ряд тригонометрических уравнений, называемых *гармониками*. Этот метод подходит для аналитического выражения сезонных колебаний, имеющих синусоидальную форму. Исходным рядом для преобразования Фурье лучше всего принять не первичный ряд за несколько лет, а усредненный ряд месячных уровней, в котором исключен тренд и (или) в основном погашены случайные колебания. Рассмотрим сезонные колебания среднего по ферме надоя молока на 1 корову (табл. 6.4).

Таблица 6.4.

Месяц i	Надой $\bar{y}_i$ кг/гол	t_i , градусо в	$\cos t_i$	$\sin t_i$	$\bar{y}_i \cos t_i$	$\bar{y}_i \sin t_i$	$\mathcal{E}_i$	$\bar{y}_i - \mathcal{E}_i$
Январь	230	0	1	0	230	0	229	1
Февраль	260	30	0,866	0,5	225	130	264	-4
Март	315	60	0,5	0,866	157,5	273	311	4
Апрель	352	90	0	1	0	352	357	-5
Май	392	120	-0,5	0,866	-196	339	391	1
Июнь	403	150	-0,866	0,5	-349	201,5	404	-1
Июль	398	180	-1	0	-398	0	391	7
Август	352	210	-0,866	-0,5	-305	-176	356	-4
Сентябрь	308	240	-0,5	-0,866	-154	-267	309	-1
Октябрь	262	270	0	-1	0	-262	263	-1
Ноябрь	225	300	0,5	-0,866	112,5	-195	229	-4
Декабрь	223	330	0,866	-0,5	193	-111,5	216	7
Σ	3720	-	0	0	-484	284	3720	0

Преобразование сезонных колебаний в ряд Фурье

Среднемесячный надой = $3720 \text{ кг} / 12 = 310 \text{ кг}$ на 1 корову

Тригонометрическое уравнение ряда Фурье для его первой гармоники, которой мы здесь и ограничимся, имеет форму:

$$\hat{\epsilon}_i = a + b_1 \cos t_i + b_2 \sin t_i$$

$$\text{где } a = \bar{y}; b_1 = \frac{\sum_{i=1}^{12} y_i \cos t_i}{6}; b_2 = \frac{\sum_{i=1}^{12} y_i \sin t_i}{6}$$

Смысл уравнения в том, что без сезонных колебаний все уровни были бы равны среднемесячному, т.е. $\bar{y}$; колебания же в равной мере разнесены на $\sin t$ и $\cos t$. В первом квадранте (т.е. от января до апреля) косинус является положительной величиной и снижается от 1 до 0, синус тоже положителен и возрастает от 0 до 1. Во втором квадранте (апрель-июль) косинус отрицателен и снижается от 0 до -1, синус положителен и снижается от 1 до 0. В третьем квадранте (июль-октябрь) косинус отрицателен, но возрастает от -1 до 0, а синус снижается от 0 до -1. В четвертом квадранте косинус возрастает от 0 до 1 (к декабрю до +0,866), а синус возрастает от -1 до 0 (к декабрю до -0,5). Цикл завершается новым январем. За счет комбинации изменений косинуса и синуса при разных значениях параметров b_1 и b_2 , удастся отобразить, как показывает табл. 6.4 (графа $\hat{\epsilon}_i$), любое синусоидальное колебание уровней временного ряда. Имеем: $b_1 = -484/6 = -80,7$; $b_2 = 284/6 = 47,3$. Уравнение сезонных колебаний продуктивности коров имеет вид:

$$\hat{\epsilon}_i = 310 - 80,7 \cos t_i + 47,3 \sin t_i,$$

где $t_i = 0^\circ$ в январе, а месяц = 30° дуги.

Отклонения фактических уровней (но усредненных за ряд лет) от расчетных по ряду Фурье очень малы: максимальное отклонение 7, среднее (по модулю) 3,33, что составляет лишь 1,07%. Такая точность вполне достаточна для прогнозов и других расчетов. Если же отклонения оказались значительными, следует на основании ряда отклонений повторить расчет, т.е. рассчитать вторую гармонику, и тогда окончательные уровни модели (ряда Фурье) будут представлять собой сумму всех гармоник:

$$\hat{\epsilon}_i = a + \sum_{k=1}^m b_{1k} \cos t_i + \sum_{k=1}^m b_{2k} \sin t_i,$$

где m - число гармоник,

k -номер гармоники.

Однако, если колебания явно не имеют синусоидальной формы, то требуется много гармоник, расчет становится трудоемким, и гораздо проще применить метод, описанный ранее в разд. 6.3.2.

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 надежного вычисления меры колеблемости необходимо как минимум 7-9 уровней первичного временного ряда, а для вычисления тренда по этим мерам колеблемости - опять 7-9 таких же 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, чтоб 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 данной ее особенности измерить тренд колеблемости по ряду отдельных отрезков времени сложно. При длине первичного ряда в 15-20 уровней получается всего два значения $S(t)$, чего явно не хватает для расчета тренда.

Не вполне корректными с математической точки зрения являются расчет скользящих показателей колеблемости со сдвигом в один период времени и последующее их аналитическое выравнивание. Конечно, скользящие показатели уже зависят друг от друга, но выявить общую тенденцию колебаний и приблизительно измерить тренд $S(t)$ все же возможно. Покажем применение этого метода на примере временного ряда урожайности зерновых культур во Франции (см. разд. 5.10. В приложении 1 вычислены отклонения уровней от тренда, с которых и начинается измерение тренда среднего квадратического отклонения (табл. 6.5).

Скользящие показатели колеблемости $S(t)_i$ будем рассчитывать по 11-летним подпериодам, т.е. первый за 1970-1980 гг., второй за 1971-1981 гг. и т.д. Первая величина $S(t)$ будет относиться к середине подпериода, т.е. 1975 г. и т.д., последняя скользящая средняя за 1985-1995 гг. относится к 1990 году. Итого получаем 16 скользящих значений показателей колеблемости, которые и выравниваем по уравнению прямой.

Тренд среднего квадратического отклонения уровней урожайности от их тренда имеет вид:

$$\hat{S}(t) = 3,42 - 0,1235 \cdot t_i; t = 0,5 \text{ в } 1983 \text{ году}$$

Таким образом, имеется тенденция снижения силы колебаний урожайности зерновых культур во Франции за рассмотренный период. 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 урожайности, т.е. распространении выборочной оценки на генеральную совокупность периодов времени.

Для указанной цели, придется использовать излагаемую только в гл. 7 методику вероятностных оценок параметров.

Средняя ошибка репрезентативности среднегодового изменения $b_{S(t)}$, т.е.

$$m_{bs(t)} = \frac{S(t)_{S(t)}}{\sqrt{\sum_{i=1}^{16} t_i^2}},$$

Таблица 6.5

Вычисление тренда показателя колеблемости

Год	U_i	U_i^2	Скольз ящие $S(t)_i$	t_i	$S(t)_i t_i$	$\mathcal{E}(t)_i$	$U_{S(t)_i}$	$U_{S(t)_i}^2$
1970	0,6	0,36	-	-	-			
1971	4,2	17,64	-	-	-			
1972	5,7	32,49	-	-	-			
1973	6,6	43,56	-	-	-			
1974	2,9	8,41	-	-	-			
1975	-3,4	11,56	4,48	-7,5	-33,6	4,34	0,12	0,0144
1976	-7,4	54,76	4,52	-6,5	-29,4	4,22	0,30	0,0900
1977	-2,8	7,84	4,30	-5,5	-23,6	4,10	0,20	0,0400
1978	1,6	2,56	3,95	-4,5	-17,8	3,97	-0,02	0,0004
1979	-1,0	1,00	3,89	-3,5	-13,6	3,85	0,04	0,0016
1980	0,9	0,81	3,85	-2,5	-9,6	3,73	0,12	0,0144
1981	-1,8	3,24	3,84	-1,5	-4,6	3,60	0,24	0,0576
1982	-0,4	0,16	2,95	-0,5	-1,5	3,48	-0,53	0,2809
1983	-2,5	6,25	2,85	0,5	1,4	3,35	-0,50	0,2500
1984	6,3	39,69	2,80	1,5	4,2	3,23	-0,43	0,1849
1985	2,4	5,76	2,86	2,5	7,2	3,11	-0,25	0,0625
1986	-3,3	10,89	2,91	3,5	10,2	2,98	-0,07	0,0049
1987	-0,8	0,64	2,85	4,5	12,8	2,86	-0,01	0,0001
1988	1,6	2,56	2,88	5,5	15,8	2,74	0,14	0,0196
1989	0,3	0,09	2,87	6,5	18,7	2,61	0,26	0,0676
1990	-1,3	1,69	2,85	7,5	21,4	2,44	0,36	0,1296
1991	1,8	3,24	-	-	-			
1992	0,1	0,01	-	-	-			
1993	-1,4	1,96	-	-	-			
1994	-2,4	5,76	-	-	-			
1995	-6,2	38,44	-	-	-			
Σ	-	301,37	54,65	0	-42,0	54,62	-	1,2185

$$\overline{S(t)} = \frac{54,65}{16} = 3,42; b_{S(t)} = \frac{-42,0}{340} = -0,1235 \text{ ц/га в год.}$$

Здесь в числителе стоит величина среднего квадратического отклонения скользящих значений $S(t)_i$ от их трендовых значений $\hat{S}(t)_i$ (вторая справа графа в табл. 6.5).

Имеем:

$$S(t)_{S(t)} = \sqrt{\frac{\sum_{i=1}^{16} U_{S(t)}^2}{16-1}} = \sqrt{\frac{1,2185}{15}} = 0,285$$

$$m_{b_{S(t)}} = \frac{S(t)_{S(t)}}{\sqrt{\sum_{i=1}^{16} t_i^2}} = \frac{0,285}{\sqrt{340}} = 0,0542$$

Критерий Стьюдента, т.е. отношение $\frac{b_{S(t)}}{m_{b_{S(t)}}} = \frac{0,1235}{0,0542} = 2,28$. Табличное

значение критерия Стьюдента при 15 степенях свободы вариации и значимости 0,05 составляет 2,13. Фактическое значение критерия больше табличного, следовательно, можно считать достаточно надежно установленным уменьшение колебаний урожайности зерновых культур во Франции за 1970-1995 гг. (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 один интервал времени уже был ранее 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:

$$r_{U_i U_{i+m}} = \frac{\sum_{i=1}^{n-m} U_i \cdot U_{i+m}}{\frac{\sum_{i=1}^{m-1} U_i^2}{2} + \sum_{i=m}^{n-m} U_i^2 + \frac{\sum_{i=n-m+1}^n U_i^2}{2}}$$

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

Регион	Тренд	Коэффициенты автокорреляции с лагами							
		1 год	2 года	3 года	4 года	5 лет	6 лет	7 лет	8 лет
Северный		-0,05	-0,07	-0,33	-0,11	0,00	0,01	0,08	0,08
Центрально-Черноземный	$17,3-0,08t+0,0071t^2$	0,20	0,18	0,04	0,09	0,12	-0,40	-0,19	0,70
Поволжский	$12,26-0,03t+0,0031t^2$	-0,35	-0,05	0,01	0,12	0,03	-0,31	0,24	0,27

($m-1$) квадратов входят с половинным весом, в сравнении со средними отклонениями. Рассмотрим пример расчета коэффициентов автокорреляции отклонений от тренда и их значения (табл. 6.6.).

Таблица 6.6.

**Тренды и коэффициенты автокорреляции отклонений от них.
Урожайность зерновых культур**

Авторы расчетов дают следующую интерпретацию серий коэффициентов автокорреляции: по Северному региону: «смешанный тип динамики колебаний, при котором какая-либо закономерность визуально не просматривается».

Мы считаем полезным добавить, что по Северному региону семь коэффициентов из восьми незначимо отличны от нуля, это говорит об отсутствии каких-либо циклов, о случайном распределении отклонений во времени.

По Центрально-Черноземному региону: «квазипериодические волны - чередование подъемов и спадов колебаний урожайности относительно тренда, различных по продолжительности».

Относительно Поволжского региона: «маятниковая» колеблемость, характеризуется последовательным чередованием подъемов и спадов колебаний урожайности относительно тренда. По нашему мнению, можно добавить, что чистой маятниковой колеблемости здесь нет, так как наблюдается и по два отклонения одного знака подряд; есть, видимо, смесь маятниковой и случайно распределенной колеблемости.

Строго циклическая колеблемость, например сезонная, в рядах коэффициентов автокорреляции отклонений от трендов проявится, как волнообразные изменения значений этих коэффициентов с алгебраическими минимумами при лагах величиной в 0,5; 1,5 и т.д. длины цикла и алгебраическими максимумами при лагах величиной в целое число длительности цикла.

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 - 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, на 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 желаем 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, максимальных 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 можно узнать 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, т.е. 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 численности занятых в народном хозяйстве России за 1990-1996 гг. (см. рис. 4.1 и табл. 5.3)

Тренд имеет вид:

$$\varphi_i = 70,5 - 1,614t_i \text{ млн чел.},$$

где $t_i=0$ в 1993 г., среднее квадратическое отклонение уровней от тренда $S(t)=0,2864$ млн чел.

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.

$$m_b = \frac{0,2864}{\sqrt{28}} = 0,0541$$

Отношение среднегодового изменения к его средней ошибке - это *t-критерий Стьюдента*:

$$t = \left| \frac{b}{m_b} \right| = \left| \frac{-1,614}{0,0541} \right| = 29,8$$

Величину критерия сравниваем с табличной величиной критерия Стьюдента для $7-2=5$ степеней свободы, которая для значимости (вероятности нулевой гипотезы) $0,05$ равна $2,57$, а для значимости $0,01$ она равна $4,07$. Фактическая величина критерия много больше табличных, следовательно, вероятность нулевой гипотезы (о равенстве параметра b нулю) чрезвычайно мала. Достоверно известно, что тренд существовал, и что численность работников народного хозяйства снижалась 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.

Годы	Темпер. y_i	Годы	Темпер. y_i	Годы	Темпер. y_i	Годы	Темпер. y_i
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
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:

$$\bar{t}_i = 5,183 + 0,02554t_i; t = 0 \text{ в } 1977 \text{ г.}$$

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 ошибки, 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, уточнить 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:

$$m_c = \frac{S(t)}{\sqrt{t_i^4 - \overline{t^2}^2}}$$

Под корнем, при условии отсчета номеров периодов (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:

$$\overline{t_i^4} - \left[\overline{t^2} \right]^2$$

Здесь черта над скобками – знак средних величин.

Рассмотрим пример по данным, представленным на рис. 4.2, – динамика экспорта Японии в 1988 – 1995 гг., имеющая параболический тренд. Его уравнение имеет вид:

$$\hat{y}_i = 323,2 + 25,2t_i + 2,40t_i^2.$$

Проверим, надежно ли отличие от нуля параметра c , половины ускорения. Колеблемость уровней экспорта измеряется величиной

$S(t) = \sqrt{\frac{67}{8-2}} = 3,66$. Находим необходимые для расчета ошибки параметра величины при изменении периодов от середины ряда при $n = 8$. Имеем:

$$\sum_{i=1}^8 t_i^2 = 42; \overline{t^2} = \frac{42}{8} = 5,25; \left(\overline{t^2} \right)^2 = 5,25^2 = 27,56$$

$$\sum_{i=1}^8 t_i^4 = 388,5; \overline{t^4} = \frac{388,5}{8} = 48,56.$$

Имеем:

$$m_c = \frac{3,66}{\sqrt{48,56 - 27,56}} = 0,7987 \approx 0,8$$

Критерий Стьюдента равен отношению $\frac{c}{m_c} = \frac{2,4}{0,8} = 3,0$. Табличное значение критерия при пяти степенях свободы составляет 2,57. Таким образом, отличие ускорения роста экспорта Японии от нуля за 1988-1995 гг. установлено с надежностью, большей, чем 0,95.

Для оценки основного параметра экспоненциального тренда - среднего коэффициента изменения уровней k - целесообразнее всего применить предложенную Е.М.Четыркиным [18, с. 173-174] 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 народонаселения Земли по десятилетиям 1950-2000 гг. (см. рис. 4.3 и табл. 5.6). Тренд имеет вид:

$$\phi_i = 4004 \cdot 1,195^t$$

В логарифмическом виде:

$$\ln \phi_i = 8,295 + 0,1783t_i$$

Дополнительно вычисляем отклонения логарифмов уровней от логарифмов тренда (табл. 7.2).

Таблица 7.2.

Определение отклонений логарифмов уровней от логарифмов тренда.

Годы	$\ln y_i$	$\ln \phi_i$	отклонен ия $U_{\ln y_i}$	$U^2_{\ln y_i}$	t_i	t_i^2
1950	7,835	7,849	-0,014	0,000196	-2,5	6,25
1960	8,026	8,027	-0,001	0,000001	-1,5	2,25
1970	8,223	8,206	+0,017	0,000289	-0,5	0,25
1980	8,396	8,384	+0,012	0,000144	0,5	0,25
1990	8,564	8,562	+0,002	0,000004	1,5	2,25
2000*	8,726	8,741	-0,015	0,000225	2,5	6,25
Σ	49,77	49,77	0	0,000859	0	17,5

* – оценка

Среднее квадратическое отклонение логарифмов:

$$S_{(t) \ln y_i} = \sqrt{\frac{0,000859}{6-2}} = 0,014654$$

Средняя ошибка логарифма коэффициента изменения:

$$m_{\ln k} = \frac{0,014654}{\sqrt{17,5}} = 0,003503$$

Критерий Стьюдента:

$$\frac{\ln k}{m_{\ln k}} = \frac{0,1783}{0,003503} = 50,9$$

Табличный критерий Стьюдента при четырёх степенях свободы и значимости 0,01 равен 4,60. Полученное значение критерия много больше табличного, так что вероятность нулевой гипотезы можно считать равной нулю, а рост населения Земли - достоверным. Понятно, что столь очевидное явление и не требовало проверки, пример приведения для показа методики надежности экспоненциального тренда, а не проверки самого факта роста населения, как это имело место в примере с ростом среднегодовой температуры.

Для кривых, не имеющих постоянного основного параметра, вышеизложенный метод проверки надежности не применим. В таких случаях можно, во-первых, проверять сам факт наличия какого-либо тренда, путем сравнения средних уровней за первую и вторую половину периода, во-вторых, с помощью 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 этот случай 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 определяется по формуле $m_a = \frac{S(t)}{\sqrt{n}}$. 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 свободного члена, т.е. $\frac{S(t)}{\sqrt{n}}$, а в любой иной точке тренда его средняя ошибка вычисляется по формуле:

$$m_{f_m} = \sqrt{m_a^2 + m_b^2 \cdot t_m^2} = \sqrt{\frac{S^2(t)}{n} + \frac{S^2(t) \cdot t_m^2}{\sum_{i=1}^n t_i^2}} = S(t) \sqrt{\frac{1}{n} + \frac{t_m^2}{\sum_{i=1}^n t_i^2}} - \text{для однократного}$$

выравнивания и при $t_i = 0$ в середине ряда. При нумерации периодов времени от начала ряда вместо t_i в формулу следует подставить величину $(i - \bar{i})$; $(m - \bar{t})$.

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

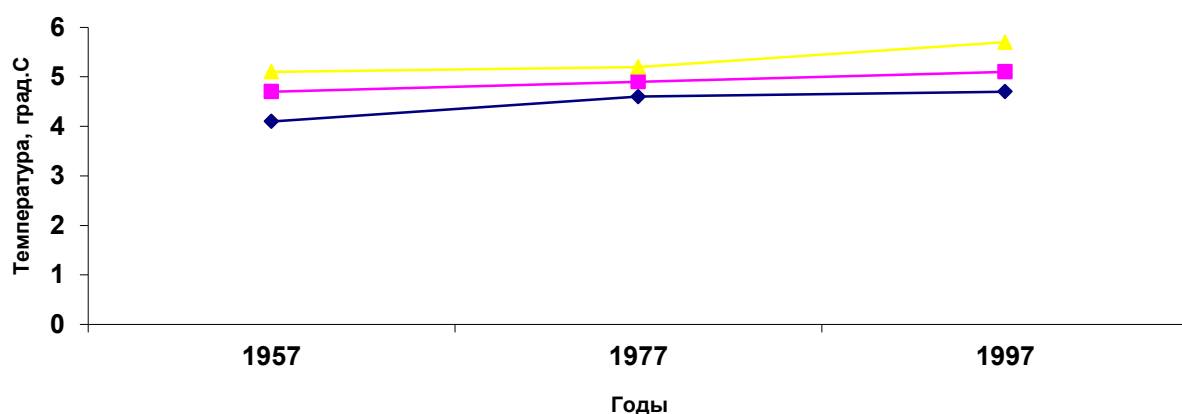
Рассчитаем среднюю ошибку тренда среднегодовой температуры воздуха в Ленинграде-Санкт-Петербурге:

$$m_{y_n} = 1,121 \cdot \sqrt{\frac{1}{41} + \frac{t_m^2}{21 \cdot 707}}$$

Для середины ряда – 1977 г. средняя ошибка тренда составила:
 $1,121 \cdot \sqrt{\frac{1}{41}} = 0,175^\circ$. А для крайних уровней - 1957 г. и 1997 г. - средняя 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 зоны вероятных значений генерального тренда.

Если эту зону мы желаем 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 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 расширится 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 отклонения или коэффициента вариации. Их можно применить 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 вид [19, с. 499-500]:

$$m_{S(t)} = \frac{S(t)}{\sqrt{2n}},$$

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 вероятности 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).

Таблица 7.3.

Вероятность отличия $S(t)$ от нуля

n	$\sqrt{2n}$	Табличный критерий			Вывод о надежности отличия колеблемости уровней от нуля
		P=0,9	P=0,95	P=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	—"

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 $S(t) \pm t_{\text{Стьюд.}} \cdot m_{S(t)}$. Например, доверительный интервал средней силы колебаний среднегодовой температуры воздуха в Ленинграде-Санкт-Петербурге за 1957–1997 гг. с вероятностью 0,95 составил:

$$1,121 \pm 2,03 \cdot \frac{1,121}{\sqrt{2 \cdot 41}} = 1,121 \pm 0,251^\circ.$$

Доверительный интервал среднего квадратического отклонения урожайности зерновых культур во Франции за 1970–1995 гг. (см. табл. 6.5) с

вероятностью 0,99 составляет: $3,54 \pm 2,80 \cdot \frac{3,54}{\sqrt{2 \cdot 26}} = 3,54 \pm 1,37$ ц/га. Ввиду довольно значительной силы колебаний, доверительный интервал оценки генерального среднего квадратического колебания довольно широк. Ошибка возрастает прямо пропорционально силе колеблемости и с ростом надежности оценки, а уменьшается обратно пропорционально корню квадратному из числа уровней ряда.

Средняя ошибка репрезентативности выборочной оценки генерального коэффициента колеблемости имеет вид [20]:

$$m_{V(t)} = \frac{V(t)}{\sqrt{2n}} \cdot \sqrt{1 + 2 \left(\frac{V(t)}{100} \right)^2}^3$$

где $V(t)$ – коэффициент колеблемости, %.

Например, коэффициент вариации урожайности зерновых во Франции за 1970–1995 гг. составил 6,9%. Если рассматривать этот показатель как выборочный для Франции вообще на больший период, то средняя ошибка коэффициента как оценки генерального равна:

$$\frac{6,9}{\sqrt{2 \cdot 26}} \cdot \sqrt{1 + 2 \cdot 0,069^2} = 0,96\%.$$

С вероятностью 0,95 при 25 степенях свободы вариации, доверительные границы генерального коэффициента вариации составят $6,9\% \pm 2,06 \cdot 0,96\%$ ил: от 4,94 до 8,86%. Таким образом, почти наверняка колеблемость слабее 10%.

Не менее, а может и более, важной задачей, чем вероятностная оценка генеральных параметров колеблемости, является вероятностная оценка крайних отклонений от тренда, например, сильных неурожаев, экстремальных температур и влажности воздуха, скорости ветра 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 распределения вероятностей отклонения от тренда близко к нормальному закону. В нашем примере распределение отклонений от тренда среднегодовой температуры воздуха в Ленинграде–Санкт-Петербурге близко к нормальному (табл. 7.4).

Таблица 7.4.

**Проверка близости распределения колебаний температуры к
нормальному закону; $\sigma(t) = 1,121$**

Отклонения, град.	i	t_i -критерий	Вероятность $P_i(t_i)$	f_{Ti}	$\frac{(-f_{Ti})^2}{f_{Ti}}$
Ниже $-1,2^\circ$		от $-\infty$ до $-1,07$	0,1423	5,8	0,11
От $-1,2^\circ$ до $-0,4^\circ$		от $-1,07$ до $-0,36$	0,2171	8,9	0,10
От $-0,4^\circ$ до $+0,4^\circ$	5	от $-0,36$ до $+0,36$	0,2812	11,5	1,06
От $0,4^\circ$ до $1,2^\circ$		от $0,36$ до $1,07$	0,2171	8,9	0,10
Выше $1,2^\circ$		от $1,07$ до $+\infty$	0,1423	5,8	0,11
ИТОГО	1	—	1	40,9	1,48

t_i – нормированное отклонения границ интервала от среднего отклонения, равного нулю.

Вероятность попасть в интервал при условии нормального распределения отклонений по их величине P_i – это половина разности интегральных функций нормального распределения: $0,5[F(t_1) - F(t_2)]$, где t_1 , t_2 – значения критерия для границ интервала. Для среднего интервала от $t_1 = -0,36$ до $t_2 = +0,36$ вероятность $P=F(0,36)$. Теоретические частоты f_{Ti} есть произведение $n \cdot P_i$, где $n=41$.

Итог последней графы – это критерий χ^2 (хи-квадрат). Табличное значение критерия для значимости 0,10 равен 4,60, при двух степенях свободы, а фактическое много ниже табличного. Следовательно, вероятность сходства распределения отклонений температуры от тренда с нормальным много больше, чем 0,1, и гипотеза о нормальном распределении не отвергается.

Другие временные ряды, рассмотренные в данном учебном пособии слишком коротки для проверки по χ^2 . В 1976–1980 гг. кафедрой статистики Ленинградского сельскохозяйственного института (ЛСХИ) было проведено по договору с Управлением статистики сельского хозяйства (ЦСУ) СССР изучение колебаний урожайности по многим культурам в областях и краях РСФСР. Среди других был получен вывод о близости распределения отклонений урожайности от трендов по величине отклонений к нормальному закону распределения [19 с. 3-9].

Этот эмпирический вывод подкрепляется теоретическими соображениями: колебания урожайности зависят от очень большого числа сравнительно независимых факторов, каждый из которых не играет определяющей роли. Следовательно, колебания урожайности отвечают условиям "предельной теоремы Ляпунова", которая устанавливает, когда случайная переменная имеет нормальное распределение вероятностей. На этом основании будем считать, что и колебания урожайности зерновых во

Франции подчинены нормальному закону. Среднее квадратическое отклонение, согласно данным табл. 6.5 равно 3,54 ц/га. Находим *вероятности рисков*, т.е. что отклонение от тренда вниз (неурожай) превышает уровни –5 ц/га; –7 ц/га; –10ц/га; - 12 ц/га (табл. 7.5).

Вероятность P равна половине разности между единицей и $F(t)$, то есть применяется *односторонний критерий* (иногда в литературе приводится готовая таблица вероятностей именно для него). Поясним определение этой вероятности с помощью графика (рис. 7.2), из которого ясно и то, что у нас обозначено, как $F(t)$.

Таблица 7.5.

Расчет вероятностей рисков (неурожая) зерновых во Франции

Отклонения вниз от тренда, ц/га U_i	Нормированное отклонение $t = U_i: \sigma(t)$	Его вероятность P
-5 и более	1,41	0,079
-7 и более	1,98	0,024
-10 и более	2,82	0,0024
-12 и более	3,39	0,00034

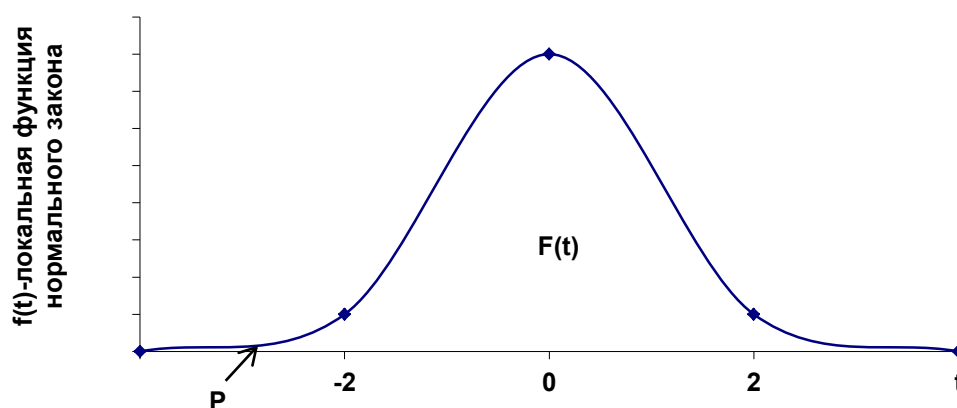


Рис. 7.2. Вероятность отрицательного отклонения, большего по величине, чем заданная граница.

Таким образом, вероятность небольшого неурожая (отклонения на 5 ц/га или больше) почти равна 8%, т.е. в среднем может случиться 8 раз за 100 лет, а вот вероятность сильного неурожая во Франции (больше, чем на 10 ц/га вниз от тренда) очень мала - всего 0,002. Таким риском можно пренебречь. Конечно, это относится к стране в целом, а для отдельного фермера и колеблемость урожаев будет гораздо больше, и вероятность риска. Для ее определения нужно анализировать временной ряд урожайности на ферме.

Логически ясно (это видно из графика, рис. 7.2), что точно такова же, как вероятность неурожая, больше, чем на $2S(t)$ от тренда вниз, так и вероятность высокого урожая, больше, чем на $2S(t)$ от тренда вверх. И с таким «сверхурожаем» тоже может быть связан коммерческий риск – риск сильного падения цены на товар.

Если же распределение колебаний по их величине далеко от нормального, а 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 следующем условном примере, приведенном в табл. 7.6.

Таблица 7.6.

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
Ниже от 20 до 10	7	0,14	0,15	0,29
Ниже от 0 до 10	12	0,24	0,18	0,42
Выше от 0 до 10	5	0,10	0,13	0,23
Выше от 10 до 20	14	0,28	0,19	0,47
Выше от 20 до 30	7	0,14	0,15	0,29
Выше более чем на 30	2	0,04	0,08	0,12
ИТОГО	50	1,00	—	—

Средняя ошибка репрезентативности выборочной доли (частости), как известно, равна:

$$m_d = \frac{\sqrt{d \cdot (1-d)}}{\sqrt{n}}.$$

Вычислив средние ошибки всех частостей, 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 частости и сделать вывод о возможной частости отклонения от тренда на

указанный процент по величине этой частоты плюс трехкратная средняя ошибка.

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 прогресс агротехники приводил к уменьшению величины колебаний, хотя бы той же урожайности. 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:

$$\overline{S(t)^2}_{геом.} = \sum_{i=1}^k n_i \sqrt[k]{\prod_{i=1}^k [S(t)^2]^{n_i}}$$

Критерий Бартлетта M имеет вид:

$$M = \ln \frac{\overline{S(t)^2}_{ap.}}{\overline{S(t)^2}_{геом.}} \cdot \sum_{i=1}^k n_i ,$$

его средняя ошибка:

$$C = 1 + \frac{\sum_{i=1}^k \frac{1}{n_i} - \frac{1}{\sum n_i}}{3(k-1)}$$

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 урожайности зерновых культур во Франции (см. гл. 5 и 6) за первые 11 лет (1970–1980 гг.) и за последние 11 лет (1985–1995 гг.):

$$S(t)_{70-80} = \sqrt{\frac{180,96}{11-1}} = 4,254 \text{ ц/га.}$$

$$S(t)_{85-95} = \sqrt{\frac{71,04}{11-1}} = 2,665 \text{ ц/га.}$$

Соответственно дисперсии равны:

$$S^2(t)_{70-82} = 18,1; \quad S^2(t)_{85-95} = 7,1$$

их арифметическая средняя равна:

$$S^2_{ap} = 12,6$$

а их геометрическая средняя

$$\sqrt{18,1 \cdot 7,1} = 11,34$$

$$M = 2 \ln \frac{12,6}{11,34} \cdot 22 = 4,636$$

$$C = 1 + \frac{\frac{1}{11} - \frac{1}{22}}{3} = 1,015$$

$M/C = 4,57$. Табличное значение критерия χ^2 при одной степени свободы и значимости 0,05 составляет 3,84. Фактическое значение 4,57 больше табличного, следовательно, можно считать, что колеблемость в последние 11 лет ниже, чем в первые 11 лет изучавшегося периода, т.е. колеблемость урожайности зерновых во Франции уменьшилась.

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 А.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. Тенденция характеризует 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 вышесказанное, что понятие устойчивости 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.

Устойчивость временного ряда можно оценивать на различных явлениях. При этом в зависимости от явления будут меняться показатели, которые используются в качестве форм выражения сущности исследуемого

процесса, но 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:

$$R_{\bar{y}} = \bar{y}_{\text{благ}} - \bar{y}_{\text{неблаг}} \quad (8.1)$$

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 отношение к единице, тем меньше колеблемость и соответственно выше устойчивость. Назовем это отношение *индексом устойчивости уровней динамических рядов* и обозначим:

$$i_{\bar{y}} = \frac{\bar{y}_{\text{благ}}}{\bar{y}_{\text{неблаг}}} \text{ или } i_{\bar{y}} = \frac{\bar{y}_в}{\bar{y}_н} \quad (8.2.)$$

– отношение средней уровней выше тренда к средней уровней ниже тренда (при тенденции роста).

Например, по данным табл. 5.7 индекс устойчивости уровней валового сбора чая в Китае за 1978-1994 гг. составил 1,02.

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):

$$\text{среднее линейное отклонение } a(t) = \frac{\sum_{i=1}^n |y_i - \tilde{y}_i|}{n - p}, \quad (8.3)$$

$$\text{среднее квадратическое отклонение } S_y(t) = \sqrt{\frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{n - p}}, \quad (8.4)$$

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

выравнивании дает информацию о снижении или повышении устойчивости уровней за период исследования. Аналитическое выравнивание $a(t)$ 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 показатели [23]:

1. *Процентный размах* (Percentage Range) - PR:

$$PR = W_M - W_m, \quad (8.5.)$$

где $W_M = \text{Max}(W_2, \dots, W_{t-1})$,

$W_m = \text{Min}(W_2, \dots, W_{t-1})$,

$$W_t = \frac{|X_t - X_{t-1}|}{X_{t-1}} * 100, \quad t = 1, \dots, n$$

PR оценивает разность между максимальным и минимальным относительными приростами в процентах.

2. Показатель скользящие средние (Moving Average) – МА, который оценивает величину среднего отклонения от уровня скользящих средних:

$$MA = \frac{\sum_{t=r+1}^{n-r} \left[\frac{x_i - x_t}{x_t} \right]}{n + 1 - m}, \quad (8.6)$$

$$x_t = \frac{\sum_{i=t-r}^{t+r} x_i}{m},$$

где $r = (m-1)/2$, m – период скользящей средней.

Пример по главе 3. (?)

3. *Среднее процентное изменение* (Average Percentage Change) - APC, которое оценивает среднее значение абсолютных величин относительных приростов и квадратов относительных приростов:

$$APC = \frac{\sum_{t=2}^n \left[\frac{X_t - X_{t-1}}{\max(X_t - X_{t-1}, X_{t-1} - X_t)} \right]^2}{n - 1} * 100 \quad (8.7)$$

Бланфорд и Оффат, анализируя выше перечисленные коэффициенты, отмечают их хорошую согласованность относительно коэффициента Спирмэна.

Относительные показатели колеблемости, чаще всего используемые в статистике, вычисляются делением абсолютных показателей на средний уровень за весь изучаемый период (см. разд. 6.2.2):

$$\text{коэффициент линейной колеблемости } V_y = \frac{d_y(t)}{\bar{y}}, \quad (8.8)$$

$$\text{коэффициент колеблемости } V_y = \frac{S_y(t)}{\bar{y}}, \quad (8.9)$$

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 (гл. 6, разд. 6.2.2), коэффициент колеблемости урожайности зерновых культур во Франции за 1970-1995 гг. составил 6,9%, следовательно, коэффициент устойчивости уровней равен 93,1%.

8.2 Методы измерения устойчивости тенденции динамики

Наиболее простым показателем устойчивости тенденции временного ряда является коэффициент Спирмэна K_p [3, с. 39]:

$$K_p = 1 - \frac{6 \sum_{i=1}^n d^2}{n^3 - n} \quad (8.11)$$

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 Спирмэна 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 необходима поправка к формуле Спирмэна:

$$K_p = 1 - \frac{6 \sum_{i=1}^n d^2 - A}{n^3 - n - 12A} \quad (8.12)$$

$$\text{где } A = \frac{1}{12} \sum_{j=1}^m (j^3 - A_j)$$

j – номера связок по порядку,

A_j – число одинаковых рангов в j -й связке (число одинаковых уравнений)

При малой вероятности совпадения уровней и достаточном их числе эта поправка незначительна.

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, $K_p = +1$. Это означает полную устойчивость самого факта роста уровней ряда, непрерывность роста.

Чем ближе K_p к $+1$, тем ближе рост уровней к непрерывному, выше устойчивость роста. При $K_p = 0$ рост совершенно неустойчив. При отрицательных значениях чем ближе K_p к -1 , тем устойчивее снижение изучаемого показателя. В рассмотренном ранее ряду динамики урожайности зерновых во Франции за 1970-1995 гг. коэффициент Спирмена составил 95,62%.

Коэффициент устойчивости роста (Kp) можно получить и по другой формуле.

$$Kp = \frac{12 \sum_{i=1}^n P_{ti} * P_{yi}}{n^3 - n} - \frac{3(n+1)}{n-1}, \quad (8.13)$$

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. Это 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.

Например, коэффициент устойчивости уровней урожайности картофеля в России за 1982-1997 гг. составил 0,919, а коэффициент устойчивости (снижения) тренда - только -0,612. Устойчивого тренда нет.

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. Каяйкиной [9] 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 $S_y(t)$:

$$K = \frac{b}{S_y(t)}. \quad (8.15)$$

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, притом 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 размах) величины колебаний, т.е. 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 и 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*:

$$O_{k_3} = \frac{\bar{k}}{\bar{k}_{Sy(t)}} \quad (8.16)$$

Если $O_{k_3} > 1$, 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 $O_{k_3} < 1$, наоборот, колебания растут быстрее уровней тренда и коэффициент колеблемости растет, коэффициент устойчивости уровней снижается. Таким образом, величина O_{k_3} 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 $b_{Sy(t)}$, полученным по достаточно длинному ряду путем выравнивания показателей $S_y(t)$, скользящих или следующих друг за другом. Имеем показатель

$$O_c = \frac{2c}{b_{Sy(t)}} \quad (8.17)$$

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 свидетельствует о повышении устойчивости динамики, тренда. Если $O_c < 1$, значит, колебания растут сильнее, чем происходит 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. $c > 0$; $b_{Sy(t)} > 0$; $2c > b_{Sy(t)}$.

Прирост уровней ряда растет, колебания тоже растут, но медленнее, 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.

$$2. c > 0; \quad b_{Sy(t)} > 0; \quad 2c < b_{Sy(t)}.$$

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.

$$3. c > 0; \quad b_{Sy(t)} < 0; \quad 2c > b_{Sy(t)}. \text{ – очевидная ситуация.}$$

Эта комбинация означает, что прирост уровней растет, а колеблемость снижается. Ясно, что при этом показатель устойчивости тенденции K возрастает.

4. $c > 0; \quad b_{Sy(t)} < 0; \quad 2c < b_{Sy(t)}$ – нереальная комбинация, третье неравенство противоречит двум первым.

5. $c < 0; \quad b_{Sy(t)} > 0; \quad 2c > b_{Sy(t)}$ – также нереальное сочетание по той же причине.

$$6. c < 0; \quad b_{Sy(t)} > 0; \quad 2c < b_{Sy(t)} \text{ – очевидная ситуация.}$$

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, относительно его устойчивости.

$$7. c < 0; \quad b_{Sy(t)} < 0; \quad 2c > b_{Sy(t)}$$

Отсюда следует, что прирост уровней сокращается, но медленнее, чем колеблемость, так как неравенство $2c > b_{Sy(t)}$ понимается по алгебраической величине, а не по модулю, т.е., например, $c = -0,05$, а $2c > b_{Sy(t)} = -0,13$, 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. $c < 0; \quad b_{Sy(t)} < 0; \quad 2c < b_{Sy(t)}$ – 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 подчеркнём, 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 $b_{Sy(t)}$.

Глава 9 Изучение динамики комплекса взаимосвязанных признаков

До этой главы 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 связью, либо корреляционной связью. В данной главе последовательно будут 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 учебника могут быть изложены только очень кратко. Желающим глубже изучить проблемы анализа и прогнозирования систем взаимосвязанных признаков рекомендуется обратиться к специальной литературе [1, 5, 6, 10, 14, 16, 18, 21].

9.1 Динамика жестко связанной системы признаков (показателей)

Насколько нам известно, в полном объеме динамика жестко связанной системы в нашей литературе впервые описана Л.Н.Кривенковой в диссертации, защищенной при Санкт-Петербургском университете экономики и финансов³. Изложение материала начнем с конкретной задачи: необходимо рассмотреть тенденции и колеблемость трех функционально взаимосвязанных признаков: площади посева зерновой культуры, ее урожайности и валового сбора зерна (табл 9.1). Если площади в разные годы обозначим как Π_i , урожайность - u_i , валовый сбор - b_i , то имеем функциональную связь: $b_i = \Pi_i \cdot u_i$, 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.

³ См.: Кривенкова Л.Н. Статистические методы анализа и моделирования свиноводства (на уровнях отдельного предприятия и региона): Дис. ...канд. экон. наук. – СПб., 1993. – С. 95–112.

Таблица 9.1.

Динамика площади, урожайности, валового сбора

Номера периодов	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_{Pi} , га	U_{Yi} , ц/га	U_{bi} , 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

Тренд площади: $\hat{P}_i = 120 + 5t_i$, $t=0$ 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 том случае, если бы колеблемость отсутствовала. Тогда валовой сбор каждого года являлся бы произведением уровней трендов площади и урожайности, которые совпадали бы с фактическими уровнями площади и урожайности, т.е. имело бы место равенство: $\hat{b}_i = \hat{P}_i \bullet \hat{y}_i = b_i$ и вектор валового сбора был бы следующим:

Таблица 9.2.

