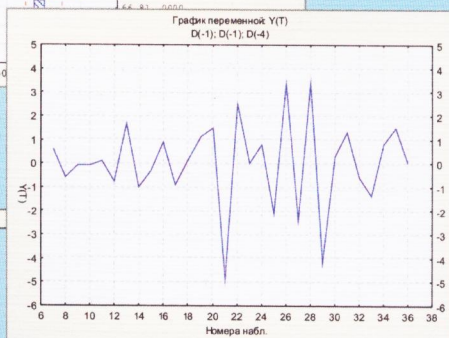
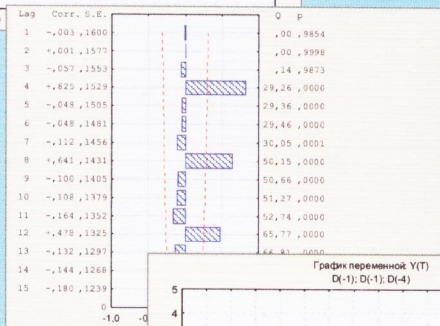
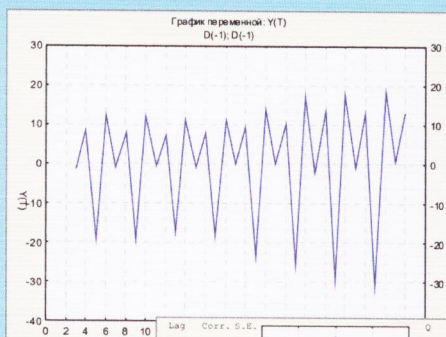




АНАЛИЗ ВРЕМЕННЫХ РЯДОВ И ПРОГНОЗИРОВАНИЕ



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«МОРДОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
им. Н. П. ОГАРЁВА»

Ю. В. САЖИН, А. В. КАТЫНЬ,
Ю. В. САРАЙКИН

АНАЛИЗ ВРЕМЕННЫХ РЯДОВ И ПРОГНОЗИРОВАНИЕ

УЧЕБНИК

*Рекомендовано Учебно-методическим объединением
по образованию в области статистики
в качестве учебника для студентов
высших учебных заведений, обучающихся
по направлению «Статистика»
и другим экономическим специальностям*

САРАНСК
ИЗДАТЕЛЬСТВО МОРДОВСКОГО УНИВЕРСИТЕТА

2013

УДК 311.13(075.8)

ББК У05

С147

Рецензенты:

кафедра статистики и эконометрики
Оренбургского государственного университета
(заведующий — доктор экономических наук профессор *В. Н. Афанасьев*);
заведующий кафедрой статистики Саратовского государственного
социально-экономического университета доктор экономических наук
профессор *В. А. Прокофьев*;
заведующий кафедрой информационных систем в экономике и управлении
доктор экономических наук профессор *Л. И. Зинина*

Сажин Ю. В.

С147

Анализ временных рядов и прогнозирование : учебник /
Ю. В. Сажин, А. В. Катень, Ю. В. Сарайкин. — Саранск :
Изд-во Мордов. ун-та, 2013. — 192 с.
ISBN 978-5-7103-2815-6

Рассмотрены методы статистического анализа основной тенденции, сезонной и случайной компонент временного ряда. Подробно изложены вопросы по прогнозированию ARIMA-процессов и рядов с периодической компонентой, особенности корреляции и регрессии временных рядов. Пособие содержит примеры с необходимыми пояснениями, тесты для осуществления самоконтроля усвоения знаний, практические задания и задания для лабораторных работ на ПЭВМ.

Предназначено для студентов (бакалавров, магистрантов) и аспирантов экономических специальностей вузов.

УДК 311.13(075.8)

ББК У05

ISBN 978-5-7103-2815-6

© Сажин Ю. В., Катень А. В.,
Сарайкин Ю. В., 2013

© Оформление. Издательство
Мордовского университета, 2013

ПРЕДИСЛОВИЕ

В процессе реформирования экономики все в большей степени возрастает быстрота реакции на конъюнктуру внешнего и внутреннего рынка, на правительственные решения, на новые социально-экономические условия. Изменения экономики страны и отдельных территорий находят отражение в закономерностях, основные тенденции и проявления которых необходимо учитывать при разработке прогнозов на различных уровнях принятия решений. К настоящему времени накоплен существенный набор средств как для долгосрочного, так и для средне- и краткосрочного прогнозирования. Правильный выбор метода, оценка его достоверности и экономическая интерпретация результатов прогнозов требуют знакомства с методологией прогнозирования и возможностями конкретного метода.

В учебнике рассмотрены вопросы методологии прогнозирования: акцентировано внимание на совместном использовании методов экспертного и статистического прогнозирования, изложены методы статистического анализа основной тенденции, сезонной и случайной компонент временного ряда, рассмотрены методы прогнозирования на основе временных рядов, особенности коррелирования и регрессии связанных рядов динамики. Содержатся примеры с необходимыми пояснениями, тесты для осуществления самоконтроля усвоения знаний. В конце каждой главы даны практические задания и задания для лабораторных работ на ПЭВМ.

В процессе обучения предлагаемое издание позволит студентам получить набор определенных навыков системного видения и анализа происходящих процессов, выявления существующих тенденций и формализованных причинно-следственных связей.

1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРОГНОЗИРОВАНИЯ В ЭКОНОМИКЕ

1.1. Понятие категорий «прогноз» и «прогнозирование». Классификация прогнозов и методов прогнозирования.

1.2. Этапы прогнозирования социально-экономических явлений и процессов.

1.3. Оценка точности моделей прогнозирования. Верификация прогнозов.

1.1. ПОНЯТИЕ КАТЕГОРИЙ «ПРОГНОЗ» И «ПРОГНОЗИРОВАНИЕ». КЛАССИФИКАЦИЯ ПРОГНОЗОВ И МЕТОДОВ ПРОГНОЗИРОВАНИЯ

Статистическое прогнозирование, наряду с другими видами прогнозирования явлений и процессов, является инструментом управления, используется для планирования и принятия решений на любом уровне. Прогнозирование — одна из наиболее значимых основ для эффективного развития как предприятий и отраслей, так и экономики региона или страны в целом. Вовремя полученный и достоверный прогноз дает возможность избегать тех или иных ситуаций, которые могут негативно повлиять на общее состояние экономики (или конкретного хозяйствующего субъекта), или смягчать их отрицательное воздействие.

Чаще всего под прогнозом понимается научно обоснованное суждение о возможных состояниях объекта в будущем или об альтернативных путях и сроках достижения этих состояний (либо о том и другом). Прогнозы обусловлены многовариантностью описания развития объекта в будущем в зависимости от воздействия условий среды его функционирования. Настоящее и будущее зависят от многих факторов, в том числе случайных, сложное переплетение и сочетание которых учесть практически невозможно. Поэтому все прогнозы имеют вероятностный характер.

Процесс разработки прогноза называется *прогнозированием*. Оно основано на системе установленных причинно-следственных связей и закономерностей и позволяет выявить состояние и вероятностные пути развития явлений и процессов в будущем.

Прогнозы можно подразделять по различным основаниям в зависимости от целей, специфики области применения, масштабности объекта прогнозирования, периода упреждения и т. д.

В зависимости от целей исследования прогнозы делятся на поисковые и нормативные. **Поисковый прогноз** отвечает на во-

прос: «Что вероятнее всего ожидать в будущем?». **Нормативный прогноз** предназначен для указания возможных путей и сроков достижения заданного, желаемого состояния прогнозируемого явления, т. е. разрабатывается на базе заранее определенных целей и задач. Поисковый прогноз не ориентируется на заданную цель, а отталкивается при определении будущего состояния объекта от его прошлого и настоящего. Нормативный же прогноз осуществляется в обратной последовательности: от заданного состояния в будущем к существующим тенденциям и закономерностям. На рис. 1.1 показаны возможные варианты развития объекта при поисковом (а) и нормативном (б) прогнозировании.

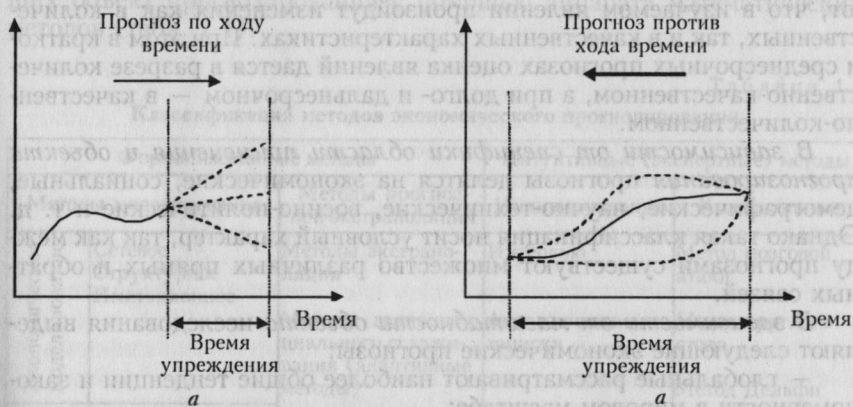


Рис. 1.1. Возможные варианты развития объекта:

а — при поисковом прогнозировании; б — при нормативном прогнозировании

При поисковом прогнозировании важной характеристикой является период, на который делается прогноз (период упреждения). **Период упреждения** — это отрезок времени от момента, для которого имеются последние статистические данные об изучаемом объекте, до момента, к которому относится прогноз (рис. 1.1, а).

По времени упреждения, как правило, выделяются следующие прогнозы социально-экономических явлений:

- оперативные (с периодом упреждения до одного месяца);
- краткосрочные (до одного-двух лет);
- среднесрочные (от трех до пяти лет);
- долгосрочные (от пяти до десяти лет);
- дальнесрочные (свыше 10 лет).

Само понятие долгосрочности прогноза не является чем-то фиксированным, оно зависит от специфики изучаемого объекта исследования, от интенсивности изменения показателей. Например, прогноз на один квартал вперед может рассматриваться как среднесрочный на фондовом рынке и краткосрочный в торговле потребительскими

товарами. Следует помнить, что чем меньше период упреждения и длиннее информационная база (период наблюдений), тем меньше ошибка статистического прогноза. Поэтому при прогнозировании следует руководствоваться таким правилом: период упреждения не должен превышать $\frac{1}{3}$ длины исходной базы.

Прогнозы по времени упреждения отличаются друг от друга по содержанию и характеру оценок исследуемых явлений. Оперативный, или текущий, прогноз основан на том, что в прогнозируемом периоде не произойдет никаких существенных изменений в исследуемом явлении. А если они и совершатся, то будут количественно несущественными. Кратко-, средне-, долгосрочные прогнозы предполагают, что в изучаемом явлении произойдут изменения как в количественных, так и в качественных характеристиках. При этом в кратко- и среднесрочных прогнозах оценка явлений дается в разрезе количественно-качественном, а при долго- и дальнесрочном — в качественно-количественном.

В зависимости от специфики области применения и объекта прогнозирования прогнозы делятся на экономические, социальные, демографические, научно-технические, военно-политические и т. д. Однако такая классификация носит условный характер, так как между прогнозами существуют множество различных прямых и обратных связей.

В зависимости от масштабности объекта исследования выделяют следующие экономические прогнозы:

- глобальные рассматривают наиболее общие тенденции и закономерности в мировом масштабе;
- макроэкономические анализируют наиболее общие тенденции явлений и процессов в масштабе экономики страны в целом;
- региональные определяют развитие отдельных регионов;
- отраслевые предсказывают развитие отраслей;
- структурные (межотраслевые и межрегиональные) рассматривают развитие экономики в разрезе отраслей;
- микроэкономические предсказывают развитие отдельных предприятий, производств и т. д.

Каждый из видов прогноза применяется в конкретной ситуации в зависимости от особенностей объекта прогнозирования, требований, предъявляемых к исходной информации, а также на основе сравнения эффективности и оптимальности решения аналогичных задач.

Отличительная черта многих социально-экономических явлений и процессов — инерционность, которая обнаруживается в сохранении, с одной стороны, взаимосвязей прогнозируемого явления с остальными, а с другой — тенденции во времени. Поэтому при прогнозировании необходимо всесторонне раскрывать причинно-следственные связи и факторы, характеризующие развитие этих явлений, изучать внутренние структурные связи, а также внешнюю среду, в которой они проявляются.

Прогнозирование социально-экономических явлений и процессов осуществляется с помощью различных *методов прогнозирования*. Под методом прогнозирования понимается способ теоретических и практических действий, направленных на разработку прогнозов.

Всю совокупность методов прогнозирования можно разделить на две большие группы: *интуитивные* и *формализованные*. Первая включает методы экспертных оценок с научно обоснованной организацией проведения всех этапов экспертизы и применением количественных методов при оценке суждений экспертов и при формальной обработке результатов. Во вторую группу входят методы прогнозирования, основанные на применении различных моделей, описывающих объект прогнозирования с помощью аппарата математических методов (табл. 1.1).

Таблица 1.1

Классификация методов экономического прогнозирования

Формализованные методы		Интуитивные (экспертные) методы		
Методы моделирования		Методы прогнозной экстраполяции	Индивидуальные	Коллективные
Экономическое моделирование	Сетевое	Методы экстраполяции	Интервью	Метод мозговой атаки
	Структурное		Аналитические записки	Метод круглого стола
	Имитационное	Методы экспоненциального сглаживания (адаптивные методы)		
Статистическое моделирование	Классическая регрессия на основе одного или системы уравнений	Модели авторегрессии и скользящего среднего	Построение сценариев	Матричный метод
Экономико-математическое моделирование	Детерминированное			
	Стохастическое			
	Линейно-динамическое			
	Нейросетевое			
	На основе распознавания объектов			

I. Интуитивные (экспертные) методы базируются на интуитивно-логическом мышлении человека в сочетании с количественными методами оценки и обработки результатов. Они применяются в основном в следующих случаях:

1) объект или экономическое явление не поддается математическому описанию;

- 2) отсутствует информационная база, позволяющая использовать аппарат формализованных методов;
- 3) возникают экстремальные ситуации, требующие принятия быстрых решений;
- 4) существует большая неопределенность среды функционирования объекта;
- 5) отсутствуют финансовые ресурсы, программное обеспечение, квалифицированные кадры для проведения формализованных исследований.

По принципу действия и способу получения информации экспертные методы можно подразделить на *индивидуальные* и *коллективные*. Суть индивидуальных оценок заключается в том, что каждый эксперт дает личную оценку вероятности наступления события. При коллективной формируется обобщенная оценка экспертной группы путем обработки индивидуальных независимых оценок, вынесенных экспертами, входящими в группу.

Существуют две категории экспертов — узкие специалисты и специалисты широкого профиля, обеспечивающие формулирование крупных проблем и построение моделей. Эксперты для прогноза выбираются на основе их репутации среди определенной категории специалистов. Однако не следует забывать и того обстоятельства, что первоклассный специалист не всегда может достаточно квалифицированно рассмотреть и понять общие, глобальные вопросы. Для этой цели нужно привлекать экспертов, хотя и недостаточно узко информированных, но обладающих способностью к дерзанию и воображению.

При индивидуальной оценке к эксперту предъявляются повышенные требования. Это связано с условиями ее проведения. Проблема выбора экспертов имеет два аспекта — выбор отдельного эксперта и выбор группы экспертов. При выборе отдельного эксперта прежде всего должны приниматься во внимание соответствие сферы его компетентности задачам экспертизы, а также уровень квалификации в рассматриваемой области. Определить сферу компетентности и уровень квалификации возможно путем разработки и совершенствования приемов взаимной оценки экспертов и самооценки этих свойств.

На выбор группы экспертов оказывают влияние такие факторы, как соответствие суммарной сферы компетентности группы в оцениваемой области, возможности прогнозистов по организации экспертизы с учетом затрат времени и средств на проведение опроса.

При решении задач формирования экспертной группы необходимо выявить и стабилизировать работоспособную сеть экспертов. Способ стабилизации заключается в следующем. На основе анализа литературы по прогнозируемой проблеме выбирается любой специалист, имеющий несколько публикаций в данной области. К нему обращаются с просьбой назвать десять наиболее компетентных, по его мнению, специалистов по данной проблеме. Затем обращаются одновре-

менно к каждому из десяти названных специалистов с просьбой указать десять наиболее крупных их коллег-ученых. Из полученного списка вычеркиваются десять первоначальных, а остальным рассылаются письма, содержащие указанную ранее просьбу. Данную процедуру повторяют до тех пор, пока ни один из вновь названных специалистов не добавит новых фамилий к списку экспертов, т. е. пока не стабилизируется экспертная сеть. Полученную сеть можно считать генеральной совокупностью специалистов, компетентных в области прогнозируемой проблемы. Однако в силу ряда практических ограничений оказывается нецелесообразным привлекать к экспертизе всех специалистов. Необходимо сформировать репрезентативную выборку из генеральной совокупности.

Индивидуальные экспертные оценки предполагают, что каждый эксперт дает свою оценку вероятности события. Их используют, как правило, при решении узких научных или практических проблем. Выделяют такие индивидуальные методы оценки: интервью, аналитические экспертные оценки, построение сценариев.

Интервью предполагает частную беседу организатора прогнозной деятельности с экспертом. Организатор заранее разрабатывает программу в виде вопросов по перспективному развитию прогнозируемого объекта. Эксперт сразу должен дать заключение по самым разным вопросам. С одной стороны, на него оказывается невольное психологическое давление, с другой — он, не обсуждая проблемы с другим специалистом, не может генерировать принципиально новую идею. При использовании этого метода в разработке прогноза участвуют корреспондент и респондент.

Аналитические записки, или аналитические экспертные оценки. При использовании этого метода эксперт индивидуально проводит все необходимые, на его взгляд, разработки по анализу и прогнозированию объекта. Он может иметь доступ к необходимой информации, справочной литературе, результатам различных исследований и т. д. В этом случае психологическое давление на эксперта минимально. Свои результаты в виде основных доказательных выводов эксперт направляет организатору разработки прогнозов.

Построение сценариев. В экономическом прогнозировании под сценарием понимают описание возможной последовательности событий, которые связывают настоящее и будущее. В сценариях учитываются возможные изменения характеристик среды функционирования (активного прогнозного фона) и их влияние на поведение объекта прогнозирования в новых условиях. Подготовка сценария обычно поручается высококвалифицированному эксперту, который в свою очередь использует методы объединения независимых прогнозов и согласования мнений экспертов. Если понимать сценарий как описание картин будущего с учетом правдоподобных предположений, то для объективного прогноза необходимо иметь несколько сценариев развития событий: оптимистический (при благоприятном воздействии

условий на объект прогнозирования), пессимистический (при неблагоприятном стечении обстоятельств) и средний, наиболее вероятный. Цель сценария — определить генеральное или стратегическое направление развития события. Для поискового прогнозирования разрабатываются трендовые сценарии, а для нормативного — нормативные. В любом случае необходимо учитывать воздействие прогнозного фона (внешней среды) на объект прогнозирования.

Для определения степени согласованности мнений всех экспертов по относительной важности рассматриваемых факторов необходимо рассчитать вес каждого фактора. Для этого можно использовать следующие методы: ранжирование, нормирование, метод последовательного сравнения и др.

После получения весов проводится статистический анализ полученных данных, который заключается в том, что определяется степень согласованности мнений всей группы экспертов по относительной важности рассматриваемых факторов. Мерой согласованности является коэффициент конкордации, который подсчитывается следующим образом.

Результаты опроса представляются в виде матрицы рангов для m экспертов и n факторов или объектов экспертиз. Подсчитываются

суммы для каждого фактора $S_j = \sum_{i=1}^m \rho_{ij}$ (ρ_{ij} — ранг i -го фактора по степени важности у j -го эксперта) и среднее значение этих сумм

по всем факторам $\bar{S} = \frac{\sum_{j=1}^n S_j}{n}$. Далее определяется сумма квадра-

тов отклонений $D = \sum_{j=1}^n (S_j - \bar{S})^2$. Коэффициент конкордации вычисляется как

$$K_k = \frac{12D}{m^2(n^3 - n)}$$

Коэффициент конкордации может находиться в пределах от 0 до 1, причем значение $K_k = 1$ получается в случае полного совпадения мнений экспертов.

Коэффициент конкордации можно подсчитать для оценки степени согласованности мнений экспертов не только по каждому методу, но и между методами; кроме того, можно оценить степень согласованности мнений экспертов при использовании нескольких методов одновременно.

Коллективные экспертные оценки. Метод коллективной экспертной оценки основан на выявлении обобщенной оценки

экспертной группы путем обработки индивидуальных независимых оценок, вынесенных экспертами, входящими в группу. При коллективных экспертных оценках используются методы мозговой атаки, круглого стола, Дельфи, матричный и др.

Одним из наиболее перспективных методов формирования групповой оценки экспертов является метод Дельфи (от названия древнегреческого города Дельфы, известного своим оракулом (прорицателем), жрецы которого предсказывали будущее). Метод представляет собой ряд последовательно осуществляемых процедур, направленных на формирование группового мнения по проблемам, по которым ощущается недостаток информации. Он был разработан и применен впервые в США в 1964 г. сотрудниками научно-исследовательской корпорации РЭНД О. Хелмером и Т. Гордоном.

Для разработки прогноза необходимо сформировать репрезентативную группу экспертов, провести экспертизу (опрос), статистически обработать результаты.

Опрос проводится по заранее разработанным анкетам в несколько туров. Целью первого тура является уточнение перечня событий для прогноза в определенной области. После того, как они идентифицированы (одинаковые объединены, второстепенные исключены), их перечень становится основой второй анкеты.

Во втором туре опроса эксперты непосредственно работают над вопросами в анкетах. Их просят также привести соображения, по которым они считают свои оценки правильными, т. е. указать причины выбора. После второго тура опроса проводят обработку полученных оценок. При статистической обработке результатов опросов экспертов рассчитываются средние значения, дисперсии и доверительные интервалы, коэффициенты вариации оценок по каждому из направлений. Полученные таким образом показатели принимаются за характеристики распределения оценок. Каждому эксперту сообщаются значения этих характеристик. Экспертов, чьи оценки оказались в крайних квартилях, просят мотивировать их, т. е. обосновать причины расхождения с групповым мнением. Эксперты могут приводить любые аргументы или возражения. Они могут пересмотреть свои мнения и при желании исправить оценки. С полученными обоснованиями знакомят остальных экспертов, не указывая при этом, чьи они. Такая процедура позволяет всем экспертам принять в расчет обстоятельства, которые они могли случайно пропустить или которыми пренебрегли во время первого и второго туров опроса.

В третьем туре опроса анкета состоит из перечня событий, групповой медианы событий и верхнего и нижнего квартилей для каждого события, а также сводных данных (аргументов) о причинах оценок. Участников экспертизы просят рассмотреть аргументы и сформулировать новые оценки событий. Если их новая оценка не попала в интервалы между квартилями, полученными во втором туре опроса, то их просят обосновать свою точку зрения и прокомментировать позицию тех, кто придерживается противоположных взглядов.

