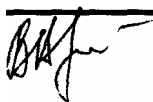


В.Н. Афанасьев, М.М. Юзбашев



АНАЛИЗ ВРЕМЕННЫХ РЯДОВ И ПРОГНОЗИРОВАНИЕ

Допущено Министерством образования Российской Федерации в
качестве учебника для студентов высших учебных заведений,
обучающихся по направлению и специальности «Статистика»

С
уважением
и на добрую
память от автора
д.т.н. Юзбашев
29.11.01 В.Н.А.



МОСКВА «ФИНАНСЫ И СТАТИСТИКА»

2001

РЕЦЕНЗЕНТЫ:

И.И. Елисеева

доктор экономических наук, профессор Санкт-Петербургского университета экономики и финансов, чл.-корр. РАН;

В.С. Мхитарян

доктор экономических наук, профессор Московского государственного университета экономики, статистики и информатики

Афанасьев В.Н., Юзбашев М.М.

А94 Анализ временных рядов и прогнозирование: Учебник. — М.: Финансы и статистика, 2001. — 228 с.: ил.

ISBN 5-279-02419-8

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

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

0702000000-082

A 010(01)-2001 ш-
2001

УДК

311.1(075.8)

ББК 60.6я73

ISBN 5-279-02419-
8

© Афанасьев В.Н.,
Юзбашев М.М., 2001

ПРЕДИСЛОВИЕ

Цель создания данного учебника - помочь студентам, изучающим статистику и особенно специализирующимся по этой дисциплине, получить более широкое и углубленное представление по всему кругу методов статистического исследования, по изменениям социально-экономических и других массовых явлений во времени.

Министерством образования Российской Федерации в государственном образовательном стандарте по специальности 061700 «Статистика» утвержден курс под названием «Анализ временных рядов и прогнозирование».

В учебниках по общей теории статистики этой теме посвящается, как правило, одна глава «Анализ динамики» («Динамические ряды»), что явно недостаточно для глубокого изложения непрерывно развивающихся методов статистического изучения разнообразных массовых процессов. В предлагаемом учебнике, в частности, подробно излагаются методы распознавания типа тренда, алгоритмы расчета логарифмического, логистического, гиперболического трендов. Описываются различные модели сезонных колебаний, рассматривается моделирование функционально и корреляционно связанных систем показателей, а также система показателей устойчивости уровней и тенденций изменения. Ряд положений (теорем) доказываются авторами впервые.

Учебник рассчитан на лиц, имеющих подготовку по общей теории статистики, хотя авторы стремились сделать изложение максимально доступным. В этом, в частности, состоит отличие предлагаемого учебника от переводных книг таких авторов, как Т. Андерсон, К. Гренджер и М. Хатанака, а также отечественных монографий по прикладной математической статистике,

Ограниченный объем пособия не позволяет углубленно рассматривать те или иные проблемы статистики и вынуждает авторов в некоторых случаях отсылать к отдельным специальным работам, указанным в списке рекомендуемой литературы. Иногда авторы отсылают читателей вперед, в последующие главы, что методически не совсем верно, но в целях экономии места и имея в виду хорошую подготовленность читателя, авторы считают оправданным.

Авторы выражают глубокую признательность проф., чл.-корр. РАН И.И. Елисейевой и проф. В.С. Мхитаряну за ценные замечания, сделанные ими при рецензировании рукописи настоящего учебника.

Компьютерный набор и выверку рукописи осуществляли И.Н. Выгорова и Т.Н. Левина, за что авторы приносят им свою благодарность.

Авторы будут благодарны всем, кто пожелает высказать свои предложения по улучшению данного учебника, впервые издаваемого в России.

Предложения просим присылать по адресу:

460795 ГСП г. Оренбург, ул. Челюскинцев, 18 Кафедра статистики и экономического анализа Оренбургского государственного аграрного университета.

Глава 1. ВИДЫ И ПОСТРОЕНИЕ ВРЕМЕННЫХ РЯДОВ

1.1. Что такое «временной ряд»

Термин *временные ряды* в статистике России пока непривычен. В учебниках по общей теории статистики преобладают термины *ряды динамики, динамические ряды, статистическое изучение динамики*. В зарубежной англоязычной литературе принят термин *time series*, в немецкой - *zeitreihen 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 слова *временной*, т.е. относящийся ко времени, связанный со временем, происходящий во времени, и слова *временный*, т.е. непостоянный, преходящий, малосущественный. В европейских языках это разные слова: в немецком, например, временный - *provisorisch*, во французском - *provisoire*, в английском - *provisional*, т.е. эти слова происходят не от корня «время».

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

Неточность последнего термина состоит в том, что не каждый ряд уровней за последовательные моменты или периоды времени содержит на самом деле (отражает) динамику какого-либо признака. Термин *динамика* правильнее относить к изменениям, направленному развитию, наличию тенденции рассматриваемых во времени показателей. Про ряд уровней, содержащих лишь колебания, но не имеющих надежно установленной тенденции, говорят: «В этом ряду, в данном процессе нет никакой динамики». Так можно характеризовать экономику застойного периода, население страны или региона, находящиеся в стационарном состоянии, любую общественную или механическую систему, находящуюся в статическом состоянии.

Следовательно, *динамические ряды* - понятие, относящееся к тем рядам уровней, в которых содержится тенденция изменения, а *временные ряды* - более общее понятие, включающее как динамические, так и статические последовательности уровней какого-либо показателя.

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

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

1.2. Классификация временных рядов

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

1) по времени - *моментные и интервальные. Интервальный ряд* (табл. 1.1) - последовательность, в которой уровень явления относят к результату, накопленному или вновь произведенному за определенный интервал времени. Таковы, например, ряды показателей объема продукции предприятия по месяцам года, количества отработанных человеко-дней по отдельным периодам (месяцам, кварталам, полугодиям, годам, пятилетиям и т.п.) и т.д. Если же уровень ряда характеризует изучаемое явление в конкретный момент времени, то совокупность уровней образует *моментный ряд*. Примерами моментных рядов могут быть последовательность показателей численности населения на начало года, поголовье скота в фермерских хозяйствах на 1 декабря или 1 июня за несколько лет, величина запаса какого-либо материала на начало периода и т.д. Важное отличие моментных рядов от интервальных состоит в том, что сумма уровней интервального ряда дает вполне реальный показатель - общий выпуск продукции за год (пятилетие, десятилетие), общие затраты рабочего времени, общий объем продаж акций и т.д., сумма же уровней моментного ряда иногда и подсчитывается, но реального содержания, как правило, не имеет;

Таблица 1.1 Производство животного и растительного масла в регионе, т

Показатель	Год			
	1990	1997	1998	1999
Животное масло	31149	7376	6137	5286
Растительное масло	35249	22944	25701	24217

2) по форме представления уровней - ряды *абсолютных* (см. табл. 1.1), *относительных* (табл. 1.2) и *средних величин* (табл. 1.3);

3) по расстоянию между датами или интервалами времени выделяют полные и неполные временные ряды. *Полные ряды* имеют место, когда даты регистрации или окончания периодов следуют друг за другом с равными интервалами (см. табл. 1.2; табл. 1.4), *неполные* - когда принцип равных интервалов не соблюдается (см. табл. 1.1 и 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 банках отражают платежеспособность населения и т.д.

Таблица 1.2

Динамика индексов цен на реализованную сельскохозяйственную и приобретенную сельскохозяйственными предприятиями промышленную продукцию, разы

Показатель	Цепные индексы цен							
	1991	1992	1993	1994	1995	1996	1997	1998
Сельскохозяйственная продукция	1,6	8,6	8,1	3,0	3,3	1,4	1,1	1,1
Промышленная продукция	1,9	16,2	10,7	4,2	3,2	1,6	1,2	1,1
Минеральные удобрения	1,7	12,8	10,7	6,5	3,4	1,9	1,2	1,1
Электроэнергия	1,5	13,3	21,6	6,5	2,9	1,6	1,2	1,2
Горючие и смазочные материалы	1,3	34,6	11,3	3,4	3,9	1,4	1,2	1,1
Топливо	2,0	16,2	11,7	4,9	3,4	1,6	1,1	1,1

Таблица 1.3

Производство сельскохозяйственной продукции в Самарской (Куйбышевской) области в расчете на одного жителя

Вид продукции	Год						
	1971-1975	1976-1980	1981-1985	1986-1990	1991-1995	1996-1998	1999
Зерно, кг	803,7	970,6	724,6	770,4	621,6	514,5	195,7
Молоко, кг	276,4	283,5	300,6	326,5	279,4	200,0	202,6
Мясо в убойном весе, кг	46,3	48,6	52,4	65,1	50,7	29,4	29,0
Яйцо, шт.	209	263	294	321	270	145	153
Картофель, кг	177	210	157	161	151	140	121
Овощи, кг	69	76	80	83	47	53	52

Таблица 1.4

Прогноз агрегированного показателя экономической конъюнктуры в России в 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 (см. табл. 1.4) основаны на частных показателях и характеризуют изучаемый процесс комплексно. Так, чтобы иметь представление о состоянии экономики в 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 экономическому содержанию* и учитывать существо изучаемого явления и цель исследования.

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

Несопоставимость по территории возникает в результате изменений границ стран, регионов, хозяйств и т.п. Для приведения 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 должны характеризовать размер того или иного явления по одному и тому же кругу входящих в состав его частей. Например, при характеристике динамики урожайности овощных культур в регионе по годам нельзя в одни годы учитывать только сельскохозяйственные предприятия, 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. Например, до недавнего времени в России урожайность сельскохозяйственных культур определяли делением валового сбора на продуктивную весеннюю площадь, в США - 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-ч, л.с.); атмосферного давления 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.

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

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

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

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

Глава 2. СОСТАВЛЯЮЩИЕ ЭЛЕМЕНТЫ ВРЕМЕННОГО РЯДА

2.1. Понятие об основной тенденции и колеблемости временных рядов

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

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

При статистическом изучении динамики необходимо четко разделять два основных ее элемента - тенденцию и колеблемость, чтобы дать каждому

из них количественную характеристику с помощью специальных показателей.

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

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

2.2. Иерархия тенденций и колебаний

В предыдущих разделах были рассмотрены две основные компоненты временного ряда - его тенденция и колеблемость отдельных уровней. При решении конкретных задач статистического исследования эти компоненты следует разделять, измерять каждую из них отдельно. В то же время при рассмотрении сложных процессов на больших интервалах времени мы наблюдаем иерархию тенденций и колебаний: то, что для времени высокого порядка, например столетия, выступает как колебания, на интервале времени низшего порядка, например трех-пяти лет, может выступать как тенденция. Например, существует 10-11-летняя циклическая колеблемость солнечной активности, одним из показателей которой служат *числа Вольфа W* (число групп солнечных пятен, умноженное на 10, плюс число отдельных 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 1994 и 1995 гг.

Год	Месяц											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII

1994	58,8	35,9	31,7	16,7	18,2	28,1	35,0	22,8	26,7	43,8	18,0	26,2
1995	23,8	29,9	31,1	14,6	14,7	15,8	14,6	15,1	12,3	21,7	9,4	10,8

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

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

Имеется «высший» уровень динамики температур — ее тенденция к повышению, в основном в результате антропогенного 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 13-14ч температура

имеет тенденцию роста, а к вечеру - тенденцию снижения ввиду дневного нагревания воздуха солнечным светом и охлаждения ночью. Но в отдельные часы температура колеблется около этих «тенденций» в зависимости от облачности, ветра, дождя и т.д., иногда за полчаса температура воздуха может измениться на 5 - 10 градусов.

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

В связи с этим знание статистических методов и изучение тенденций и колебаний для экономиста-менеджера, для статистика-аналитика имеют 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.

Исторический метод. Периодизация осуществляется на основе «узаконенной» структуры динамики. При этом обращают внимание на значимые даты и события, а именно: время принятия управленческих решений по данному показателю, смену хозяйственного механизма, смену руководства, войны и т.п. Недостатком этого метода является то, что точные временные границы периодов путем теоретического анализа удается получить крайне редко.

Метод параллельной периодизации. Идея этого метода 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 «однокачественных» периодов развития нужно взять периоды x .

Рассмотрим условный пример.

Показатель	Год								
	1981	1982	1983	1984	1985	1986	1987	1988	1989
x	10	9	11	13	12	18	17	20	21
y	20	19	21	24	24	35	34	40	41

Периоды однокачественной динамики показателей x легко выделить: это 1981-1985 и 1986-1989 гг. Линейный коэффициент корреляции между этими рядами очень высок: $R = 0,995$. Таким образом, можно считать, что ряд x полностью определяет значение уровней ряда y . Теперь, если предстоит качественный скачок показателя x , с очень большой степенью вероятности можно ожидать аналогичных изменений показателя y . В качестве недостатка метода параллельной периодизации следует отметить сложности в определении x — детерминирующего показателя. Более того, во многих случаях такой параметр вообще невозможно найти, так как он должен обладать весьма редкими свойствами - связью с анализируемым показателем и, главное, неоспоримыми временными границами периодов.

Методы многомерного статистического анализа. Часто требуется выделить однокачественные 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 др. 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 вследствие недоступности статистической информации. На основе комплексных динамических рядов (системы показателей) периодизация реализуется методом многомерной средней и методами факторного и кластерного анализа.

Однокачественность уровней временного ряда означает, что в пределах всего изучаемого периода, к которому относятся уровни, должна быть проведена типологическая группировка.

После выделения однородных групп могут использоваться и анализироваться уровни ряда. Это требование может быть сформулировано как обеспечение сравнимости по структуре совокупности, для чего обычно применяется стандартная, нормативная структура.

Глава 3. ПОКАЗАТЕЛИ ВРЕМЕННОГО РЯДА И МЕТОДЫ ИХ ИСЧИСЛЕНИЯ

3.1. Показатели, характеризующие тенденцию динамики

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

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

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

$$\begin{aligned} \text{цепное: } \Delta_{ц} &= y_n - y_{n-1}; \\ \text{базисное: } \Delta_0 &= y_n - y_0. \end{aligned}$$

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 тыс. т. 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

Номер периодов или моментов времени	Уровень ряда, тыс. т	Абсолютное изменение уровней, тыс. т в год	Ускорение абсолютного изменения, тыс. т-год	Темп роста уровня к периоду, %	
				предыдущему	начальному
0 (начальный)	100	-	-	-	-
1	112	12	112	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 из табл. 3.1, ускорение является константой тенденции данного ряда, что свидетельствует о параболической форме этой тенденции. Ее уравнение имеет вид:

$$y_i = y_0 + at_i + bt_i^2, \tag{3.1}$$

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 производства (или добыча 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:

$$S = S_0 + V_0t + \frac{at^2}{2},$$

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 уве-

личивало 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 происходит 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 составляет приблизительно 15% в год), и за шесть лет объем продукции увеличился в 2,32 раза, а на втором предприятии, вычислив также шесть уровней параболического тренда, читатель убедится, что в среднем за год объем продукции возрастал примерно на 20%, а за шесть лет он 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:

цепной темп роста в период
с номером n

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

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

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

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

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

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

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

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

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

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

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

Значения цепных темпов прироста, рассчитанных каждый к своей базе, различаются не только числом процентов, но и величиной абсолютного изменения, составляющей каждый процент. Следовательно, 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 и нередко выражаются в процентах. Следовательно, абсолютные изменения (и ускорения) тоже оказываются относительными величинами и 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 свидетельствует 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 дол.).

Например, «Биржевые ведомости» за 5.05.90 сообщают: индекс Доу Джонса на 3.05.90 составил 2689,64% по сравнению 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%), то создается ложное впечатление о чудовищном крахе на биржах США, потому что снижение на 70% воспринимается как темп изменения - будто от прежней цены акций осталось только 30%.

На самом деле снижение показателей с 2760 до 2690% никакой катастрофой экономике США не грозит: это обычная на рынке ценных бумаг колеблемость курсов. «Биржевые ведомости» далее сообщали, что индекс Доу Джонса на 7.06.90 достиг 2911,6%, т.е. по сравнению с 5.05.90 возрос на 222 единицы - во избежание путаницы их 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 и 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. ее 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; 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 равна единице, или 100%. Изменение, происшедшее с одной из долей, 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 не оборудованной 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:

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

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

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

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

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

k - средний темп роста;

m - число групп.

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

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

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 и 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.

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 в N-й области

Год	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
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
1982—1986	157,0
1988—1992	186,6

Средние уровни принято относить к середине усредняемого 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:

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

В моментном ряду смысл среднего уровня заключается 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 усредняемого 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 па число усредняемых отрезков времени от базисного до сравниваемого периода:

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

Например, производство телевизоров в Российской Федерации в 1980 г. составило 4013 тыс. шт., а в 1990 г. - 4717 тыс. шт.

Среднегодовой абсолютный прирост производства телевизоров за 10 лет составил:

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

Для правильной интерпретации показатель среднего 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 гг.); 2) время, на которое показатель рассчитан, время, входящее в его единицу измерения, - 1 год. Можно рассчитать среднемесячный абсолютный прирост за те же 10 лет - он будет в 12 раз меньше среднегодового прироста.

Среднее ускорение абсолютного изменения применяется реже. Для его надежного расчета даже при слабых колебаниях уровней требуется применять методику аналитического выравнивания по параболе II порядка. Не рекомендуется измерять среднее ускорение без абстрагирования от колебаний уровней. Для более грубого, приближенного расчета среднего ускорения можно воспользоваться средними годовыми уровнями, сглаживающими колебания. Например, среднегодовое производство мяса в Российской Федерации составляло:

Пятилетие	1976-1980	1981-1985	1986-1990
Прирост, млн	7,40	8,09	9,68

Абсолютный прирост за второе пятилетие по сравнению с первым составил 0,69 млн. т, за третье по сравнению со вторым 1,59 млн. т. Следовательно, ускорение в третьем пятилетии по сравнению со вторым составило: $1,59 - 0,69 = 0,90$ млн. т в год за пять лет, а среднегодовое ускорение прироста равно: $0,90 : 5 = 0,18$ 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, 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 ввиду асимметрии темпа прироста и темпа сокращения при равных их величинах общий темп прироста всегда отрицателен. Так, если за первый год объем производства вырос на 20%), а за второй снизился на 20%> (темпы цепные), то за два года имеем:

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

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

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

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

где m - число суммируемых уровней;

y_0 - базисный уровень.

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

$$\begin{aligned} k_{\text{пар}} &= 1 - \frac{3}{8} + \sqrt{\frac{9}{64} + \frac{1}{20} \left(\frac{\sum y_i}{y_0} - 5 \right)} = \\ &= 1 - 0,375 + \sqrt{0,1406 + 0,05(8 - 5)} = 1,16407, \text{ или } 116,4\%. \end{aligned}$$

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:

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

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

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

Та же задача может быть решена на основе *ускорений*. Имеем первый ряд с базисным уровнем y_0^I , базисным абсолютным изменением a_0^I и средним ускорением b^I ; второй ряд - с показателями y_0^{II} , a_0^{II} , b_0^{II} . При каком числе n периодов (лет) после базисного уровни рядов сравниваются? Тенденции рядов параболические:

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

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

Приравняв правые части уравнений, получим:

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

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

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

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

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

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; 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 сравниются, 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, получим:

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

откуда

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

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

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: $y_0^1 = 300$; $\bar{k}^1 = 1,09$; второй ряд - $y_0^{11} = 100$; $\bar{k}^{11} = 1,2$, 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 сравняется с первым при сохранении экспоненциальных трендов обоих рядов.

Глава 4. ОСНОВНЫЕ ТИПЫ ТЕНДЕНЦИЙ И УРАВНЕНИЙ ТРЕНДА

В главе 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 отражаются тем или иным уравнением, линией тренда.

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

4.1. Прямолинейный тренд и его свойства

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

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

где \hat{y}_i - выровненные, т.е. 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, т.е. для

$t = 0$;

b - средняя величина изменения уровней ряда за единицу изменения времени;

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

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

народном хозяйстве очень хорошо отражается прямолинейным трендом. Его параметры вычислены в гл. 5 (табл. 5.3).

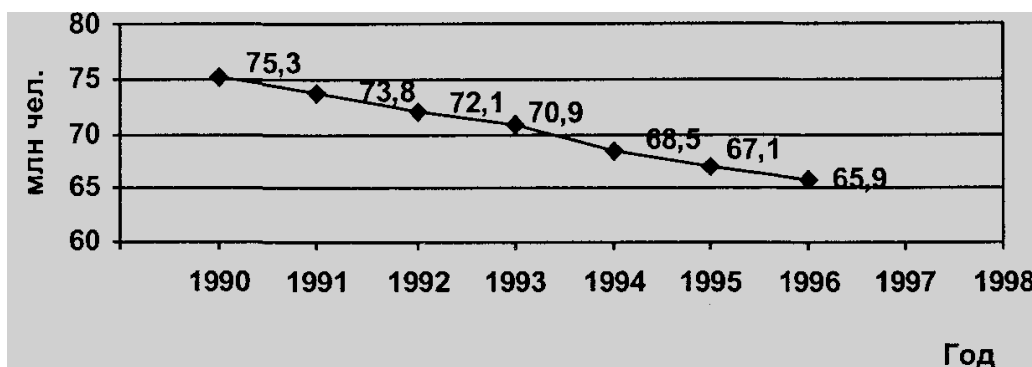


Рис. 4.1. Динамика числа занятых в народном хозяйстве в России на 31 декабря каждого года

Основные свойства тренда в форме прямой линии таковы:

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

Свойства линейного тренда иллюстрирует табл. 4.1. Уравнение тренда:

$$y_i = 100 + 20 * t_i.$$

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

Таблица 4.1

Показатели динамики при линейном тренде к увеличению уровней $y_i = 100 + 20 * t_i$.

Номер периода t_i	Уровень y_i	Абсолютное изменение к	Темпы (цеп- ные), %	Ускорение
1	120	+20	120,0	-
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 уровней: $y_i = 200 - 20 * t_i$.

Номер периода t_i	Уровень y_i	Абсолютное изменение к предыдущему	Темп к предыдущему периоду, %	Ускорение
1	180	-20	90,0	-
2	160	-20	88.9	0
3	140	-20	87.5	0
4	120	-20	85.7	0
5	100	-20	83.3	0
6	80	-20	80,0	0

4.2. Параболический тренд и его свойства

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

$$y_i = a + b * t + c * t^2.$$

Параболы III порядка и более высоких порядков редко применимы для выражения тенденции динамики и слишком сложны для получения надежных оценок параметров при ограниченной длине временного ряда. Прямую линию, с точки зрения математики, можно также считать одним из видов 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 непродолжительном, этапе развития имеют примерно постоянное ускорение абсолютного прироста уровней. Такими бывают рост населения отдельных городов или регионов, ускоренное увеличение объема продукции в фазе циклического подъема, как, например, динамика экспорта Японии в 1988-1995 гг. на рис. 4.2.