Вектор валового сбора

Номера периодов от середины ряда	4	3	2	1					
Валовой сбор	500	730	970	220	480	750	030	320	620
Абсолютный прирост к предыдущему		30	40	50	60	70	80	90	00
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 так же, как сам показатель валового сбора с показателями площади и урожайности:

Тренд признака-произведения есть произведение трендов признаков-сомножителей, если колеблемость равна нулю:

$$\begin{aligned}\hat{B} &= \hat{A} * \hat{C} = (120 + 5t) * (29 + t) = 120 * 29 + 5t * 29 + 120 * t + 5t * t = \\ &= 3480 + (145 + 120)t + 5t^2,\end{aligned}$$

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 табл. 9.1., имеет вид: $\hat{B}_i = 3472,2 + 264,3t_i + 4,94t_i^2$

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 произведением 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 $\bar{x}, \bar{z}$ полагаем линейными, следовательно, тренд $\bar{y}$ – парабола II-го порядка. Будем также для упрощения записи вести отсчет номеров периодов времени t_i от середины временных рядов. Фактические уровни признаков можно представить как сумму уровня тренда и отклонения от него, обозначаемого $U_{x_i}, U_{z_i}, U_{y_i}$, так что

$$x_i = \bar{x}_i + U_{x_i}; z_i = \bar{z}_i + U_{z_i}; y_i = \bar{y}_i + U_{y_i}$$

Так как

$$y_i = x_i z_i, \text{ то } y_i = (\bar{x}_i + U_{x_i}) * (\bar{z}_i + U_{z_i}). \quad (9.1)$$

Рассмотрим произведение трендов сомножителей:

$$\begin{aligned} \bar{x}_i \cdot \bar{z}_i &= (\bar{x} + b_x t_i) * (\bar{z} + b_z t_i) = \bar{x} \cdot \bar{z} + \bar{x} b_z t_i + \bar{z} b_x t_i + b_x t_i \cdot b_z t_i = \\ &= \bar{x} \cdot \bar{z} + (\bar{x} b_z + \bar{z} b_x) t_i + b_x b_z t_i^2 \end{aligned} \quad (9.2.)$$

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 изменения признаков имеют разные знаки, ускорения их произведения – отрицательная величина. При наличии большего, чем двух сомножителей тренд их произведения будет параболой более высокого порядка со значительно более сложным поведением, и 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 полагают, что если число работников будет возрастет на 10%, а производительность их труда на 8% в год, то выпуск продукции будет увеличиваться 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%; прирост на 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 $\bar{y}_i = a_y + b_y t_i + c_y t_i^2$, 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, так как $y_i = x_i z_i$, имеем:

$$b_y = \frac{\sum x_i z_i t_i}{\sum t_i^2} = \frac{\sum (\bar{x}_i + U_{xi}) (\bar{z}_i + U_{zi}) t_i}{\sum t_i^2} = \frac{\sum (\bar{x} + b_x t_i + U_{xi}) (\bar{z} + b_z t_i + U_{zi}) t_i}{\sum t_i^2} = \quad (9.3)$$

$$= \frac{\sum (\bar{x} \cdot \bar{z} t_i + \bar{x} b_z t_i^2 + \bar{z} b_x t_i^2 + \bar{x} U_{zi} t_i + \bar{z} U_{xi} t_i + b_x U_{zi} t_i^2 + b_z U_{xi} t_i^2 + b_x b_z t_i^3 + U_{xi} U_{zi} t_i)}{\sum t_i^2}$$

Рассмотрим суммы каждого из слагаемых в числителе (9.3):

$$1. \frac{\sum \bar{x} \cdot \bar{z} t_i}{\sum t_i^2} = \frac{\bar{x} \cdot \bar{z} \sum t_i}{\sum t_i^2} = 0, \text{ т.к. } \sum t_i = 0$$

2 и 3.

$$\frac{\sum \bar{x} b_z t_i^2}{\sum t_i^2} + \frac{\sum \bar{z} b_x t_i^2}{\sum t_i^2} = \frac{\bar{x} \sum b_z t_i^2}{\sum t_i^2} + \frac{\bar{z} \sum b_x t_i^2}{\sum t_i^2} = \frac{\bar{x} b_z \sum t_i^2}{\sum t_i^2} + \frac{\bar{z} b_x \sum t_i^2}{\sum t_i^2} = \bar{x} b_z + \bar{z} b_x, \text{ что}$$

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: так как сумма или математическое ожидание произведений величин, математические ожидания (или суммы) каждого из которых равны нулю, тоже равна нулю:

$$\sum U_{zi} t_i = 0; \sum U_{xi} t_i = 0$$

6 и 7.

$$\frac{\sum b_x U_{zi} t_i^2}{\sum t_i^2} + \frac{\sum b_z U_{xi} t_i^2}{\sum t_i^2} = \frac{b_x \sum U_{zi} t_i^2}{\sum t_i^2} + \frac{b_z \sum U_{xi} t_i^2}{\sum t_i^2}$$

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.

$$8. \frac{\sum b_x b_z t_i^3}{\sum t_i^2} = \frac{b_x b_z \sum t_i^3}{\sum t_i^2} = 0, \text{ т.к. как } \sum t_i^3 = 0$$

9. $\frac{\sum U_{xi} U_{zi} t_i}{\sum t_i^2}$, 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. гл. 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

$$\begin{aligned} y_i &= \frac{\sum (\bar{x} + b_x t_i + U_{xi}) (\bar{z} + b_z t_i + U_{zi})}{n} - \frac{\sum (\bar{x} + b_x t_i + U_{xi}) (\bar{z} + b_z t_i + U_{zi}) t_i^2}{\sum t_i^2} = \\ &= \frac{\sum \bar{x} \cdot \bar{z} + \sum \bar{x} b_z t_i + \sum \bar{z} b_x t_i + \sum b_x b_z t_i^2 + \sum \bar{x} U_{zi} + \sum \bar{z} U_{xi} + \sum b_x U_{zi} t_i + \sum b_z U_{xi} t_i + \sum U_{xi} U_{zi}}{n} - \\ &- \frac{\sum \bar{x} \cdot \bar{z} t_i^2 + \sum \bar{x} b_z t_i^3 + \sum \bar{z} b_x t_i^3 + \sum b_x b_z t_i^4 + \sum \bar{x} U_{zi} t_i^2 + \sum \bar{z} U_{xi} t_i^2 + \sum b_x U_{zi} t_i^3 + \sum b_z U_{xi} t_i^3 + \sum U_{xi} U_{zi} t_i^2}{\sum t_i^2} \end{aligned}$$

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.

$$2 \text{ и } 3. \frac{\sum \bar{x} b_z t_i}{n} + \frac{\sum \bar{z} b_x t_i}{n} = \frac{\bar{x} b_z \sum t_i}{n} + \frac{\bar{z} b_x \sum t_i}{n} = 0, \sum t_i = 0.$$

$$4. \frac{\sum b_x b_z t_i^2}{n} = \frac{b_x b_z \sum t_i^2}{n}.$$

$$5 \text{ и } 6. \frac{\sum \bar{x} U_{zi}}{n} + \frac{\sum \bar{z} U_{xi}}{n} = \frac{\bar{x} \sum U_{zi}}{n} + \frac{\bar{z} \sum U_{xi}}{n} = 0, \sum U_{xi} = \sum U_{zi} = 0$$

$$7 \text{ и } 8. \frac{\sum b_x U_{zi} t_i}{n} + \frac{\sum b_z U_{xi} t_i}{n} = \frac{b_x \sum U_{zi} t_i}{n} + \frac{b_z \sum U_{xi} t_i}{n} = 0, \text{ как уже показано}$$

ранее.

9. $\frac{\sum U_{xi} U_{zi}}{n}$. Этот член произведения в общем случае не равен нулю

при наличии корреляции между отклонениями от тренда.

$$10. -\frac{\sum \bar{x} \cdot \bar{z} t_i^2}{\sum t_i^2} = -\frac{\bar{x} \cdot \bar{z} \sum t_i^2}{\sum t_i^2} = -\bar{x} \bar{z}.$$

$$11 \text{ и } 12. -\frac{\sum \bar{x} b_z t_i^3}{\sum t_i^2} - \frac{\sum \bar{z} b_x t_i^3}{\sum t_i^2} = -\frac{\bar{x} b_z \sum t_i^3}{\sum t_i^2} - \frac{\bar{z} b_x \sum t_i^3}{\sum t_i^2} = 0, \text{ т.к. } \sum t_i^3 = 0.$$

$$13. -\frac{\sum b_x b_z t_i^4}{\sum t_i^2} = -b_x b_z \frac{\sum t_i^4}{\sum t_i^2}.$$

$$14 \text{ и } 15. -\frac{\sum \bar{x} U_{zi} t_i^2}{\sum t_i^2} - \frac{\sum \bar{z} U_{xi} t_i^2}{\sum t_i^2} = -\frac{\bar{x} \sum U_{zi} t_i^2}{\sum t_i^2} - \frac{\bar{z} \sum U_{xi} t_i^2}{\sum t_i^2} \text{ — в общем случае,}$$

как ранее показано, не равны нулю, т.к. 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.

$$18. -\frac{\sum U_{xi} U_{zi} t_i^2}{\sum t_i^2}, \text{ в общем случае и этот член не равен нулю при}$$

наличии корреляции между отклонениями.

Суммируя члены разложения 1, 4, 10 и 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 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 таблицы 9.1.

Свободный член тренда признака-произведения вычисляется системно вместе с квадратическим членом, а, значит, расхождение последнего с таковым в произведении трендов сомножителей означает, что и свободные члены также расходятся. Следовательно, в общем случае $a_y \neq \bar{x} \cdot \bar{z}$, свободный член уравнения параболического 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: $r_{ny} = -0,326$; $r_{nb} = -0,047$; $r_{yb} = 0,725$.

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:

$S(t)n = 7,4$ га; $v(t)n = 6,2\%$; колеблемость площади слабая;

$S(t)y = 4,4$ ц/га; $v(t)y = 15,2\%$; колеблемость умеренная;

$S(t)b = 633,6$ ц; $v(t)b = 15,3\%$; 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 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 влияют причины природного характера.

Итак, можно сделать лишь качественные выводы о связи и силе колебаний жестковзаимосвязанных 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 и 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.Поповой [13, с. 57-61] и в общем случае В.Н.Афанасьевым [2].

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:

$$X_i = \sum_{j=1}^k x_{ij} .$$

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:

$$B_x = \frac{\sum_{i=1}^n x_i t_i}{\sum_{i=1}^n t_i^2} = \frac{\sum_{i=1}^n \sum_{j=1}^k x_{ij} t_i}{\sum_{i=1}^n t_i^2} = \sum_{j=1}^k \frac{\sum_{i=1}^n x_{ij} t_i}{\sum_{i=1}^n t_i^2} = \sum_{j=1}^k b_j ,$$

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 тренды 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 непостоянной, 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 простейшего 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 $\bar{X}_i$ по совокупности есть сумма уровней трендов по единицам совокупности:

$$\bar{X}_i = \sum_{j=1}^k \bar{x}_{ji}.$$

Тогда и каждое отклонение от тренда по совокупности в целом:

$$U_i = X_i - \bar{X}_i = \sum_{j=1}^k x_{ij} - \bar{x}_{ij} = \sum_{j=1}^k u_{ij}.$$

Квадрат отклонения в i -м году от тренда по совокупности в целом равен: $U_i^2 = \left(\sum_{j=1}^n u_{ij} \right)^2$; сумма квадратов отклонений по совокупности в целом:

$$\sum_{i=1}^n U_i^2 = \sum_{i=1}^n \left(\sum_{j=1}^k u_{ij} \right)^2 = \sum_{i=1}^n \left(\sum_{j=1}^k u_{ij}^2 + 2 \sum_{1 \leq m < p \leq k} u_{im} u_{ip} \right) \quad (9.5)$$

$m, p \in j; m \neq p$

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 (сочетанием 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

$$\sum_{i=1}^n U_{mi} U_{pi} = r_{u_m u_p} \sqrt{\sum_{i=1}^n U_m^2 \cdot \sum_{i=1}^n U_p^2} = r_{u_m u_p} \sqrt{\gamma S_m^2 \gamma S_p^2} = r_{u_m u_p} \gamma S_m S_p, \quad (9.6)$$

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 $\gamma \cdot S^2(t)_{сов.}$ по совокупности в целом. Учитывая это и результат (9.6), можно записать вместо (9.5):

$$\gamma S^2 C_{об} = \sum_{j=1}^k \gamma \cdot S_j^2 C_j + 2 \sum_{1}^{C_k^2} r_{u_m u_p} \cdot \gamma S_m C_m \cdot \gamma S_p C_p$$

Сократив обе части равенства на число степеней свободы γ , имеем окончательный результат для объемных признаков:

$$S^2 C_{об} = \sum_{j=1}^k S_j^2 C_j + 2 \sum_{1}^{C_k^2} S_m C_m \cdot S_p C_p r_{u_m u_p} \quad (9.7)$$

Итак, можно сделать вывод: дисперсия колебаний признака в целом по совокупности с 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 ц^2 , то дисперсия валового сбора по району была бы равна 180000 ц^2 . В таком случае имели бы: $S(t)_{сов.} = \sqrt{180000} = 424,26 \text{ ц}$, в то время, как по каждому предприятию $S(t)_j = \sqrt{9000} = 94,87 \text{ ц}$, 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 снизила 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_{umnp} в преобладающей части – положительные величины. Если предположить, что в среднем общие факторы объясняют половину колебаний, т.е. $r^2=0,5$, $r\approx0,7$ то получим следующий результат по (9.7):

$$S(t)_{cov.} = \sqrt{180000 + C_{20}^2 \cdot 94,87^2} = \sqrt{180000 + 190 \cdot 94,87^2} = 13748 \text{ ц.}$$

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. В результате 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 нулю¹.

Данный закон справедлив и для вторичных признаков, таких, как урожайность. Если бы колебания урожайности у всех единиц совокупности были жестко связаны (т.е. все r_{umnp} 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, в хозяйствах есть не только общие для совокупности факторы колеблемости, но и специфические, все $r_{umnp} \rightarrow 1$, а, значит, колебания средней урожайности

¹ Проблема взаимопогашения колебаний валового сбора и урожайности была для России исследована в ряде работ А.И. Манелли, например, в монографии [20, гл. 2].

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, конечно, не строго согласованные по всем единицам совокупности, также 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, как известно, всегда меньше скалярной суммы векторов, не учитывающей их направленности.

9.3 Корреляция между временными рядами: сущность, ограничения

Предполагается, что читатель знаком с теорией корреляции в пространственных совокупностях и ее показателями, которые здесь используются. Корреляция временных рядов применяется для решения следующих задач:

- 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 наличия трендов? Если, например, в стране растет производство и применение минеральных удобрений, растет и урожайность сельскохозяйственных культур, но последняя растет не только по причине увеличения применения удобрений, а также и за счет других факторов - селекции новых сортов, мелиорации, орошения, механизации производства, роста экономической заинтересованности фермеров и др. 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 изучать, измерять, 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, 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 корреляции между колебаниями признаков, о методиках 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 же 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. В связи с этим в дальнейшем рассматриваются лишь случайно-распределенные во времени колебания, например, колебания урожайности.

Классический пример, иллюстрирующий отличие корреляции отклонений от тренда и корреляции уровней ряда, – это 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: сельскохозяйственные машины, энергоносители, удобрения. Рассмотрим пример, 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 г.

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 $-0,055$, 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:

$$r_{U_x U_y} = \frac{\sum_{i=1}^n (u_{x_i} - \bar{u}_x) \cdot (u_{y_i} - \bar{u}_y)}{\sqrt{\sum_{i=1}^n (u_{x_i} - \bar{u}_x)^2 \cdot \sum_{i=1}^n (u_{y_i} - \bar{u}_y)^2}}$$

Таблица 9.3.

**Корреляция урожайности картофеля с его себестоимостью
совхоза им. Ленина Волосовского района Ленинградской области**

Годы	Ур- сть, ц/га x_i	себ- ть, руб/ц y_i	Тренды		Отклонения от трендов		Их квадраты		Произвед. $U_{xi}U_{yi}$	$\bar{y}_i +$ $(-0,124U_x)$ руб/ц
			$\bar{x}_i$	$\bar{y}_i$	U_{xi}	U_{yi}	U_{xi}^2	U_{yi}^2		
1977	108	11,8	97	11,7	+11	+0,1	121	0,01	+1,1	10,3
1978	81	15,4	101	12,9	-20	+2,5	400	6,25	-50,0	15,4
1979	106	13,0	105	14,1	+1	-1,1	1	1,21	-1,1	14,0
1980	124	13,9	108	15,3	+16	-1,4	256	1,96	-22,4	13,3
1981	103	15,1	112	16,6	-9	-1,5	81	2,25	+13,5	17,7
1982	106	19,6	116	17,8	-10	+1,8	100	3,24	-18,0	19,0
1983	149	16,2	120	19,0	+29	-2,8	841	7,84	-81,2	15,4
1984	148	17,2	124	20,2	+24	-3,0	576	9,00	-72,0	17,2
1985	102	24,0	128	21,4	-26	+2,6	676	6,76	-67,6	24,6
1986	130	22,4	131	22,7	-1	-0,3	1	0,09	+0,3	22,8
1987	80	32,3	135	23,9	-55	+8,4	3025	70,56	-462,0	30,7
1988	139	24,7	139	25,1	0	-0,4	0	0,16	0	25,1
1989	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

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:

$$f(x)_i = (19,0 + 1,22t_i) + (-0,124u_{x_i}).$$

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 одно 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. 9.3 имеем:

$$\begin{aligned}\bar{\Delta}_x &= +5,57 \approx +5,8 \\ \bar{\Delta}_y &= +0,738 \approx +0,74\end{aligned}$$

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:

$$r_{\Delta_x \Delta_y} = \frac{\sum_{i=1}^n (\Delta_{x_i} - \bar{\Delta}_x) \cdot (\Delta_{y_i} - \bar{\Delta}_y)}{\sqrt{\sum_{i=1}^n (\Delta_{x_i} - \bar{\Delta}_x)^2 \cdot \sum_{i=1}^n (\Delta_{y_i} - \bar{\Delta}_y)^2}} = \frac{-1805,6}{\sqrt{14328,2 \cdot 261,6}} = -0,933,$$

Таблица 9.4.

Корреляция отклонений от средних отклонений

Годы	i	i	$x_i =$ $x_{i+1} -$ x_i	$y_i =$ $y_{i+1} -$ y_i	$\Delta_{x_i} - \bar{\Delta}_x$	$\Delta_{y_i} - \bar{\Delta}_y$	$(\Delta_{x_i} - \bar{\Delta}_x) \cdot$ $(\Delta_{y_i} - \bar{\Delta}_y)$	$(\Delta_{x_i} - \bar{\Delta}_x)^2$	$(\Delta_{y_i} - \bar{\Delta}_y)^2$
1977	08	1,8	...	...	...	...	...	...	...
1978	1	5,4	-27	3,6	-32,8	+2,86	-93,8	1075,8	8,2
1979	06	3,0	+25	2,4	+19,2	-3,14	-60,3	368,6	9,9
1980	24	3,9	+18	0,9	+12,2	+0,16	+2,0	148,8	0,0
1981	03	5,1	-21	1,2	-26,8	+0,46	-12,3	718,2	0,2

1982	06	9,6	+3	4,5	-2,8	+3,76	-10,5	7,8	14,1
1983	49	6,2	+43	3,4	+37,2	-4,14	-154,0	1383,8	17,1
1984	48	7,2	-1	1,0	-6,8	+0,26	-1,8	46,2	0,1
1985	02	4,0	-46	6,8	-51,8	+6,06	-313,9	2683,2	36,7
1986	30	2,4	+28	1,6	+22,2	-2,34	-51,9	492,8	5,5
1987	0	2,3	-50	9,9	-55,8	+9,16	-511,1	3113,6	83,9
1988	39	4,7	+59	7,6	+53,2	-8,34	-443,7	2830,2	69,6
1989	83	1,4	+44	3,3	+38,2	-4,04	-154,3	1459,2	16,3
Σ	559	47,0	+75	9,6	—	—	-1805,6	14328,2	261,6

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 предложит свои ответы на эти вопросы. Еще раз, и не последний, авторы подчеркивают, что наука – открытая система, продолжающийся процесс познания, открытие новых «материки» (реже) и «островов» (чаще) в бесконечном океане неведомого.

В заключении данного раздела напомним, что метод корреляции отклонений от трендов 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 – the 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 уровням одновременным, 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.

Вначале рассмотрим 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 вносится большая 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, расчет корреляции приведен в табл. 9.5. При этом будем считать, что тренд дозы внесенных органических удобрений отсутствует или несущественен.

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.

Таблица 9.5.

Корреляция с лагом в 1 год

Годы	Доза x_i т/га	Ур-ть y_i с 1 га	Тренд $\bar{y}_i$	Откл. U_{yi}	$x_i - \bar{x} = \Delta x_i$	$\Delta x_i U_{y_{i+1}}$	Δx_i^2	$U_{y_i}^2$	ϵ'
1987	45	13,0	15,0	-2,0	+4	+9,6	16	4,00	15,0
1988	36	18,0	15,6	+2,4	-5	-1,5	25	5,76	16,5
1989	47	16,5	16,2	+0,3	+6	-10,2	36	0,09	15,0
1990	33	15,1	16,8	-1,7	-8	+23,2	64	2,89	18,2
1991	42	14,5	17,4	-2,9	+1	0	1	8,41	15,6
1992	51	18,0	18,0	0	+10	+38,0	100	0	18,2
1993	36	22,4	18,6	+3,8	-5	+2,5	25	14,44	20,9
1994	42	18,7	19,2	-0,5	+1	+3,2	1	0,25	18,0
1995	35	23,0	19,8	+3,2	-6	+3,6	36	10,24	20,0
1996	40	19,8	20,4	-0,6	-1	+2,0	1	0,36	19,0
1997	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

Коэффициент корреляции с учетом лага в 1 год имеет вид:

$$r_{\Delta x_i U_{y(i+1)}} = \frac{\sum_{i=1}^{n-1} \Delta x_i U_{y_{i+1}}}{\sqrt{\sum_{i=1}^{n-1} \Delta x_i^2 \cdot \sum_{i=2}^n U_{y_i}^2}} = \frac{70,4}{\sqrt{305 \cdot 46,44}} = +0,5915, \quad r^2 = 0,35 \text{ или } 35\%$$

Связь колебаний дозы удобрений под предшественник 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 0,23 ц/га.

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, т.е. трендовые значения $\bar{y}_i + U_{y(x)i+1}$ 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 две 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 $\bar{x}$ и $\bar{y}$, отклонения отдельных уровней от трендов U_{xi} , U_{yi} , и начинаем вычислять корреляцию между ними, сначала без лага, затем с лагом в один период, с лагом в 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 3 года, то через 4 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 (табл. 9.6):

$$r_{u_{xi}u_{yi}} = \frac{\sum_{i=1}^n U_{xi} U_{yi}}{\sqrt{\sum_{i=1}^n U_{xi}^2 \cdot \sum_{i=1}^n U_{yi}^2}} = \frac{109,5}{\sqrt{1058 \cdot 100,72}} = 0,335$$

Таблица 9.6.

Корреляция отклонений от тренда с неизвестным заранее лагом

Годы	t_i	x_i	ξ_i	U_{xi}	y_i	η_i	U_{yi}	$U_{x_i}^2$	$U_{y_i}^2$
1983	-7	67	80	-13	18,2	22,2	-4,0	169	16,0
1984	-6	84	82	+2	26,7	23,4	+3,3	4	10,89
1985	-5	95	84	+11	27,8	24,6	+3,2	121	10,24
1986	-4	77	86	-9	22,3	25,8	-3,5	81	12,25
1987	-3	100	88	+12	27,5	27,0	+0,5	144	0,25
1988	-2	98	90	+8	30,6	28,2	+2,4	64	5,76
1989	-1	93	92	+1	26,8	29,4	-2,6	1	6,76
1990	0	91	94	-3	31,9	30,6	+1,3	9	1,69
1991	1	91	96	-5	32,3	31,8	+0,5	25	0,25
1992	2	111	98	+13	31,9	33,0	-1,1	169	1,21
1993	3	94	100	-6	30,7	34,2	-3,5	36	12,25
1994	4	87	102	-15	38,4	35,4	+3,0	225	9,00
1995	5	107	104	+3	39,5	36,6	+2,9	9	8,41
1996	6	106	106	0	37,8	37,8	0	0	0
1997	7	109	108	+1	36,6	39,0	-2,4	1	5,76
Σ	0	1410	1410	0	459	459	0	1058	100,72

$$r_{u_{xi}u_{yi+1}} = \frac{\sum_{i=1}^{n-1} U_{x_i} U_{y_{i+1}}}{\sqrt{\sum_{i=1}^{n-1} U_{x_i}^2 \cdot \sum_{i=2}^n U_{y_i}^2}} = \frac{-173,2}{\sqrt{1057 \cdot 84,72}} = -0,579$$

$$r_{u_{xi}u_{yi+2}} = \frac{\sum_{i=1}^{n-2} U_{x_i} U_{y_{i+2}}}{\sqrt{\sum_{i=1}^{n-2} U_{x_i}^2 \cdot \sum_{i=3}^n U_{y_i}^2}} = \frac{-49,8}{\sqrt{1057 \cdot 73,83}} = -0,178$$

$$r_{u_{xi}u_{yi+3}} = \frac{\sum_{i=1}^{n-3} U_{x_i} U_{y_{i+3}}}{\sqrt{\sum_{i=1}^{n-3} U_{x_i}^2 \cdot \sum_{i=4}^n U_{y_i}^2}} = \frac{184}{\sqrt{1048 \cdot 63,59}} = 0,713$$

$$r_{u_{xi}u_{yi+4}} = \frac{\sum_{i=1}^{n-4} U_{x_i} U_{y_{i+4}}}{\sqrt{\sum_{i=1}^{n-4} U_{x_i}^2 \cdot \sum_{i=5}^n U_{y_i}^2}} = \frac{-57,4}{\sqrt{823 \cdot 51,34}} = -0,279$$

$$r_{u_{xi}u_{yi+5}} = \frac{\sum_{i=1}^{n-5} u_{xi} u_{yi+5}}{\sqrt{\sum_{i=1}^{n-5} u_{xi}^2 \cdot \sum_{i=6}^n u_{yi}^2}} = \frac{-104,7}{\sqrt{787 \cdot 51,09}} = -0,522$$

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 3 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: обратная 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 признаков¹.

Очень упрощенно проблема состоит в следующем: имеем пространственную многофакторную регрессионную модель:

$$y_i = a + \sum_{j=1}^k b_j x_{ji} ,$$

выражающую корреляционную связь результативного признака y от 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

¹ 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 и авторы данного учебника.

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 значения. Рассмотрим пример по изменению среднего надоя молока на корову в совхозах Ленинградской области в 1983 г. по сравнению с 1982 г. 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 1982 г. регрессионная модель имела вид:

$$\bar{y}_{82} = 3,36x_1 + 11,02x_2 + 16,69x_3 - 6,07x_4 - 2,92x_5 + 8,08x_6 + 62,51x_7 + 3,48x_8 - 64,50$$

За текущий 1983 г. регрессионная модель приняла следующий вид:

$$\bar{y}_{83} = 2,95x_1 + 18,4x_2 + 19,55x_3 - 5,70x_4 - 2,95x_5 + 8,54x_6 + 52,74x_7 + 4,37x_8 + 124,1$$

Средние значения признаков за оба периода приведены в табл. 9.7.

Таблица 9.7.

Значение факторных признаков

Признаки	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	y	R^2
1982г.	145,3	34,8	9,0	74,2	105,6	41,7	42,1	29,6	3270,2	0,926
1983г.	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:
 $\frac{3440,7}{3270,2} \cdot 100\% = 105,2\%$, абсолютное изменение: +170,5 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, состоит в получении более надежных параметров связи, абстрагированных от колебаний под влиянием условий 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. Крастинь в монографии [11]. Им показано, что лучше применять усреднение коэффициентов годичных моделей 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 завода-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 фактических 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 относятся изучение динамики регрессионной модели и формирование ее прогнозируемой формы на будущий период. Эта проблема в нашей статистике подробно рассмотрена в работах А.А. Френкеля, в частности [15]. В кратком изложении разработанный им метод 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. Вычисляются 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 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, чтобы их системный характер разрушился. 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» [15, с.173].

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 была разработана русскими статистиками: С.П. Бобровым и Б.С. Ястремским, Н.С. Четвериковым и получила у них название *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, а 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.

С.П. Бобров и Б.С. Ястремский в 1922-1923 г. 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. Ястремский предложил [22] 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 ученый 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.

динамического ряда.

С.П. Бобров справедливо упрекает Ястремского за принятие гипотезы о постоянстве «сигм», так как они 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.

С.П. Бобров предложил вычислять *скользящий коэффициент корреляций* между временными рядами [4], аналогично скользящей средней. 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 экономических факторов. Заметим, что у С.П. Боброва нет также определенного указания на то, что речь должна идти о корреляции между отклонениями уровней от трендов, а не самих уровней двух рядов.

Идея 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. Лунеевой в 1924 г. [17], 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 $\sum_{i=1}^m U_{xi}U_{yi}$, где m от 1 до n ;

- нарастающие итоги квадратов отклонений каждого признака от своего тренда:

$$\sum_{i=1}^m U_{xi}^2 \text{ и } \sum_{i=1}^m U_{yi}^2 ;$$

- все три нарастающие ряда выравниваются 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 m , 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 выравнивать 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. Четвериков в [17] 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 одного месяца до целого года до окончания сбора следующего урожая. Лаг зависит от конкретной задачи, сущности изучаемого процесса, о чем упомянуто ранее. Метод Н.С. Четверикова и сейчас сохраняет свое значение, является важным вкладом отечественной статистики в методологию изучения связей и динамики.

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 и деятельности 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. Иван Ефремов предусмотрел в далеком будущем 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 прогнозирования по тренду: чтобы знать, какого уровня достигнет тот или иной «идуший» процесс, например, 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 оправдается. 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, чтобы 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 энергопотребления 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). Если, например, тренд урожайности зерновых культур во Франции был рассчитан за 1970-1995 гг. (база в 25 лет), то прогноз урожайности нежелательно строить более, чем на восемь лет вперед, т.е. до 2003 года. 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: 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 ее основании 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. Но 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 прогнозиста 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 и этап разработки 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 урожайности зерновых культур во Франции, рассчитанный в гл. 5, имеет вид:

$$\bar{Y}_t = 51,25 + 1,452 t_i, t = 0 \text{ в } 1983 \text{ г.}$$

Прогноз по этому тренду на 2000 г., номер которого от 1983 г. равен 17, составит:

$$\bar{Y}_{2000} = 51,25 + 1,452 * 17 = 75,93 \text{ ц/га.}$$

Интерпретация этого прогноза должна быть следующей: если урожайность зерновых во Франции будет возрастать до 2000 г. с той же средней скоростью (среднегодовым приростом), с каким она росла в период с

1970 по 1995 г., то тренд урожайности в среднем пройдет в 2000 г. через точку 75,93 ц/га. Такой прогноз и называется *точечным прогнозом*. Разумеется, точечный прогноз - это скорее абстракция, чем реальность. Если уровни урожайности и параметры тренда можно было бы определять с бесконечной степенью точности, то и вероятность точного осуществления точечного прогноза урожайности, составляющего 75,9324501387455603279... ц/га, была бы равна нулю. Поскольку мы дали прогноз с двумя знаками за запятой, то реально это уже не строго точечный прогноз, а прогноз попадания тренда в интервал от 75,9250001 до 75,9349999 ц/га, т.е. в интервал шириной 0,01 ц/га. Если точечный прогноз дать в целых центнерах с гектара, то это означает прогноз на прохождение линии тренда в прогнозируемом периоде в интервале от 75,500001 до 76,49999..., т.е. в интервал шириною в 1 ц/га. Вероятность этого события уже не мала.

От строго математических дефиниций перейдем к более практическим свойствам точечного прогноза. Он означает, что при нормальном законе распределения отклонений от тренда вероятности того, что урожайность окажется ниже точечного прогноза или 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, урожайность пять лет не может возрастать без малейших колебаний точно на 1,452 ц/га. В гл. 7 уже 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. При этом следует 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:

$$m_b = \frac{S(t)}{\sqrt{\sum_1^n t_i^2}},$$

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:

$$m_{\epsilon_k} = \sqrt{m_a^2 + m_b^2 \cdot t_k^2} = \sqrt{\frac{S(t)^2}{n} + \frac{S(t)^2 \cdot t_k^2}{\sum t_i^2}} = S(t) \sqrt{\frac{1}{n} + \frac{t_k^2}{\sum t_i^2}} \quad (10.2)$$

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 $+\frac{n-1}{2}$.

Например, тренд численности занятых в народном хозяйстве РФ за

1990-1996 гг. составил: $\hat{y}_i = 70,5 - 1,62 t_i$, где $t = 0$ в 1993 г., а $S(t) = 0,286$ млн чел. Точечный прогноз на 1998 г. составил: $\hat{y}_{98} = 70,51 - 1,62 \cdot 5 = 62,41$ млн чел. Средняя ошибка прогноза тренда на 1998 г. составит:

$$m_{\hat{y}_{98}} = 0,286 \cdot \sqrt{\frac{1}{7} + \frac{5^2}{28}} = 0,291 \text{ млн чел.}$$

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 t -критерия Стьюдента для вероятности 0,95 и пяти степеней свободы вариации (7-2 параметра линейного тренда). Получаем вероятную ошибку $\alpha_{\hat{y}_{98}} = t_{\text{Ст.}} \cdot m_{\hat{y}_{98}} = 2,8 \cdot 0,291 = 0,81$.

Итак, с вероятностью 0,95 тренд численности занятых в народном хозяйстве РФ в 1998 г. пройдет в границах: $62,41 \pm 0,81$ или от 61,6 до 63,22 млн чел. человек.

При линейном тренде и многократном выравнивании средняя ошибка прогноза для линии тренда на период с номером t_k от середины базы примет вид:

$$m_{\hat{y}_n} = S(t) \sqrt{\frac{1}{N} + \frac{t_k^2}{\ell \cdot \sum_{i=1}^n t_i^2}}, \quad (10.3)$$

где ℓ - число сдвигов базы расчета среднегодового прироста «b»;

n - длина каждой базы расчета;

N - общая длина временного ряда.

Например, тренд урожайности зерновых культур во Франции, рассчитанный в гл. 5 имел вид: $\hat{y}_i = 51,25 + 1,452 t_i$, где $t = 0$ в 1983 г.

$N = 26$, $\ell = 12$, $n = 15$, $S(t) = 3,54$ ц/га.

Средняя ошибка прогноза для линии тренда на 2000 г. с номером 17

составит: $m_{\hat{y}_{2000}} = 3,54 \sqrt{\frac{1}{26} + \frac{17^2}{12 \cdot 280}} = 1,25$ ц/га.

Критерий Стьюдента при 24 степенях свободы вариации составляет для вероятности 0,95 составит 2,08. Таким образом, с вероятностью 0,95 тренд при сохранении прежней скорости роста урожайности в 2000 г. пройдет в интервале: $75,93 \pm 2,08 \cdot 1,25$ или от 72,33 до 78,53 ц/га.

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 вид [18, с. 171]:

$$m_{\xi_k} = S(t) \sqrt{\frac{t_k^2}{\sum t_i^2} + \frac{\sum t_i^4 - (2\sum t_i^2) \cdot t_k^2 + n t_k^4}{n \sum t_i^4 - (\sum t_i^2)^2}}.$$

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 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 или момента

Неопределенность прогноза уровня отдельного периода складывается из двух элементов: ошибки линии тренда для прогнозируемого периода и колебания 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 их сознательном подавлении, например, колебаний урожайности при прогрессе агротехники, мелиорации земель.

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 $\xi_{(t)k}$, 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 неабсолютная истина, тренд имеет свою ошибку, которую также нужно учесть, используя в прогнозе уровня не $\xi_{(t)k}$, а доверительный интервал $\xi_{(t)k} \pm t_{Cr} \cdot m_{\xi_{(t)k}}$ и т.д. На

эти теоретически верные рассуждения следует ответить тем, что ограниченная точность и надежность исходных уровней временного ряда приводят 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 $\xi(t)_k$, 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, нужно 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 незабываемый 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:

$$m_{yk} = \sqrt{m_{fk}^2 + S^2(t)_k} \cdot (10.5)$$

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 (10.1) и (10.3), имеем:

$$m_{yk} = \sqrt{\frac{S(t)^2}{n} + \frac{S(t)^2 \cdot t_k^2}{\sum_{i=1}^n t_i^2} + S(t)_k^2} \cdot (10.6)$$

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:

$$m_{yk} = S(t) \sqrt{1 + \frac{1}{n} + \frac{t_k^2}{\sum_{i=1}^n t_i^2}} \cdot (10.7)$$

Именно данная формула приводится обычно в учебниках.

Соответственно при многократном расчете среднегодового прироста $\bar{b}$, получаем формулу

$$m_{yk} = S(t) \sqrt{1 + \frac{1}{N} + \frac{t_k^2}{\ell \sum_{i=1}^n t_i^2}}. \quad (10.8)$$

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 численности занятых в народном хозяйстве РФ на 1998 г. без учета тренда колеблемости:

$$m_{y98} = 0,286 \sqrt{1 + \frac{1}{7} + \frac{5^2}{28}} = 0,408 \text{ млн. чел.}$$

С вероятностью 0,95 численность занятых должна составить: $62,41 \pm 2,8 \cdot 0,408 = 62,41 \pm 1,14$, или от 61,27 до 63,55 млн чел.

В этом случае ввиду слабой колеблемости и малого срока прогноза ошибка прогноза отдельного уровня лишь на 40% больше ошибки положения тренда на 1998 г.

2. Прогноз урожайности зерновых культур во Франции на 2000 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 2000 г. составляет: $S(t)_k = \bar{S}(t) + b_{S(t)} \cdot t_k = 3,54 + (-0,1235) \cdot 17 = 1,44$. Подставляя его в расчет по формуле для многократного выравнивания, имеем:

$$m_{y_{2000}} = \sqrt{\frac{3,54^2}{26} + \frac{3,54^2 \cdot 17^2}{12 \cdot 280} + 1,44^2} = 1,906 \text{ ц/га.}$$

Заметим, что из трех слагаемых подкоренного выражения наибольшим является последнее, т.е. $S(t)_k^2$, составляющее 57% суммы. Ошибка прогноза уровня урожайности в основном вызвана колебаниями, возможными в 2000 г., а не ошибкой прогноза тренда, существенно сниженной применением многократного выравнивания и длинной базой.

С вероятностью 0,95 доверительные границы прогноза урожайности зерновых культур во Франции, при условии сохранения до 2000 г. прежней скорости ее увеличения, составляют: $75,93 \pm 2,08 \cdot 1,906 = 75,93 \pm 3,96$ ц/га

10.3.3 Прогноз доверительного интервала для среднего уровня за ряд периодов времени

Предметом прогнозирования может быть не только уровень отдельного года или месяца, но и средний уровень за ряд периодов или моментов

времени. Рассмотрим, что в этом случае можно считать *точечным прогнозом* и какова формула средней ошибки прогноза среднего уровня.

При линейной форме тренда прироста 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 1998-2002 гг. за точечный прогноз следует взять прогнозируемый на 2000 г. уровень. Он равен среднему уровню за 1998-2002 гг., так как

$$\frac{\bar{Y}_{98} + (\bar{Y}_{98} + b) + (\bar{Y}_{98} + 2b) + (\bar{Y}_{98} + 3b) + (\bar{Y}_{98} + 4b)}{5} = \frac{5\bar{Y}_{1998} + 10b}{5} = \bar{Y}_{1998} + 2b = \bar{Y}_{2000}$$

Если 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, что и 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, 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 среднего уровня для отрезка m единичных периодов после периода с номером t_k от середины базы расчета тренда:

$$m_{\bar{Y}_{k-k+m}} = \sqrt{\frac{S^2(t)}{n} + \frac{S^2(t)t_{k+0,5m}^2}{\sum_{i=1}^n t_i^2} + \frac{S^2(t)_{k+0,5m}}{m}}. \quad (10.9)$$

Если тренд $S(t)$ и его прогноз не вычислялся или несущественен, то имеем:

$$m_{\bar{Y}_{k-k+m}} = S(t) \sqrt{\frac{1}{n} + \frac{t_{k+0,5m}^2}{\sum_{i=1}^n t_i^2} + \frac{1}{m}} - \quad (10.10)$$

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 1970-1990 гг. тренд валового сбора пшеницы в России имел вид: $\hat{y}_t = 45,5 - 0,5857t$ млн т, $t = 0$ в 1980 г.

Рассчитаем по этому тренду прогноз среднегодового валового сбора на 1991-1995 гг. и его доверительные границы, без учета тренда колеблемости, при $S(t) = 7,18$ млн т. За точечный берем прогноз тренда на 1993 г.:

$$t_{k+\frac{m}{2}} = 13. \hat{y}_{91-95} = 45,5 + 13 \cdot (-0,5857) = 37,88 \text{ млн т};$$

$$m_{\bar{y}_{k-k+m}} = 7,18 \sqrt{\frac{1}{21} + \frac{13^2}{770} + \frac{1}{5}} = 4,91 \text{ млн т.}$$

С вероятностью 0,95 доверительный интервал прогноза среднегодового валового сбора зерна в России при условии сохранения до 1995 г. прежнего тренда составил: $37,88 \pm 2,09 \cdot 4,91$, или от 27,62 до 48,14 млн т. Широкий интервал вызван значительной силой колеблемости и требованием высокой надежности прогноза.

С учетом тенденции колебаний к их уменьшению ошибка прогноза была бы меньше, но для дальнейшего изложения это приведет к чрезмерному усложнению расчетов¹.

Рассмотрим, как меняются прогноз и его ошибка при постепенном получении фактических уровней прогнозируемого отрезка времени, т.е. после получения в нашем примере данных за 1991 и 1992 гг. и т.д. Ясно, что при каждом поступлении взамен прогнозируемой величины фактического уровня должна уменьшаться и ошибка прогноза среднего уровня на пятилетие и при замене последнего прогноза на 1995 г. на фактический уровень ошибка становится нулем. Рассмотрим этот процесс подробно. Пришел 1991 г., и получен его валовой сбор 38,9 млн т. Теперь среднегодовой уровень на 1991-1995 гг. равен сумме $38,9 + \text{четыре прогноза на 1992-1995 гг., деленной на пять}$. Если исходить из того, что уравнение тренда не будет ежегодно пересматриваться, имеем:

$$\bar{y}_{91-95} = (38,9 + 38,5 + 37,9 + 37,3 + 36,7) : 5 = 37,86 \text{ млн т.}$$

Средняя ошибка прогноза среднегодового сбора равна сумме нуля (для 1991 г.) и четырех ошибок для 1992-1995 гг., деленной на пять:

$$(0 + 4,73 + 4,91 + 5,09 + 5,28) : 5 = 4,00$$

Соответственно уменьшится ширина доверительного интервала

¹ Такой расчет мог бы стать темой дипломного исследования по статистике.