При необходимости может быть проведено еще несколько туров опроса. Практика показывает, что мнения экспертов достаточно быстро сходятся к единому.

При использовании данного метода следует учитывать следующее:

1) группы экспертов должны быть стабильными и численность их должна удерживаться в разумных рамках;

2) время между турами опросов должно быть не более месяца;

3) вопросы в анкетах должны быть тщательно продуманы и четко сформулированы;

4) число туров должно быть достаточным, чтобы обеспечить всех участников возможностью ознакомиться с причиной той или иной оценки, а также критики этих причин;

5) отбор экспертов должен проводиться систематически;

6) необходимо иметь самооценку компетенции экспертов по рассматриваемым проблемам;

7) нужна формула согласованности оценок, основанная на данных самооценок;

8) следует установить влияние различных видов передачи информации экспертам по каналам обратной связи;

9) необходимо определить влияние общественного мнения на экспертные оценки и на их сходимость.

В некоторых областях науки, техники, технологии, организации производства может быть применен метод эвристического прогнозирования. Так называется метод получения и специализированной обработки прогнозных оценок объекта путем систематизированного опроса высококвалифицированных специалистов (экспертов). Прогнозные экспертные оценки отражают индивидуальное суждение специалиста о перспективах развития его области и основаны на мобилизации профессионального опыта и интуиции. Ход прогнозирования и матрицы ответов, компетенции, специализации и предпочтительности представлены в прил. 1.

II. Формализованные методы базируются на математической теории, которая обеспечивает повышение достоверности и точности прогнозов, значительно сокращает сроки их выполнения, позволяет автоматизировать обработку информации и оценку результатов. Классификация группы формализованных методов пока еще не сложилась. Это объясняется бурным развитием в последнее десятилетие экономико-математических методов, а также появлением новых программных продуктов, значительно облегчающих применение трудоемких формализованных методов.

Обычно среди формализованных методов выделяют две самостоятельные группы: методы прогнозной экстраполяции и методы моделирования. Под экстраполяцией понимается метод научного исследования, заключающийся в распространении выводов, полученных из наблюдений за одной частью явления, на другую его часть. Формаль-

ная экстраполяция базируется на предположении, что в будущем сохранится такая же тенденция развития явления, что наблюдалась в прошлом и настоящем.

Методы моделирования основаны на выявлении и описании взаимосвязей и взаимозависимостей между различными явлениями и прогнозировании с учетом такого влияния. В связи с этим формализованные методы в зависимости от используемой информации можно разделить на два класса: методы прогнозирования по одномерному временному ряду и многофакторное прогнозирование.

1.2. ЭТАПЫ ПРОГНОЗИРОВАНИЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ И ПРОЦЕССОВ

Процесс прогнозирования включает одновременное исследование объекта прогнозирования и среды его функционирования. При этом методология прогнозирования предполагает реализацию следующих этапов.

1-й этап — предпрогнозная ориентация. На данном этапе ставятся цели и определяются задачи прогнозирования. Выделяется объект прогнозирования, рассматриваются его состояние, основные элементы, взаимосвязи и факторы, формирующие исследуемый объект и оказывающие влияние на него; устанавливаются причинно-следственные связи как внутри явления, так и вне него, определяется их статистическое выражение. При характеристике объекта прогнозирования должны соблюдаться следующие методологические принципы:

— принцип системности анализа, который требует рассмотрения объекта прогнозирования как системы взаимосвязанных характеристик объекта и прогностического фона с позиций целей и задач прогнозного исследования;

— принцип природной специфичности, требующий обязательного учета особенностей природы объекта прогнозирования, специфики закономерностей его развития, абсолютных и расчетных значений предельного развития;

— принцип оптимизации описания, который сводится:

а) к реализации оптимального соотношения в использовании формализованных и неформализованных методов;

б) описанию объекта минимальным числом переменных и параметров, обеспечивающих заданную точность и достоверность прогноза;

в) учету информационной ценности исходных данных, т. е. при анализе объекта большее значение придается последней, новой информации о нем и меньшее — устаревшей во времени.

2-й этап — эмпирический. Предполагаются сбор и первичная обработка фактических данных об объектах прогнозирования и факторах, влияющих на него. Определяется возможность получения не-

обходимой и надежной информационной базы, которая может быть как количественной, так и качественной. Исходная информация должна быть как можно более полной. Проверяется сопоставимость данных; выявляются аномальные значения («выбросы»); устраняются, если есть, ошибки регистрации; восполняются пропущенные значения.

3-й этап — аналитический. Включает анализ собранной информации и выбор методов прогнозирования. Определяется круг возможных моделей прогнозирования и осуществляется оценка их параметров. Выбор моделей строится на выявлении особенностей и сложившихся закономерностей в исследуемом явлении. При этом определяются периоды однородности во временном ряду данных и (или) формируются однородные совокупности на основе анализа вариации прогнозируемого объекта.

Исследуется качество выбранных моделей, проводится проверка адекватности их реальному процессу. Выбирается наилучшая из моделей. Возможно, в результате проверки будут выявлены некоторые новые факты, для подтверждения которых необходимо собрать дополнительную информацию, т. е. вернуться ко второму этапу.

4-й этап — этап непосредственного прогнозирования. Выполняется собственно прогноз, т. е. определяются точечные значения и указывается доверительный интервал характеристики объекта прогнозирования. Кроме формализованных методов, чаще всего используемых для прогнозирования, следует применять также экспертные. Если данные прогнозов не противоречат друг другу, их обрабатывают и строят комбинированный прогноз. В противном случае необходимо проанализировать причины, вызвавшие противоречивые результаты, и провести повторное прогнозирование.

5-й этап — верификация и корректировка прогноза. Проводится проверка достоверности, точности прогноза на основе расчета и анализа абсолютных, относительных и средних показателей точности. Надежность прогноза определяется, как правило, величиной доверительных интервалов. Корректировка и уточнение прогноза проводятся на основе его верификации и (или) по мере необходимости при поступлении новых данных об объекте прогнозирования и среды его функционирования.

6-й этап — завершающий. Состоит в принятии решений на основе прогнозной модели и выработке рекомендаций о возможностях ее использования для получения прогнозных оценок (т. е. осуществляется синтез прогнозов). Не стоит забывать о том, что прогнозирование имеет смысл только в том случае, если его результаты помогут лучше управлять процессом и планировать его.

В ходе планирования может обнаружиться, что факторы, которым придавалось большое значение, никак не участвовали в моделях. В этом случае необходимо проверить факторы, оказывающие

влияние на изменчивость исследуемого явления. Важно помнить, что не только результаты прогноза влияют на процесс принятия решений, но и сами решения могут радикальным образом изменить прогноз.

Следует учитывать, что модели прогнозирования не только дают диапазон будущих значений характеристик исследуемого объекта, но и выполняют важную стратегическую функцию: предоставляют спектр различных развитий явления в будущем, предупреждая работников управления о возможных последствиях того или иного события.

1.3. ОЦЕНКА ТОЧНОСТИ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ. ВЕРИФИКАЦИЯ ПРОГНОЗОВ

Основными характеристиками качества статистических моделей, выбранных для прогнозирования, являются показатели точности. Понятие точности характеризует степень близости смоделированных (расчетных по определенной модели прогноза) значений (\tilde{y}) в их совокупности к исходным фактическим данным (y). Чем меньше отклонения теоретических значений от эмпирических данных, тем выше качество модели. Для оценки точности на практике используются коэффициент детерминации, среднее абсолютное отклонение (Mean Absolute Derivation, MAD), средний квадрат ошибок (Mean Squared Error, MSE), средняя ошибка аппроксимации, или средняя относительная ошибка по модулю (Mean Absolute Percentage Error, MAPE) (табл. 1.2).

Таблица 1.2

Формулы для расчета показателей точности модели прогнозирования

Показатель	Формула для расчета
Коэффициент детерминации	$R^2 = 1 - \frac{\sum (y - \tilde{y})^2}{\sum (y - \bar{y})^2}$, или $R^2 = 1 - \frac{\sum (\tilde{y} - \bar{y})^2}{\sum (y - \bar{y})^2}$
Среднее абсолютное отклонение	$\bar{\Delta} = \frac{1}{n} \sum y - \tilde{y} $
Средний квадрат ошибок	$S^2 = \frac{\sum (y - \tilde{y})^2}{n}$
Средняя относительная ошибка по модулю	$\bar{\delta} = \frac{1}{n} \sum \left \frac{y - \tilde{y}}{y} \right \cdot 100\%$

Считается, что чем ближе коэффициент детерминации к единице, тем точнее модель описывает исходный набор данных. При сравнении альтернативных моделей с помощью других показателей необходимо исходить из следующего: чем меньше значение этих характеристик, тем выше точность модели.

В стандартных программах вычисляется скорректированный коэффициент детерминации $R_{ск}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2)$, в котором учитывается количество факторных признаков. Он устроен так, что добавление нового параметра в модель приводит к снижению его значения (штрафуется). Эта особенность выгодно отличает скорректированный коэффициент детерминации от ранее представленных характеристик точности, которые фокусируют внимание исследователя на величине отклонения смоделированных значений от реальных данных и не учитывают степень оптимальности структуры построенной модели.

Под **верификацией прогноза** понимается оценка его достоверности и точности. Ценность прогноза в большей степени определяется его точностью, которая зависит от степени совпадения будущего значения процесса с оценкой этого значения, сделанной заранее. В связи с тем что наблюдение за реальным процессом, как правило, всегда проводится в условиях различных помех, а протекание самих процессов подвержено влиянию разных случайных факторов, нельзя рассчитывать на то, что прогноз будущего значения процесса будет абсолютно точным.

О точности прогноза принято судить по величине его **погрешности (ошибке)** — разности между прогнозируемым и фактическим значениями исследуемого признака (абсолютная ошибка). Также широко используется и относительная ошибка прогноза, выраженная в процентах относительно фактического значения показателя,

$$\delta = \frac{y_t - y_t^*}{y_t} \cdot 100 \%,$$

где y_t, y_t^* — соответственно фактическое и прогнозное значения показателя.

Если значение ошибки больше 0, то это свидетельствует о завышенной прогнозной оценке, если меньше 0 — о заниженной.

Очевидно, что оценка прогноза возможна только в случаях:

- 1) когда период упреждения закончился и исследователь имеет фактические значения прогнозируемого показателя;
- 2) при разработке ретроспективного прогноза.

При ретроспективном (*ex post*) прогнозировании исходные данные временного ряда делятся на две части, так чтобы во второй находились более поздние данные, составляющие обычно примерно 15 %

всей информации. Данные первой части используются для построения модели прогноза, а второй — для оценки точности построенной модели. При ретроспективном прогнозировании следует иметь в виду, что величину ошибки прогноза нельзя рассматривать как доказательство пригодности или непригодности применяемой модели прогнозирования, так как она получена при использовании лишь части имеющихся данных.

На формирование исследуемого явления влияет множество разнообразных факторов, поэтому полное совпадение или значительное расхождение прогнозируемого уровня с фактическим может быть следствием благоприятного или неблагоприятного стечения обстоятельств. В связи с этим при построении прогноза необходимо давать его в виде интервала значений, который зависит как от колеблемости уровней ряда, так и от его исходной длины и периода упреждения. Единичный хороший прогноз может быть получен и по плохой модели, и наоборот. Следовательно, о качестве прогнозов, применяемых методов и моделей можно судить лишь по совокупности сопоставлений прогнозов и их реализаций.

Наиболее простой мерой качества прогнозов при условии, что имеются данные об их реализации, является μ — относительное число случаев, когда фактическая реализация охватывается интервальным прогнозом, к общему числу прогнозов, т. е.

$$\mu = \frac{p}{p+q},$$

где p — число прогнозов, подтвержденных фактическими данными; q — число прогнозов, не подтвержденных фактическими данными.

Когда все прогнозы подтверждаются, то $q = 0$ и $\mu = 1$, если же все прогнозы не подтвердились, то $p = 0$, а следовательно, и μ равны нулю. Таким образом, чем ближе значение μ к единице, тем надежнее считается модель прогноза. Сопоставление μ для разных моделей имеет смысл только при условии, что доверительные вероятности приняты одинаковыми при построении прогнозов по всем моделям.

Для оценки качества прогнозов можно использовать коэффициент несоответствия (расхождения) Гейла:

$$v = \frac{\sqrt{\sum (y_t - y_t^*)^2}}{\sqrt{\sum y_t^2}}.$$

Коэффициент расхождения может быть использован при сопоставлении качества прогнозов, получаемых на основе различных методов и моделей. Коэффициент $v = 0$ в случае совершенного прогнозирования; $v = 1$, когда процесс прогнозирования приводит к той же среднеквадратической ошибке, что и прогноз, полученный путем

экстраполяции; $v > 1$, когда прогноз дает худшие результаты, чем предположение о неизменности исследуемого явления [3; 4].

Надежность прогноза определяется вероятностью реализации соответствующей прогностической оценки. Чем больше такая вероятность, тем выше надежность. Вероятность реализации может быть оценена с помощью экспертного прогнозирования или связана с доверительными интервалами прогноза при использовании статистических методов прогнозирования.

Практика разработки прогнозов опирается на систему методов, среди которых статистические занимают важное место. Решающую роль при статистическом подходе к прогнозированию играет выбор соответствующей модели. Располагая различными моделями, можно получить варианты прогноза, отвечающие определенным условиям и гипотезам, учтенным при построении модели. Вместе с тем, наряду с формализованным прогнозированием, необходимо применять и методы экспертного прогнозирования, так как цель построения статистической модели — не заменить суждение и опыт специалиста, а снабдить его разнообразной статистической информацией, специфическим образом обобщенной и приведенной в систему.

Тест и задания для самоконтроля

Тест

1. Что понимается под категорией «прогноз»:

а) описание, картина, модель намечаемого будущего состояния экономической системы, хозяйства страны, регионов, отраслей, предприятий, компаний;

б) научное, основанное на эмпирических данных вероятностное представление о состоянии объекта прогнозирования на конкретный момент времени при определенных условиях среды его функционирования;

в) предвидение, предсказание, основанное на определенных данных?

2. Какие этапы не включаются в прогнозирование:

а) предпрогнозная ориентация;

б) верификация прогноза;

в) идентификация прогноза?

3. По продолжительности периода упреждения выделяют прогнозы:

а) долгосрочный;

б) интервальный;

в) глобальный.

4. По масштабности объекта прогнозирования выделяют прогнозы:

а) количественные;

- б) макропрогнозы;
- в) общего назначения.

5. По цели проведения выделяют прогнозы:

- а) нормативные;
- б) прогнозы специального назначения;
- в) оперативный прогноз.

6. В каких случаях целесообразно применять экспертные методы прогнозирования:

- а) объект или экономическое явление не поддается математическому описанию;
- б) отсутствует достаточно представительная статистическая выборка;
- в) нет программного обеспечения?

7. Какие требования предъявляются к эксперту:

- а) определенный практический и исследовательский опыт;
- б) отсутствие заинтересованности в конкретных результатах;
- в) высокий уровень владения современными методами прогнозирования?

8. Под методом эвристического прогнозирования понимается метод получения и специализированной обработки прогнозных оценок объекта путем:

- а) систематизированного опроса высококвалифицированных специалистов;
- б) формирования группы экспертов во главе с ведущим;
- в) организации круглого стола, в рамках которого будут согласовываться мнения экспертов с целью выработки единого.

9. Какие коэффициенты можно использовать при определении степени согласованности мнений множества экспертов:

- а) ранговый коэффициент корреляции Спирмена;
- б) ранговый коэффициент корреляции Кендалла;
- в) корреляционное отношение;
- г) коэффициент конкордации?

10. Какие из перечисленных методов не относятся к формализованным:

- а) линейная регрессия по панельным данным;
- б) построение сценариев;
- в) метод эвристического прогнозирования?

11. Какие из перечисленных методов не относятся к экспертным:

- а) имитационное моделирование;
- б) экспоненциальное сглаживание;
- в) метод мозговой атаки?

12. Верификация прогноза — это:

- а) оценка его достоверности и точности;
- б) оценка его значимости;
- в) оценка его объективности.

13. Коэффициент детерминации рассчитывается по формуле:

а) $1 - \frac{\sum(y - \tilde{y})^2}{\sum(y - \bar{y})^2}$; б) $\frac{\sum(y - \tilde{y})^2}{\sum(y - \bar{y})^2}$; в) $\sqrt{1 - \frac{\sum(y - \tilde{y})^2}{\sum(y - \bar{y})^2}}$.

14. Средняя относительная ошибка вычисляется так:

а) $\frac{1}{n} \sum |y - \tilde{y}|$; б) $\frac{1}{n} \sum \left| \frac{y - \tilde{y}}{y} \right| \cdot 100\%$; в) $\frac{y_t - y_t^*}{y_t} \cdot 100\%$.

15. Среднеквадратическая ошибка рассчитывается по формуле:

а) $\frac{1}{n} \sum \left| \frac{y - \tilde{y}}{y} \right| \cdot 100\%$; б) $\frac{\sum(y - \tilde{y})^2}{n}$; в) $\sqrt{\frac{\sum(y - \tilde{y})^2}{n}}$.

16. Абсолютная ошибка прогноза вычисляется так:

а) $\frac{1}{n} \sum |y - \tilde{y}|$; б) $\frac{y_t - y_t^*}{y_t} \cdot 100\%$; в) $y_t - y_t^*$.

17. При каком значении средней относительной ошибки модель имеет более высокую точность:

- а) менее 10 %; б) выше 10 %; в) выше 33 %?

18. Какая из характеристик точности учитывает количество включенных в модель параметров:

- а) коэффициент детерминации;
б) средняя относительная ошибка;
в) среднеквадратическая ошибка;
г) скорректированный коэффициент детерминации?

Практические задания

1. Составьте схему поэтапного прогнозирования для какого-либо социально-экономического явления или процесса (прил. 2).

2. По списку тем, представленных в прил. 2, определите объект исследования; показатели, характеризующие его состояние и развитие; охарактеризуйте факторы, влияющие на развитие исследуемого явления (факторы прогнозного фона), и установите показатели, характеризующие влияние этих факторов.

3. Предложите методы для проведения статистического анализа и прогнозирования какого-либо социально-экономического явления или процесса.

4. По приведенным в таблице фактическим и расчетным по разным моделям данным определите, какая из моделей является более точной?

Год	2006	2007	2008	2009	2010
Фактические значения	22,0	25,0	26,0	28,0	28,0
Расчетные значения по модели 1	22,8	24,3	25,8	27,3	28,8
Расчетные значения по модели 2	22,1	24,7	26,5	27,7	28,0

5. В таблице приведены данные об объеме перевозок грузов железнодорожным транспортом в регионе и расчетные значения, полученные по двум моделям (млн т).

Квартал	Исходные значения	Расчетные значения	
		I модель	II модель
I квартал 2010 г.	267	275	260
II квартал 2010 г.	267	253	275
III квартал 2010 г.	258	250	253
IV квартал 2010 г.	262	269	278
I квартал 2011 г.	253	253	263
II квартал 2011 г.	257	248	251
III квартал 2011 г.	263	250	269

Сравните точность I и II моделей на основе средней относительной ошибки по модулю. Сделайте вывод.

2. ВРЕМЕННЫЕ РЯДЫ И ИХ ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ

- 2.1. Понятие и виды временных рядов.
- 2.2. Требования, предъявляемые к временным рядам.
- 2.3. Основные показатели динамики экономических явлений.
- 2.4. Компонентный состав временного ряда.

2.1. ПОНЯТИЕ И ВИДЫ ВРЕМЕННЫХ РЯДОВ

Основной задачей, решаемой при проведении любого статистического исследования, выступает определение объективных закономерностей развития социально-экономических явлений и процессов. Для этого используется информация, характеризующая их развитие во времени, т. е. временные ряды данных.

Временным рядом (time series, динамическим рядом, рядом динамики) называется ряд расположенных в хронологической последовательности значений статистического показателя, характеризующего изменение явления во времени. Процесс изменения социально-экономических явлений во времени заключается в том, что происходит изменение воздействия на это развитие многих факторов социально-экономического, технологического и любого другого характера, а фактор времени аккумулирует их влияние. Всякий временной ряд включает конкретные значения показателя, которые называются уровнями ряда и обозначаются через y_t . Характеристика t указывает период, за который или по состоянию на который приводятся числовые значения показателя.

Анализ временных рядов дает возможность проследить развитие явления, показать его основные тенденции. Выбор соответствующих приемов и способов анализа зависит от задач исследования и определяется характером исходных данных. Поэтому, приступая к анализу временных рядов, важно правильно их классифицировать.

1. В зависимости от способа выражения показателя временные ряды подразделяются на ряды **абсолютных, относительных и средних** величин. Если уровни ряда представляют собой не значения, непосредственно наблюдаемые, а величины средние или относительные, то такие ряды называются производными. Уровни этих временных рядов получаются с помощью некоторых вычислений на основе абсолютных показателей (например, временные ряды индекса цен).

2. В зависимости от характера изучаемых показателей временные ряды разделяют на **интервальные** и **моментные**. В интервальных

каждый уровень характеризует величину изучаемого явления за некоторый интервал времени (день, месяц, год и т. д.), в моментных — на определенный момент времени (например, стоимость основных фондов на начало года или на начало каждого квартала). Из различного характера интервальных и моментных рядов вытекают некоторые особенности их уровней. Так, отдельные уровни моментного ряда абсолютных величин содержат элементы повторного счета, т. е. в каждом последующем уровне содержится полностью или частично значение предыдущего уровня. Все это делает бессмысленным суммирование значений уровней моментного ряда. Значения же уровней абсолютных интервальных временных рядов, в отличие от уровней моментного ряда, не содержат повторного счета, их можно просуммировать, что позволяет получать ряды более укрупненных периодов, или ряды с нарастающими итогами, которые получаются путем последовательного суммирования смежных уровней. Эти особенности определяют способы расчета средних величин за исследуемый период. В табл. 2.1 представлен интервальный временной ряд, в табл. 2.2 — моментный.

Таблица 2.1

Ввод в действие жилых домов по региону, м²

Год, t	2007	2008	2009	2010	2011	2012
y_t	213,2	242,8	284,1	276,6	282,3	290,6

Таблица 2.2

Численность постоянного населения региона на 1 января, тыс. чел.

Год, t	2007	2008	2009	2010	2011	2012
y_t	856,8	847,6	840,4	833,0	826,5	821,0

3. В зависимости от расстояния между уровнями временные ряды подразделяются на ряды с **равностоящими** и **неравностоящими** уровнями во времени. На практике для анализа и дальнейшего прогнозирования социально-экономических явлений чаще всего приходится иметь дело с рядами с равностоящими уровнями.