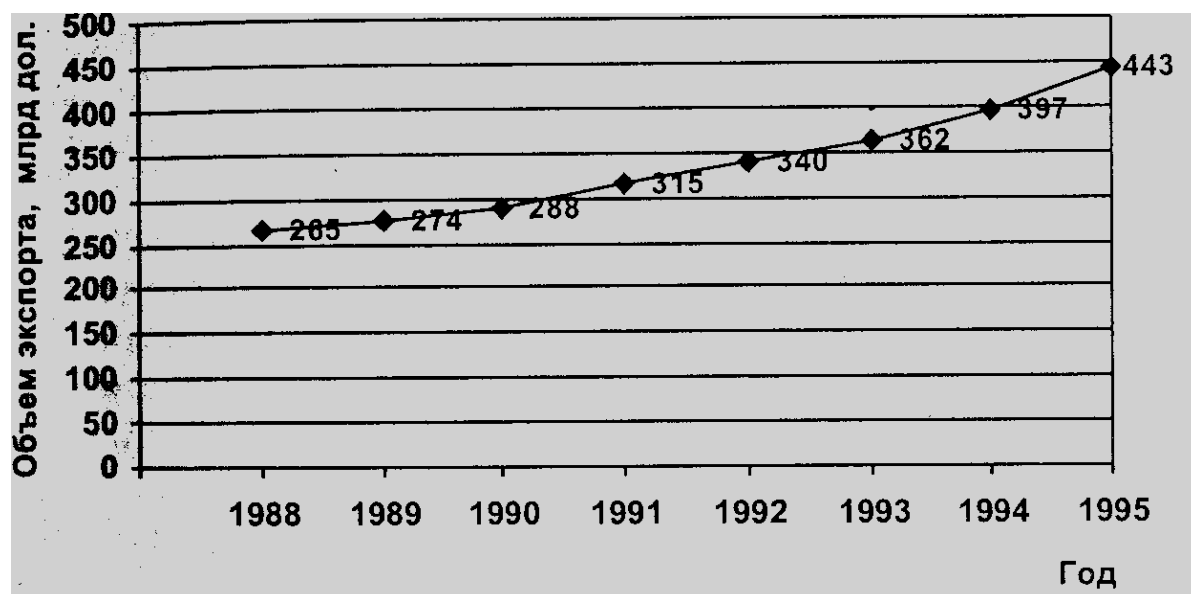


Рис. 4.2. Динамика экспорта Японии

Расчет уравнения этой параболы приведен в гл. 5. Основные свойства тренда в форме параболы II порядка таковы:

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

2) парабола, рассматриваемая относительно ее математической формы, имеет две ветви: восходящую с увеличением уровней признака и нисходящую с их уменьшением. Но относительно статистики по содержанию изучаемого процесса изменений трендом, выражающим определенную тенденцию развития, чаще всего можно считать только одну из ветвей:

либо восходящую, либо нисходящую. В особых, более конкретных, ситуациях мы не отрицаем возможности объединения обеих ветвей в единый тренд;

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

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

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

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

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

4) при параболической форме тренда, в зависимости от соотношений между его параметрами, цепные темпы изменений могут либо уменьшаться, либо некоторое время возрастать, но при достаточно длительном периоде рано или поздно темпы роста 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 четыре случая параболических трендов, а лишь два первых (табл. 4.3 и 4.4).

Таблица 4.3

Показатели динамики при параболическом тренде,

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

Номер периода	Уровень y_i	Абсолютное изменение	Цепные темпы, % к предыдущему периоду	Ускорение
1	122	+22	122,0	-
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 $b < 0, c < 0$: $\hat{y}_i = 200 - 20t - 2t^2$

Номер периода t_i	Уровень y_i	Абсолютные изменения	Цепные темпы, % к предыдущему периоду	Ускорение	Цепное относительное изменение, % к предыдущему периоду
1	178	-22	89,0	-	-11,0

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 $\hat{y} = a + bt + ct^2$ достигается при нулевом значении первой производной:

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

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

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

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

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

Если имеем параболу при $b < 0, a > 0$, например:

$\hat{y}_i = 200 - 20t + 2t^2$, то минимальное значение тренда достигается при $t = \frac{-b}{2c} = \frac{20}{2 \cdot 2} = 5$, и это минимальное значение составит:

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

4.3. Экспоненциальный тренд и его свойства

Экспоненциальным трендом называют тренд, выраженный уравнением: $\hat{y}_i = a \cdot k^{t_i}$ или в форме: $\hat{y}_i = \exp[\ln a + \ln k \cdot t_i]$. Свободный член экспоненты a равен выровненному уровню, т.е. уровню тренда в момент или период, принятый за начало отсчета времени, т.е. при $t=0$. Основным параметром экспоненциального тренда k является постоянным темпом изменения уровней (ценным). Если $k > 1$, имеем тренд с возрастающими уровнями, причем это возрастание не просто ускоренное, а с

возрастающим ускорением и возрастающими производными всех более высоких порядков. Если $k < 1$, то имеем тренд, выражающий тенденцию постоянного, но замедляющегося сокращения уровней, причем замедление непрерывно усиливается. Экстремума экспонента не имеет и при $t \rightarrow \infty$ стремится либо к ∞ при $k > 1$, либо к 0 при $k < 1$.

Экспоненциальный тренд характерен для процессов, развивающихся в среде, не создающей никаких ограничений для роста уровня. Из этого следует, что на практике он может развиваться только на ограниченном промежутке времени, так как любая среда рано или поздно создает ограничения, любые ресурсы со временем исчерпаемы. Однако практика показала что, например, численность населения Земли на протяжении 1950-1985 гг. возрастала примерно по экспоненте со среднегодовым темпом роста $k = 1,018$ и за это время возросла вдвое - с 2,5 до 5 млрд. чел. (рис. 4.3). В настоящее время темп роста населения постепенно уменьшается.

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

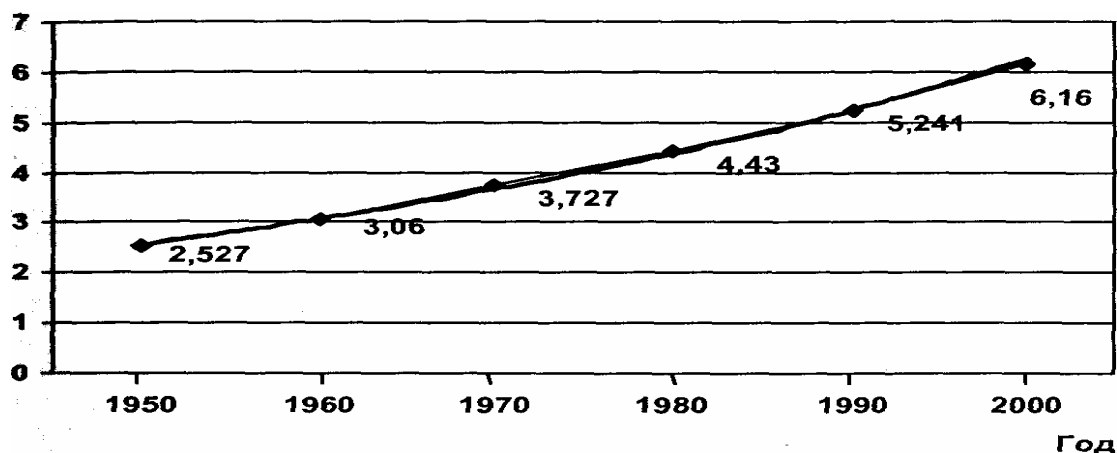


Рис. 4.3. Рост народонаселения Земли

Расчет экспоненциального тренда дан в гл. 5. Основные свойства экспоненциального тренда:

1. Абсолютные изменения уровней тренда пропорциональны самим уровням.

2. Экспонента экстремумов не имеет: при $k > 1$ тренд стремится к $+\infty$, при $k < 1$ тренд стремится к нулю.

3. Уровни тренда представляют собой геометрическую прогрессию: уровень периода с номером $t = m$ есть $a \cdot k^m$.

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

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

Таблица 4.5

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

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

Таблица 4.6

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

Номер периода t_i	Уровень	Абсолютные изменения	Цепные темпы, % к	Ускорение	Замедление ускорения
---------------------	---------	----------------------	-------------------	-----------	----------------------

	€_i	(цепные)	предыдуще му периоду		
1	160,00	40,00	80	-	-
2	128,0	-32,00	80	+8,00	-
3	102,40	-25,60	80	+6,40	-1,60
4	81,92	-20,48	80	+5,12	-1,28
5	65,54	-16,38	80	+4,10	-1,02
6	52,43	-13,11	80	+3,27	-0,83

Читатель может заинтересоваться и таким вопросом: как назвать тенденцию динамики, при которой и темп изменения был бы непостоянен, а имел постоянное абсолютное или относительное изменение, например, уравнение типа $\hat{y} = a(k + bt_i)^{t_i}$ или $\hat{y} = ak^{t^2}$ и т.д. Подобные «гиперэкспоненты» не применяются статистикой, ибо любой, сколь угодно быстрый, сколь угодно ускоряющийся рост может быть отображен обычной экспонентой - стоит лишь уменьшить период, 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 млрд. лет развивается по экспоненте.

4.4. Гиперболический тренд и его свойства

Из различных форм гипербол рассмотрим только наиболее простую:

$$\hat{y} = a + \frac{b}{t}.$$

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

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

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 развития 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 $1/t_i$ должны быть всегда положительными.

Основные свойства гиперболического тренда:

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

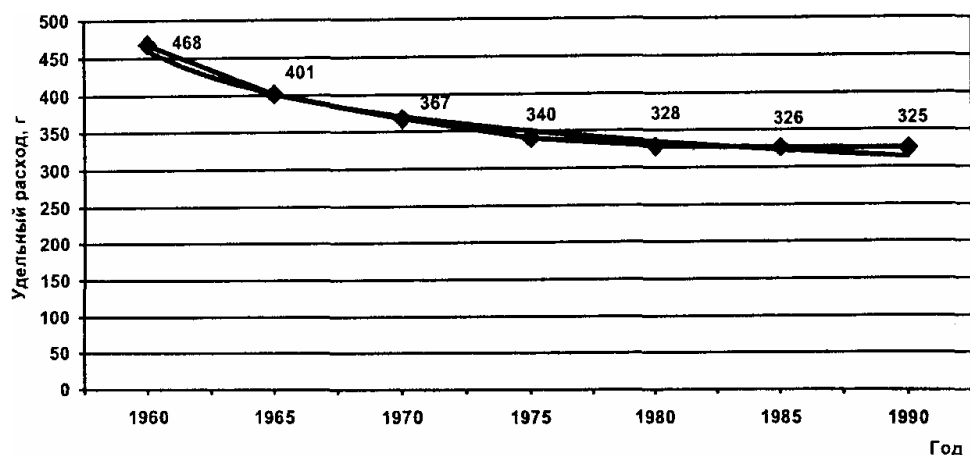


Рис. 4.4. Динамика расхода условного топлива на производство электроэнергии (г на 1 кВт-ч) на электростанциях региона

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

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

Таблица 4.7

Показатели динамики при гиперболическом тренде:

$$\hat{y}_i = 100 + \frac{100}{t_i}$$

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

4.5. Логарифмический тренд и его свойства

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

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

Примером тенденций, соответствующих логарифмическому тренду, может служить динамика рекордных достижений в спорте: известно, что увеличение на 1 см рекорда прыжка в высоту или снижение на 0,1 с времени бега на 200 или 400 м требует все больших и больших затрат времени, каждый рекорд дается все большим и большим трудом. В то же время нет и «вечных» рекордов, все спортивные достижения улучшаются, но медленнее и медленнее, т.е. 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.

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 рис. 4.5 для убедительности нанесен и логарифмический тренд, расчет

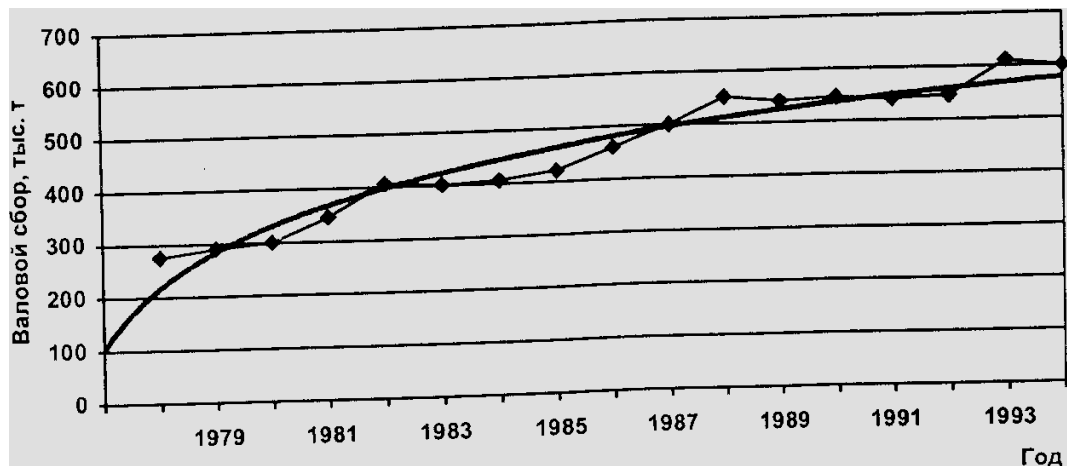


Рис. 4.5. Динамика валового сбора чая в Китае

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, т.е. рост происходит по линейному тренду, затем, в завершающей части цикла, рост замедляется по гиперболе по мере приближения к предельному значению показателя.

Примером такого цикла динамики может служить изменение доли грамотного населения в стране, например в России, с 1800 г. до наших дней, или изменение доли семей, имеющих телевизоры, примерно с 1945 до 2000 г. в России, доли жилищ в городах, имеющих горячее водоснабжение или центральное отопление (процесс, еще не законченный). В некоторых зарубежных программах для компьютеров логистическая кривая называется *S-образной кривой*.

Можно, конечно, логистическую тенденцию считать объединением трех разных по типу 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, не оборудованных 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 в 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 было только для принципиально новых типов фотоаппаратов, насыщенность которыми еще находится в самом начале первого этапа.

В вышеописанном диапазоне изменения уровней, т.е. от нуля до единицы, уравнение логистического тренда имеет вид:

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

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

a_0 должно быть примерно равно -10. Чем больше a_1 , тем быстрее будут снижаться уровни, например, при $a_0 = -10$; $a_1 = 1$, уже при $t_i = 20$ уровни снизятся почти до нуля.

Если же диапазон изменения уровней ограничен не нулем и единицей, а любыми значениями, определяемыми исходя из существа задачи,

обозначаемыми y_{\max} и y_{\min} , то формула логистического тренда принимает вид:

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

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

Таблица 4.8

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

Показатели динамики при логистическом тренде:

Номер периода t_i	Уровень y_i	Абсолютные изменения к предыдущему периоду	Ускорение	Темп роста к предыдущему 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

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 $t \rightarrow \infty$. Ускорения в первой половине периода отрицательные и по модулю возрастающие; во второй половине периода ускорения положительные и уменьшающиеся в пределе до нуля. Темпы изменений все меньше 100%, в конце первой половины периода наименьшие, во второй половине возрастающие с замедлением до 100% в пределе. Графическое изображение логистического тренда приведено на рис. 5.2.

Глава 5. МЕТОДЫ РАСПОЗНАВАНИЯ ТИПА ТРЕНДА И ОЦЕНКИ ЕГО ПАРАМЕТРОВ

При изучении методов распознавания типа тренда не следует забывать о существовании изучаемого процесса, который отображается временным рядом. Как правило, тип тренда должен соответствовать характерным особенностям процесса. В гл. 4 для каждого типа тренда приведены примеры выражаемых этим типом процессов. Определяя другие процессы по временным рядам, полезно по указанным примерам подобрать подходящие типы тренда. Если, например, изучается динамика продуктивности коров или валового надоя молока, то эти процессы аналогичны представленной в гл. 4 динамике урожайности, и скорее всего они 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, вероятнее всего, он будет отображаться 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 предыдущей главе. Кроме того, характер тенденции часто маскируется значительной колеблемостью уровней ряда, поэтому требуется специальная методика распознавания типа тренда, наилучшим образом отражающего тенденцию фактического ряда уровней, чему и посвящена эта глава. После определения типа тренда необходимо вычислить оценки его параметров, как правило, по методу наименьших квадратов, а также с использованием специфических приемов для 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, чтобы равные отрезки означали различие уровней в одинаковое число раз. Интерпретация вида графика будет другой: при линейном масштабе график, близкий к прямой линии, означает линейную тенденцию, а при логарифмическом масштабе оси ординат прямая линия показывает экспоненциальную тенденцию.

Необходимо строго соблюдать равенство промежутков времени на равных отрезках оси абсцисс. Логарифмический масштаб по времени не рекомендуется, так как он крайне затруднит интерпретацию графика. Рассмотрим пример графического изображения, представленный на рис. 5.1.

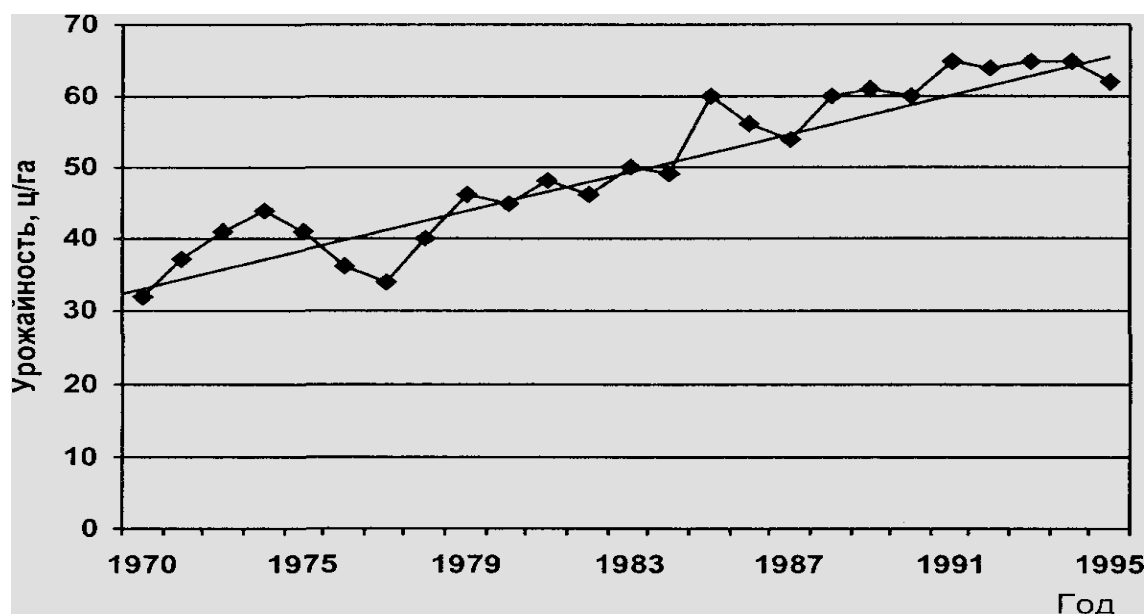


Рис. 5.1. Динамика урожайности зерновых во Франции

—●— фактические уровни
 — тренд

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

Но «преимущество» параболы над прямой может быть невелико. Следовательно, нужно применить опять же статистико-математические критерии существенности уменьшения среднего отклонения при переходе от прямой к параболе. Не отрицая допустимости указанной методики с дополнительной проверкой существенности снижения среднего отклонения от тренда, рассмотрим и другие методы выбора типа тренда без вычисления последнего, 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. Для того чтобы проверить данную гипотезу,

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

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

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

2) по ряду сглаженных уровней вычисляются 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. На основании графика, представленного на рис. 5.1, предложена гипотеза о линейном тренде (табл. 5.1).

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

Полученное значение F-критерия значительно ниже табличного для значения 0,05, следовательно, различия между средними значениями цепных абсолютных изменений в разных подпериодах не являются существенными; вероятность нулевой гипотезы (о случайном характере этих различий) много больше 0,05, и она не может быть отклонена. Принимается исходная гипотеза о том, что средние значения абсолютных приростов 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, представленного 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, который дает оценки параметров, отвечающие принципу максимального правдоподобия: сумма квадратов отклонений фактических уровней от тренда (от выровненных 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 -

Таблица 5.1

Проверка гипотезы о линейном тренде урожайности зерновых культур,
ц/га

Год	Урожайность	Скользкая 5- летняя средняя	Абсолютные изменения	Средние по подпериодам абсолютные изменения
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	$\bar{\Delta}_2 = +1,71$
1981	47,3	48,12	+0,64	
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	$\bar{\Delta}_2 = +1,04$
1988	60,8	58,52	+0,70	
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 1995	65,5 63,2		—	Общая средняя величина $\bar{\Delta}_{общ} = +1,18$
Итого	1332,6	-	-	

Источник данных. Развитие рынка зерна в России.-М.: Изд-во ЦЭЖ при
Правительстве РФ, 1997. - С. 53.

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

Вид вариации	Сумма квадратов отклонений	Число степеней свободы	Дисперсия	F-критерий	
				фактический	табличный для P = 0,05
Между периодами	3,1682	2	1,5841	1,7365	3,55
Остаточная	16,4198	18	0,9122	-	-
Вся вариация	19,5880	20	0,9794	-	-

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 выровненными (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.

5.3.1. Уравнение прямой линии тренда

Уравнение имеет вид:

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

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

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

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

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

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

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

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

После алгебраических преобразований получаем два «нормальных уравнения» метода наименьших квадратов (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 \mathcal{E}_1 равен $a+b$, для второго

$$\mathcal{E}_2 = a + 2b \text{ и т.д.}$$

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 с номером $n/2$ и $(n/2)+1$. В последнем случае все номера периодов t_i будут дробными. При нумерации периодов времени точно от середины ряда половина номеров t_i будет 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:

$$\begin{aligned} na &= \sum_{i=1}^n y_i, \\ b \sum_{i=1}^n t_i^2 &= \sum_{i=1}^n y_i t_i. \end{aligned} \quad (5.5), (5.6)$$

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 численности занятых в народном хозяйстве России с 1990 по 1996 г. представлена в табл. 5.3. В целях экономии места в той же таблице приведены и

другие показатели, необходимые для измерения колеблемости, описываемые в гл.6.

Таблица 5.3 Расчет линейного тренда

Год	Уровень, y_i , млн. чел.	Номер года, t_i	$y_i t_i$	Тренд \hat{y}_i млн. чел.	Отклонение от тренда, $y_i - \hat{y}_i = u_i$	u^2_i	$u_i u_{i+1}$
1990	75,3	-3	-225,9	75,3	0,0	0,00	0,00
1991	73,8	-2	-147,6	73,7	0,1	0,01	0,00
1992	72,1	-1	-72,1	72,1	0,0	0,00	0,00
1993	70,9	0	0,0	70,5	0,4	0,16	-0,16
1994	68,5	1	68,5	68,9	-0,4	0,16	0,08
1995	67,1	2	134,2	67,3	-0,2	0,04	-0,04
1996	65,9	3	197,7	65,7	0,2	0,04	-
Σ	493,6	0	-45,2	493,5	0,1	0,41	-0,12

$$a = \bar{y} = \frac{493,6}{7} = 70,5 \text{ млн чел.}; \quad b = \frac{-45,2}{28} = -1,615 \text{ млн чел. в год.}$$

Уравнение тренда: $\hat{y}_i = 70,5 - 1,615t_i$, $t_i = 0$ в 1993 г. В среднем численность занятых сокращалась на 1615 тыс. чел. 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: $\hat{y}_i = 70,5 - 1,615t_i$, $t_i = 0$ в 1993 г. Для вычисления параметров a , b , c по методу наименьших квадратов три частные производные

функции: $f(a,b,c) = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ приравниваются 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:

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

Уравнения (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, описываемые в гл. 6.

Таблица 5.4. Расчет параболического тренда объема экспорта Японии

Год	Уровень y_i , млрд. дол.*	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	$u_i \cdot u_{i+1}$
1	2	3	4	5	6	7	8	9
1988	265	-3,5	-927,5	3246	266	-1	1	1
1989	274	-2,5	-685,0	1712	275	-1	1	2
1990	288	-1,5	-432,0	648	290	-2	4	-10
1991	315	-0,5	-157,5	79	310	5	25	25
1992	340	0,5	-170,0	85	335	5	25	-15
1993	362	1,5	-543,0	814	365	-3	9	12
1994	397	2,5	-992,5	2481	401	-4	16	-4
1995	443	3,5	1550,5	5427	442	1	1	-
Σ	2684	0,0	1054,0	14492	2684	0	82	11

* Округлены до целых.

$$\sum_{i=1}^8 t_i^2 = \frac{8^3 - 8}{12} = 42; \quad \sum_{i=1}^8 t_i^4 = \frac{3 \cdot 8^5 - 10 \cdot 8^3 + 7 \cdot 8}{240} = 388,5.$$

Вычисляем параметры параболы

$$b = \frac{1054}{42} \cong 25,1;$$

$$8a + 42c = 2684,$$

$$42a + 388,5c = 14492;$$

$$a + 5,25c = 335,5,$$

$$a + 9,25c = 345,0;$$

$$4c = 10,5; c = 2,625,$$

$$a = 321,7.$$

Уравнение тренда:

$$\hat{y}_i = 321,7 + 25,1t_i + 2,625t_i^2,$$

где $t = 0,5$ в 1992 г.