прогноза. Прошел 1992 г., валовой сбор этого года равен 46,2 млн т. Теперь точечный прогноз среднегодового уровня 1991-1995 гг. составит $(38,9 + 46,2 + 37,9 + 37,3 + 36,7) : 5 = 39,4$ млн т. Средняя ошибка этого прогноза равна $m_{\bar{x}} = (0 + 0 + 4,91 + 5,09 + 5,28) : 5 = 3,06$, доверительный интервал с вероятностью 0.95 равен $39,4 \pm 2,09 \cdot 3,06$ или от 33,0 до 45,8 млн т. После 1993 г., валовой сбор в котором равен фактически 43,5 млн т, имеем:

точечный прогноз

$$(38,9 + 46,2 + 43,5 + 37,3 + 36,7) : 5 = 40,52;$$

среднюю ошибку

$$(0 + 0 + 0 + 5,09 + 5,28) : 5 = 2,07;$$

доверительный интервал прогноза среднегодового валового сбора

$$40,52 \pm 2,09 \cdot 2,07 \text{ или от } 36,19 \text{ до } 44,85 \text{ млн т.}$$

После 1994 г. с его фактическим валовым сбором 32,1 млн т. имеем:

точечный прогноз

$$(38,9 + 46,2 + 43,5 + 32,1 + 36,7) : 5 = 39,48;$$

среднюю ошибку

$$(0 + 0 + 0 + 0 + 5,28) : 5 = 1,06;$$

доверительный интервал:

$$39,48 \pm 2,09 \cdot 1,06 \text{ или от } 37,26 \text{ до } 41,70 \text{ млн т.}$$

И после 1995 г. имеем фактический среднегодовой сбор:

$$(38,9 + 46,2 + 43,5 + 32,1 + 30,1) : 5 = 38,16 \text{ млн т.}$$

Как видим, фактический среднегодовой валовой сбор находится в границах всех постепенно сужающихся доверительных интервалов прогноза. Ретроспективная проверка методики прогноза показала ее состоятельность, несмотря на явно кризисные годы сельского хозяйства РФ.

10.3.4 Расчет страхового запаса для преодоления риска отклонения от тренда

Доверительные интервалы прогноза могут быть основанием для расчета рисков больших отклонений от тренда и страхового запаса для преодоления этого риска. В предыдущем разделе вычислены границы доверительного интервала прогноза валового сбора пшеницы в России на 1991-1995 гг.: среднегодовой валовой сбор ожидался с вероятностью 0,95 в границах от 27,62 до 48,14 млн т. Это означает, что при близком к нормальному закону распределения отклонений от тренда, вероятность того, что валовой сбор окажется в среднем за год ниже 27,62 млн т. равна: $(1 - 0,95) : 2 = 0,025$, или риск 2,5%. Однако если для нормального обеспечения населения России продовольствием необходимо 30 млн т. пшеницы в год, то вероятность риска будет больше. Для ее определения следует вычислить кратность отклонения 30 млн т. от точечного прогноза тренда 37,88 млн т. в единицах $S(t)$, т.е. $7,88 : 7,18 = 1,097 \approx 1,10$. Вероятность того, что отклонение от тренда превысит $-1,1 S(t)$ есть $(1 - F(1,1)) : 2 = (1 - 0,714) : 2 = 0,143$, или 14,3%. Такой высокий риск неприемлем, и для его перекрытия и доведения

до приемлемого уровня в 2,5% необходимо иметь страховой запас зерна пшеницы $30-27,62=2,38$ млн т на год, или $2,38 \cdot 5=11,9$ млн т на все пятилетие. Либо иметь резерв валюты и договора о поставке пшеницы из-за рубежа на сумму стоимости этих 11,9 млн т плюс стоимость перевозки. Как видим, расчет доверительных границ валового сбора и на отдельный год, и на большие сроки может иметь важное народно-хозяйственное значение. По верхней границе доверительного интервала можно рассчитать вероятность иного, не столь драматического риска: риск перепроизводства, нехватки емкостей для хранения зерна, снижение цены на рынке ниже уровня рентабельности.

10.4 Прогнозирование по модели тренда и сезонных колебаний

В гл. 6 были рассмотрены две модели сезонных колебаний: с помощью средних индексов сезонности *каждого месяца за ряд лет* и с помощью ряда Фурье, или тригонометрических функций. Прогнозирование по каждой из этих моделей с учетом тренда либо месячных уровней, либо среднегодовых уровней излагается в данном разделе.

10.4.1 Модель тренда месячных уровней и средних «индексов сезонности»

Имеем тренд месячных уровней затрат труда в сельскохозяйственном предприятии за 1995-1997 гг. (в тыс. ч) $\tilde{Y}_i = 60,03 + 1,085t_i$, $t=0$ в июле 1996 г., и вектор средних индексов сезонности для каждого месяца:

месяц	Янв.	Фев.	Март	Апр	Май	Июн	Июл	Авг	Сен	Окт	Ноя	Дек
$\bar{i}_{сезj}$	0,433	0,394	0,517	0,777	1,617	1,481	1,219	1,680	1,617	1,206	0,734	0,471

Среднее квадратическое отклонение фактических уровней от расчетных по тренду с учетом сезонных колебаний, т.е. мера случайной колеблемости $S(t)_{случ}$, равно 4,08 тыс.ч.

Модель, учитывающая тренд и сезонность, имеет вид:

$$\Psi'_{ij} = \Psi_i \cdot \bar{i}_{сезj} = (a + bt_i) \cdot \bar{i}_{сезj}.$$

Подстановкой в эту модель i -го номера месяца от середины базы и соответствующего этому месяцу j -го номера в году получаем точечный прогноз затрат труда в ij -м месяце. Например, для апреля 1999 г.:

$$i=33; j=4; \Psi'_{ij} = (60,03 + 1,085 \cdot 33) \cdot 0,777 = 74,46 \text{ тыс. ч.}$$

Для декабря 1998 г.:

$$i=29; j=12. \Psi'_{ij} = (60,03 + 1,085 \cdot 29) \cdot 0,471 = 43,09 \text{ тыс. ч.}$$

Средняя ошибка прогноза месячного уровня, вызванная наличием случайной колеблемости вычисляется по ранее приведенным в разд. 10.3.2

формулам. Для апреля 1999 г.:

$$m_{\bar{y}_{33}} = S(t) = \sqrt{1 + \frac{1}{n} + \frac{t_k^2}{\sum t_i^2}};$$

$$m_{\bar{y}_{33}} = 4,08 \sqrt{\frac{1}{37} + \frac{33^2}{4218}} + 1 = 4,63 \text{ тыс. ч.}$$

Доверительные границы затрат труда в апреле 1999 г. составят с вероятностью 0,95 (t -критерий Стьюдента $t \approx 2$):

$$74,46 \pm 2 \cdot 4,63 \text{ или от } 65,20 \text{ до } 83,72 \text{ тыс. ч.}$$

Для декабря 1998 г.:

$$m_{\bar{y}_{29}} = 4,08 \sqrt{1 + \frac{1}{37} + \frac{29^2}{4218}} = 4,52 \text{ тыс. ч.}$$

Доверительные границы с вероятностью 0,9 составят: $43,09 \pm 1,65 \cdot 4,52$ или от 35,63 до 50,55 тыс. ч.

Без учета сезонных колебаний прогноз был бы лишен всякого смысла, так как среднее квадратическое отклонение за счет сезонных колебаний равно 30,8, т.е. в 7,5 раза больше, чем случайная колеблемость, и ошибки прогнозов были бы во столько же раз больше, а доверительные интервалы - от нуля и до 100 с лишним. Такой «прогноз» можно сделать и без научных методов.

Следует, однако, подчеркнуть, что изложенная выше методика не лишена дискуссионных моментов: как, например, разделить степени свободы между сезонной колеблемостью и случайной? Измерять ли силу сезонной колеблемости по первичному ряду, через $\bar{y}_i \cdot \bar{i}_{\text{сез}} - \bar{y}_i$ или в относительном выражении, через средние индексы сезонности (их отклонения от 100% или от единицы)?

10.4.2 Модель тренда готовых уровней и тригонометрического уравнения сезонности

В разд. 6.3. приведен расчет модели сезонных колебаний месячных надоев молока на среднюю корову в форме тригонометрического уравнения первой гармоники ряда Фурье. Если продолжить анализ за 9 лет, получим модель:

$$\bar{y}_j = 316,7 + 45,4 \sin t_j - 83_{125} \cos t_j$$

где $\bar{y}_j$ - надой молока за j -месяц;

316,7 - средний месячный надой в среднем году базы расчета тренда 1989-1997 гг., т.е в 1993 .;

t_j - номер месяца года, начиная от января, для которого $t_j = 0$, умноженный на 30 градусов дуги ($360^\circ: 12$).

Имеем так же меру случайной колеблемости:

$$S(t)_{\text{случ}} = 11,0$$

Тренд имеет вид:

$$\bar{y}_i = 316,7 + 15,86 \cdot t_i$$

Чтобы построить по этим данным объединенную модель тренда и сезонности, необходимо учесть, что коэффициенты при синусе и косинусе зависят от величины свободного члена уравнения - среднего за данный год месячного надоя. Следовательно, эти коэффициенты каждый год необходимо корректировать на изменение по тренду среднемесячного надоя, т.е. для года с номером i они примут вид:

$$45,4 \left(1 + \frac{15,86 \cdot t_i}{316,7} \right); 83,25 \left(1 + \frac{15,86 \cdot t_i}{316,7} \right).$$

Сама же объединенная модель месячных надоев будет иметь следующее выражение:

$$\bar{y}_{ij} = 316,7 + 15,86 t_i + 45,4 \left(1 + \frac{15,86 t_i}{316,7} \right) \sin t_j - 83,25 \left(1 + \frac{15,86 t_i}{316,7} \right) \cos t_j.$$

Рассчитаем по этой модели прогноз (точечный) надоя на корову в апреле 1999 г.:

$$t_i = 6; t_j = 3 \cdot 30^\circ = 90^\circ$$

$$\bar{y}_{\text{апр}1999} = 316,7 + 15,86 \cdot 6 + 45,4 \left(1 + \frac{15,86 \cdot 6}{316,7} \right) \sin 90^\circ - 83,25 \left(1 + \frac{15,86 \cdot 6}{316,7} \right) \cos 90^\circ = 316,7 +$$

$$95,2 + 59,04 \sin 90^\circ - 108,26 \cos 90^\circ = 470,9 \text{ кг/гол.}$$

С учетом того, что средняя колеблемость не будет возрастать пропорционально надоем года от года, имеем среднюю ошибку прогноза:

$$m_{\bar{y}_{\text{апр}1999}} = 11 \sqrt{1 + \frac{1}{9} + \frac{6^2}{60}} = 14,4 \text{ кг/гол.}$$

С вероятностью 0,95 продуктивность коров в хозяйстве в апреле 1999 г. составит при семи степенях свободы:

$$470,9 \pm 2,36 \cdot 14,4 \text{ или от } 436,9 \text{ до } 504,9 \text{ кг/гол.}$$

Рассчитаем по той же модели прогноз надоя молока на среднюю корову на ноябрь 2000 г.:

$$t_i = 7, t_j = 10 \cdot 30^\circ = 300^\circ.$$

Точечный прогноз:

$$\bar{y}_{\text{ноябрь}2000} = 316,7 + 15,86 \cdot 7 + 45,4 \left(1 + \frac{15,86 \cdot 7}{316,7} \right) \sin 300^\circ - 83,25 \left(1 + \frac{15,86 \cdot 7}{316,7} \right) \cos 300^\circ = 427,7 +$$

$$+ 61,3 \cdot \sin 300^\circ - 112,4 \cos 300^\circ = 302,4 \text{ кг/гол.}$$

Средняя ошибка прогноза:

$$m_{\bar{y}_{\text{ноябрь}2000}} = 11 \cdot \sqrt{1 + \frac{1}{9} + \frac{49}{60}} = 15,3 \text{ кг/гол.}$$

С вероятностью 0,95 средний надой молока на корову в ноябре 2000 г. составит:

$$302,4 \pm 2,36 \cdot 15,3 \text{ или от } 266,3 \text{ до } 338,5 \text{ кг/гол.}$$

Если же принять гипотезу о росте случайных колебаний по абсолютной величине пропорционально росту среднего надоя, то средняя ошибка получит для ноября 2000 г. вид:

$$m_{\text{у ноября 2000}} = 11 \left(1 + \frac{15,86 \cdot 7}{316,7} \right) \sqrt{1 + \frac{1}{9} + \frac{49}{60}} = 14,86 \cdot 1,388 = 20,6 \text{ кг/гол.},$$

т.е. больше на треть, чем без учета тенденции роста абсолютного показателя случайной колеблемости. Соответственно, станут шире и доверительные интервалы прогнозов. Для проверки существования или отсутствия тренда случайной колеблемости необходим достаточно длительный временной ряд, а это, как уже отмечалось в разд. 6.4, не всегда имеется в условиях задачи. Если такая возможность есть, следует для расчета ошибок прогноза учесть тенденцию случайной колеблемости.

10.5 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, затем урожайность по их трендам, после чего перемножаем прогнозы:

$$\bar{x}_6 = 120 + 5 \cdot 6 = 150 \text{ га} ; \quad \bar{y}_6 = 29 + 6 = 35 \text{ ц/га};$$

$$\bar{y}_0 = \bar{x}_6 \cdot \bar{y}_6 = 150 \cdot 35 = 5250 \text{ ц};$$

б) перемножаем тренды площади и урожайности, и на основе полученной параболы вычисляем прогноз валового сбора:

$$\begin{aligned} \hat{y}_i &= [20 + 5t] \cdot [9 + t] = 3480 + 265t + 5t^2 \\ \hat{y}_6 &= 3480 + 265 \cdot 6 + 5 \cdot 6^2 = 5250 \text{ ц} \end{aligned}$$

Средняя ошибка выборочной оценки или прогноза произведения двух переменных вычисляется по формулам:

$$\text{а) } m_{\Pi \cdot y_k} = \sqrt{m_{\Pi_k}^2 \cdot \bar{y}_k^2 + m_{y_k}^2 \cdot \bar{\Pi}_k^2}, \quad (10.11)$$

если признаки Π и y независимы друг от друга;

$$\text{б) } m_{\Pi \cdot y_k} = \sqrt{m_{\Pi_k}^2 \cdot \bar{y}_k^2 + m_{y_k}^2 \cdot \bar{\Pi}_k^2 + 2 m_{\Pi_k} \cdot \bar{y}_k m_{y_k} \cdot \bar{\Pi}_k r'_{\Pi y}}, \quad (10.12)$$

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, вычисление 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:

$$m_{n_6} = S(t)_n \cdot \sqrt{1 + \frac{1}{9} + \frac{6^2}{60}} = 7,4 \sqrt{1 + \frac{1}{9} + \frac{6^2}{60}} = 13,16 \text{ га}$$

$$m_{y_6} = S(t)_y \cdot \sqrt{1 + \frac{1}{9} + \frac{6^2}{60}} = 4,4 \sqrt{1 + \frac{1}{9} + \frac{6^2}{60}} = 5,76 \text{ ц/га}$$

Подставляя в формулу средней ошибки прогноза валового сбора с учетом корреляции колебаний площади и урожайностью, имеем:

$$m_{y_6} = \sqrt{(13,16 \cdot 35)^2 + (5,76 \cdot 150)^2 - 2(13,16 \cdot 35)(5,76 \cdot 150)(-0,326)} = \sqrt{6991795} = 836,2 \text{ ц}$$

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 колеблемостью урожайности.

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. *Афанасьев В.Н.* Моделирование и прогнозирование временных рядов: учеб.-метод. пособие для вузов / В.Н. Афанасьев, Т.В. Лебедева. – М.: Финансы и статистика, 2009. – 292 с. ISBN 978-5-279-03402-4

5. *Афанасьев В.Н.* Статистические методы прогнозирования в экономике: учеб.-. пособие для вузов / В.Н. Афанасьев, Т.В. Лебедева. – М.: Финансы и статистика, 2009. – 180 с. ISBN 978-5-279-03401-7

6. *Афанасьев В.Н.* Статистическое исследование банковского кредитования: учебное пособие для вузов / В.Н. Афанасьев, Т.В. Лебедева, Т.В. Леушина; Оренбургский гос. ун-т. – Оренбург: ОГУ, 2011. – 300 с. ISBN 978-5-7410-1136-2

7. *Афанасьев В.Н., Цыпин А.П.* Эконометрика в пакете STATISTICA: учебное пособие по выполнению лабораторных работ / В.Н. Афанасьев, А.П. Цыпин, – Оренбург: ИП Кострицын, 2010. – 292 с. ISBN 978-5-91933-004-2

8. *Бобров С.П.* Конъюнктурные кривые 1922– 1923 гг. //Экономическое обозрение. – 1924. – № 5 – 7.

9. *Вайну Я.Я.Ф.* Корреляция рядов динамики. – М.: Финансы и статистика, 1982.

10. *Дубров А.М., Мхитарян В.С, Трошин Л.И.* Многомерные статистические методы. – М.: Финансы и статистика, 1999.

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

12. *Калинина М.С.* Статистические методы изучения динамики урожайности. – Л.: ЛСХИ, 1969.

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

14. *Крастинь О.П.* Изучение статистических зависимостей по многолетним данным. – М.: Финансы и статистика, 1981.

15. *Лапин Г.Ф.* Биометрия. – М.: Высшая школа, 1968.

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

17. *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.

18. Статистический анализ экономических временных рядов и

прогнозирование: ЦЭМИ АН СССР. – М.: Наука, 1973.

19. *Френкель А.А.* Математические методы анализа динамики и прогнозирования производительности труда. – М.: Экономика, 1972.

20. *Хенан Э.* Многомерные временные ряды: Пер. с англ. – М.: Мир, 1974.

21. *Четвериков И.С.* Статистические и стохастические исследования. – М.: Госстатиздат, 1963.

22. *Четыркин Е.М.* Статистические методы прогнозирования. – М.: Статистика, 1977.

23. *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*

24. *Юзбашев М.М.* Экспериментальная проверка закономерности распределения отклонений урожайности отдельных лет от тренда по их величине // Науч. труды ЛСХИ. – Т. 322. –Л. –Пушкин. – 1981.

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

26. *Юл Д.Э., Кендэл М.Д.* Теория статистики: Пер. с англ. – М.: Госстатиздат, 1960.

27. *Ястремский Б.С.* Переменная корреляция // Вестник статистики. – 1924. – Кн. XVII.

28. *Blanford D. Offut S.* A Review of Empirical Techniques for the Analysis of Commodity Instability. – USsl. – 1983.

Приложение № 1. Многократное скользящее выравнивание.
Урожайность зерновых культур во Франции

1979	1978	1977	1976	1975	1974	1973	1972	1971	1970	Годы
45.2	46.3	40.5	34.4	37.0	41.8	44.1	41.7	38.8	33.7	Ур-сть ц/га V_1
90.4	46.3	0	-34.4	-74.0	-125.4	-176.4	-208.5	-2,32.8	-235.9	Y_{it_1}
45.2	0	-40.5	-68.8	-111.0	-167.2	-220.5	-250.2	-271.6	-	Y_{it_2}
0	-46.3	-81.0	-103.2	-148	-209	-264.6	-291.9	-	-	Y_{it_3}
-45.2	-92.6	-121.5	-137.6	-185.0	-250.8	-308.7	-	-	-	Y_{it_4}
-90.4	-138.9	-162.0	-172.0	-222.0	-292.6	-	-	-	-	Y_{it_5}
-135.6	-185.2	-202.5	-206.4	-259.0	-	-	-	-	-	Y_{it_6}
-180.8	-231.5	-243	-204.8	-	-	-	-	-	-	Y_{it_7}
-226	-277.8	-283.5	-	-	-	-	-	-	-	Y_{it_8}
-271.2	-324.1	-	-	-	-	-	-	-	-	Y_{it_9}
-316.4	-	-	-	-	-	-	-	-	-	$Y_{it_{10}}$
-	-	-	-	-	-	-	-	-	-	$Y_{it_{11}}$
-	-	-	-	-	-	-	-	-	-	$Y_{it_{12}}$
46.2	44.7	43.3	41.8	40.4	38.9	37.5	36.0	34.6	33.1	Тренд
-1	1.6	-2.8	-7.4-	-3.4	2.9	6.6	5.7	-4.2	0.6	$Y_i - \epsilon$
1.00	2.56	7.84	54.76	11.56	8.41	43.56	32.49	17.64	0.36	$(Y_i - \epsilon)^2$

1991	1990	1989	1988	1987	1986	1985	1984	1983	1982	1981	1980
65.4	60.8	61.0	60.8	57.0	53.0	57.3	59.7	49.5	50.1	47.3	48.5
-	-	-	-	-	-	-	417.9	297.0	250.5	189.2	145.5
-	-	-	-	-	-	401.1	358.2	247.5	200.4	141.9	97.0
-	-		-	-	371	343.8	298.5	198	150.3	94.6	48.5
-	-		-	399.0	318.0	286.5	238.8	148.5	100.2	47.3	0
			425.6	342.0	265.0	289.2	179.1	99	50.1	0	-48.5
-	-	427.0	364.8	285.0'	212.0	171.9	119.4	49.5	0	-47.3	-97.0
-	425.0	366.0	304.0	228.0	159.0	114.6	59.7	0	-50.1	-94.6	-145.5
457.8	364.8	305.0	243.2	171.0	106.0	57.3	0	-49.5	-100.2	-141.9	-194
392.4	304.0	244.0	182.4	114.0	53.0	0	-59.7	-99	-150.3	-189.2	-242.5
327.0	243.2	183.0	121.6	57.0	0	-57.3	-119.4	-148.5	-200.4	-236.5	-291
261.6	184.4	122.0	60.8	0	-53.0	-114.6	-179.1	-198.5	-250.5	-283.8	-339.5
196.2	121.6	61.0	0	-57.0	-106.0	-171.9	-238.8	-247.5	-300.6	-331.1	-
63.6	62.1	60.7	59.2	57.8	56.3	54.9	53.4	52.0	50.5	49.5	47.6
1.8	1.3	0.3	1.6	-0.8	-3.3	2.4	6.3	-2.5	-0.4	-1.8	0.9
3.24	1.69	0.09	2.56	0.64	10.89	5.76	39.69	6.25	0.16	3.24	0.81

Σ	1995	1994	1993	1992
1332.6	63.2	65.5	65.1	64.9
349.9	-	-	-	-
361.4	-	-	-	-
359.7	-	-	-	-
396.9	-	-	-	-
463.6	-	-	-	-
492.6	-	-	-	-
470.0	-	-	-	-
432.2	-	-	-	-
408.1	-	-	-	454.3
407.4	-	-	455.7	389.4
379.9	-	458.5	390.6	324.5
346.4	442.4	393.0	325.5	259.6
1334.5	69.4	67.9	66.5	65.0
0	-6.2	-2.4	-1.4	-0.1
301.37	38.44	5,76	1.96 j	0.01

$b_1=1.25; b_2=1.29; b_3=1.29; b_4=1.42; b_5=1.66; b_6=1.77; b_7=1.68; b_8=1.54;$
 $b_9=1.46; b_{10}=1.46; b_{11}=1.36; b_{12}=1.2$

$$b_i = \frac{\sum Y_i t_i}{280}$$

$$\bar{b} = \frac{\sum_{i=1}^{12} b_i}{12} = 1,452. \text{ Тренд } \hat{y} = 51,25 + 1,452t, t=0,5 \text{ в } 1983$$

ЧАСТЬ 2. ДОПОЛНИТЕЛЬНЫЕ ГЛАВЫ

Введение

Современное общественное развитие характеризуется нарастанием роли информации, глобализацией национальных экономик, а вследствие этого — высокими требованиями к анализу и прогнозированию динамики процессов. Проблема получения достоверных и адекватных результатов включает в себя задачу развития существующей методологии статистического анализа и прогнозирования временных рядов.

В данном разделе учебника представлен ряд основных предпосылок и гипотез, выполнение которых является критически важным для возможности применения методов анализа. К этим предпосылкам относятся: сама возможность прогнозирования явления; требование сопоставимости уровней временного ряда; вероятностный характер изучаемого явления; осведомлённость о предыстории изучаемого процесса; наложение на природу исходных данных какой-либо структуры (стационарность, эргодичность и т.п.); выделение компонентов динамики (тренд, циклическая, сезонная и случайная колеблемость) и определение модели их взаимосвязи (мультипликативной, аддитивной или смешанной).

Нами перечислены принципы, являющиеся крайне важными, а иногда и критическими, для получения достоверных результатов: предварительное изучение графика и коррелограммы исходного временного ряда; использование формализованного тестирования для определения характеристик временного ряда; верификация использованного метода или модели на данных, не участвовавших в их спецификации и идентификации.

Данная работа знакомит исследователей (магистрантов, аспирантов) с альтернативными и продвинутыми методами анализа, нацеленными на причинное понимание наблюдаемых экономических процессов. Большинство из представленных методов реализованы в современных программных продуктах и проиллюстрированы примерами на основе реальных статистических данных. Последний раздел посвящен подробному описанию этапов анализа и прогнозирования индекса реальных инвестиций в Российской Федерации и может служить читателю отправной точкой собственного практического исследования.

В дополнительных главах учебника представлена информация, дополняющая содержание учебника В.Н. Афанасьева и М.М. Юзбашева «Анализ временных рядов и прогнозирование» (2001 и 2010 гг. издания, изд. «Финансы и статистика») с грифом «Рекомендовано Министерством образования Российской Федерации в качестве учебника для студентов высших учебных заведений по направлению «Статистика». Примеры взяты из диссертации на соискание ученой степени кандидата экономических наук, Любича Вячеслава, выполненной под моим руководством. Переиздание учебника посвящено памяти моего учителя Михаила Михайловича Юзбашева.

*Избранный член Международного статистического института (ISI)
доктор экономических наук профессор В.Н. Афанасьев*

Глава 11 Международный опыт в развитии методологии статистического анализа временных рядов

Что необходимо знать из главы 11:

1. Основные практические принципы моделирования временного ряда и причины, по которым они необходимы.

2. Подходы к периодизации временных рядов и анализу лаговых зависимостей.

3. Как использовать альтернативные методы сглаживания, уметь интерпретировать результаты рекурсивных расчётов и вейвлет-анализа.

4. Основы анализа синхронности колебаний, коинтеграции временных рядов и их роль в исследовании причинности.

11.1 Принципы практического моделирования временного ряда

Теоретические положения построения временных рядов, указанные в главе 1 учебника В.Н. Афанасьева и М.М. Юзбашева [3], учитываются при непосредственном применении методов и моделей временных рядов⁴. Тем не менее, повышение качества анализа напрямую зависит от использования следующих эмпирических принципов:

- 1) графическое отображение временного ряда;
- 2) формализованное определение свойств временного ряда;
- 3) проверка качества моделирования и прогнозирования на основе данных, не вошедших в выборку.

Один из самых важных и первостепенных этапов анализа временного ряда — его графическое отображение. Благодаря современному программному обеспечению построение сколь угодно сложного графика (в том числе, трёхмерного) не составляет особого труда. По графику предварительно определяется тенденцию развития явления, её однородность, характер колеблемости, наличие выбросов. Это значительно облегчает труд исследователя и способствует направленным поискам необходимого метода анализа.

В работах по анализу временных рядов и прогнозированию значительное внимание уделяется графическому анализу. В пособии Европейской экономической комиссии ООН справедливо отмечается, что изображение может заменить собой тысячу слов, но если график слишком перегружен, то он становится головоломкой и способен привести читателя к ложным выводам [44, с. 8]. Поэтому тема грамотного построения графиков остаётся актуальной даже при использовании современных программных средств.

Внимания требуют следующие аспекты построения графика:

⁴Здесь и далее: «модель» — это математическая интерпретация реальности, в то время как «метод» — это способ, правило или образ действия.

- выбор нулевых точек, так как это важное решение для формирования общего вида графика, возможности оценить наличие, характер тенденции и колебаний;
- определение масштаба, шкалы (линейной, логарифмической, вероятностной или другой), обозначений шкал;
- использование сплошных линий вместо пунктира и закрашенных областей, отказ от маркеров точек данных с целью снизить «зашумлённость». Легенду к графику можно заменить необходимыми подписями рядом с объектами (данное замечание не относится к картограммам);
- подписи значений (если они необходимы) не должны мешать читателю воспринимать общую картину;
- отображение крупным планом важного участка графика в отдельном окне;
- присутствие на графике координатной сетки, асимптот, нормативных значений и прочих линий;
- наложение графиков других показателей того же или иного типа (например, наложение столбиковой диаграммы и линейного графика);
- отображение на графике линии, уравнения тренда и других расчётных характеристик временного ряда;
- использование двумерных графиков для двумерного массива данных, чтобы не возникало искажение пропорций.

Цветовое, шрифтовое, композиционное оформление графика должно способствовать его лёгкому прочтению, вставке в материалы научной работы, доклад или презентацию с сохранением свойств читаемости и наглядности (в том числе, при публикации в чёрно-белом формате).

Для лёгкости прочтения текста на графиках требуется соблюдать правила грамматики, делать надписи слева направо, не использовать аббревиатуры и акронимы, а также написание целых слов строчными буквами [44, с. 9].

Простой временной график — наиболее используемый, несложный и полезный тип графика. Другие типы, например, гистограммы, достаточно трудно применить ко временным рядам. Тем не менее, гистограмма распределения временного ряда, в котором исключена тенденция и сезонность, применяется при изучении выбросов. После простого временного графика рассматривают график корреляционной функции — коррелограмму.

Коррелограмма используется как наглядное средство описания свойств динамики. В общем случае коррелограмма является оценкой значений автокорреляционной функции генерирующего ряд процесса, если исследуемый временной ряд стационарный. Если он таковым не является, то коррелограмма не показывает ничего, кроме нестационарности временного ряда и необходимости применения способов удаления тренда для достижения стационарности.

Графики автокорреляционной и частной автокорреляционной функции используются в методологии Бокса-Дженкинса при спецификации моделей

ARMA. При этом высокие значения (пики на графике), повторяющиеся с фиксированным интервалом, говорят о наличии сезонности и приводят к моделям SARMA.

Исключительно для отображения сезонных колебаний в анализе временных рядов применяется лепестковая диаграмма. Это удобное средство сравнения средних показателей сезонности различных временных рядов или их подпериодов (при наложении нескольких графиков друг на друга), однако неэффективное при изменяющейся сезонной волне.

Таким образом, графическое изображение временного ряда широко используется в анализе, так как значительно помогает исследователю. Тем не менее, в отдельных случаях даже правильно построенный график не даёт чёткого представления, является ли временной ряд стационарным или «близким к нестационарному», является ли он реализацией линейного или нелинейного процесса и так далее. Поэтому необходимы специальные информационные критерии и тесты. В этом и состоит следующий принцип: состоятельные результаты требуют формализованных способов определения свойств временного ряда.

Выяснение того, стационарен ли временной ряд или нет, — одна из важных практических проблем. Она осложняется тем, что процесс может оказаться «близким к нестационарному». Примером является стационарный процесс $AR(1)$ с параметром ϕ , близким к единице:

$$X_t = \phi X_{t-1} + Z_t, \quad (11.1)$$

где $0 < |\phi| < 1$;

Z_t — белый шум.

Временной ряд (особенно, короткий), сгенерированный таким процессом, имеет график и свойства, схожие с (нестационарным) случайным блужданием [29, с. 46]:

$$X_t = X_{t-1} + Z_t. \quad (11.2)$$

Краткосрочные прогнозы по моделям (1.1) и (1.2) близки друг к другу, но не долгосрочные. Существенное различие в том, что долгосрочный прогноз по модели $AR(1)$ стремится к общему среднему, в то время как все прогнозы случайного блуждания равны наиболее позднему наблюдаемому значению.

Автокорреляционная функция нестационарных процессов, процессов с длинной памятью и других близких к «нестационарным» процессов затухает очень медленно. В связи с этим, при работе с данными процессами трудно определить, равен ли единице параметр d в модели $ARIMA(p, d, q)$, или $d = 0$. Поиск решения указанной проблемы привёл исследователей к созданию тестов на единичный корень.

Тест на единичный корень проверяет гипотезу о нестационарности временного ряда с использованием авторегрессионной модели. Перед

вычислением тестовых статистик решается вопрос спецификации: включать ли константу, константу и линейный тренд или ничего из этого не включать в тестовое уравнение регрессии. Один из подходов — выполнить тестирование вместе и с константой, и с линейным трендом, так как два других варианта — лишь частные случаи этой более общей формы. Но включение неподходящих регрессоров в уравнение снижает мощность теста и увеличивает вероятность ошибочных выводов о наличии единичного корня, в то время как фактически его нет. Поэтому общим принципом спецификации является правдоподобное описание данных как в нулевой, так и в альтернативной гипотезах. Если во временном ряду прослеживается тренд, то в тестируемую регрессию необходимо включить тренд и свободный член. Если тренда не прослеживается, но среднее значение не равно нулю, то учитывают лишь свободный член. Если ряд колеблется около нулевого среднего, то в тестовую регрессию не включается ни тренд, ни константа.

Тест Дикки-Фуллера (Dickey-Fuller; DF) пригоден, только если временной ряд является реализацией процесса $AR(1)$. При наличии автокорреляции более высоких порядков применяют тесты ADF (AugmentedDickey-Fuller) и PP (Phillips-Perron). Отличие состоит в подходах к контролю корреляции высоких порядков в рядах: ADF-тест выполняет добавление в тестируемое уравнение членов в виде лаговых разностей, в то время как PP-тест корректирует тестовую статистику для коэффициента уравнения $AR(1)$ -регрессии, чтобы учесть серийную корреляцию в остатках.

В работе Ф. Дайболда и Л. Килиана [31] представлены результаты исследования, подтверждающие повышение точности прогнозирования канонической модели $AR(1)$ с трендом для различных размеров выборки, горизонтов прогнозирования и степеней присутствия тенденции. Ими оценены преимущества предварительного тестирования на единичный корень по сравнению с обычным взятием первых разностей и построением модели по исходным уровням временного ряда.

Однако Р. Тейлор отмечает, что большинство тестов на единичный корень имеет малую мощность даже для выборок среднего размера, потому что альтернативная гипотеза «близка» к нулевой гипотезе, а сама процедура тестирования чувствительна к моделируемой лаговой структуре [47].

Рассуждая над проблемой тестирования, К. Четфилд делает вывод, что выбор подходящего описания временного ряда крайне важен для моделирования и прогнозирования, однако формальный тест на единичный корень помогает лишь незначительно. Факт, что нулевая гипотеза (наличие единичного корня) не отклонена, не означает, что она обязательно должна быть принята, например, если необходимо точно оценить тренд, а не исключить тенденцию. С другой стороны, могут существовать практические причины для вычисления первых разностей, даже если наличие единичного корня отрицается: например, если модель, полученная для разностей, менее чувствительна к выборокам. Ключевой вопрос для прогнозирования не в том, помогает ли тест на единичный корень выбрать «правильную модель», а в том, даёт ли выбранная модель вневыборочные прогнозы лучше, чем

альтернативные модели [29, с. 48]. Не менее важной при построении и выборе моделей должна оставаться экономическая сущность изучаемых данных, чтобы не променять интерпретируемость результатов на их математическую обоснованность.

Эконометристы тестируют временные ряды не только на наличие единичного корня, но и на постоянство дисперсии, автокорреляцию остатков и другие свойства. Статистикам, напротив, свойственно выбирать общий класс моделей, а затем отдавать предпочтение той, которой соответствует минимальное значение информационного критерия. Такие критерии помогают автоматически выбрать модель из уже имеющихся кандидатов, основываясь одновременно на величине ошибок и числе оцениваемых параметров. Например, информационный критерий Акайки (Akaike's information criterion; AIC) в несмещённом виде рассчитывается по формуле:

$$AICc = \ln \hat{\sigma}_k^2 + \frac{n+k}{n-k-2} \quad (11.3)$$

где c означает, что изначальная формула критерия скорректирована;

$\hat{\sigma}_k^2$ — дисперсия ошибок модели;

n — размер выборки;

k — число параметров в модели.

Критерий Акайки базируется на обобщении принципа максимального правдоподобия. Приведённое выражение подразумевает, что случайное возмущение является гауссовым.

Используя те же данные, получают значения информационного критерия Шварца (Schwarz's information criterion; SIC):

$$SIC = \ln \hat{\sigma}_k^2 + \frac{k \ln n}{n}. \quad (11.4)$$

Критерий Шварца имеет более фундаментальное теоретическое обоснование. Он базируется на байесовском подходе, поэтому его называют байесовским информационным критерием (Bayesian information criterion; BIC). Оценка порядка модели по этому критерию является состоятельной, но на практике чаще используется AICc [12, с. 255]. Исследования подтверждают, что SIC лучше работает при больших выборках, в то время как AICc — при малых выборках, где относительное число параметров больше [45, с. 54]. Данные критерии, в отличие от описанных тестов, применяются уже после оценки параметров альтернативных моделей.

Следующим принципом моделирования временных рядов является проверка качества моделирования и прогнозирования на основе данных, не вошедших в выборку. Основной целью моделирования является получение *интерпретируемой модели* какого-либо процесса, отражение присущих

связей и взаимозависимостей, а целью прогнозирования — получение *прогноза*. Временной ряд — это выборка, частный случай, и главным является не доскональное описание и приближение к данным выборки, а изучение генеральной совокупности. Поэтому оценка качества моделей и точности прогнозов не должна проводиться на тех же выборочных данных. Формально статистическая теория требует, чтобы модель формулировалась на одном массиве данных, оценивалась на втором и проверялась на третьем. Однако это невозможно в случае анализа временных рядов, где присутствует лишь одна реализация стохастического процесса. Разделение одного временного ряда на три части неравноценно трём независимым выборкам. К тому же, ряд может оказаться слишком коротким для дробления на три пригодные для анализа части. На практике применяется своеобразный компромиссный вариант разделения исходного временного ряда на две части: первая используется для оценки параметров, а вторая (так называемая *послевыборочная* или *вневыборочная*) — для тестирования, оценки качества модели и прогнозов. Свойства получаемых прогнозов при этом сходны с настоящими прогнозами, и на практике они не так точны, как внутри используемой выборки.

К сожалению, во многих исследованиях данный принцип проверки качества моделирования (которое стремится раскрыть некие *общие* зависимости) и прогнозирования (которое должно предсказывать будущие значения показателя, *заранее неизвестные*) не используется в достаточной мере. Выводы и предложения зачастую формулируются на основе результатов, демонстрирующих ложно хорошее приближение к данным. Существуют способы сомнительной обоснованности, с помощью которых результаты «улучшаются», поэтому при изучении аналитических работ важно убедиться, действительно ли все прогнозы выполнены при одних и тех же условиях. Например, если подразумевается, что прогнозы являются *послевыборочными*, то все сравниваемые методы должны использовать лишь историческую информацию анализируемого массива данных, не выходя за его пределы [29, с. 8]. «Улучшению» результатов способствует:

- оценка параметров модели на всём массиве данных, включая тестовый массив;
- использование известных «будущих» значений объясняющих переменных тестового массива при многомерном прогнозировании. Очевидно, что это улучшает прогноз зависимой переменной в тестовой части временного интервала, однако в момент построения реального прогноза данные значения неизвестны. Экономисты называют подобные прогнозы *экспост* прогнозами (*ex-postforecasts*), чтобы отличать их от *экс-анте* прогнозов (*ex-anteforecasts*). Последние, являясь истинными *послевыборочными* прогнозами, порой подразумевают получение прогнозов объясняющих переменных, для прогнозирования зависимой переменной [29, с. 9].

Стоит отметить, что некоторые методы прогнозирования явно или неявно используют дополнительную информацию, характер которой не

принимается во внимание другими методами. Чтобы учесть такие преимущества при сравнении требуется исчерпывающее описание того, как прогнозы были получены. В отдельных случаях прогнозирование, альтернативное истинно послевыборочному, оказывается предпочтительным. Например, в сценарном прогнозировании используют экс-пост прогнозы при оценке эффекта различных объясняющих переменных. Однако исследователь должен понимать, что сравнивать такие прогнозы с истинными послевыборочными неправомерно.

11.2 Скользящие и рекурсивные оценки в системе методов эконометрического анализа одномерного временного ряда

Анализ одномерных временных рядов традиционно широко применяется в практической деятельности. Одномерный анализ имеет ряд ценных преимуществ перед многомерным анализом:

- сокращение трудовых и материальных затрат на сбор и подготовку первичной информации, оценку её значимости, нормирование;
- отсутствие необходимости обоснования включения тех или иных факторов, изучения факторной структуры;
- исключение многих проблем, связанных с мультиколлинеарностью;
- возможность использовать более простые и понятные методы;
- лёгкость расчётов, оперативное получение результата;
- высокая наглядность и интерпретируемость результатов.

Важным шагом анализа временного ряда является определение длины периода времени, за который или по состоянию на который приводятся уровни временного ряда. Выбор длины интервала должен соотноситься с целями анализа. Затем, с учётом характера полученных данных, подбираются методы анализа. При этом однородность развития явления определяет, можно ли анализировать весь временной ряд целиком, или требуется предварительно разделить его на части. Данная задача решается с помощью методов периодизации временных рядов.

Наиболее распространённые методы периодизации:

- историческая. Выделение периодов осуществляется в соответствии с теоретическим анализом применительно к той проблеме, в рамках которой рассматривается изучаемый процесс. При этом учитывают, прежде всего, крупные аномалии (войны, эпидемии, землетрясения), радикальные реформы управления, целевые комплексные программы, кризисные явления (в том числе экономические депрессии, дефолты), другие радикальные рыночные изменения в изучаемой области (демонополизация отрасли, создание продукта нового поколения);

– параллельная. Данный метод основывается на использовании второго, детерминирующего показателя, периодизация которого заранее известна. Однако осуществить параллельную периодизацию достаточно сложно, так как возникают трудности с нахождением такого показателя, который, кроме известной объективной периодизации, должен обладать редкими свойствами связи с анализируемым показателем.

– фазовый анализ. В соответствии с методикой, описанной в работе Ю. П. Лукашина [17, глава 10], флуктуации — отклонения текущих значений показателя от некоторого установленного исследователем уровня — образуют фазы. Фаза представляет собой период положительных или отрицательных флуктуаций. Основная идея метода заключается в постепенной очистке временного ряда от маломощных фаз, отождествляемых со случайными или второстепенными колебаниями. После определения критерия остановки итеративного процесса агрегирования фаз и непосредственного осуществления агрегирования получают некоторое разбиение временного ряда точками смены фаз на подпериоды. В каждом из подпериодов движение ряда можно представить упрощённым образом.

В работе В. В. Глинского [8] рассмотрена возможность применения многомерной средней величины и факторного анализа (метода главных компонентов и метода главных факторов) для периодизации состояний процесса при обработке уровней ряда динамики, а также периодизации динамики процесса. Во втором случае в качестве исходных данных берутся показатели динамики: абсолютные приросты, темпы роста и другие. Однако факторный анализ не учитывает хронологическую последовательность уровней временных рядов. Результаты факторного анализа, представленные в работе [8], не приводят к очевидным выводам относительно периодов динамики; периодизация временного ряда определяется во многом не только результатами метода главных компонентов, но и знанием особенностей развития экономики в изучаемом периоде, экспертными суждениями автора.