4. В зависимости от наличия тенденции в изучаемом явлении временные ряды подразделяют на **стационарные** и **нестационарные**. Стационарные процессы протекают во времени приблизительно однородно и имеют вид непрерывных случайных колебаний вокруг некоторого среднего значения, причем ни характеристика этих колебаний, ни средняя амплитуда со временем существенно не изменяются. Любой стационарный процесс можно рассматривать как продолжающийся во времени неопределенно долго. Поэтому при проведении исследований в качестве начала отсчета можно выбрать любой момент, причем на любом временном интервале должны быть получены одни и те

же вероятностные характеристики, т. е. математическое ожидание и дисперсия у стационарного процесса постоянны.

В экономической практике в большинстве случаев приходится анализировать процессы, имеющие вполне определенную тенденцию развития во времени. Такие процессы называются нестационарными, и временные ряды тоже именуют нестационарными. Характеристики этих процессов меняются во времени, т. е. зависят от начала отсчета. В ряде случаев такие процессы на определенных интервалах времени условно могут быть приняты за стационарные. В этом случае на подобные условно стационарные процессы могут быть перенесены все основные выводы стационарных случайных процессов. Те процессы, которые содержат явно выраженную тенденцию, можно преобразовать в стационарные путем исключения основной тенденции и периодических колебаний, если они имеются.

2.2. ТРЕБОВАНИЯ, ПРЕДЪЯВЛЯЕМЫЕ К ВРЕМЕННЫМ РЯДАМ

Анализ развития явления и его дальнейшее прогнозирование, как правило, опираются на методы, которые предъявляют определенные требования к исходной информации.

1. Основным из условий, необходимых для правильного отражения временным рядом процесса, является **сопоставимость** его уровней. Для несопоставимых уровней проводить исследование динамики неправомерно. К несопоставимости приводят изменение состава изучаемой совокупности, переход к другим единицам измерения, изменение методологии учета и расчета показателей, инфляционные процессы и т. п. Временные ряды являются несопоставимыми и в том случае, когда показатели относятся к различным периодам года или приведены на разные даты.

При обнаружении несопоставимости уровней ряда должна применяться процедура смыкания рядов. Устранить несопоставимость, вызванную указанными причинами, можно путем умножения значений показателя в предыдущие периоды на коэффициент перевода, равный отношению показателей на тот момент времени, когда изменились условия формирования уровней ряда. Однако такая процедура не всегда может быть проведена, и не всегда обеспечивает требуемую точность, а в ряде случаев способна привести к снижению ценности исходной информации, а следовательно, и к затруднению дальнейшего анализа. Другой способ устранения несопоставимости: уровень переходного периода принимается для 2-й части ряда за 100 % и от этого уровня определяются соответствующие показатели (вперед или назад). При этом получается сопоставимый ряд относительных величин.

2. Некоторые методы анализа и прогнозирования предъявляют достаточно **жесткие требования к длине ряда**, поэтому иногда

вместо годовых данных желательнее взять поквартальные или помесечные. Но, безусловно, вопрос о выборе интервала времени между уровнями ряда должен решаться исходя из целей каждого конкретного исследования. Так, для успешного изучения сезонных колебаний на базе поквартальных или помесечных данных желательнее иметь информацию не менее чем за 3 года. Использование стандартных программных пакетов для исследования периодичности иногда требует большей длины информации — не менее пяти полных периодов колебаний.

Применение математического аппарата также накладывает ограничение на допустимую длину временных рядов при использовании их в регрессионном анализе. Для такого анализа нужно иметь временные ряды, длина которых в несколько раз (не менее чем в 4 раза) превосходит количество независимых переменных в уравнении регрессии.

3. Для анализа динамики явления необходимо, чтобы информация была **полной**, т. е. временные ряды не должны иметь пропущенных значений, которые могут объясняться как недостатком данных при сборе информации, так и произошедшими изменениями в системе отчетности.

4. Важный этап при построении качественных прогнозных моделей — выявление во временном ряду **аномальных** (нехарактерных) значений. Обычно их можно обнаружить визуально, при помощи графического представления временных рядов, но, прежде чем «исправлять» выявленные таким образом значения ряда, их необходимо подвергнуть качественному и количественному анализу.

Нехарактерные уровни во временном ряду можно подразделить на три группы:

- значения, отражающие объективное развитие процесса, но сильно отличающиеся от общей тенденции, так как они проявляют свои экстремальные воздействия крайне редко;

- значения, возникающие вследствие изменений методики расчета;

- значения, обусловленные ошибками при измерении показателя, при записи и передаче информации, а также связанные с различными катастрофическими явлениями, не влияющими на дальнейшее развитие процесса.

Аномальные значения первой группы не всегда должны исключаться из временного ряда и могут даже оказаться полезными на этапе исследования причинно-следственного механизма развития явления. Наличие нехарактерных пиковых значений для одного и того же момента времени в различных временных рядах свидетельствует, как правило, о причинных связях между соответствующими показателями. Аномальные значения второй группы следует не исключать из рассмотрения, а принимать за «поворотные» (пороговые), начиная с которых должны быть пересчитаны по новой методике все предыдущие значения временного ряда.

Аномальные значения третьей группы должны быть исключены из рассмотрения в любом случае, так как они искажают представление о характере развития явления и могут оказать существенное влияние на выводы, полученные в результате анализа ряда, содержащего такую информацию. Выявление указанных значений, замена их истинными или расчетными — необходимый этап первичной обработки данных. Если аномальные значения отражают реальное развитие процесса, то при построении моделей они заменяются расчетными, но учитываются при вычислении возможных величин отклонений фактических данных от полученных по модели.

Соответствие исходной информации всем указанным требованиям проверяется на этапе предварительного анализа временных рядов. Лишь после этого переходят к расчету и анализу основных показателей динамики, построению моделей, получению прогнозных оценок.

2.3. ОСНОВНЫЕ ПОКАЗАТЕЛИ ДИНАМИКИ ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ

После приведения временных рядов к сопоставимому виду (если это необходимо) и обнаружения аномальных наблюдений нужно определить сложившиеся закономерности в динамике исследуемых явлений (это одна из главных целей при статистическом анализе социально-экономических процессов). Закономерности выявляются с помощью показателей динамики, формулы которых представлены в табл. 2.3. Все эти показатели могут быть рассчитаны либо базисным, либо цепным способом. В основе расчета лежит сравнение уровней временного ряда. Если оно осуществляется с одним и тем же показателем (y_0), принятым за базу сравнения, то показатели называются базисными (во временных рядах часто за базу сравнения берется первый уровень ряда y_1). Если каждый уровень (y_t) сопоставляется с предыдущим (y_{t-1}), то вычисленные показатели называются цепными.

Таблица 2.3

Основные показатели динамики

Показатель динамики	Способ расчета	
	базисный	цепной
Абсолютный прирост	$\Delta y_t = y_t - y_0$	$\Delta y_t = y_t - y_{t-1}$
Темп роста	$T_p = \frac{y_t}{y_0} \cdot 100\%$	$T_p = \frac{y_t}{y_{t-1}} \cdot 100\%$
Темп прироста	$T_{пр} = T_p - 100\% = \frac{y_t - y_0}{y_0} \cdot 100\%$	$T_{пр} = T_p - 100\% = \frac{y_t - y_{t-1}}{y_{t-1}} \cdot 100\%$

Абсолютный прирост определяется как разность двух сравниваемых уровней и показывает скорость изменения уровней ряда. Если цепные абсолютные приросты со временем возрастают, то исследуемое явление развивается с ускорением. **Ускорение** абсолютного прироста — это разность между абсолютными приростами, вычисленными цепным способом. Оно позволяет увидеть, насколько данная скорость (абсолютный прирост) больше или меньше предыдущей.

Темп роста характеризует отношение двух сравниваемых уровней ряда, выраженное в процентах. Показатели, представленные коэффициентами, принято называть коэффициентами роста. Цепные темпы роста характеризуют интенсивность изменения уровней от года к году, базисные — фиксируют интенсивность роста (снижения) за весь интервал времени между текущими и базисными уровнями.

Темп прироста характеризует абсолютный прирост в относительных величинах. Определенный в процентах, он показывает, на сколько процентов изменился сравниваемый уровень по отношению к принятому за базу сравнения.

Если систематически повышаются цепные темпы роста или прироста, то ряд развивается с **относительным ускорением**, которое можно определить как разность следующих друг за другом темпов роста или прироста. Полученная величина выражается в процентных пунктах (п. п.).

Чтобы знать, что скрывается за каждым процентом прироста, рассчитывается **абсолютное значение 1 % прироста** как отношение абсолютного прироста уровня за интервал времени к темпу прироста за этот же промежуток времени:

$$\Pi = \frac{\Delta y_t}{T_{\text{пр}}} = \frac{\Delta y_t}{\frac{\Delta y_t}{y_{t-1}} \cdot 100} = \frac{y_{t-1}}{100} = 0,01 y_{t-1},$$

т. е. абсолютное значение 1 % прироста в данном периоде есть сотая часть уровня предыдущего периода.

Для получения обобщающих показателей динамики развития определяются средние величины: средний абсолютный прирост, средние темпы роста и прироста. Эти величины часто используются при прогнозировании.

Величина **среднего абсолютного прироста** показывает скорость развития явления за изучаемый интервал времени и вычисляется по формуле

$$\overline{\Delta y} = \frac{\Delta y_2 + \Delta y_3 + \dots + \Delta y_n}{n-1} = \frac{y_n - y_1}{n-1}, \quad (2.1)$$

где n — длина временного ряда (количество его уровней).

Для определения средней скорости изменения явления за рассматриваемый период времени вычисляют **средний темп роста**.

Этот показатель рассчитывают по формуле средней геометрической:

$$\bar{T}_p = n\sqrt{\frac{y_2 y_3 y_4 \dots y_n}{y_1 y_2 y_3 \dots y_{n-1}}} \cdot 100 \% = n\sqrt{\frac{y_n}{y_1}} \cdot 100 \%. \quad (2.2)$$

Средний темп прироста соответственно равен:

$$\bar{T}_{пр} = \bar{T}_p - 100 \%.$$

Показатель среднего темпа роста, рассчитываемый по формуле средней геометрической, основан на сопоставлении конечного и начального уровней временного ряда, а промежуточные уровни во внимание не принимаются. Поэтому для практических расчетов показатель, полученный по этой формуле, можно использовать в случае более или менее равномерного изменения явления.

К расчету среднего уровня чаще всего прибегают для рядов, изменение которых стабилизируется в течение большого периода времени (например, средний уровень урожайности за ряд лет), а также для рядов с колеблющимися уровнями за короткие промежутки времени (например, численность работников предприятия изменяется каждый день и поэтому для отражения его работы рассчитывается показатель средней численности работников).

Средний уровень интервального ряда с равностоящими уровнями определяется по формуле средней арифметической:

$$\bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n} = \frac{\sum y_t}{n}. \quad (2.3)$$

Средний уровень моментного ряда с равностоящими уровнями рассчитывается по формуле средней хронологической:

$$\bar{y} = \frac{1/2 y_1 + y_2 + y_3 + \dots + y_{n-1} + 1/2 y_n}{n-1}. \quad (2.4)$$

Если определяются средние уровни для рядов с неравностоящими уровнями, то используются формулы взвешенных средних величин.

Исчерпывающий анализ уровней ряда и его динамики требует параллельного использования всех приведенных выше показателей. Анализ, основанный на каком-либо одном из них, имеет односторонний характер и может привести к ошибочным выводам.

При анализе многолетней динамики рядов на базе поквартальных или помесячных данных показатели, представленные в табл. 2.3, рассчитываются не по исходным уровням ряда, а по итоговым значениям за каждый год.

2.4. КОМПОНЕНТНЫЙ СОСТАВ ВРЕМЕННОГО РЯДА

Временной ряд есть последовательность, в которой каждое значение содержит в себе прошлое для последующих состояний. Очевидно, что любая попытка предвидеть будущее без исследования динамических рядов прошлого является малообоснованной и ошибочной. Поэтому для получения достаточно точных и надежных прогнозов необходимо подробно изучить настоящее и прошлое состояния явления или процесса.

Исследование скорости и интенсивности развития временных рядов при помощи показателей динамики часто не позволяет сразу определить основную тенденцию поступательного движения изучаемого явления. Это зависит от того, что уровни временного ряда со временем меняются, колеблются, но подобная колеблемость неодинакова и может быть вызвана влиянием различного рода факторов. Принято выделять 4 типа таких факторов:

1) **долговременные**, формирующие общее направление развития (тенденцию) в изменении анализируемого явления; обычно эта тенденция описывается с помощью той или иной монотонной неслучайной функции, которая называется функцией тренда или просто трендом;

2) **сезонные**, формирующие периодически повторяющиеся в определенное время года колебания анализируемого явления;

3) **циклические**, формирующие изменения изучаемого явления, обусловленные действием долговременных циклов. Такие изменения похожи на сезонные тем, что они являются повторяющимися и волнообразными, но отличаются большей длительностью циклов. Циклические колебания в экономических процессах примерно соответствуют циклам конъюнктуры;

4) **нерегулярные**, не поддающиеся учету и регистрации факторы, которые для социально-экономических явлений делятся на 2 группы:

а) внезапные, приводящие к скачкообразным структурным изменениям в изучаемом процессе (например, война, эпидемия, экологические катастрофы и т. д.);

б) случайные, являющиеся результатом действия большого количества относительно слабых второстепенных факторов.

Таким образом, временной ряд можно представить как функцию четырех компонент, характеризующих действие вышеприведенных факторов. В зависимости от того, как взаимосвязаны факторы между собой, выделяют следующие структурные схемы формирования значений ряда:

$$Y_t = F_t + S_t + C_t + E_t \text{ — аддитивная схема;}$$

$$Y_t = F_t \cdot S_t \cdot C_t \cdot E_t \text{ — мультипликативная схема;}$$

$$Y_t = F_t \cdot S_t \cdot C_t + E_t \text{ — смешанная схема.}$$

Здесь F_t — трендовая компонента, характеризующая действие долговременных факторов;

S_t — сезонная компонента;
 C_t — циклическая компонента;
 E_t — случайная компонента.

Следует отметить, что уровни временного ряда не всегда являются составляющими всех компонент одновременно. Однако во всех случаях во временных рядах предполагается непрямое участие случайной компоненты, которая может быть представлена в сочетании с определенной тенденцией или с периодическими колебаниями. Примером временного ряда, не содержащего основную тенденцию и периодические составляющие (т. е. когда $Y_t = E_t$), может служить ряд, отображающий урожайность ячменя (рис. 2.1).

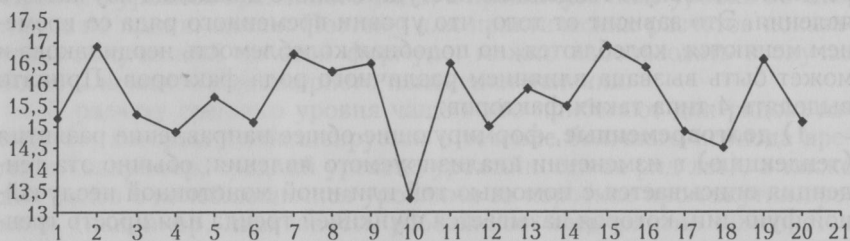


Рис. 2.1. Урожайность ячменя в Англии и Уэльсе с 1884 по 1904 г. [1]

Компоненты временных рядов, имеющих в своем развитии периодические колебания, нарастающих на тренд, могут иметь аддитивный или мультипликативный тип связи. Особенность аддитивной модели заключается в том, что характер периодических колебаний, т. е. их амплитуда, не изменяется во времени (рис. 2.2).

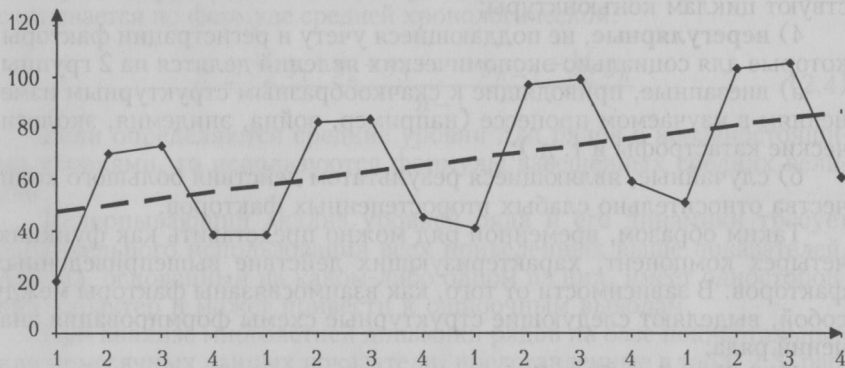


Рис. 2.2. Аддитивная модель временного ряда

При мультипликативном типе связи амплитуда колебаний растет с увеличением среднего значения ряда (рис. 2.3) и, наоборот, уменьшается при снижающейся тенденции.

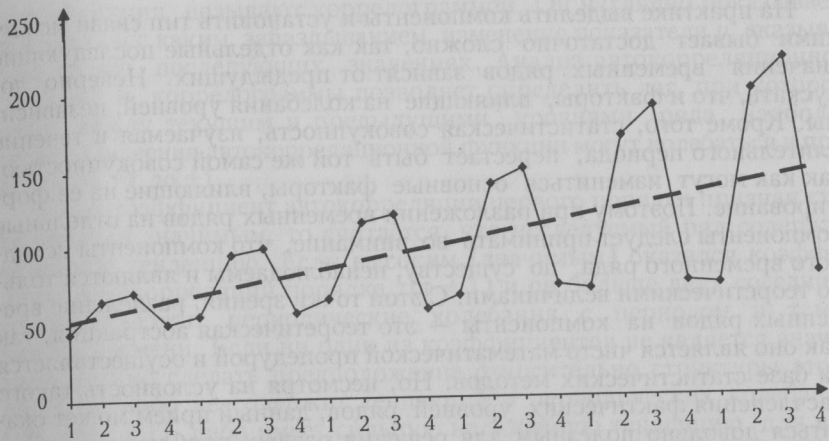


Рис. 2.3. Мультипликативная модель временного ряда

Тип связи между компонентами временного ряда можно определить по нормальному распределению отклонений фактических уровней временного ряда от теоретических, полученных по уравнению тренда. В случае нормальности распределения абсолютных отклонений связь является аддитивной, а относительных — мультипликативной.

Чаще на практике рассматриваются временные ряды, в которых можно выделить основную тенденцию и случайную составляющую (особенно при использовании годовых данных, в которых влияние сезонности не отражается). Аналитически данное положение можно выразить через модель вида $Y_t = F_t + E_t$.

На рис. 2.4 представлен временной ряд со снижающейся тенденцией.

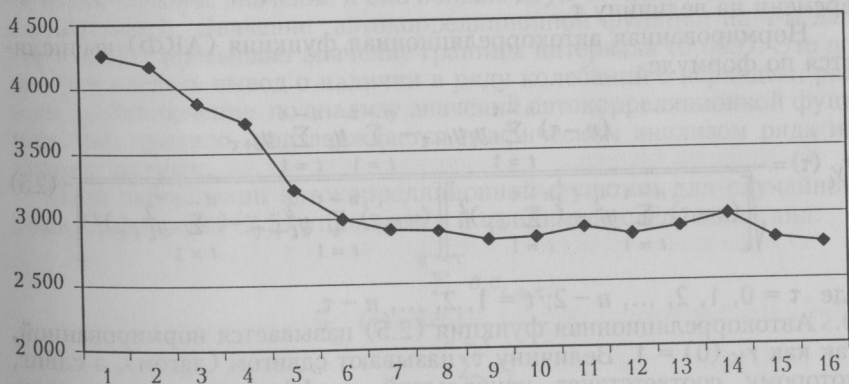


Рис. 2.4. Временной ряд с явно выраженной тенденцией к снижению

На практике выделить компоненты и установить тип связи между ними бывает достаточно сложно, так как отдельные последующие значения временных рядов зависят от предыдущих. Неверно допускать, что и факторы, влияющие на колебания уровней, независимы. Кроме того, статистическая совокупность, изучаемая в течение длительного периода, перестает быть той же самой совокупностью, так как могут измениться основные факторы, влияющие на ее формирование. Поэтому при разложении временных рядов на отдельные компоненты следует принимать во внимание, что компоненты исходного временного ряда, по существу, ненаблюдаемы и являются только теоретическими величинами. С этой точки зрения, разделение временных рядов на компоненты — это теоретическая абстракция, так как оно является чисто математической процедурой и осуществляется на базе статистических методов. Но, несмотря на условность такого расчленения фактических уровней рядов, данный прием может оказаться довольно полезным для решения разных проблем анализа и прогнозирования на базе временных рядов.

Практическое решение любой задачи по анализу и прогнозированию временных рядов, как правило, начинается с построения графика динамики исследуемого показателя, тем более что современные программные средства предоставляют пользователю большие возможности для этого. На стадии проведения графического анализа можно исследовать компонентный состав временных рядов, а также сделать первые шаги к выбору модели для описания их динамики и последующего прогнозирования.

При наличии тенденции и периодических колебаний значения каждого последующего уровня ряда зависят от предыдущих. Поэтому при характеристике структуры ряда (т. е. при определении компонентного состава) можно применять автокорреляционную функцию $r_{Y_t}(\tau)$, которая представляет собой множество коэффициентов корреляции между временным рядом Y_t и этим же рядом, сдвинутым во времени на величину τ .

Нормированная автокорреляционная функция (АКФ) вычисляется по формуле

$$r_{Y_t}(\tau) = \frac{(n-\tau) \sum_{t=1}^{n-\tau} y_t y_{t+\tau} - \sum_{t=1}^{n-\tau} y_t \sum_{t=1}^{n-\tau} y_{t+\tau}}{\sqrt{\left[(n-\tau) \sum_{t=1}^{n-\tau} y_t^2 - \left(\sum_{t=1}^{n-\tau} y_t \right)^2 \right] \left[(n-\tau) \sum_{t=1}^{n-\tau} y_{t+\tau}^2 - \left(\sum_{t=1}^{n-\tau} y_{t+\tau} \right)^2 \right]}}, \quad (2.5)$$

где $\tau = 0, 1, 2, \dots, n-2$; $t = 1, 2, \dots, n-\tau$.

Автокорреляционная функция (2.5) называется нормированной, так как $r_{Y_t}(0) = 1$. Величину τ называют сдвигом (лагом), а сдвиг, которому соответствует наибольший коэффициент автокорреляции, — временным лагом. График нормированной автокорреляцион-

ной функции называют коррелограммой. Он наглядно показывает, как часто и с каким запаздыванием изменение показателя y_t сказывается на его последующих значениях. Анализ автокорреляционной функции и коррелограммы позволяет определить лаг, при котором связь между текущим и предыдущими уровнями ряда наиболее тесная. Значения автокорреляционной функции могут колебаться в пределах от -1 до $+1$.

Если коэффициент автокорреляции первого порядка признан статистически значимым, то считается, что исследуемый ряд содержит основную тенденцию. Если высоким (значимым) оказался коэффициент автокорреляции порядка τ ($\tau \neq 1$) и он больше двух соседних, то ряд содержит периодические колебания с периодом в τ моментов времени. Если ни один из коэффициентов не является значимым, можно сделать предположение относительно структуры этого ряда: он либо не содержит тенденций и периодических колебаний (см. рис. 2.1), либо имеет сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ [11; 36].