Интерпретация параметров тренда такова: экспорт Японии в 1988-1995 гг. возрастал в номинальной оценке ускоренно, со средним ускорением: $2 \cdot 2,625 = 5,25$ млрд. дол, в год за год, средний за весь период прирост объема экспорта составил 25,1 млрд. дол. в год, средний уровень экспорта на середину периода был равен 321,7 млрд. дол.

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

$$\hat{y}_i = 261,9 + 1,475t_i + 2,625t_i^2,$$

где $t_i = 0$ в 1987 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 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 расхода условного топлива на производство электроэнергии на электростанциях региона (г. на 1 кВт-ч).

Таблица 5.5. Расчет гиперболического уравнения тренда

Год	y_i , Г на 1 кВт-ч	t_i	$\frac{1}{t_i}$	$1/t_i^2$	$A t_i$	Тренд, \hat{y}_i	$y_i - \hat{y}_i =$ u_i	u_i^2
1960	468	1	1,000	1,000	468	473	-5	25
1965	401	2	0,500	0,250	200	387	+14	196
1970	367	3	0,333	0,111	122	359	+8	64
1975	340	4	0,250	0,062	85	344	-4	16
1980	328	5	0,200	0,040	66	336	-8	64
1985	326	6	0,167	0,028	54	330	-4	16
1990	325	7	0,143	0,020	46	326	-1	1
Σ	2555	-	2,593	1,511	1041	2555	0	382

Нормальные уравнения МНК:

$$\begin{aligned} 7a + 2,593b &= 2555, \\ 2,593a + 1,511b &= 1041. \end{aligned}$$

Решая систему уравнений, получаем:

$$a = 301,3; b = 171,9.$$

Уравнение гиперболического тренда удельного расхода топлива имеет вид:

$$\hat{y}_i = 301,3 + \frac{171,9}{t_i},$$

где $t_i = 0$ в 1965г.

Величина удельного расхода 301,3 - это предел, к которому стремится экономия топлива при данной технологии тепловых электростанций региона. Существенного резерва экономии уже нет.

5.4. Оценка параметров экспоненциального, логарифмического и логистического уравнений тренда

Данные типы трендов объединены в одну группу в связи с необходимостью при оценке их параметров прибегать к логарифмированию. При расчете логарифмического уравнения тренда логарифмируют номера периодов (моментов) времени, а при расчете параметров экспоненциального и логистического трендов - сами уровни. Поскольку отрицательные числа не имеют

действительных логарифмов, если нужно логарифмировать номера периодов времени, то нельзя переносить начало их отсчета в середину ряда. Если же сами уровни могут принимать отрицательные значения, например, уровни финансового результата от реализации, уровни температуры воздуха или почвы, то необходимо перенести начало отсчета уровней на величину, алгебраически меньшую реального наименьшего уровня. Например, температуру следует выразить не в градусах Цельсия, а в Кельвинах, финансовый результат при наибольшем убытке 83 млн. руб., отсчитывать от -100 млн. руб., чтобы наименьший уровень выразился как 17 млн. руб. По окончании расчета тренда нетрудно восстановить обычные единицы измерения. Так, получив тренд финансового результата при отсчете от -100 млн. руб. как

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

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:

$$47 < \lfloor \ln(100 : 27) : \ln 1,028 \rfloor < 48.$$

5.4.1. Экспоненциальное уравнение тренда

Формула уравнения имеет вид: $\hat{y}_i = a \cdot k^{t_i}$.

Для нахождения параметров a и k уравнение логарифмируем:

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

В такой форме, т.е. для логарифмов, уравнение соответствует линейному, 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 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 численности населения Земли в 1950-2000 гг.

Год	y_i млн. чел.	t_i	$y_i t_i$	$\ln y_i$	$t_i \ln y_i$	Тренд, \hat{y}_i млн. чел.	$y_i - \hat{y}_i$ $= u_i$	u_i^2
1950	2527	-2,5	-6317,5	7,835	-19,588	2565	-38	1444
1960	3060	-1,5	-4590,0	8,026	-12,039	3065	-5	25
1970	3727	-0,5	-1863,5	8,223	-4,112	3663	+64	4096
1980	4430	0,5	2215,0	8,396	4,198	4377	+53	2809
1990	5241	1,5	7711,5	8,564	12,846	5231	+10	100
2000	6160*	2,5	15400,0	8,726	21,815	6250	-90	8100
Σ	25145	0,0	12555,5	49,770	3,120	25151	-6	16574

* Оценка по данным до 1997 г.

Источники данных. Демографический энциклопедический словарь. - М.: 1985. - С. 271; Demographical Yearbook UNO (до 1997 г.).

$$\ln a = \frac{49,77}{6} = 8,295; \quad a = 4004;$$

$$\ln k = \frac{3,12}{17,5} = 0,1783; \quad k = 1,195.$$

Уравнение тренда:

$$\hat{y}_i = 4004 \cdot 1,195^{t_i},$$

где $t = 0,5$ в 1980г.

Итак, население Земли в период с 1950 по 2000 г. возрастало со среднегодовым темпом роста, равным корню десятой степени из среднего темпа за десятилетие, найденного по данным табл. 5.6, т.е. $\sqrt[10]{1,195} = 1,01797$, или 1,8% прироста в год. Прогнозировать дальнейшую динамику численности населения по рассчитанному тренду не следует, так как уже в десятилетии 1990-2000 гг. темп стал замедляться, и этот процесс, очевидно, будет продолжаться. По данным Венского Международного института прикладного системного анализа, наиболее вероятный вариант роста населения Земли в XXI в. - постепенное замедление роста до полного его прекращения к 2100 г. при уровне населения 11,5 млрд. чел. Крайними и наименее вероятными вариантами к 2100 г. являются: очень слабо замедляющийся рост до 18 млрд. чел. или переход к снижению числа жителей Земли, начиная примерно с середины XXI в., до 5 млрд. чел.

5.4.2. Логарифмическое уравнение тренда

Особенность этого типа тренда заключается в том, что логарифмировать необходимо номера периодов (моментов) времени: $\mathcal{F} = a + b \ln t$. 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 вначале 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).

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

1978-1983 гг.-331,7;

1984-1989 г. 482,7;

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:

Таблица 5.7

Расчет логарифмического тренда валового сбора чая в Китае

Год	y_i тыс.т.	t_i	$\ln t_i$	\bar{y}_i	$y_i - \bar{y}_i = u_i$	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	-
<i>Итого</i>	7716	-	-	7650	$k=6$	652,6	1470

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

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

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

Вычитая (5.18) из (5.19), имеем:

$$b [\ln(x + 6) - \ln(x)] = 151. \quad (5.21)$$

Вычитая (2) из (3), имеем:

$$b [\ln(x + 11,5) - \ln(x + 6)] = 83,3. \quad (5.22)$$

Делим второй результат на первый:

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, поэтому лучше ограничиться приближением $x \approx 6,5$ лет, следовательно, середина между 1980 и 1981 гг. - это номер 6,5 от начала отсчета номеров лет, тогда 1978г. - это номер $t = 4$. Исходя из этого нумеруем все года в табл. 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) и (5.20), получим три независимые 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 1974г.

По этому уравнению рассчитаны уровни тренда \hat{y}_i 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, эту нумерацию поэтому рациональнее проводить от середины ряда. Особенностью логистического тренда является этап обоснования значений максимального и минимального уровней временного ряда. Это обоснование осуществляется 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

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

и обозначим его левую часть, т.е.

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

Условие метода наименьших квадратов:

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

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

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

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

$$na_0 + a_1 \sum_{i=1}^n t_i = \sum_{i=1}^n \ln \zeta_i, \quad (5.23)$$

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

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

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

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

Итак, алгоритм расчета логистической кривой состоит из десяти этапов:

- 1) обоснование величин y_{\max} , y_{\min} ;
- 2) вычисление по фактическому временному ряду значений

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

- 3) вычисление $\ln \zeta_i$;
- 4) нумерация периодов или моментов времени от середины ряда;

5) умножение $\ln \zeta_i$ на t_i ;

$$\sum_{i=1}^n \ln \zeta_i; \sum_{i=1}^n t_i \ln \zeta_i;$$

6) подсчет итоговых сумм

7) вычисление

8) вычисление

9) вычисление для всех периодов

a_0 и a_1 ;

$$\ln \hat{\zeta}_i = a_0 + a_1 t_i;$$

$$\hat{\zeta}_i = \exp(a_0 + a_1 t_i)$$

$$\hat{y}_i = \frac{\hat{y}_{\max} - \hat{y}_{\min}}{\hat{\zeta}_i + 1} + \hat{y}_{\min}.$$

10) вычисление уровней тренда

Проведем расчет логистического тренда по данным рис. 5.2.

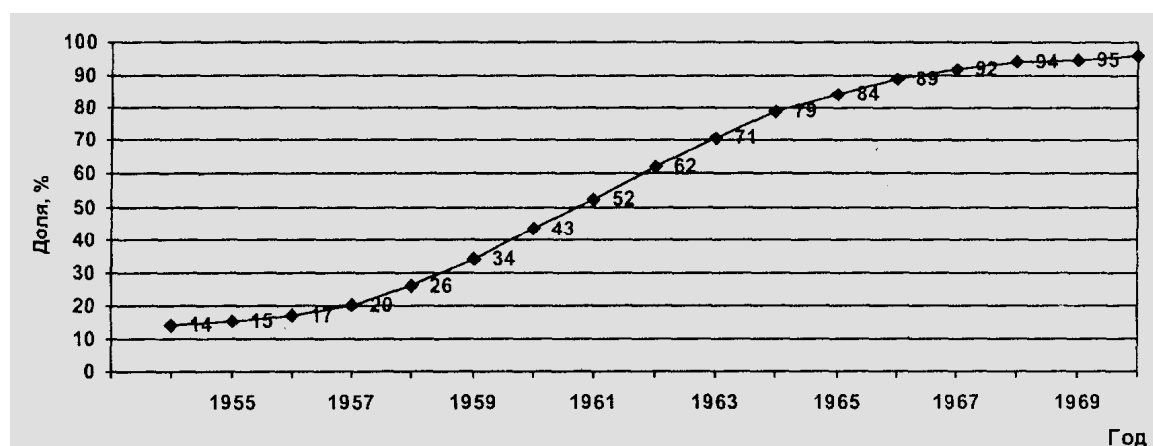


Рис. 5.2. Логистическая тенденция динамики доли тепловозной и электровозной тяги в грузообороте железных дорог СССР

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

Исходя из границ периода времени и фактических уровней ряда получаем:

$$\hat{y}_{min} = 10\%; \hat{y}_{max} = 100\%;$$

$$a_0 = \frac{-0,228}{14} = -0,016286; \quad a_1 = \frac{-99,325}{227,5} = -0,436593;$$

$$\hat{\xi}_i = \exp[-0,016286 + t_i(-0,436593)]$$

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

$$\hat{y}_i = \frac{100 - 10}{e^{-0,016286 - 0,436593 t_i} + 1} + 10.$$

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.

5.5. Многократное скользящее выравнивание

Как видно из табл. 5.3-5.5, при расчете параметров тренда разные уровни имеют неодинаковые веса, так как умножаются на разные величины t_i .

Таблица 5.8

Расчет логистического тренда

Год		$\zeta_i = \frac{100-10}{y_i-10} - 1$	$\ln \zeta_i$	t_i	$t_i * \ln \zeta_i$	ξ	$\hat{y}_i = \frac{100-10}{\hat{\zeta}_i+1} + 10, \%$	u_i	u^2_i
1955	15,0	17,00	2,833	-6,5	-18,416	16,803	15,1	-0,10	0,01
1956	17,0	11,86	2,473	-5,5	-13,601	10,854	17,6	-0,60	0,36
1957	20,0	8,00	2,079	-4,5	-9,357	7,017	21,2	-1,20 •	1,44
1958	26,4	4,49	1,501	-3,5	-5,255	4,535	26,3	+0,10	0,01
1959	33,5	2,83	1,040	-2,5	-2,601	2,931	32,9	+0,60	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	+1,30 •	1,69
1962	61,8	0,74	-0,305	0,5	-0,152	0,791	60,3	+1,50 •	2,25
1963	70,7	0,48	-0,728	1,5	-1,093	0,510	69,6	+1,10 •	1,21
1964	78,9	0,31	-1,183	2,5	-2,958	0,330	77,7	+1,20 •	1,44
1965	84,5	0,21	-1,570	3,5	-5,495	0,213	84,2	+0,30	0,09
1966	88,8	0,14	-1,951	4,5	-8,779	0,138	89,1	-0,30	0,09
1967	92,4	0,09	-2,383	5,5	-13,109	0,089	92,6	-0,20	0,04
1968	94,4	0,07	-2,713	6,5	-17,633	0,058	95,1	-0,70	0,49
Σ	778,4	-	-0,228	0,0	-99,325	-	773,3	-	13,89

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 то же самое, что и номера периодов от середины ряда.

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

Следующий шаг в освобождении параметров тренда от влияния случайного распределения положительных и отрицательных колебаний уровней на протяжении временного ряда можно сделать, применяя методику многократного скользящего выравнивания [20].

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 (см. табл. 5.1), состоящий из 26 уровней ($N = 26$), необходимо выравнивать по 15 уровням: сначала - с 1970 по 1984 г., затем - с 1971 по 1985 г. и т.д., скользя по ряду на 1 год, вплоть до последних 15 уровней с 1981 по 1995г. При этом каждый раз вычисляется среднегодовой прирост, например, b линейного тренда, а на концах будут года, то благоприятные для урожая зерновых, то неблагоприятные и по метеорологическим, и по

экономическим условиям. Всего получим 12 разных баз выравнивания по 15 лет; обозначив буквой n длину каждой базы, т.е. число уровней, по которым производится расчет параметра, а число таких баз расчета, укладываемых в ряд длиной N уровней, - буквой l , составим равенство:

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

В ряду из 26 уровней уложатся 12 баз по 15 уровней в каждой. Получим 12 значений среднегодового прироста 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 качественной тенденции. Если для однократного расчета параметра достаточным (минимальным) можно считать ряд из 9 - 11 уровней, а для достаточной степени взаимопогашения в средней величине следует иметь не менее 6—8 заниженных и завышенных значений параметра, т.е. минимальное значение будет $l \approx 6 - 8$, то минимальная длина исходного временного ряда, т.е. N , должна составлять $m + l - 1 = (9 + 6 - 1) - (11 + 8 - 1)$, или от 14 до 18 уровней. При более коротких рядах применение многократного скользящего выравнивания нецелесообразно.

Во-вторых, многократное выравнивание следует применять, если колеблемость исходных уровней достаточно существенная, скажем, коэффициент колеблемости (см. о нем в гл. 6) хотя бы не ниже 5%. При более слабой колеблемости искажения параметра невелики и при однократном выравнивании, поэтому нет необходимости «стрелять из пушки по воробьям», применяя сложную методику многократного выравнивания.

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

- 1) длина базы выравнивания, т.е. n , должна быть равна или кратна длине цикла;
- 2) число баз скользящего выравнивания, т.е. l , должно быть равно длине цикла.

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

Наконец, следует помнить, что искажающее влияние распределения случайных отклонений по длине временного ряда относится только к параметру динамики - среднегодовому (месячному и т.д.) приросту, ускорению и т.д., но не к среднему уровню ряда, не к свободному члену уравнения тренда. В связи с этим не нужно усреднять значения свободного члена по скользящим базам, а в качестве свободного члена для прямой следует взять общую среднюю величину уровней

исходного ряда, т.е. $a = \sum_{i=1}^N y_i / N$.

Для параболы, экспоненты и т.д. свободный член определяется расчетом на основе этой же средней величины. Так, для параболы:

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

Рассмотрим пример многократного скользящего выравнивания по данным табл. 5.1. Тренд, как показано ранее, линейный, но колеблемость существенная. Сделаем 12 скользящих баз расчета среднегодового прироста по 15 уровней в каждой. Вид таблицы для расчета без помощи ЭВМ приведен в приложении 1.

Результаты расчетов по многократному выравниванию

Скользящие базы расчета, годы	Среднегодовой прирост, b_i ц/га в год
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/га.

Среднее среднегодовое изменение (прирост):

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

Уравнение тренда:

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

где $t_i = 0,5$ в 1983 г.

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

Как видно из табл. 5.9, среднегодовой прирост по скользящим базам расчета сначала несколько возрастает, а затем снижается. 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. В связи с этим допустимо усреднение значений среднегодового прироста.

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

Особенно сложно оценить параметры тренда при несинусоидальных и сезонных колебаниях (см. разд. 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 пакетах статистических программ для ЭВМ. Данное пособие не позволяет по своему объему включить много таких методик, поэтому изложена одна - для линейного тренда в разд. 6.3.

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

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

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

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

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

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

$$\hat{y}_{80} - 2\hat{y}_{79} + \hat{y}_{78} = 314 - 2 \cdot 272 + 221 = -9 \text{ тыс. т в год за год}$$

$$\left(-9 \frac{\text{тыс. т}}{\text{год}^2} \right).$$

Темп роста - основной параметр экспоненциального тренда k . Для всех других типов тренда для вычисления темпа роста за некоторый период t следует разделить \mathcal{F}_m на \mathcal{F}_0 , т.е. на выровненный уровень базисного года. Так, для линейного тренда (см. табл. 5.3) темп изменения числа занятых в народном хозяйстве России в 1996 г. по сравнению с 1990 г. 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 ее можно прогнозировать или учитывать как фактор ошибки прогноза (гл. 10), т.е. сделать прогноз наиболее надежным и (или) точным;

3) на основе параметров и прогнозов колебаний можно рассчитать резервы, страховой запас, необходимый для преодоления 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 выходит за пределы статистического исследования и должна выполняться наукой, изучающей те явления и процессы, динамика которых отражена временным рядом.

6.1. Графическое отображение и основные свойства разных типов колебаний

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

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

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

Графическое изображение каждого из этих типов и описание основных свойств каждого типа колеблемости, во-первых, помогают по виду фактического ряда определить, каков преобладающий в нем тип колебаний, во-вторых, помогают экономисту, менеджеру, другому специалисту понять, какие последствия могут иметь колебания для его сферы деятельности и как с этими колебаниями (если нужно) бороться.

6.1.1. Пилообразная колеблемость

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

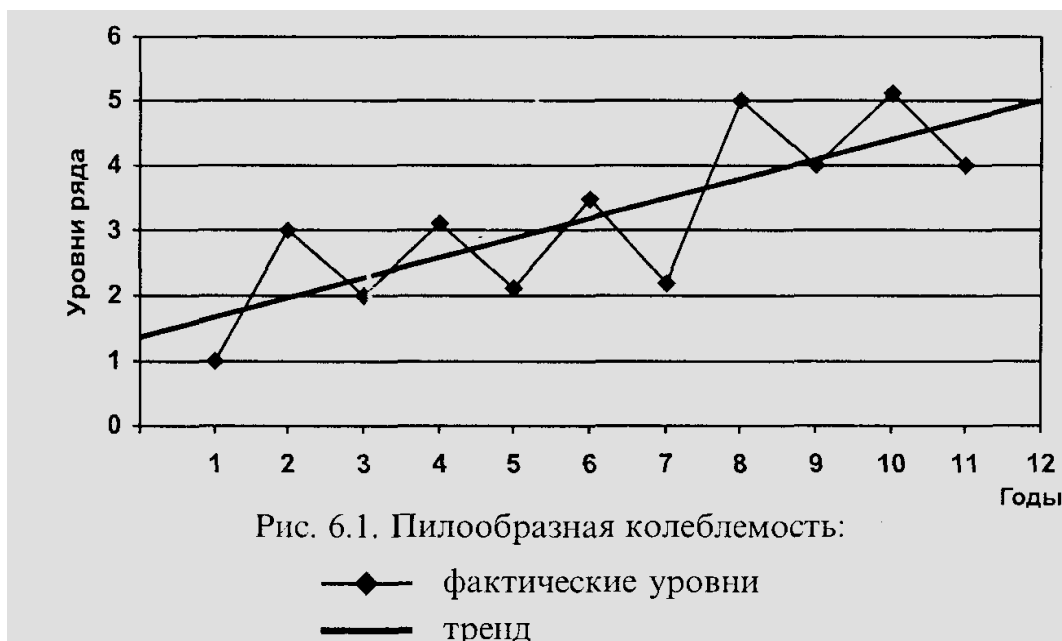


Рис. 6.1. Пилообразная колеблемость:

—◆— фактические уровни
 — тренд

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 надежное прогнозирование: если в данный период отклонение отрицательное, то в периоде 5 вперед оно будет положительным (данный период считать нулевым номером). Число положительных отклонений при достаточно большой длине ряда равно (точнее, стремится к равенству) числу отрицательных отклонений, а общее количество локальных экстремумов (отклонений от тренда, которые либо меньше, либо больше двух соседних по алгебраической величине) равно числу уровней.

Причины пилообразной колеблемости зависят как от внутренних факторов системы, так и от внешних. Внутренние для агротехнической системы причины пилообразной колеблемости урожайности - это колебания содержания питательных веществ в почве. Если по какой-либо внешней причине в данном году получен особо высокий урожай, то он выносит из почвы больше питательных веществ, чем в среднем в ней образуется за счет деятельности микроорганизмов, им 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 выравнивания периода) опять будет выше среднего и т.д. Конечно, рациональная агротехника подавит пилообразные колебания, например, увеличив после высокого урожая внесение в почву удобрений, и компенсирует повышенные этим 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 пилообразных колебаний как элемента во временном ряду можно, во-первых, по виду графика, во-вторых, подсчетом числа локальных экстремумов в ряду отклонений от тренда: чем это число ближе к числу уровней ряда, тем большую роль играют пилообразные колебания в их общем комплексе. Третий способ распознавания - по знаку и величине коэффициента автокорреляции отклонений от тренда I порядка, т.е. со сдвигом (лагом) на 1 год.

Коэффициент автокорреляции отклонений имеет формулу

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

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

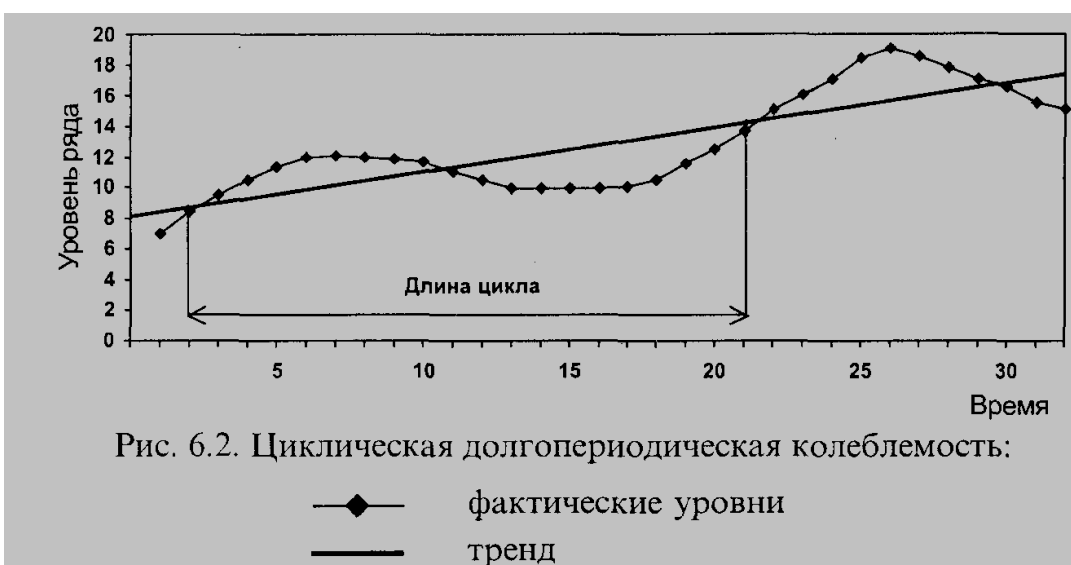
Чем ближе коэффициент автокорреляции к -1 , тем большую роль играет пилообразная составляющая в общей колеблемости изучаемого временного ряда.