Более простым и обоснованным, на наш взгляд, является фазовый анализ. Например, временной ряд, представленный на рисунке 11.1, неоднороден по характеру тенденции: наблюдавшийся длительное время рост сменяется падением производства. В соответствии с этим проведено агрегирование фаз до получения двух подпериодов. Размер флуктуации определён как цепной абсолютный прирост производства молока, применена стандартная схема замены наименее мощной фазы и ближайших соседних фаз одной объединённой фазой. Полученная фазовая диаграмма наглядно отображает кардинальные изменения динамики производства молока в Оренбургской области. В соответствии с рисунком 11.1, перелом произошёл в 1990 г. Полученные выводы совпадают с теми, которые могли быть получены при использовании метода исторической периодизации: потрясения в социально-экономической жизни страны в результате распада Советского Союза привели к упадку экономики.

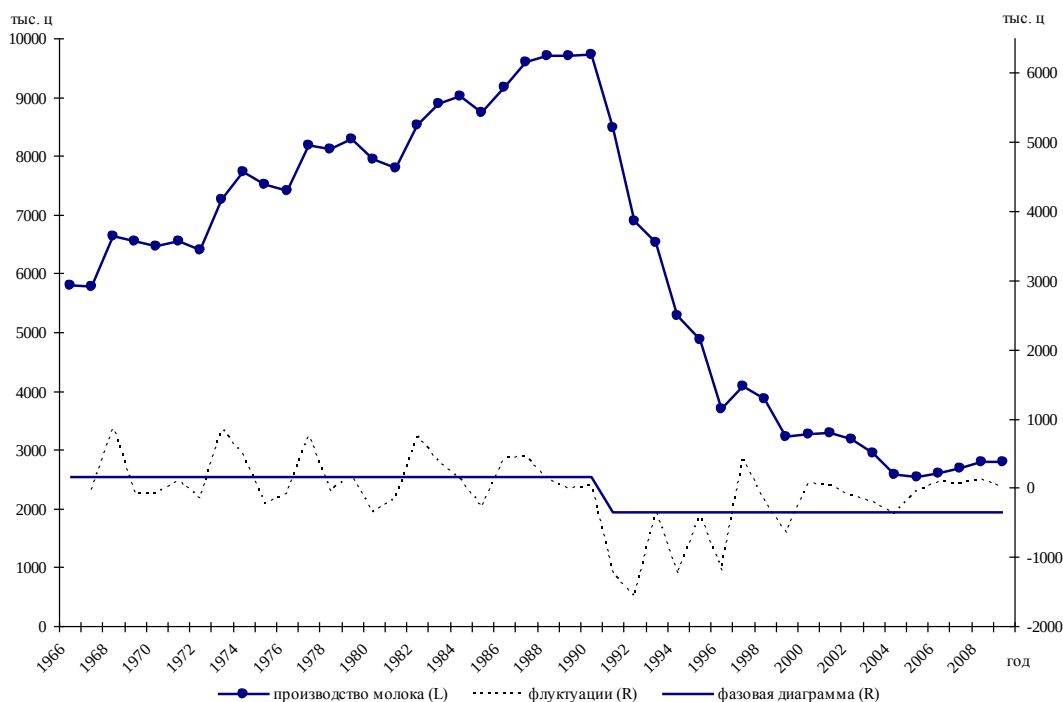


Рисунок 11.1 — Динамика производства, график флуктуаций и фазовая диаграмма производства молока сельскохозяйственными организациями Оренбургской области в 1966–2009 гг.

Таким образом, фазовый анализ способен выделить во временном ряду длинные и мощные волны для их индивидуального или совместного анализа, позволяет формализованным путём определить точки смены фаз на временной оси. Метод отличается гибкостью, простотой и ясностью концепции.

При переходе к сезонным данным (дневным, еженедельным, месячным или квартальным) эффективный метод спектрального анализа и процедура сезонной корректировки «Х-11». Для изучения тренда и циклической составляющей рекомендуется использовать годовые уровни, максимально освобождённые от случайных и сезонных колебаний.

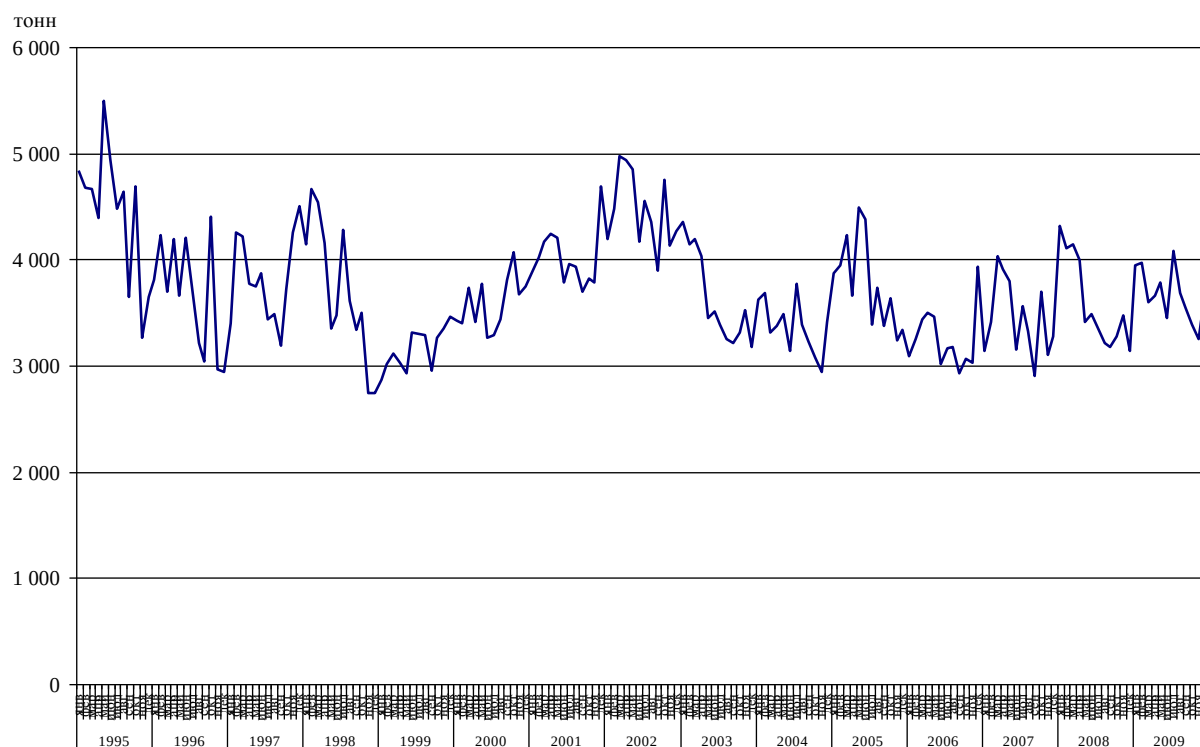


Рисунок 11.2 — Динамика производства цельномолочной продукции в Оренбургской области (в пересчёте на молоко)

Рассмотрим временной ряд производства молочной продукции в крупных и средних сельскохозяйственных организациях Оренбургской области, собранный на основе статистических бюллетеней за 1995–2009 гг. Исходя из месячных данных, представленных на рисунке 11.2, чётко выраженная тенденция рассматриваемого показателя не прослеживается, сезонность выражена слабо. Для выявления закономерностей, скрытых за случайной колеблемостью, применено агрегирование (рисунки 11.3 и 11.4), но после взаимопогашения колебаний тренд не наблюдается, что указывает на стационарность временного ряда.

Графики на рисунках 11.2–11.4 отличаются разной степенью сглаженности, придающей временному ряду визуально естественный вид. В процессе сглаживания уровней повышаются коэффициенты автокорреляции, а вместе с тем, и предсказуемость временного ряда. При этом в результате погашения колебаний путём укрупнения интервалов утеряно значительное число уровней ряда (исходные 180 уровней были преобразованы в 60 при агрегировании по кварталам и в 15 — при агрегировании по годам). Альтернативные процедуры сглаживания позволяют избежать данной проблемы.

В статистике широко известны возможности механического сглаживания с помощью скользящей средней и аналитического сглаживания с использованием полиномиальных функций. Оба этих метода применимы для определения наличия тенденции, её направлении развития (рисунок 11.5), однако каждый из них имеет недостатки.

Скользящая средняя сокращает количество уровней ряда с обоих концов. Особенно отрицательным моментом является нехватка наиболее поздних, актуальных сглаженных уровней ряда, что негативно сказывается на его восприятии и возможностях принятия перспективных решений.

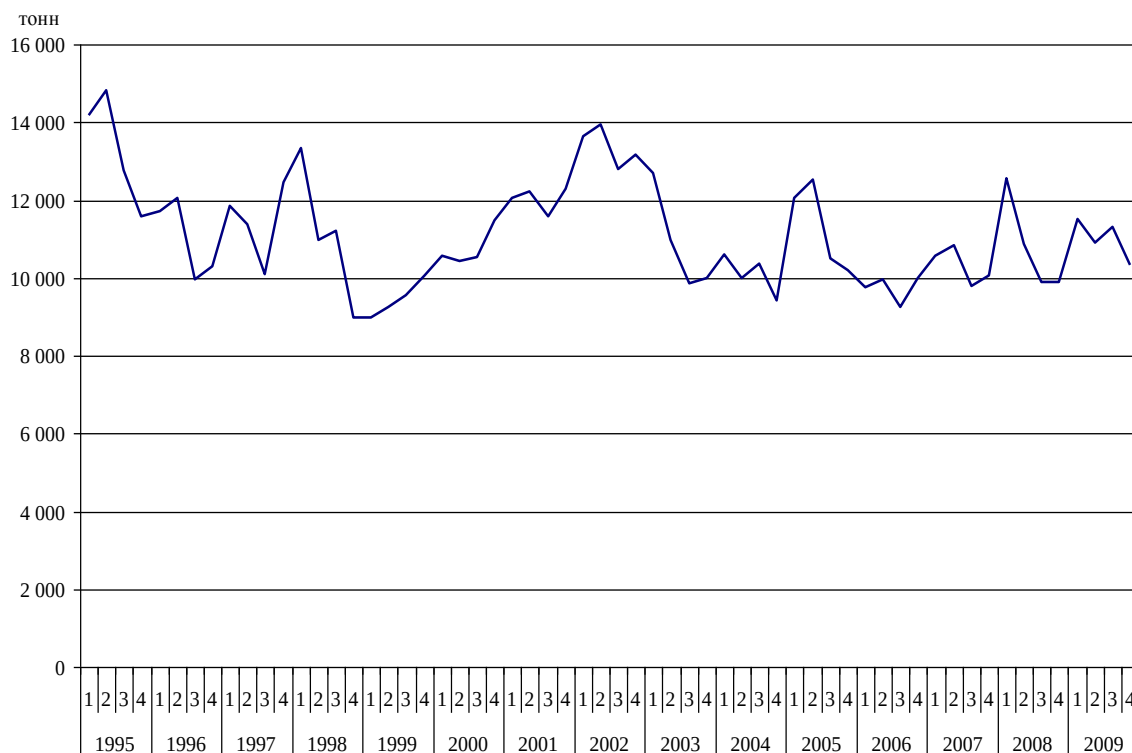


Рисунок 11.3 — Динамика производства цельномолочной продукции в Оренбургской области (в пересчёте на молоко), агрегированная по кварталам

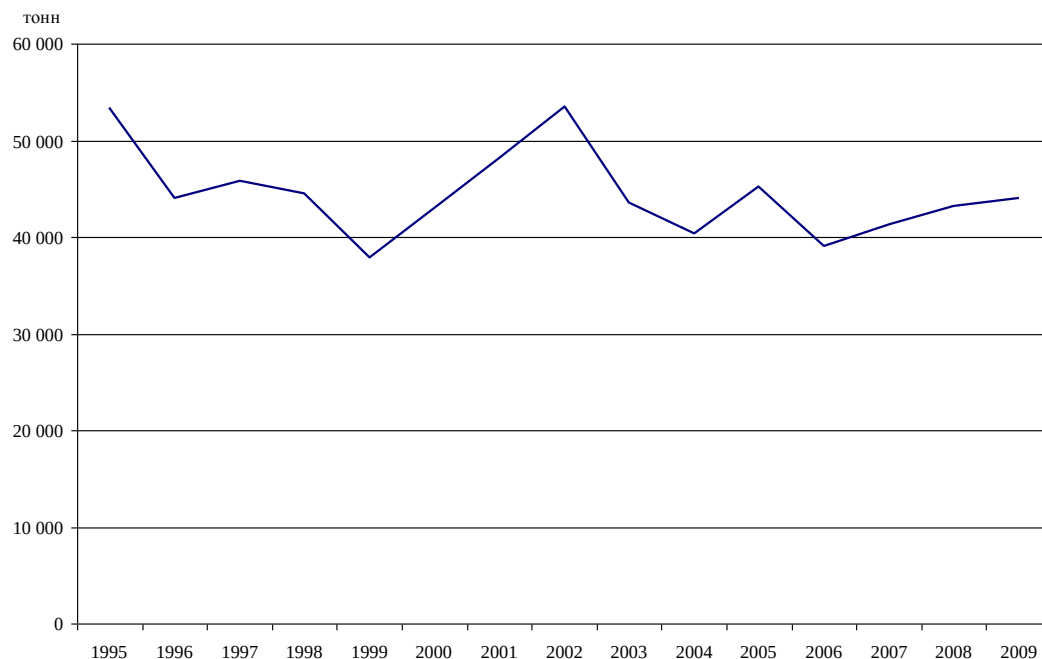


Рисунок 11.4 — Динамика производства цельномолочной продукции в Оренбургской области (в пересчёте на молоко), агрегированная по годам

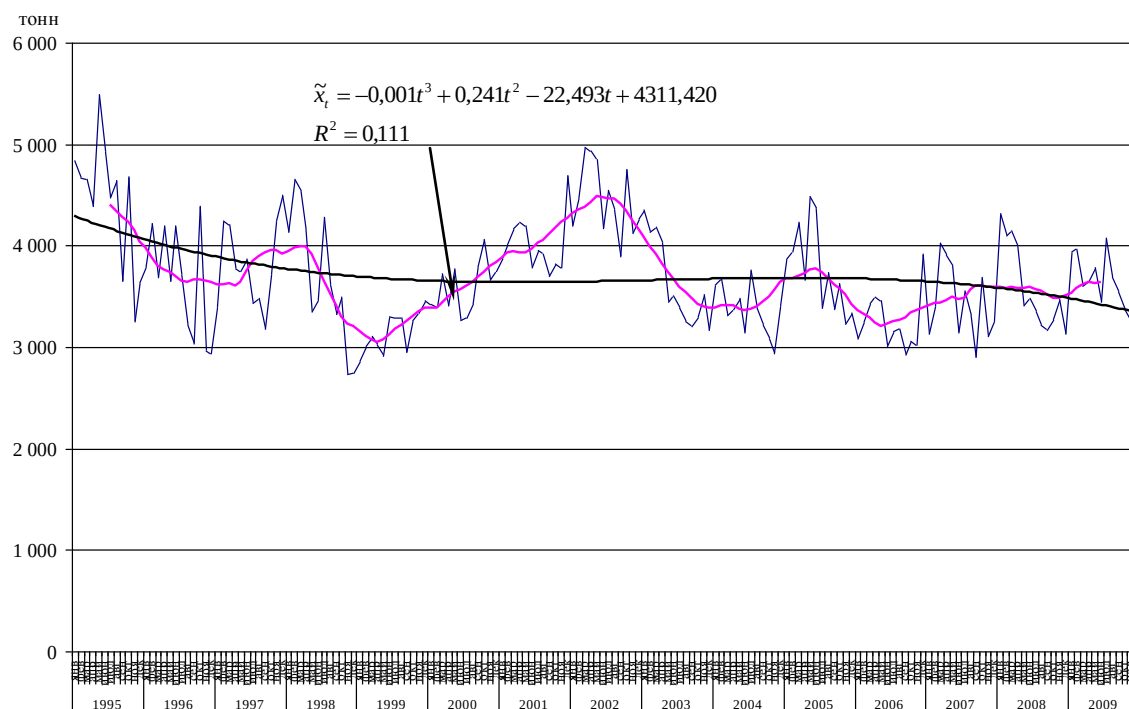


Рисунок 11.5 — Динамика производства цельномолочной продукции в Оренбургской области (в пересчёте на молоко), сглаженная с помощью 12-месячной скользящей средней и полинома третьей степени

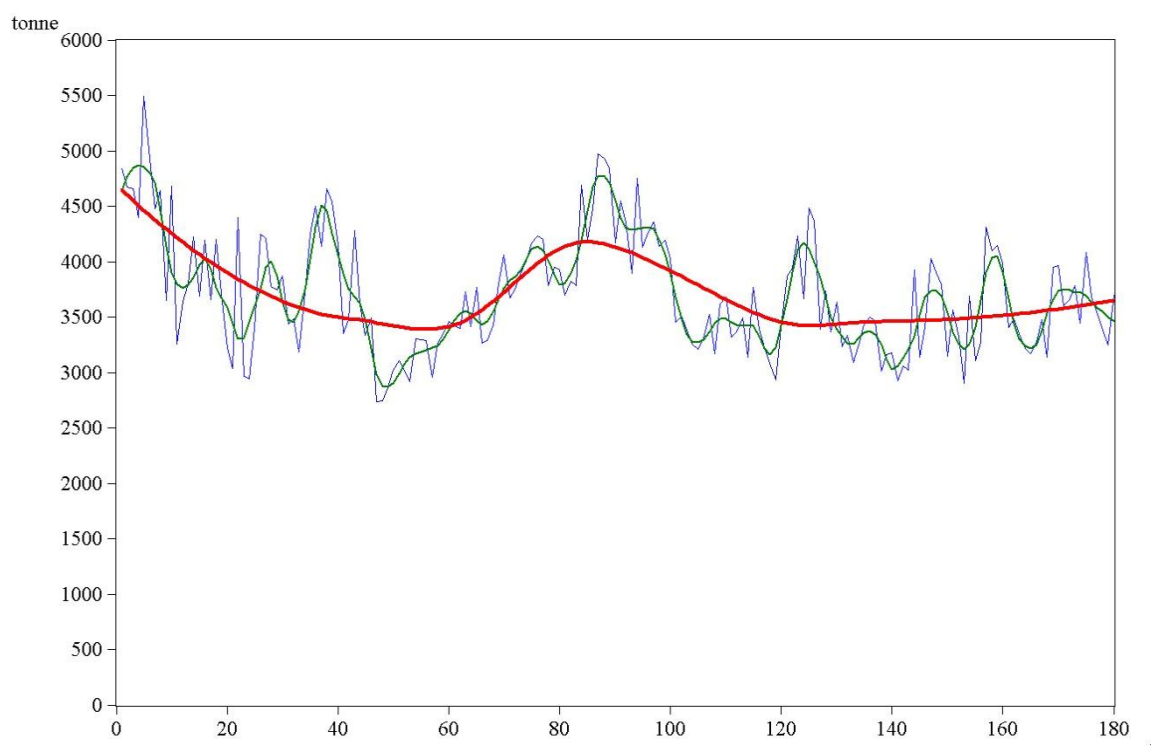


Рисунок 11.6 — Сглаживание динамики производства цельномолочной продукции в Оренбургской области (в пересчёте на молоко) с помощью локально взвешенной регрессии

При сглаживании по аналитической функции временной ряд представляется в виде:

$$x_t = f_t + y_t, \quad (11.5)$$

где f_t — функция от времени, а y_t — стационарный процесс [45, с. 72].

При этом уровни ряда не теряются, полученное уравнение пригодно для исключения тенденции, прогнозирования, при малых степенях полиномов параметры имеют экономическую интерпретацию. Всё это делает аналитическое сглаживание временных рядов с использованием функций от времени (в том числе и периодических) очень популярным на практике. Но применение подобных функций имеет свои ограничения. Во-первых, исследуемая реализация процесса должна иметь однородную тенденцию. Функция подбирается ко всему временному ряду в целом, усреднённо, и приближение к фактическим данным на некоторых промежутках оказывается недостаточным (рисунок 11.5). Во-вторых, при оценке параметров уравнения должны выполняться предпосылки метода наименьших квадратов. В-третьих, необходимо найти компромисс между повышением доли объяснённой вариации за счёт увеличения числа параметров и сохранением интерпретируемости, пригодности модели для анализа. В-четвёртых, при добавлении новых уровней временного ряда требуется повторять все шаги построения модели, при этом последняя, может кардинально измениться. В-пятых, полученные результаты должны быть статистически значимыми.

Альтернативными способами сглаживания служат:

1) применение весовой функции (ядра) для усреднения наблюдений, когда оценка f_t в (11.5) имеет вид:

$$\hat{f}_t = \sum_{i=1}^n \omega_t(i) x_t, \quad (11.6)$$

где $\omega_t(i)$ — функция Нарадая-Уотсона (Naradaya-Watson estimator):

$$\omega_t(i) = K\left(\frac{t-i}{b}\right) / \sum_{j=1}^n K\left(\frac{t-j}{b}\right), \quad (11.7)$$

а функция ядра $K(\cdot)$ типичной нормальной формы имеет вид [45, с. 74]:

$$K(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-z^2/2\right); \quad (11.8)$$

2) сглаживание с помощью локальной скользящей регрессии. Данный метод основан на построении регрессии $\hat{f}_t$ на k ближайших друг к другу уровнях временного ряда $\{x_{t-n/2}, \dots, x_t, \dots, x_{t+m/2}\}$, $n+m=k$, при $n=m$ интервал является симметричным. Развитие и усложнение этого метода основано на применении локально взвешенной регрессии, при которой сначала с помощью схемы весов значениям временного ряда, наиболее близким к x_t , присваивается больший вес, а затем применяется робастная взвешенная

регрессия для сглаженной оценки f_t . Чем больше ширина подпериода k , тем более сглаженной получается линия. На рисунке 11.6 представлены результаты сглаживания линейной локально взвешенной регрессией с использованием 50% данных для оценки трендовой составляющей и 7% для оценки сезонных колебаний;

3) сглаживание сплайнами — представляет собой изменённый метод полиномиальной регрессии, когда период времени $t=1, \dots, n$ делится узловыми значениями $t_0, t_1, \dots, t_k$ на k интервалов $[t_0=1, t_1], [t_1+1, t_2], \dots, [t_{k-1}+1, t_k=n]$, и на каждом интервале оценивается регрессия вида:

$$f_t = \beta_0 + \beta_1 t + \dots + \beta_p t^p. \quad (11.9)$$

Широко распространены кубические сплайны ($p=3$).

В поисках компромисса между приближением сглаженной линии к реальным данным и степенью её сглаженности минимизируют сумму

$$\sum_{i=1}^n [y_i - f_t]^2 + \lambda \int_0^1 f_t''^2 dt, \quad (11.10)$$

где f_t — кубический сплайн с узлом в каждом t , а степень сглаживания контролируется параметром $\lambda > 0$;

4) сглаживание одного временного ряда как функции от другого. При этом к полигону распределения двух одновременно наблюдаемых признаков подбирается кривая с помощью локально взвешенной регрессии или сплайнов. Данный метод выходит за рамки одномерного анализа и применим лишь в предположении, что пары наблюдений изучаемых признаков представляют собой независимые одинаково распределённые величины;

5) фильтр Ходрика-Прескотта (Hodrick-Prescott filter), широко используемый в макроэкономике для получения сглаженной оценки трендовых компонентов временных рядов. Его изобретение связано с анализом послевоенных бизнес-циклов в начале 1980-х гг. С технической точки зрения, это двусторонний линейный фильтр, который рассчитывает сглаженные значения s_t , минимизируя сумму:

$$\sum_{t=1}^T (x_t - s_t)^2 + \lambda \sum_{t=2}^{T-1} ((s_{t+1} - s_t) - (s_t - s_{t-1}))^2. \quad (11.11)$$

Чем больше параметр λ , тем более сглаженным получится временной ряд. При $\lambda = \infty$ ряд s представляет собой линейный тренд. Авторами теста предложены значения:

$$\lambda = \begin{cases} 100 & \text{для годовых данных;} \\ 1600 & \text{для квартальных данных;} \\ 14400 & \text{для месячных данных.} \end{cases} \quad (11.12)$$

На рисунке 11.7 приведён пример использования фильтра. Выровненные уровни отражают наблюдавшуюся тенденцию во временном ряду, а отклонения обрабатываются, как случайная составляющая обычной полиномиальной модели.

Таким образом, на примере динамики производства молока показано, как современные методы сглаживания при меньших трудозатратах и объёмах предпосылок относительно характера исходной информации способны давать пригодные результаты, превосходящие результаты традиционных методов. На графиках видно, что локально взвешенная регрессия (рисунок 11.6), более адекватно описывает динамику, чем полином (рисунок 11.5). При использовании локально взвешенной регрессии не теряются уровни на концах временного ряда, как это происходит при использовании скользящей средней. Фильтр Ходрика-Прескотта (рисунок 11.7) и локально взвешенная регрессия (рисунок 11.6) дали схожие результаты. Преимуществами фильтра являются меньшие трудозатраты на подбор необходимых параметров (значения уже предложены авторами фильтра для временных рядов различных периодичностей) и быстрота получения результата.

Намеренно не указаны в данном перечне способ экспоненциального сглаживания и модели авторегрессии. Они дают возможность получить предварительное суждение о тренде и периодической составляющей, а также смоделировать их, но представляют собой более высокий (по сравнению со сглаживанием) уровень анализа в связи со своими возможностями и особенностями применения.

В качестве следующего этапа анализа рассмотрим коррелограмму временного ряда (рисунок 11.8). Из представленного рисунка следует важный вывод: отсутствие выраженной сезонности производства молочной продукции (коэффициент автокорреляции на лаге 12 месяцев составляет 0,157, а коэффициент частной автокорреляции равен –0,001).

Автокорреляционные коэффициенты отражают тесноту линейной взаимосвязи между последовательными значениями временного ряда. Визуальный поиск возможных нелинейных зависимостей осуществим при помощи матрицы полигонов распределения текущих и лаговых значений. На рисунке 11.9 представлены зависимости лаговых значений производства цельномолочной продукции (ось ординат) от текущих значений (ось абсцисс). По данному рисунку прослеживается увеличение рассеяния значений с ростом лага и снижением коэффициентов автокорреляции r . Нелинейные зависимости в данном случае не наблюдаются, что упрощает спецификацию модели одномерного временного ряда.

Характер затухания коррелограммы (рисунок 11.8) свидетельствует о том, что временной ряд является реализацией авторегрессионного стационарного процесса. Следуя изложенным ранее положениям, проведём тестирование временного ряда на наличие единичного корня с помощью ADF-теста. Так как изначальной уверенности в существовании тренда нет, а значения ряда имеют ненулевое среднее, то в тестовое уравнение в качестве регрессора включён только свободный член.

Рассчитанные значения тестовой статистики (таблица 11.1) подтверждают отсутствие единичного корня на уровне значимости 5,0%: нулевая гипотеза о наличии единичного корня отвергается в пользу альтернативной, так как статистика ADF-теста меньше (лежит левее) соответствующего критического значения.

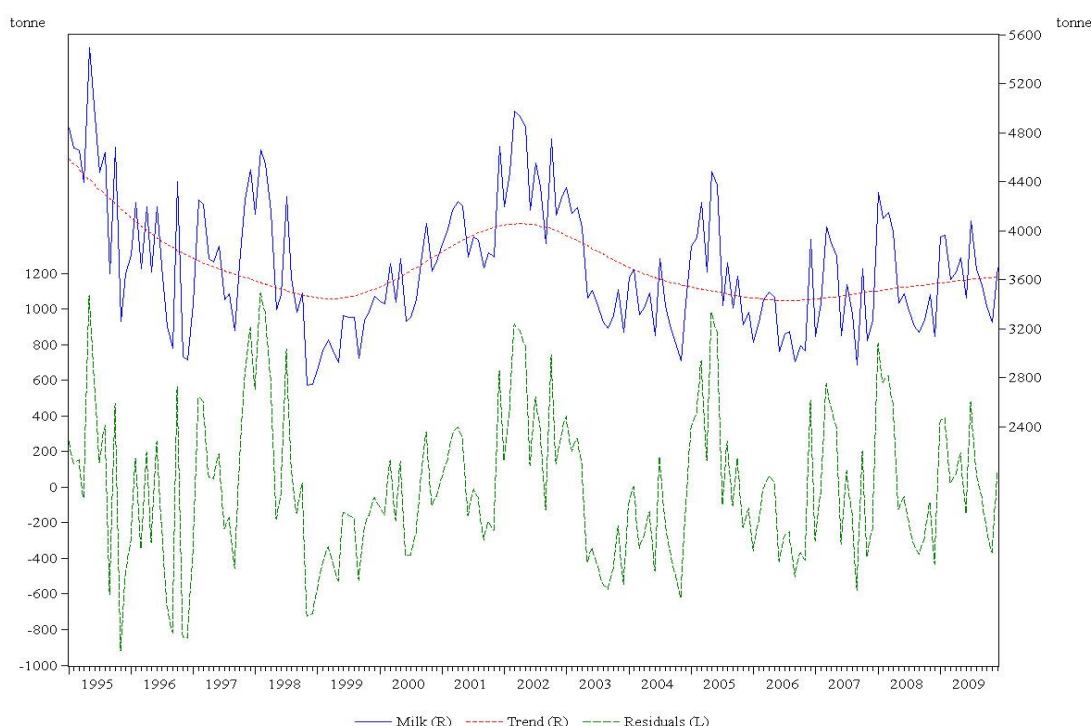


Рисунок 11.7 — Сглаживание динамики производства цельномолочной продукции в Оренбургской области (в пересчёте на молоко) с помощью фильтра Ходрика-Прескотта, $\lambda = 14400$

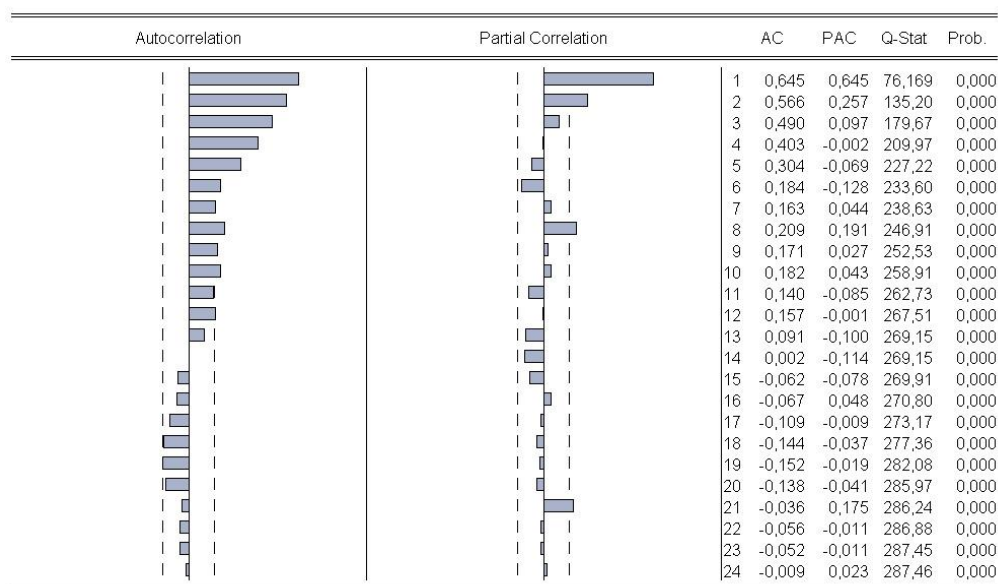


Рисунок 11.8 — Коррелограмма временного ряда производства цельномолочной продукции в Оренбургской области (в пересчёте на молоко) с января 1995 г. по декабрь 2009 г.

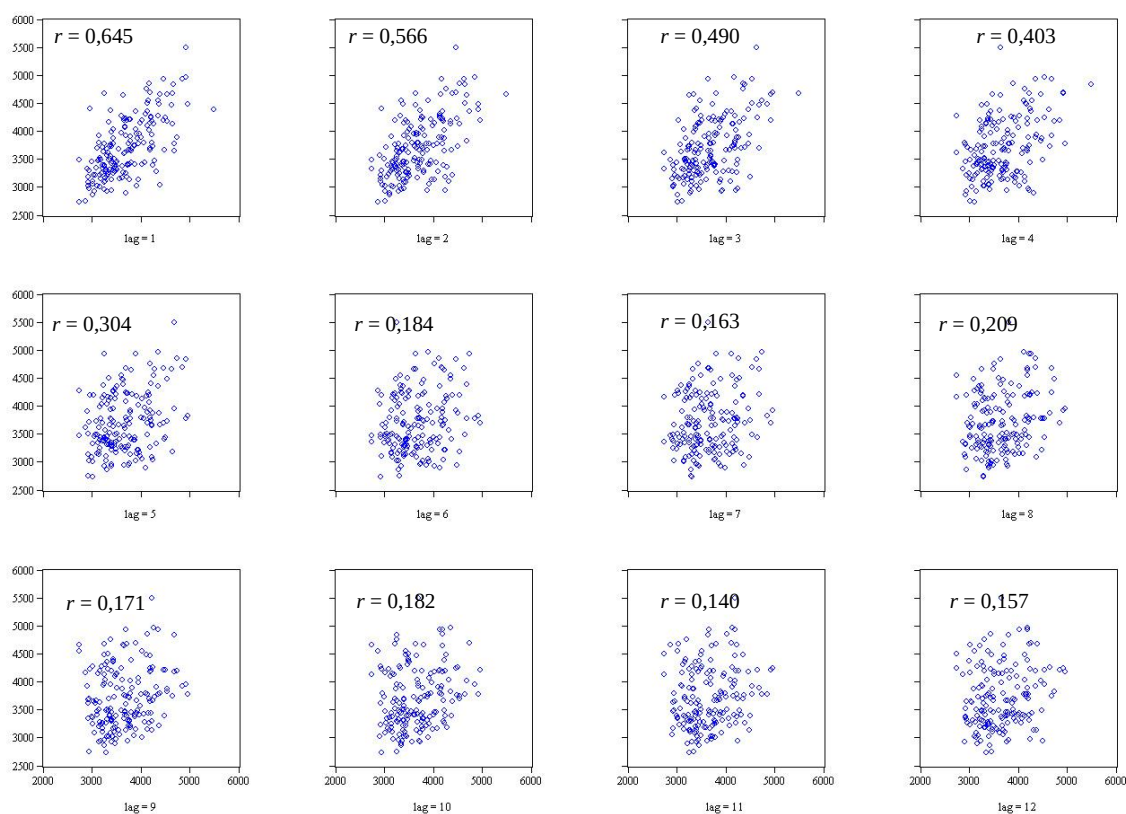


Рисунок 11.9 — Матрица лаговых зависимостей динамики производства цельномолочной продукции в Оренбургской области с января 1995 г. по декабрь 2009 г.

Таблица 11.1 — ADF-тест уровней временного ряда производства молока в Оренбургской области с января 1995 г. по декабрь 2009 г.

	<i>t</i> -статистика	Вероятность
--	----------------------	-------------

Статистика ADF-теста		-4,477704	0,0003
Критические значения:	1,0%	-3,467205	
	5,0%	-2,877636	
	10,0%	-2,575430	

Результаты проведённого предварительного анализа временного ряда производства молочной продукции делают необоснованным использование традиционных трендовых моделей, поэтому рассмотрим применение моделей для стационарных временных рядов. В соответствии с подходом Бокса-Дженкинса, при идентификации модели $ARIMA(p,d,q)$ сначала определяется d , а затем p и q . Для рассматриваемого временного ряда $d=0$, что следует из результатов ADF-теста на единичный корень и характера затухания выборочной автокорреляционной функции. Далее определим, после какого лага начинается убывание по абсолютной величине выборочной автокорреляционной и частной автокорреляционной функций. Принимая во внимание, что автокорреляционная функция МА-процесса равна нулю при лаге, большем q ; а частная автокорреляционная функция AR-процесса обрывается после лага p , делаемся предположения о значениях q и p . Из рисунка 11.8 следует, что $p=2$, $q=0$, то есть мы имеем дело с реализацией процесса $AR(2)$:

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \varepsilon_t. \quad (11.13)$$

Результаты оценки коэффициентов представлены в таблице 11.2.

Таблица 11.2 — Результаты оценки авторегрессионной модели динамики производства цельномолочной продукции в Оренбургской области с января 1995 г. по декабрь 2009 г.

Параметр	Коэффициент	Стандартная ошибка	t-статистика	Вероятность
c	3668,768	107,37	34,169	0,0000
AR(1)	0,460	0,073	6,326	0,0000
AR(2)	0,271	0,072	3,761	0,0002
R^2	0,461	Средняя величина зависимой переменной		3692,556
Нормированный R^2	0,454	Стандартное отклонение зависимой переменной		520,599
Стандартная ошибка регрессии	384,506	Критерий Акайки		14,759
Сумма квадратов отклонений	25872804	Критерий Шварца		14,812
Функция правдоподобия	-1310,507	Критерий Ханнана-Куина		14,780

F-статистика	74,735	Статистика Дарбина-Уотсона	2,061
Вероятность (F-статистика)	0,0000		
Обратные AR-корни	0,80	-0,34	

Так как расчёты проводились в Eviews, где модель авторегрессии второго порядка имеет вид:

$$x_t = c + u_t, \quad (11.14)$$

$$\text{и } u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} = \rho_1 (x_{t-1} - c) + \rho_2 (x_{t-2} - c),$$

$$\text{то } x_t = c(1 - \rho_1 - \rho_2) + \rho_1 x_{t-1} + \rho_2 x_{t-2}.$$

(11.15)

Подставляя значения из таблицы 11.2, получим модель, статистически значимую с вероятностью 95,0%:

$$x_t = 987,36 + 0,46x_{t-1} + 0,27x_{t-2}. \quad (11.16)$$

Важным этапом анализа является определение устойчивости полученной модели во времени, для чего строится динамический ряд её коэффициентов. В качестве примера проведён анализ на двухлетних скользящих периодах. Из рисунка 11.10 следует, что коэффициенты модели указанной спецификации претерпевали значительные изменения в изучаемом периоде, однако после 1999 г. зависимости вида AR(2) стали укрепляться, в связи с чем колебания значительно сократились, а скользящие значения стали близки к коэффициентам 0,46 и 0,27, полученным для всей 15-тилетней выборки. Таким образом, коэффициенты модели (11.16) признаны устойчивыми во времени, а сама модель пригодна для использования в дальнейшем анализе.

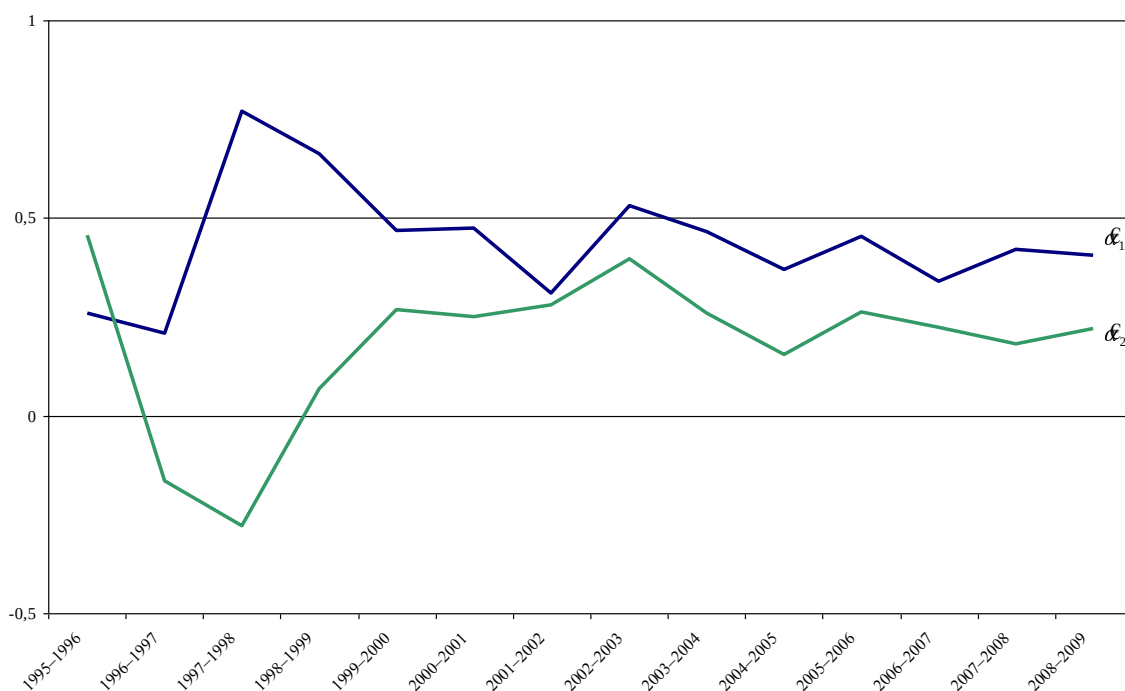


Рисунок 11.10 — Скользящие коэффициенты модели AR(2) динамики производства молока в Оренбургской области

Отметим, что изучение реакции на импульс, ступенчатое изменение или линейную функцию наглядно показывает адаптивные свойства полученной модели. Рассмотрим график реакции модели (11.16) на единичный импульс (рисунок 11.11).

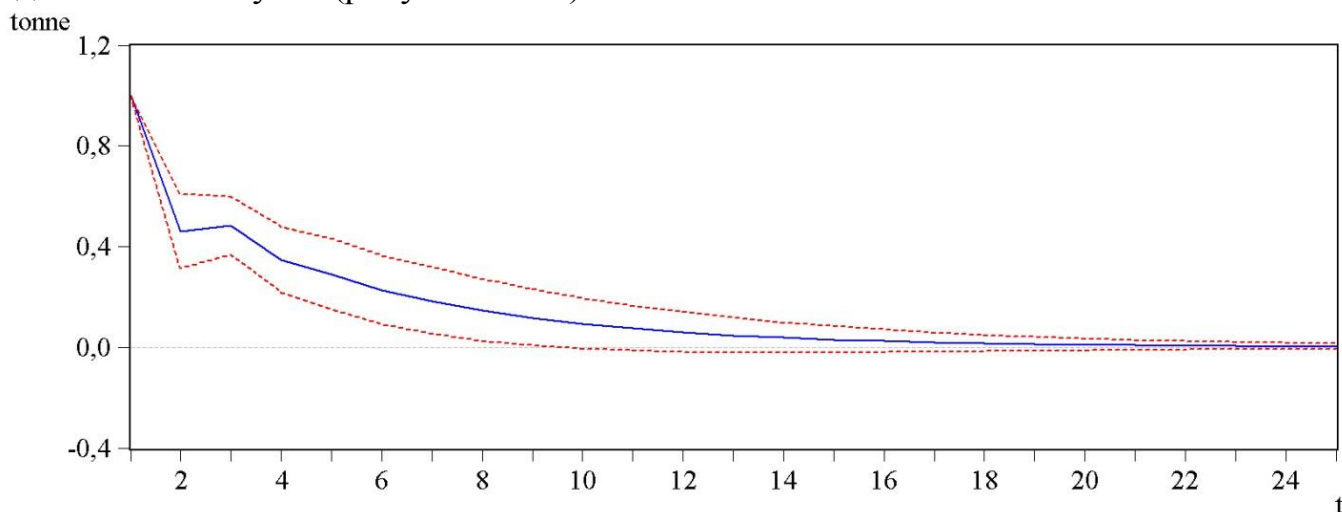


Рисунок 11.11 — Реакция модели (11.16) на единичный импульс (с доверительными границами ± 2 среднеквадратических отклонения)

Во-первых, функция затухает немонотонно, так как авторегрессионный член второго порядка получает информацию об импульсе с запаздыванием. В результате этого наблюдается рост при $t=3$. Во-вторых, значения медленно стремятся к нулю, и даже через год ($t=12$) влияние импульса остаётся заметным. Для более простой модели AR(1) функция убывает монотонно и быстрее, чем для AR(2). Выбор подходящей теоретической

модели определяется знаниями предметной области, зависимостями между последовательными значениями показателя. С другой стороны, реакция на импульс верно подобранной модели содействует расширению знаний об изучаемом процессе, раскрывает существующие взаимосвязи.