Например, для некоторого ряда Y_t были рассчитаны значения автокорреляционной функции $r(\tau)$ по формуле (2.5). Результаты приведены в табл. 2.4.

Таблица 2.4

Значения автокорреляционной функции

Сдвиг (лаг)	1	2	3	4	5	6	7
Значения автокорреляционной функции	0,208	-0,464	0,115	0,611	0,030	0,456	0,053
95% граница	0,490	0,511	0,603	0,609	0,742	0,742	0,806

Из таблицы видно, что автокорреляционная функция на 4-м сдвиге имеет высокое значение и оно больше двух соседних (т. е. является пиковым). Значение автокорреляционной функции на 4-м лаге ($r_4 = 0,611$) превышает значение границы интервала (0,609). Это позволяет сделать вывод о наличии в ряду колебаний с периодом, равным 4. Заключение по анализу значений автокорреляционной функции, как правило, подтверждается графическим анализом ряда исходных данных.

При вычислении автокорреляционной функции для случайного ряда E_t формула (2.5) упрощается и принимает следующий вид:

$$r_{E_t}(\tau) = \frac{\sum_{t=1}^{n-\tau} e_t e_{t+\tau}}{\sum_{t=1}^{n-\tau} e_t^2} \quad (2.6)$$

Если присутствие основной тенденции во временном ряду прослеживается нечетко, то, прежде чем перейти к определению типа тенденции и выделению тренда, нужно выяснить, существует ли вообще тенденция в исследуемом процессе.

Тест и задания для самоконтроля

Тест

1. Временной ряд характеризует:

- а) структуру совокупности по какому-либо признаку;
- б) изменение характеристики совокупности во времени;
- в) изменение характеристики совокупности в пространстве.

2. Показатели изменения уровней ряда динамики, исчисленные с переменной базой сравнения, называются:

- а) базисными; б) цепными.

3. Разность уровней временного ряда именуется:

- а) абсолютным приростом;
- б) темпом роста;
- в) темпом прироста;
- г) коэффициентом роста.

4. Отношение уровней ряда динамики называется:

- а) абсолютным приростом;
- б) средним уровнем;
- в) коэффициентом роста;
- г) абсолютным значением одного процента прироста.

5. Абсолютное значение одного процента прироста характеризует:

- а) абсолютную скорость роста (снижения) уровней ряда динамики;
- б) интенсивность изменения уровней;
- в) относительное изменение абсолютного прироста уровня ряда динамики;
- г) содержание одного процента прироста в абсолютном выражении.

6. Каждый базисный абсолютный прирост равен:

- а) сумме последовательных цепных абсолютных приростов;
- б) разности соответствующих базисных абсолютных приростов;
- в) произведению цепных абсолютных приростов.

7. По формуле $T = \frac{\Delta y}{y_{i-1}}$ определяется:

- а) базисный темп роста;
- б) цепной темп роста;
- в) базисный темп прироста;
- г) цепной темп прироста;
- д) абсолютное значение 1 % прироста.

8. При исчислении среднегодового темпа роста верной является формула:

а) $\bar{T}_p = \sqrt[m-1]{T_1 T_2 \dots T_m}$;

б) $\bar{T}_p = \sqrt[m]{T_1 T_2 \dots T_m} 100\%$;

в) $\bar{T}_p = n-1 \sqrt{\frac{y_n}{y_1}} \cdot 100\%$.

9. Средний уровень моментного ряда динамики с равными временными промежутками исчисляется по формуле средней:

- а) арифметической простой; б) арифметической взвешенной;
 в) гармонической простой; г) гармонической взвешенной;
 д) хронологической простой; е) хронологической взвешенной.

10. Среднегодовой темп роста среднедушевого денежного дохода населения за период определяется по средней:

- а) арифметической простой; б) арифметической взвешенной;
 в) гармонической простой; г) гармонической взвешенной;
 д) геометрической.

11. По формуле $T = \frac{y_t - y_1}{y_1} \cdot 100\%$ определяется:

- а) базисный темп роста; б) цепной темп роста;
 в) базисный темп прироста; г) цепной темп прироста;
 д) абсолютное значение 1 % прироста.

Практические задания

1. Производство сельскохозяйственной продукции за 2006–2012 гг. в регионе в расчете на одного жителя характеризуется следующими данными.

Продукция	2006	2007	2008	2009	2010	2011	2012
Молоко, кг	276	283	300	326	279	200	203
Яйцо, шт.	209	263	294	321	270	200	210
Картофель, кг	80	83	76	69	70	53	60

Определите показатели динамики (цепные и базисные) производства сельскохозяйственной продукции. Рассчитайте средние и за период показатели. Сформулируйте выводы.

2. Динамика цепных индексов цен на реализованную сельскохозяйственную и промышленную продукцию для сельского хозяйства представлена в таблице.

Показатель	2006	2007	2008	2009	2010	2011	2012
Сельскохозяйственная продукция	1,6	2,6	2,9	3,0	3,3	1,4	1,1
Промышленная продукция	1,9	2,0	2,7	4,2	3,2	1,6	1,2
Минеральные удобрения	1,7	2,8	3,0	6,5	3,4	1,9	1,2

Проанализируйте динамику индексов цен за период с 2006 по 2012 г.

3. Среднегодовые темпы роста продукции фермерского хозяйства за период 2007–2012 гг. в земледелии составили 102,6 %, а в животноводстве — 105,3 %. Стоимость продукции в 2012 г. (в условных единицах): в земледелии — 7 820, в животноводстве — 8 590.

Определите среднегодовой темп роста продукции в целом за период 2007–2012 гг.

4. Имеются следующие данные о мощности электростанций региона.

Год	Мощность электростанций, млн кВт	Ценные показатели динамики			
		Абсолютный прирост, млн кВт	Коэффициент роста	Темп прироста, %	Абсолютное значение 1 % прироста, млн кВт
2007	22,3	1,3			
2008					
2009				2,12	0,24
2010			1,04		
2011			1,07		
2012		2,3			

Найдите отсутствующие в таблице сведения за 2007–2012 гг., а также определите, в каком периоде (2007–2009 или 2010–2012 гг.) были более высокие абсолютный и относительный приросты мощности электростанций региона.

5. Определите недостающие уровни ряда и средние показатели динамики.

Год	Товарооборот, млн руб.	Базисные показатели динамики			
		Абсолютный прирост	Темп роста, %	Темп прироста, %	Абсолютное значение 1 % прироста
2007					
2008		-2			
2009			110		
2010				-5	
2011		+30			
2012			98		0,18

6. Данные по перевозке грузов железнодорожным транспортом за 2009–2012 гг. представлены в таблице (млн т).

Квартал	2009	2010	2011	2012
I	253	267	345	440
II	257	267	356	442
III	264	258	334	383
IV	251	262	311	368

Проанализируйте, используя показатели динамики, произошедшие изменения в перевозке грузов железнодорожным транспортом.

7. Имеются следующие данные о товарообороте магазина за 2005–2012 гг., млн руб.

Год	2005	2006	2007	2008	2009	2010	2011	2012
Товарооборот	600	630	610	580	650	670	630	690

Представьте графически динамику товарооборота магазина. Определите компонентный состав временного ряда.

8. Данные о реализации молочной продукции в регионе за 2009–2012 гг. представлены в таблице (тыс. руб.).

Квартал	2009	2010	2011	2012
I	54,6	53,2	56,4	66,6
II	80,6	97,5	112,1	126,6
III	78,8	98,7	114,5	132,8
IV	53,5	60,5	74,6	76,7

Представьте графически динамику реализации молочной продукции в регионе. Визуально определите компонентный состав временного ряда.

9. Имеются условные данные об объемах потребления электроэнергии жителями региона за 16 кварталов (млн кВт · ч).

t	y_t	t	y_t	t	y_t	t	y_t
1	6,0	5	7,2	9	8,0	13	9,0
2	4,4	6	4,8	10	5,6	14	6,6
3	5,0	7	6,0	11	6,4	15	7,0
4	9,0	8	10,0	12	11,0	16	10,8

Представьте графически динамику объемов потребления электроэнергии. Сделайте вывод по графику исходных данных и по коррелограмме о наличии во временном ряду основной тенденции и сезонных колебаний.

10. Имеются условные поквартальные данные о прибыли компании за четыре года (2009—2012 гг.), тыс. долл. США.

Квартал	2009	2010	2011	2012
I	72	100	90	64
II	70	92	80	58
III	62	80	68	48
IV	52	60	50	30

Представьте графически динамику прибыли компании. Сделайте вывод о структуре временного ряда, используя автокорреляционную функцию.

11. Ниже приведены значения коэффициентов автокорреляции уровней некоторого временного ряда Y_t за 30 мес.

τ	1	2	3	4	5	6	7
r_τ	0,63	0,38	0,72	0,97	0,55	0,40	0,65

Охарактеризуйте структуру временного ряда, дав его графическое изображение.

3. АНАЛИЗ И МОДЕЛИРОВАНИЕ ОСНОВНОЙ ТЕНДЕНЦИИ ВРЕМЕННОГО РЯДА

- 3.1. Виды тенденций и проверка гипотезы о существовании основной тенденции.
- 3.2. Методы выявления основной тенденции временного ряда.
 - 3.2.1. Механическое сглаживание.
 - 3.2.2. Аналитическое выравнивание временных рядов.
- 3.3. Критерии адекватности моделей тренда.

3.1. ВИДЫ ТЕНДЕНЦИЙ И ПРОВЕРКА ГИПОТЕЗЫ О СУЩЕСТВОВАНИИ ОСНОВНОЙ ТЕНДЕНЦИИ

Во временных рядах, описывающих социально-экономические процессы, могут наблюдаться тенденции трех видов: тенденция среднего уровня, тенденция дисперсии и тенденция автокорреляции.

В статистической литературе под **тенденцией среднего уровня** (основной тенденцией) понимают некоторое общее направление развития явления. Обычно основную тенденцию можно представить графически более или менее гладкой траекторией. Предполагается, что такая траектория, которую можно описать с помощью некоторой математической функции времени, характеризует основную закономерность развития во времени и в некоторой мере свободна от случайных воздействий [32] (в дальнейшем эту функцию времени будем обозначать через \hat{y}_t и называть **трендом**).

Тенденция дисперсии представляет собой тенденцию в отклонениях между уровнями фактическими и расчетными, найденными по математической формуле тренда. Этот вид тенденции легко представить графически. Если в отклонениях от тренда прослеживается закономерность в виде периодических колебаний, то можно предположить, что явление подвержено влиянию осциллятивных факторов [30].

Тенденцией автокорреляции является тенденция изменения связи между отдельными уровнями временного ряда, которую можно проследить с помощью АКФ по коррелограмме.

Часто во временных рядах с незначительной тенденцией развития во времени и существенной колеблемостью уровней не всегда удается (даже при использовании современных программных средств) визуально определить, присутствует ли тренд в данном ряду или нет.

На рис. 3.1 видны периодические сезонные колебания: очевидны повторяющиеся подъем выпуска продукции с сентября до конца каждого года и спад в производстве с начала каждого года. Если в ряду с сильной колеблемостью уровней присутствие основной тенденции прослеживается нечетко, то, прежде чем перейти к выделению и описанию тенденции, нужно выяснить, существует ли она вообще в исследуемом ряду.

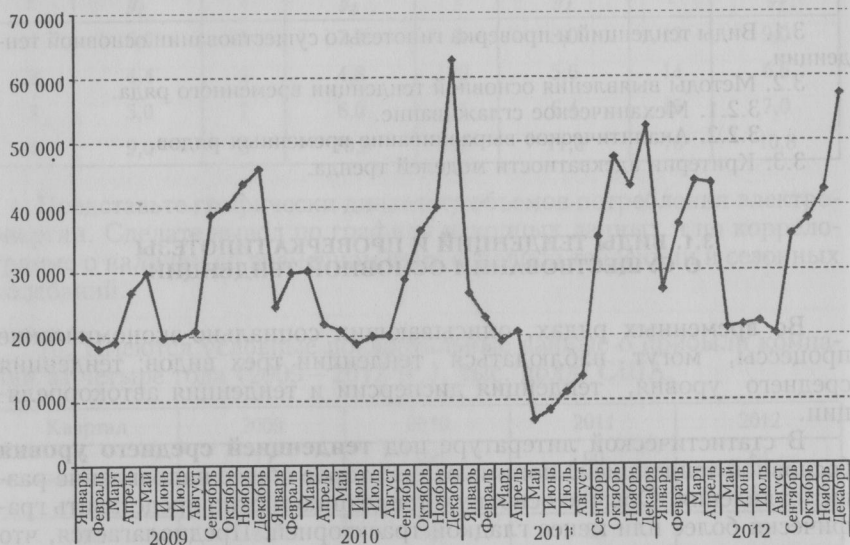


Рис. 3.1. Динамика объема выпуска продукции на предприятии, тыс. руб.

Основные подходы к решению этой задачи основаны на статистической проверке гипотезы о неизменности среднего уровня временного ряда [1; 30; 32]. Для этого используются критерии, отличающиеся друг от друга как мощностью, так и сложностью математического аппарата: метод проверки разностей средних уровней, метод Фостера—Стюарта; непараметрические критерии — «восходящих» и «нисходящих» серий; критерий серий, основанный на медиане выборки; критерий квадратов последовательных разностей (критерий Аббе) и другие. Не стоит забывать об автокорреляционной функции, значение первого коэффициента которой позволяет делать вывод о наличии основной тенденции исходного ряда.

На практике чаще всего используют метод Фостера—Стюарта, который позволяет обнаружить тенденцию как среднего уровня, так и дисперсии, что очень важно для практического анализа. Этот метод можно реализовать в виде следующей последовательности шагов.

1. Каждый уровень ряда, начиная со второго, сравнивается со

всеми предыдущими, при этом определяются значения вспомогательных характеристик u_t и m_t :

$$u_t = \begin{cases} 1, & \text{если } y_t > y_{t-1}, y_{t-2}, \dots, y_1; \\ 0 & \text{в остальных случаях,} \end{cases}$$

$$m_t = \begin{cases} 1, & \text{если } y_t < y_{t-1}, y_{t-2}, \dots, y_1; \\ 0 & \text{в остальных случаях,} \end{cases}$$

т. е. $u_t = 1$, если y_t больше всех предшествующих уровней, а $m_t = 1$, — если меньше.

2. Вычисляются значения $s_t = u_t + m_t$ и $d_t = u_t - m_t$ для всех $t = 2, 3, \dots, n$.

3. Находятся характеристики

$$S = \sum s_t; D = \sum d_t.$$

Показатель S применяется для обнаружения тенденции дисперсии; D — тенденции среднего уровня.

4. После того, как для исследуемого ряда найдены фактические значения S и D , с помощью критерия Стьюдента проверяется гипотеза о том, что величины $S - \mu$ и $D - 0$ случайны. Для этого определяются расчетные значения критерия t_S и t_D :

$$t_S = \frac{S - \mu}{\sigma_S}; t_D = \frac{D - 0}{\sigma_D},$$

где μ — математическое ожидание величины S , определяемое для случайного распределения уровней во времени;

σ_S — среднеквадратическая ошибка величины S ;

σ_D — среднеквадратическая ошибка величины D (значения μ , σ_S , σ_D занесены в табл. 3.1).

Если в таблице нет искомого значений для n (например, $n = 12$), то необходимые μ , σ_D , σ_S приближенно находят с помощью интерполирования, используя табличные данные для ближайших n ($n = 10$ и $n = 15$).

Таблица 3.1

Значения μ и стандартных ошибок σ_S и σ_D для n от 10 до 50 [32]

n	μ	σ_S	σ_D
1	2	3	4
10	3,858	1,288	1,964
15	4,636	1,521	1,964
20	5,195	1,677	2,279
25	5,632	1,791	2,373

1	2	3	4
30	5,990	1,882	2,447
35	6,294	1,956	2,509
40	6,557	2,019	2,561
45	6,790	2,072	2,606
50	6,998	2,121	2,645

5. Расчетные значения t_D и t_S сравниваются с критическим t_α , взятым из таблицы t -распределения Стьюдента для определенного уровня значимости α . Если расчетное значение t_D больше критического t_α , тогда гипотеза об отсутствии тенденции среднего уровня отвергается. Если расчетное значение меньше критического t_α , тогда гипотеза об отсутствии тенденции среднего уровня во временном ряду подтверждается (аналогично проверяется гипотеза об отсутствии тенденции дисперсии).

Критерий «восходящих» и «нисходящих» серий реализуется в виде следующей последовательности шагов.

1. Для временного ряда определяется последовательность γ_t из плюсов и минусов следующим образом:

$$\gamma_t = \begin{cases} +, & \text{если } y_t > y_{t-1} \text{ для } t = 2, 3, \dots, n; \\ -, & \text{если } y_t < y_{t-1} \text{ для } t = 2, 3, \dots, n. \end{cases}$$

Если последующее наблюдение оказалось равным предыдущему, то учитывается только одно из них.

2. В совокупности плюсов и минусов подсчитывается количество серий $v(n)$, где под серией понимается последовательность подряд идущих плюсов или минусов. Один плюс или минус тоже считается серией.

3. Определяется протяженность самой длинной серии $k_{\max}(n)$.

4. Для того, чтобы гипотеза об отсутствии тенденции среднего уровня не была отвергнута, должны выполняться следующие неравенства (при 5% уровне значимости):

$$\begin{cases} v(n) > \left[\frac{1}{3}(2n-1) - 1,96\sqrt{\frac{16n-29}{90}} \right]; \\ k_{\max}(n) \leq k_0(n). \end{cases} \quad (3.1)$$

Квадратные скобки в неравенстве означают целую часть числа. Если хотя бы одно из неравенств не выполняется, то гипотеза об отсутствии тренда отвергается.

Величина $k_0(n)$ зависит от длины исходного ряда и определяется по табл. 3.2.

Значение k_0

Длина ряда	$n \leq 26$	$26 < n \leq 153$	$153 < n \leq 170$
Значение k_0	5	6	7

Рассмотрим теперь критерий серий, основанный на медиане выборки.

1. Исходный ряд преобразуется в ранжированный y'_1, y'_2, \dots, y'_n , где y'_1 — наименьшее значение ряда Y_t .

2. Определяется медиана ранжированного ряда \mathbf{Me} . Для нечетного значения n ($n = 2m + 1$)

$$\mathbf{Me} = y'_{m+1};$$

для четного n ($n = 2m$)

$$\mathbf{Me} = (y'_m + y'_{m+1}) / 2.$$

3. Образуется последовательность из плюсов и минусов по следующему правилу:

$$\delta_t = \begin{cases} +, & \text{если } y_t > \mathbf{Me} \text{ для } t = 1, 2, \dots, n; \\ -, & \text{если } y_t < \mathbf{Me} \text{ для } t = 1, 2, \dots, n. \end{cases}$$

Если значение y_t равно медиане, то оно опускается.

4. Подсчитываются общее число серий $v(n)$ и протяженность самой длинной серии $k_{\max}(n)$ аналогично тому, как это делалось для критерия «восходящих» и «нисходящих» серий.

5. Для того, чтобы гипотеза о случайности временного ряда не была отвергнута, необходимо выполнение следующих неравенств (для 5% уровня значимости):

$$\begin{cases} v(n) > \left[\frac{1}{2}(n+1 - 1,96\sqrt{n-1}) \right]; \\ k_{\max}(n) < [1,43 \ln(n+1)]. \end{cases} \quad (3.2)$$

Если хотя бы одно из неравенств нарушается, то гипотеза об отсутствии тенденции среднего уровня отвергается.

При решении конкретных прикладных задач анализа социально-экономических явлений исследователь может столкнуться с временными рядами показателей, в которых отсутствует тенденция развития, т. е. изменение значений уровней исходного ряда динамики носит стационарный характер. В этой связи, прежде чем приступить к прогнозированию, необходимо всеми известными методами убедить-ся в том, что тенденция в исследуемом временном ряду действительно отсутствует. Только после того, как это установлено и гипотезы о неизменности среднего уровня приняты всеми методами, следует использовать те методы для прогнозирования, которые дают возможность определить развитие явления при отсутствии тенденции.

Пример 3.1. По данным об изменении урожайности зерновых культур, которые представлены в табл. 3.3, проверим с помощью метода Фостера – Стюарта наличие тенденций в исследуемом ряду.

Таблица 3.3

Расчет вспомогательных характеристик

t	y_t	u_t	m_t	d_t	s_t	γ_t
1	2	3	4	5	6	7
1	6,7					
2	7,3	1	0	1	1	+
3	7,6	1	0	1	1	+
4	7,9	1	0	1	1	+
5	6,6	0	1	-1	1	-
6	8,6	1	0	1	1	+
7	7,8	0	0	0	0	-
8	7,7	0	0	0	0	-
9	7,9	0	0	0	0	+
10	8,2	0	0	0	0	+
11	8,4	0	0	0	0	+
12	9,1	1	0	1	1	+
13	8,3	0	0	0	0	-
14	8,7	0	0	0	0	+
15	8,9	0	0	0	0	+
16	9,1	0	0	0	0	+
17	9,5	1	0	1	1	+
18	10,4	1	0	1	1	+
19	10,5	1	0	1	1	+
20	10,2	0	0	0	0	-

1. Каждый уровень ряда y_t сравниваем со всеми предыдущими и определяем значения вспомогательных характеристик u_t и m_t (графы 3, 4).

2. Находим значения $d_t = u_t - m_t$, $s = u_t + m_t$ для всех t начиная со второго (графы 5, 6).

3. Вычисляем характеристики $D = \sum d_t = 7$; $S = \sum s_t = 9$.

4. Значения μ , σ_D , σ_S для $n = 20$ берем из табл. 3.1: $\mu = 5,195$, $\sigma_S = 1,677$, $\sigma_D = 2,279$; тогда $t_S = (9 - 5,195) : 1,677 = 2,269$ и $t_D = 7 : 2,279 = 3,07$.

Значение t_α для $\alpha = 0,05$ берем из таблицы t -распределения Стьюдента: $t_\alpha = 2,093$.

Так как расчетное значение критерия t_D больше критического t_{α} , то с вероятностью 95 % можно утверждать, что гипотеза об отсутствии тенденции среднего уровня отвергается, т. е. в исследуемом временном ряду существует основная тенденция развития. Так как расчетное значение критерия t_S также больше табличного t_{α} , то можно утверждать, что гипотеза об отсутствии тенденции дисперсии тоже отвергается. Это свидетельствует о том, что в исследуемом ряду могут присутствовать периодические колебания [30].

Пример 3.2. По данным предыдущего примера рассмотрим применение критерия «восходящих» и «нисходящих» серий.

1. В графе γ_t ставим знак «+», если последующее значение больше предыдущего, и знак «-» — если меньше (графа 7).