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 1 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.



Для прогнозирования циклическая колеблемость благоприятна, особенно если длина цикла строго постоянна. Прогноз на любой будущий период состоит

из прогноза тренда и циклического отклонения от него, соответствующего фазе цикла в прогнозируемый период. Например, зная, что солнечная активность имеет 10-11-летнюю периодичность и что предыдущий цикл имел максимум в 1990-1991 гг., можно уверенно прогнозировать следующий максимум на 2000-2001 гг.

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

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

где l - длина цикла.

Причиной циклической колеблемости является какая-либо основная сила, влияющая на уровень изучаемого явления. Иначе говоря, есть главный фактор, вызывающий колебания. Сезонные колебания температуры, осадков, а следовательно, 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 не известна.

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 пересечении 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 абсолютной 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 отрицательных отклонений, 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 «интерференция колебаний» (термин, заимствованный из физики).

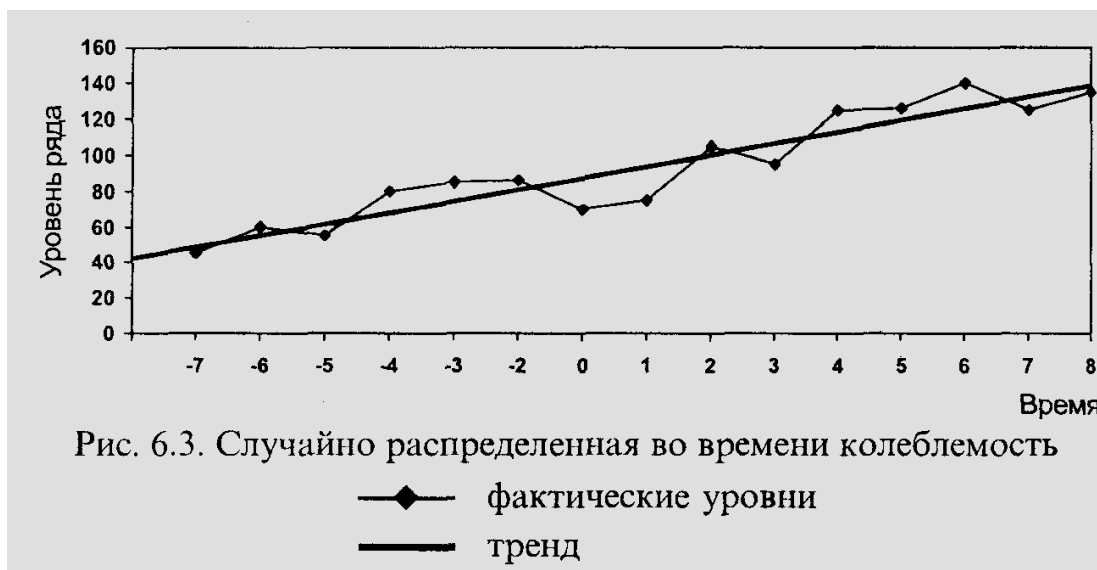


Рис. 6.3. Случайно распределенная во времени колеблемость

—◆— фактические уровни
— тренд

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

- из-за хаотического чередования знаков отклонений от тренда их взаимопогашение наступает только на достаточно длительном периоде, а на коротких отрезках отклонения могут накапливаться, 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 еще от многих других факторов. Практика статистических исследований колеблемости урожаев показала, что преобладают именно случайно распределенные колебания. Наличие множества примерно равноправных и независимых факторов означает также, что нельзя существенно уменьшить колеблемость, воздействуя только на какой-либо отдельный фактор. Необходимо, если это возможно, регулировать все основные факторы, как, например, и делается в защищенном грунте (теплицах).

Распознать случайно распределенную во времени колеблемость по виду графика труднее, чем два других типа колебаний. Число локальных экстремумов может также колебаться. В среднем, как доказал английский статистик М. 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, следовательно, эта гипотеза не может быть отклонена.

Коэффициент автокорреляции отклонений от тренда при случайно распределенной колеблемости стремится к нулю при $n \rightarrow \infty$. Если ряд состоит

менее чем из 19-22 уровней, коэффициенты автокорреляции I порядка, не превышающие 0,3 по абсолютной величине, свидетельствуют о преобладании случайной компоненты в общем комплексе колебаний. В случае, изображенном на рис. 6.3, $r_U^a = -0,025$.

6.2. Измерение показателей силы и интенсивности колебаний

Показатели силы и интенсивности колебаний аналогичны по построению, по форме показателям силы и интенсивности вариации признака в пространственной совокупности. По существу они отличаются тем, что показатели вариации вычисляются на основе отклонений от постоянной средней величины, а показатели, характеризующие колеблемость уровней временного ряда, - по отклонениям отдельных уровней от тренда, который можно считать «подвижной 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 1 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 того, что в нем встретится особенно большое отклонение от тренда. Поэтому с увеличением длины изучаемого периода возрастает в среднем и амплитуда колебаний в отличие от всех других показателей колеблемости, которые не зависят от длины ряда.

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

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

Знак t отличает указанный и все последующие показатели от аналогичного среднего по модулю отклонения от постоянной средней величины, меры силы вариации в пространственной совокупности. Средний модуль отклонений измеряется в тех же единицах, что 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 не пригодны для прогнозирования доверительных границ возможных колебаний с заданной вероятностью (см. гл. 10).

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

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

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

Если же речь идет о вычислении *оценки генерального показателя*

колеблемости, а исходный временной ряд рассматривается как *выборка* из генерального ряда, продолжаемого и в прошлое и в будущее, то следует учитывать потерю степеней свободы колеблемости и применять показатель:

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

где p - число параметров в уравнении тренда.

Причину учета числа параметров тренда можно проиллюстрировать следующими примерами.

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

6.2.2. Показатели относительной интенсивности колебаний

Показатели относительной интенсивности вариации рассчитываются как отношение ее абсолютных показателей к постоянной средней величине, относительной интенсивности колебаний - как отношение индивидуальных отклонений отдельного периода к уровню тренда за этот же период, а обоб-

шающие показатели - как отношение 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 том же 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 считать не сильным, а умеренным неурожаем.

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

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

- для оценки генеральной величины и прогнозов или

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

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

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

В заключение необходимо подчеркнуть, что любая погрешность в определении типа тренда или при 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 заработной платы и среднедушевого дохода в декабре (13-я 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 температуры (табл. 6.1).

Таблица 6.1

Динамика средних месячных температур в Ленинграде - Санкт-Петербурге

Месяц	1995г. Y_{i95}	1996г. Y_{i96}	1997г. Y_{i97}	В среднем за 1988-1997г. \bar{y}_i	В % к среднегодо- вым температурам, $\frac{\bar{y}_i}{\bar{y}_i} \cdot 100\%$	$(\frac{\bar{y}_i}{\bar{y}_i} - \frac{\bar{y}_i}{\bar{y}_i})^2$
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	0,0	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	+6,46	+5,19	+5,75	+5,93	100,0	813

Данные табл. 6.1 позволяют 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 показали, что тенденция существует, но слабая: среднегодовой абсолютный прирост температуры составил $0,0255^\circ$ 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:

$$\sigma_{uz} = \sqrt{\frac{\sum_{i=1}^{12} (\bar{y}_i - \bar{\bar{y}})^2}{12}} = \sqrt{\frac{813}{12}} = 8,23^{\circ}.$$

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 $\sigma \approx 3^{\circ}$, а в Узбекистане (г. Бухара) $\sigma \approx 12^{\circ}$.

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

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

Выше 15°C - дни считаются летними, ниже 0°C - зимними.

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

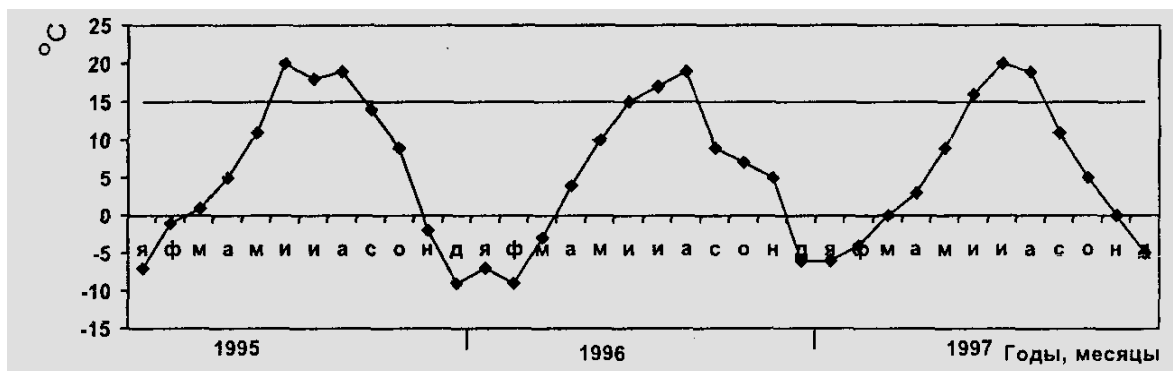


Рис. 6.4. Колебания месячной температуры воздуха в Санкт-Петербурге за 1995-1997 гг.

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

В качестве примера такого вида сезонных колебаний рассмотрим динамику реализации свиней после откорма, имеющую пик в 4-м квартале года (табл. 6.2), и сезонные колебания затрат труда на развивающемся предприятии с двумя пиками в мае-июне и в августе-сентябре (табл. 6.3).

Таблица 6.2

Расчет параметров тренда при асимметричных сезонных колебаниях

Год	Квартал	Уровень, y_i	t_i	$y_i t_i$	Тренд		$i_{\text{сез}}$
					первичный, €_i	корректированный, $\text{€}_i \text{ корр}$	
1	1-й	20	-4	-80	22	29	0,690
	2-й	26	-3	-78	27	32	0,812
	3-й	28	-2	-56	32	36	0,778
	4-й	60	-1	-60	38	39	1,538
2	1-й	30	0	0	43	43	0,698
	2-й	38	1	38	48	46	0,826
	3-й	42	2	84	53	50	0,840
	4-й	100	3	300	58	53	1,887
3	1-й	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 $t_i = 0$ в 1-м квартале II года.

При наличии сезонных колебаний, не имеющих синусоидального характера, особенно для рядов, имеющих резкий 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 будем считать, что резкий пик уровней приходится ежегодно на 4-й 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 значения t_i , равные соответственно -4; 0; 4, в среднем нулевые веса; низкие значения уровней вторых кварталов приходятся на значения 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 на 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:

$$y_{i \text{ кор}} = 42,67 + 3,53 t_i, \quad t = 0 \text{ в 1-м квартале II года.}$$

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 бы 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 для 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	Ме- сяц, i	Уров- ни, тыс. ч, Y _{ij}	Тренд f _i	i _{сезi}	f _i · i _{сезi}	Y _i - f _i × i _{сезi} = u _{сл}	u ² _{сл}	f _i - f _i × i _{сезi} = u _{сез}	u ² _{сез}
I	Я	17	40	0,425	17	0	0	23	529
	Ф	16	42	0,381	16	0	0	26	676
	М	22	43	0,512	22	0	0	21	441
	А	35	44	0,795	34	1	1	10	100
	Май	76	45	1,689	72	4	16	-27	729
	И	67	46	1,457	67	0	0	-21	441
	Ил.	57	47	1,213	57	0	0	-10	100
	Ав.	78	48	1,625	80	-2	4	-32	1024
	С	84	49	1,714	78	6	36	-29	841
	О	64	50	1,280	60	4	16	-10	100
	Н	35	51	0,686	37	-2	4	14	196
	Д	24	52	0,462	23	1	1	29	841
II	Я	20	53	0,377	23	-3	9	30	900
	Ф	19	55	0,345	21	-2	4	34	1156
	М	29	56	0,518	29	0	0	27	729
	А	42	57	0,737	44	-2	4	13	169
	Май	96	58	1,655	93	3	9	-35	1225
	И	91	59	1,542	86	5	25	-27	729
	Ил.	71	60	1,183	72	-1	1	-12	144
	Ав.	102	61	1,672	101	1	1	-40	1600
	С	102	62	1,645	99	3	9	-37	1369
	О	76	63	1,206	75	1	1	-12	144
	Н	48	64	0,750	47	1	1	17	289
	Д	29	65	0,446	30	-1	1	35	1225
III	Я	33	67	0,493	29	4	16	38	1444
	Ф	31	68	0,456	27	4	16	41	1681
	М	36	64	0,522	35	1	1	34	1156
	А	56	70	0,800	54	2	4	16	256
	Май	107	71	1,507	114	-7	49	-43	1849
	И	104	72	1,444	106	-2	4	-34	1156
	Ил.	92	73	1,260	88	4	16	-15	225
	Ав.	129	74	1,743	123	6	36	-49	2401
	С	112	75	1,493	120	-8	64	-45	2025
	О	86	70	1,132	91	-5	25	-15	225
	Н	59	77	0,766	56	3	9	21	441
	Д	40	79	0,506	37	3	9	42	1764
IV	Я	35	80	0,438	34	1	1	46	2116
У		2220	2221	—	2197	23	393	+24	32436

Примечание. Я- январь, Ф - февраль, М - март, А - апрель, И - июнь. Ил. - июль, Ав. - август, С - сентябрь, О - октябрь, Н - ноябрь, Д - декабрь.

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:

$\bar{i}_{\text{сез корр}}$	Месяц	$\bar{i}_{\text{сез}}$
0,429	Январь	$(0,425 + 0,377 + 0,493 + 0,438) : 4 = 0,433$ и т.д.
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
<hr/>		
$\sum_1^{12} = 12,025$		$\sum_1^{12} = 12,146$

Сумма индексов составила 12,146, хотя средний индекс должен быть равен единице. Следует откорректировать 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, $\bar{y}_i - \bar{y}_{\text{сез}_i}$ округлены в табл. 6.3 до целых. То, что $\sum_{i=1}^{37} \bar{y}_i - \bar{y}_{\text{сез}_i}$ меньше

$\sum_{i=1}^{37} \bar{y}_i$, не является недостатком расчета: дело в «лишнем» январе, уровень

которого с учетом сезонного колебания в среднем за три года ниже тренда на 30, в

результате даже с учетом этого остается небольшой избыток $\sum_{i=1}^{37} \bar{y}_i - \bar{y}_{\text{сез}_i}$,

объясняемый округлением. Избыток на 6 при сумме уровней 2220, разумеется, несуществен.

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

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

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, так как 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:

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

Сезонные колебания за год имели 11 степеней свободы вариации, но в ряду отклонений $\bar{y}_i - \bar{y}_{\text{сез}_i}$ 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 $V(t)_{\text{сез}} = 31,35/60 = 0,522$, или 52,2%. Сезонная колеблемость сильная.

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

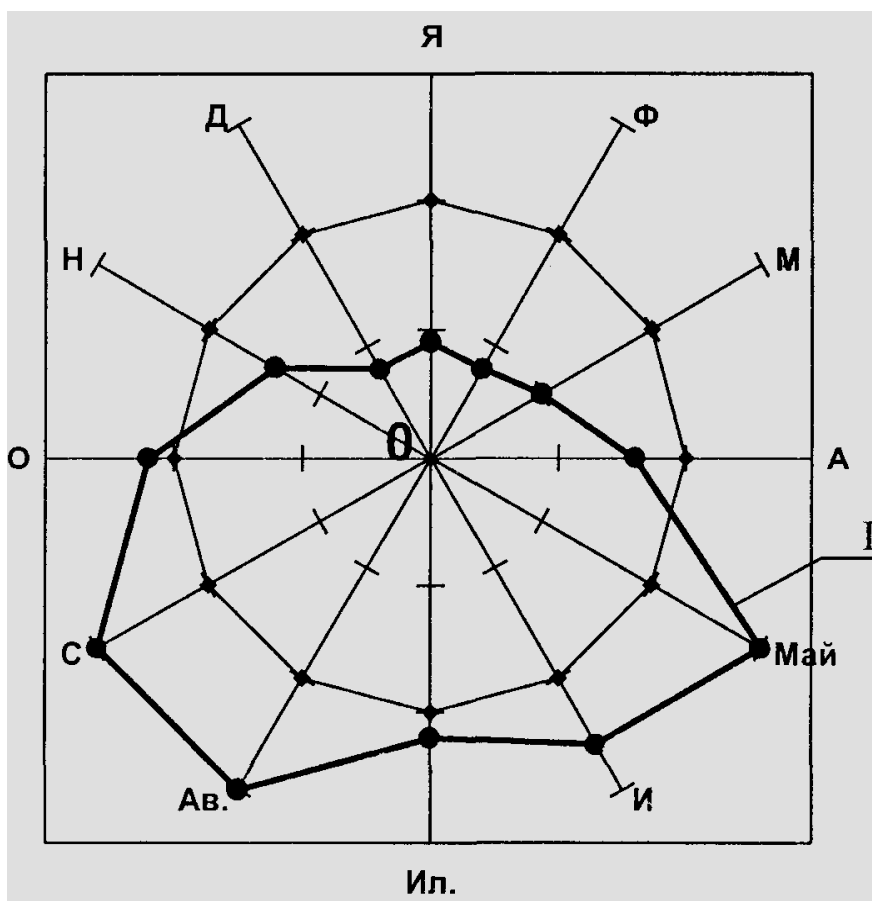


Рис. 6.5. Сезонные колебания затрат труда на сельскохозяйственном предприятии

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

Преобразование сезонных колебаний в ряд Фурье

Месяц, i	Надой, \bar{y}_i , кг/гол	t_i , гра- дусов	$\cos t_i$	$\sin t_i$	$\bar{y}_i \cos t_i$	$\bar{y}_i \sin t_i$	f_i	$\bar{y}_i - f_i$
Январь	230	0	1,000	0,000	230	0	229	1
Февраль	260	30	0.866	0.500	225	130	264	-4
Март	315	60	0.500	0.866	157.5	273	311	4
Апрель	352	90	0.000	1.000	0	352	357	-5
Май	392	120	-0.500	0.866	-196	339	391	1
Июнь	403	150	-0.866	0.500	-349	201.5	404	-1
Июль	398	180	-1.000	0.000	-398	0	391	7
Август	352	210	-0.866	-0.500	-305	-176	356	-4
Сентябрь	308	240	-0.500	-0.866	-154	-267	309	-1
Октябрь	262	270	0.000	-1.000	0	-262	263	-1
Ноябрь	225	300	0.500	-0.866	112.5	-195	229	-4
Декабрь	223	330	0.866	-0.500	193	-111,5	216	7
У	3720*	-	0,000	0,000	-484	284	3720	0

***Среднемесячный надой составил 310 кг на 1 корову (3720 : 12).

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

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

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

Смысл уравнения состоит 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 удается отобразить, как показывает табл. 6.4 (графа \hat{y}_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:

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

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:

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

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 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 $8 \times 8 = 64$ уровня. Такие ряды анализируются нечасто, а значит, нет и условий для расчета тренда мер колеблемости.

Положение отчасти спасает то, что для вычисления тренда колеблемости вовсе необязательно, чтобы за весь изучаемый период существовал единый тренд уровней показателя. Вполне допустимо для расчета тренда колеблемости объединить отрезки времени с разными по типу трендами или с *кусочно-линейным* трендом. От изменения скорости роста или даже типа роста, или направления тенденции динамики колеблемость зависит мало или совсем не зависит. Но и с учетом этой ее особенности измерить тренд колеблемости по ряду отдельных отрезков времени сложно. При длине первичного ряда в 15-20 уровней получается всего 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 изменения силы колебаний и приближенно измерить тренд $S(t)$ все же возможно. Покажем применение этого метода на примере временного ряда урожайности зерновых

культур во Франции (см. разд. 5.1). 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 г. Итого получаем 16 скользящих значений показателей колеблемости, которые и выравниваем по уравнению прямой.

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

$$\mathbf{S}(t) = 3,42 - 0,1235 \times t_i; \quad t_i = 0,5 \quad \text{в} \quad 1983\text{г.}$$

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 $\mathbf{bs}(t)$ со средней ошибкой репрезентативности. Это необходимо для применения полученного тренда силы колебаний в прогнозировании урожайности, т.е. для распространения 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 - $\mathbf{bs}(t)$, т.е.

$$m_{b_{s(t)}} = \frac{S(t)_{S(t)}}{\sqrt{\sum_1^{16} t_i^2}}.$$

Вычисление тренда показателя колеблемости

Год	U_i	U_i^2	Скользящие значения $S(t)_i$	t_i	$S(t)_i \times t_i$	$\mathcal{S}(t)_i$	$U_{S(t)_i}$	$U^2_{S(t)_i}$
1970	0.6	0.361	-	-	-			
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	-	-	-			
У	-	301,37	54,65	0	-42,0	54,62	-	1,2185

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

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

$$S(t)_{S(t)} = \sqrt{\frac{\sum_1^{16} U_{S(t)}^2}{16-1}} = \sqrt{\frac{1,2185}{15}} = 0,285;$$

$$m_{b_{S(t)}} = \frac{S(t)_{S(t)}}{\sqrt{\sum_1^{16} t_i^2}} = \frac{0,285}{\sqrt{340}} = 0,0542.$$

$$\frac{b_{S(t)}}{m_{b_{S(t)}}} = \frac{0,1235}{0,0542} = 2,28$$

Критерий Стьюдента равен отношению

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 был рассмотрен в разд. 6.1. Аналогично строятся и формулы коэффициентов автокорреляции со сдвигом в два, три и т.д. периодов времени. В общем виде коэффициент автокорреляции порядка m , т.е. со сдвигом на m периодов времени, вычисляется по формуле:

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

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 $(m - 1)$ 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

Регион	Тренд	Коэффициенты автокорреляции с лагами							
		1 год	2 года	3 года	4 года	5 лет	6 лет	7 лет	8 лет
Северный	$\bar{y} = 8,61 + 0,27t - 0,0068t^2$	-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

Источник. Развитие рынка зерна в России (тенденции производства зерна в Российской Федерации за 1970-1996 гг.). - М.: ЦЭК при Правительстве Российской Федерации, 1997. - С. 111.

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, которая 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 Санкт-Петербурге за последние 40 лет и измерен линейный тренд. Но нас интересует среднегодовой прирост не только как факт, относящийся к 1957-1997 гг., но и как характеристика процесса потепления климата города вообще для распространения ее на будущее, например, 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 вносит неизбежный элемент случайности во все показатели динамики, если их хотят распространить на генеральную совокупность.

Само наличие тренда или его отсутствие на изучаемом отрезке времени может быть доказано лишь с некоторой вероятностью, для чего используются специальные критерии. При изучении случайной колеблемости очень важно определить вероятность крайних, 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 путем 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 оцепить 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:

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

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 достигает 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 Санкт-Петербурге за 1957-1997 гг. (табл. 7.1).