При оценке параметров эконометрической модели делаются допущения о характере случайных отклонений. Корректность этих допущений выявляется на этапе проверки (верификации) модели [18, гл. 5]. Отклонения исследуют на случайность, нормальность распределения, стабильность дисперсии, несмещённость и отсутствие автокорреляции.

Протестируем остатки полученной модели (рисунок 11.12) на серийную корреляцию (автокорреляцию высоких порядков). Для данного тестирования широко распространённый критерий Дарбина-Уотсона не подходит, так как он определяет значимость автокорреляции только первого порядка. Поэтому используется Q-статистика и тест множителей Лагранжа на наличие автокорреляции.

Q-статистика автоматически рассчитывается многими статистическими программами при построении коррелограммы (рисунок 11.8). Она имеет распределение χ^2 с числом степеней свободы, равным величине лага. Если Q-статистика меньше $\chi^2_{табл.}$, то нулевая гипотеза об отсутствии автокорреляции не отвергается. В противном случае принимается альтернативная гипотеза о наличии автокорреляции.

Тест множителей Лагранжа, альтернативный Q-тесту, проверяет наличие автокорреляции в остатках модели ARMA(p,q) вплоть до лага

$$\tau = \max(p, q). \quad (11.17)$$

Данный тест проводится как при обычным методе наименьших квадратов (МНК), так и при двухшаговом методе наименьших квадратов (ДМНК). Для остатков модели (11.16) результаты теста до лага 12 месяцев свидетельствуют об отсутствии автокорреляции на данных лагах с вероятностью 95,0% (таблица 11.3).

Указанные выше методы доказали свою эффективность при анализе макроэкономических данных с низкой частотой наблюдений (с периодом в год, квартал или месяц) и значительным агрегированием (во времени, и по объектам). Подобные временные ряды являются сравнительно сглаженными и простыми для анализа.

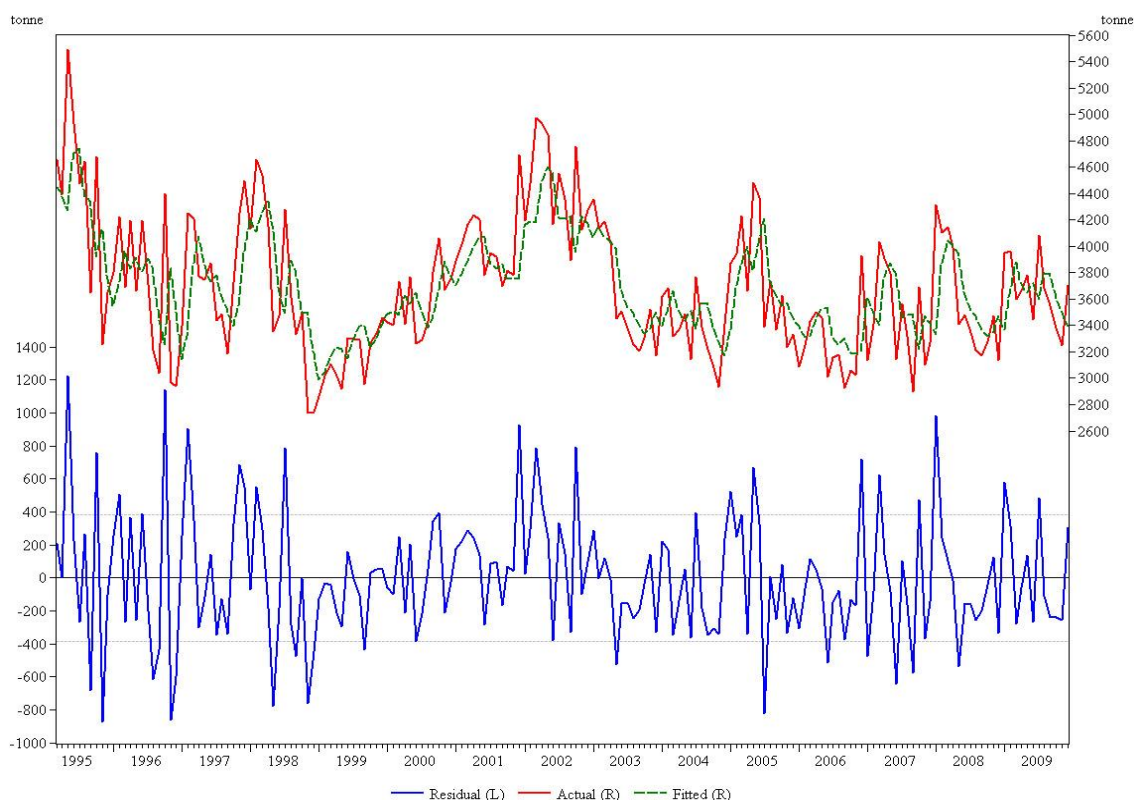


Рисунок 11.12 — Динамика фактического производства молочной продукции в Оренбургской области, её теоретические значения, рассчитанные по модели (11.16), и график остатков

Таблица 11.3 — Результаты теста множителей Лагранжа на наличие автокорреляции в остатках модели (11.16)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-32,22586	107,8176	-0,298892	0,7654
AR(1)	1,757934	5,637103	0,311851	0,7556
AR(2)	-1,181613	4,594044	-0,257205	0,7973
RESID(-1)	-1,805472	5,632250	-0,320560	0,7490
RESID(-2)	0,289850	2,009152	0,144265	0,8855
RESID(-3)	-0,208638	0,615827	-0,338793	0,7352
RESID(-4)	0,001981	0,290435	0,006820	0,9946
RESID(-5)	-0,077175	0,112692	-0,684832	0,4944
RESID(-6)	-0,211152	0,109156	-1,934402	0,0548
RESID(-7)	-0,180414	0,089764	-2,009882	0,0609
RESID(-8)	0,056323	0,088233	0,638349	0,5241
RESID(-9)	-0,025931	0,084235	-0,307839	0,7586
RESID(-10)	0,134090	0,082243	1,630417	0,1049
RESID(-11)	-0,021943	0,081079	-0,270640	0,7870
RESID(-12)	0,092846	0,081080	1,145110	0,2538

Однако современный экономический анализ, особенно финансовая эконометрика, имеет дело с более дезагрегированными данными микроуровня, наблюдаемыми со значительно большей частотой. Примером

служат данные с фондовых и валютных рынков, которые доступны не только с ежедневной периодичностью, но и практически непрерывно. Спектральный анализ и частотное представление временного ряда в данном случае имеют ряд значительных преимуществ [39, с. 624].

Использование спектральных характеристик актуально не только потому, что они раскрывают информацию, незаметную при временном представлении сигнала, но и потому, что во многих экономических временных рядах объективно присутствует периодическая составляющая. Например, результатом преобразования Фурье является спектральное представление, при котором на оси абсцисс графика обозначаются частоты, а ось ординат отражает амплитуды той или иной частоты в сигнале. Альтернативой преобразования Фурье служит преобразование Гильберта, оконное преобразование Фурье, распределение Вигнера, преобразование Уолша, вейвлет-преобразование [10, с. 6].

Рассмотрим общий случай, когда частота колебаний заранее неизвестна и переменна во времени. В оконном преобразовании Фурье размер окна выбирается единожды: узкое окно обеспечивает лучшее временное представление, а широкое — частотное. Поэтому критичным является вопрос о размерах окна. Вейвлет-преобразование решает эту проблему разрешения с помощью кратномасштабного анализа: временной ряд анализируется на различных частотах и различном разрешении одновременно, что особенно эффективно позволяет выделить и локализовать на временной оси высокочастотные компоненты короткой длительности с протяжёнными низкочастотными компонентами.

Результаты вейвлет-анализа динамики производства молока в Оренбургской области представлены на рисунке 11.13.

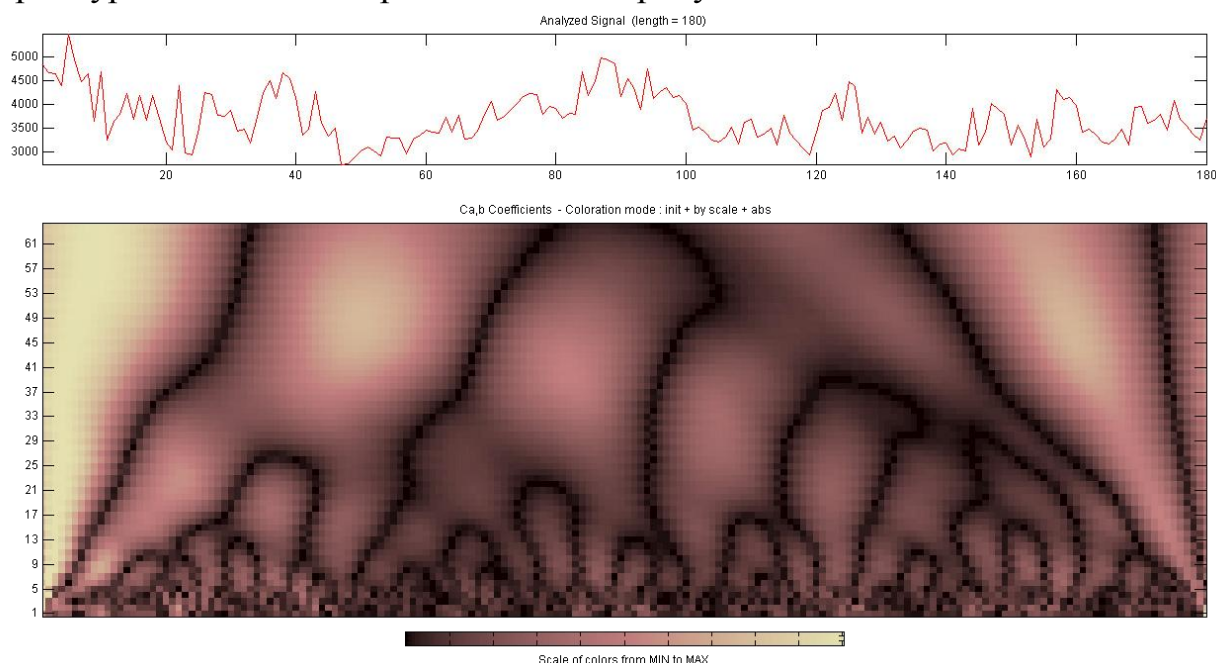


Рисунок 11.13 — Результаты вейвлет-анализа динамики производства молока в Оренбургской области в 1995–2009 гг.

Помимо временного графика, на рисунке 11.13 с помощью различных оттенков показаны коэффициенты совпадения формы выбранного вейвлета Добеши с фактической динамикой процесса на различных уровнях частот и временных отрезках одновременно. Колебания с периодичностью 12 месяцев встречаются на интервалах с 12-го по 36-й; с 84-го по 96-й; со 132-го по 144-й и со 166-го по 172-й месяцы. На промежутке с 60-го по 84-й месяцы период колебаний более длительный и достигает двух лет (при этом однолетние периодичности отсутствуют). Наибольший период для расчёта коэффициентов (64 месяца) определён автоматически, исходя из длины исходного ряда (180 уровней). Однако коэффициенты для периодов высоких порядков не интерпретируются, так как мала уверенность в их статистической значимости.

Методология анализа постоянно совершенствуется. Рассмотренные методы сглаживания, аспекты анализа лаговых зависимостей, исследования свойств модели, а также методы частотно-временного анализа служат средством построения новых эконометрических моделей и прогнозов.

11.3 Применение двуходового объединения и теории коинтеграции в анализе взаимосвязи временных рядов

Не подлежит сомнению тот факт, что развитие социально-экономических процессов происходит во взаимосвязи друг с другом. Среди способов, позволяющих изучить существующие связи и закономерности, выделяют графический анализ, индексы, корреляционно-регрессионный анализ, факторный анализ. Развитие методологии идёт как по пути совершенствования свойств простых методов, так и по пути раскрытия возможностей и продвижения более сложных и формализованных подходов.

Одним из специфических методов является выявление синхронности колебаний временных рядов. В работе В. Н. Афанасьева [2] приводятся результаты изучения синхронности колебаний производства сельскохозяйственной продукции в различных странах. В ходе данного анализа определяются отклонения от трендов динамики производства в каждой рассматриваемой стране, а затем находятся парные коэффициенты корреляции. Чем ближе найденный коэффициент к 1, тем бóльшая синхронность наблюдается в колебаниях; чем ближе его значение к -1 , тем сильнее асинхронность колебаний; при значениях, близких к нулю, связь отсутствует.

Специфика данного метода заключается в том, что в большинстве случаев, с точки зрения качественного анализа проблемы, взаимосвязь показателей отсутствует. В силу этого, при обнаружении значимых коэффициентов корреляции регрессионная модель не строится. Однако синхронность определяет не взаимосвязь, а механическое совпадение колебаний: их силу и направленность. Изучение синхронности колебаний производства в административных районах (субъектах федерации, различных

государствах) имеет большое социально-экономическое значение, так как способствует рациональному, выгодному распределению ресурсов между участниками, а также выравниванию обеспеченности населения необходимыми товарами.

С целью дальнейшего развития методологии изучения синхронности колебаний временных рядов на базе монографического исследования В. Н. Афанасьева [2] предложен и апробирован метод двуходового объединения для выделения из исследуемой совокупности групп объектов с синхронными и асинхронными колебаниями.

Двухходовое объединение — один из методов кластерного анализа, учитывающий при формировании групп одинаковую важность как наблюдений, так и переменных. Его применение к матрице парных коэффициентов корреляции позволяет объединить объекты, используя информацию и по строкам, и по столбцам. В матрице представлены *парные* коэффициенты корреляции, поэтому и кластеризация проводится сразу по двум направлениям.

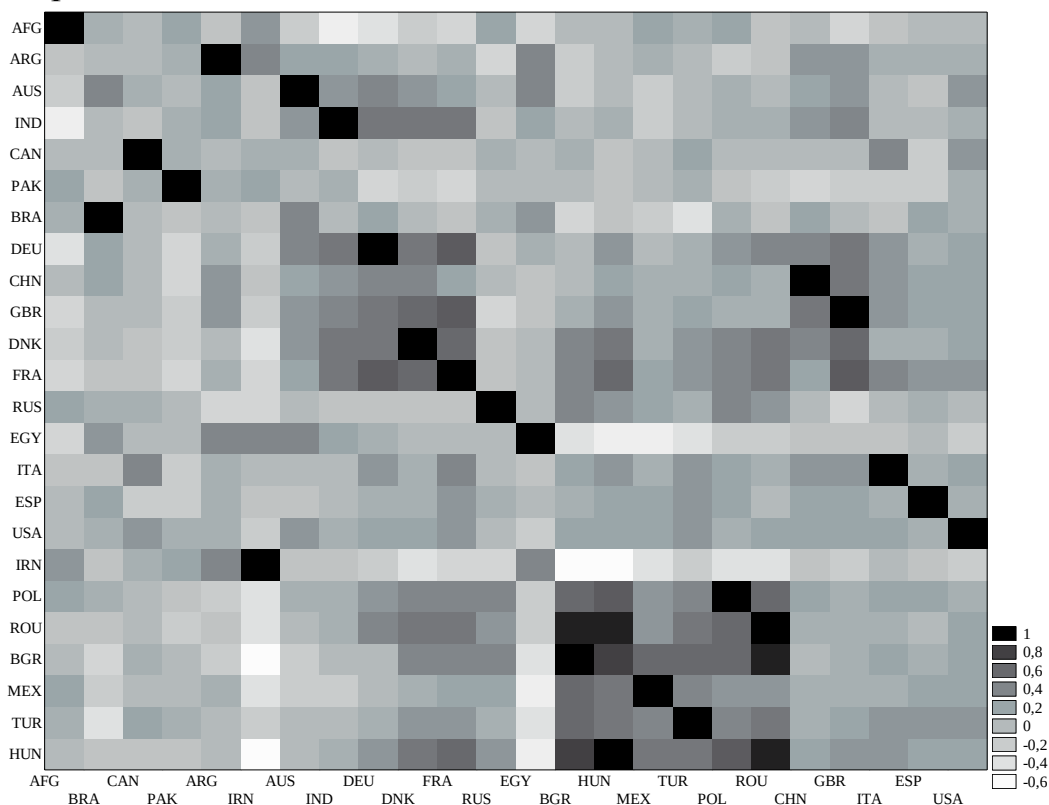


Рисунок 11.14 — Результаты двуходового объединения коэффициентов парной корреляции колебаний урожайности пшеницы

На рисунке 1.14 представлены результаты анализа динамики урожайности пшеницы в 1961–2009 гг. в 24 странах, являющихся мировыми лидерами по её производству⁵. Исходные временные ряды получены с

⁵В анализе не участвовали, входящие в первую двадцатку, Украина, Казахстан и Узбекистан, так как отсутствуют данные об урожайности пшеницы в этих странах за период с 1961 по 1991 гг.

официального сайта Продовольственной и сельскохозяйственной организации Объединённых Наций (ФАО) [48]; обозначения стран даны в соответствии с кодами ISO 3166.

Первый кластер сформировали шесть стран (Болгария, Венгрия, Мексика, Польша, Румыния, Турция) с синхронными колебаниями урожайности. Урожайность пшеницы в России имеет с ними положительную, но слабую синхронность колебаний в отличие от Ирана и Египта, урожайность пшеницы в которых колеблется в противофазе со странами первого кластера.

Ядро второго кластера составляют Дания и Франция, к которым примыкают Великобритания и Германия. Со странами этой европейской группы колебания российской урожайности асинхронны, что потенциально способствует развитию выгодного сотрудничества в торговле зерном.

Кроме того, Великобритания и Германия входят в следующий кластер, третий участник которого — Китай. Колебания урожайности пшеницы в России и в Китае также асинхронны.

Особое место занимают Австралия и Индия, имеющие слабую (около 0,4) синхронность колебаний урожайности с Европой (Германией, Данией, Францией). Такие государства, как Аргентина, Бразилия, Канада и США не имеют однозначно определённого места в каком-либо из кластеров, поэтому привлекательность сделок с ними должна оцениваться на основе других критериев.

Цена и возможность наращивания экспорта во многом зависят от объёма производства зерна, который в краткосрочном периоде определяется урожайностью. Асинхронность колебаний урожайности (производства) между странами позволяет взаимовыгодно экспортировать зерно из стран с относительно высоким урожаем в страны с относительно низким в текущем году урожаем. Перераспределение продовольственных запасов в соответствии с данным принципом способствует снижению социальной напряжённости, рассматривается как инструмент диверсификации, обеспечения продовольственной безопасности и борьбы с голодом.

Итак, изучение синхронности колебаний основано на сопоставлении силы и направленности колебаний во временных рядах, качественные зависимости между которыми могут и не прослеживаться. Если же такие зависимости существуют, то применяются методы для построения эконометрических моделей.

С целью исключения ложной корреляции временных рядов, возникающей при наличии тенденции в каждом временном ряду, применяют методы исключения тенденции:

- 1) преобразование уровней исходного временного ряда в новые переменные, не содержащие тенденции (метод последовательных разностей и метод отклонений от трендов);

- 2) элиминирование воздействия фактора времени на зависимую и независимую переменные модели (включение в модель регрессии фактора времени).

Общим недостатком указанных методов является модификация модели экономических взаимосвязей вследствие либо замены переменных, либо добавления в эту модель фактора времени. Большая часть соотношений, выведенных из экономической теории, сформирована на основе уровней временных рядов, а не их последовательных разностей или отклонений от трендов. Поэтому видоизменение моделей является нежелательным.

В ряде случаев наличие тенденции во временном ряду является следствием именно того, что другой ряд, включённый в модель, тоже содержит тенденцию. Поэтому одинаковая или противоположная направленность тенденций рядов может иметь устойчивый характер и наблюдаться на протяжении длительного промежутка времени. Коэффициент корреляции, рассчитанный по уровням таких временных рядов, соответственно, не содержит ложной корреляции и характеризует истинную причинно-следственную зависимость между ними [23, с. 447].

В 80-х гг. XX века К. Грэнджер и Р. Ингл изложили результаты разработки нового подхода к анализу взаимосвязанных временных рядов и ввели термин «коинтеграция». Коинтеграция представляет собой причинно-следственную зависимость в уровнях двух (или более) временных рядов, которая выражается в совпадении или противоположной направленности их тенденций и случайной колеблемости. За вклад в развитие методологии коинтеграции К. Грэнджер в 2003 г. был удостоен половины премии Шведского государственного банка по экономическим наукам памяти Альфреда Нобеля.

Теория коинтеграции активно развивается. Наличие коинтеграционных отношений между интегрированными (часто называемыми нестационарными⁶) экономическими переменными оценивается перед проведением запланированного анализа, такого как расчёт интересующих параметров или проверка гипотезы. Если данные коинтегрированы, то применяются модели корректировки ошибок (errorcorrectionmodels; ECMs), в противном случае — модели векторной авторегрессии (VAR) на основе последовательных разностей временных рядов.

В отечественной научной литературе представлены краткие основные сведения о коинтеграции — в работах Г. Г. Канторовича [13], Ю. П. Лукашина [17], И. И. Елисеевой и других [23]. Из англоязычных публикаций К. Грэнджера и Н. Свансона [37], Р. Дэвидсона и Д. Мак-Кинона [30], В. Грина [39] можно получить массу дополнительной информации, в частности, из работы Н. Бэлка и Т. Фомби [28] — о пороговой коинтеграции; В. Эндерса и К. Грэнджера [32], В. Эндерса и П. Сиклоса [33] — о нелинейном механизме корректировки с асимметричной корректировкой ошибок.

⁶Процесс $I(1)$ обычно называют нестационарным. Однако, если временной ряд $I(1)$ может быть нестационарным, не все нестационарные ряды являются интегрированными $I(1)$. Аналогично, стационарный временной ряд можно обозначить как интегрированный ряд нулевого порядка $I(0)$, но не все ряды $I(0)$ являются стационарными.

Не смотря на то, что многие экономические временные ряды признаются учёными коинтегрированными, необходимо подчеркнуть, что коинтеграция — это очень специфический феномен. Коинтеграция проявляется по причине наличия у экономических данных общих стохастических трендов, которые элиминируются коинтеграционными линейными комбинациями. Общие стохастические тренды обычно выражаются как линейная комбинация скачков системы. Экономические данные коинтегрированы, так как они отвечают на скачки совместно, поэтому коинтеграции не прослеживается, если временные ряды отвечают на импульсы отдельно.

Однако в отдельных случаях временные ряды имеют схожее поведение только при импульсах некоторого конкретного типа. Следовательно, если временной ряд устойчив к снижению, то его реакция на негативные изменения не будет симметрична реакции на положительные скачки. Например, розничные цены на бензин в Российской Федерации всегда растут при повышении цены на нефть, в то время как снижение последней не обеспечивает падения цен на бензин. В компонентах таких некоинтегрированных временных рядов может быть заключена важная информация, необходимая для понимания их динамических взаимосвязей. Указанный аспект упускается обычной теорией коинтеграции, но учитывается при изучении коинтеграции компонентов данных, то есть скрытой коинтеграции. С помощью данного подхода становится доступной оценка долговременных взаимосвязей между некоинтегрированными нестационарными временными рядами. Таким образом, стандартная коинтеграция является частным случаем скрытой коинтеграции, в то время как скрытая коинтеграция является простым примером нелинейной коинтеграции (рисунок 11.15).



Рисунок 11.15 — «Иерархия» коинтеграции

Автор метода К. Грэнджер рассматривает возникновение скрытой коинтеграции на примере двух случайных блужданий без смещений:

$$X_t = X_{t-1} + \varepsilon_t = X_0 + \sum_{i=1}^t \varepsilon_i \quad (11.18)$$

и

$$Y_t = Y_{t-1} + \eta_t = Y_0 + \sum_{i=1}^t \eta_i, \quad (11.19)$$

где X_0 и Y_0 — начальные значения;

ε_i и η_i — белый шум с нулевым средним.

На данном этапе не обсуждается, коинтегрированы X_t и Y_t , или нет. Новые переменные позволяют разложить возмущения на бóльшие некоторого уровня и меньшие этого уровня:

$$\varepsilon_i^\vee = \max(\varepsilon_i, d) \text{ и } \varepsilon_i^\wedge = \min(\varepsilon_i, d). \quad (11.20)$$

Отметим, что $\varepsilon_i = \varepsilon_i^\vee + \varepsilon_i^\wedge - d$; d называют порогом. Популярный выбор порога таков, что положительные значения $\varepsilon_i^+ = \max(\varepsilon_i, 0)$, а отрицательные $\varepsilon_i^- = \min(\varepsilon_i, 0)$. Порог выбирается таким образом, чтобы были исключены ситуации, когда при всех i $\varepsilon_i = \varepsilon_i^\wedge$ или $\varepsilon_i = 0$. Теперь предположим, что $\sum_{i=1}^t \varepsilon_i^\wedge$, $\sum_{i=1}^t \eta_i^\wedge$, $\sum_{i=1}^t \varepsilon_i^\vee$ и $\sum_{i=1}^t \eta_i^\vee$ — величины $I(1)$. Может быть и больше компонентов X_t и Y_t , но для простоты рассматривается лишь два. Мы имеем:

$$X_t = X_{t-1} + \varepsilon_t = X_0 + \sum_{i=1}^t \varepsilon_i^\wedge + \sum_{i=1}^t \varepsilon_i^\vee - dt \quad (11.21)$$

и

$$Y_t = Y_{t-1} + \eta_t = Y_0 + \sum_{i=1}^t \eta_i^\wedge + \sum_{i=1}^t \eta_i^\vee - dt. \quad (11.22)$$

Скрытая коинтеграция наблюдается во временных рядах при условии, что их компоненты коинтегрированы друг с другом. Только при выполнении специальных условий скрытая коинтеграция между нестационарными компонентами X_t и Y_t подразумевает наличие стандартных коинтеграционных связей для самих временных рядов.

Для простоты теоретических выкладок предполагается, что ни $\left\{ \sum_{i=1}^t \varepsilon_i^+ \quad \sum_{i=1}^t \eta_i^- \right\}$, ни $\left\{ \sum_{i=1}^t \varepsilon_i^- \quad \sum_{i=1}^t \eta_i^+ \right\}$ не являются коинтегрированными (на практике данное положение подлежит проверке). Рассмотрим оставшиеся случаи взаимосвязи компонентов временных рядов.

Случай 1. Ни $\left\{ \sum_{i=1}^t \varepsilon_i^+ \quad \sum_{i=1}^t \eta_i^+ \right\}$, ни $\left\{ \sum_{i=1}^t \varepsilon_i^- \quad \sum_{i=1}^t \eta_i^- \right\}$ не являются коинтегрированными. Из этого следует, что X_t и Y_t также некоинтегрированы, то есть имеют различные стохастические тренды.

Случай 2. Или $\left\{ \sum_{i=1}^t \varepsilon_i^+ \quad \sum_{i=1}^t \eta_i^+ \right\}$, или $\left\{ \sum_{i=1}^t \varepsilon_i^- \quad \sum_{i=1}^t \eta_i^- \right\}$ коинтегрированы, но не одновременно. Следовательно, X_t и Y_t имеют общие либо позитивные, либо негативные колебания. X_t и Y_t в данном случае всё ещё не являются коинтегрированными, но из структуры их взаимосвязи можно извлечь больше информации, чем в первом случае. Эти сведения о скрытых коинтеграционных связях не используются, если исследователь заинтересован лишь коинтеграцией между X_t и Y_t .

Случай 3. И $\left\{ \sum_{i=1}^t \varepsilon_i^+ \quad \sum_{i=1}^t \eta_i^+ \right\}$, и $\left\{ \sum_{i=1}^t \varepsilon_i^- \quad \sum_{i=1}^t \eta_i^- \right\}$ коинтегрированы, но с различными коинтеграционными векторами. Временные ряды X_t и Y_t всё ещё остаются некоинтегрированными. Они хотя и имеют общие положительные и отрицательные возмущения, но общие возмущения некоинтегрированы.

Случай 4. И положительные $\left\{ \sum_{i=1}^t \varepsilon_i^+ \quad \sum_{i=1}^t \eta_i^+ \right\}$, и отрицательные возмущения $\left\{ \sum_{i=1}^t \varepsilon_i^- \quad \sum_{i=1}^t \eta_i^- \right\}$ коинтегрированы с одинаковыми коинтеграционными векторами, то есть присутствует один общий импульс. Это интерпретируется как наличие общего стохастического тренда для X_t и Y_t , отвечающего за их долговременное поведение в динамике. В данном случае ряды X_t и Y_t являются коинтегрированными.

Скрытая коинтеграция является простым примером нелинейной коинтеграции. Например, две $I(1)$ переменные X_t и Y_t линейно коинтегрированы, если существует такое α , что $\varepsilon_t - \alpha X_t \sim I(0)$. Следовательно, они нелинейно коинтегрированы, если существует такое β , что $f(\varepsilon_t) - \beta g(\varepsilon_t) \sim I(0)$ для некоторых нелинейных функций f и g .

Указанные случаи показывают, насколько специфичным явлением является коинтеграция. Стандартная коинтеграция между временными рядами наблюдалась лишь в четвёртом случае, однако в первых трёх случаях степень её отсутствия была различной. Если уделять внимание лишь четвёртому случаю, то множество ценной информации теряется. Даже если временные ряды не являются коинтегрированными, их скрытая структура может быть плодотворно использована для изучения динамики и получения усовершенствованных прогнозов.

Тем не менее, существуют причины, по которым не удаётся найти коинтеграционные взаимосвязи между экономическими временными рядами, даже если их наличие ожидается на основании экономической теории. Среди таких причин — используемая процедура тестирования, недостаточная длина

временного ряда, выбор длины лага при определении тестовой статистики, наличие структурных сдвигов.

Многие экономические теории связаны с равновесными отношениями, при которых один или более показателей являются аттракторами для других. В частности, коинтеграционные тесты активно применяются в финансовой литературе, так как в отсутствие принудительного регулирования рынка устремляются к равновесным отношениям, хорошо известным из теории. В качестве примера рассматривают тестирование на наличие коинтеграции номинальных процентных ставок [27], реальных процентных ставок [43], [35] и биржевых цен [42].

По мнению П. Сиклоса и К. Грэнджера [46], в некоторых случаях более реалистичным будет предположение, что временные ряды коинтегрированы лишь на протяжении некоторых периодов и некоинтегрированы в остальное время (временная коинтеграция). Причина такого поведения экономических временных рядов в том, что отдельные явления или важные изменения институциональных свойств экономики могут расстроить существующие равновесные взаимосвязи на некоторый (возможно, длительный) период времени.

Коинтеграционные взаимосвязи отражают сложные социально-экономические зависимости, приближают исследователя к пониманию причинно-следственных связей. Изучение причинности зависит от источников познания окружающего мира. С одной стороны, это наш собственный опыт. С другой стороны, это механистическое понимание: с помощью системы умозаключений можно предположить некоторые последствия или изобрести способы их предотвращения. Знающий руководитель предлагает и обосновывает свои действия, чтобы вывести предприятие из кризиса, хороший агроном получает высокий урожай — во многом благодаря знаниям механизмов.

Обычное, основанное на опыте понимание, не требует глубокого понимания механизмов. Фактически, при столь огромном периоде существования человеческих знаний, механистическое понимание возникло сравнительно недавно. Опыт севооборотов с незапамятных времён приносит свои плоды, но лишь недавно изучены вещества, содержащиеся в почве и влияющие на развитие сельскохозяйственных культур. Бездумная и безудержная эмиссия денег всегда оборачивалась крахом государственных финансов, но теоретические работы, посвящённые изучению причин инфляции, соотношению товарной и денежной масс появились гораздо позже.

На данный момент механистические знания в социальных науках очень ограничены или вовсе отсутствуют. Этому способствовало доминирующее на протяжении большей части XX века мнение учёных, что статистик должен сторониться причинности. Основателями статистики, особенно Р. Фишером и К. Пирсоном, было строго установлено, что статистика лишь ассоциативна [25, с. 157]. В соответствии с этим течением действовало правило «correlation

does not imply causation»⁷, антонимом которому служило латинское высказывание-заблуждение «Cum Нос Ergo Propter Нос»⁸. Осторожное отношение к причинности может показаться обоснованным, так как причинные выводы сложны и полны опасностей. Но многие области знаний, особенно экономика и общественные науки, ждут от статистики руководства. В последние годы причинность стала главным предметом спора среди статистиков, развиваются новые подходы по её изучению. Статистический анализ играет важнейшую роль в формировании связей с основанной на опыте причинностью. Экономические исследования — одна из областей, где статистика является обязательной, особенно благодаря необходимости учётных данных для анализа и управления.

В моделировании поведения экономических субъектов популярна причинность К.Грэнджера, зародившаяся в эконометрике и распространяющаяся в другие области знаний. Причинность К. Грэнджера сфокусирована на измерениях во времени и на том, как они влияют друг на друга, как настоящее и прошлое влияют на дальнейшее развитие. Данный метод стремится к пониманию того, как различные процессы взаимосвязаны. Это не обязательно означает, что механизмы могут быть разъяснены детально только благодаря статистическим данным, но анализ К. Грэнджера — шаг к расширенному пониманию зависимостей. Рост доступности детальных статистических данных открывает большие возможности для данного подхода. Привлекательная черта причинности К. Грэнджера — анализ взаимосвязи между прошлым, настоящим и будущим, который приводит к механистическому пониманию.

Причинность К. Грэнджера определена для двух временных рядов: X и Y . Отвечая на вопрос, является ли X причиной для Y , находят, насколько текущие значения временного ряда Y могут быть объяснены его лаговыми значениями, а затем, определяют, улучшается ли аппроксимация Y при добавлении лаговых значений временного ряда X . Считают, что X является Грэнджер-причиной для Y , если X улучшает прогноз Y , или, что равнозначно, если коэффициенты при лаговых значениях X являются статистически значимыми. Двусторонняя причинность также часто встречается: X является Грэнджер-причиной Y , а Y — Грэнджер-причиной X .

Важно отметить, что выражение « X является Грэнджер-причиной Y » не подразумевает, что Y является эффектом или результатом от действия X . Причинность К. Грэнджера измеряет предшествование и информационную составляющую, но не саму причинность в общем смысле этого слова.

С технической точки зрения, первым шагом в проведении теста причинности К. Грэнджера является определение длины лага в тестовой регрессии. Предпочтительным является выбор более длинных лагов, так как теория построена на предположении о важности всей информации прошлых периодов времени. Таким образом, длина лага (p) выбирается в соответствии

⁷ «корреляция не подразумевает причинность»

⁸ «вместе с этим значит по причине этого»

с мотивированными ожиданиями наиболее длительного времени, на протяжении которого одна из переменных может предсказывать значения другой. Затем строится двумерная регрессия вида

$$\begin{aligned} Y_t &= \alpha_0 + \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \beta_1 X_{t-1} + \dots + \beta_p X_{t-p} + \varepsilon_t; \\ X_t &= \alpha_0 + \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \beta_1 Y_{t-1} + \dots + \beta_p Y_{t-p} + u_t, \end{aligned} \quad (11.23)$$

и для каждого уравнения рассчитывается F -статистика (статистика Вольда) для объединённой гипотезы:

$$\beta_1 = \beta_2 = \dots = \beta_p = 0. \quad (11.24)$$

Нулевая гипотеза теста состоит в том, что X не является Грэнджер-причиной Y в первом уравнении, и что Y не является Грэнджер-причиной X во втором уравнении регрессии (1.23).

В работе А. Хатеми-Джей и Г. Шукура [40] представлены результаты анализа взаимосвязи между бюджетным дефицитом и дефицитом торгового баланса США. Анализ квартальных данных с 1-го квартала 1975 г. по 2-ой квартал 1998 г., скорректированных на сезонность, показал, что Грэнджер-причинность между указанными временными рядами не наблюдается. Однако после определения наличия структурного сдвига в выборке и разделения временных рядов на два однородных подпериода результаты оказались следующими: в первом подпериоде бюджетный дефицит являлся Грэнджер-причиной дефицита торгового баланса; во втором подпериоде зависимость изменилась на противоположную [40, с. 10]. Данный пример подчёркивает неоднозначность Грэнджер-причинности и важность периодизации временного ряда с целью выявления однородных участков динамики, использования предварительных тестов, указанных ранее.

Статистическая зависимость между двумя стохастическими процессами может быть двух основных типов. Первый тип — когда два процесса отражают (частично) одно и то же явление. Это не является причинной зависимостью. Экономическим примером служит повышение производительности труда и капиталоотдачи при повышении урожайности культур в сельскохозяйственном производстве.

Второй тип взаимосвязи — когда один процесс правдоподобно влияет на изменения в другом процессе причинным образом. Это может быть взаимосвязь с односторонним влиянием, или же двусторонняя взаимосвязь, в которой оба процесса влияют друг на друга. Например, уровень государственных расходов на социальные программы, рассматриваемый как стохастический процесс, причинно влияет на повышение благосостояния населения, главным образом, в односторонней форме. Реальность более сложна, так как повышение государственных расходов стимулирует инфляционные процессы, негативно отражающиеся на всей экономике.

Основанием осторожного отношения к причинности является то, что причинные выводы имеют абсолютное качество относительно глубоких взаимосвязей, и что статистические данные просто не могут подвергнуть сомнению настолько авторитетные утверждения. С другой стороны, причинное моделирование означает, что исследователь выявляет влияние различных процессов друг на друга, стремится познать сущность взаимодействия субъектов и на основе этого построить модели их поведения.

Новая роль статистики возникает в области моделирования экономических систем, где механистические модели анализируются и оцениваются математическими и статистическими методами. Рост доступности, количества и качества статистических данных, а также развитие математических методов создаёт качественно новую ситуацию, предвосхищающую широкое распространение причинных моделей.

Глава 12 Повышение качества прогнозирования временных рядов

Что необходимо знать из главы 2:

- 1. Порядок и особенности прогнозирования с помощью моделей Бокса-Дженкинса, ARFIMA и стейт-спейс моделей.*
- 2. Отличия и методы идентификации разностно-стационарных с тренд-стационарных процессов.*
- 3. Способы прогнозирования при наличии причинных зависимостей между показателями, достоинства и недостатки таких прогнозов.*
- 4. Критерии и показатели для сравнения альтернативных прогнозов.*

12.1 Потенциал моделей с длинной памятью и стейт-спейс моделей в прогнозировании одномерных временных рядов

Статистические методы решают одну из основных задач науки — прогнозирование возможных значений признаков изучаемого объекта. Наиболее распространённым методом прогнозирования экономических временных рядов является экстраполяция, то есть распространение прошлых и настоящих закономерностей, связей и соотношений на будущее.

Прогноз определяет ожидаемые варианты экономического развития, исходя из гипотезы, что основные факторы и тенденции прошлого периода сохраняются на период прогноза, или что можно обосновать и учесть направление их изменений в рассматриваемой перспективе. Подобная гипотеза выдвигается, исходя из инерционности экономических явлений и процессов. Инерционность в социально-экономических явлениях проявляется двояким образом. Во-первых, как инерционность взаимосвязей, то есть сохранение зависимости (корреляции) прогнозируемой переменной от совокупности факторных признаков. Во-вторых, как инерционность в развитии отдельных сторон явлений, то есть как некоторая степень сохранения их характеров — темпов, направления, колеблемости основных количественных показателей на протяжении сравнительно длительного времени [9, с. 175].

Статистический прогноз предполагает не только качественное предсказание, но и достаточно точное количественное измерение ожидаемых значений признака. Считается, что ошибка прогнозирования будет тем меньше, чем меньше срок упреждения и чем длиннее база прогноза — прошлый период, однородный по закономерностям развития [9, с. 176].

Прогнозирование временного ряда осуществляется без использования или с использованием его модели. Одним из наиболее известных методов прогнозирования является экспоненциальное сглаживание. Оно представляет собой основу целого семейства адаптивных моделей с различными свойствами, такими как наличие тренда, сезонности (аддитивной или мультипликативной), адаптацией параметров сглаживания. Среди работ, посвящённых применению данного метода, отметим [9], [11], [15], [17], [20],

[22]. В работе Г. С. Кильдишева и А. А. Френкеля [15] отмечается, что экспоненциальное сглаживание, во-первых, часто «не срабатывает» при изучении коротких эконометрических временных рядов (15–20 наблюдений). Во-вторых, для нахождения оценок коэффициентов сглаживающего полинома используется рекуррентная процедура, позволяющая при конечном числе наблюдений получить приближённое решение задачи. Причём приближение тем точнее, чем больше число наблюдений. В-третьих, проблема выбора начальных условий принципиально сводится к оценке погрешности метода, а вопрос выбора оптимального значения параметра сглаживания для своего решения требует, прежде всего, чёткой постановки задачи [15, с. 61]. Тем не менее, экспоненциальное сглаживание успешно применяется как в автоматическом режиме для большого числа временных рядов, так и в обычном анализе, с подробным изучением результатов, выбросов, реакции на импульс.

Среди альтернативных методов прогнозирования, не основанных на построении модели, стоит рассмотреть следующие:

- пошаговую авторегрессию, представляющую собой процедуру применения методологии Бокса-Дженкинса для автоматического определения лагов авторегрессии. Однако полная процедура Бокса-Дженкинса значительно проще в применении, поэтому данный метод используется редко;

- метод ARARMA, при котором к исходным временным рядам сначала применяется модель AR для удаления тренда (коэффициенты такой модели могут приближаться или превосходить единицу), а затем оценивается модель ARMA;

- общее экспоненциальное сглаживание (general exponential smoothing; GES). Данный метод был предложен Р. Брауном в 1963 г. и являлся очень востребованным. Временной ряд представляется в виде:

$$X_t = \sum_{i=1}^k a_i f_i(t) + \varepsilon_t, \quad (12.1)$$

где $\{a_i\}$ — постоянные величины; $f_i(t)$ — известные функции от времени, такие как полиномы или синусоиды.

Коэффициенты $\{a_i\}$ меняются со временем, поэтому модель (12.1) применяется локально. В соответствии с этим, для оценки модели используется метод взвешенных наименьших квадратов, присваивающий больший вес последним наблюдениям. Если в роли функции выступает константа, то GES сводится к простому экспоненциальному сглаживанию; если оценивается константа и линейный тренд, то GES сводится к двойному экспоненциальному сглаживанию. При наличии сезонных колебаний данный метод значительно отличается от сглаживания по методу Холта-Уинтерса [29, с. 102];

– другие методы, не нашедшие широкого применения по причине теоретических проблем или сомнительных выгод от их практического применения.