2. Определяем число серий: $v(20) = 8$.

3. Протяженность самой длинной серии $k_{\max}(20) = 6$.

4. Табличное значение $k_0(20) = 5$.

5. В соответствии с формулой (3.1) вычисляем систему неравенств:

$$\begin{cases} 8 \neq 9, \\ 6 \neq 5. \end{cases}$$

Оба неравенства не выполняются, следовательно, гипотеза об отсутствии тенденции среднего уровня в исследуемом временном ряду отвергается (что подтверждает вывод, сделанный в предыдущем примере).

3.2. МЕТОДЫ ВЫЯВЛЕНИЯ ОСНОВНОЙ ТЕНДЕНЦИИ ВРЕМЕННОГО РЯДА

После того, как установлено (графически или с применением вышеописанных критериев), что временной ряд содержит основную тенденцию развития, необходимо определить и описать ее поведение. Для этого применяются следующие методы, которые условно делятся на две группы.

1. Методы механического выравнивания (сглаживания) основаны на том, что фактические уровни ряда заменяются расчетными, обладающими меньшей колеблемостью, чем исходные данные. Механическое сглаживание ряда рассматривается как средство, помогающее определить, как выглядит основная тенденция развития, но оно не позволяет получить ее описание с помощью какого-либо аналитического уравнения.

2. Методы аналитического выравнивания основаны на том, что приблизительно известен вид тренда данного ряда. Например, рассматривая график временного ряда (см. рис. 2.4), исследователь может предположить, что тенденция описывается линейной функцией времени t , т. е. $y_t = a_0 + a_1 t$, где a_0, a_1 — неизвестные параметры, которые нужно определить (оценить).

3.2.1. Механическое сглаживание

В основе методов сглаживания трендовой составляющей ряда лежит следующая идея: если «индивидуальный» разброс значений временного ряда Y_t около среднего значения a характеризуется дисперсией σ^2 , то разброс среднего из m членов временного ряда $(y_1 + y_2 + \dots + y_m)/m$ около того же значения a будет характеризоваться дисперсией, равной σ^2/m [1]. А уменьшение меры случайного разброса (дисперсии) и означает как раз сглаживание соответствующей траектории.

Распространенный прием выявления основной тенденции развития — сглаживание временного ряда с помощью простых и взвешенных скользящих средних. Они позволяют сгладить периодические и случайные колебания и тем самым выявить имеющую тенденцию в развитии.

Алгоритм сглаживания по простой скользящей средней можно представить в виде следующих этапов.

1. Для временного ряда Y_t определяют интервал сглаживания m , который включает m последовательных уровней ряда ($m \leq \frac{n}{3}$). Выбор интервала сглаживания зависит от целей исследования. При этом надо иметь в виду, что чем шире интервал сглаживания, тем в большей степени погашаются колебания и тем плавнее получается линия тренда.

2. Рассчитываются средние арифметические из уровней ряда, входящих в интервал сглаживания.

3. Вычислив значение средней для первых m уровней, переходят к расчету средних для уровней y_2, y_3, \dots, y_{m+1} , затем для y_3, y_4, \dots, y_{m+2} и т. д. Таким образом, интервал сглаживания как бы скользит по временному ряду.

4. Заменяют фактические значения ряда, стоящие в центре интервала сглаживания, на соответствующие средние значения.

Если m — нечетное число ($m = 2p + 1$), то сглаженное значение относится к середине выбранного интервала и вычисляется по формуле

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

где \bar{y}_t — значение скользящей средней для момента t ;

y_i — фактическое значение уровня в момент i (здесь i — порядковый номер уровня в интервале сглаживания).

Если $m = 2p$ — четное число, тогда сглаженное значение будет

находиться между двумя датами. Чтобы ликвидировать этот сдвиг, применяется центрирование, которое заключается в нахождении двух скользящих средних и определении средней из двух соседних средних (пример сглаживания ряда методом 5- и 4-членной скользящей средней представлен в табл. 3.4, графы 3–5).

Таблица 3.4

**Сглаживание урожайности зерновых культур в хозяйстве
методом скользящей средней**

Год	y_t	5-летние скользящие средние	Скользящие 4-летние средние нецентри- рованные	Центриро- ванные скользящие средние	Взвешенные 5-летние скользящие средние
1	2	3	4	5	6
1997	9,5	—	—	—	—
1998	13,7	—	12,3	—	—
1999	12,1	12,5	13,2	12,8	12,8
2000	14,0	13,7	13,7	13,5	13,4
2001	13,2	14,1	14,6	14,1	14,1
2002	15,6	14,4	14,6	14,6	14,7
2003	15,4	15,2	15,7	15,1	15,1
2004	14,0	15,6	15,6	15,6	15,4
2005	17,6	14,7	14,5	15,0	15,6
2006	15,4	15,1	15,3	14,9	14,9
2007	10,9	15,3	14,7	15,0	14,3
2008	17,5	15,5	15,5	15,1	15,2
2009	15,0	15,2	16,3	15,8	16,2
2010	18,5	16,0	15,65	15,97	16,3
2011	14,2	—	—	—	—
2012	14,9	—	—	—	—

При работе с рядами квартальной и месячной динамики можно использовать следующие скользящие средние:

$$\bar{y}_t = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4}; \quad (3.3)$$

$$\bar{y}_t = \frac{\frac{1}{2}y_{t-6} + y_{t-5} + \dots + y_t + \dots + y_{t+5} + \frac{1}{2}y_{t+6}}{12}.$$

При использовании метода скользящих средних последние и первые p значений теряются, что является существенным недостатком этого метода, так как последние данные обладают наибольшей информативностью. Для того, чтобы восстановить потерянные значения в конце ряда, следует вычислить средний абсолютный прирост на последнем участке сглаживания и путем его последовательного прибавления к последнему сглаженному уровню получить недостающие p значений в конце временного ряда. Аналогичную процедуру можно реализовать для оценивания первых уровней ряда.

Если для рассматриваемого процесса характерно нелинейное развитие или для исследователя желательно сохранить мелкие волны, применение простой скользящей средней нецелесообразно. В этих случаях более надежным является использование взвешенной скользящей средней.

При сглаживании с помощью взвешенной скользящей средней на каждом участке выравнивание осуществляется по полиномам невысоких порядков (по параболам 2-го или 3-го порядка). Выравнивание по полиному 1-й степени соответствует сглаживанию по простой скользящей средней. Взвешенная средняя приписывает каждому уровню вес, зависящий от удаления данного уровня от стоящего в середине интервала сглаживания (простая скользящая средняя учитывает все уровни, входящие в интервал сглаживания, с равными весами). Выравнивание с помощью взвешенной скользящей средней осуществляется следующим образом.

Для каждого участка сглаживания подбирается полином вида

$$y_t = a_0 + a_1 i + a_2 i^2 + \dots,$$

параметры которого оцениваются по методу наименьших квадратов (здесь i — порядковый номер уровня в пределах интервала сглаживания). Если отсчет времени в пределах участка сглаживания производить от его середины, т. е. $i = -p, \dots, -2, -1, 0, 1, 2, \dots, p$, то сглаженное значение уровня будет равно параметру a_0 выбранного полинома. Весовые коэффициенты при уровнях ряда, входящих в интервал сглаживания, будут одинаковыми для всех участков, причем при сглаживании по полиному k -й нечетной степени будут такими же, как по полиному $(k-1)$ -й степени.

Расчет взвешенных скользящих средних осуществляется по следующим формулам (полиномы имеют 2-ю или 3-ю степень, а m — нечетное число):

$$m = 5 \quad \bar{y}_t = \frac{1}{35} (-3y_{t-2} + 12y_{t-1} + 17y_t + 12y_{t+1} - 3y_{t+2});$$

$$m = 7 \quad \bar{y}_t = \frac{1}{21} (-2y_{t-3} + 3y_{t-2} + 6y_{t-1} + 7y_t + 6y_{t+1} + 3y_{t+2} - 2y_{t+3});$$

$$m = 9 \bar{y}_t = \frac{1}{231} (-21y_{t-4} + 14y_{t-3} + 39y_{t-2} + 54y_{t-1} + 59y_t + 54y_{t+1} + 39y_{t+2} + 14y_{t+3} - 21y_{t+4}).$$

Как видно из приведенных формул, веса симметричны относительно центрального уровня y_t и их сумма с учетом вынесенного за скобки общего множителя равна единице. Наличие в системе весов как положительных, так и отрицательных величин дает сглаженной кривой возможность сохранять различные изгибы.

3.2.2. Аналитическое выравнивание временных рядов

Для описания основной тенденции временного ряда, как правило, используется аналитическое выравнивание с помощью тех или иных функций времени (кривых роста).

Аналитическое выравнивание состоит из двух основных этапов:

- 1) выбора типа кривой, форма которой соответствует характеру изменения временного ряда;
- 2) определения численных значений параметров кривой (оценки).

Наиболее ответственным этапом при аналитическом выравнивании ряда является выбор формы кривой роста, адекватно описывающей динамику развития процесса. К выбору типа кривой можно подойти различными путями (графический метод, метод конечных разностей и т. д.). Однако какой бы путь ни был принят, он обязательно предполагает знакомство с основными свойствами кривых роста. Поэтому остановимся на характеристике отдельных кривых, наиболее часто применяемых на практике. К ним относятся функции, которые условно можно разбить на три класса. К 1-му относятся используемые для описания процессов с монотонным характером развития и отсутствием пределов роста (многочлены различных порядков, показательные (экспоненциальные) кривые, логарифмические кривые). 2-й класс представлен кривыми, описывающими процесс, который в исследуемом периоде имеет предел роста. Функции, относящиеся к этому классу, называются кривыми насыщения (модифицированная экспонента, гиперболические кривые). Если кривые насыщения имеют точку перегиба, то их относят к 3-му классу и называют S-образными. Эти кривые рассматривают два последовательных процесса: один с ускорением в развитии, другой — с замедлением. Такие процессы часто рассматриваются в демографии, страховании, при определении спроса на новый вид продукции. К S-образным относятся логистические кривые и кривая Гомперца.

Методы выбора кривых роста

Выбор формы кривой является одной из основных проблем, решение которых во многом определяет результаты прогнозирования.

Существует несколько практических подходов, облегчающих процесс выбора формы кривой роста.

Наиболее простой способ — визуальный, опирающийся на графическое изображение уровней временного ряда. Подбирают такую кривую роста, форма которой подходит к описанию тенденции исследуемого явления. Применение современных пакетов статистической обработки данных существенно облегчает для исследователя проведение данного этапа. При этом следует иметь в виду, что использование полиномов высоких порядков для описания основной тенденции и дальнейшего прогнозирования (экстраполирования) может привести к большим ошибкам. Этого можно избежать, если проводить качественный логический анализ изучаемого явления [36]. Он может быть основан на рассчитанных показателях динамики тенденции ряда.

В том случае, если уровни временного ряда могут быть представлены в виде суммы систематической составляющей F_t и случайной компоненты E_t , можно использовать метод последовательных разностей, который помогает при выборе кривой полиномиального вида $y_t = a_0 + a_1t + \dots + a_p t^p$. Данный метод предполагает вычисление первых, вторых и последующих разностей уровней ряда:

$$\Delta y_t = y_t - y_{t-1};$$

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} \text{ и т. д.}$$

Расчет ведется до тех пор, пока разности не будут примерно одинаковыми. Порядок разностей принимается за степень выравнивающего полинома [32].

При относительно стабильных темпах роста или темпах прироста для описания основной тенденции можно использовать экспоненциальную (показательную) кривую.

Тип тенденции можно определить путем сравнения коэффициентов автокорреляции первого порядка, рассчитанных по исходным и преобразованным данным. Если временной ряд имеет линейную тенденцию, то коэффициент автокорреляции первого порядка исходного ряда должен иметь близкое к единице значение. Если временной ряд содержит нелинейную тенденцию в форме экспоненты, то коэффициент автокорреляции первого порядка по логарифмам уровней будет выше, чем коэффициент, рассчитанный по исходным уровням ряда. Чем сильнее выражена нелинейная тенденция, тем в большей степени будут различаться значения коэффициентов автокорреляции.

При использовании простейших компьютерных средств для обработки информации (MS Word и MS Excel) для определения типа тренда выбор наилучшего уравнения можно осуществить путем перебора основных форм тренда, расчета по каждому уравнению коэффициента детерминации R^2 и выбора уравнения тренда с максимальным значением этого коэффициента. При этом необходимо помнить,

для каких целей выделяется наилучшая форма тренда: для экстраполирования или интерполирования [36].

Часто на практике при выборе формы кривой исходят из значения критерия, в качестве которого принимают сумму квадратов отклонений фактических значений уровней от расчетных, полученных выравниванием. Из рассматриваемых кривых выбирается та, которой соответствует минимальное значение критерия. Применение данного критерия дает практические результаты в том случае, если отбор функций проходит в два этапа:

— на первом — выполняется ограничение приемлемых функций исходя из содержательного анализа задачи;

— на втором — рассчитывается критерий и на его основе выбирается наиболее подходящая кривая роста.

Например, возможна ситуация, когда наилучшей функцией по данному критерию признана прямая, однако полученное на ее основе прогнозное значение отрицательно. Если из экономической сути показателя вытекает невозможность отрицательных значений (например, при прогнозировании объема выпускаемой продукции), то следует отказаться от линейной функции, выбрав менее «удачную» по данному критерию, но более соответствующую содержательному смыслу показателя. Более подходящей в этом случае может оказаться экспоненциальная кривая при $b < 1$ (см. рис. 3.2, *в*) или гипербола.

В качестве критерия выбора наилучшей формы кривой роста для описания основной тенденции можно использовать показатели, представленные в табл. 1.2.

Свойства и оценка параметров основных видов кривых роста (трендов)

Среди кривых роста наиболее часто на практике применяют полиномы (многочлены) различных порядков, которые имеют вид

$$y_t = a_0 + a_1t + a_2t^2 + \dots + a_pt^p, \quad (3.4)$$

где a_i — параметры многочлена ($i = 0, 1, 2, \dots, p$);

t — независимая переменная (время);

p — порядок полинома.

Коэффициенты многочленов невысоких степеней имеют конкретную интерпретацию в зависимости от содержания временного ряда. Например, их можно трактовать так: коэффициент a_0 характеризует уровень ряда при $t = 0$, a_1 — скорость роста, a_2 — ускорение (его значение равно половине ускорения значения показателя), a_3 — изменение ускорения. В экономических исследованиях обычно применяют полиномы не выше третьего порядка, так как полиномы более высоких степеней отражают случайные отклонения (что противоречит смыслу основной тенденции).

Полиному первой степени $y_t = a_0 + a_1t$ на графике соответствует прямая линия, и он используется для описания процессов, разви-

вающихся во времени равномерно (рис. 3.2, а). В этом случае первые разности (абсолютные приросты) имеют приблизительно одинаковые значения.

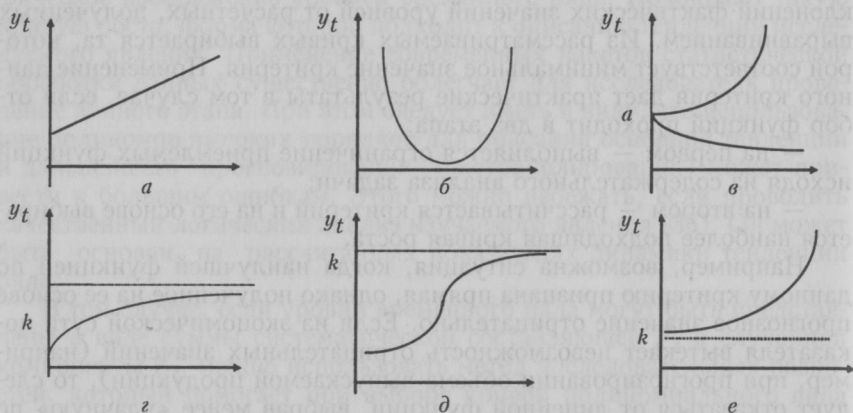


Рис. 3.2. Кривые роста:

а — полином 1-го порядка $y_t = a_0 + a_1 t$, при $a_1 > 0$; б — полином 2-го порядка $y_t = a_0 + a_1 t + a_2 t^2$; в — экспонента $y_t = ab^t$; г — модифицированная экспонента $y_t = k + ab^t$; д — кривая Гомперца, если $\log a < 0, b < 1$;
е — кривая Гомперца, если $\log a > 0, b > 1$

Полином второй степени $y_t = a_0 + a_1 t + a_2 t^2$ можно применять в тех случаях, когда процесс развивается равноускоренно (рис. 3.2, б). На это указывает относительное постоянство вторых разностей. Полином третьей степени $y_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$ используют, если постоянны третьи разности.

Оценки параметров модели (3.3) определяются при помощи метода наименьших квадратов (МНК), суть которого состоит в нахождении таких параметров, при которых сумма квадратов отклонений расчетных значений уровней от фактических была бы минимальной, т. е. оценки параметров находятся в результате минимизации выражения

$$\sum_{t=1}^n (y_t - \hat{y}_t)^2 \rightarrow \min, \quad (3.5)$$

где n — длина временного ряда;

y_t — фактические значения уровней временного ряда;

\hat{y}_t — расчетные значения.

Система нормальных уравнений для определения параметров модели (3.3) выглядит следующим образом:

Система нормальных уравнений для определения параметров A и B строится исходя из минимизации:

$$\sum (\log y_t - \log \hat{y}_t)^2 \rightarrow \min.$$

Следовательно, в нормальных уравнениях вместо фактических уровней выступают их логарифмы:

$$\begin{cases} nA + B\sum t = \sum \log y_t; \\ A\sum t + B\sum t^2 = \sum t \cdot \log y_t. \end{cases} \quad (3.8)$$

Если уровни исходного ряда могут принимать отрицательные значения (например, значения финансового результата от реализации продукции), то необходимо перенести начало отсчета уровней на величину, алгебраически меньшую минимального реального уровня, а по окончании расчета тренда восстановить обычные единицы измерения [3]. Кроме того, следует иметь в виду, что при оценивании МНК оценки коэффициентов a и b оказываются смещенными, так как в расчете участвуют не исходные уровни, а их логарифмы. Смещение будет тем значительнее, чем больше разность между последовательными уровнями временного ряда. Не приводит к смещению в подобных случаях нелинейный метод наименьших квадратов [32].

Более сложным вариантом экспоненциальной кривой является логарифмическая парабола:

$$y_t = ab^t c^{t^2}. \quad (3.9)$$

Прологарифмировав выражение (3.9), получим параболу

$$\log y_t = \log a + t \log b + t^2 \log c,$$

коэффициенты которой можно также оценить с помощью метода наименьших квадратов.

Все рассмотренные типы кривых используются для описания монотонно возрастающих или убывающих процессов без насыщения.

Когда процесс характеризуется насыщением, т. е. существует некоторое предельное значение в исследуемом явлении (потребление каких-либо продуктов), то его следует описывать при помощи кривой, имеющей отличную от нуля асимптоту. Примером такой кривой может служить модифицированная экспонента

$$y_t = k + ab^t. \quad (3.10)$$

Функция (3.10) имеет горизонтальную асимптоту $y = k$.

Если параметр $a < 0$, то асимптота находится выше кривой, если $a > 0$, то ниже. При решении экономических задач чаще всего приходится иметь дело с кривой, у которой $a < 0$, $b < 1$. В этом случае рост уровней происходит с замедлением и стремится к некоторому пределу (рис. 3.2, z).

При решении экономических задач значение асимптоты можно оп-

ределить исходя из свойств исследуемого процесса (например, коэффициент использования оборудования не может превышать 1). Иногда значение асимптоты задается экспертным путем. В этих случаях другие параметры кривой могут быть определены с помощью метода наименьших квадратов после приведения уравнения (3.10) к линейному виду

$$y_t - k = ab^t, \quad (3.11)$$

где k — заданное значение асимптоты.

Прологарифмировав выражение (3.11), получим:

$$\log(y_t - k) = \log a + t \log b,$$

где параметры $\log a$ и $\log b$ оцениваются исходя из системы нормальных уравнений (3.8).

Для описания убывающих процессов, имеющих асимптоту при $y = 0$, можно использовать гиперболические функции вида

$$y_t = a_0 + \frac{a_1}{t}. \quad (3.12)$$

При нахождении параметров гиперболической функции в системе (3.6 а) необходимо t заменить на $1/t$.

Итак, гипербола и модифицированная экспонента хорошо описывают процессы, на развитие которых воздействует некоторый ограничивающий фактор, причем это влияние растет вместе с достигнутым уровнем. Если воздействие ограничивающего фактора начинает проявляться только после определенного момента (точки перегиба), до которого процесс развивался по некоторому экспоненциальному закону, то для выравнивания используют S -образные кривые. Самые известные из них — кривая Гомперца и логистические кривые. Кривая Гомперца названа в честь английского статистика и математика, впервые применившего ее для прогнозирования в страховании. Кривая Гомперца является несимметричной и имеет следующий вид:

$$y_t = ka^{b^t}. \quad (3.13)$$

Если $\log a < 0$, кривая имеет S -образный вид и асимптота $y = k$ проходит выше кривой. Если $\log a > 0$, то асимптота $y = k$ проходит ниже кривой, которая изменяется монотонно: при $b < 1$ убывает; при $b > 1$ возрастает (рис. 3.2, е). Наибольший интерес для решения экономических задач представляет вариант, когда $\log a < 0$, $b < 1$ (рис. 3.2, д).

Если в модифицированной экспоненте вместо y_t ввести обратную величину, т. е. $1/y_t$, то получим логистическую кривую:

$$\frac{1}{y_t} = k + ab^t, \quad (3.14)$$

которую называют кривой Перла — Рида.

С помощью логистической функции и ее различных форм хорошо описывается развитие новых производств. Сначала, когда технические методы недостаточно разработаны, издержки производства высоки и спрос на данный товар еще очень мал, оно развивается медленно. В дальнейшем благодаря совершенствованию методов изготовления, переходу к массовому производству и увеличению емкости рынка для данного товара производство растет быстрее. Затем наступает период насыщения рынка, рост производства все более замедляется и, наконец, почти прекращается, т. е. стабилизируется на определенном уровне. Для коротких временных периодов выявленные закономерности следует обобщать с определенной осторожностью, так как на небольшом промежутке ряд может быть хорошо выравнен экспоненциальной кривой, хотя на самом деле это может оказаться первой половиной (до точки перегиба) логистической кривой.

Для оценивания параметров модифицированной экспоненты, кривой Гомперца и логистической кривой можно применять упрощенные методы: трех сумм, трех точек [3; 32; 36], МНК, если известно значение асимптоты. (Метод трех сумм рассмотрен далее, в п. 4.2.)

Пример 3.3. Для определения тренда временного ряда, представленного на рис. 3.3 и в табл. 3.5, были оценены параметры линейной, параболической и показательной функций с помощью метода наименьших квадратов.