Таблица 7.1

Среднегодовая температура воздуха в Санкт-Петербурге, °С

Год	Температура, y_i	Год	Температура, y_i	Год	Температура, y_i	Год	Температура, 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:

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

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 значения 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 всей 21 выборки, 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;

$\sum_{i=1}^{21} t^2_i$ - 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 г-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. Необходимо, 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

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

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:

$$\overline{(t_i - \bar{t})^4} - \left[\overline{(t_i - \bar{t})^2} \right]^2.$$

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 измерении периодов от середины ряда при $n = 8$. Имеем:

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

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

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:

$$\hat{y}_i = 4004 \cdot 1,195^t.$$

В логарифмическом виде

$$\ln \hat{y}_i = 8,295 + 0,1783t_i.$$

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

Среднее квадратическое отклонение логарифмов:

Таблица 7.2

Определение отклонений логарифмов уровней от логарифмов тренда

Год	$\ln y_i$	$\ln \hat{y}_i$	Отклонение $U_{\ln y_i}$	$U^2_{\ln y_i}$	t_i	t^2_i
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
У	49,770	49,770	0	0,000859	0,0	17,50
* Оценка						

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

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 четырех 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 приведен для показа методики надежности экспоненциального тренда, а не для 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 неприменим. В таких случаях можно, во-первых, проверять сам факт наличия какого-либо тренда путем сравнения средних уровней за первую и за вторую половины периода, во-вторых, с помощью обычной методики 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 $a + b t_i$, 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. $S(t) / \sqrt{n}$, а в любой иной точке тренда его средняя ошибка вычисляется по формуле

$$m_{\hat{y}_m} = \sqrt{m_a^2 + m_b^2 \cdot t_m^2} = \sqrt{\frac{S^2(t)}{n} + \frac{S^2(t) \cdot t_m^2}{\sum_{i=1}^n t_i^2}} = S(t) \sqrt{\frac{1}{n} + \frac{t_m^2}{\sum_{i=1}^n t_i^2}}$$

- для однократного выравнивания и при $t_i = 0$ в середине ряда. При нумерации периодов времени от начала ряда вместо t_i в формулу следует подставить величину $(t_i - \bar{t}); (t_m - \bar{t})^2$.

При многократном скользящем определении параметра b второе слагаемое подкоренного выражения примет вид:

$$\frac{t_m^2}{l \sum_{i=1}^n t_i^2}$$

где n - длина одной базы расчета тренда;

l - число баз.

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

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

Для середины ряда - 1977 г. - средняя ошибка тренда составила:

$$1,121 \cdot \sqrt{\frac{1}{41}} = 0,175^\circ.$$

А для крайних уровней-1957г. и 1997г.-

$$\text{средняя ошибка тренда составляет: } 1,121 \cdot \sqrt{\frac{1}{41} + \frac{20^2}{21 \cdot 707}} = 0,254^\circ.$$

Таким образом, ошибка тренда возрастает от середины базы его расчета (середина ряда) к его краям, образуя конусообразную зону вероятных значений

генерального тренда.

Если эту зону мы хотим определить с достаточно большой вероятностью, то среднюю ошибку следует умножить на величину t -критерия Стьюдента для соответствующей вероятности. Границы доверительной зоны тренда среднегодовой температуры с вероятностью 0,95 изображены на рис. 7.1.

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 расширяется при этом гораздо сильнее, чем для линейного тренда.

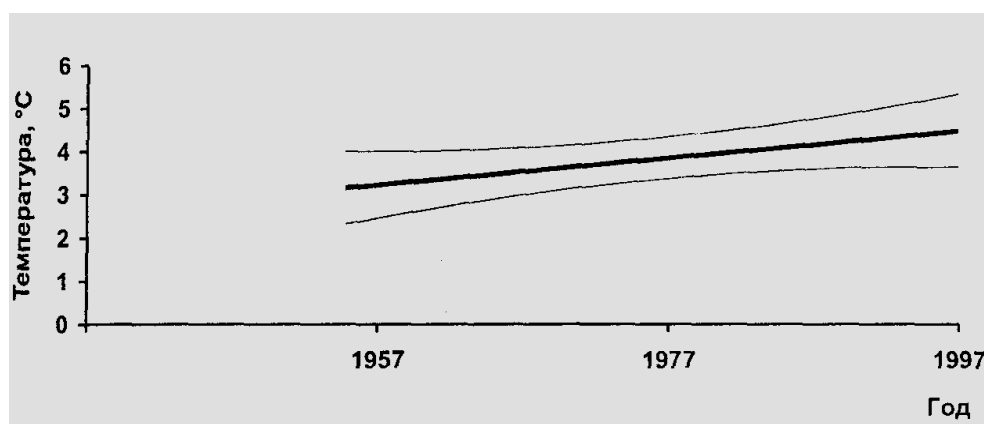


Рис. 7.1. Доверительные границы генерального тренда среднегодовой температуры воздуха в Санкт-Петербурге

————— средний тренд

————— границы тренда с вероятностью 0,95

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 применять для вероятностных оценок среднего квадратического отклонения уровней ряда от тренда и коэффициента колеблемости.

Средняя ошибка репрезентативности выборочной оценки генерального среднего квадратического отклонения от тренда при их нормальном распределении имеет 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 вероятностей 0,95 и 0,99, можно свести в одну таблицу, получаем готовую таблицу для оценки надежности отличия генерального среднего квадратического отклонения уровней от нуля (табл. 7.3).

Таблица 7.3

Вероятность отличия колеблемости $S(t)$ от нуля

n	$\bullet jin$	Табличный критерий t для вероятности			Вывод о надежности отличия
		0,90	0,95	0,99	
2	2,20	2,92	4,30	9,92	Значительно ниже 0,9
3	2,45	2,35	3,18	5,84	
5	3,16	2,02	2,57	4,03	Выше 0,9, но ниже 0,95
8	4,0	1,86	2,30	3,35	Выше 0,95, но ниже 0,99
9	4,24	1,83	2,26	3,25	Выше 0,99
10	4,47	1,81	2,23	3,17	То же
12	4,90	1,78	2,18	3,06	Практически достоверно
15	5,48	1,75	2,13	2,95	
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 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:

$$3,54 \pm 2,80 \cdot \frac{3,54}{\sqrt{2 \cdot 26}} = 3,54 \pm 1,37 \text{ ц/га}.$$

Ввиду довольно значительной силы колебаний, доверительный интервал оценки генерального среднего квадратического колебания также довольно широк, Ошибка возрастает прямо пропорционально силе колеблемости и росту надежности оценки, а уменьшается обратно пропорционально корню квадратному из числа уровней ряда.

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

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

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

Например, коэффициент 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 $6,9\% \pm 2,06 \cdot 0,96\%$,

или 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 распределение вероятностей отклонения от тренда близко к нормальному закону. В нашем примере распределение отклонений от тренда среднегодовой температуры воздуха в Санкт-Петербурге близко к нормальному (табл. 7.4).

Таблица 7.4

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

Отклонение, градус	f_i	t_i -критерий	Вероятность P_i	f_{T_i}	$\frac{(f_i - f_{T_i})^2}{f_{T_i}}$
Ниже $-1,2^\circ$	5	от $-\infty$ до $-1,07$	0,1423	5,8	0,11
От $-1,2$ до $-0,4^\circ$	8	от $-1,07$ до $-0,36$	0,2171	8,9	0,10
От $-0,4$ до $0,4^\circ$	15	от $-0,36$ до $0,36$	0,2812	11,5	1,06
От $0,4$ до $1,2^\circ$	8	от $0,36$ до $1,07$	0,2171	8,9	0,10
Выше $1,2^\circ$	5	от $1,07$ до $+\infty$	0,1423	5,8	0,11
Итого	41	-	1	40,9	1,48

t_i - нормированное отклонение 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_{T_i} 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 равно 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 Центрального статистического управления (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 «пределной 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, равно 3,54 ц/га. Находим *вероятности рисков*, т.е. что отклонение от тренда вниз (неурожай) превышает уровни -5 ц/га; -7 ц/га; -10 ц/га; -12 ц/га (табл. 7.5).

Вероятность P равна половине разности между единицей и $F(t)$, т.е. применяется *односторонний критерий* (иногда в литературе приводится готовая таблица вероятностей именно этого критерия). 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 (неурожаев) зерновых во Франции

Отклонение вниз от тренда, U_i ц/га	Нормированное отклонение $t = U_i : S(t)$	Вероятность отклонения, 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

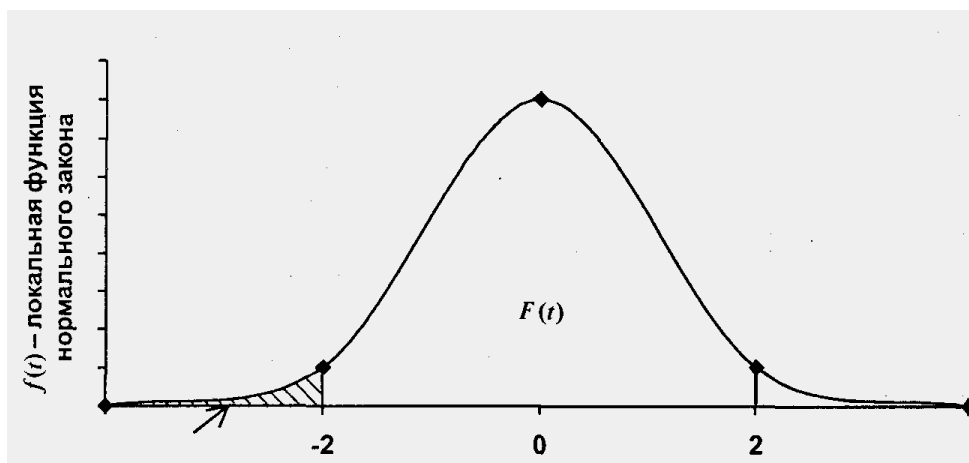


Рис. 7.2. Вероятность отрицательного отклонения, большего по величине, чем заданная граница

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

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 и получаем вероятные ошибки приблизительно с вероятностью 0,95 или на 3 и тогда получаем 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 частости 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*. 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:

$$\overline{S(t)^2}_{\text{арифм}} = \frac{\sum_{i=1}^k S(t)_i^2 \cdot n_i}{\sum_{i=1}^k n_i},$$

где k - число дисперсий;

n_i - их веса, число уровней в подпериодах.

Взвешенная геометрическая средняя:

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

Критерий Бартлетта имеет вид:

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

его средняя ошибка:

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

Отношение M/C имеет распределение χ^2 (хи-квадрат) с числом степеней свободы $k - 1$.

При сравнении двух дисперсий и равном числе уровней в каждом подпериоде (средние будут невзвешенные) формулы упрощаются:

$$M = 2 \ln \frac{S(t)_{\text{арифм}}^2}{S(t)_{\text{геом}}^2} \cdot \sum_{i=1}^k n_i;$$

$$C = 1 + \frac{\frac{1}{n} - \frac{1}{2n}}{3}.$$

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:

$$S^2_{\text{арифм}} = 12,6,$$

а их геометрическая средняя:

$$S^2_{\text{геом}} = \sqrt{18,1 \cdot 7,1} = 11,34;$$

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

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

$M/C = 4,57$. 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 ТRENDA

Устойчивость временного ряда - понятие многоплановое. Его следует рассматривать с двух позиций:

- устойчивости уровней временного ряда;
- устойчивости тенденции (тренда).

Вопрос определения понятия *устойчивости* 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. Тенденцией характеризуется процесс эволюции. В явном виде невозможно видеть все причины, порождающие тенденцию (тренд). Полное разделение элементов случайного и необходимого существует только в виде научной абстракции.

Расчленение динамического ряда на составляющие элементы - условный описательный прием. Тем не менее, несмотря на взаимозависимость тенденции и колеблемости, решающим фактором, обуславливающим тенденцию, является целенаправленная деятельность человека, а главной причиной колеблемости - изменение условий жизнедеятельности. Исходя из вышеизложенного можно отметить следующее. Устойчивость не означает обязательное повторение одинакового уровня из года в год; такое понимание устойчивости приравнивало бы ее к застойному состоянию изучаемого явления. Слишком узким и жестким было бы понятие устойчивости ряда - как полное отсутствие в динамическом ряду всяких колебаний, так как полностью устранить влияние случайных факторов на показатель невозможно. Сокращение колебаний уровней ряда - одна из главных задач при повышении устойчивости, но этим она не исчерпывается, необходимо развитие явления. Отсюда и следует, что устойчивость временного ряда - понятие не простое, а многоплановое.

Устойчивость временного ряда - это наличие необходимой тенденции изучаемого статистического показателя с минимальным влиянием на него неблагоприятных условий.

Из этого 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 существа исследуемого процесса, но содержание понятия *устойчивость* будет оставаться 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:

$$i_{\bar{y}} = \frac{\bar{y}_{\text{благ}}}{\bar{y}_{\text{неблаг}}}, \text{ или } i_{\bar{y}} = \frac{\bar{y}_{\text{б}}}{\bar{y}_{\text{н}}} \quad (8.2)$$

- 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):

среднее линейное отклонение

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

среднее квадратическое отклонение

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

где y_i - фактический уровень;

\tilde{y}_i - выровненный 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 скольжения при многократном аналитическом выравнивании дает информацию о снижении или о 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**:

$$PR = W_M - W_m,$$

где $W_M = \max(W_2, \dots, W_{t-1});$
 $W_m = \min(W_2, \dots, W_{t-1});$

$$W_t = \frac{|X_t - X_{t-1}|}{X_{t-1}} \cdot 100, \quad t = 1, \dots, n.$$
(8.5)

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

2. Показатель *скользящее среднее* (Moving Average) - **МА**, который оценивает величину среднего отклонения от уровня скользящих средних:

$MA = \frac{\sum_{t=r+1}^n \left[\frac{x_i - x_t}{x_t} \right]}{n + 1 - m},$ $X_t = \frac{\sum_{i=t-r}^{t+r} x_i}{m},$ <p>где $r = (m - 1)/2;$ m – период скользящей средней.</p>	(8.6)
---	-------

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

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

Бланфорд и Оффат, анализируя вышперечисленные коэффициенты,

отмечают их хорошую согласованность относительно коэффициента Спирмена.

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

коэффициент линейной колеблемости

$$V_y^d(t) = \frac{d_y(t)}{\bar{y}}, \quad (8.8)$$

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

$$(8.9)$$

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 процентах. 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) не превзойдет средней величины колебаний $S_y(t)$, составляет лишь 0,68, если распределение колебаний по их величине близко к нормальному.

Например (см. гл. 6, разд. 6.2.2), коэффициент колеблемости 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 Спирмена K_p [3, с. 39]:

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

где d - разность рангов уровней изучаемого ряда (P_y) и рангов номеров периодов или моментов времени в ряду (P_t);

n - число таких периодов или моментов.

Для определения коэффициента Спирмена величины уровней изучаемого явления y^{\wedge} 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 Спирмена:

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

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

j - номера связок по порядку (см. нижнюю формулу);

A_j - число одинаковых рангов в j -й связке (число одинаковых уровней).

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

Чем ближе Kp к $+1$, тем ближе рост уровней к непрерывному, выше устойчивость роста. При $Kp = 0$ рост совершенно неустойчив. При отрицательных значениях чем ближе Kp к -1 , тем устойчивее снижение изучаемого показателя. В рассмотренном ранее ряду динамики урожайности зерновых во Франции за 1970-

1995 г. коэффициент Спирмепа составил 95,62%.

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

$$Kp = \frac{12 \sum_{i=1}^n P_{t_i} \cdot P_{y_i}}{n^3 - n} - \frac{3(n+1)}{n-1}, \quad (8.13)$$

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;

\vec{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 $\vec{y}_i = a + bt_i$, 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 -

$F(t):2 = 0,5 - 0,34 = 0,16$. 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 $b = -0,4$ $Sy(t)$, 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:

$\tilde{y} = ak^t \cdot k$ - величина отвлеченная, притом всегда положительная (знакопеременные уровни здесь не рассматриваются).

Недопустимо сопоставлять темпы с абсолютным показателем колеблемости $Sy(t)$, 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 выравнять его тоже по экспоненте, чтобы определить величину среднегодового темпа (в размах) величины колебаний, т.е. показатель $Ks(t)$. Так как для одноразового надежного вычисления показателя колеблемости уже необходимо иметь не менее 11-15 уровней, то для получения динамического ряда $Sy(t)$ и его среднегодового темпа изменения необходим динамический ряд исходных уровней значительной длины (не менее 11-15 плюс еще 9-11), т.е. более 20 уровней, а лучше около или более 30. Далеко не всегда можно получить такой длинный ряд достаточно однокачественных уровней с единым трендом.

Сопоставляя темпы роста уровней ряда с темпами изменения колеблемости, получим *показатель опережения*:

$$O_{kэ} = \frac{\bar{k}}{k_{S_y(t)}}. \quad (8.16)$$

Если $O_{kэ} > 1$, это свидетельствует, что уровни ряда в среднем растут быстрее колебаний (или снижаются медленнее колебаний). В таком случае, как понятно без доказательства, коэффициент колеблемости уровней будет снижаться, а коэффициент устойчивости уровней повышаться. Если $O_{kэ} < 1$, наоборот, колебания растут быстрее уровней тренда и коэффициент колеблемости растет, а коэффициент устойчивости уровней снижается. Таким образом, величина $O_{kэ}$ определяет направление динамики коэффициента устойчивости уровней.

Параболический тренд $\tilde{y}_i = a + bt_i + ct_i^2$ имеет два динамических параметра: среднегодовой прирост b и половину ускорения прироста c . Величина b в параболе не является константой, и для построения показателей комплексной устойчивости W нужно взять среднюю за весь ряд величину \bar{b} . В остальном интерпретация та же, что и для прямой. Второй показатель - половину ускорения c или ускорение прироста $2c$ - логично сопоставлять уже не с самой величиной колеблемости $S_y(t)$, а с ее среднегодовым приростом $b_{S_y(t)}$, полученным по достаточно длинному ряду путем выравнивания показателей $S_y(t)$, скользящих или следующих друг за другом. Имеем показатель

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

Интерпретация показателя O_c такова: если $O_c > 1$; значит, положительное ускорение (прирост абсолютного прироста уровней) больше, чем прирост среднего квадратического отклонения от тренда. Значит, отношение прироста уровней к среднему отклонению от тренда станет увеличиваться, т.е. показатель K будет возрастать, что 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:

$$1. c > 0; b_{S_y(t)} > 0; 2c > b_{S_y(t)}.$$

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.

$$2. c > 0; b_{S_y(t)} > 0; 2c < b_{S_y(t)}.$$

Хотя прирост уровней возрастает (ускоряется), но колеблемость растет еще быстрее, а, значит, показатель устойчивости тенденции K снижается. Это менее благоприятный тип динамики, чем случай 1.

3. $c > 0; b_{S_y(t)} < 0; 2c > b_{S_y(t)}$ - очевидная ситуация. Эта комбинация означает, что прирост уровней растет, а колеблемость снижается. Ясно, что при этом показатель устойчивости тенденции K возрастает.

4. $c > 0; b_{S_y(t)} < 0; 2c < b_{S_y(t)}$ - нереальная комбинация, третье неравенство противоречит двум первым.

5. $c < 0; b_{S_y(t)} > 0; 2c > b_{S_y(t)}$ - также нереальное сочетание по той же причине.

6. $\hat{c} < 0; b_{Sy(t)} > 0; 2c < b_{Sy(t)}$ - очевидная ситуация. Это означает, что прирост уровней снижается, а колебания возрастают. Естественно, показатель устойчивости тенденции уменьшается и за счет знаменателя, устойчивость падает, это самый неблагоприятный тип динамики производства относительно его устойчивости.

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

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.

$$8. \hat{c} < 0; b_{Sy(t)} < 0; 2c < b_{Sy(t)} - \text{также понимается по алгебраической величине.}$$

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 относительно устойчивости.

Еще раз подчеркнем, 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

В предыдущих главах рассматривалась динамика одного признака, выраженного тем или иным показателем, но фактически наука и практика всегда имеют дело не с изолированными признаками, а с их системами, жестко связанными функциональной либо корреляционной связью. В данной главе будут последовательно рассмотрены методики анализа таких систем признаков, а также свойства трендов и колеблемости при агрегировании объектов по совокупности, описаны связи, особенно корреляционные, в динамике. Все эти проблемы на порядок сложнее ранее изложенных и ввиду ограниченности объема учебника могут быть изложены только очень кратко. Желаящим 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 финансов¹. Изложение материала начнем с конкретной задачи: необходимо рассмотреть тенденции и колеблемость трех функционально взаимосвязанных признаков: площади посева зерновой культуры, ее урожайность и валовой сбор зерна (табл. 9.1). Если площади в разные годы обозначим как Π_i урожайность - y_i , валовой сбор - b_i , то имеем функциональную связь: $b_i = \Pi_i \cdot y_i$, справедливую для каждого года (ошибки регистрации не принимаем во внимание). Для соблюдения жесткости связи численные значения округлим до целых (табл. 9.1). Тренды площади и урожайности берем линейные.

¹ 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.

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:

$$\begin{aligned}\hat{b} &= \hat{P} \cdot \hat{y} = (120 + 5t) \cdot (29 + t) = 120 \cdot 29 + 5t \cdot 29 + 120 \cdot t + 5t \cdot t = \\ &= 3480 + (145 + 120)t + 5t^2,\end{aligned}$$

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 порядка, вычисленная по данным ряда валового сбора табл. 9.1, имеет вид:

$$\hat{B}_i = 3472,2 + 264,3t_i + 4,9t_i^2.$$

И если в данном примере расхождения параметров невелики, то при более сильной колеблемости они могут оказаться уже значительно большими. Главный результат наших исследований состоит в том, что установлен факт несовпадения тренда произведения с 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 $\mathbf{x}_i, \mathbf{z}_i$ полагаем линейными, следовательно, тренд \mathbf{y}_i - парабола II 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 соответственно $U_{x_i}, U_{z_i}, U_{y_i}$, так что

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

Так как

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

Рассмотрим произведение трендов сомножителей:

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

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, ускорение их произведения - отрицательная величина. При наличии более двух сомножителей тренд их произведения будет параболой более высокого порядка со значительно сложным поведением, в данном учебнике подробно не рассматривается.

Упомянем все же, что если оба признака-сомножителя изменяются по параболе II порядка, то тренд их произведения будет уже параболой IV порядка. Если тренды сомножителей - экспоненты, то и тренд их произведения - тоже экспонента, но вот каков ее параметр, об этом часто судят неверно. Многие руководители предприятий полагают, что если число работников будет возрастать на

10%, а производительность их труда - на 8% в год, то выпуск продукции будет увеличиваться на $10 + 8 = 18\%$ или даже на $10 \cdot 8 = 80\%$) в год! Оба эти ответа неправильны. Тренд производства будет иметь среднегодовой темп роста, равный произведению темпов сомножителей, т.е. $1,08 \cdot 1,10 = 1,188$, или $118,8\%$; следовательно, прирост продукции составит $18,8\%$) в год к предыдущему уровню.

Далее рассмотрим свойства тренда признака-произведения при наличии колебаний каждого из признаков-сомножителей, опишем структуру каждого из параметров его параболического тренда $\hat{y}_i = a_y + b_y t_i + c_y t_i^2$, начиная со среднего абсолютного прироста b_y который и вычисляется первым из уравнения МНК:

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

Далее не будем указывать границ суммирования, они всегда проходят по всем уровням ряда (по всем периодам). При этом, так как $y_i = x_i z_i$, имеем:

$$b_y = \frac{\sum x_i z_i t_i}{\sum t_i^2} = \frac{\sum (\hat{x}_i + U_{x_i}) \cdot (\hat{z}_i + U_{z_i}) t_i}{\sum t_i^2} = \frac{\sum (\bar{x} + b_x t_i + U_{x_i}) \cdot (\bar{z} + b_z t_i + U_{z_i}) t_i}{\sum t_i^2} = \frac{\sum (\bar{x} \cdot \bar{z} t_i + \bar{x} b_z t_i^2 + \bar{z} b_x t_i^2 + \bar{x} U_{z_i} t_i + \bar{z} U_{x_i} t_i + b_x U_{z_i} t_i^2 + b_z U_{x_i} t_i^2 + b_x b_z t_i^3 + U_{x_i} U_{z_i} t_i)}{\sum t_i^2} \quad (9.3)$$

Рассмотрим суммы каждого из слагаемых в числителе (9.3):

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

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

$$= \frac{\bar{x} b_z \sum t_i^2}{\sum t_i^2} + \frac{\bar{z} b_x \sum t_i^2}{\sum t_i^2} = \bar{x} b_z + \bar{z} b_x, \text{ что равно второму члену произ-$$

ведения трендов сомножителей (см. 9.2).

$$4 \text{ и } 5. \frac{\sum \bar{x} U_{z_i} t_i}{\sum t_i^2} + \frac{\sum \bar{z} U_{x_i} t_i}{\sum t_i^2} = \frac{\bar{x} \sum U_{z_i} t_i}{\sum t_i^2} + \frac{\bar{z} \sum U_{x_i} t_i}{\sum t_i^2} = 0,$$

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

$$\sum U_{z_i} t_i = 0; \quad \sum U_{x_i} t_i = 0.$$

$$6 \text{ и } 7. \frac{\sum b_x U_{z_i} t_i^2}{\sum t_i^2} + \frac{\sum b_z U_{x_i} t_i^2}{\sum t_i^2} = \frac{b_x \sum U_{z_i} t_i^2}{\sum t_i^2} + \frac{b_z \sum U_{x_i} t_i^2}{\sum t_i^2}.$$

Эти члены разложения (9.3) в общем случае не равны нулю, так как $U_{z_i} t_i^2$, $U_{x_i} t_i^2$ – случайные величины, зависящие от распределения отклонений от тренда по периодам времени.