С точки зрения наглядности, более привлекательным является прогнозирование по модели временного ряда. В данной ситуации прогноз $x_N(h)$, является оценкой величины X_{N+h} , зависящей от модели (M) и информации, доступной на момент прогнозирования (I_N):

$$\hat{x}_N(h) = E(X_{N+h} | M, I_N). \quad (12.2)$$

В одномерном прогнозировании I_N представляет собой наблюдаемый временной ряд: уровни и моменты (периоды) времени, к которым относятся значения показателя. Модель M предполагается неизвестной, в прогнозировании используется её оценка.

Модели, разработанные Дж. Боксом и Г. Дженкинсом, не вызывают трудностей у статистиков, как и применение экспоненциального сглаживания. Прогнозы ARMA рассчитываются рекурсивно путём замены:

- 1) будущих значений шума — нулями;
- 2) будущих значений временного ряда — их условными ожиданиями;
- 3) текущих и прошлых значений временного ряда и шума — их фактически наблюдававшимися значениями.

Тем не менее, внимания требуют следующие замечания.

1 В связи со сложностью интерпретации графиков автокорреляционной и частной автокорреляционной функции аналитику необходим значительный опыт для выбора модели на их основе. Обойти данную проблему возможно, оценив сразу ряд моделей, а затем выбрав наиболее подходящую по информационным критериям, обсуждавшимся ранее.

2 Альтернативные процедуры оценки моделей ARIMA различаются, главным образом, обработкой первых значений временного ряда. К таким процедурам относится метод максимального правдоподобия (maximum likelihood; ML) и условных наименьших квадратов (conditional least squares; CLS). Например, при использовании CLS для оценки модели $AR(p)$, первые p значений временного ряда считаются фиксированными. При реализации метода ML, наоборот, несколько первых значений ряда усредняются. Для длинных временных рядов (несколько сотен наблюдений) выбор алгоритма оценки не несёт значительных различий, но для коротких временных рядов предпочтительно применение метода ML, в то время как использование оценок Юла-Уокера (метода моментов) не рекомендуется, особенно когда корни AR-процесса близки к единичному кругу [29, с. 91].

3 Процедуры тестирования, основанные на исследовании оценённых остатков, являются маломощными, в результате чего даже плохие модели редко отклоняются. Причиной служит малая длина временного ряда,

а также то, что модель формулируется, оценивается и проверяется на одном и том же массиве данных.

4 Для нестационарных временных рядов крайне значимым является порядок разностей, которыми удаляется тенденция. При этом следует рассматривать альтернативные методы удаления тренда перед оценкой модели ARMA (например, с помощью полиномиальных функций).

Существует различие между временными рядами, приводимыми к стационарному путём взятия разностей (разностно-стационарными, difference-stationary; DS), и рядами, у которых стационарными являются отклонения от детерминистического тренда (тренд-стационарными, trend-stationary; TS). В отличие от TS временных рядов, DS временные ряды имеют *стохастический* тренд. В работе П. Франсеса отмечается, что большинство экономических временных рядов являются DS, и DS модели на практике дают более точные послевыборочные прогнозы [34].

Временной ряд TS имеет конечную память о шоках, величина предыдущей ошибки никак не учитывается при подсчёте последующих значений. Для сохранения памяти о шоках (хотя бы на некоторое время) Г. Г. Канторовичем предложено заменить белый шум на процесс ARMA(p, q) для моделирования структуры ошибок в TS временных рядах [12, с. 288]. С другой стороны, в разложении DS-процесса присутствует сумма всех предыдущих ошибок, поэтому влияние шоков длится бесконечно долго. Экономически это не совсем целесообразно, так как предыдущие отклонения не должны сказываться постоянно, что особенно относится к длинным и высокочастотным временным рядам.

Для сохранения информации о шоках на длительное время (но не навсегда) разрабатывается теория процессов с длинной памятью как промежуточный вариант между процессами с короткой и бесконечной памятью.

Стационарный процесс X_t является процессом с длинной памятью, если существует действительное число α , $0 < \alpha < 1$, и константа c , $c > 0$, для которых выполняется условие [19, с. 14]:

$$\lim_{k \rightarrow \infty} \frac{\rho_k}{c \cdot k^{-\alpha}} = 1, \quad (12.3)$$

где ρ — коэффициент автокорреляции;

k — номер лага.

Длинная память в ряду определяется асимптотически. При её наличии каждая из автокорреляций в отдельности может быть слабой, тогда как их сумма должна оказаться высокой. Это делает ещё более затруднительным обнаружение длинной памяти, поскольку выявления отдельных лагов с высокой автокорреляцией недостаточно. Важным считается темп стремления автокорреляции к нулю по мере увеличения лага: чем сильнее влияние

длинной памяти, тем более медленной будет конвергенция к нулю, и тем большей будет сумма автокорреляций [19, с. 15].

Поведение временных рядов с длинной памятью описывается авторегрессионными частично интегрированными моделями скользящего среднего ARFIMA(p, d, q), допускающими нецелый параметр d [36]. Простая модель ARFIMA(0, d , 0) имеет вид:

$$(1-B)^d X_t = \varepsilon_t, \quad (12.4)$$

где B — лаговый оператор⁹.

С другой стороны,

$$\varepsilon_t = (1-B)^d X_t = \sum_{j=0}^{\infty} \pi_j B^j X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad (12.5)$$

где $\pi_0 = 1$;

$$\pi_{j+1} = \frac{(j-d)\pi_j}{j+1}, \quad (12.6)$$

то есть модель (12.4) можно представить в виде:

$$X_t - dX_{t-1} + \frac{d(d-1)}{2} X_{t-2} - \dots = \varepsilon_t. \quad (12.7)$$

Прогнозирование процессов с длинной памятью сходно с прогнозированием ARIMA. Уравнения (12.5) и (12.6) дают усечённый прогноз [45, с. 276]:

$$\hat{x}_N(h) = \sum_{j=1}^N \pi_j x_{N+h-j}. \quad (12.8)$$

При обнаружении длинной памяти не в уровнях временного ряда, а в его волатильности, используется модель FIGARCH(r, ξ, s):

$$\phi(L)(1-L)^{\xi} \varepsilon_t^2 = \omega + \beta(L) \bar{\varepsilon}_t^2, \quad (12.9)$$

где L — лаговый оператор;

$$\phi(L) = 1 - \alpha(L) - \beta(L) \bar{\varepsilon}_t^2; \quad (12.10)$$

$$\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_r L^r; \quad (12.11)$$

$$\beta(L) = \beta_1 L + \beta_2 L^2 + \dots + \beta_s L^s. \quad (12.12)$$

⁹Например, $(1-B)^2 = 1 - 2B + B^2$, где $B^2 X_t = X_{t-2}$.

Представленные модели частично интегрированных процессов являются действенными инструментами анализа и прогнозирования временных рядов, поскольку дают возможность одновременного моделирования эффектов длинной и короткой памяти. Моделирование длинной памяти на основе использования традиционных процессов ARMA с включением крайне большого количества лагов также возможно, но такой подход не позволяет оценить степень персистентности ряда. Кроме того, в этом случае для персистентного ряда корни полиномов близки к единице, вследствие чего асимптотические распределения оказываются плохими аппроксимациями распределений на конечные выборки [19, с. 16].

Стэйт-спэйс модели (state-space models) схожи по своей широте и применимости с линейной регрессией ([21], [29], [41], [45]). Данный класс моделей разработан инженерами и в настоящее время активно внедряется в статистику для анализа сигнала, сопровождающегося помехами:

$$\text{Наблюдаемое значение} = \text{сигнал} + \text{шум}. \quad (12.13)$$

Рассмотрим в доступной форме алгоритм их применения для прогнозирования временных рядов. В стэйт-спэйс моделях сигнал в момент времени t задаётся как линейная комбинация набора переменных, называемых стэйт-переменными, которые составляют так называемый стэйт-вектор (вектор состояний) в момент времени t . Если обозначить число переменных через m , то уравнение (12.13) можно записать в виде:

$$X_t = h_t^T \theta_t + n_t, \quad (12.14)$$

где h_t — известный вектор размерности $(m \times 1)$;

θ_t — стэйт-вектор размерности $(m \times 1)$;

n_t — ошибка наблюдения, имеющая нулевое среднее значение.

Набор стэйт-переменных — это такой минимальный объём текущей и прошлой информации, что будущее поведение системы полностью определяется текущими значениями стэйт-переменных. Таким образом, будущее не зависит от прошлых значений, что говорит о наличии у стэйт-вектора свойств марковского процесса: для его прогноза требуется знание лишь последнего, наиболее позднего значения. Следовательно, стэйт-вектор изменяется во времени в соответствии со стэйт-уравнением (уравнением системы):

$$\theta_t = G_t \theta_{t-1} + \omega_t, \quad (12.15)$$

где матрица G_t размерности $(m \times m)$ предполагается известной, а ω_t обозначает m -вектор шума, имеющий нулевое среднее.

Подразумевается, что на протяжении всего периода времени отсутствует автокорреляция и корреляция ошибок в моделях (12.14) и

(12.15). Дополнительным условием является нормальность для n_t : $n_t \sim N(0, \sigma_n^2)$. При этом ω_t имеет многомерное нормальное распределение с нулевым вектором средних значений и известной вариационно-ковариационной матрицей¹⁰ W_t .

Уравнения (12.14) и (12.15) представляют собой общую форму одномерной стэйт-спэйс модели, которая адаптируется для оценки текущего стэйт-вектора и прогнозов каждый раз, когда становится доступной новая информация. Данная процедура называется фильтром Калмана.

Если известна точная форма стэйт-спэйс модели, включая текущее значение стэйт-вектора, то прогноз по (12.14) имеет вид:

$$\hat{x}_N(h) = h_{N+h}^T \theta_{N+h}, \quad (12.16)$$

где h_{N+h} предполагается известным, а $\theta_{N+h} = G_{N+h} G_{N+h-1} \dots G_{N+1} \theta_N$ при условии, что известны будущие значения G .

На практике θ_N оценивается на основе информации, доступной вплоть до момента времени N . Важным специальным случаем модели является тот, когда h_t и G_t представляют собой известные постоянные функции, например, h и G . Тогда формула прогноза принимает вид:

$$\hat{x}_N(h) = h^T G^h \theta_N, \quad (12.17)$$

При известной структуре модели вычисление прогнозов зависит лишь от возможности получить качественные текущие оценки стэйт-вектора θ_N . Фильтр Калмана проводится в два этапа. Первый этап, называемый этапом экстраполяции, связан с прогнозированием θ_t на основе данных, доступных в момент времени $(t-1)$:

$$\hat{\theta}_{t|t-1} = G_t \hat{\theta}_{t-1}. \quad (12.18)$$

Вариационно-ковариационная матрица данной оценки определяется уравнением:

$$P_{t|t-1} = G_t P_{t-1} G_t^T + W_t. \quad (12.19)$$

Когда становится доступной новая информация x_t в момент времени t , осуществляется второй этап фильтрации Калмана, называемый коррекцией:

¹⁰ Диагональные значения вариационно-ковариационной матрицы являются дисперсиями элементов θ , а значения вне диагонали — ковариациями между различными элементами, с. 93].

$$\hat{\theta}_t = \hat{\theta}_{t|t-1} + K_t e_t, \quad (12.20)$$

$$P_t = P_{t|t-1} - K_t h_t^T P_{t|t-1}, \quad (12.21)$$

где e_t — ошибка прогноза в момент t :

$$e_t = x_t - h_t^T \hat{\theta}_{t|t-1}; \quad (12.22)$$

где K_t — матрица коэффициентов усиления:

$$K_t = \frac{P_{t|t-1} h_t}{h_t^T P_{t|t-1} h_t + \sigma_n^2}. \quad (12.23)$$

Следующие особенности важны при спецификации стэйт-уравнений в статистических программных продуктах:

- стэйт-уравнения не содержат зависимых переменных уравнений сигнала;

- в каждом стэйт-уравнении должна быть линейная связь с предыдущим состоянием (лаг=1). Нелинейность, наличие других лагов или одновременных взаимосвязей не допускается. Данное требование, тем не менее, не ограничивает использование лагов высоких порядков: они выражаются в виде новой, стэйт-переменной.

- стэйт-уравнения могут содержать экзогенные переменные и неизвестные коэффициенты, а также быть нелинейными по этим элементам.

Аспекты спецификации уравнений сигнала:

- уравнения сигнала не содержат текущих или будущих значений сигнальных переменных;

- уравнение сигнала должно быть линейным по отношению к текущим состояниям. Нелинейность и лаги не допускаются. Данное ограничение обходят путём включения новых детерминированных переменных состояния для выражения лаговых значений;

- уравнения сигнала могут включать экзогенные переменные и быть нелинейными по этим параметрам.

Ошибки или их дисперсия в каждом уравнении стэйт-спэйс модели задаются отдельно. Если они не указываются, то данное уравнение предполагается детерминированным. Различные варианты спецификации позволяют варьировать наличие корреляции между ошибками уравнений.

Нами представлены, главным образом, линейные модели для прогнозирования временных рядов. Это объясняется трудностями выбора подходящего типа нелинейной модели и метода оценки её параметров, а также тем, что нелинейные модели (как и многомерные) улучшают точность прогнозирования незначительно. Тем не менее, нелинейные модели интересны применяемым математическим аппаратом, возможностью более глубокого исследования сущности изучаемого процесса.

12.2 Анализ причинности взаимосвязей в многофакторном прогнозировании временных рядов

Одномерные модели применяются для многих целей, включая прогнозирование большого количества временных рядов и обеспечение стандарта в сравнительных прогнозных исследованиях. В то же время, многомерные модели могут многое предложить в достижении лучшего понимания глубинной структуры рассматриваемой системы и в получении лучших прогнозов.

Многомерные модели часто дают лучшее приближение к данным выборки, чем одномерные модели, но имеются причины, по которым прогнозы не всегда подтверждаются на практике [29, с. 110]:

- чем больше оцениваемых параметров, тем больше возможностей для увеличения вариации, роста неопределённости параметров;
- чем больше переменных для измерения, тем больше вероятность ошибок и появления выбросов;
- наблюдаемые многомерные данные не всегда пригодны для построения многомерной модели;
- вычисление прогнозов зависимой переменной может потребовать будущие значения объясняющих переменных, которые недоступны на момент прогнозирования. В таком случае объясняющие переменные должны быть предсказаны каким-либо способом до того, как прогноз зависимой переменной будет найден, что неизбежно ведёт к уменьшению точности. Если прогнозы объясняющих переменных имеют низкую точность, то результирующие прогнозы зависимой переменной могут иметь точность ниже, чем одномерные прогнозы;
- вычисление многомерных прогнозов зависит от наличия хорошей многомерной модели, что не всегда гарантировано. Как и одномерная модель, многомерные аналоги содержат ошибки идентификации, чувствительны к длине анализируемых временных рядов и могут существенно меняться с течением времени. В целом, многомерные модели сложны и более уязвимы к неправильной спецификации, чем одномерные модели.

Опыт в поиске подходящей многомерной модели обосновывает важность правильной постановки проблемы, получения достаточной дополнительной информации с целью учесть контекст и идентифицировать все значимые объясняющие переменные. В общем случае применяется итеративный метод построения. При этом всегда присутствует некоторое противоречие между:

- а) включением лишних объясняющих переменных, которые, как кажется, улучшают адекватность модели, но в действительности ведут к плохим прогнозам за пределами выборки; и

б) исключением из рассмотрения переменных, которые действительно нужны.

Целью является поиск лаконичной модели (с меньшим числом оцениваемых параметров) при обеспечении включения в анализ важных переменных.

Подбор многомерных моделей к временным данным всё ещё не лёгок, несмотря на значительные улучшения в программном обеспечении. Исследования в данном направлении стимулируются совершенствованием вычислительных ресурсов, но остаётся много нерешённых проблем.

При многомерном прогнозировании перед исследователем стоят два важных вопроса:

1) существует ли обратная связь?

Отвечая на данный вопрос, определяют, должна ли модель включать в себя одно уравнение или несколько. Если обратная связь отсутствует, то переменные образуют незамкнутую систему, описываемую, например, уравнением множественной регрессии. Если же «результат» влияет на «факторы», то система является замкнутой, и модель с одним уравнением регрессии может ввести в заблуждение при её ошибочном применении.

2) действительно ли прогнозы лежат вне выборки?

Оценивая прогнозные качества модели на тестовой выборке, необходимо убедиться, что получаемые значения действительно являются вневыборочными прогнозами, исключить возможное использование «будущих» значений объясняющих переменных тестовой выборки, недоступных при истинном вневыборочном прогнозировании. Желание отказаться от прогнозирования объясняющих переменных и получить при этом истинно вневыборочные прогнозы объясняет поиск так называемых ведущих индикаторов. Например, в регрессионном уравнении

$$Y_t = a + bX_{t-d} + \varepsilon_t \quad (12.24)$$

объясняющая переменная X_t является ведущим индикатором для результативной Y_t , так как модель способна предсказывать значения Y_t до d шагов вперёд без необходимости прогнозировать ряд X_t . При горизонте прогноза больше d необходимое значение X_t недоступно. В результате возникнет необходимость прогноза значений X_t , но необходимый период прогнозирования ведущего индикатора, по крайней мере, меньше, чем для результативной переменной. Таким образом, многомерная модель способна давать «хорошие» прогнозы зависимой переменной при условии, что значения всех необходимых объясняющих переменных известны (если они — ведущие индикаторы) или могут быть точно спрогнозированы.

Модель множественной регрессии для временных рядов имеет вид:

$$Y_t = \beta^T x_t + u_t, \quad (12.25)$$

где x_t — вектор объясняющих переменных в момент времени t ;

u_t — ошибки наблюдения в момент времени t .

На первый взгляд, оценить данную модель на основе временных рядов просто. Если допустить, что x_t — предопределённые переменные, которые не включают в себя никаких лаговых значений (ни объясняемой, ни объясняющих переменных), и что переменные u_t независимы, $N(0, \sigma^2)$ и некоррелированы со значениями x_t , то оценка модели (12.25) может быть проведена обычным МНК. Однако эти предположения неправдоподобны для временных рядов. В частности, ошибки наверняка будут коррелированы.

Тем не менее, аналитики до сих пор используют МНК, даже когда необходимые предпосылки не могут быть выполнены. Это происходит частично из-за доступности и лёгкости использования программного обеспечения для МНК, а также трудности выбора подходящей альтернативной процедуры оценки. Чтобы решить данный вопрос отметим, что обобщённый метод наименьших квадратов (ОМНК) используется, когда предполагается (авто)корреляция u_t с известной структурой, тогда как итеративная процедура Кохрейна-Окутта (Cochrane-Orcutt procedure) применяется, когда u_t представляют собой AR(1) процесс с неизвестным параметром. Если предполагается корреляция u_t с одной из переменных x_t , то используют ДМНК. Также существуют теоретические выкладки для работы с такими отклонениями от основных регрессионных предпосылок, как наличие стохастических регрессоров вместо предопределённых, негауссовские ошибки в модели и другими.

Благодаря современному программному обеспечению практически отсутствуют вычислительные трудности при подборе регрессионных моделей на основе перебора доступных объясняющих переменных и их преобразований, лаговых структур и так далее. В результате исследователь получает модель, демонстрирующую ложно хорошее приближение к исходным данным, неспособную давать качественные прогнозы.

Чтобы избежать ложных регрессий, необходимо:

- провести строгий отбор объясняющих переменных, ограничив их число;
- удалить тренд и сезонные колебания перед построением регрессионной модели (однако при наличии коинтеграции это приведёт к неправильной спецификации);
- подробно идентифицировать структуру ошибок для использования подходящей процедуры оценки коэффициентов;
- провести тщательную проверку каждой оценённой модели.

Тем не менее, даже с учётом указанных предосторожностей, использование моделей множественной регрессии для прогнозирования временных рядов не гарантирует получение оптимального результата. Поэтому наилучшей мерой предосторожности может оказаться отказ от использования регрессионных моделей в пользу альтернативных классов.

Модели передаточных функций — более безопасный вариант для моделей с одним уравнением, так как они допускают более общую структуру ошибок и более общую смещённую во времени взаимосвязь:

$$Y_t = \nu_0 X_t + \nu_1 X_{t-1} + \dots + n_t = \nu(B)X_t + n_t, \quad (12.26)$$

где $\nu(B) = \nu_0 + \nu_1 B + \nu_2 B^2 + \dots$ — полиномиальный оператор обратного сдвига B ; $\{n_t\}$ — ошибки наблюдения (шум).

На ошибки не накладывается условие независимости, но они должны быть некоррелированы с анализируемыми временными рядами. Объясняющие переменные при этом являются стохастическими или предопределёнными величинами.

Модели вида (12.26) называются моделями передаточно-шумовой функции, передаточной функции или моделями с распределённым лагом (потому что они «распределяют смещённую во времени взаимосвязь» на множество лагов). По сравнению с регрессионными, такие модели более подходят для описания динамики разомкнутого причинного взаимодействия между одной факторной переменной и одной результативной [29, с. 124].

В отечественных работах по эконометрике ранее освещались вопросы, касающиеся свойств моделей с распределённым лагом, оценки их параметров и прогнозирования по ним. В частности, изучение структуры лага, метод Алмон и метод Койка, подробная интерпретация параметров моделей представлены в работе коллектива авторов под редакцией И. И. Елисеевой [23]; особенности свойств ошибок модели и оценки параметров описаны в учебнике Н. П. Тихомирова и Е. Ю. Дорохиной [24].

Тем не менее, отметим, во-первых, что в правой части модели (12.26) отсутствует константа. Её включение усложняет вычисления, а также форма анализируемых данных обычно делает наличие свободного члена необязательным (если переменные скорректированы на среднее значение или являются отклонениями от равновесного значения). Во-вторых, если $\nu_0 \neq 0$, то присутствует одновременная взаимосвязь между X_t и Y_t . Чтобы избежать рассмотрения обратной связи аналитик должен быть готов приписывать это отношение зависимости Y от X , а не наоборот.

Наличие обратной связи в системе учитывается VAR-моделями. Прогнозы с минимальной средней квадратической ошибкой вычисляются для моделей VAR и VARMA (VMA-модели обычно не используются на практике). Процесс моделирования весьма затруднителен, для снижения числа оцениваемых параметров на модели накладываются ограничения, в том числе используется подход, называемый байесовской векторной авторегрессией (Bayesian vector autoregression; BVAR). Данный метод используется вне зависимости от того, придерживается ли аналитик байесовской теории или нет. Сущность техники заключается в стремлении предотвратить переоценку количества параметров с помощью сведения числа параметров больше первого порядка к нулю. Априорное правило Миннесота

(Minnesotaprior), гласит, что значения, априорно совместимые с любыми другими рядами, являются случайным блужданием [29, с. 136].

Использование данных методов расширяет возможности использования векторных авторегрессионных моделей. По аналогии с одномерным случаем, они строятся и для нестационарных данных (VARIMA-модели) после приведения временных рядов к стационарному виду. Следует подчеркнуть, что для различных рядов могут потребоваться различные порядки разностей, поэтому на практике неуместно находить одинаковые разности для всех компонентов. Как и в случае множественной регрессии, взятие разностей (удаление тренда) приводит к неправильной спецификации при присутствии коинтеграционных взаимосвязей в системе. Таким образом, взятие разностей не является универсальным способом подбора модели VAR к нестационарным данным. Главной причиной VAR-моделирования часто служит попытка получить большее понимание изучаемой системы, а не желание получить более точные прогнозы.

Учесть наличие коинтеграции между переменными позволяют модели корректировки ошибок. Если для временных рядов $X_t \sim I(1)$ и $Y_t \sim I(1)$ (то есть разностно-стационарных временных рядов, имеющих единичный корень) существует линейная комбинация $Y_t + \beta X_t \sim I(0)$, то временные ряды являются коинтегрированными, а вектор с компонентами (α, β) — коинтеграционным вектором. Уникальность коинтеграционного вектора обеспечивается равенством единице первого его компонента.

Другими словами, если в остатках регрессии

$$Y_t = b_0 + b_1 X_t + u_t \quad (12.27)$$

отсутствует единичный корень, то $X_t, Y_t \sim CI(1,1)$, то есть коинтегрированы¹¹. Коинтеграционные взаимосвязи прослеживаются в долгосрочной динамике, длинных экономических временных рядах (например, между потреблением и ВВП, потреблением и доходом, ценами и заработной платой). Отклонения от равновесного состояния оцениваются по формуле:

$$e_t = Y_t - \hat{b}_0 - \hat{b}_1 X_t. \quad (12.28)$$

Будучи включёнными в регрессионную модель, они обеспечивают корректировку краткосрочного отклонения от равновесия:

$$\Delta Y_t = c_0 + c_1 \Delta X_t + c_2 e_{t-1} + u_t. \quad (12.29)$$

¹¹Если процессы $X_t \sim I(d)$ и $Y_t \sim I(d)$ имеют одинаковый порядок интеграции d , и существует такой вектор (α, β) , что $\alpha X_t + \beta Y_t \sim I(d-b)$, то $X_t, Y_t \sim CI(d, b)$, где $b > 0$ и обозначает, на сколько снижается порядок интеграции. Также запись $X_t, Y_t \sim CI(1, 1)$ равнозначна записи $X_t, Y_t \sim CI$.

В работе Г. Г. Канторовича [13] показано, что данная простейшая модель корректировки ошибок соответствует авторегрессионной модели с распределёнными лагами¹² ADL(1,1). При этом оценки ЕСМ-модели (12.29) заведомо состоятельны, так как все её регрессоры являются стационарными. Более того, оценка МНК коэффициента наклона в модели (12.27) также является состоятельной, и скорость её сходимости выше (пропорциональна T), чем в случае оценки МНК для стационарных регрессий (пропорциональна $\frac{1}{\sqrt{T}}$). Данное свойство оценок МНК при наличии коинтеграции называют суперсостоятельностью [13, с. 90].

Для случая скрытой коинтеграции К. Грэнджер [38] представил модель корректировки ошибок (crouching error correction model), которая при сохранении обозначений к формулам (11.18)–(11.22) и наличия коинтеграционного вектора (1,–1) для величин $\left\{ \sum_{i=1}^t \varepsilon_i^+ \quad \sum_{i=1}^t \eta_i^+ \right\}$ имеет вид:

$$\varepsilon_t^+ = \gamma_0 + \gamma_1 \left(\sum_{i=1}^{t-1} \varepsilon_i^+ - \sum_{i=1}^{t-1} \eta_i^+ \right) + \text{lags} \left(\varepsilon_{t-1}^+, \eta_{t-1}^+ \right) \zeta_t, \quad (12.30)$$

$$\eta_t^+ = \delta_0 + \delta_1 \left(\sum_{i=1}^{t-1} \varepsilon_i^+ - \sum_{i=1}^{t-1} \eta_i^+ \right) + \text{lags} \left(\varepsilon_{t-1}^+, \eta_{t-1}^+ \right) \zeta_t, \quad (12.31)$$

где $\text{lags} \left(\varepsilon_{t-1}^+, \eta_{t-1}^+ \right)$ — добавочные регрессоры с различными лагами ε_{t-1}^+ и η_{t-1}^+ ; ξ_t и ζ_t — белый шум.

Для наличия коинтеграции хотя бы один из коэффициентов (γ_1 или δ_1) должен быть отличен от нуля. Через приращения модель записывается следующим образом:

$$\Delta X_t^+ = \gamma_0 + \gamma_1 \left(X_{t-1}^+ - Y_{t-1}^+ \right) + \text{lags} \left(X_{t-1}^+, \Delta Y_{t-1}^+ \right) \xi_t. \quad (12.32)$$

Аналогичный вид имеет запись для ΔY_t^+ .

Предположим, что $\left\{ \sum_{i=1}^t \varepsilon_i^- \quad \sum_{i=1}^t \eta_i^- \right\}$ коинтегрированы с вектором (1,– k), $k \neq 1$, и имеют следующие модели корректировки ошибок:

$$\varepsilon_t^- = \gamma'_0 + \gamma_2 \left(\sum_{i=1}^{t-1} \varepsilon_i^- - k \sum_{i=1}^{t-1} \eta_i^- \right) + \text{lags} \left(\varepsilon_{t-1}^-, \eta_{t-1}^- \right) \xi_t, \quad (12.33)$$

$$\eta_t^- = \delta'_0 + \delta_2 \left(\sum_{i=1}^{t-1} \varepsilon_i^- - k \sum_{i=1}^{t-1} \eta_i^- \right) + \text{lags} \left(\varepsilon_{t-1}^-, \eta_{t-1}^- \right) \zeta_t. \quad (12.34)$$

¹² autoregressivedistributedlag или ARMAX («X» в данном акрониме обозначает «Xogenous» — экзогенный).

Временные ряды X_t и Y_t остаются при этом неинтегрированными, однако имеет место следующая зависимость, полученная из представленных выше сжатых моделей корректировки ошибок:

$$\Delta X_t = \gamma_0 + \gamma_1 (X_{t-1}^+ - Y_{t-1}^+) + \gamma_2 (X_{t-1}^- - kY_{t-1}^-) + \text{lags}(X_{t-1}^+, \Delta Y_{t-1}^+, \Delta X_{t-1}^-, \Delta Y_{t-1}^-) + \xi_t \quad (12.35)$$

и

$$\Delta Y_t = \delta_0 + \delta_1 (X_{t-1}^+ - Y_{t-1}^+) + \delta_2 (X_{t-1}^- - kY_{t-1}^-) + \text{lags}(X_{t-1}^+, \Delta Y_{t-1}^+, \Delta X_{t-1}^-, \Delta Y_{t-1}^-) + \zeta_t. \quad (12.36)$$

Модели (12.35) и (12.36) являются более общими, чем обычные модели корректировки ошибок, так как накладывают меньше ограничений на свои коэффициенты. При этом они сохраняют свойства данного класса моделей за исключением того, что показывают долговременные равновесные взаимосвязи и кратковременную динамику нестационарных компонентов, а не самих временных рядов.

Развитие прогностических свойств моделей корректировки ошибок связано, в частности, со спецификацией коинтеграционного соотношения: при необходимости в него кроме константы включается детерминированный тренд, позволяющий учесть характер нестационарных временных рядов. Ключевой вопрос в том, требуется ли прогнозировать отдельные переменные (обычный случай) или линейную комбинацию переменных, которая делает систему коинтегрированной. Различия в том, что индивидуальные переменные обычно являются нестационарными и, следовательно, имеют интервалы прогноза, расширяющиеся с увеличением горизонта прогнозирования. Стационарная линейная комбинация переменных, напротив, имеет конечный интервал прогнозирования при увеличении горизонта прогноза.

Мнения учёных относительно выгод использования коинтеграционных соотношений в прогнозировании существенно варьируются. Теоретически очевидно, что данный подход предпочтительнее подхода Бокса-Дженкинса. Последний во избежание ложной регрессии прибегает к регрессии первых разностей, однако при переходе к разностям теряется информация о долгосрочной взаимосвязи временных рядов, а также нарушается спецификация модели, если она выражена в терминах экономической теории. Модели корректировки ошибок, оперирующие с коинтеграционным равновесием, описывают как краткосрочное, так и долгосрочное поведение временных рядов, используя их начальные уровни.

К. Четфилд [29] отмечает, что улучшенные прогнозы не всегда доступны на реальных данных в связи с тем, что стандартным методам вычисления точности прогноза не удаётся оценить сохранение коинтеграционных взаимосвязей между переменными. Авторы рекомендуют анализировать коинтеграцию при стремлении смоделировать и понять

многомерные экономические данные, хотя уверенность в том, что использование коинтеграции приведёт к улучшенным вневыборочным прогнозам, пока ещё невелика.

12.3 Сравнительная оценка альтернативных методов прогнозирования и обобщение прогнозов

Богатство методологии анализа временных рядов делает непростым выбор конкретного метода для проведения исследования и прогнозирования (в конечном счёте — для принятия решения). Естественное стремление выбрать «лучший» метод из всех, но их растущее многообразие приводит к мысли, что одного метода, лучшего, чем все остальные, не существует.

Практики подтверждают, что метод, подходящий для одномерного временного ряда, может оказаться непригодным для анализа большого массива данных. Метод, применяемый одним аналитиком, не будет принят другим исследователем из-за их различных уровней осведомлённости и (или) имеющегося в распоряжении программного обеспечения. В некоторых случаях даже применение одного и того же метода для анализа одного и того же процесса становится со временем неуместным, например, в связи с изменением характера тенденции.

Для выбора метода прогнозирования необходимо учесть результаты, полученные во время структурирования проблемы и определения цели исследования. Они определяют критерии для сравнения и выбора оптимального метода:

- 1) точность прогноза,
- 2) стоимость прогноза,
- 3) экспертная оценка аналитика,
- 4) доступность программного обеспечения,
- 5) свойства прогнозируемых временных рядов,
- 6) направление использования результатов,
- 7) другие критерии, зависящие от обстоятельств.

Большинство исследователей сходятся во мнении, что под лучшим методом понимается тот, который даёт наиболее точные оценки на заданном горизонте прогнозирования. Это не всегда так, и любой из оставшихся критериев может оказаться решающим в выборе. Тем не менее, критерии 2–7 являются контекстно-зависимыми, поэтому в данной работе не рассматриваются.

Качества прогноза оценивается с помощью абсолютных, сравнительных или качественных показателей [9, с. 198]. Все они базируются на расчёте абсолютной ошибки прогноза (для данных тестовой выборки):

$$e_t = x_t - \hat{x}_{t-1} \quad (12.37)$$

К абсолютным показателям также относится средняя квадратическая ошибка прогноза (prediction means square error; PMSE):

$$PMSE = \frac{\sum_{t=N+1}^{N+h} e_t^2}{h}, \quad (12.38)$$

где N — время составления прогноза (номер последнего уровня выборки),
 h — число прогнозируемых уровней;

и средняя абсолютная ошибка (mean absolute error; MAE):

$$MAE = \frac{\sum_{t=N+1}^{N+h} |e_t|}{h}. \quad (12.39)$$

Данные показатели широко используются на практике, но обоснование выбора одного из них зачастую опускается. Причиной служит недостаточное внимание к определению функции потерь.

Концепция функции потерь (или функции стоимости) является ключевой в прогнозировании. Цель данной функции — оценить потери, связанные с ошибками прогноза различных величин и знаков. Например, если ошибки прогноза возрастут в два раза, то это будет в два раза хуже (линейная функция) или в четыре (квадратическая)? Равнозначна ли положительная по знаку ошибка такой же по величине ошибке, но с отрицательным знаком (другими словами, симметрична ли функция потерь)? Эти и другие вопросы должны быть решены аналитиком (возможно, субъективно) в соответствии с контекстом конкретной анализируемой ситуации. Например, формула (12.38) подразумевает квадратическую функцию потерь, а (12.39) — линейную.

Недостатком рассматриваемых показателей является то, что их значения существенно зависят от масштаба измерения уровней исследуемых явлений, поэтому используется средняя абсолютная процентная ошибка (mean absolute percentage error; MAPE) [1, с. 69], [9, с. 200]:

$$MAPE = \frac{1}{h} \sum_{t=N+1}^{N+h} \frac{|e_t|}{x_t} \cdot 100. \quad (12.40)$$

Степень точности обычно определяют по значению MAPE, применяя следующую шкалу [1, с. 69]:

MAPE < 10% — высокая точность;

10% < MAPE < 20% — хорошая точность;

20% < MAPE < 50% — удовлетворительная точность;

$MAPE > 50\%$ — неудовлетворительная точность.

$MAPE$ используется для сравнения точности прогнозов разнородных объектов прогнозирования, так как характеризует относительную точность прогноза. Данный показатель нацелен на наличие положительных значений показателя: нулевые значения должны пропускаться в связи с тем, что x_t находится в знаменателе дроби, а чередующиеся положительные и отрицательные уровни (например, если ряд колеблется вокруг нулевого среднего) в сумме компенсируют значения дроби $\frac{|e_t|}{x_t}$. В связи с этим применима модифицированная формула:

$$MAPE_M = \frac{1}{h} \sum_{t=N+1}^{N+h} \left| \frac{e_t}{x_t} \right| \cdot 100. \quad (12.41)$$

Среди сравнительных показателей точности прогноза следует выделить коэффициент несоответствия прогнозов:

$$K = \frac{\sum_{t=N+1}^{N+h} |e_t|}{\sum_{t=N+1}^{N+h} |x_t - x_{t-1}^*(1)|}, \quad (12.42)$$

где $x_{t-1}^*(1)$ — значение эталонного прогноза.

Тем не менее, какой бы показатель точности прогноза ни был выбран, результаты для различных методов часто оказываются сходными, близкими друг к другу. Возникает вопрос, являются ли различия между методами значительными, и что именно значительно в конкретном случае. Поэтому сравнительная оценка различных прогнозов завершается качественными показателями, отображением прогнозных значений на временном графике и визуальным определением возможности такого развития процесса. В том числе, в работе А. Г. Гранберга [9, с. 202] предложено использование диаграммы «прогноз-реализация», представляющей собой облако точечных прогнозов (ось ординат) в координатах фактических значений переменной (ось абсцисс). Использование диаграммы позволяет содержательно оценить качество различных прогнозов (в том числе прогнозирования поворотных точек), наиболее типичные ошибки (недооценки или переоценки изменений).

Если исследователь всё же испытывает затруднения в выборе того или иного метода или модели прогнозирования, то используется объединение прогнозов. Эмпирические исследования подтверждают, что даже простое усреднение индивидуальных прогнозов способствует улучшению результатов, так как учитывает больше информации, полученной от прогнозирования различными способами. Усреднение проводится с применением простой или взвешенной средней (веса устанавливаются

экспертным путём или на основе качества модели), а также с помощью байесовского усреднения моделей. Последний способ является наиболее теоретически обоснованным, но и наиболее сложным. Если традиционно «наилучшая» модель выбирается из конкурирующих на основе её показателей точности, информационных критериев (AIC, SIC и других), а затем используется для прогнозирования, то байесовский подход подразумевает, что временной ряд генерируется не одной, а несколькими вероятностными моделями. Однако байесовское усреднение моделей трудно реализовать на практике, в том числе из-за того, что необходимо априорно определить круг подходящих для его прогнозирования моделей и присвоить им априорные вероятности (апостериорные вероятности рассчитываются при последующем анализе временного ряда).

В данной работе предложен более простой метод усреднения прогнозов, представляющий собой формализованную процедуру объединения, основанного на показателе точности:

$$\bar{x}_N(h) = \sum_{i=1}^n \hat{x}_i(h) d_i, \quad (12.43)$$

где n — число сравниваемых методов, а вес каждого прогноза в усреднённом рассчитывался по формуле:

$$d_i = \frac{1/C_i}{\sum_{i=1}^n 1/C_i} \cdot 100, \quad (12.44)$$

где C_i — показатель качества для каждого из прогнозов.

Тем не менее, применение усреднения прогнозов лишает аналитика таких преимуществ, как простота модели, обоснованность доверительных интервалов прогнозирования, экономическая интерпретируемость параметров.

Глава 13 Учебный пример

Что необходимо знать из главы 3:

- 1. Порядок проведения корректировок исходных временных рядов и подготовки их для анализа.*
- 2. Как на практике применять концепцию Грэнджер-причинности.*
- 3. Уметь правильно интерпретировать результаты тестирования гипотез и оценки параметров.*
- 4. Эвристический метод усреднения прогнозов и возможность его модификации для альтернативных функций потерь.*

В качестве примера рассмотрим применение методов и моделей, представленных ранее в данной работе, используя данные о состоянии и развитии экономики Российской Федерации [49].

В качестве основного показателя для прогноза выбран Π_0 — индекс реальных инвестиций в основной капитал, % (01.1994 г.=100%). Факторные показатели выбирались на основе изучения работ по анализу инвестиций [4, глава 6], [5, глава 4], [6, глава 2], [7, глава 3], [16] с учётом доступной информации:

Π_1 — индекс реального объёма промышленного производства, % (01.1993 г.=100%);

Π_2 — денежный агрегат М2 (на конец периода), млрд. (трлн.) руб.;

Π_3 — официальный курс доллара, руб./долл.;

Π_4 — задолженность предприятий в бюджет, млрд. руб.;

Π_5 — индекс цен на строительно-монтажные работы, % к предыдущему периоду;

Π_6 — индекс потребительских цен, % к предыдущему периоду;

Π_7 — индекс промышленного производства в электроэнергетике, % (01.1993 г.=100%);

Π_8 — индекс промышленного производства в газовой промышленности, % (01.1993 г.=100%);

Π_9 — индекс промышленного производства в нефтедобывающей промышленности, % (01.1993 г.=100%).

Базовые индексы (Π_0 , Π_1 , Π_7 , Π_8 , Π_9) пересчитаны в цепные, среднемесячные значения моментных показателей (Π_2 , Π_3 , Π_4) рассчитаны по формуле средней хронологической простой, стоимостные показатели (Π_2 , Π_4) переведены в постоянные цены декабря 2007 г. В результате получена система сопоставимых показателей, состоящая из базовой выборки (с 02.1994 г. по 12.2007 г.) и тестовой выборки (01.2008–02.2010 гг.):

Y — индекс реальных инвестиций в основной капитал, % к предыдущему периоду;

X_1 — индекс реального объёма промышленного производства, % к предыдущему периоду;

X_2 — среднемесячная величина денежного агрегата М2 в ценах декабря 2007 г., млрд. руб.;

X_3 — среднемесячный официальный курс доллара, руб./долл.;
 X_4 — среднемесячная задолженность предприятий в бюджет в ценах декабря 2007 г., млрд. руб.;
 X_5 — индекс цен на строительные-монтажные работы, % к предыдущему периоду;
 X_6 — индекс потребительских цен, % к предыдущему периоду;
 X_7 — индекс промышленного производства в электроэнергетике, % к предыдущему периоду;
 X_8 — индекс промышленного производства в газовой промышленности, % к предыдущему периоду;
 X_9 — индекс промышленного производства в нефтедобывающей промышленности, % к предыдущему периоду.

Временной график изучаемого показателя представлен на рисунке 13.1.