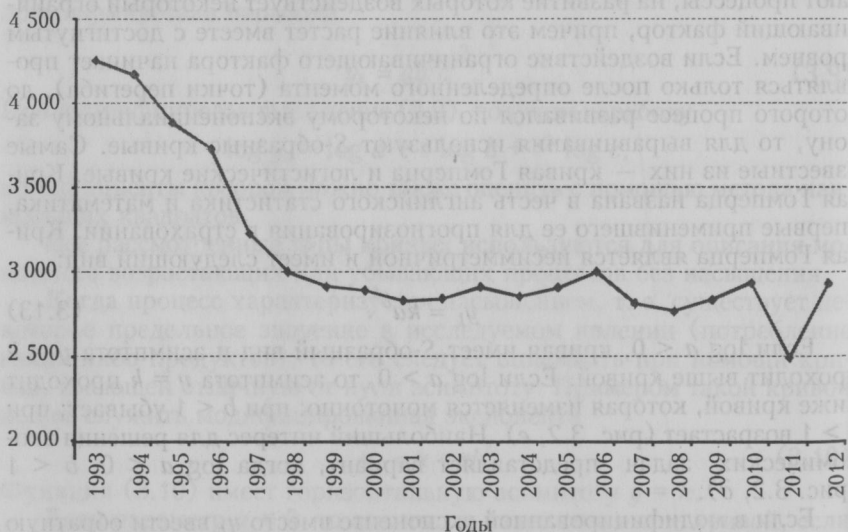


Рис. 3.3. Потребление электроэнергии в регионе, млн кВт · ч

Для расчета параметров функций на основе МНК составляются системы нормальных уравнений. Исходные данные и необходимые расчеты для вычисления коэффициентов представлены в табл. 3.5.

Матрица определения параметров математических функций

Год	y_1	Условные обозначения времени				$t y_t$	$t^2 y_t$	$\lg(y_t)$	$t \lg(y_t)$
		t	t^2	t^3	t^4				
1993	4 252	1	1	1	4 252	4 252	8,36	8,36	
1994	4 167	2	4	8	8 334	16 668	8,33	16,67	
1995	3 884	3	9	27	11 652	34 956	8,26	24,79	
1996	3 730	4	16	64	14 920	59 680	8,22	32,90	
1997	3 227	5	25	125	16 135	80 675	8,08	40,40	
1998	2 999	6	36	216	17 994	107 964	8,01	48,04	
1999	2 912	7	49	343	20 384	142 688	7,98	55,84	
2000	2 889	8	64	512	23 112	184 896	7,97	63,75	
2001	2 829	9	81	729	25 461	229 149	7,95	71,53	
2002	2 840	10	100	1 000	28 400	284 000	7,95	79,52	
2003	2 909	11	121	1 331	31 999	351 989	7,98	87,73	
2004	2 852	12	144	1 728	34 224	410 688	7,96	95,47	
2005	2 903	13	169	2 197	37 739	490 607	7,97	103,66	
2006	2 999	14	196	2 744	41 986	587 804	8,01	112,08	
2007	2 812	15	225	3 375	42 180	632 700	7,94	119,12	
2008	2 768	16	256	4 096	44 288	708 608	7,93	126,81	
2009	2 836	17	289	4 913	48 212	819 604	7,95	135,15	
2010	2 933	18	324	5 832	52 794	950 292	7,98	143,71	
2011	2 492	19	361	6 859	47 348	899 612	7,82	148,60	
2012	2 930	20	400	8 000	58 600	1 172 000	7,98	159,66	
Итого	62 163	210	2 870	44 100	610 014	8 168 832	160,6247	1 673,77	

Система нормальных уравнений для линейной функции исходя из (3.6а) примет вид

$$\begin{cases} 20a_0 + 21a_1 = 62\,163; \\ 210a_0 + 2870a_2 = 610\,014. \end{cases}$$

Решая эту систему, получим, что $a_0 = 3\,782,32$ и $a_1 = -64,21$, т. е. уравнение прямой линии имеет вид

$$\hat{y}_t = 3\,782,32 - 64,21t.$$

Аналогичные расчеты проводятся для параболы и показательной функции. В результате расчетов были получены следующие уравнения:

для параболы

$$\hat{y}_t = 4\,418,68 - 237,76t + 8,26t^2;$$

для показательной функции

$$\hat{y}_t = 3\,763,74 \cdot 0,981^t.$$

На основе полученных уравнений путем подстановки в уравнение тренда значений $t = 1, 2, \dots, n$ определяются теоретические уровни для каждого периода анализируемого ряда динамики. В частности, для прямой они составят:

$$\hat{y}_1 = 3\,782,32 - 64,21 \cdot 1 = 3\,718,1;$$

$$\hat{y}_2 = 3\,782,32 - 64,21 \cdot 2 = 3\,653,9;$$

.....

$$\hat{y}_{20} = 3\,782,32 - 64,21 \cdot 20 = 2\,498,12.$$

Аналогичные расчеты проводятся для параболы и для показательной функции. Результаты расчетов приведены в табл. 3.6. По итоговым данным этой таблицы определяем значения среднеквадратических ошибок:

$$\text{для прямой } S = \sqrt{\frac{1\,817\,900,1}{20}} = 301,5;$$

$$\text{для параболы } S = \sqrt{\frac{618\,833,3}{20}} = 175,9;$$

$$\text{для показательной функции } S = \sqrt{\frac{1\,674\,890}{20}} = 289,4.$$

Из сравнения полученных значений следует, что предпочтение в выборе функции следует отдать параболе.

Матрица определения средней квадратической ошибки

Год	t	y_t	Теоретические уровни по моделям			Квадрат отклонений фактических уровней y_t от теоретических значений \hat{y}_t		
			прямая	показательная функция	парабола	прямолинейная функция	показательная функция	парабола
1993	1	4252	3718,1	3692,1	4189,2	285 034,0	313 543,2	3946,0
1994	2	4167	3653,9	3621,7	3976,2	263 263,9	297 326,5	36 398,0
1995	3	3884	3589,7	3552,7	3779,8	86 612,0	109 735,9	10 861,6
1996	4	3730	3525,5	3485,1	3599,9	41 822,7	59 994,5	16 933,0
1997	5	3227	3461,3	3418,7	3436,5	54 890,5	36 740,4	43 887,8
1998	6	2999	3397,1	3353,6	3289,6	158 468,0	125 711,3	84 473,8
1999	7	2912	3332,9	3289,7	3159,3	177 134,7	142 640,9	61 168,3
2000	8	2889	3268,7	3227,0	3045,5	144 147,0	114 254,5	24 501,4
2001	9	2829	3204,5	3165,5	2948,3	140 970,3	113 263,5	14 224,2
2002	10	2840	3140,3	3105,2	2867,5	90 152,1	70 356,6	757,9
2003	11	2909	3076,0	3046,1	2803,3	27 904,6	18 796,0	11 167,6
2004	12	2852	3011,8	2988,1	2755,6	25 548,8	18 516,5	9 284,3
2005	13	2903	2947,6	2931,2	2724,5	1992,1	792,9	31 863,8
2006	14	2999	2883,4	2875,3	2709,9	13 357,3	15 295,7	83 593,3
2007	15	2812	2819,2	2820,6	2711,8	52,1	73,2	10 043,5
2008	16	2768	2755,0	2766,8	2730,2	168,7	1,4	1 427,3
2009	17	2836	2690,8	2714,1	2765,2	21 081,3	14 853,7	5 014,7
2010	18	2933	2626,6	2662,4	2816,7	93 881,4	73 211,0	13 530,5
2011	19	2492	2562,4	2611,7	2884,7	4955,1	14 330,5	154 215,1
2012	20	2930	2498,2	2562,0	2969,3	1 864 63,6	135 452,4	1 540,9
Итого	210	62 163	62 163,0	61 889,6	62 163,0	1 817 900,1	1 674 890,4	618 833,3

Для определения наилучшей формы тренда можно использовать и значение коэффициента детерминации R^2 , которое можно быстро найти, воспользовавшись функцией «Добавить диаграмму» в MS Excel или MS Word. Для этого необходимо построить график и выделить линию тренда с указанием на диаграмме уравнения тренда и ошибки аппроксимации. Для нашего примера исходный ряд данных, линия тренда, уравнение полинома 2-го порядка и коэффициент детерминации представлены на рис. 3.4.

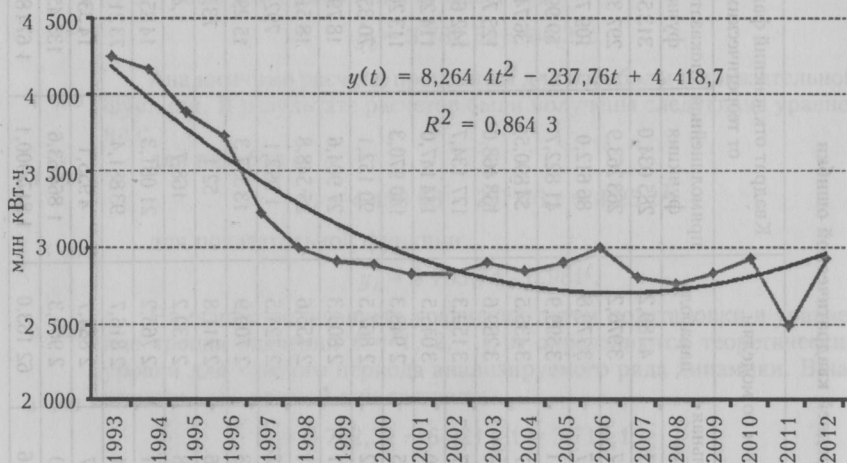


Рис. 3.4. Исходные данные, линия тренда, уравнение полинома 2-го порядка и коэффициент детерминации

Для компактного представления полученных уравнений, их ошибок и/или коэффициентов детерминации можно воспользоваться табличной формой (табл. 3.7).

Таблица 3.7

Основные характеристики кривых роста

Вид кривой роста	Аналитическое уравнение	Средне-квадратическая ошибка	Коэффициент детерминации
Прямая	$\hat{y}_t = 3 782,32 - 64,21t$	301,5	0,60
Полином 2-го порядка	$\hat{y}_t = 4 418,68 - 237,76t + 8,26t^2$	175,9	0,86
Показательная	$\hat{y}_t = 3 763,74 \cdot 0,981^t$	289,4	0,62

3.3. КРИТЕРИИ АДЕКВАТНОСТИ МОДЕЛЕЙ ТRENDA

Проверка адекватности моделей реальному процессу проводится на основе анализа случайной (остаточной) компоненты. В случае если исходные уровни ряда формируются как сумма двух составляющих: $Y_t = F_t + E_t$, а основная тенденция описана уравнением тренда \hat{y}_t , то ряд остатков (остаточная компонента) вычисляется как разности $e_t = y_t - \hat{y}_t$ для всех моментов времени. Принято считать, что модель адекватна описываемому процессу, если значения остаточной компоненты удовлетворяют свойствам случайности, независимости и нормальности распределения.

Проверка гипотезы о случайности распределения остаточной компоненты

При правильном выборе тренда отклонения от него должны носить случайный характер. Это означает, что изменение случайной величины ε_t не связано с изменением времени. По выборке e_t проверяется гипотеза о случайности значений величины ε_t . Для проверки этой гипотезы можно использовать непараметрические критерии — «восходящих» и «нисходящих» серий или основанный на медиане выборки (см. п. 3.1). Также можно воспользоваться АКФ, вычисленной для ряда остатков. Если гипотеза о неизменности среднего уровня ряда остатков e_t (гипотеза о случайности) будет отвергнута, это означает, что уравнение тренда для описания основной тенденции ряда выбрано неудачно.

Проверка гипотезы о независимости распределения остаточной компоненты

Если вид функции выбран неудачно, то последовательные значения ряда остатков не будут обладать свойствами независимости, так как они будут коррелировать между собой. В этом случае говорят об автокорреляции остатков.

Наиболее распространенным приемом обнаружения автокорреляции является метод, предложенный Д. Дарбиным и Д. Уотсоном. Использование критерия Дарбина — Уотсона направлено на выявление автокорреляции остатков 1-го порядка. Фактическое значение этого критерия определяется по формуле

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}. \quad (3.15)$$

Полученное значение сравнивается с критическими границами,

найденными авторами критерия, что и позволяет принять или опровергнуть гипотезу об отсутствии автокорреляции в ряду остатков e_t . Значения критерия Дарбина — Уотсона при 5% уровне значимости приведены в прил. 3.

Расчетное значение критерия Дарбина — Уотсона можно приближенно получить через коэффициент автокорреляции 1-го порядка (2.5) как

$$d \approx 2(1 - r_1). \quad (3.16)$$

Таким образом, если в остатках существует полная положительная автокорреляция и $r_1 = 1$, то $d = 0$; если полная отрицательная — $r_1 = -1$ и соответственно $d = 4$. Если автокорреляция остатков отсутствует, то $r_1 = 0$ и $d = 2$. Следовательно, значение d может находиться в пределах от 0 до 4.

При сравнении расчетной величины d с верхней d_B и нижней d_H границами критерия возможны следующие варианты:

1) если $d < d_H$, то гипотеза о независимости случайных отклонений отвергается (т. е. в ряду остатков есть автокорреляция);

2) если $d > d_B$, то гипотеза о независимости случайных отклонений принимается;

3) если $d_H < d < d_B$, то нет достаточных оснований для принятия решения по данному критерию.

Рассмотренные варианты относятся к случаю, когда в остатках предполагается наличие положительной автокорреляции, т. е. когда $d < 2$.

Когда же расчетное значение $d > 2$, то предполагают наличие в ряду остатков отрицательной автокорреляции. Для проверки отрицательной автокорреляции с критическими значениями d_H и d_B сравнивается не сам коэффициент d , а $4 - d$.

Проверка гипотезы о нормальности распределения остаточной компоненты

В ряде случаев при применении методов прогнозирования явлений необходимо проверить гипотезу о том, что отклонения от тренда или соответствующей модели подчиняются закону нормального распределения. А так как временные ряды, как правило, невелики (не более 50 наблюдений), то проверка распределения на нормальность может быть лишь приближенной и проводить ее можно с помощью исследования показателей асимметрии и эксцесса.

При нормальном распределении эти показатели должны быть равны нулю. Но так как отклонения от тренда представляют собой лишь выборку из некоторой генеральной совокупности, то определяются выборочные характеристики асимметрии и эксцесса и их ошибки:

$$\hat{A} = \frac{\frac{1}{n} \sum e_t^3}{\sqrt{\left(\frac{1}{n} \sum e_t^2\right)^3}}; \hat{\Theta} = \frac{\frac{1}{n} \sum e_t^4}{\left(\frac{1}{n} \sum e_t^2\right)^2} - 3; \quad (3.17)$$

$$\sigma_{\hat{A}} = \sqrt{\frac{\sigma_n(n-1)}{(n-2)(n+1)(n+3)}}; \sigma_{\hat{\Theta}} = \sqrt{\frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}}.$$

В этих формулах \hat{A} — выборочная характеристика асимметрии;
 $\hat{\Theta}$ — выборочная характеристика эксцесса;
 $\sigma_{\hat{A}}$ — среднеквадратическая ошибка выборочной характеристики асимметрии;
 $\sigma_{\hat{\Theta}}$ — среднеквадратическая ошибка выборочной характеристики эксцесса.

Если одновременно выполняются следующие неравенства:

$$\begin{aligned} |\hat{A}| &< 1,5\sigma_{\hat{A}}; \\ \left| \hat{\Theta} + \frac{6}{n+1} \right| &< 1,5\sigma_{\hat{\Theta}}, \end{aligned} \quad (3.18)$$

то гипотеза о нормальном характере распределения случайной компоненты принимается. Если же выполняется хотя бы одно из неравенств

$$\begin{aligned} |\hat{A}| &\geq 2\sigma_{\hat{A}}; \\ \left| \hat{\Theta} + \frac{6}{n+1} \right| &\geq 2\sigma_{\hat{\Theta}}, \end{aligned} \quad (3.19)$$

то гипотеза отвергается.

Пример 3.4. Рассчитать по данным предыдущего примера ряд остатков и проверить модель полинома 2-го порядка на адекватность (т. е. проверить ряд остатков на случайность, независимость и нормальность распределения).

Проверим с помощью критерия «восходящих» и нисходящих» серий случайность уровней ряда остатков e_t . Число серий для ряда длиной $n = 20$: $v(20) = 11$, протяженность самой длинной серии $k_{\max}(20) = 5$, табличное значение $k_0(20) = 5$. Таким образом, подставляя полученные значения в систему (3.11), получим

$$\begin{cases} v(20) > \left[\frac{1}{3}(2 \cdot 20 - 1) - 1,96 \sqrt{\frac{16 \cdot 20 - 29}{90}} \right]; & \begin{cases} 11 > 9; \\ 5 \leq 5. \end{cases} \\ k_{\max}(20) \leq k_0(20); \end{cases}$$

Оба неравенства выполняются, следовательно, нулевая гипотеза об отсутствии тенденции среднего уровня принимается, а ряд остатков e_t можно признать случайным.

Год	t	y_t	\hat{y}_t	$e_t = y_t - \hat{y}_t$	γ_t	$(e_t - e_{t-1})^2$	e_t^2
1993	1	4 252	4 189,2	62,8			3 943,8
1994	2	4 167	3 976,2	190,8	+	16 384,0	36 404,6
1995	3	3 884	3 779,8	104,2	-	7 499,6	10 857,6
1996	4	3 730	3 599,9	130,1	+	670,8	16 926,0
1997	5	3 227	3 436,5	-209,5	-	115 328,2	43 890,3
1998	6	2 999	3 289,6	-290,6	-	6 577,2	84 448,4
1999	7	2 912	3 159,3	-247,3	+	1 874,9	61 157,3
2000	8	2 889	3 045,5	-156,5	+	8 244,6	24 492,3
2001	9	2 829	2 948,3	-119,3	+	1 383,8	14 232,5
2002	10	2 840	2 867,5	-27,5	+	8 427,2	756,3
2003	11	2 909	2 803,3	105,7	+	17 742,2	11 172,5
2004	12	2 852	2 755,6	96,4	-	86,5	9 293,0
2005	13	2 903	2 724,5	178,5	+	6 740,4	31 862,3
2006	14	2 999	2 709,9	289,1	+	12 232,4	83 578,8
2007	15	2 812	2 711,8	100,2	-	35 683,2	10 040,0
2008	16	2 768	2 730,2	37,8	-	3 893,8	1 428,8
2009	17	2 836	2 765,2	70,8	+	1 089,0	5 012,6
2010	18	2 933	2 816,7	116,3	+	2 070,3	13 525,7
2011	19	2 492	2 884,7	-392,7	-	259 081,0	154 213,3
2012	20	2 930	2 969,3	-39,3	+	124 891,6	1 544,5
Итого		62 163	62 163,0	0	X	629 900,7	614 836,8

Расчетное значение критерия Дарбина — Уотсона получено по формуле (3.15):

$$d = \frac{629\,900,7}{614\,836,8} = 1,025.$$

Сравнивая это значение с критическими границами d_H и d_B при 5% уровне значимости, можно сделать вывод, что ряд остатков содержит автокорреляцию, так как расчетное значение критерия меньше нижней границы ($1,025 < 1,20$).

К такому же выводу можно прийти, проанализировав график автокорреляционной функции (рис. 3.5). По коррелограмме видно, что значение коэффициента автокорреляции на первом лаге существенно. (По этому значению можно приближенно получить расчетное значение критерия Дарбина — Уотсона $d = 2 \cdot (1 - 0,487) = 1,026$.)

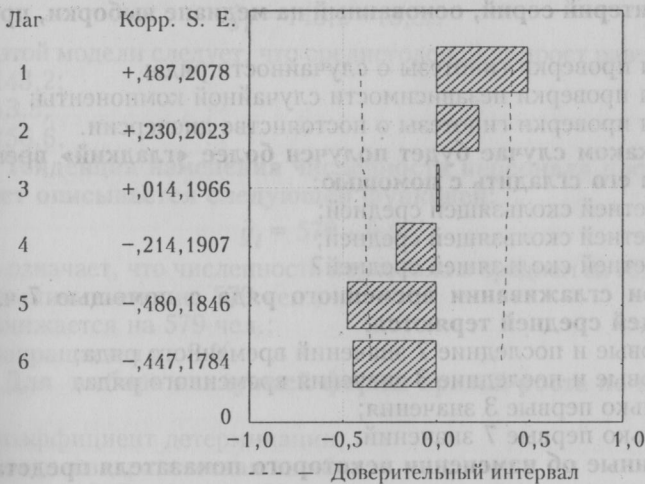


Рис. 3.5. Автокорреляционная функция остатков модели параболического тренда

Проверим нормальность распределения ряда остатков, используя выборочные показатели распределения и эксцесса их среднеквадратических ошибок (3.17).

$$\hat{A} = -0,692; \quad \hat{\hat{\epsilon}} = -0,294;$$

$$\sigma_A = \sqrt{\frac{6(20-1)}{(20+1)(20+3)}} = 0,512;$$

$$\sigma_{\hat{\hat{\epsilon}}} = \sqrt{\frac{24 \cdot 20(20-2)(20-3)}{(20+1)^2(20+3)(20+5)}} = 0,992.$$

Неравенства (3.18) выполняются, следовательно, гипотеза о нормальном распределении случайного компонента не отвергается.

Тесты и задания для самоконтроля

Тест

1. Какой способ неприемлем для выявления основной тенденции во временном ряду:

- метод Фостера — Стюарта;
- вычисление автокорреляционной функции;
- метод наименьших квадратов;
- непараметрические критерии серий?

$$\hat{y}_t = 143,2 + 13,3t.$$

Из этой модели следует, что среднегодовой прирост равен:

- а) 143,2;
- б) 13,3;
- в) 156,6.

11. Тенденция изменения численности персонала предприятия за 10 лет описывается следующей функцией:

$$\hat{y}_t = 579 \cdot 0,97^t.$$

Это означает, что численность персонала в среднем за год:

- а) увеличивается на 579 чел.;
- б) снижается на 579 чел.;
- в) сокращается на 3 %.

12. Для выбора наилучшей формы кривой роста не используются:

- а) коэффициент детерминации;
- б) среднеквадратическая ошибка;
- в) критерий Дарбина — Уотсона;
- г) средняя ошибка аппроксимации.

13. Критерий Дарбина — Уотсона применяется для:

- а) проверки подчинения случайной компоненты нормальному закону распределения;
- б) проверки случайности остаточной компоненты;
- в) проверки независимости случайной компоненты;
- г) для обнаружения периодической компоненты.

14. Добыча и производство газа в регионе характеризуется следующими данными, млн м³.

Год	2006	2007	2008	2009	2010	2011	2012
t	1	2	3	4	5	6	7
y_t	157	169	181	198	212	221	236

В результате аналитического выравнивания временного ряда получена линейная функция

$$\hat{y}_t = 143,14 + 13,3 \cdot t.$$

Теоретическое значение показателя в 2009 г. равно:

- а) 156,44;
- б) 183,04;
- в) 196,34.