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

$$9. \frac{\sum U_{x_i} U_{z_i} t_i}{\sum t_i^2}.$$

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 \mathbf{b}_y содержит еще три члена, в общем случае не равных нулю. Следовательно, в общем случае $\mathbf{b}_y \neq \mathbf{b}_x \mathbf{b}_z$, что мы и наблюдаем на примере табл. 9.1.

Рассмотрим далее квадратический параметр тренда признака-произведения, т.е. c . Из расчета по методу наименьших квадратов (см. гл. 6) для параболы II порядка имеем:

$$c_y = \left(\frac{\sum y_i}{n} - \frac{\sum y_i t_i^2}{\sum t_i^2} \right) \cdot \left(\frac{\sum t_i^2}{n} - \frac{\sum t_i^4}{\sum t_i^2} \right). \quad (9.4)$$

Выражение (9.4) во второй скобке не содержит величин признаков и не нуждается в анализе. В первую скобку подставляем значения:

$$y_i = \frac{\sum (\bar{x} + b_x t_i + U_{x_i}) \cdot (\bar{z} + b_z t_i + U_{z_i})}{n} - \frac{\sum (\bar{x} + b_x t_i + U_{x_i}) \cdot (\bar{z} + b_z t_i + U_{z_i}) \cdot t_i^2}{\sum t_i^2} =$$

$$= \frac{\sum \bar{x} \cdot \bar{z} + \sum \bar{x} b_z t_i + \sum \bar{z} b_x t_i + \sum b_x b_z t_i^2 + \sum \bar{x} U_{z_i} + \sum \bar{z} U_{x_i} + \sum b_x U_{z_i} t_i + \sum b_z U_{x_i} t_i + \sum U_{x_i} U_{z_i}}{n} -$$

$$\frac{\sum \bar{x} \cdot \bar{z} t_i^2 + \sum \bar{x} b_z t_i^3 + \sum \bar{z} b_x t_i^3 + \sum b_x b_z t_i^4 + \sum \bar{x} U_{z_i} t_i^2 + \sum \bar{z} U_{x_i} t_i^2 + \sum b_x U_{z_i} t_i^3 + \sum b_z U_{x_i} t_i^3 + \sum U_{x_i} U_{z_i} t_i^2}{\sum t_i^2}.$$

Рассмотрим каждый из 18 членов разложения, используя уже известные из предыдущего анализа равенства.

$$1. \sum \frac{\bar{x} \cdot \bar{z}}{n} = \bar{x} \cdot \bar{z},$$

т.е. равен первому (свободному) члену произведения трендов сомножителей.

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

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

$$5 \text{ и } 6. \frac{\sum \bar{x} U_{z_i}}{n} + \frac{\sum \bar{z} U_{x_i}}{n} = \frac{\bar{x} \sum U_{z_i}}{n} + \frac{\bar{z} \sum U_{x_i}}{n} = 0, \text{ так как}$$

$$\sum U_{x_i} = \sum U_{z_i} = 0.$$

$$7 \text{ и } 8. \frac{\sum b_x U_{z_i} t_i}{n} + \frac{\sum b_z U_{x_i} t_i}{n} = \frac{b_x \sum U_{z_i} t_i}{n} + \frac{b_z \sum U_{x_i} t_i}{n} = 0, \text{ как}$$

уже показано ранее.

$$9. \frac{\sum U_{x_i} U_{z_i}}{n}.$$

Этот член произведения в общем случае не равен нулю при наличии корреляции между отклонениями от тренда.

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

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

так как $\sum t_i^3 = 0$.

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

$$14 \text{ и } 15. -\frac{\sum \bar{x} U_{z_i} t_i^2}{\sum t_i^2} - \frac{\sum \bar{z} U_{x_i} t_i^2}{\sum t_i^2} = -\frac{\bar{x} \sum U_{z_i} t_i^2}{\sum t_i^2} - \frac{\bar{z} \sum U_{x_i} t_i^2}{\sum t_i^2} - \text{В}$$

общем случае, как ранее показано, не равны нулю, так как зависят от распределения отклонений от трендов по времени.

$$16 \text{ и } 17. - \frac{\sum b_x U_{z_i} t_i^3}{\sum t_i^2} - \frac{\sum b_z U_{x_i} t_i^3}{\sum t_i^2} = - \frac{b_x \sum U_{z_i} t_i^3}{\sum t_i^2} - \frac{b_z \sum U_{x_i} t_i^3}{\sum t_i^2}.$$

В общем случае эти члены не равны нулю при асимметричном распределении отклонений от тренда по длине периода, особенно при ограниченной длине ряда.

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

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 члены разложения 9,14,15, 16, 17 и 18, квадратический член параболы - тренда признака-произведения не равен аналогичному члену произведения трендов сомножителей, что и видим по данным табл. 9.1.

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 расходятся. Следовательно, в общем случае $a_y \neq \bar{x} \times \bar{z}$, свободный член уравнения параболического тренда при неравенстве нулю квадратического параметра вообще никогда не равен средней арифметической величине признака:

$$a_y \neq \bar{y} = \bar{x} \times \bar{z}.$$

Итак, на вопрос о причинах отличия параметров тренда признака-

произведения от произведения соответствующих параметров трендов сомножителей можно дать ответ: *параметры тренда признака-произведения при наличии колебаний уровней признаков-сомножителей относительно их трендов содержат дополнительные случайные члены, зависящие от распределения отклонений признаков-сомножителей от тренда по длине ряда и от наличия корреляции между этими отклонениями.*

Можно сказать, что тренд произведения больше зависит от случайностей, чем зависело бы произведение трендов сомножителей. Это положение необходимо учитывать при обсуждении методики прогнозирования системы жестко связанных признаков.

Теперь кратко рассмотрим связи между колебаниями признаков.

Из табл. 9.1 видно, что лишь четыре раза из девяти позиций знак отклонения от тренда валового сбора соответствует знаку произведения отклонений от тренда площади и урожайности. Представляется на первый взгляд, что колебания признаков вообще никак не связаны.

Более точный анализ связи показал, что коэффициенты корреляции между отклонениями от трендов составили:

$$r_{U_n U_y} = -0,326;$$

$$r_{U_n U_b} = -0,047; \quad r_{U_y U_b} = 0,725.$$

Следовательно, колебания валового сбора в основном были вызваны колебаниями урожайности, а колебания размеров площади слабо связаны и с колебаниями урожайности, и с колебаниями валового сбора.

Что касается интенсивности и силы колебаний, то имеем следующие показатели:

$S(t)n = 7,4$ 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 па трендовый уровень урожайности плюс произведение отклонения урожайности на трендовую величину площади, как «должно было бы быть». Между отклонениями от тренда нет жесткой функциональной связи: множественный коэффициент детерминации колебаний валового сбора колебаниями площади и урожайности равен лишь 0,566, или 56,6%. Жесткая связь колебаний была бы только при такой же жесткой связи колебаний площади и урожайности. Но такой связи не может быть на практике, ибо причины колебаний размера посевной площади в основном имеют экономическую или организационно-хозяйственную основу, а па колебания урожайности влияют причины природного характера.

Итак, можно сделать лишь качественные выводы о связи и силе колебаний жестко взаимосвязанных признаков:

1) при существенной и прямой связи колебаний факторов-сомножителей колебания признака-произведения будут в среднем сильнее, чем каждого из сомножителей, а при обратной и существенной связи колебаний сомножителей колеблемость признака-произведения будет в среднем слабее, чем колеблемость сомножителей;

2) при слабой связи между колебаниями сомножителей колебания признака-произведения приблизительно такие же, как колебания сомножителя с наибольшей колеблемостью по величине коэффициента $V(t)$,

3) ввиду случайного распределения колебаний сомножителей во времени для изучения их связи необходимо рассмотреть достаточно длинные ряды, не менее 13-15 уровней в каждом.

9.2. Агрегирование трендов и колебаний по совокупности объектов

9.2.1. Тренды объемных признаков

Рассмотрим проблему соотношения тренда и колеблемости по совокупности объектов (например, тренда и колеблемости валового сбора по району в целом) и соотношения трендов 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:

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

Средний уровень за ряд лет по совокупности - свободный член линейного тренда - равен, следовательно, сумме свободных членов линейных трендов валового сбора по всем единицам совокупности.

Далее покажем, из чего складывается среднегодовой прирост валового сбора по совокупности:

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

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:

$$nA + C \sum_{i=1}^n t_i^2 = \sum_{i=1}^n X_i,$$

$$A \sum_{i=1}^n t_i^2 + C \sum_{i=1}^n t_i^4 = \sum_{i=1}^n X_i t_i^2.$$

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

Подставляя в правые части X_i имеем

$$nA + C \sum_{i=1}^n t_i^2 = \sum_{i=1}^n \sum_{j=1}^k x_{ij},$$

$$A \sum_{i=1}^n t_i^2 + C \sum_{i=1}^n t_i^4 = \sum_{i=1}^n \sum_{j=1}^k x_{ij} t_i^2.$$

Решая эту систему уравнений, получаем:

$$C = \left(\frac{\sum_{i=1}^n \sum_{j=1}^k x_{ij} - \frac{\sum_{i=1}^n \sum_{j=1}^k x_{ij} t_i^2}{\sum_{i=1}^n t_i^2}}{n} \right) \cdot \left(\frac{\sum_{i=1}^n t_i^2}{n} - \frac{\sum_{i=1}^n t_i^4}{\sum_{i=1}^n t_i^2} \right).$$

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 у 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 пулевом ускорении.

В случае экспоненциальных трендов по каждой единице совокупности тренд по совокупности в целом также является экспонентой, коэффициент роста которой k является не постоянной, а переменной величиной, в каждом периоде равной средней арифметической взвешенной из индивидуальных темпов k_{ij} по величине уровней предыдущего периода. С течением времени общий темп роста по совокупности асимптотически приближается к величине темпа роста, являющегося наибольшим из всех индивидуальных темпов, так как уровень признака у единицы совокупности с наибольшим темпом роста со временем становится преобладающим в совокупности, его доля стремится к единице. Разумеется, теорема сложения трендов к экспонентам неприменима. Она заменяется теоремой усреднения трендов, которую здесь излагать не будем.

9.2.2. Тренды качественных признаков

Более сложная проблема - агрегирование трендов качественных признаков, таких, как урожайность, производительность труда, коэффициент рентабельности и т.д. Очевидно, что величина каждого уровня качественного признака по совокупности в целом есть средняя взвешенная арифметическая величина, из значений данного признака по единицам совокупности; весами являются значения

объемного признака - знаменателя изучаемого качественного показателя; для урожайности - это площадь посева.

Кратко изложим результат исследования, начиная с 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, состоящая из трендов по отдельным хозяйствам. При малой колеблемости долей хозяйств в общей площади культуры по совокупности тренд урожайности в совокупности будет приблизительно равен среднему взвешенному тренду отдельных хозяйств. При существенных изменениях в распределении площадей между хозяйствами с разными трендами общий тренд урожайности по совокупности уже не будет равен среднему из трендов по хозяйствам.

Если бы число единиц совокупности было достаточно большим, а изменения их долей в общем объеме признака-веса были случайными, не связанными или слабо связанными с уровнями урожайности и со скоростями ее изменения в отдельных хозяйствах, то, в силу закона больших чисел, параметры тренда урожайности по совокупности в целом в вероятностном смысле приближались бы к их математическому ожиданию, т.е. к среднему из всех индивидуальных трендов. Насколько реальное изменение площадей в совокупности хозяйств отвечает этим условиям, необходимо конкретно исследовать в каждой отдельной задаче.

9.2.3. Агрегирование показателей колеблемости

Ранее доказано, что каждый фактический уровень объемного признака X_i по совокупности в целом равен сумме уровней этого признака для всех единиц совокупности:

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

Точно так же каждый уровень тренда \hat{X}_i по совокупности есть сумма уровней трендов по единицам совокупности:

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

Тогда и каждое отклонение от тренда по совокупности в целом:

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

Квадрат отклонения в i -м году от тренда по совокупности в целом равен:

$$U_i^2 = \left(\sum_{j=1}^k u_{ij} \right)^2;$$

сумма квадратов отклонений по совокупности в целом:

$$\sum_{i=1}^n U_i^2 = \sum_{i=1}^n \left(\sum_{j=1}^k u_{ij} \right)^2 = \sum_{i=1}^n \left(\sum_{j=1}^k u_{ij}^2 + 2 \sum_1^{C_k^2} u_{im} u_{ip} \right), \quad (9.5)$$

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

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 уровней для разных единиц совокупности от своих трендов. Эта последняя удвоенная сумма парных отклонений по всем C_k^2 (сочетание из k по 2) есть удвоенная сумма ковариаций колебаний по всем возможным парам единиц совокупности. Так как коэффициент каждой парной

корреляции колебаний - величина

$$r_{u_m u_p} = \frac{\sum_{i=1}^n U_{mi} U_{pi}}{\sqrt{\sum_{i=1}^n U_{mi}^2 \sum_{i=1}^n U_{pi}^2}},$$

то

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

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

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

$$\gamma S^2(t)_{\text{сов}} = \sum_{j=1}^k \gamma \cdot S_j^2(t) + 2 \sum_1^{C_k^2} r_{u_m u_p} \cdot \gamma S_m(t) \cdot S_p(t).$$

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

$$S^2(t)_{\text{сов}} = \sum_{j=1}^k S^2(t)_j + 2 \sum_1^{C_k^2} S_m(t) \cdot S_p(t) \cdot r_{u_m u_p}. \quad (9.7)$$

Итак, можно сделать вывод: дисперсия колебаний признака в целом по совокупности 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 $r_{umup} = 0$), дисперсия признака по совокупности в целом была бы равна сумме дисперсий признака для всех единиц совокупности.

Например, если в каждом из 20 предприятий района валовой сбор имел бы дисперсию колебаний, равную 9000 ц², то дисперсия валового сбора по району была бы равна 180000 ц². В таком случае имели бы: $S(t)_{сов} = \sqrt{180000} = 424,26$ ц, в то время, как по каждому предприятию $S(t)_j = \sqrt{9000} = 94,87$ ц, и их сумма по 20 предприятиям составила бы: $94,87 \cdot 20 = 1897,49$. Отсутствие связи колебаний у разных единиц совокупности, независимость их распределения во времени более чем вчетверо снизили бы величину колебаний признака по совокупности в целом. К сожалению, в границах не только административного района, но даже и области, края, небольшого государства многие факторы колебаний валового сбора сельскохозяйственных культур являются общими, действующими на всей территории более или менее согласованно. Это означает, что коэффициенты корреляции r_{umup} в преобладающей части - положительные величины. Если предположить, что в среднем общие факторы объясняют половину колебаний, т.е. $\bar{r}^2 = 0,5$, $\bar{r} \approx 0,7$, то получим следующий результат по (9.7):

$$S(t)_{сов} = \sqrt{180000 + C_{20}^2 \cdot 94,87^2} = \sqrt{180000 + 190 \cdot 94,87^2} = 1374,8 \text{ ц.}$$

Как видим, и эта величина все еще существенно меньше, чем сумма колебаний по 20 единицам. Так как на практике невозможно, чтобы все факторы колеблемости для всех единиц совокупности были только общими, всегда есть и часть специфических факторов колеблемости для отдельных предприятий, то коэффициенты корреляции отклонений от трендов всегда в среднем меньше

единицы, а тогда правая часть выражения (9.7) меньше, чем квадрат суммы колебаний. 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:*

$$S(t)_{\text{сов}} < \sum_{j=1}^k S(t)_j,$$

$$V(t)_{\text{сов}} < \bar{V}(t)_j.$$

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 нулю².

Данный закон справедлив и для вторичных признаков, таких, как урожайность. Если бы колебания урожайности у всех единиц совокупности были жестко связаны (т.е. все $r_{um_{up}}$ были равны единице), то колебания урожайности по совокупности были равны средней из показателей $S(t)_j$ каждой единицы совокупности. Но так как на разных предприятиях, в хозяйствах есть не только общие для совокупности факторы колеблемости, но и специфические, все $r_{um_{up}} < 1$, а, значит, колебания средней урожайности по совокупности хозяйств, даже если взять простую среднюю, будут меньше, чем среднее квадратическое отклонение по всем единицам. А если еще среднюю урожайность по совокупности вычислить как взвешенную по площадям, то их колебания, конечно, не строго согласованные по всем единицам совокупности, также будут

² Проблема взаимопогашения колебаний валового сбора и урожайности для России была исследована в ряде

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:

- 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, поэтому в данном учебнике не рассматривается. Она решается специалистом в той сфере знаний, которая изучает объект, процесс, - агрономом, инженером, экономистом, социологом, биохимиком, астрономом и т.д. Без причинного обоснования лучше не начинать измерение связи в динамике.

Но даже и после такого обоснования остается открытым вопрос: при наличии одинаково направленных трендов двух причинно-связанных признаков не преувеличится ли теснота связи за счет трендов? Если, например, в стране растет производство и применение минеральных удобрений, растет и урожайность сельскохозяйственных культур, но 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 большие 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, например, таких, как Оренбургская область, причинно связаны не с какой-либо тенденцией изменения суммы осадков, а с ее колебаниями в отдельные годы. К тенденции же роста урожайности осадки никакого отношения (причинной связи) не имеют. Снимается ложная корреляция тем, что колебания других факторов, влияющих на урожайность, - экономических, организационных - не связаны или слабо связаны с колебаниями осадков. Тенденции факторов связаны часто, колебания - почти никогда. Поэтому связь между колебаниями одного фактора с результативным показателем (его колебаниями) почти всегда свободна от ложной корреляции, наведенной другими факторами.

В последующих разделах данной главы в основном будут рассматриваться корреляция между колебаниями признаков, а также методики ее измерения и моделирования. Что же касается измерения связи между тенденциями, между самими уровнями временных рядов, включающих тенденцию, а не только колеба-

ния, то эта проблема не может считаться решенной. Некоторые указания читатели учебника могут найти в разделе о смешанных прогностических моделях (гл. 10). Излагаемые здесь методики решают только ограниченный класс задач - измерение связи между колебаниями факторного (факторных) признака и колебаниями результативного признака.

Строго говоря, это жесткое ограничение относится и к пространственной корреляции в том смысле, что и в ней измеряется связь вариации результативного признака с вариацией фактора. Например, за счет вариации дозы минеральных удобрений объясняется 38% вариации урожайности пшеницы между хозяйствами области ($r^2 = 0,38$), а не 38% уровня урожайности, как иногда неверно считают.

9.4. Методы измерения корреляции между колебаниями признаков

Итак, в предыдущем разделе было установлено, что единственная «чистая» задача об измерении корреляции временных рядов - это измерение связи между колебаниями их уровней. *Колебания* - это, как правило, случайная составляющая, в отличие от тренда. Если же и колебания не случайны, а строго упорядочены, как, например, сезонные, то и задача о связи таких колебаний не является «чистой», так как содержит риск ложной связи. В связи с этим далее 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, представленный в табл. 9.3.

Средние: $\bar{x} = 119,92 \approx 120; \bar{y} = 19,0$.

Уравнения трендов:

урожайности: $\hat{x} = 119,9 + 3,81t$;

себестоимости: $\hat{y} = 19,0 + 1,22t$, где $t = 0$ в 1983 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 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**

Год	Урожайность, ц/га, X_i	Себесто- имость, руб/ц, Y_i	Тренды		Отклонения от трендов		Квадраты отклонений		Произве- дения $U_{xi} U_{yi}$	t i руб/ц
			\bar{x}_i	\bar{y}_i	U_{xi}	U_{yi}	U^2_{xi}	U^2_{yi}		
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	0,0	25,1
1989	183	21,1	143	26,3	+40	-4,9	1600	24,00	-196,0	21,3
У	1559	247,0	1559	247,0	0	0,0	7678	133,30	-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:

$$\mathbf{a} = \bar{\mathbf{y}} - \mathbf{b}\bar{\mathbf{x}}, \text{ т.е. для отклонений от трендов: } \mathbf{a} = \bar{U}_y - \mathbf{b}\bar{U}_x = 0.$$

Уравнение регрессии имеет вид:

$$U_{y_i} = b U_{x_i}. \quad (9.10)$$

Подставляя данные из табл. 9.3, получаем:

$$r_{U_x U_y} = \frac{-952,7}{\sqrt{7678 \cdot 133,3}} = -0,941; \quad r^2 = 88\%.$$

Таким образом, колебания себестоимости картофеля в совхозе почти целиком были связаны с колебаниями урожайности, связь обратная, как и требуют законы экономики. И вся она была подавлена тем, что оба тренда имели одно и то же направление по совершенно разным причинам: прогресс

агротехники - не причина инфляции и роста цен. Равно как и наоборот: инфляция скорее тормозила прогресс урожайности.

Коэффициент регрессии: $b = \frac{-952,7}{7678} = -0,124$, уравнение регрессии:
 $U_{y_i} = -0,124U_{x_i}$. Смысл этого уравнения таков: в среднем отклонение себестоимости от ее тренда в i -м году составляет 0,124 величины отклонения урожайности от своего тренда с обратным знаком. Значения себестоимости, рассчитанные по модели с учетом тренда себестоимости и колебаний урожайности, приведены в последней графе табл. 9.3:

$$\hat{y}(x)_i = (19,0 + 1,22t_i) + (-0,124U_{x_i}).$$

Как видим, полученные по этой модели уровни себестоимости довольно близки к фактическим.

Другим методом измерения корреляции между временными рядами служит метод корреляции цепных показателей динамики, которые являются константами трендов. Для линейных трендов - это абсолютные цепные изменения. Метод предпочтительно применять для таких рядов, в которых среднее изменение (параметр b) существенно меньше, чем среднее колебание $S(t)$, иначе говоря, показатель K значительно меньше единицы.

Логика применения метода заключается в том, что если колеблемость намного больше изменения тренда за единицу времени, то цепные абсолютные изменения, т.е. разности соседних уровней, в основном состоят из колебаний. В связи с этим корреляция абсолютных изменений будет мало отличаться от корреляции отклонений от тренда. Метод имеет и преимущество: не нужно вычислять тренд, ошибка в выборе типа тренда не влияет на конечный результат. Расчет идет непосредственно по исходным временным рядам. По данным табл. 9.3 имеем:

$$\Delta_x = +5,57 \approx +5,8; \quad \Delta_y = +0,738 \approx +0,74.$$

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 приведены в табл. 9.4. Исходя из них имеем:

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

что почти совпадает с ранее полученной величиной коэффициента корреляции отклонений от трендов.

Если тренды признаков являются экспонентами, то вместо корреляции отклонений от трендов можно применить метод корреляции цепных темпов роста уровней, поскольку именно темпы роста - основной параметр экспоненциальных трендов.

Остаются недостаточно проработанными следующие 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 из них предложит свои ответы на эти вопросы. Еще раз, и не последний, авторы подчеркивают, что наука - открытая система, продолжающийся процесс познания, открытия новых «материков» (реже) и «островов» (чаще) в бесконечном океане неведомого.

В заключение 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.