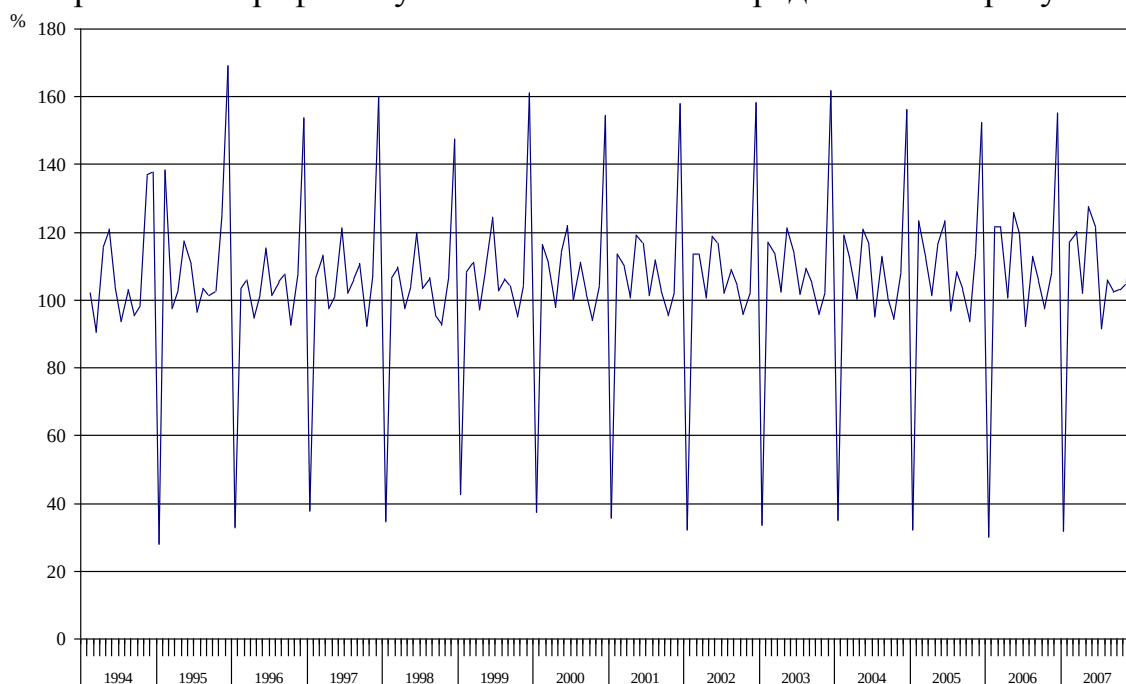


Рисунок 13.1 — Динамика индекса реальных инвестиций в основной капитал, % к предыдущему периоду

Далее рассмотрим процесс прогноза Y различными способами на период 01.2008–02.2010 гг.

13.1 Наивная модель

Получим самые первые прогнозы Y_c помощью простой наивной модели, для которой не требуется никаких расчётов. Значения наивных прогнозов будут использоваться для расчёта коэффициента несоответствия K по формуле (12.42).

В обычной наивной модели все прогнозы равны наиболее позднему значению. Учитывая, что мы имеем дело с ежемесячными данными, в которых может присутствовать сезонность, прогнозным значением будем

считать наиболее поздние фактические значения для соответствующего месяца. Например, прогнозом на 07.2009 г. служит фактическое значение за 07.2007 г.; прогнозом на 01.2008 г., 01.2009 г. и 01.2010 г. — фактическое значение 01.2007 г. Аналитически представим это следующим образом:

$$\hat{y}_N(h) = Y_{N-\tau+h-\tau \cdot \left[\frac{h}{\tau} \right]}, \quad (13.1)$$

где τ — период сезонных колебаний, $N \geq \tau$ (для ежемесячных данных $\tau = 12$); $\left[\frac{h}{\tau} \right]$ — целая часть от деления номера прогнозного значения на период сезонных колебаний.

Формула (3.1) позволяет «прогнозировать» на любой период вперёд, несмотря на то, за какой месяц известны последние фактические данные.

13.2 Детерминированный полиномиальный тренд

Для оценки традиционного полиномиального тренда проведена предварительная корректировка ряда на аддитивную сезонность с помощью процедуры сезонной корректировки «X-11» в STATISTICA 8.0.

Результаты дисперсионного теста (результаты D8 «X-11») подтвердили наличие стабильной сезонности во временном ряду Y на уровне значимости 5% (таблица 13.1).

Таблица 13.1 — Дисперсионный анализ наличия сезонности в ряду Y

Эффект	Сумма квадратов	Число степеней свободы	Средний квадрат	F-критерий	Вероятность, p
Между месяцами	110542,48	11	10049,32	250,89	0,00
Остаточный	6208,53	155	40,06		
Общий	116751,01	166			

Средние показатели сезонности получены путём усреднения окончательных значений сезонных факторов, представленных в результатах D10 метода. Окончательно скорректированный на сезонность временной ряд $Y_{(sa)}$ (отражается в результатах D11 процедуры) использовался для оценки параметров линейного тренда:

$$\hat{y}_{(sa)t} = a + b \cdot t, \quad (13.2)$$

где $\hat{y}_{(sa)t}$ — прогноз уровня (оценка) временного ряда, скорректированного на сезонность.

Как видно из таблицы 13.2, уравнение тренда является статистически значимым на уровне 0,05; в том числе значимы все его коэффициенты. Тем не менее, лишь 4,3% колеблемости десезонализованных уровней временного

ряда (3,7% — в генеральной совокупности) объясняются наличием линейного тренда.

Таблица 13.2 — Результаты оценки параметров линейного тренда в MS Excel

Регрессионная статистика						
Множественный R	0,207					
R^2	0,043					
Нормированный R^2	0,037					
Стандартная ошибка	5,622					
Наблюдения	167					

Дисперсионный анализ						
	df	SS	MS	F	Значимость F	
Регрессия	1	232,571	232,571	7,358	0,007	
Остаток	165	5215,050	31,606			
Итого	166	5447,621				

	Коэффи- циенты	Стандартн ая ошибка	t - статисти ка	P - значени е	Нижни е 95%	Верхни е 95%
Константа	103,423	0,874	118,333	0,000	101,698	105,149
Переменн ая t	0,024	0,009	2,713	0,007	0,007	0,042

Прогнозирование на тестовую выборку 01.2008–02.2010 гг. осуществлялось суммированием трендовых уровней и средних значений сезонных факторов для соответствующих месяцев:

$$\hat{y}_i(h) = a + b \left[\frac{h-1}{12} \right] + \bar{S}_i, \quad (13.3)$$

где $\bar{S}_i$ — средний показатель сезонности для соответствующего месяца, $i = \overline{1,12}$.

13.3 Экспоненциальное сглаживание

Программа STATISTICA 8.0 имеет возможность автоматического поиска необходимых параметров экспоненциального сглаживания в зависимости от выбранного критерия (минимизации внутривыборочных показателей MAE, PMSE или MAPE). Предварительно необходимо лишь указать, предполагается ли наличие во временном ряду тренда и (или) сезонности, а также тип и период сезонных колебаний при их наличии.

Для прогнозирования индекса реальных инвестиций при отсутствии тенденции выбрана аддитивная сезонность с периодом $\tau = 12$ месяцев:

$$\hat{y}_t(h) = \hat{a}_t + \hat{g}_{t-\tau+h}, \quad (13.4)$$

где $\hat{y}_t(h)$ - текущий уровень ряда после элиминирования сезонных колебаний

$$\hat{a}_t = \alpha(Y_t - \hat{g}_{t-\tau}) + (1-\alpha)\hat{a}_{t-1}; \quad (13.5)$$

аддитивная сезонная составляющая

$$\hat{g}_t = \delta(Y_t - \hat{a}_t) + (1-\delta)\hat{g}_{t-\tau}. \quad (13.6)$$

Так как минимизация показателей точности при автоматической оценке происходит на данных основной выборки (а целью нашего исследования является получение качественных *вневыборочных* прогнозов), то экспоненциальное сглаживание к исходному временному ряду применялось трижды (таблица 13.3). В результате получено три пары параметров сглаживания $\alpha; \delta$.

Как видно из таблицы 13.3, минимизация MAPE не дала ожидаемых результатов, и минимальная средняя абсолютная процентная ошибка соответствует способу сглаживания при минимизации средней абсолютной ошибки. Однако данные прогнозы являются внутривыборочными. В сравнении вневыборочных прогнозов будут участвовать все три способа экспоненциального сглаживания.

Таблица 13.3 — Оценка точности внутривыборочного прогноза индекса реальных инвестиций методом экспоненциального сглаживания

Внутривыборочные характеристики	Критерий выбора параметров сглаживания		
	MAE → min	PMSE → min	MAPE → min
Параметры сглаживания $\alpha; \delta$	(0,020; 0,00)	(0,016; 0,00)	(0,012; 0,00)
Средняя ошибка, п. п.	0,369	0,383	0,389
Средняя абсолютная ошибка, п. п.	4,454	4,463	4,472
Сумма квадратов отклонений, (п. п.) ²	6724,313	6722,676	6724,989
Средняя квадратическая ошибка, (п. п.)²	40,265	40,256	40,269
Средняя процентная ошибка, %	0,052	0,067	0,072
Средняя абсолютная процентная ошибка, %	4,436	4,438	4,438

13.4 Модели авторегрессии

Получив прогнозы традиционными способами, обратимся к альтернативным методам с учётом рекомендаций, изложенных ранее, в том числе изучение коррелограммы временного ряда Y (рисунок 13.2).

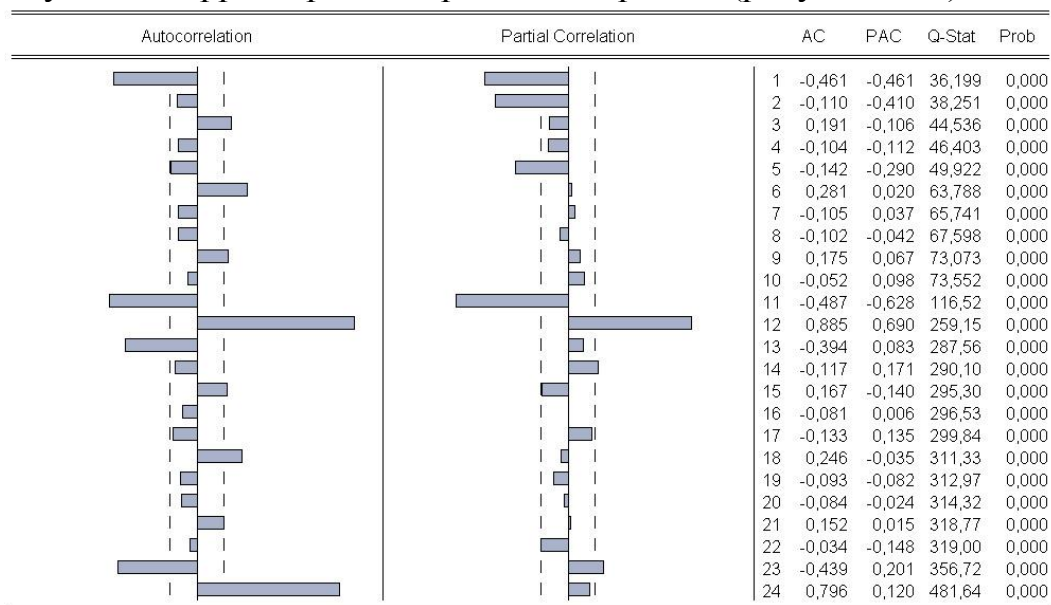


Рисунок 13.2 — Коррелограмма динамики индекса реальных инвестиций в основной капитал (02.1994–12.2007 гг.)

Поведение автокорреляционной функции подтверждает отсутствие тенденции (рисунок 13.2). Также видны значительные выбросы на лаге в 12 и 24 месяца. Коэффициент частной автокорреляции, нивелирующий влияние уровней ряда с промежуточными лагами, не подтверждает существенную зависимость текущих значений от значений двухлетней давности ($PAC_{24}=0,120$).

При подборе сезонной ARMA-модели в EViews 6 оценивались уравнения с различной спецификацией. Предельные значения порядков авторегрессии и скользящего среднего определены по значениям коэффициентов автокорреляции и частной автокорреляции (рисунок 3.2). Величина p и q тестировалась в следующих шести комбинациях: (1,0), (1,1), (2,0), (2,1), (2*,0), (2*,1), где последние две имеют первый параметр авторегрессии, равный нулю. Каждому из указанных случаев мог соответствовать один из двух вариантов спецификации сезонности: (1,0) или (1,1). Таким образом, всего оценивалось 12 моделей, из которых выбирались те, которые имеют значимые коэффициенты при сохранении стационарности модели (коэффициенты меньше единицы). Модели, удовлетворяющие данным условиям, представлены в таблице 13.4.

Таблица 13.4 — Сезонные ARMA-модели динамики индекса реальных инвестиций с 02.1994 г. по 12.2007 г.

Модель $SARMA(p,q)(P,Q)_{12}$	Число уровней	Число параметров	Стандартная ошибка	AIC	SIC	Статистика Дарбина- Уотсона
(1,1)(1,1)	154	4	5,197	6,166	6,265	1,672
(2,1)(1,0)	153	4	5,985	6,449	6,548	1,994
(2,0)(1,0)	153	3	6,072	6,471	6,550	1,873
(2*,0)(1,0)	153	2	6,154	6,492	6,551	2,227

Из таблицы 13.4 видно, что наилучшие внутривыборочные показатели имеет модель $SARMA(1,1)(1,1)$.

Число наблюдений в уравнениях с учётом наличия сезонных эффектов сократилось и составило 153–154 уровня. При уровне значимости 0,05 и 150 уровнях критические значения статистики Дарбина-Уотсона равны:

$$DW_L(2;150)=1,706, DW_U(2;150)=1,760;$$

$$DW_L(3;150)=1,693, DW_U(3;150)=1,774;$$

$$DW_L(4;150)=1,679, DW_U(4;150)=1,789.$$

В таблице 13.4 выделены фактические значения статистики, при которых подтверждается гипотеза об отсутствии автокорреляции первого порядка ($DW_{факт.} > DW_U$). Если $DW_{факт.} > 2$, то рассчитывается величина

$$DW'_{факт.} = 4 - DW_{факт.}, \quad (13.7)$$

которая затем сравнивается с табличными значениями.

Применение других методов тестирования автокорреляции даёт возможность эффективно обойти ограничения критерия Дарбина-Уотсона, в том числе его неэффективность при наличии лаговых значений переменной в качестве регрессоров. Таким образом, остатки проверялись на независимость с помощью Q-статистики, позволяющей тестировать наличие автокорреляции высоких порядков (рисунок 13.3).

Из рисунка 13.3а видно, что в остатках модели $SARMA(1,1)(1,1)_{12}$ только три первых коэффициента автокорреляции признаются статистически незначимыми с вероятностью 95,0%. Значимая серийная корреляция до лага 12 месяцев присутствует в остатках моделей $SARMA(2,0)(1,0)_{12}$ и $SARMA(2^*,0)(1,0)_{12}$ (рисунок 13.3в и 13.3г). Лишь модель $SARMA(2,1)(1,0)_{12}$ имеет статистически незначимые коэффициенты автокорреляции остатков вплоть до лага 12 (рисунок 13.3б). Особенно важным является отсутствие автокорреляции остатков с лагом 12 в условиях построения сезонной модели.

Таким образом, не смотря на то, что модель $SARMA(1,1)(1,1)_{12}$ имеет наименьшую ошибку, её остатки автокоррелированы. В том числе присутствует сравнительно сильная автокорреляция с лагом в 12 месяцев, то есть модели не удалось полностью описать сезонную колеблемость.

Из автокорреляции остатков вытекает, что коэффициенты регрессии не смещены, но стандартные ошибки недооценены. Следовательно, проверка

статистической значимости коэффициентов ненадёжна. Поэтому в дальнейшем анализе будем использовать модель SARMA(2,1)(1,0)₁₂.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0,131	0,131	2,6911	0,101
		2 -0,037	-0,055	2,9088	0,234
		3 0,031	0,044	3,0591	0,383
		4 -0,254	-0,272	13,401	0,009
		5 -0,165	-0,094	17,802	0,003
		6 0,003	0,007	17,804	0,007
		7 0,002	0,015	17,805	0,013
		8 -0,144	-0,218	21,211	0,007
		9 0,068	0,053	21,976	0,009
		10 -0,070	-0,140	22,804	0,011
		11 0,056	0,140	23,340	0,016
		12 0,359	0,264	45,114	0,000

а) SARMA(1,1)(1,1)₁₂

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0,014	-0,014	0,0322	0,857
		2 0,049	0,049	0,4071	0,816
		3 0,016	0,018	0,4490	0,930
		4 -0,164	-0,166	4,7141	0,318
		5 -0,105	-0,115	6,4857	0,262
		6 0,092	0,109	7,8449	0,250
		7 -0,007	0,017	7,8519	0,346
		8 0,045	0,008	8,1784	0,416
		9 0,121	0,086	10,609	0,303
		10 -0,090	-0,072	11,941	0,289
		11 -0,067	-0,068	12,693	0,314
		12 -0,207	-0,216	19,891	0,069

б) SARMA(2,1)(1,0)₁₂

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0,194	-0,194	6,3924	0,011
		2 -0,039	-0,079	6,6499	0,036
		3 0,123	0,104	9,2490	0,026
		4 -0,029	0,014	9,3953	0,052
		5 -0,041	-0,035	9,6880	0,085
		6 0,139	0,116	13,055	0,042
		7 0,081	0,139	14,226	0,047
		8 -0,042	0,021	14,546	0,069
		9 -0,073	-0,108	15,492	0,078
		10 0,174	0,127	20,930	0,022
		11 -0,158	-0,100	25,469	0,008
		12 -0,122	-0,181	28,173	0,005

в) SARMA(2,0)(1,0)₁₂

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0,223	-0,223	8,4662	0,004
		2 -0,031	-0,085	8,6265	0,013
		3 0,172	0,154	13,692	0,003
		4 -0,031	0,045	13,854	0,008
		5 -0,022	-0,007	13,935	0,016
		6 0,165	0,142	18,725	0,005
		7 0,086	0,168	20,018	0,006
		8 -0,045	0,031	20,379	0,009
		9 -0,067	-0,130	21,190	0,012
		10 0,185	0,112	27,324	0,002
		11 -0,152	-0,093	31,528	0,001
		12 -0,087	-0,163	32,895	0,001

г) SARMA(2*,0)(1,0)₁₂

Рисунок 13.3 — Коррелограммы остатков SARMA-моделей индекса реальных инвестиций в основной капитал

Наличие тенденции в остатках модели проверялось с помощью ADF-теста. Из таблицы 13.5 видно, что расчётное значение t -статистики (-6,537) меньше критических значений, поэтому нулевая гипотеза о наличии единичного корня отвергается в пользу противоположной односторонней гипотезы о его отсутствии.

Таблица 13.5 — Результаты ADF-теста остатков модели SARMA(2,1)(1,0)₁₂

Null Hypothesis: RESID2110 has a unit root

Exogenous: Constant

Lag Length: 2 (Automatic based on SIC, MAXLAG=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-6,537048	0,0000
Test critical values: 1% level	-3,474265	
5% level	-2,880722	
10% level	-2,577077	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(RESID2110)

Method: Least Squares

Sample (adjusted): 1995M07 2007M12

Included observations: 150 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESID2110(-1)	-0,934804	0,143001	-6,537048	0,0000
D(RESID2110(-1))	-0,068097	0,117919	-0,577491	0,5645
D(RESID2110(-2))	-0,017372	0,083274	-0,208614	0,8350
C	0,017182	0,484949	0,035431	0,9718

R-squared	0,495443	Mean dependent var	0,049418
Adjusted R-squared	0,485075	S.D. dependent var	8,274544
S.E. of regression	5,937668	Akaike info criterion	6,426815
Sum squared resid	5147,361	Schwarz criterion	6,507098
Log likelihood	-478,0111	Hannan-Quinn criter.	6,459431
F-statistic	47,78759	Durbin-Watson stat	1,963511
Prob(F-statistic)	0,000000		

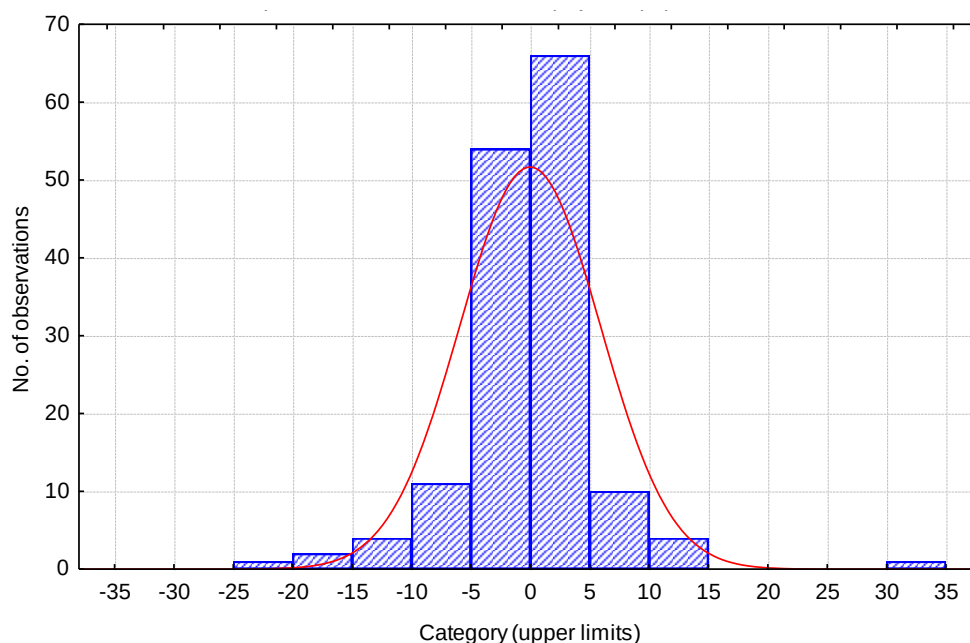


Рисунок 13.4 — Гистограмма остатков модели SARMA(2,1)(1,0)₁₂ и кривая нормального распределения, имеющего такое же среднее и дисперсию, как и остатки

Нормальность распределения остатков проверяется с помощью соответствующих тестов и визуально (рисунок 13.4). Фактическая и теоретическая плотности распределения достаточно хорошо согласуются друг с другом. Критерий согласия Колмогорова-Смирнова и его модификация — тест Лиллифорса — с вероятностью 95% подтверждают, что распределение остатков подчиняется нормальному закону.

Данная модель оценивалась в EViews 6 (таблица 13.6), при этом оценки коэффициентов авторегрессии и скользящего среднего определяются с учётом следующей спецификации:

$$Y_t = c + u_t, \quad (13.8)$$

где оцениваемая структура ошибок имеет вид:

$$(-\alpha_1 B - \alpha_2 B^2) \cdot (-\theta B^{12}) \cdot u_t = (+\beta B) \cdot \varepsilon_t. \quad (13.9)$$

Выразив u_t из уравнения (8) и подставив в (9), получаем модель:

$$Y_t = c \cdot (-\alpha_1 - \alpha_2 - \theta + \alpha_1 \theta + \alpha_2 \theta) \cdot \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \theta Y_{t-12} - \alpha_1 \theta Y_{t-13} - \alpha_2 \theta Y_{t-14} + \beta \varepsilon_{t-1} + \varepsilon_t, \quad (3.10)$$

то есть модель ARMA(14,1) с нелинейными ограничениями, наложенными на коэффициенты. Подставляя числовые значения из таблицы 13.6, получим:

$$Y_t = 5,222 - 0,689Y_{t-1} - 0,305Y_{t-2} + 0,977Y_{t-12} + 0,673Y_{t-13} + 0,298Y_{t-14} + 0,592\varepsilon_{t-1} + \varepsilon_t. \quad (13.11)$$

Таблица 13.6 — Оценка модели SARMA(2,1)(1,0)12 в EViews 6

Dependent Variable: Y
Method: Least Squares
Sample (adjusted): 1995M04 2007M12
Included observations: 153 after adjustments
Convergence achieved after 13 iterations
MA Backcast: 1995M03

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	115,1532	18,77809	6,132316	0,0000
AR(1)	-0,688634	0,165533	-4,160106	0,0001
AR(2)	-0,305011	0,073769	-4,134687	0,0001
SAR(12)	0,977255	0,019256	50,75123	0,0000
MA(1)	0,591890	0,165243	3,581940	0,0005
R-squared	0,950509	Mean dependent var	105,9487	
Adjusted R-squared	0,949171	S.D. dependent var	26,54701	
S.E. of regression	5,985089	Akaike info criterion	6,448553	
Sum squared resid	5301,551	Schwarz criterion	6,547587	
Log likelihood	-488,3143	Hannan-Quinn criter.	6,488782	
F-statistic	710,6075	Durbin-Watson stat	1,993603	
Prob(F-statistic)	0,000000			
Inverted AR Roots	1.00	.86- .50i	.86+ .50i	.50+ .86i
	.50- .86i	.00+1.00i	-.00-1.00i	-.34- .43i
	-.34+ .43i	-.50+ .86i	-.50- .86i	-.86- .50i
	-.86+ .50i	-1.00		
Inverted MA Roots	-.59			

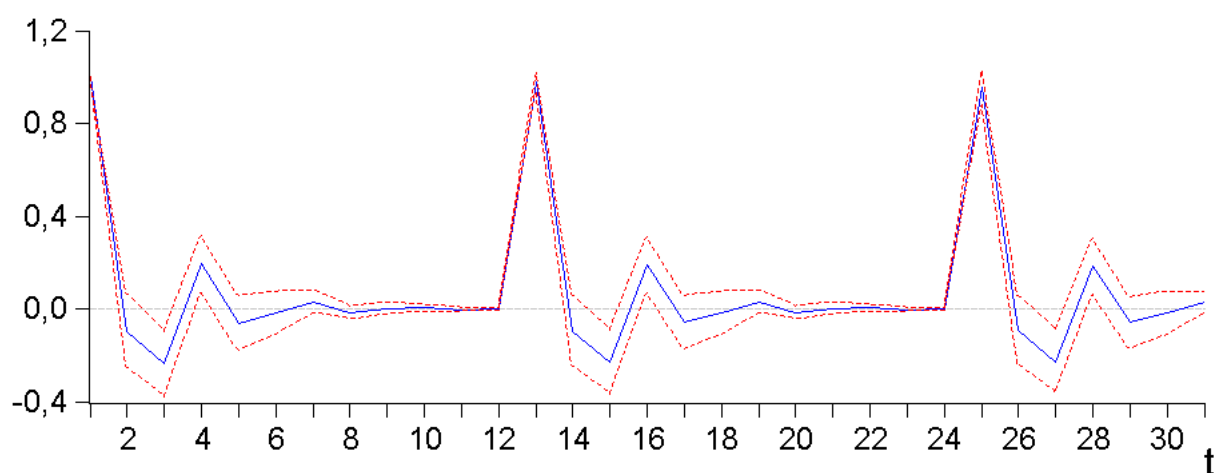


Рисунок 13.5 — Реакция модели SARMA(2,1)(1,0)12 на единичный импульс с доверительными границами ± 2 стандартных отклонения

Из рисунка 13.5 видно, что единичный импульс на первом шаге практически полностью воспроизводится данной моделью через 12 месяцев ($t=13$), а также через два года ($t=25$). Это объясняется тем, что коэффициент θ модели (13.9) близок к единице.

Прогнозирование по модели (13.10) осуществлялось в соответствии со стандартным алгоритмом прогнозирования ARMA-моделей.

13.5 Многофакторная модель

Для спецификации многофакторных модели рассмотрим графики исследуемых показателей. Из рисунка 13.6 видно, что многие временные ряды факторных показателей являются нестационарными (предположительно X_2 , X_3 , X_4 , X_5 , X_6), содержат ярко выраженные сезонные колебания.

С целью формализованной оценки наличия сезонности, к каждому временному ряду применена процедура сезонной корректировки «X-11» в STATISTICA 8.0. Дисперсионный анализ (результат D8 процедуры) показал, что на уровне 0,05 нулевая гипотеза об отсутствии сезонности во временных рядах X_5 и X_6 не отвергается. Динамика других показателей имеет статистически значимые сезонные колебания, для нивелирования которых проведена замена на скорректированные на сезонность уровни (результат D11 процедуры «X-11»).

Поиск единичных корней во временных рядах новой системы показателей (десезонализированной) осуществлён с помощью ADF-теста. На основе визуального изучения графиков тестовое уравнение специфицировано с включением константы, без использования детерминированного линейного тренда. Тестирование проводилось для начальных уровней, первых и вторых разностей до получения разностно-стационарного ряда. По результатам тестирования все исследуемые временные ряды являются стационарными, кроме $X_{2(sa)}$, $X_{3(sa)}$, $X_{4(sa)}$, которые имеют единичные корни (таблица 13.7).

Таблица 13.7 — Результаты ADF-теста для временных рядов $X_{2(sa)}$, $X_{3(sa)}$, $X_{4(sa)}$

Параметр	Тестируемые значения		
	$D(X_{2(sa)})$	$D(X_{3(sa)})$	$D(X_{4(sa)})$
Экзогенные переменные в тестовом уравнении	константа	константа	константа
Длина лага, автоматически установленная по SIC	2	3	2
Значение t -статистики	-3,812	-3,701	-4,486
Вероятность (p-значение)	0,0041	0,0049	0,0003

Из таблицы 13.7 следует, что представленные временные ряды с вероятностью 95% являются интегрированными первого порядка, поэтому мы переходим к их первым разностям $D(X_{2(sa)})$, $D(X_{3(sa)})$, $D(X_{4(sa)})$.

Таким образом, определены временные ряды, содержащие и стохастический тренд, и сезонную составляющую (X_2 , X_3 , X_4); сезонную составляющую без тренда (Y , X_1 , X_7 , X_8 , X_9); не имеющие систематических изменений (X_5 , X_6). Графики новых показателей, из которых исключена сезонность и тенденция, представлены на рисунке 13.7.

С целью спецификации многомерной модели оценена матрица коэффициентов парной корреляции стационарных временных рядов и разностей соответствующих порядков разностно-стационарных временных рядов (таблица 13.8).

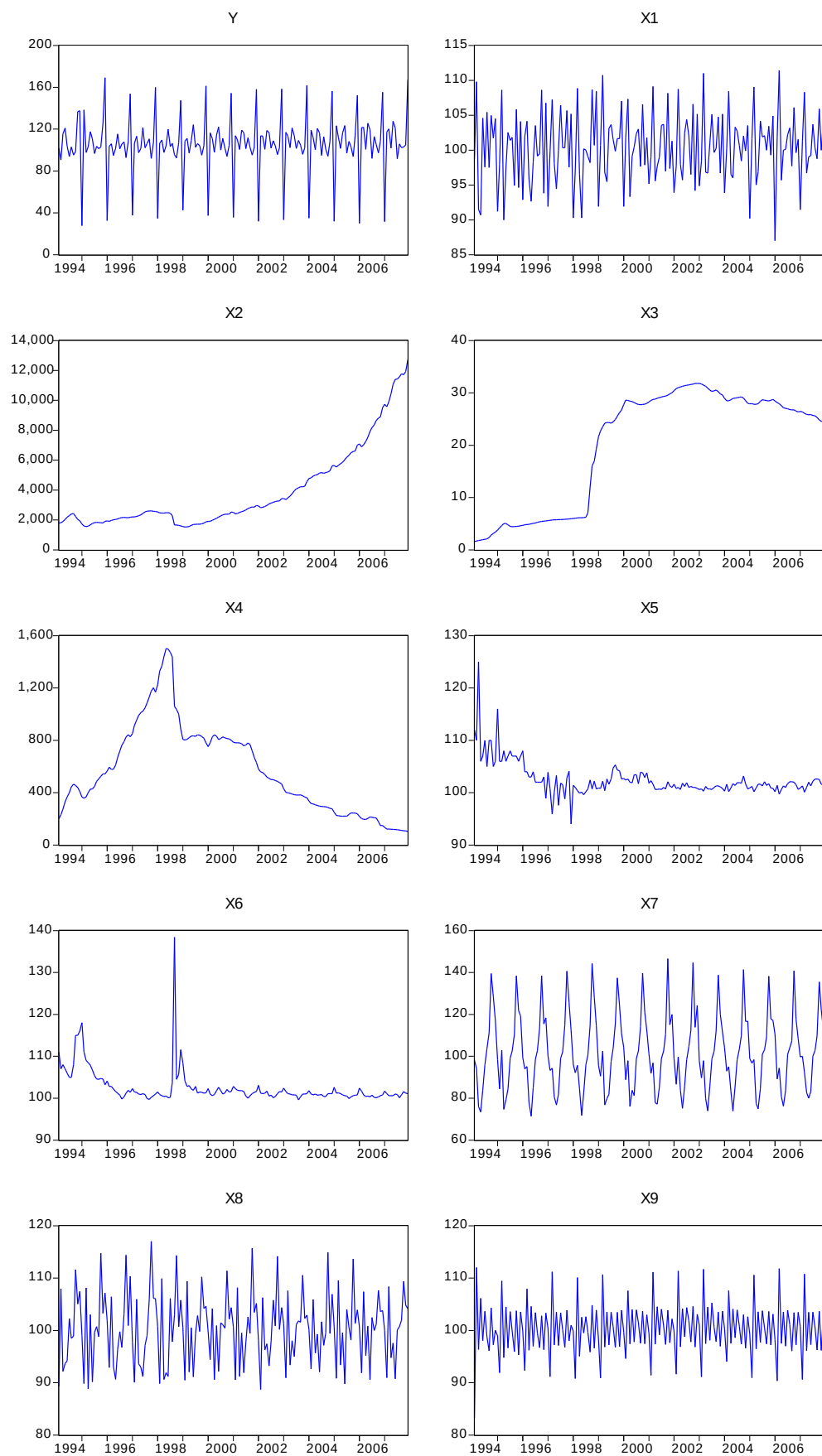


Рисунок 13.6 — Динамика изучаемых показателей с 02.1994 г. по 12.2007 г.

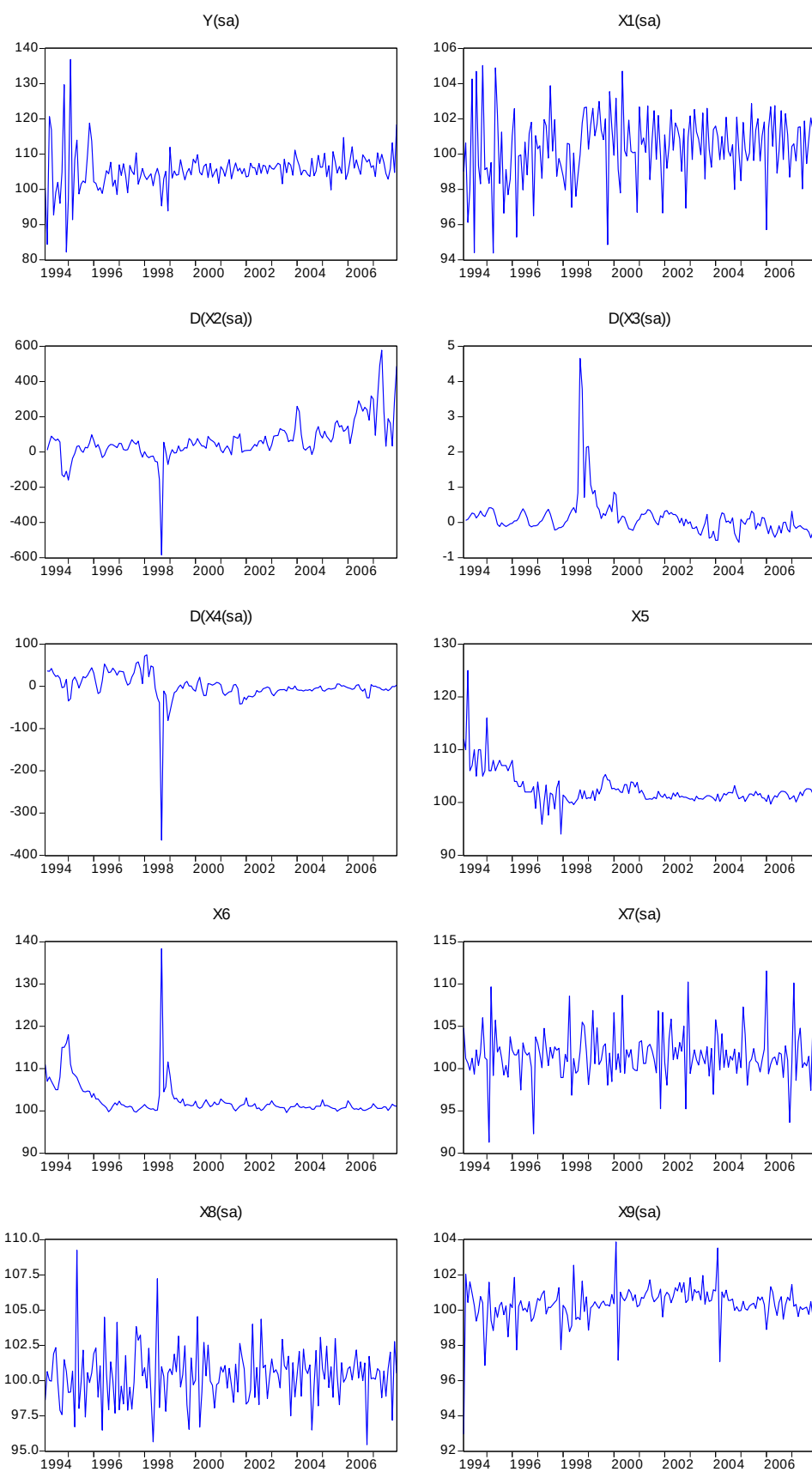


Рисунок 13.7 — Графики временных рядов, из которых исключена тенденция и сезонность (при их наличии в исходных уровнях)

Таблица 13.8 — Матрица парных коэффициентов корреляции (выделены коэффициенты, статистически значимые на уровне 5%)

	$Y_{(sa)}$	$X_{1(sa)}$	$D(X_{2(sa)})$	$D(X_{3(sa)})$	$D(X_{4(sa)})$	X_5	X_6	$X_{7(sa)}$	$X_{8(sa)}$	$X_{9(sa)}$
$Y_{(sa)}$	1,000	0,092	0,233	-0,158	0,053	— 0,038	— 0,137	— 0,149	0,003	0,198
$X_{1(sa)}$	0,092	1,000	0,102	0,050	-0,108	— 0,227	— 0,065	0,303	0,234	0,308
$D(X_{2(sa)})$	0,133	0,102	1,000	-0,509	0,317	— 0,173	— 0,578	— 0,000	0,040	0,108
$D(X_{3(sa)})$	— 0,158	0,050	-0,509	1,000	-0,614	0,006	0,602	0,068	— 0,005	— 0,018
$D(X_{4(sa)})$	0,053	— 0,108	0,317	-0,614	1,000	0,139	— 0,562	— 0,028	— 0,024	— 0,021
X_5	— 0,038	— 0,227	-0,173	0,006	0,139	1,000	0,462	— 0,040	— 0,070	— 0,074
X_6	— 0,137	— 0,065	-0,578	0,602	-0,562	0,462	1,000	0,014	— 0,029	— 0,192
$X_{7(sa)}$	— 0,149	0,303	-0,000	0,068	-0,028	— 0,040	0,014	1,000	0,291	— 0,013
$X_{8(sa)}$	0,003	0,234	0,040	-0,005	-0,024	— 0,070	— 0,029	0,291	1,000	0,186
$X_{9(sa)}$	0,198	0,308	0,108	-0,018	-0,021	— 0,074	— 0,192	— 0,013	0,186	1,000

Из таблицы 13.8 следует, что с индексом реальных инвестиций $Y_{(sa)}$ имеют значимую корреляционную связь индекс промышленного производства в нефтедобывающей промышленности $X_{9(sa)}$ и первые разности среднемесячного официального курса доллара $D(X_{3(sa)})$.

Значимость линейного коэффициента корреляции проверяется на основе t -критерия Стьюдента

$$t_r = \frac{r \cdot \sqrt{N-2}}{\sqrt{1-r^2}}, \quad (13.12)$$

который сравнивается с табличным значением при $\alpha=0,05$ и числе степеней свободы, меньшим числа уровней временного ряда на две единицы ($N-2$). Корректируя значение N , определим значимые коэффициенты корреляции между $Y_{(sa)}$ и лаговыми значениями факторных признаков, варьируя продолжительность лага от одного месяца до года (таблица 13.9).

Таблица 13.9 — Коэффициенты корреляции между текущими значениями $Y_{(sa)}$ и лаговыми значениями факторных признаков, статистически значимые на уровне 5% (в скобках указана величина лага l)

	$X_{1(sa),t-l}$	$D(X_{2(sa)},t-l)$	$D(X_{3(sa)},t-l)$	$D(X_{4(sa)},t-l)$	$X_{5,t-l}$	$X_{6,t-l}$	$X_{7(sa),t-l}$	$X_{8(sa),t-l}$	$X_{9(sa),t-l}$
		l	$-l$	$-l$		l	$-l$	$-l$	
$Y_{(sa)}$ t	-0,169 (1)				0,259 (1)				
	0,221 (2)	0,166 (5)			-0,168				
	-0,190 (3)	0,184 (7)			(3)				
	0,269 (5)	0,178 (8)	-0,158	-0,194	0,215 (7)				0,198(0
	0,196 (8)	0,165 (9)	(0)	(4)	-0,166	-	-	-	)
	-0,228	0,184			(9)				
	(10)	(12)			0,253				
	0,199 (11)				(10)				

Из таблицы 13.9 видно, что показатели X_3 , X_6 , X_7 , X_8 и X_9 не имеют распределённой во времени статистически значимой корреляционной связи с Y . Показатели X_1 , X_2 и X_5 , напротив, имеют продолжительную, практически не прерывающуюся во времени статистически значимую связь. Однако остаётся неясным, почему направление корреляционной связи, например, между $Y_{(sa)}$ и лаговыми значениями X_5 меняется на противоположное при переходе от одного лага к другому.

Таблица 13.10 — Результаты поиска причинности К. Грэнджера между результативной и факторными переменными

Нулевая гипотеза	Максимальный лаг	F-статистика	Вероятность (p -значение)
$X_{1(sa)}$ не является Грэнджер-причиной $Y_{(sa)}$	11	2,719	0,003
$Y_{(sa)}$ не является Грэнджер-причиной $X_{1(sa)}$		0,892	0,550
$D(X_{2(sa)})$ не является Грэнджер-причиной $Y_{(sa)}$	12	1,977	0,031
$Y_{(sa)}$ не является Грэнджер-причиной $D(X_{2(sa)})$		0,416	0,955
$D(X_{3(sa)})$ не является Грэнджер-причиной $Y_{(sa)}$	2	4,287	0,015
$Y_{(sa)}$ не является Грэнджер-причиной $D(X_{3(sa)})$		0,671	0,512
$D(X_{4(sa)})$ не является Грэнджер-причиной $Y_{(sa)}$	4	1,979	0,101
$Y_{(sa)}$ не является Грэнджер-причиной $D(X_{4(sa)})$		0,373	0,827

X_5 не является Грэнджер-причиной $Y_{(sa)}$	10	3,164	0,001
$Y_{(sa)}$ не является Грэнджер-причиной X_5		2,458	0,010
$X_{9(sa)}$ не является Грэнджер-причиной $Y_{(sa)}$	2	0,754	0,472
$Y_{(sa)}$ не является Грэнджер-причиной $X_{9(sa)}$		0,545	0,581

Для построения многомерной прогностической модели, основанной на причинных взаимосвязях, применима концепция причинности К. Грэнджера. При спецификации тестового уравнения (11.23) использовался лаг, соответствующий максимальному лагу l , указанному для каждой переменной в таблице 13.9. Наличие причинной взаимосвязи с показателями, имеющими лишь одновременную корреляцию с результативной переменной ($D(X_{3(sa)})$ и $X_{9(sa)}$), проверялось на наименьшем технически возможном лаге, равном двум. Результаты тестирования представлены в таблице 13.10.