15. В результате аналитического выравнивания ряда, представленного в таблице, получено линейное уравнение $\hat{y}_t = 172,1 + 4,42t$.

Год	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
t	-5	-4	-3	-2	-1	0	1	2	3	4	5
y_t	149	145	168	146	177	176	190	186	176	211	170

Теоретическое значение показателя в 2008 г. равно:

а) 172,1;

б) 176,52;

в) 176,0.

16. Для временного ряда длиной $n = 25$ в результате аналитического выравнивания по кривым роста было получено уравнение параболы. Расчетное значение критерия Дарбина — Уотсона $d = 1,5$ показывает, что:

а) в остатках обнаружена автокорреляция;

б) в остатках нет автокорреляции;

в) нет достаточных оснований для принятия решения по данному критерию.

Практические задания

1. Имеются данные об изменении некоторого показателя y_t .

t	1	2	3	4	5	6	7	8	9	10
y_t	7,1	7,3	7,6	7,9	7,0	8,2	7,8	7,7	7,9	8,1
t	11	12	13	14	15	16	17	18	19	20
y_t	8,0	8,6	8,3	8,5	8,6	9,0	9,5	8,5	9,4	9,7

Представьте графически динамику этого показателя. Используя метод Фостера — Стюарта, с вероятностью 95 % определите наличие (отсутствие) тенденции среднего уровня во временном ряду.

2. Изменение курса акций промышленной компании в течение месяца представлено в таблице.

t	y_t	t	y_t	t	y_t	t	y_t
1	509	6	515	11	517	16	510
2	507	7	520	12	524	17	516
3	508	8	519	13	526	18	518
4	509	9	512	14	519	19	524
5	518	10	511	15	514	20	521

Представьте графически динамику курса акций компании. Проверьте гипотезу об отсутствии тенденции среднего уровня (с вероятностью $p = 0,95$) в изменении курса акций, используя:

а) метод Фостера — Стюарта;

б) критерий серий, основанный на медиане выборки.

3. Годовые данные об изменении урожайности зерновых культур представлены в таблице.

t	y_t	t	y_t	t	y_t	t	y_t
1	15,2	6	15,1	11	16,5	16	16,4
2	16,9	7	16,7	12	15,0	17	14,9
3	15,3	8	16,3	13	15,9	18	14,5
4	14,9	9	16,5	14	15,5	19	16,6
5	15,7	10	13,3	15	16,9	20	15,1

Представьте графически динамику урожайности зерновых культур. Проверьте гипотезу об отсутствии тенденции среднего уровня (с вероятностью $p = 0,95$) в изменении урожайности, используя:

- метод Фостера – Стюарта;
- критерий «восходящих» и «нисходящих» серий.

4. Имеются данные о ежесуточной добыче угля бригадой шахтеров, т.

Число месяца	Добыча угля	Число месяца	Добыча угля	Число месяца	Добыча угля
1	151	6	145	11	146
2	146	7	149	12	154
3	152	8	147	13	150
4	151	9	150	14	156
5	154	10	153	15	149

Для выявления основной тенденции временного ряда произведите его сглаживание с помощью:

- укрупнения интервалов;
- пятидневной скользящей средней.

5. Имеются данные по району о розничном товарообороте, млн руб.

Год	Квартал			
	I	II	III	IV
2009	263	269	258	266
2010	279	270	268	276
2011	280	275	272	282
2012	281	283	290	284

Для выявления основной тенденции временного ряда произведите его сглаживание с помощью:

- укрупнения интервалов;
- четырёхквартальной скользящей средней.

6. Производство сахара в регионе характеризуется следующими данными, тыс. т.

t	1	2	3	4	5	6
y_t	14,8	14,5	11,5	17,5	18,3	18,9

Выполните аналитическое выравнивание по полиному 2-го порядка.

7. Имеются месячные данные о темпах роста номинальной заработной платы по отношению к декабрю предыдущего года.

Месяц	Темп роста	Месяц	Темп роста
Январь	82,9	Июнь	121,6
Февраль	87,3	Июль	118,6
Март	99,4	Август	114,1
Апрель	104,8	Сентябрь	123,0
Май	107,2	Октябрь	127,3

Выберите наилучший тип тренда (используя графическое изображение и коэффициенты автокорреляции по уровням ряда и их логарифмам), определите его параметры.

8. Производство некоторых видов продукции в 2007 – 2012 гг. в регионе характеризуется следующими данными, тыс. т.

Вид продукта	2007	2008	2009	2010	2011	2012
Молоко	95	93	94	102	119	115
Колбасные изделия	52	47	47,5	55	70	85

Для выявления основной тенденции развития производства для каждого временного ряда:

- 1) постройте график динамики;
- 2) выбрав наиболее подходящее уравнение основной тенденции развития, произведите аналитическое выравнивание ряда;
- 3) сделайте выводы о значении параметров аналитического уравнения;
- 4) изобразите на графике полученные теоретические уровни (линию тренда).

9. Экспорт, импорт и внешнеторговый оборот страны характеризуются условными данными, млн долл. США.

Год	Экспорт	Импорт	Внешне-торговый оборот	Год	Экспорт	Импорт	Внешне-торговый оборот
1987	44	43	87	2000	204	206	410
1988	47	46	93	2001	209	205	414
1989	51	51	102	2002	236	247	483
1990	56	56	112	2003	257	278	535
1991	62	63	135	2004	281	280	561
1992	67	71	138	2005	328	332	660
1993	72	74	146	2006	366	386	752
1994	79	80	159	2007	405	419	824
1995	95	91	186	2008	431	412	843
1996	117	131	248	2009	450	432	882
1997	129	126	255	2010	498	496	994
1998	146	144	290	2011	549	540	1 089
1999	166	164	330	2012	590	510	1 100

По каждому ряду:

- 1) постройте график динамики;
- 2) проведите расчет параметров уравнений кривых роста разной формы;
- 3) определите точность каждого уравнения при помощи средней ошибки аппроксимации;
- 4) выберите лучшую форму тренда.

10. Добыча и производство газа в регионе характеризуются следующими данными, млн м³.

t	1	2	3	4	5	6	7
y_t	157	169	181	198	212	221	236

В результате выравнивания временного ряда получены следующие уравнения кривых роста:

- линейной $\hat{y}_t = 143,14 + 13,286t$;
- полинома 2-го порядка $\hat{y}_t = 142,29 + 13,86t - 0,072t^2$.

Выберите уравнение кривой роста, наилучшим образом описывающей основную тенденцию временного ряда.

11. Динамика урожайности сельскохозяйственной культуры представлена в таблице.

t	y_t	t	y_t	t	y_t	t	y_t
1	13,5	6	15,1	11	16,1	16	22,8
2	11,6	7	16,9	12	20,4	17	23,1
3	14,7	8	16,8	13	17,8	18	19,6
4	16,2	9	12,9	14	18,3	19	27,0
5	15,2	10	13,8	15	18,9		

Тенденция ряда урожайности сельскохозяйственной культуры описывается линейным уравнением

$$\hat{y}_t = 11,646 + 0,576t.$$

Выделите ряд остатков. Проверьте ряд на случайность, независимость и нормальность распределения. Сделайте выводы.

12. По данным таблицы определите наличие автокорреляции в ряду остатков, если известно, что тенденция описывается линейным уравнением $\hat{y}_t = 172,1 + 4,42t$.

Год	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
t	-5	-4	-3	-2	-1	0	1	2	3	4	5
Y_t	149	145	168	146	177	176	190	186	176	211	170

13. Динамика валового сбора озимой пшеницы представлена в таблице, т.

t	y_t	t	y_t	t	y_t	t	y_t
1	47,3	6	64,3	11	59,7	16	99,7
2	67,4	7	66,5	12	100,5	17	98,8
3	58,1	8	70,8	13	77,4	18	86,0
4	76,6	9	49,7	14	93,4	19	109,8
5	69,1	10	74,4	15	79,9		

Тенденция ряда валового сбора пшеницы описывается линейным уравнением

$$\hat{y}_t = 51,595 + 2,469t.$$

Выделите ряд остатков. Проверьте его на случайность, независимость и нормальность распределения. Сделайте выводы.

14. Программа дала следующие характеристики ряда остатков: длина ряда $n = 20$, коэффициент асимметрии $A = 0,6$, коэффициент эксцесса $\Theta = 0,7$.

Сделайте вывод о нормальности распределения случайной компоненты временного ряда.

4. ПРОГНОЗИРОВАНИЕ ДИНАМИКИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ

- 4.1. Простейшие приемы экстраполяции.
- 4.2. Прогнозирование на основе моделей кривых роста.
- 4.3. Прогнозирование рядов динамики, не имеющих тенденции.

В данном разделе рассмотрено, как простейшие методы и модели используются при прогнозировании экономических показателей. Прежде всего следует отметить, что универсальных методов, пригодных на все случаи жизни, не существует. Выбор метода прогнозирования и его эффективность зависят от многих условий, а именно:

- от поставленных целей и задач исследования;
- от длины анализируемого временного ряда n : чем длиннее ряд исходных данных, тем выше точность прогноза;
- от периода упреждения l , т. е. от того, на сколько временных тактов l строится прогноз. Под одним временным тактом может подразумеваться день, неделя, месяц, квартал, год. Период прогноза (упреждения) не должен превышать величины трети исходной длины ряда $\left[\frac{n}{3} \right]$. Иначе говоря, для прогноза на 5 уровней желательно иметь

временной ряд по длине не менее 15 уровней;

- от наличия или отсутствия в анализируемом временном ряду сезонной (циклической) компоненты или каких-либо резких изменений в поведении тренда.

Выбор метода прогнозирования следует производить с учетом всех особенностей как целей прогноза, так и анализируемого временного ряда.

4.1. ПРОСТЕЙШИЕ ПРИЕМЫ ЭКСТРАПОЛЯЦИИ

Один из наиболее распространенных методов прогнозирования заключается в **экстраполяции**, т. е. в продлении в будущее тенденции, наблюдавшейся в прошлом. Основная цель данного прогноза заключается в том, чтобы показать, к каким результатам можно прийти в будущем, если явление будет развиваться со скоростью, ускорением и т. д., аналогичными для прошлого периода.

Теоретической основой распространения тенденции на будущее

выступает свойство социально-экономических процессов, называемое инерционностью, которое проявляется в сохранении во времени тенденций, закономерностей, взаимосвязей с другими явлениями [3; 30]. Широкое практическое применение методов экстраполяции объясняется их простотой, сравнительно небольшим объемом информации и четкостью механизма реализации лежащих в их основе предпосылок.

В зависимости от того, какие исходные данные положены в основу прогноза, выделяют следующие элементарные методы экстраполяции.

1. Экстраполяция на основе среднего уровня. При предположении о том, что средний уровень не имеет тенденции к изменению или оно незначительно, можно принять, что прогнозируемый уровень равен среднему значению уровней в прошлом, т. е.

$$y_{n+l}^* = \bar{y},$$

где y_{n+l}^* — прогнозируемое значение уровня в точке $n + l$;

l — период упреждения;

\bar{y} — средний уровень ряда Y_t .

Однако, рассматривая временной ряд как выборку из некоторой генеральной совокупности, сложно предположить, что прогнозная точечная оценка полностью совпадает с эмпирическими значениями признака. В этом случае целесообразно определить доверительный интервал путем построения интервального прогноза по выражению вида

$$\bar{y} \pm t_\alpha s_y \sqrt{1 + \frac{1}{n}},$$

где t_α — табличное значение t -статистики Стьюдента с $n - 1$ степенями свободы и уровнем значимости α ;

s_y — среднее квадратическое отклонение для выборки,

$$s_y = \sqrt{\frac{\sum_{t=1, n} (y_t - \bar{y})^2}{n - 1}}.$$

Недостаток этой прогностической оценки заключается в том, что доверительный интервал не связан с периодом упреждения.

Если рассматриваемая совокупность является стационарной, то для оценки средней нужно брать возможно большее число уровней. Если же ряд имеет некоторую тенденцию к изменению, то чем он продолжительнее, тем меньше оснований применения среднего уровня для прогностической оценки.

2. Экстраполяция на основе среднего абсолютного прироста может быть выполнена в том случае, если есть уверенность в равно-

мерном изменении уровней (на равномерный характер развития указывают примерно одинаковые значения цепных абсолютных приростов). Кроме того, необходимо выполнение следующего условия:

$$s_{\text{ост}}^2 \leq \rho^2,$$

где $s_{\text{ост}}^2$ — остаточная дисперсия, которая вычисляется по формуле

$$s_{\text{ост}}^2 = \frac{\sum (y_t - \bar{y}_{\Delta})^2}{n}$$

(\bar{y}_{Δ} — теоретические значения ряда, выравненные методом среднего абсолютного прироста);

$$\rho^2 = \frac{1}{2} \frac{\sum \Delta_t^2}{n}, \text{ где } \Delta_t \text{ — цепные абсолютные приросты.}$$

В этом случае прогнозирование осуществляется по следующей формуле:

$$y_{n+l}^* = y_n + l\bar{\Delta}y, \quad (4.1)$$

где y_{n+l}^* — прогнозная оценка значения уровня в точке $n + l$;

y_n — фактическое значение в последней n -й точке ряда;

l — период упреждения;

Δ — значение среднего абсолютного прироста.

Пример 4.1. Обоснуем правомерность использования среднего абсолютного прироста для прогнозирования и рассчитаем прогноз процентной ставки банка на IV квартал 2012 г. (табл. 4.1).

Таблица 4.1

Процентная ставка банка

Квар- тал	I (2011)	II (2011)	III (2011)	IV (2011)	I (2012)	II (2012)	III (2012)
y	17,0	16,5	15,9	15,5	14,9	14,5	13,8
Δy	—	-0,5	-0,6	-0,4	-0,6	-0,4	-0,7

Легко заметить, что цепные абсолютные приросты примерно одинаковы, что свидетельствует о близости процесса к линейному. Поэтому представляется правомерной возможность оценить прогнозное значение с помощью среднего абсолютного прироста

$$\bar{\Delta}y = -0,53.$$

Кроме того, значение остаточной дисперсии $s_{\text{ост}}^2 = 0,0058$ намного меньше, чем $\rho^2 = 0,1483$.

Следовательно исходя из формулы (4.1) прогноз на IV квартал 2012 г. составит:

$$y_{n+1}^* = 13,8 - 0,53 \cdot 1 = 13,27\%.$$

3. Экстраполяция на основе среднего темпа (коэффициента) роста применяется для прогнозирования процессов, описание динамики которых соответствует представлению в виде показательной или экспоненциальной кривой (на это указывают примерно одинаковые значения цепных темпов роста или прироста). В этом случае прогнозное значение на l шагов вперед определяется по следующей формуле:

$$y_{n+l}^* = y_n \bar{K}^l, \quad (4.2)$$

где \bar{K} — средний коэффициент роста, рассчитанный для ряда Y_t .

Сумма теоретических значений, полученных в результате выравнивания по среднему темпу роста, должна совпадать с суммой эмпирических значений исходного временного ряда. Несовпадение может быть вызвано следующими причинами:

- исходному временному ряду свойственна другая закономерность, а не экспоненциальная;
- существенное и значимое влияние на изучаемое социально-экономическое явление оказывают случайные факторы.

Пример 4.2. Рассчитаем прогноз процентной ставки банка в IV квартале 2012 г., если известно, что изменение ее ежеквартальной динамики происходило примерно с постоянным темпом роста (с I квартала 2011 г. по III квартал 2012 г.), который был равен 92,7 %. В III квартале 2012 г. ставка составляла 11 %.

Прогноз процентной ставки, определяемый по формуле (4.2), в IV квартале 2012 г. равен:

$$y_{t+1}^* = 11 \cdot 0,927 = 10,2\%.$$

Рассмотренные методы прогнозирования являются простейшими, и поэтому прогнозы, полученные на их основе, оказываются приближенными и не всегда надежны при увеличении периода упреждения. Как правило, эти методы используются только при краткосрочном прогнозировании. Применение их в средне- и долгосрочном прогнозировании нецелесообразно, так как они не учитывают колеблемость, изменяемость уровней временного ряда, а в основе построения моделей прогноза и получения прогнозных оценок на всем периоде упреждения лежит принцип равномерного увеличения или уменьшения (в зависимости от знака абсолютного прироста или допустимых границ темпа роста) исследуемого явления.

4.2. ПРОГНОЗИРОВАНИЕ НА ОСНОВЕ МОДЕЛЕЙ КРИВЫХ РОСТА

При прогнозировании социально-экономических показателей с помощью кривых роста необходимо выполнить следующие действия:

1) выбрать одну или несколько кривых, форма которых соответствует характеру изменения выбранного ряда (см. п. 3.2.2);

2) оценить их параметры (см. п. 3.2.2);

3) проверить адекватность выбранных кривых исследуемому процессу, оценить точность моделей и произвести окончательный выбор кривой роста;

4) рассчитать точечный и интервальный прогнозы.

Проверка адекватности выбранных моделей реальному процессу строится на анализе случайной компоненты. Принято считать, что модель тренда адекватна описываемому процессу, если значения остаточной компоненты ε_t ($\varepsilon_t = y_t - \hat{y}_t$) удовлетворяют свойствам случайности, независимости, нормальности распределения (см. п. 3.3). Если случайная компонента удовлетворяет этим свойствам, то выбранную кривую роста можно использовать для определения прогноза. В то же время применение кривых роста для прогнозирования должно базироваться на предположении о неизменности, сохранении основной тенденции как на всем периоде наблюдений, так и в прогнозируемом периоде [32]. Если есть основания принять эти допущения, то прогнозные значения можно найти путем подстановки в уравнение кривой значений времени t , соответствующих периоду упреждения ($t = n + 1, \dots, n + l$). Полученный таким образом прогноз называют точечным. Но, как правило, при прогнозировании социально-экономических показателей фактическое попадание в точку маловероятно, и поэтому прогноз должен быть определен в виде интервала значений.

Несовпадение фактических данных с точечным прогнозом, полученным путем экстраполяции, может быть вызвано: 1) ошибочностью выбора формы линии тренда; 2) погрешностью оценивания параметров кривой роста; 3) погрешностью, связанной с отклонениями отдельных наблюдений от тренда.

Погрешности, связанные со вторым и третьим ее источниками, могут быть отражены в виде доверительного интервала прогноза, который в общем виде определяется как

$$y_{n+l}^* \pm t_{\alpha} S_p, \quad (4.3)$$

где y_{n+l}^* — точечный прогноз на период $t = n + l$;

t_{α} — значение t -статистики Стьюдента;

S_p — средняя квадратическая ошибка прогноза.

Принимая во внимание, что параметры a_i являются выборочными оценками, доверительный интервал можно представить в виде [32]

$$y_{n+l}^* \pm t_{\alpha} S_y \sqrt{\frac{n+1}{n} + \frac{(t_l - \bar{t})^2}{\sum (t - \bar{t})^2}}, \text{ или } y_{n+l}^* \pm t_{\alpha} S_y \sqrt{\frac{n+1}{n} + \frac{3(n+2l-1)^2}{n(n^2-1)}}, \quad (4.4)$$

где S_y — среднее квадратическое отклонение фактических данных от расчетных,

$$S_y = \sqrt{\frac{\sum (y_t - \hat{y}_t)^2}{n - k}}$$

(k — число оцениваемых параметров выравнивающей кривой);

t_l — время упреждения, для которого делается экстраполяция,

$t_l = n + l$;

t — порядковый номер уровня ряда;

\bar{t} — порядковый номер уровня, стоящего в середине ряда,

$$\bar{t} = \frac{n + 1}{2}.$$

Обозначив корень в выражении (4.4) через K , можно определить доверительный интервал следующим образом:

$$y_{n+l}^* \pm t_\alpha S_y K, \text{ или } y_{n+l}^* \pm S_y K^*, \text{ где } K^* = t_\alpha K.$$

В прил. 4 приведены значения K^* в зависимости от длины n и периода упреждения l для прямой и параболы. Очевидно, что при увеличении длины ряда значения K^* уменьшаются, с ростом периода упреждения l — увеличиваются. При этом чем больше длина ряда, тем меньше влияние оказывает период упреждения l .

Пример 4.3. Прогноз показателя следует строить по модели кривой роста, если известно, что она адекватно описывает реальный процесс. Поэтому, прежде чем найти прогнозные значения по данной модели, следует выяснить случайный характер отклонений от тренда. Для проверки этого предположения можно использовать критерий «восходящих» и «нисходящих» серий (п. 3.1). Если остатки e_t признаются случайными и для ряда остатков также выполняются условия независимости и нормальности, то при построении точечных и интервальных прогнозов можно использовать выбранное уравнение кривой роста.

Предположим, что для ряда длиной $n = 24$ было найдено уравнение параболы, которое лучше всего подходит для описания основной тенденции ряда:

$$\hat{y}_t = 161,82 - 6,72t + 0,138t^2.$$

Проверка ряда остатков показала, что они являются случайными, независимыми и нормально распределенными. Таким образом, модель полинома 2-го порядка можно признать адекватно описывающей динамику показателя, и ее можно использовать для построения прогнозных значений. Точечный прогноз на следующий период, т. е. для $t = 25$

$$\hat{y}_{25} = 161,82 - 6,72 \cdot 25 + 0,138 \cdot 25^2 = 80,07.$$

На практике результат экстраполяции прогнозируемых уровней экономических явлений обычно выполняется не точечным, а интер-

вальным прогнозом. Для этого определяется доверительный интервал прогноза (4.4).

В нашем примере при 5% уровне значимости и числе степеней свободы $\nu = 24 - 3$ табличное значение критерия Стьюдента составит: $t_{0,05;21} = 2,08$; среднеквадратическое отклонение $S_y = 8,935$;

$$(t_l - \bar{t})^2 = \left(25 - \frac{300}{24}\right)^2 = 156,25; \quad \Sigma(t - \bar{t})^2 = 1\,150.$$

Тогда нижняя граница доверительного интервала прогноза

$$80,07 - 2,08 \cdot 8,935 \cdot \sqrt{\frac{25}{24} + \frac{156,25}{1\,150}} = 59,9,$$

а верхняя —

$$80,07 + 2,08 \cdot 8,935 \cdot \sqrt{\frac{25}{24} + \frac{156,25}{1\,150}} = 100,24.$$

Таким образом, с вероятностью 95 % можно утверждать, что фактическое значение показателя на следующий год будет находиться в интервале

$$59,9 \leq y_{25} \leq 100,24.$$

Аналогичные расчеты выполняются и для последующих периодов упреждения.

Наиболее существенным вопросом прогнозирования по кривым роста является проблема точности прогноза, которая весьма условна в силу следующих причин.

1. Выбранная для прогнозирования функция дает лишь приближенную оценку тенденции, так как она не является единственно возможной.

2. Статистическое прогнозирование осуществляется на основе ограниченного объема информации, что в свою очередь сказывается на величине доверительных интервалов прогноза.