Таблица 9.4

Корреляция отклонений от средних отклонений

Год	γ	y_i	$\Delta x_i =$ $= x_{i+1} - x_i$	$\Delta y_i =$ $= y_{i+1} - y_i$	$\Delta_{x_i} - \bar{\Delta}_x$	$\Delta_{y_i} - \bar{\Delta}_y$	$(\Delta_{x_i} - \bar{\Delta}_x) \times$ $(\Delta_{y_i} - \bar{\Delta}_y)$	$(\Delta_{x_i} - \bar{\Delta}_x)^2$	$(\Delta_{y_i} - \bar{\Delta}_y)^2$
1977	108	11,8							
1978	81	15,4	-27	+3,6	-32,8	+2,86	-93,8	1075,8	8,2
1979	106	13,0	+25	-2,4	+19,2	-3,14	-60,3	368,6	9,9
1980	124	13,9	+18	+0,9	+12,2	+0,16	+2,0	148,8	0,0
1981	103	15,1	-21	+1,2	-26,8	+0,46	-12,3	718,2	0,2
1982	106	19,6	+3	+4,5	-2,8	+3,76	-10,5	7,8	14,1
1983	149	16,2	+43	-3,4	+37,2	-4,14	-154,0	1383,8	17,1
1984	148	17,2	-1	+1,0	-6,8	+0,26	-1,8	46,2	0,1
1985	102	24,0	-46	+6,8	-51,8	+6,06	-313,9	2683,2	36,7
1986	130	22,4	+28	-1,6	+22,2	-2,34	-51,9	492,8	5,5
1987	80	32,3	-50	+9,9	-55,8	+9,16	-511,1	3113,6	83,9
1988	139	24,7	+59	-7,6	+53,2	-8,34	-443,7	2830,2	69,6
1989	183	21,4	+44	-3,3	+38,2	-4,04	-154,3	1459,2	16,3
У	1559	247,0	+75	+9,6	-	-	-1805,6	14328,2	261,6

Эти методы лучше применять только при явном преобладании колеблемости над тенденцией изменения за единицу времени, т.е. при малом показателе K для линейных трендов или малых аналогичных показателях для других типов трендов (см. разд. 8.3).

9.5. Корреляция с учетом лага и циклов

Среди природных и общественных явлений нередко встречаются такие, которые связаны между собой не в одном и том же периоде времени, а с некоторым запозданием - по-английски - lag, откуда пошел термин *лаг*. Например, капиталовложения в создание машиностроительного, автомобильного завода отразятся в росте объема производства не в том году, когда они произведены, а через два-три и более лет, капиталовложения в строительство крупной гидроэлектростанции - через 6-8 лет. При наличии лага в реальной связи изучаемых явлений измерять корреляцию факторного признака с результативным нужно, конечно, не по одновременным уровням, а с учетом лага. Например, отклонение от тренда капиталовложений скажется на отклонении от тренда выпуска продукции через k лет. Значит, измерять корреляцию нужно через произведения $U_{x(t_1)}$, $U_{y(t_1+k)}$.

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.

Таблица 9.5

Корреляция с лагом в 1 год

Год	Доза, т/га, x_i	Урожайнос- ть, Ц/га, y_i	Тренд \bar{x}_i	Отклоне- ние, U_{y_i}	$x_i - \bar{x} =$ $= \Delta_{x_i}$	$\Delta_{x_i}^*$ $* U_{y_{i+1}}^2$	$\Delta^2_{x_i}$	$\Delta^2_{y_i}$	\bar{y}'_i
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	0,0	1	8,41	15,6
1992	51	18,0	18,0	0,0	+10	+38,0	100	0,00	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
У	451	198,0	198,0	0,0	0	+70,4	314	50,44	197,2

При этом будем считать, что тренд дозы внесенных органических удобрений отсутствует или несуществен.

Средняя доза удобрений: $\bar{x} = 451:11 = 41$ т/га.

Тренд урожайности: $\bar{y}_i = 18,0 + 0,6 \cdot t_i$; $t = 0$ в 1992 г.

Коэффициент корреляции с учетом лага в 1 год имеет вид:

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

Связь колебаний дозы удобрений под 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 удобре-

ний 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

щих 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 три года, то через четыре года. Лагом в этом случае считается средний промежуток времени между локальными максимумами коэффициентов корреляции, между отклонениями от трендов.

Рассчитываем коэффициенты корреляции отклонений от тренда, начиная с нулевого лага (табл. 9.6):

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

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

Таблица 9.6

Корреляция отклонений от тренда с неизвестным заранее лагом

Год	t_i	x_i	ϵ_i	U_{x_i}	y_i	ϵ_i	U_{y_i}	$U_{x_i}^2$	$U_{y_i}^2$
1983	-7	67	80	-13	18,2	22,2	-4,0	169	16,00
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	Л	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	0,00
1997	7	109	108	+1	36,6	39,0	-2,4	1	5,76
У	0	1410	1410	0	459	459	0	1058	100,72

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

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

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

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

Нет смысла продолжать корреляцию, так как остается все меньше и меньше слагаемых в суммах и коэффициент становится все более случайным. Можно сделать достаточно уверенно вывод о том, что лаг равен трем годам, так как коэффициент с лагом три года довольно резко выделяется. Такой вывод будет справедлив, если по существу известно, что связь должна быть прямой, например, x - капиталовложения, млрд руб., y - ввод в эксплуатацию жилой площади, млн м². Если же неизвестен априори не только лаг, но даже и направление связи, то

следует проверить и альтернативную гипотезу: обратную связь при лаге в два года.

9.6. Понятие о динамике комплекса статистически взаимосвязанных признаков

В разд. 9.1 была рассмотрена методика изучения динамики комплекса жестко взаимосвязанных признаков. Здесь мы представим лишь пути исследования динамики статистически взаимосвязанных, в основном корреляционно-связанных, признаков³.

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

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

выражающую корреляционную зависимость результативного признака y от ряда факторных признаков x^1, \dots, x^k . С развитием процесса во времени могут изменяться значения факторных признаков, теснота их связи с результативным признаком¹, а также относительная роль данного комплекса факторов в общей вариации результативного показателя.

Самая простая задача - разложить изменения результативного признака в текущем периоде в сравнении с базисным за счет:

- изменения величин факторных признаков;
- изменения коэффициентов регрессии при факторах;
- прочих причин.

Чтобы разложить общее изменение средней величины результативного признака на элементы, используем обычную последовательность индексирования: сначала изменяем значения объемных показателей, которыми в рассматриваемом случае являются средние величины факторных признаков при сохранении

³ Эту 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.

базисных значений качественных показателей. В данной задаче это коэффициенты регрессии, измеряющие эффект каждого фактора. Далее индексируем коэффициенты регрессии при неизменных на уровне отчетного периода средних значениях. Рассмотрим пример по изменению среднего надоя молока на корову в совхозах Ленинградской области в 1983 г. по сравнению с 1982 г. и системы факторных признаков:

y - средний надой молока на корову, кг/год;

l'' , - затраты труда на 1 среднегодовую корову, ч;

x^{\wedge} - доля концентратов в рационе коров, %;

$l'z$ - доля корнеплодов в рационе коров, %;

$l''4$ - средняя длительность сухостойного периода, дни;

x^{\wedge} - средняя длительность сервис-периода, дни;

V_g - доля коров классов «элита», «элита-рекорд» в стаде, %;

v , - общий объем рациона коров, ц корм. ед./год;

X_g - оценка сельхозугодий, баллы. За базисный 1982 г. регрессионная

модель имела вид:

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

За текущий 1983 г. регрессионная модель приняла следующий вид:

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

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

Таблица 9.7 Значения факторных признаков

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^{\wedge} и X_u снизился коэффициент при $x-u$, 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

k

$\hat{A}_0 \hat{A}_1 \dots \hat{A}_k$

$$/ \quad = \frac{M}{\dots} = 341a^2 = 10428, \text{ или } 104,28\%.$$

$$y^{(i)} \wedge \quad \underline{\quad} \quad 3270,2$$

$\hat{A}_0, \hat{A}_1 \dots \hat{A}_k$

$Y=1$

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 па 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 равнодействующим) 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 годовых моделей за ряд лет, чем предварительное усреднение информации за ряд лет, а затем вычисление по ней показателей связи. В последнем приеме возможно возникновение неинтерпретируемых «парадоксальных» коэффициентов регрессии, либо больших, чем все годовые, либо меньших. Также нарушает условия применения МНК так называемый *метод заводов-лет* (или *хозяйство-лет*) - простое суммирование (без усреднения) первичной информации за ряд лет, особенно если совокупность велика.

Условиям задачи наиболее полно отвечает ковариационный анализ, при котором различия, обусловленные особенностями периодов времени, полностью отделяются от вариаций результативного показателя за счет факторных признаков. (В данном учебнике сложные методы ковариационного анализа, введение дополнительных переменных для периодов времени не рассматривается.)

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_j - для каждого фактора, а также L .
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 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 -, и л., 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. Персоне предложил, чтобы убедиться в надежности коэффици-

ента корреляции между динамическими рядами, лучше не ограничиваться вычислением его за весь период в целом, и, кроме этого разделить период па 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. Персоне не ставил вопроса об изучении динамики силы связи.

С.П. Бобров и Б.С. Ястремский в 1922-1923 гг. пошли по другому пути. Они впервые поставили вопрос о важном значении, которое может иметь в экономике, биологии и других науках факт изменения тесноты связи между факторным и результативным признаком, и разработали свои методики решения этой задачи.

Б.С. Ястремский предложил [22] для изучения изменения тесноты связи во времени построить ряд нарастающих чисел-

\underline{n}

телей коэффициента, т.е. сумм $\sum_{i=1}^n x_i y_i$, если эти суммы с ростом n и 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 a^{\wedge} и o'' он считал неизменными на уровне не всего динамического ряда.

С.П. Бобров справедливо упрекает Ястремского за принятие гипотезы о постоянстве «сигм», так как они тоже могут эволюционировать, как и числитель. Кроме этого у Б.С. Ястремского нет ясности в том, какие величины он принял за x и y - 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 дол-

жны, 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 скользящий расчет 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

$$\sum_{i=1}^m U_{x_i} U_{y_i}, \text{ где } m \text{ от } 1 \text{ до } n;$$

- нарастающие итоги квадратов отклонении каждого признака от своего тренда:

$$\sum_{i=1}^m U_{x_i}^2 \text{ и } \sum_{i=1}^m U_{y_i}^2 ;$$

- 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 / делением ординаты первой выровненной линии, т.е.

$$\sum_{i=1}^m U_{x_i} U_{y_i}$$

для года номер /л, на корень квадратный из произведения ординат нарастающих сумм квадратов отклонений, т.е.

J_m

Это построение полностью

m

отвечает изло

"а WW.

l=1

;=i

женной ранее в данной главе методике измерения корреляции временных рядов. Но Н.С. Четвериков идет дальше, вычисляя 'и,н, для многих периодов скользящим способом: конкретно для периодов от t^{\wedge} до $^{\wedge}$, где $n > m$, а m достаточно велико для получения надежной меры тесноты связи. В результате исследователь получает ряд коэффициентов корреляции для периодов от номера г до г 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, пригод-

на 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 триады (см. гл. 5) уже есть модель временного ряда. В гл. 6, в частности, в разделе о сезонных (и иных циклических) колебаниях получены и некоторые модели колеблемости. Остается свести их в общую модель изменения изучаемого показателя с течением времени и оценить возможность прогнозирования его будущих значений.

Прогноз (в переводе с греческого языка - предвидение, предсказание, предзнание) - неотъемлемая составляющая всей человеческой деятельности, в том числе экономической. Это промежуточное звено между познанием объективной реальности и деятельностью людей по ее преобразованию. Один из основоположников позитивизма Огюст Конт (1798-1857) говорил: *«Savoir pour prévoir; prévoir pour agir» (знать, чтобы предвидеть; предвидеть, чтобы действовать).*

Самые разные прогнозы - от прогноза погоды на завтра до прогноза результатов президентских выборов - составляют значительную часть информации, циркулирующей в обществе. Разработкой прогнозов рынка сбыта, финансовых потоков, курса

валют и других важнейших показателей деятельности заняты тысячи, если не миллионы работников банков, фирм, государственных органов, частных компаний.

Создание методов прогнозирования - одна из главных проблем науки и, может быть, труднейшая их них. Не случайно ученый-геолог, писатель-фантаст и один из самых глубоких мыслителей России XX в. Ив. Аит. Ефремов (1907-1972) 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 прогнози-

рования 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 заключается 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 по делу. 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 энер-

гаи, 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 г. Чем дальше удален прогнозный уровень от базы расчета тренда, тем больше ошибка прогноза, как будет показано в дальнейшем.

Прогноз по тренду - лишь один из статистических методов прогнозирования. Полезно сравнить его свойства, положительные и негативные, со свойствами прогнозирования на основе многофакторных регрессионных моделей. Начнем с положительных свойств прогноза по тренду. Коэффициент при номере периода в уравнении тренда (B - в линейном уравнении) - это комплексный коэффициент регрессии при всех реальных факторах, влияющих на уровень изменяющегося показателя, которые сами изменяются во времени. Подчеркнем: при всех факторах! Ни в одну (факторную регрессионную модель мы не можем включить все факторы, влияющие на изучаемый показатель, например на урожайность. Во-первых, часть факторов вообще неизвестна, так как наши знания, наука не имеют статуса абсолютной, полной истины. Во-вторых, часть факторов теоретически известна, но на практике по ним нет достаточно надежной или даже вообще никакой информации. В-третьих, если число известных факторов велико, то всех их явно невозможно включить в уравнение регрессии по математическим ограничениям:

мультиколлинеарность, гетероскедастичность, превышение чис-

ла 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 тренда соответствует задаче. Напротив, уравнение многофакторной регрессии - это модель вариации уровня показателя в статической совокупности. Эта модель объясняет не изменение, например, урожайности во времени, а ее различия в совокупности хозяйств в данный период. Логическая база прогноза по многофакторной регрессии в статике неадекватна задаче прогнозирования. Конкретный пример: один из главных факторов вариации урожайности в регрессионной модели - тип почвы, почвенная разность, но почвы области не будут в динамике за несколько лет меняться, и на динамику этот фактор не влияет. Зато в регрессионной модели за данный год по всем хозяйствам области средняя температура месяца почти одинакова, и из регрессионной модели этот фактор исключается. Однако в динамике температура месяца может сильно колебаться, и в прогнозе это следовало бы учитывать.

Последнее, хотя и не очень существенное преимущество прогноза по тренду заключается в том, что для него не требуется большого объема исходной информации о факторах. Достаточно однородного по характеру тенденции периода за 20-25 лет, т.е. всего два десятка уровней, например, урожайности.

Но у прогнозирования по тренду есть, конечно и свои недостатки. Неявность факторов динамики, скрытых за «номером периода», лишает исследователя возможности учесть ожидаемый или планируемый перелом, скачок в развитии того или иного фактора. Нет возможности проигрывать разные варианты прогноза при разных сочетаниях значений факторов, что обычно делается при прогнозе по регрессионной модели с управляемыми факторами.

Прогноз по тренду несет в себе как бы черты фатализма: будет то-то, изменить ничего нельзя. Ведь мы не можем изменить

или отменить ход времени, а аргумент уравнения тренда - это время. Конечно, на самом деле тренд образовался как под влиянием природных факторов, так и деятельности человека. Но слитность этих 4-х факторов все равно оставляет впечатление, что человек устранен из процесса, так что психологически данный метод нередко отторгается именно по причине своего фаталистического имиджа. Особенно это чувствовалось в планово-командной экономике, когда считалось, что в будущем будет то и столько, сколько мы запланируем. Прогнозирование в этой системе управления было подавлено «прямым директивным планированием».

Теперь ясно, что прогнозирование - неотъемлемый элемент менеджмента, оно составляет этап и 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:

$$\hat{y}_t = 51,25 + 1,452 t, \quad t = 0 \text{ в } 1983 \text{ г.}$$

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

$$\hat{y}_{2000} = 51,25 + 1,452 \cdot 17 = 75,93 \text{ ц/га.}$$

Интерпретация этого прогноза должна быть следующей: если урожайность зерновых во Франции будет возрастать до 2000 г. с той же средней скоростью (среднегодовым приростом), с какой 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 интервале шириной 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 был 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 было показано, что, распространяя уравнение тренда на будущее, мы обязаны считать его лишь выборочной оценкой генеральных параметров, точно нам не известных. Наличие случайной колеблемости уровней порождает ошибку репрезентативности выборочных оценок тренда, которую следует принимать 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. И так, при несущественности колебаний процесса точный прогноз оказывается самодостаточным и не требует каких-либо дополнений. В экономике, «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:

$$m_b = \frac{S(t)}{\sqrt{\sum_{i=1}^n t_i^2}}, \quad (10.1)$$

где n - число уровней базы тренда;

$S(t)$ - среднее квадратическое отклонение уровней ряда от тренда.

Объединяя эти ошибки как независимые по правилу для дисперсий независимых переменных и учитывая, что ошибка среднегодового прироста за t лет (или иных отрезков времени) возрастет в t раз, получаем формулу средней ошибки прогноза для линии тренда на период с удалением t от середины базы прогноза:

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

при однократном выравнивании ряда за n периодов, где $\sum t_i^2$

— это сумма квадратов номеров лет t_i от $-\frac{n-1}{2}$ до $+\frac{n-1}{2}$.

Например, тренд численности занятых в народном хозяйстве в РФ за 1990–1996 гг. составил: $\hat{y}_t = 70,5 - 1,62 t$, 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 од ного среднего квадратического отклонения, т.е. m , 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). По

лучаем вероятную ошибку: $\alpha_{\hat{y}_{98}} = t_{Ст} \cdot m_{\hat{y}_{98}} = 2,8 \cdot 0,291 = 0,81$.

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

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

$$m_{\hat{y}_n} = S(t) \sqrt{\frac{1}{N} + \frac{t_k^2}{\ell \cdot \sum_1^n t_i^2}}, \quad (10.3)$$

где f - число сдвигов базы расчета среднегодового прироста B ;

u - длина каждой базы расчета;

N - общая длина временного ряда.

Например, тренд урожайности зерновых культур во Франции, рассчитанный в гл. 5, имеет вид:

$$\hat{y}_i = 51,25 + 1,452 t, \quad \text{где}$$

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

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

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

Критерий Стьюдента при 24 степенях свободы вариации для вероятности 0,95 составит 2,08. Таким образом, с вероятностью 0,95 тренд при сохранении прежней скорости роста урожайности в 2000 г. проходит 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 не зависимыми 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^{\wedge} от середины базы расчета тренда приобретает вид [18, с. 171]:

$$m_{\hat{y}_k} = S(t) \sqrt{\frac{t_k^2}{\sum t_i^2} + \frac{\sum t_i^4 - (2 \sum t_i^2) \cdot t_k^2 + n t_k^4}{n \sum t_i^4 - (\sum t_i^2)^2}}. \quad (10.4)$$

При использовании многократного скользящего выравни-

вания для расчета параметров параболы II порядка знаменатели обеих дробей подкоренного выражения умножаются на число сдвигов базы расчета C , а суммы, стоящие в формулах, исчисляются за одну базу. В числителе же последней дроби будет $N-f^{\wedge}$.

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 (момента)

Неопределенность прогноза уровня отдельного периода складывается из двух элементов: ошибки линии тренда для прогнозируемого периода и колебаний уровня около тренда. Первый элемент рассмотрен в предыдущем разделе. Колеблемость отдельных уровней относительно линии тренда измеряется средним квадратическим отклонением $S(t)$. Однако необходимо ответить на вопрос: допустимо ли переносить значение этой величины, полученное за период-базу, на прогнозируемый период? Теоретически могут иметь место изменения величины колебаний и в сторону их роста при тенденции роста уровней и постоянстве факторов колеблемости (постоянном коэффициенте колеблемости), и в сторону сокращения абсолютной величины колебаний при 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 $S(t)$, 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 уровень не $S(t)$, а доверительный интервал $S(t) \pm t^* m$, и т.д. На эти теоретически верные рассуждения следует ответить тем, что ограниченная точность и надежность исходных уровней временного ряда $\hat{S}(t)_k$ 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 $S(t)$, если тренд колеблемости надежно установлен (см. разд. 7.3), или величину $S(t)$ по базе прогноза, если изменение этого показателя со временем не установлено надежно.

Кроме этого необходимо быть осторожным с линейными трендами уменьшения колеблемости. Пока это сокращение незначительно и значение $S(t)$ 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 $\hat{S}(t)_k$ и вариация признаков в пространстве: она не может быть уничтожена. Для прогнозов показателей колеблемости на более далекую перспективу при тенденции ее сокращения следует применять гиперболическую форму тренда.

Определив величину показателя колеблемости для прогнозируемого периода, подставляем этот показатель вместе с ошибкой прогноза тренда в единую формулу средней ошибки прогноза конкретного отдельного уровня опираясь на правило сложения независимых дисперсий:

$$m_{Y_k} = \sqrt{m_{\hat{S}_k}^2 + S^2(t)_k}. \quad (10.5)$$

Эта формула является общей для любых типов линии тренда. Для каждого типа различны первые слагаемые - ошибки

тренда на период $\hat{}$. Для линейного тренда при однократном его расчете, используя формулы (10.1) и (10.3), имеем:

$$m_{y_k} = \sqrt{\frac{S(t)^2}{n} + \frac{S(t)^2 \cdot t_k^2}{\sum_1^n t_i^2} + S(t)_k^2}. \quad (10.6)$$

Эта формула должна применяться, если приведен расчет величины $S(l)^\wedge$ на прогнозируемый период, так как в этом случае в первых двух дробях в числитель входит величина $S(l)$ за период-базу, а третье слагаемое подкоренного выражения - это прогнозируемая величина колеблемости на прогнозный период. Таким образом, «вынести за знак корня величину» $S(t)$ нельзя, так как они под корнем различные.

Если же на период прогноза принята та же величина показателя колеблемости, как и за период-базу расчета тренда, то эта величина выносится из-под корня, тогда имеем:

$$m_{y_k} = S(t) \sqrt{1 + \frac{1}{n} + \frac{t_k^2}{\sum_1^n t_i^2}}. \quad (10.7)$$

Именно данная формула приводится обычно в учебниках. Соответственно при многократном расчете среднегодового прироста B получаем формулу

$$m_{y_k} = S(t) \sqrt{1 + \frac{1}{N} + \frac{t_k^2}{\ell \sum_1^n t_i^2}}. \quad (10.8)$$

Значения обозначений те же, что в формулах (10.1 и 10.3).

Для всех других типов тренда средняя ошибка отдельного уровня вычисляется по общей формуле.

По ранее рассмотренным примерам имеем.

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

$$m_{y_{98}} = 0,286 \sqrt{1 + \frac{1}{7} + \frac{5^2}{28}} = 0,408 \text{ млн чел.}$$

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: $S(t)^{\wedge} = 5(f) + s(r) - h = 3,54 + (-0,1235) \cdot 17 = 1,44$. 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. $S(t)^{\wedge}$, 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:

$$75,93 \pm 2,08 \cdot 1,906 = 75,93 \pm 3,96 \text{ ц/га.}$$

10.3.3. Прогноз доверите

льного интервала для среднего уровня за ряд периодов

Предметом прогнозирования может быть не только уровень отдельного года или месяца, но и средний уровень за ряд периодов (моментов). Рассмотрим, что в этом случае можно считать *точечным прогнозом* и какова формула средней ошибки прогноза среднего уровня.

При линейной форме тренда приросты уровней или их сокращения - постоянная величина. В этом случае средняя вели-

чипа 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

$$\frac{\hat{y}_{98} + (\hat{y}_{98} + b) + (\hat{y}_{98} + 2b) + (\hat{y}_{98} + 3b) + (\hat{y}_{98} + 4b)}{5} =$$

$$= \frac{5\hat{y}_{1998} + 10b}{5} = \hat{y}_{1998} + 2b = \hat{y}_{2000}.$$

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 ..., 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 k от середины базы расчета тренда:

$$k + \frac{m}{2}$$

$$m\bar{y}_{k-k+m} = \sqrt{\frac{S^2(t)}{n} + \frac{S^2(t)t_{k+0,5m}^2}{\sum_l^n t_l^2} + \frac{S^2(t)}{m}}. \quad (10.9)$$

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

$$m\bar{y}_{k-k+m} = S(t) \sqrt{\frac{1}{n} + \frac{t_{k+0,5m}^2}{\sum_l^n t_l^2} + \frac{1}{m}}. \quad (10.10)$$

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

$$\frac{S^2(t)}{m}$$

и вычисляется квадратный

ошибки тренда добавляете?