Из таблицы 3.10 следует, что с вероятностью 95% между временными рядами $X_{1(sa)}$ и $Y_{(sa)}$; $D(X_{2(sa)})$ и $Y_{(sa)}$; $D(X_{3(sa)})$ и $Y_{(sa)}$ существует односторонняя причинная связь по направлению к $Y_{(sa)}$, а между X_5 и $Y_{(sa)}$ присутствует двусторонняя причинная зависимость. Наличие двусторонней связи с X_5 , частично объясняет знакопеременное поведение кросс-корреляционной функции (таблица 13.9).

Таблица 13.11 — Результаты поиска причинности К. Грэнджера между X_5 и другими факторными переменными, максимальный лаг — 12 месяцев

Нулевая гипотеза	F-статистика	Вероятность (p-значение)
$X_{1(sa)}$ не является Грэнджер-причиной X_5	1,231	0,294
X_5 не является Грэнджер-причиной $X_{1(sa)}$	1,583	0,156
$D(X_{2(sa)})$ не является Грэнджер-причиной X_5	0,548	0,771
X_5 не является Грэнджер-причиной $D(X_{2(sa)})$	0,741	0,618
$D(X_{3(sa)})$ не является Грэнджер-причиной X_5	0,375	0,894
X_5 не является Грэнджер-причиной $D(X_{3(sa)})$	0,334	0,918
$D(X_{4(sa)})$ не является Грэнджер-причиной X_5	0,668	0,676
X_5 не является Грэнджер-причиной $D(X_{4(sa)})$	0,489	0,816
X_6 не является Грэнджер-причиной X_5	0,896	0,500
X_5 не является Грэнджер-причиной X_6	1,181	0,320
$X_{7(sa)}$ не является Грэнджер-причиной X_5	1,931	0,079
X_5 не является Грэнджер-причиной $X_{7(sa)}$	1,084	0,375
$X_{8(sa)}$ не является Грэнджер-причиной X_5	0,804	0,568
X_5 не является Грэнджер-причиной $X_{8(sa)}$	1,427	0,208
$X_{9(sa)}$ не является Грэнджер-причиной X_5	1,430	0,207
X_5 не является Грэнджер-причиной $X_{9(sa)}$	1,359	0,235

Отбор факторов для уравнения обратной связи проведён аналогично. Содержание таблицы 13.11 свидетельствует о том, что включение в уравнение для X_5 дополнительных регрессоров, представленных факторными переменными, не обосновано, так как имеющиеся временные ряды не подтверждают наличия причинных взаимосвязей (прямых или обратных) на уровне значимости 0,05.

В процессе спецификации проведено объединение лаговых значений экзогенных переменных в отдельные факторы:

– среднемесячный процент роста реального объёма промышленного производства в предшествующие три месяца сформировал фактор F_1 :

$$F_{1,t} = \sqrt[3]{X_{1(sa),t-1} \cdot X_{1(sa),t-2} \cdot X_{1(sa),t-3}}, \quad (13.13)$$

то есть объединил начальные лаговые значения $X_{1(sa)}$, имеющие значимую связь с результатом (таблица 13.9). Лаги более высоких порядков для данного показателя не использовались, так как ранее проводимые исследования показывают, что реакция инвестиций на повышение выпуска достаточно короткая и быстрая [4, с. 296]. Так как фактор X_1 является индексом, то (13.13) имеет форму средней геометрической с целью сохранить экономическую сущность и интерпретируемость результатов;

– лаговые приросты объёмов денежной массы образовали фактор F_2 :

$$F_{2,t} = D\Delta_{2(sa)}^5 + D\Delta_{2(sa)}^7 + D\Delta_{2(sa)}^8 + D\Delta_{2(sa)}^9 + D\Delta_{2(sa)}^{12}. \quad (13.14)$$

В результате получена модель:

$$\hat{y}_{(sa)t} = c_1 F_{1,t} + c_2 F_{2,t} + c_3 D\Delta_{3(sa)}^9 + c_4 X_{5,t-1}. \quad (13.15)$$

Из таблицы 13.12 видно, что все коэффициенты парной линейной связи между признаками не превышают 0,4. Таким образом, объединение лаговых значений в общие факторы позволило предвосхитить появление мультиколлинеарности.

Результаты оценки модели (13.15) представлены в таблице 13.13. Значения вероятностей в последнем столбце таблицы свидетельствуют о статистической значимости коэффициентов на уровне 0,05.

Таблица 13.12 — Матрица парных коэффициентов корреляции факторных переменных модели (13.15)

	$F_{1,t}$	$F_{2,t}$	$D(X_{3(sa)})_t$	$X_{5,t-1}$
$F_{1,t}$	1,000	0,027	-0,110	-0,227
$F_{2,t}$	0,027	1,000	-0,344	-0,255
$D(X_{3(sa)})_t$	-0,110	-0,344	1,000	-0,024
$X_{5,t-1}$	-0,227	-0,255	-0,024	1,000

Таблица 13.13 — Оценка коэффициентов модели (13.15) в EViews 6

Dependent Variable: Y(sa)
Method: Least Squares
Sample (adjusted): 1995M03 2007M12
Included observations: 154 after adjustments
 $Y(sa)=C(1)*F1+C(2)*F2+C(3)*D(X3(sa))+C(4)*X5(-1)$

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0,749	0,125	5,985	0,000
C(2)	0,002	0,001	3,194	0,002
C(3)	-0,965	0,515	-2,043	0,043
C(4)	0,292	0,123	2,379	0,019
R-squared	0,093	Mean dependent var		105,534
Adjusted R-squared	0,074	S.D. dependent var		3,728
S.E. of regression	3,587	Akaike info criterion		5,418
Sum squared resid	1930,050	Schwarz criterion		5,497
Log likelihood	-413,199	Hannan-Quinn criter.		5,450
Durbin-Watson stat	1,862			

Модель второй эндогенной переменной (X_5) специфицировалась с учётом отсутствия других predetermined переменных из анализируемого массива, так как поиск причинных зависимостей между ними не дал результатов (таблица 13.11). Для определения подходящего лага между $Y_{(sa)}$ и X_5 оценивалась выборочная кросс-корреляционная функция (рисунок 13.8).

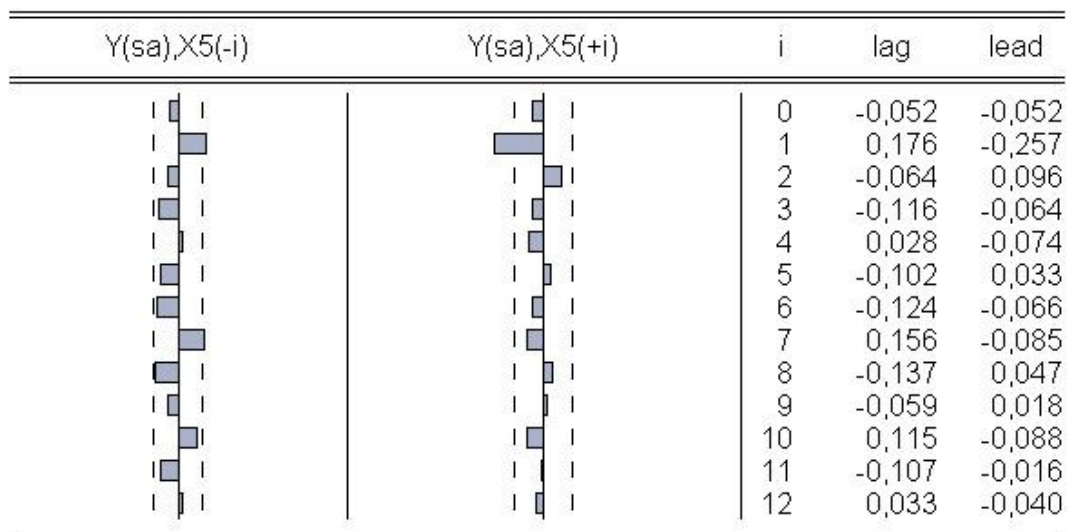


Рисунок 3.8 — График выборочной кросс-корреляционной функции

Из рисунка 13.8 видно, что влияние роста инвестиций ($Y_{(sa)}$) на стоимость строительно-монтажных работ (X_5) происходит с задержкой в один месяц. Коэффициент корреляции при данном лаге наибольший по абсолютной величине и равен $-0,257$. При этом анализ показал, что ошибки регрессии

$$X_{5,t} = a_1 + a_2 Y_{(sa)t-1} + u_t \quad (13.16)$$

имеют структуру ARMA(1,1):

$$\textcircled{(-} a_3 L \textcircled{)} \hat{u}_t = \textcircled{(+} a_4 L \textcircled{)} \hat{\varepsilon}_t. \quad (13.17)$$

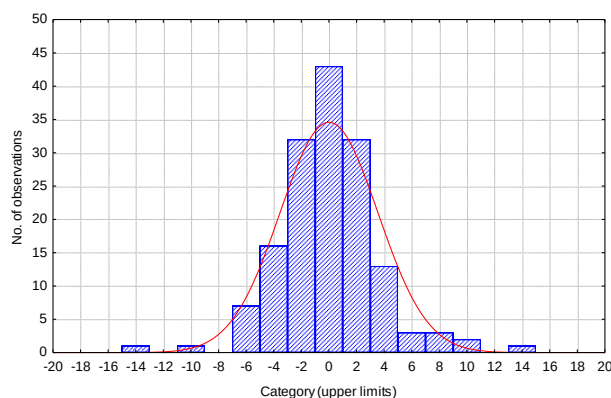
Из таблицы 13.14 с вероятностью 95% следует, что модель в целом является статистически значимой (F -критерий), как и каждый её коэффициент (t -тест). Подставив ошибку u_t из (13.16) в выражение (13.17), получим уравнение после выборочного прогноза по уровням временных рядов:

$$X_{5,t} = a_1(1 - a_3) + a_2 \textcircled{(-} a_3 \textcircled{)} Y_{(sa)t-1} - a_3 Y_{(sa)t-2} \textcircled{+} a_3 X_{5,t-1} + a_4 \varepsilon_{t-1} + \varepsilon_t. \quad (13.18)$$

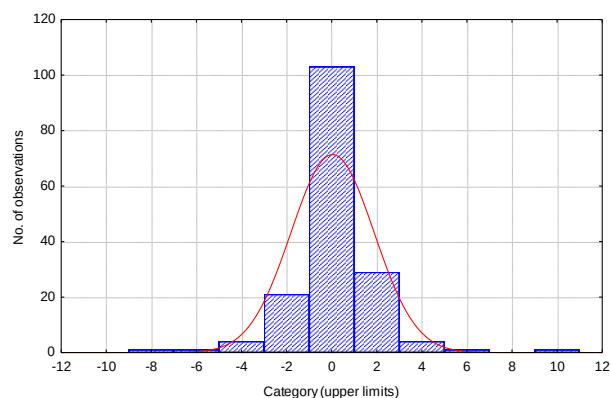
Результаты тестирования нормальности распределения остатков полученных моделей представлены на рисунке 13.9, откуда видно, что фактическое распределение остатков достаточно хорошо согласуется с нормальным распределением. Данный вывод подтверждается критерием Колмогорова-Смирнова и тестом Лиллифорса, которые указывают на низкую вероятность отличия фактического распределения от нормального (вероятность менее 0,05).

Таблица 13.14 — Результаты оценки модели (13.16) с авторегрессионной структурой ошибок

Dependent Variable: X5					
Method: Least Squares					
Sample (adjusted): 1994M04 2007M12					
Included observations: 165 after adjustments					
Convergence achieved after 12 iterations					
MA Backcast: 1994M03					
Variable	a_1 a_2 Coefficient	Std. Error	t-Statistic	Prob.	
C	a_3 114,398	2,642	43,292	0,000	
Y(sa)(-1)	a_4 -0,123	0,025	-5,004	0,000	
AR(1)	0,928	0,018	52,066	0,000	
MA(1)	-0,775	0,057	-13,492	0,000	
R-squared	0,650	Mean dependent var		102,398	
Adjusted R-squa...	0,643	S.D. dependent var		3,118	
S.E. of regression	1,862	Akaike info criterion		4,105	
Sum squared resid	558,211	Schwarz criterion		4,180	
Log likelihood	-334,675	Hannan-Quinn criter.		4,136	
F-statistic	99,647	Durbin-Watson stat		1,921	
Prob(F-statistic)	0,000				



а) модели (15)



б) модели (18)

Рисунок 13.9 — Гистограммы остатков регрессионных моделей и кривые нормального распределения, имеющего такое же среднее и дисперсию, как и остатки

Для получения прогноза индекса инвестиций по модели (13.15) необходимы будущие значения объясняющих переменных: факторов, $D(X_{3(sa)})$, и X_5 . Так как прогнозная модель сформулирована в терминах десеզоналированных переменных, и прогнозирование временных рядов в исходных уровнях требует проведения повторной сезонной корректировки (результаты которой могут отличаться от проведённой перед оцениванием параметров моделей), то прогнозы объясняющих переменных строятся на основе уровней, уже скорректированных на сезонность. Априорно не удалось установить, следует ли при этом прогнозировать непосредственно сами факторы F_1 , F_2 и разности $D(X_{3(sa)})$, или же сначала необходимо получить прогноз десеզоналированных уровней, а затем — рассчитать значения факторов и приростов для подстановки в уравнение модели. Поэтому прогнозирование экзогенных переменных проводилось двумя способами. Полученные прогнозы $Y_{(sa)}$ скорректированы на сезонность с помощью средних коэффициентов сезонности, аналогично модели (13.3).

13.6 Сравнение методов и объединение результатов

Результаты прогноза Y различными способами на период 01.2008–02.2010 гг. объединены в таблице 13.15.

Индекс инвестиций является неотрицательной величиной, поэтому показатель $MARE_M$ не использовался, так как его значения в данном случае идентичны значениям $MARE$.

Из таблицы 13.15 следует, что третий вариант экспоненциального сглаживания, параметры которого подбирались для минимизации $MARE$, имеет лучшие послевыборочные прогнозы по данному показателю (6,3%) и по размеру среднего квадрата ошибки (34,9 (п. п.)²) не только среди результатов экспоненциального сглаживания, но и среди всех оценивавшихся моделей. Также необходимо отметить, что по внутривыборочным оценкам

ошибок данный вариант сглаживания не только не превосходил, но даже уступал предыдущим (таблица 13.3). Сопоставление результатов сглаживания показывает занижение ошибок при внутривыборочном оценивании качества прогнозирования, что подтверждает необходимость оценки и верификации моделей на различных частях временного ряда.

Таблица 13.15 — Сравнительная оценка прогнозов индекса реальных инвестиций в основной капитал на тестовую выборку (01.2008–02.2010 гг.)¹³

Модель (метод)	Дополнительные условия	MAE, п. п.	PMSE, (п. п.) ²	MAPE, %	K
1. Наивная	—	6,3	45,3	8,1	x
2. Тренд-сезонная	—	5,3	39,0	7,2	0,847
3. Экспоненциальное сглаживание с различными параметрами адаптации (α ; δ)	(0,020; 0,00)	5,6	35,4	6,5	0,885
	(0,016; 0,00)	5,4	35,1	6,4	0,850
	(0,012; 0,00)	5,3	34,9	6,3	0,832
4. SARMA(2,1)(1,0) ₁₂	—	5,2	43,9	6,7	0,826
5. Многомерная	прогнозирование факторов	58,3	3442,1	73,0	9,237
	прогнозирование уровней рядов	6,2	51,2	8,8	0,983

Наименьшую среднюю абсолютную ошибку послевыборочного прогноза (5,2 п. п.) имеет модель SARMA(2,1)(1,0)₁₂ (таблица 13.15). В частности, это объясняется тем, что линейные ошибки данной модели имеют тщательно специфицированную структуру, и подробная проверка внутривыборочных ошибок на наличие автокорреляции, тенденции, нормальность распределения проходила для линейных данных (а, например, не для квадратов ошибок). Такой способ обработки временного ряда способствовал получению модели с наименьшими абсолютными ошибками.

Из таблицы 13.15 также видно, что сравнительный показатель точности прогноза K , отдаёт предпочтение модели с минимальными линейными ошибками, так как в его формуле используются модули отклонений. На самом деле, если в (12.42) модули заменить квадратами отклонений, то предпочтение будет отдано результатам экспоненциального сглаживания, имеющим минимальную среднюю квадратическую ошибку прогноза. То есть, мы вновь обращаемся к вопросу о виде функции потерь, равнозначности положительных и отрицательных отклонений, значительности тех или иных отклонений в понимании аналитика и заказчика (пользователя).

¹³Если показатель представляет собой разность двух величин, измеряемых в процентах, то он измеряется в процентных пунктах (п. п.). Индекс инвестиций измеряется в процентах, поэтому MAE измеряется в п. п., а PMSE — в процентных пунктах в квадрате ((п. п.)²).

Точность прогнозов по многомерной модели, основанной на причинных зависимостях между переменными, в данном случае уступает другим методам (таблица 13.15). Это объясняется тем, что в процессе спецификации не были найдены ведущие индикаторы, позволяющие получать прогнозы без необходимости прогнозирования объясняющих переменных, хотя бы при некотором ограниченном горизонте прогнозирования. При этом прогнозирование непосредственно факторов дало результаты значительно хуже (средняя абсолютная процентная ошибка составила 73,0%, а линейная аппроксимация оказалась в 9,2 раза хуже, чем у наивной модели), чем вариант с прогнозированием входящих в них показателей. Прогнозирование входящих в модель переменных в виде уровней временных рядов, ещё не объединённых в факторы, более обоснованно, так как динамика показателей при спецификации многомерной модели исследовалась подробнее динамики факторов. Следовательно, предварительное исследование временного ряда (графический анализ, изучение коррелограммы, определение наличия единичных корней, сезонности, выбросов, пропущенных значений и так далее) способствует получению улучшенных прогнозов даже при использовании автоматических методов прогнозирования (в данном случае — экспоненциального сглаживания).

Таким образом, точность многомерной модели признаётся высокой, а её прогностические свойства будут выглядеть ещё лучше, если рассмотреть экономические взаимосвязи, которые она учитывает.

Уравнения (13.15) и (13.16) представляют собой систему с обратной связью:

$$\begin{cases} Y_t = 0,749F_{1,t} + 0,002F_{2,t} - 0,965D_{3(sa)} + 0,292X_{5,t-1} + \bar{S}_i + \varepsilon_{1,t} \\ X_{5,t} = 8,237 - 0,123C_{(sa)t-1} - 0,928Y_{(sa)t-2} + 0,928X_{5,t-1} - 0,774\varepsilon_{2,t-1} + \varepsilon_{2,t} \end{cases} \quad (13.19)$$

Так как в правых частях уравнений системы (3.19) представлены эндогенные переменные только за предшествующие периоды времени, то их можно рассматривать как экзогенные [23, с. 252]. Таким образом, уравнения являются независимыми, и для их оценки применим обычный МНК. Действительно, из первого уравнения следует, что индекс инвестиций в период t зависит, в частности, от индекса цен на строительно-монтажные работы¹⁴ X_5 в предыдущий период $(t-1)$. Поэтому ошибка $\varepsilon_{1,t}$ формируется позже, чем значение $X_{5,t-1}$, и связь между ними отсутствует, что даёт несмещённые оценки при использовании обычного метода наименьших квадратов.

¹⁴ Строительно-монтажные работы учитываются в технологической структуре инвестиций в основной капитал. В 2007 г. затраты на строительно-монтажные работы в Российской Федерации составили 46,9% от общего объёма инвестиций в основной капитал; затраты на машины и оборудование, транспортные средства (без работ по монтажу оборудования) — 35,0%; на прочие капитальные работы и затраты — 18,1% [50]. В более поздние годы технологическая структура инвестиций не публиковалась.

Важным является тот факт, что система (13.19) построена не только на основе тщательного отбора признаков по их экономическому содержанию, но и основывается на причинных взаимосвязях, подтверждённых тестированием статистических данных. Выявленные взаимосвязи представляют собой:

- влияние среднемесячного процента роста реального объёма промышленного производства в предшествующие три месяца на индекс реальных инвестиций в текущем периоде. Экономически данное влияние является полностью обоснованным: когда повышается выпуск, немедленно возникает потребность в новом капитальном оборудовании. Данное оборудование приобретается за счёт имеющейся дополнительной нераспределённой прибыли, полученной от увеличения объёма производства. То есть, уже имеются свободные ликвидные средства, позволяющие осуществить покупку без задержек с получением кредита или реализацией свободных неликвидных активов. Причём установка нового, дополнительного оборудования не задерживается необходимостью демонтировать старое, чем объясняется достаточно короткий временной лаг. Из модели (13.19) следует, что увеличение среднемесячного роста промышленного производства на 1 п. п. приведёт к росту индекса инвестиций в основной капитал в среднем на 0,749 п. п.;

- влияние роста денежной массы страны на индекс реальных инвестиций в основной капитал. Фактор F_2 представляет собой сумму распределённых во времени приростов объёма денежной массы (лаги равны 5, 7, 8, 9, 12 месяцев), имеющих статистически значимую корреляционную связь с результатом текущего периода. Задержка во взаимосвязи здесь немного больше, чем у фактора F_1 . В результате можно сказать, что на январский индекс инвестиций оказывает влияние прирост денежной массы, наблюдавшийся в первой половине предыдущего года¹⁵. Данная связь также объясняется экономически: увеличение объёма доступных денежных средств позволяет направить их на инвестиционные вложения. Следует отметить, что увеличение денежного агрегата измерялось в реальном выражении, то есть была проведена корректировка на инфляцию. Из первого уравнения системы (13.19) следует, что рост фактора F_2 от его средней величины на 100 млрд. руб. вызовет рост индекса реальных инвестиций на 0,2 п. п. (также от средней величины);

- влияние текущего прироста стоимости доллара на снижение текущего индекса реальных инвестиций. Экономика России сильно зависит

¹⁵ Строго говоря, такая формулировка является обобщением, так как суммарный прирост в первой половине предыдущего года характеризуется месячными приростами с лагами в 7, 8, 9, 10, 11, 12 месяцев. Однако в нашем случае при использовании реальных экономических данных наблюдается некоторое временное смещение связи, в результате чего на отдельных лагах она оказалась статистически незначимой при заданном уровне 0,05. Тем не менее, это не нарушает экономической сущности представленного фактора F_2 .

от курса американской валюты, что подтверждается результатами анализа причинности. При увеличении курса доллара возрастает цена импорта, что увеличивает издержки производства на предприятиях, где он применяется. В России чаще всего невозможно заменить импорт отечественными товарами (оборудованием). Ввиду этого рост курса доллара приводит к росту уровня цен и сокращению инвестиционных расходов: увеличение фактора $D(X_{3(sa)})$ на 1 руб./долл. приведёт к сокращению индекса реальных инвестиций в основной капитал на 0,965 п. п. (как и в предыдущих случаях, изменения отмеряются от средних значений показателя). Данные выводы необходимо учитывать при определении политики Центрального банка в области валютного регулирования;

– взаимовлияние индекса реальных инвестиций в основной капитал и индекса цен на строительно-монтажные работы. Изучение поведения выборочной кросс-корреляционной функции (рисунок 13.8) показало различную направленность взаимосвязи: положительная корреляция возникает между $Y_{(sa)t}$ и $X_{5,t-1}$, а отрицательная — между $Y_{(sa)t}$ и $X_{5,t+1}$. Так как спрос рождает предложение, то рост цен (а цены растут при увеличении спроса) указывает на рост доходности; а чем выше доходность, тем выше привлекательность для инвесторов. Поэтому увеличение индекса цен на строительно-монтажные работы в предшествующем месяце на 1 п. п. обусловит рост индекса реальных инвестиций на 0,292 п. п., о чём говорит первое уравнение системы (13.19). Интерпретация второго уравнения сложнее из-за того, что на ошибки была наложена структура ARMA. Вследствие этого в его правой части появилась переменная $Y_{(sa)t-2}$. Таким образом, рост инвестиционных затрат стимулирует выход на рынок дополнительных участников, предлагающих свои услуги (строительство, монтаж оборудования, пуск и наладка), в результате чего средняя стоимость таких услуг в конкурентных условиях снижается: предложение превысило спрос. Из второго уравнения следует, что рост индекса инвестиционных затрат в предыдущем месяце на 1 п. п. обусловит снижение индекса цен на строительно-монтажные работы на 0,123 п. п. в текущем периоде.

Изучение устойчивости первого уравнения модели (13.19) для прогнозирования индекса реальных инвестиций проведено с помощью рекурсивных расчётов. В частности, колеблемость остатков вокруг нулевого среднего близка к случайной, при этом они практически не выходят за доверительные границы в два среднеквадратических отклонения (рисунок 13.10). Данный факт подтверждает несмещённость оценок коэффициентов модели. Из рисунка 13.11 следует, что при увеличении объёма выборки коэффициенты асимптотически стремятся к своим значениям, характерным в целом для всей выборочной совокупности. При этом желаемое приближение достигается уже к концу 1998 г. (менее $\frac{1}{4}$ длины временного ряда рекурсивных коэффициентов), что говорит о высокой устойчивости, состоятельности оценок коэффициентов модели.

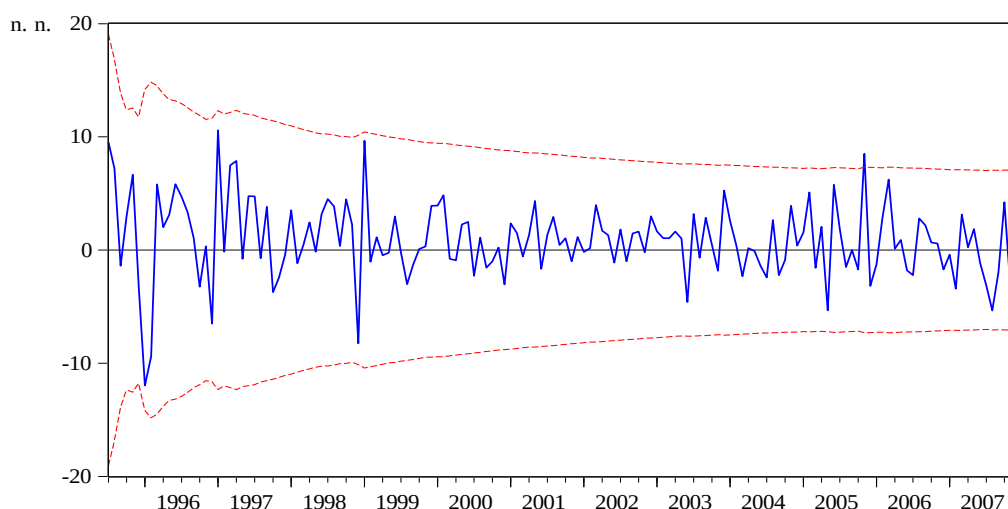


Рисунок 13.10 — Рекурсивные остатки первого уравнения модели (13.19) с границами ± 2 стандартных отклонения

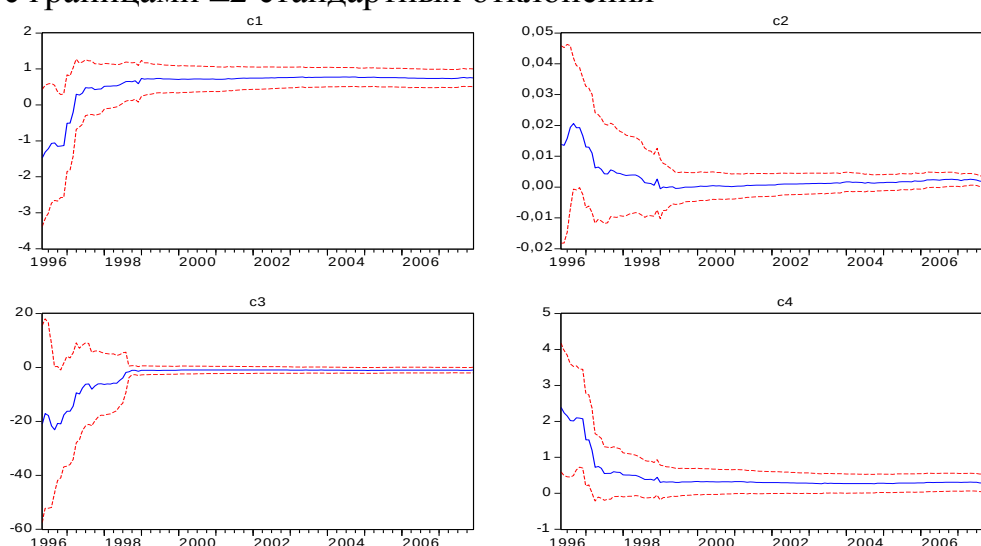


Рисунок 13.11 — Динамика рекурсивных коэффициентов модели (13.19) с границами ± 2 стандартных отклонения

Таким образом, нами получено несколько прогнозов, имеющих высокое качество, но уступающих друг другу по одному или нескольким показателям точности. Предполагая, что функция потерь является линейной (то есть, «наиболее значимым» показателем точности является MAE), определим веса для усредняемых прогнозов по формуле (12.44). Данный подход позволил присвоить наибольший вес прогнозам по модели SARMA(2,1)(1,0)₁₂, имеющим наименьшую ошибку MAE (таблица 13.16).

Таблица 13.16 — Определение доли индивидуальных прогнозов в усреднённом прогнозе

Модель (метод)	MAE _i , п. п.	$\frac{1}{MAE_i}$	Доля индивидуального прогноза в усреднённом прогнозе (d_i), %
Тренд-сезонная (TSM)	5,3	0,187	25,6
Экспоненциальное	5,3	0,190	26,1

сглаживание (ES)			
SARMA(2,1)(1,0) ₁₂	5,2	0,192	26,3
Многомерная (ММ)	6,2	0,161	22,1
Итого	22,0	0,730	100,0

Применение формулы (12.43) дало следующие результаты: ошибка усреднённого прогноза $MAE_c=4,9$ п. п.; $PMSE_c=32,0$ (п. п.)²; $MAPE_c=6,3\%$, что превосходит результаты каждой из индивидуальных моделей (таблица 13.15). Графически результаты представлены на рисунке 13.12.

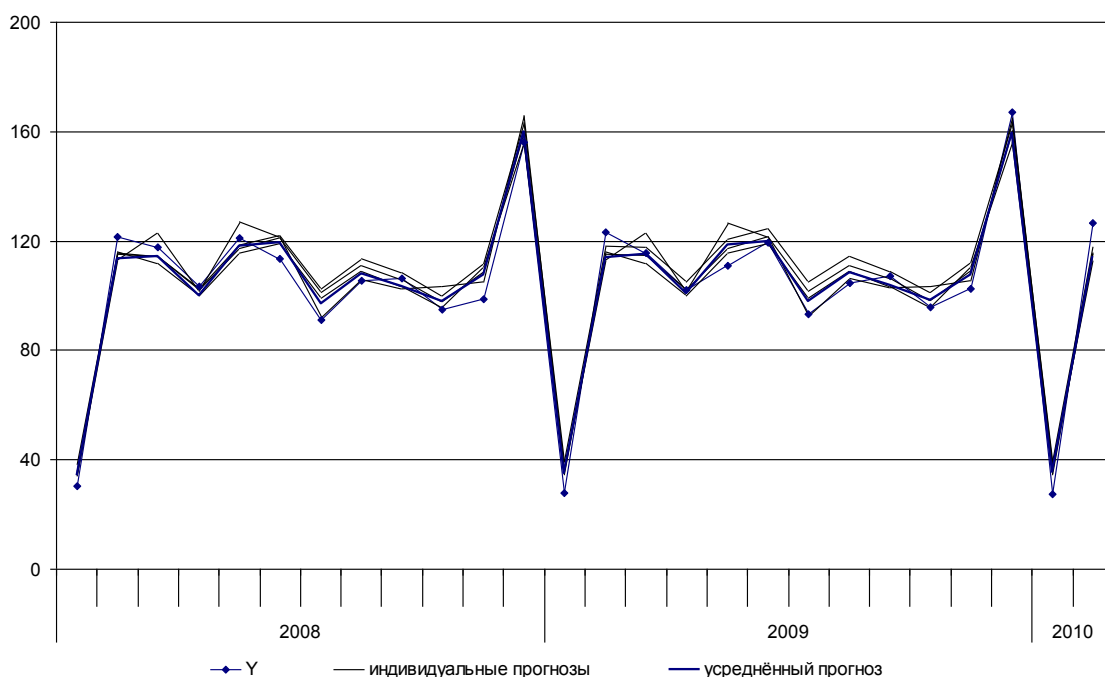


Рисунок 13.12 — Взвешенное усреднение индивидуальных прогнозов

На рисунке 13.12 видно, что усреднённый прогноз имеет лучшее приближение к фактическим послевыборочным данным, чем индивидуальные прогнозы.

Рекомендуемая литература

- 1 Арженовский, С.В. Статистические методы прогнозирования : учебное пособие / С.В. Арженовский, И.Н.Молчанов; Ростовск. гос. ун-т. – Ростов-на-Дону, 2001. – 74 с.
- 2 Афанасьев, В.Н. Статистическое обеспечение устойчивости сельскохозяйственного производства / В.Н. Афанасьев. – М.: Финансы и статистика, 1996. – 320 с.
- 3 Афанасьев, В.Н. Анализ временных рядов и прогнозирование : учебник / В.Н. Афанасьев, М.М. Юзбашев – М.: Финансы и статистика, 2001. – 228 с.
- 4 Берндт, Э.Р. Практика эконометрики: классика и современность : учебник для студентов вузов, обучающихся по специальности 060000 экономики и управления : пер. с англ. / Э.Р. Брендт; под ред. проф. С.А. Айвазяна. – М.: ЮНИТИ-ДАНА, 2005. – 863 с.
- 5 Богомолов, В.А. Экономическая безопасность: учеб. пособие для студентов вузов, обучающихся по специальности экономики и управления (060000) / В.А. Богомолов. – М.: ЮНИТИ–ДАНА, 2006. – 303 с.
- 6 Винн, Р. Введение в прикладной эконометрический анализ / Р. Винн, К. Холден. - М.: Финансы и статистика, 1981. – 294 с.
- 7 Гладилин, А.В. Эконометрика: учебное пособие / А.В. Гладилин, А.Н. Герасимов, Е.И. Громов. – М.: КНОРУС, 2006. – 232 с.
- 8 Глинский, В.В. Методы типологии данных в социально-экономических исследованиях:автореф.дисс. д-ра эконом.наук / В.В. Глинский. – СПб., СПбГУЭФ, 2009. – 44 с.
- 9 Гранберг, А.Г. Статистическое моделирование и прогнозирование:учебное пособие / Г.М. Гамбаров, Н.М. Журавель, Ю.Г. Королев и др.; под ред. А.Г. Гранберга. – М.: Финансы и статистика, 1990. – 383 с.
- 10 Грибунин, В.Г. Введение в вейвлет-преобразование. [Электронныйресурс]/ В.Г. Грибунин; фирма АВТЭКС. - СПб. – Режим доступа: <http://www.autex.spb.ru> – 59 с.
- 11 Дуброва, Т.А. Статистические методы прогнозирования:учебное пособие для вузов / Т.А. Дуброва. – М.: ЮНИТИ-ДАНА, 2003. – 206 с.
- 12 Канторович, Г.Г. Анализ временных рядов / Г.Г. Канторович // Экономический журнал ВШЭ. 2002. – №2. – С. 251–273.
- 13 Канторович, Г.Г. Анализ временных рядов / Г.Г. Канторович // Экономический журнал ВШЭ. 2003. – №1. – С. 79–103.
- 14 Кендалл, М. Многомерный статистический анализ и временные ряды / М. Кендалл, А. Стьюарт; пер. с англ. Э.Л. Пресмана, В.И. Ротаря; Под ред. А.Н. Колмогорова, Ю.В. Прохорова. – М.: Наука, 1976. – 736 с.
- 15 Кильдишев, Г.С. Анализ временных рядов и прогнозирование / Г.С. Кильдишев, А.А.Френкель. – М.: Статистика, 1973. – 102 с.
- 16 Левин, В.С.Методология статистического исследования инвестиций в основной капитал: пространственно-временной аспект:

монография / В.С. Левин, В.Н. Афанасьев, Т.Н. Левина.-М.: ИД «ФИНАНСЫ и Кредит», 2010. – 256с.

17 Лукашин, Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов: учеб. пособие / Ю.П. Лукашин. – М.: Финансы и статистика, 2003. – 416 с.

18 Любчик, В.В Развитие системы методов статистического анализа временных рядов: автореф. дис. ... канд. эконом. наук / В.В. Любчик. – Оренбург, 2011. – 22 с.

19 Любчик, В.В Развитие системы методов статистического анализа временных рядов: дис. ... канд. эконом. наук : 08.00.12 : защищена 14.07.2011 : утв. 24.09.0211 / В.В. Любчик. – Оренбург, 2011. – 180 с. – 04201163186.

20 Новак, Э. Введение в методы эконометрики: Сборник задач; пер. с польск. / Э. Новак; под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2004. – 248 с.

21 Перуновский, О.Е. Моделирование валютных рынков на основе процессов с длинной памятью / О.Е. Перуновский. – М.: ГУ-ВШЭ, 2004. – 46 с.

22 Садовникова, Н.А. Анализ временных рядов и прогнозирование : учебное пособие / Н.А. Садовникова, Р.А. Шмойлова; Московский государственный университет экономики, статистики и информатики – М., 2001. – 67 с.

23 Справочник по прикладной статистике. В 2-х т. Т. 2: Пер. с англ. / Под ред. Э. Ллойда, У. Ледермана, С.А. Айвазяна, Ю.Н. Тюрина. – М.: Финансы и статистика, 1990. – 526 с.

24 Френкель, А.А. Прогнозирование производительности труда: методы и модели / А.А. Френкель. – М.: Экономика, 1989. – 213 с.

25 Эконометрика: учебник / И.И. Елисеева, С.В. Курышева, Т.В. Костеева[и др.]; под ред. И.И. Елисеевой. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2007. – 576 с.

26 Эконометрика: учебник / Н.П. Тихомиров, Е.Ю. Дорохина. – 2-е изд., стереотип. – М.: Издательство «Экзамен», 2007. – 512 с.

27 Балаш, В.А. Модели линейной регрессии для панельных данных: учеб. пособие / В.А. Балаш., О.С. Балаш – М., 2002. – 65 с.

28 Ратникова, Т.А. Введение в эконометрический анализ панельных данных / Т.А. Ратникова // Экономический журнал ВШЭ. -2006. - № 2. – С. 267-316.

29 Статистическое моделирование и прогнозирование : учебное пособие / Г.М. Гамбаров [и др.]; под ред. А.Г. Гранберга. – М.: Финансы и статистика, 1990. – 383 с.

30 Мхитарян, В.С. Эконометрика : учебно-методический комплекс / В.С. Мхитарян, М.Ю. Архипова, В.П. Сиротин. – М. : Изд. центр ЕАОИ, 2008. – 144 с.

31 Вербик, Марно. Путеводитель по современной эконометрике : пер. с англ. / М. Вербик; научн. ред. и предисл. С. А. Айвазяна — М : Научная книга, 2008. – 616 с. – ISBN 978-5-913-035-4.

- 32 Джонстон, Дж. Эконометрические методы : пер. с англ. / Дж. Джонстон. – М. : Статистика, 1980. – 444 с.
- 33 Aalen O.O. What can statistics contribute to a causal understanding? / O.O. Aalen, A. Frigessi // Board of the Foundation of the Scandinavian Journal of Statistics, 2007. – № 4. – P. 155–168.
- 34 Anderson H.M. A cointegration analysis of treasury bills / H.M. Anderson, C.W.J. Granger, A. Haal// The review of Economics and Statistics, 1992. – № 74. – P. 116–126.
- 35 Balke N.S. Threshold cointegration / N.S. Balke, T.B. Fomby// International Economic Review, 1997. – № 38. – P. 627–645.
- 36 Chatfield C. Time series forecasting / C. Chatfield // London: Chapman and Hall, 2000. – 267 p.
- 37 Davidson R. Econometric theory and methods / R. Davidson, J.G. MacKinnon // New York: Oxford University Press, 2004. – 693 p.
- 38 Diebold F.X. Unit root tests are useful for selecting forecasting models / F.X. Diebold, L. Kilian// Journal of business and economic statistics, 2000. – № 18. – P. 265–273.
- 39 Enders W. Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates / W. Enders, C.W.J. Granger // Journal of Business and Economic Statistics, 1998. – № 16. – P. 304–311.
- 40 Enders W. Cointegration and threshold adjustment / W. Enders, P.L. Siklos// Journal of Business and Economic Statistics, 2001. – № 19. – P. 166–176.
- 41 Franses P.H. Unit roots in the Nelson-Plosser data: Do they matter for forecasting? / P.H. Franses, F. Kleibergen// Int. J. Forecasting, 1996. – № 12. – P. 283–288.
- 42 Goodwin B.K. Real interest rate equalization and the integration of international financial markets / B.K. Goodwin, T.J. Grennes// Journal of International Money and Finance, 1994. – № 13. – P. 107–124.
- 43 Granger C.W.J. An introduction to long-memory time series and fractional differencing / C.W.J. Granger, R. Joyeux// Journal of Time Series Analysis, 1980. – № 1. – P. 15–29.
- 44 Granger C.W.J. Further developments in the study of cointegrated variables/ C.W.J. Granger, N.R. Swanson// Oxford Bulletin of Economics and Statistics, 1996. – № 58. – P. 537–553.
- 45 Greene W.H. Econometric analysis. Fifth edition / W.H. Greene. - New York: Pearson Education International, 2003. – 1026 p.
- 46 Hatemi-J A. Multivariate-based tests of twin deficits in the US / A. Hatemi-J, G. Shukur// Journal of Applied Statistics, 2002. – № 29. – P. 817–824.
- 47 Kalman R.E., Bucy R.S. New results in filtering and prediction theory. Trans. / R.E. Kalman, R.S. Bucy// ASME J. Basic Eng., 1961. – № 83. – P. 95–108.
- 48 Kasa K. Common stochastic trends in international stock markets / K. Kasa // Journal of Monetary Economics, 1992. – № 29. – P. 95–124.

49 Kugler P. International real interest rate parity equalization: a multivariate time series approach / P.Kugler, K.Neusser// Journal of Applied Econometrics, 1993. – № 8. – P. 163–174.

50 Making data meaningful: a guide to writing stories about numbers. United Nations Economic Commission for Europe. – Geneva, 2006. – 21 p.

51 Shumway R.H. Time series analysis and its applications. With R examples. Second edition / R.H.Shumway, D.S. Stoffer -New York: Springer, 2006. – 576 p.

52 Siklos P.L. Temporary cointegration with an application to interest rate parity / P.L.Siklos, C.W.J.Granger // Macroeconomic Dynamics, 1997. – № 1. – P. 640–657.

53 Taylor A.M.R. On the practical problems of computing seasonal unit root tests/ A.M.R. Taylor // Int. J. Forecasting, 1997. – № 13. – P. 307–318.