корень из суммы подкоренных дробей.

Например, поданным за 1970-1990 гг. тренд \hat{B}_t - среднего сбора

$$\hat{B}_t = 45,5 - 0,5857 t, \text{ млн т, } t = 0$$

ра пшеницы в России имел вид:

в 1980г.

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

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

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

С вероятностью 0,95 доверительный интервал прогноза среднегодового валового сбора зерна пшеницы в России при условии сохранения до 1995 г. прежнего тренда составил: $37,88 + + 2,09 \cdot 4,91$, 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 $38,9 +$ четыре прогноза на 1992-1995 гг., деленной на пять. Если исходить из того, что уравнение тренда не будет ежегодно пересматриваться, имеем:

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

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 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: $u = (0 + + 0 + 4,91 + 5,09 + 5,28) : 5 = 3,06$, доверительный интервал с вероятностью 0,95 равен: $39,4 \pm 2,09 \cdot 3,06$, или от 33,0 до

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

45,8 млн т. После 1993 г., валовой сбор в котором равен фактически 43,5 млн т, имеем:

точный 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:

точный прогноз

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

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 в РФ.

10.3.4. Расчет страхового запаса для преодоления риска отклонения от тренда

Доверительные интервалы прогноза могут служить основанием для расчета рисков больших отклонений от тренда и страхового запаса для преодоления этого риска. В предыдущем разделе вычислены границы доверительного интервала прогноза 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. $7,88 : 7,18 = 1,097 = 1,10$. Вероятность того, что отклонение от тренда превысит 1,15 (0 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 народнохозяйственное значение. По верхней границе доверительного интервала можно рассчитать вероятность иного, не столь драматического риска: риск перепроизводства, нехватки емкостей для хранения зерна, снижения 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. (тыс. ч) $y^{\wedge} = 60,03 + 1,085t$, $t = 0$ в июле 1996 г., и вектор средних индексов сезонности для каждого месяца:

Месяц 'сеч	Месяц 'сез.	Месяц 'сез;
Январь 0,433 Февраль 0,394 Март 0,517 Апрель 0,777	Май 1,617 Июнь 1,481 Июль 1,219 Август 1,680	Сентябрь 1,617 Октябрь 1,206 Ноябрь 0,734 Декабрь 0,471

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

Модель, учитывающая тренд и сезонность, имеет вид:

$$\hat{y}'_{ij} = \hat{y}_i \cdot \bar{i}_{сез} = (a + bt_i) \cdot \bar{i}_{сез}.$$

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

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

Для декабря 1998 г.:

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

Средняя ошибка прогноза месячного уровня, вызванная наличием случайной колеблемости, вычисляется по ранее приведенным в разд. 10.3.2 формулам. Для апреля 1999 г.:

$$m_{\hat{y}'_{ij}} = S(t) \cdot \sqrt{1 + \frac{1}{n} + \frac{t_k^2}{\sum t_i^2}};$$

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

Доверительные границы затрат труда в апреле 1999 г. составят с вероятностью 0,95 (-критерий Стьюдента $t_{0,95}^{(2)}$):

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

1998 г.:

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

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 j ; $I^{\wedge} - y$, или в относительном выражении через средние индексы сезонности (их отклонения от 100% или от единицы)?

10.4.2. Модель тренда годовых 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:

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

где y_j - надой молока за j -й месяц;

\hat{y}_j - 316,7 - средний месячный надой в среднем году базы расчета тренда 1989-1997 гг., т.е. в 1993 г.;

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

Имеем также меру случайной колеблемости:

$$S(t)_{\text{случ}} = 11,0.$$

Тренд имеет вид:

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

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:

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

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

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

$$\begin{aligned} \hat{y}_{\text{апр}1999} &= 316,7 + 15,86 \cdot 6 + 45,4 \left(1 + \frac{15,86 \cdot 6}{316,7} \right) \sin 90^\circ - \\ &- 83,25 \left(1 + \frac{15,86 \cdot 6}{316,7} \right) \cos 90^\circ = 316,7 + 95,2 + \\ &+ 59,04 \sin 90^\circ - 108,26 \cos 90^\circ = 470,9 \text{ кг/гол.} \end{aligned}$$

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

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

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:

$$\begin{aligned} \hat{y}_{\text{ноябрь } 2000} &= 316,7 + 15,86 \cdot 7 + 45,4 \left(1 + \frac{15,86 \cdot 7}{316,7} \right) \sin 300^\circ - \\ &\quad - 83,25 \left(1 + \frac{15,86 \cdot 7}{316,7} \right) \cos 300^\circ = 427,7 + \\ &\quad + 61,3 \cdot \sin 300^\circ - 112,4 \cos 300^\circ = 302,4 \text{ кг/гол.} \end{aligned}$$

Средняя ошибка прогноза:

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

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:

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

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 трепда 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 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:

$$\hat{n}_6 = 120 + 5 \cdot 6 = 150 \text{ га}; \quad \hat{y}_6 = 29 + 6 = 35 \text{ ц/га};$$

$$\hat{B}_6 = \hat{I}_6 \cdot \hat{y}_6 = 150 \cdot 35 = 5250 \text{ ц};$$

б) перемножаем тренды площади и урожайности и на основе полученной параболы вычисляем прогноз валового сбора:

$$\hat{B}_t = [(120 + 5t) \cdot (29 + t)] = 3480 + 265t + 5t^2;$$

$$\hat{B}_6 = 3480 + 265 \cdot 6 + 5 \cdot 6^2 = 5250 \text{ ц.}$$

Средняя ошибка выборочной оценки или прогноза произведения двух переменных вычисляется по формулам:

$$\text{а) } m_{\Pi_k, y_k} = \sqrt{(m_{\Pi_k} \cdot \hat{y}_k)^2 + (m_{y_k} \cdot \hat{\Pi}_k)^2}, \quad (10.11)$$

если признаки Π и y независимы друг от друга;

$$(10.12)$$

$$\text{б) } m_{\Pi_k, y_k} = \sqrt{(m_{\Pi_k} \cdot \hat{y}_k)^2 + (m_{y_k} \cdot \hat{\Pi}_k)^2 + 2(m_{\Pi_k} \cdot \hat{y}_k)(m_{y_k} \cdot \hat{\Pi}_k) \cdot r'_{\Pi y}},$$

если признаки Π и y в динамике коррелированы.

Здесь w/\wedge и $ny\wedge$ - средние ошибки прогнозов признаков-сомножителей на период $t\wedge$; r' , - 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: $r'\wedge = -0,326$. Средние ошибки прогнозов отдельного признака рассмотрены в разд. 10.3.2. Имеем:

$$m_{\Pi_k} = S(t)_{\Pi} \cdot \sqrt{1 + \frac{1}{9} + \frac{6^2}{60}} = 7,4 \sqrt{1 + \frac{1}{9} + \frac{6^2}{60}} = 13,16 \text{ га};$$

$$m_{y_k} = S(t)_y \cdot \sqrt{1 + \frac{1}{9} + \frac{6^2}{60}} = 4,4 \sqrt{1 + \frac{1}{9} + \frac{6^2}{60}} = 5,76 \text{ ц/га.}$$

Подставляя в формулу средней ошибки прогноза валового сбора с учетом корреляции колебаний площади с урожайностью, имеем:

$$\begin{aligned} m_{B_6} &= \sqrt{(13,16 \cdot 35)^2 + (5,76 \cdot 150)^2 - 2(13,16 \cdot 35)(5,76 \cdot 150) \cdot 0,326} = \\ &= \sqrt{699179,5} = 836,2 \text{ ц.} \end{aligned}$$

С вероятностью 0,9 ((-критерий Стюдента при семи степенях свободы равен 1,95) доверительный интервал валового сбора составит:

$5250 \pm 1,95 \cdot 836,2$, или от 3619 до 6881 ц.

Широкий интервал прогноза, во-первых, связан с короткой базой расчета трендов, во-вторых, с довольно значительной колеблемостью урожайности.

10.6. Прогнозирование по смешанной трендово-факторной модели

Общий вид трендово-факторной прогностической модели:

$$\tilde{b}_{y, jk} = f(b_{x_1, jk-t}; b_{x_2, jk-t}; \dots; b_{x_i, jk-t}; \dots; b_{x_n, jk-t}), \quad (10.13)$$

где b - средневременной прирост результативного показателя;

b^1, b^2, \dots, b^n - приросты факторов;

k - номер прогнозируемого периода времени;

t - период упреждения.

Тренд-факторные модели позволяют выявить силу влияния приростов 9 СНОВПЫХ факторов и величину прибавки результативного показателя. Смешанное прогнозирование устраняет часть недостатков факторного и авторегрессионного. Оно позволяет включить скачкообразное изменение управляемых факторов.

Рассмотрим трендово-факторную модель прироста урожайности зерновых культур в хозяйствах Оренбургской области:

$$\tilde{b} = 0,0115b_{x_2} + 1,028b_{x_3} + 0,4544b_{x_4} - 0,0794,$$

где b^1 - среднегодовой абсолютный прирост материально-денежных затрат без заработной платы и затрат на удобрения;

b^2 - среднегодовой абсолютный прирост энергообеспеченности;

b^3 - среднегодовой абсолютный прирост доз удобрений.

Подставляем ожидаемые значения приростов 4 факторов в модель:

$$\tilde{b}_{y, 2000} = 0,0115 \cdot 4,0 + 1,028 \cdot 0,055 + 0,4544 \cdot 0,50 - 0,0794 = 0,26 \text{ ц/га.}$$

Среднее квадратическое отклонение среднегодового абсолютного прироста урожайности зерновых культур в хозяйствах Оренбургской области S^{\wedge} равно 0,132 ц/га, а средняя ошибка многофакторной трендово-факторной модели - 0,086 ц/га, тогда средняя ошибка прогноза

$$m_{y_{2000}} = \sqrt{m^2 \bar{b}_{yk} + S^2 \bar{b}_{y(t)k}} = 0,15 \text{ ц/га.}$$

Прогнозируемый среднегодовой абсолютный прирост урожайности зерновых культур в Оренбургской области с вероятностью 0,95 заключен в границах:

$$\tilde{b}_{y_{2000}} \pm 1,96 \cdot 0,15 = 0,26 \pm 0,29 \text{ ц/га.}$$

Из-за большой колеблемости ($U^{\wedge, l} = 62,8\%$) среднегодового абсолютного прироста вероятная ошибка его прогноза превышает величину самого прироста.

Данную методику можно использовать при прогнозировании других результативных показателей например рентабельности, производительности труда, показателей устойчивости производства и финансов.

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. *Андерсон Т.* Статистический анализ временных рядов: Пер. с англ. – М.: Мир, 1976.
2. *Афанасьев В.Н.* Статистическое обеспечение проблемы устойчивости сельскохозяйственного производства. – М.: Финансы и статистика, 1996.
3. *Афанасьев В.Н., Юзбашев М.М.* Коэффициент корреляции рангов как показатель устойчивости динамики // Вестник статистики. – 1983. – № 11.
4. *Бобров С.П.* Конъюнктурные кривые 1922–1923 гг. // Экономическое обозрение. – 1924. – № 5–7.
5. *Вайну Я.Я.Ф.* Корреляция рядов динамики. – М.: Финансы и статистика, 1982.
6. *Гренджер К., Хатакака М.* Спектральный анализ временных рядов в экономике. – М.: Статистика, 1972.
7. *Дубров А.М., Мхитарян В.С., Трошин Л.И.* Многомерные статистические методы. – М.: Финансы и статистика, 1999.
8. *Казинец Л.С.* Темпы роста и абсолютные приросты. – М.: Статистика, 1975.
9. *Каяйкита М.С.* Статистические методы изучения динамики урожайности. – Л.: ЛСХИ, 1969.
10. *Кендэл М.* Временные ряды: Пер. с англ. – М.: Финансы и статистика, 1981.
11. *Красущий О.П.* Изучение статистических зависимостей по многолетним данным. – М.: Финансы и статистика, 1981.
12. *Латин Г.Ф.* Биометрия. – М.: Высшая школа, 1968.
13. *Попова И.* Агрегирование трендов объемных показателей и их колеблемости по совокупности предприятий // Экономико-математические и статистические методы в прогнозировании и управлении сельскохозяйственного производства. – Сб. трудов. – Спб.: Спб. ГАУ, 1993.

14. Статистический анализ экономических временных рядов и прогнозирование: ЦЭМИ АН СССР. – М.: Наука, 1973.
15. Френкель А.А. Математические методы анализа динамики и прогнозирования производительности труда. – М.: Экономика, 1972.
16. Хенан Э. Многомерные временные ряды: Пер. с англ. – М.: Мир, 1974.
17. Четвериков Н.С. Статистические и стохастические исследования. – М.: Госстатиздат, 1963.
18. Четыркин Е.М. Статистические методы прогнозирования. – М.: Статистика, 1977.
19. Юзбаишев М.М. Экспериментальная проверка закономерности распределения отклонений урожайности отдельных лет от тренда по их величине // Науч. труды ЛСХИ. – Т. 322. – Л. – Пушкин. – 1981.
20. Юзбаишев М.М., Манелля А.И. Статистический анализ тенденций и колеблемости. – М.: Финансы и статистика, 1983.
21. Юл Д.Э., Кендэл М.Д. Теория статистики: Пер. с англ. – М.: Госстатиздат, 1960.
22. Ястремский Б.С. Переменная корреляция // Вестник статистики. – 1924. – Кн. XVII.
23. Blanford D. Offut S. A Review of Empirical Techniques for the Analysis of Commodity Instability. – USsl. – 1983.

ПРИЛОЖЕНИЯ

1. Многократное скользящее выравнивание. Урожайность зерновых культур во Франции

Год	Урожайность, ц/га, Y_i	$Y_i t_1$	$Y_i t_2$	$Y_i t_3$	$Y_i t_4$	$Y_i t_5$	$Y_i t_6$	$Y_i t_7$
1970	33,7	-235,9	—	—	—	—	—	—
1971	38,8	-232,8	-271,6	—	—	—	—	—
1972	41,7	-208,5	-250,2	-291,9	—	—	—	—
1973	44,1	-176,4	-220,5	-264,6	-308,7	—	—	—
1974	41,8	-125,4	-167,2	-209	-250,8	-292,6	—	—
1975	37,0	-74,0	-111,0	-148	-185,0	-222,0	-259,0	—
1976	34,4	-34,4	-68,8	-103,2	-137,6	-172,0	-206,4	-204,8
1977	40,5	0	-40,5	-81,0	-121,5	-162,0	-202,5	-243,0
1978	46,3	46,3	0	-46,3	-92,6	-138,9	-185,2	-231,5
1979	45,2	90,4	45,2	0	-45,2	-90,4	-135,6	-180,8
1980	48,5	145,5	97,0	48,5	0	-48,5	-97,0	-145,5
1981	47,3	189,2	141,9	94,6	47,3	0	-47,3	-94,6
1982	50,1	250,5	200,4	150,3	100,2	50,1	0	-50,1
1983	49,5	297,0	247,5	198,0	148,5	99,0	49,5	0
1984	59,7	417,9	358,2	298,5	238,8	179,1	119,4	59,7
1985	57,3	—	401,1	343,8	286,5	289,2	171,9	114,6
1986	53,0	—	—	371,0	318,0	265,0	212,0	159,0
1987	57,0	—	—	—	399,0	342,0	285,0	228,0
1988	60,8	—	—	—	—	425,6	364,8	304,0
1989	61,0	—	—	—	—	—	427,0	366,0
1990	60,8	—	—	—	—	—	—	425,0
1991	65,4	—	—	—	—	—	—	—
1992	64,9	—	—	—	—	—	—	—
1993	65,1	—	—	—	—	—	—	—
1994	65,5	—	—	—	—	—	—	—
1995	63,2	—	—	—	—	—	—	—
Σ	1332,6	349,9	61,4	59,7	96,9	463,6	492,6	470,0

Год	Y, h
1970	-
1971	-
1972	-
1973	-
1974	-
1975	-
1976	-
1977	-283,5
1978	-277,8 -
1979	-226,0 -
1980	-194,0 -
1981	-141,9 -
1982	-100,2 -
1983	-49,5
1984	0
1985	57,3
1986	106,0
1987	171,0
1988	243,2
1989	305,0
1990	364,8
1991	457,8
1992	-
1993	-
1994	-
1995	-
£	432,2

$$b_1 = 1,25; \\ b^{\wedge} = 1,6S; b^{\wedge}$$

$$v_{Y, I} \cdot b_1 = \wedge \\ ; b = -280$$

Продолжение

Год	$Y_i t_8$	$Y_i t_9$	$Y_i t_{10}$	$Y_i t_{11}$	$Y_i t_{12}$	Тренд	$Y_i - y$	$(Y_i - y)^2$
1970	—	—	—	—	—	33,1	0,6	0,36
1971	—	—	—	—	—	34,6	-4,2	17,64
1972	—	—	—	—	—	36,0	5,7	32,49
1973	—	—	—	—	—	37,5	6,6	43,56
1974	—	—	—	—	—	38,9	2,9	8,41
1975	—	—	—	—	—	40,4	-3,4	11,56
1976	—	—	—	—	—	41,8	-7,4	54,76
1977	-283,5	—	—	—	—	43,3	-2,8	7,84
1978	-277,8	-324,1	—	—	—	44,7	1,6	2,56
1979	-226,0	-271,2	-316,4	—	—	46,2	-1,0	1,00
1980	-194,0	-242,5	-291,0	-339,5	—	47,6	0,9	0,81
1981	-141,9	-189,2	-236,5	-283,8	-331,1	49,5	-1,8	3,24
1982	-100,2	-150,3	-200,4	-250,5	-300,6	50,5	-0,4	0,16
1983	-49,5	-99,0	-148,5	-i 98,5	-247,5	52,0	-2,5	6,25
1984	0	-59,7	-119,4	-179,1	-238,8	53,4	6,3	39,69
1985	57,3	0	-57,3	-114,6	-171,9	54,9	2,4	5,76
1986	106,0	53,0	0	-53,0	-106,0	56,3	-3,3	10,89
1987	171,0	П 4,0	57,0	0	-57,0	57,8	-0,8	0,64
1988	243,2	182,4	121,6	60,8	0	59,2	1,6	2,56
1989	305,0	244,0	183,0	122,0	61,0	60,7	0,3	0,09
1990	364,8	304,0	243,2	184,4	121,6	62,1	1,3	1,69
1991	457,8	392,4	327,0	261,6	196,2	63,6	1,8	3,24
1992	—	454,3	389,4	324,5	259,6	65,0	-0,1	0,01
1993	—	—	455,7	390,6	325,5	66,5	-1,4	1,96 j
1994	—	—	—	458,5	393,0	67,9	-2,4	5,76
1995	—	—	—	—	442,4	69,4	-6,2	38,44
Σ	432,2	408,1	407,4	379,9	346,4	1334,5	0	301,37

$b_1 = 1,25; b_2 = 1,29; b_3 = 1,29; b_4 = 1,42; b_5 = 1,66; b_6 = 1,77;$
 $b_7 = 1,68; b_8 = 1,54; b_9 = 1,46; b_{10} = 1,46; b_{11} = 1,36; b_{12} = 1,2$

$b_i = \frac{\sum Y_i t_i}{280}; \bar{b} = \frac{\sum_{i=1}^{12} b_i}{12} = 1,452$. Тренд $\hat{y}_i = 51,25 + 1,452t$, $t = 0,5$ в 1983 г.

**2. Значение *t*-критерия Стьюдента при уровне
значимости 0,10; 0,05; 0,01**

Число степеней свободы, <i>df.</i>	<i>P</i>			<i>df.</i>	<i>P</i>		
	0,10	0,05	0,01		0,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7633
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

3. Значение F-критерия Фишера при уровне значимости 0,05

df_2	df_1																
	1	2	3	4	5	6	7	8	9	10	11	12	14	15	20	30	∞
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248	250	254
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41	19,42	19,43	19,44	19,46	19,50
3	10,13	9,55	9,28	9,19	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74	8,71	8,69	8,66	8,62	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91	5,87	5,84	5,80	5,74	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68	4,64	4,60	4,56	4,50	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,96	3,92	3,87	3,81	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57	3,52	3,49	3,44	3,38	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28	3,23	3,20	3,15	3,08	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07	3,02	2,98	2,93	2,86	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91	2,86	2,82	2,77	2,70	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79	2,74	2,70	2,65	2,57	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69	2,64	2,60	2,54	2,46	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60	2,55	2,51	2,46	2,38	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53	2,48	2,44	2,39	2,31	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48	2,43	2,39	2,33	2,25	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42	2,37	2,33	2,28	2,20	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38	2,33	2,29	2,23	2,15	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,29	2,25	2,19	2,11	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	2,34	2,31	2,26	2,21	2,15	2,07	1,88

Продолжение

df_2	df_1																
	1	2	3	4	5	6	7	8	9	10	11	12	14	15	20	30	∞
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,31	2,28	2,23	2,18	2,12	2,04	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	2,20	2,15	2,09	2,00	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30	2,26	2,23	2,18	2,13	2,07	1,98	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32	2,28	2,24	2,20	2,14	2,10	2,04	1,96	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,22	2,18	2,13	2,09	2,02	1,94	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,41	2,34	2,26	2,24	2,20	2,16	2,11	2,06	2,00	1,92	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,10	2,05	1,99	1,90	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,30	2,25	2,20	2,16	2,13	2,08	2,03	1,97	1,88	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,19	2,15	2,12	2,06	2,02	1,96	1,87	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,35	2,28	2,22	2,18	2,14	2,10	2,05	2,00	1,94	1,85	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,16	2,12	2,09	2,04	1,99	1,93	1,84	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	2,04	2,00	1,95	1,90	1,84	1,74	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,98	1,95	1,90	1,85	1,78	1,69	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,86	1,81	1,75	1,65	1,39
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,88	1,85	1,79	1,75	1,68	1,57	1,28
∞	3,84	2,99	2,60	2,37	2,21	2,09	2,01	1,94	1,88	1,83	1,79	1,75	1,69	1,64	1,57	1,46	1,00

Примечание: df_1 – число степеней свободы для большей дисперсии; df_2 – число степеней свободы для меньшей дисперсии.

4. Значение χ^2 -критерия Пирсона при
уровне значимости 0,10; 0,05; 0,01

<i>d.f.</i>	<i>P</i>			<i>d.f.</i>	<i>P</i>		
	0,10	0,05	0,01		0,10	0,05	0,01
1	2,71	3,84	6,63	20	28,41	31,41	37,57
2	4,61	5,99	9,21	21	29,62	32,67	38,93
3	6,25	7,81	11,34	22	30,81	33,92	40,29
4	7,78	9,49	13,28	23	32,01	35,17	41,64
5	9,24	11,07	15,09	24	33,20	36,42	42,98
6	10,64	12,59	16,81	25	34,38	37,65	44,31
7	12,02	14,07	18,48	26	35,56	38,89	45,64
8	13,36	15,51	20,09	27	36,74	40,11	46,96
9	14,68	16,92	21,67	28	37,92	41,34	48,28
10	16,01	18,31	23,21	29	39,09	42,56	49,59
11	17,28	19,68	24,72	30	40,26	43,77	50,89
12	18,55	21,03	26,22	40	51,80	55,76	63,69
13	19,81	22,36	27,69	50	63,17	67,50	76,15
14	21,06	23,68	29,14	60	74,40	79,08	88,38
15	22,31	25,00	30,58	70	85,53	90,53	100,42
16	23,54	26,30	32,00	80	96,58	101,88	112,33
17	24,77	27,59	33,41	90	107,56	113,14	124,12
18	25,99	28,87	34,81	100	118,50	124,34	135,81
19	27,20	30,14	36,19				