3. В исходном временном ряду всегда присутствует компонента, характеризующая влияние случайных факторов. Это приводит к тому, что любой прогноз осуществляется лишь с определенной долей вероятности.

Применение трендовых моделей прогнозирования социально-экономических явлений имеет большую значимость, и, несмотря на определенную простоту их реализации, они часто используются для прогнозирования сложных процессов. Если выбранная модель тренда достаточно правильно отражает основную тенденцию развития, то полученные на ее основе прогнозы практически всегда надежны.

Следует учитывать, что экстраполяция по тренду основывается только на анализе основной тенденции развития одномерных временных рядов социально-экономических явлений и процессов. Однако прогноз по аналитическому выражению тренда имеет один существен-

ный недостаток, который может приводить к большим ошибкам. Его суть заключается в том, что прогнозируется только детерминированная составляющая временного ряда F_t и не учитывается случайная компонента. Чтобы избежать этой ошибки и сделать прогноз более точным, надо обязательно проверять трендовую модель на адекватность исследуемому процессу (отклонения от тренда ϵ_t не должны иметь закономерных изменений во времени, т. е. в ряду остатков не должно быть автокорреляции). Если в отклонениях от тренда обнаружена автокорреляция, то их можно описать при помощи модели авторегрессии (см. п. 7.4), а затем делать прогноз. Общий прогноз показателя может быть осуществлен по обоим составляющим ряда: по тренду — путем простой экстраполяции, а по отклонениям — по модели авторегрессии. Сумма двух прогнозов даст общий (суммарный) прогноз для исследуемого ряда. Также при обнаружении автокорреляции в остатках можно для их описания использовать гармоники Фурье или же другие методы прогнозирования (например, адаптивные).

Прогнозирование социально-экономических явлений на основе кривых насыщения стало применяться сравнительно недавно. Впервые эти методы были использованы в начале XX в. Применение кривых насыщения требует соблюдения определенных условий. Во-первых, исходный временной ряд должен быть достаточно длинным (30—40 лет). Во-вторых, он не должен иметь скачков, и тенденция такого ряда должна описываться достаточно плавной кривой. В-третьих, использование этих кривых в прогнозировании социально-экономических явлений может давать достаточно хорошие результаты, если предел насыщения будет определен сравнительно точно. Следует отметить, что кривые насыщения отражают кумулятивные возрастания к определенному заранее максимальному пределу. Особенностью этих кривых роста является то, что абсолютные приращения уменьшаются по мере приближения к пределу, однако процесс роста идет до конца.

Значение использования кривых насыщения в статистическом прогнозировании социально-экономических явлений состоит в том, что они способствуют эмпирически правильному воспроизведению тенденции развития исследуемого явления.

Наиболее распространенными кривыми насыщения, используемыми в статистической практике прогнозирования, являются кривые Гомперца и Перла — Риды. Особенность уравнений этих кривых состоит в том, что их параметры могут быть определены методом наименьших квадратов лишь приближенно. Для расчета данных параметров используется ряд искусственных методов, основанных на разбиении исходного ряда динамики на отдельные группы. Для того чтобы осуществить прогноз на основе кривой роста Гомперца, необходимо выполнить следующее.

1. Прологарифмировать уравнение (3.13):

$$\lg y = \lg k + (\lg a) \cdot b^t,$$

где $\lg k$ — логарифм максимального значения, к которому приближается прогнозный уровень явления;

$\lg a$ — расстояние, которое отделяет в каждый данный момент значение уровня от максимального;

b — параметр, который имеет значение от нуля до единицы.

2. Ряд динамики разбить на три приблизительно равные части.

3. Для каждой части рассчитать суммы логарифмов значений уровней

$$S_1, S_2, S_3.$$

4. Определить первые разности по этим суммам:

$$d_1 = S_2 - S_1;$$

$$d_2 = S_3 - S_2.$$

5. На основании этих расчетов найти параметры уравнения $b, \lg a, \lg k$ следующим образом:

$$b = \sqrt[m]{\frac{d_2}{d_1}}, \quad (4.6)$$

где m — число уровней для каждой части ряда;

$$\lg a = \frac{d_1(b-1)}{(b^m-1)^2}; \quad (4.7)$$

$$\lg k = \frac{1}{m} \left(S_1 - \frac{d_1}{b^m-1} \right). \quad (4.8)$$

Чтобы использовать данную кривую для экстраполяции за пределы исходного ряда динамики, достаточно подставить соответствующее значение t в полученное уравнение.

Пример 4.4. Произведем прогноз показателя на основе кривой Гомперца. В табл. 4.2 и на рис. 4.1 представлены исходные и расчетные значения (нумерация t начинается с нуля).

Ряд разбиваем приблизительно на три части и для каждой из них определяем сумму: $S_1 = 13,215$; $S_2 = 13,589$; $S_3 = 13,705$ и первые разности этих сумм, т. е.

$$d_1 = S_2 - S_1 = 13,589 - 13,215 = 0,374;$$

$$d_2 = S_3 - S_2 = 13,705 - 13,589 = 0,116.$$

Коэффициент $\lg b$ находим по формуле (4.6) как

$$b = \sqrt[5]{\frac{0,116}{0,374}} = \sqrt[5]{0,3102} = 0,7913; b^5 = 0,3102.$$

Коэффициент a определяем по формуле (4.7):

$$\lg a = \frac{0,374 \cdot (0,7913 - 1)}{(0,7913^5 - 1)^2} = \frac{0,374 \cdot (-0,2087)}{(0,3102 - 1)^2} = -0,164.$$

Таблица 4.2

Данные для прогнозирования

Порядковый номер	Исходный уровень ряда	$\lg y_t$	t	Расчетные значения	
				$\lg \hat{y}_t$	\hat{y}_t
1	377,3	2,577	0	2,587 4	380,723 0
2	413,2	2,616	1	2,621 6	418,408
3	446,0	2,649	2	2,648 7	445,348 5
4	473,9	2,676	3	2,670 1	467,842 9
5	497,2	2,697	4	2,687 1	486,519 2
6	514,3	2,711	5	2,700 5	501,764 6
7	529,9	2,724	6	2,711 1	514,162
8	520,8	2,717	7	2,719 5	524,203 6
9	519,7	2,716	8	2,726 2	532,353 4
10	525,7	2,721	9	2,731 5	538,889 8
11	536,6	2,730	10	2,735 6	544,001 4
12	542,6	2,734	11	2,738 9	548,150 7
13	551,4	2,741	12	2,741 5	551,442 2
14	555,0	2,744	13	2,743 6	554,115 1
15	570,4	2,756	14	2,745 2	556,160 3

По формуле (4.8) находим $\lg k$:

$$\lg k = \frac{1}{5} \cdot \left(13,215 - \frac{0,374}{0,3102 - 1} \right) = 2,751 4.$$

Отсюда получаем уравнение

$$\lg y_t = 2,7514 - 0,164 \cdot 0,791 3^t.$$

Расчетные значения находим путем подстановки в полученное уравнение значений t . Так, при $t = 14$ $\lg \hat{y}_t = 2,745 2$ и $\hat{y}_t = 556,160 3$. Прогнозное значение на следующий период $y^* = 559,113 5$ (при $t = 15$).

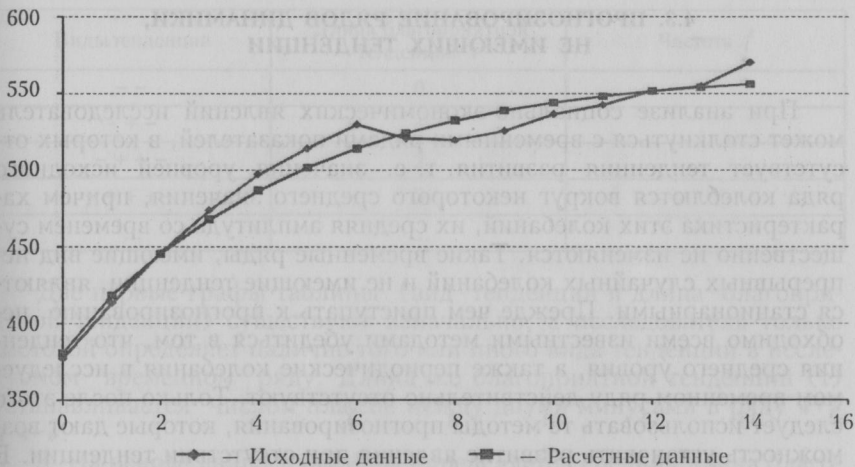


Рис. 4.1. Фактические и расчетные по кривой Гомперца значения условного показателя

Наряду с кривой роста Гомперца, широкое распространение получила кривая роста Перла — Рида, которая в прогнозировании социально-экономических явлений впервые была использована для демографических расчетов американским учеными — биологом Р. Перлом и математиком Л. Ридом. Эта кривая выражает модифицированную геометрическую прогрессию, в которой возрастание затухает по мере приближения к некоторому пределу. Максимальный предел устанавливается прежде всего на основании конкретного изучения исследуемого социально-экономического явления. Так же как и кривая Гомперца, кривая Перла — Рида использует тот же искусственный прием для определения параметров кривой. Однако следует отметить, что, по сравнению с кривой Гомперца, прогнозные данные, полученные в этом случае, имеют некоторую неопределенность.

Кривая роста Перла — Рида описывается уравнением (3.14).

Параметры уравнения находятся следующим образом:

$$b^m = \frac{d_2}{d_1}; \quad b = \frac{d_1(b-1)}{(b^m-1)}; \quad k = \frac{1}{m} \left(S_1 - \frac{d_1}{b^m-1} \right).$$

Как видно, параметры уравнения кривой роста Перла — Рида определяются так же, как и для кривой Гомперца, за исключением того, что в последнем случае не используется прием логарифмирования. Кроме того, нужно иметь в виду, что в зависимости от масштаба данных величина $1/y_t$ умножается на 10 000, 100 000 или 1 000 000.

4.3. ПРОГНОЗИРОВАНИЕ РЯДОВ ДИНАМИКИ, НЕ ИМЕЮЩИХ ТЕНДЕНЦИИ

При анализе социально-экономических явлений исследователь может столкнуться с временными рядами показателей, в которых отсутствует тенденция развития, т. е. значения уровней исходного ряда колеблются вокруг некоторого среднего значения, причем характеристика этих колебаний, их средняя амплитуда со временем существенно не изменяются. Такие временные ряды, имеющие вид непрерывных случайных колебаний и не имеющие тенденции, являются стационарными. Прежде чем приступать к прогнозированию, необходимо всеми известными методами убедиться в том, что тенденция среднего уровня, а также периодические колебания в исследуемом временном ряду действительно отсутствуют. Только после этого следует использовать те методы прогнозирования, которые дают возможность установить развитие явления при отсутствии тенденции. В настоящее время для прогнозирования стационарных процессов, представленных в виде временных рядов, используются модели из класса *ARIMA* (см. п. 7.1–7.4).

Для прогнозирования таких рядов применяются вероятностные статистические методы прогнозного оценивания. Они не позволяют получить точечную количественную характеристику прогнозируемого явления, а дают возможность лишь оценить вероятность того, что его значение на каждый следующий (с отдалением) период упреждения будет больше или меньше значения последнего уровня исходного временного ряда. Эти методы прогнозирования дают менее точные прогнозные оценки и обладают большей степенью неопределенности.

Наибольшее распространение из вероятностных методов прогнозирования получил основанный на использовании закона распределения Пуассона (распределения редких явлений) с плотностью $p = e^{-x}$. Особенность метода заключается в том, что всегда прогнозируется благоприятная тенденция. Этапы реализации данного метода следующие.

1. Осуществляется последовательное сравнение каждого следующего значения уровня исходного временного ряда со значением предыдущего уровня. Если последующий уровень больше предыдущего, то ставится знак «+», меньше — «-». Первый уровень всегда отмечается знаком «-». Знак «+» показывает, сколько периодов исследуемое явление возрастает, и этот временной отрезок принято считать благоприятной тенденцией.

2. Строится таблица, в которой отражаются виды тенденции, длины благоприятной тенденции (τ) и частота повторения благоприятной тенденции (f).

Виды тенденции	Длина благоприятной тенденции τ	Частота f
--	0	
-+-	1	
-++-	2	
-+++	3	
...	...	

Две первые графы таблицы (вид тенденции и длина благоприятной тенденции) существуют изначально, а исследователь только частотой определяет наличие того или иного вида тенденции в исследуемом временном ряду. Длина же благоприятной тенденции (τ) устанавливается числом плюсов между двумя минусами в ряду «+» и «-».

3. На основе данных таблицы определяется средняя длина благоприятной тенденции по формуле средней арифметической взвешенной:

$$\bar{\tau} = \frac{\sum \tau \cdot f}{\sum f}.$$

Средняя длина благоприятной тенденции показывает, сколько времени в среднем она наблюдалась в рассматриваемом ряду. На основе такой средней длины определяется показатель, характеризующий интенсивность прерывания этой благоприятной тенденции:

$$\lambda = \frac{1}{\bar{\tau}}.$$

Данный показатель отражает, сколько в среднем раз за рассматриваемый период времени прерывалась благоприятная тенденция.

4. Вероятность благоприятной тенденции определяется на основе следующей модификации закона распределения Пуассона:

$$p = e^{-\lambda l},$$

где p — вероятность совершения благоприятной тенденции;

l — период упреждения (число лет сохранения благоприятной тенденции).

Пример 4.5. На основе фактических данных за 9 мес. получена следующая таблица.

Виды тенденции	Длина благоприятной тенденции τ	Частота f
--	0	2
-+-	1	1
-++-	2	0
-+++	3	1

На основе данных этой таблицы определяется средняя длина благоприятной тенденции:

$$\bar{\tau} = \frac{0 \cdot 2 + 1 \cdot 1 + 2 \cdot 0 + 3 \cdot 1}{4} = 1.$$

Интенсивность прерывания благоприятной тенденции также равна 1: $\lambda = \frac{1}{\bar{\tau}} = \frac{1}{1} = 1$. Вероятность этой тенденции на следующие два периода представлена в таблице.

Период	l	λ	λl	Вероятность благоприятной тенденции p
10	1	1	1	0,368
11	2	1	2	0,135

Таким образом, с вероятностью 36,8 % можно утверждать, что значение показателя в следующем (десятом) месяце возрастет относительно показателя предыдущего месяца.

Тест и задания для самоконтроля

Тест

1. Продление в будущее тенденции, наблюдавшейся в прошлом, — это есть:

- а) интерполяция; б) экстраполяция; в) прогнозирование.

2. Если значения цепных темпов роста временного ряда примерно одинаковы, то для вычисления прогнозного значения корректно использовать:

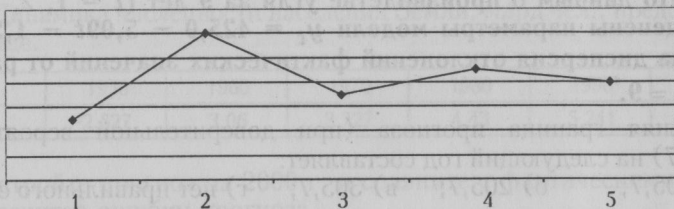
- а) средний темп прироста;
б) средний абсолютный прирост;
в) средний коэффициент роста.

3. Если значения цепных абсолютных приростов временного ряда примерно одинаковы, то для вычисления прогнозного значения можно использовать:

- а) средний темп прироста;
б) средний абсолютный прирост;
в) средний темп роста.

4. На рисунке представлен временной ряд данных за пять лет. Прогноз показателя на следующий год корректнее получить при помощи:

- а) среднего темпа роста;
б) среднего абсолютного прироста;
в) среднего уровня ряда.



5. При исчислении среднегодового коэффициента роста верной является формула:

а) $\overline{K_p} = \sqrt[m]{K_1 K_2 \dots K_m} \cdot 100\%$;

б) $\overline{K_p} = \sqrt[m]{K_1 + K_2 + \dots + K_m}$;

в) $\overline{K_p} = n \sqrt[n]{\frac{y_n}{y_1}}$.

6. При экстраполяции на основе среднего уровня это значение определяется по формуле:

а) средней арифметической;

в) средней геометрической;

д) средней хронологической.

7. В таблице представлены данные об объеме производства (млн руб.) в течение шести кварталов (с I квартала 2011 г. по II квартал 2012 г.).

t	1	2	3	4	5	6
y_t	16,40	15,36	14,31	13,28	12,23	11,18

Прогноз производства в IV квартале 2012 г., полученный с помощью среднего прироста, равен:

а) 10,31; б) 9,44; в) 12,05; г) нет правильного ответа.

8. Ежеквартальная динамика процентной ставки банка представлена в таблице.

t	I квартал 2011 г.	II квартал 2011 г.	III квартал 2011 г.	IV квартал 2011 г.	I квартал 2012 г.	II квартал 2012 г.	III квартал 2012 г.
y_t	17,0	15,8	12,7	11,7	10,8	9,0	8,3

Для приведенных данных средний коэффициент роста равен:

а) 88,3 %; б) 11,7 %; в) 0,883; г) нет правильного ответа.

9. По данным задания № 8 рассчитайте прогноз процентной ставки банка в IV квартале 2012 года:

а) 9,18; б) 7,42; в) 7,33; г) нет правильного ответа.

10. По данным о производстве угля за 9 лет ($t = 1, 2, \dots, 9$) были оценены параметры модели $y_t = 425,0 - 5,09t - 1,59t^2$ и получена дисперсия отклонений фактических значений от расчетных $S_y^2 = 9$.

Нижняя граница прогноза (при доверительной вероятности $p = 0,997$) на следующий год составляет:

- а) 105,7; б) 205,7; в) 305,7; г) нет правильного ответа.

Практические задания

1. По приведенным данным определите первые и вторые разности, покажите правомерность применения для прогнозирования среднего абсолютного прироста, сделайте прогноз на два года вперед.

t	2006	2007	2008	2009	2010	2011	2012
y_t	17,0	16,5	15,9	15,5	14,9	14,5	13,8

2. В таблице представлены данные об объеме производства, млрд руб.

t	1	2	3	4	5	6
y_t	11,18	12,23	13,38	14,31	15,36	16,40

Определите цепные абсолютные приросты и средний прирост. Найдите прогнозируемый объем производства на следующие два года.

3. Динамика процентной ставки банка представлена в таблице.

t	1	2	3	4	5	6	7
y_t	7,3	8,0	8,8	9,7	10,7	11,8	13,0

Определите цепные коэффициенты роста, средний коэффициент роста. Найдите прогнозируемую ставку банка на следующие два периода.

4. Динамика численности занятых, тыс. чел., представлена в таблице.

t	2006	2007	2008	2009	2010	2011	2012
y_t	399,5	398,1	396,6	395,3	393,9	392,2	390,8

Покажите правомерность использования среднего прироста для прогнозирования.

5. Динамика численности населения Земли, млрд чел., представлена в таблице.

t	1950	1960	1970	1980	1990	2000
y_t	2,527	3,06	3,727	4,43	5,241	6,16

Постройте прогноз на 2000 г. и сравните с фактическими данными, рассчитав ошибку прогноза.

6. Имеются следующие данные.

t	2009	2010	2011	2012
y_t	10			14

Определите неизвестные уровни ряда, предполагая их линейное изменение.

7. Ежеквартальная динамика процентной ставки банка за семь кварталов (с I квартала 2011 г. по III квартал 2012 г.) имела примерно постоянный темп роста. Процентная ставка банка в I квартале 2011 г. составляла 8,3 %, в III квартале 2012 г. — 14 %.

Определите прогноз процентной ставки банка в IV квартале 2012 г.

8. Имеются данные о производстве тканей, млн m^2 .

t	2007	2008	2009	2010	2011	2012
y_t	534	537	504	495	485	379

Определите параметры линейного тренда с учетом переноса начала координат в середину ряда. Постройте точечный прогноз на два года вперед. На графике изобразите фактические, теоретические и прогнозные значения ряда.

9. Имеются данные об изменениях некоторого показателя y_t .

t	1	2	3	4	5	6	7	8
y_t	2,3	2,8	3,6	4,8	6,3	8,1	10,3	12,9

Рассчитайте первые и вторые разности. Определите наилучшую форму линии тренда и рассчитайте его параметры. Постройте точечный и интервальный прогнозы на следующие два года, гарантируя результат с вероятностью 0,954. На графике изобразите фактические, теоретические и прогнозные значения.

10. Для временного ряда розничного товарооборота региона, млн руб., длиной $n = 20$ ($t = 1, 2, \dots, 20$) определена линейная модель тренда

$$\hat{y}_t = 10,2 + 1,2t$$

и получена дисперсия отклонений фактических значений от расчетных:

$S_y^2 = 0,25$. Используя эту модель, рассчитайте точечный и интервальный прогнозы на три шага вперед. Оформите прогнозные значения в таблицу.

11. Имеются поквартальные данные о численности занятых на предприятиях машиностроения в регионе, тыс. чел.

t	1	2	3	4	5	6	7	8
y_t	232	220	209	197	187	175	164	170

Определите наилучшую форму линии тренда и рассчитайте его параметры. Дайте прогноз численности занятых на следующие два года. Оформите прогнозные значения в таблицу.

12. По данным, представленным в таблице, оцените параметры логистической функции и определите прогнозные значения на два шага вперед.

t	1	2	3	4	5	6	7	8	9	10
y_t	2 842	3 084	3 105	3 293	3 369	3 682	4 040	4 071	4 543	4 546
t	11	12	13	14	15	16	17	18	19	20
y_t	5 144	5 168	5 413	7 574	7 588	7 608	7 594	8 256	8 256	8 289

5. АДАПТИВНЫЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ

- 5.1. Сущность адаптивных методов.
- 5.2. Метод экспоненциального сглаживания.
- 5.3. Метод экспоненциального сглаживания с трэкинг-сигналом.

5.1. СУЩНОСТЬ АДАПТИВНЫХ МЕТОДОВ

Характерной чертой адаптивных методов прогнозирования является их способность непрерывно учитывать эволюцию динамических характеристик изучаемых процессов, подстраиваться под эту эволюцию, придавая, в частности, тем большие вес и информативную ценность имеющимся наблюдениям, чем ближе они к текущему моменту. Все это делает эффективным их применение для прогнозирования неустойчивых рядов с изменяющейся тенденцией. Отличие этих методов от методов экстраполяции заключается в следующем.

1. Простейшие методы экстраполяции и прогнозирования по кривым роста основаны на равнозначной оценке исходной информации независимо от того, отражает она прошлые или последние тенденции развития социально-экономических явлений или процессов. Для получения достоверных прогнозов наиболее ценны данные последних известных уровней ряда, и поэтому целесообразно оценивать исходную информацию по-разному: наиболее позднюю (в конце исходного временного ряда) учитывать с более высокими весами, чем раннюю (в начале исходного ряда данных).

2. Одним из основных условий применения экстраполяции по кривым роста выступает инерционность развития. В последнее время в стране возрастает подвижность социально-экономических систем. Даже наиболее инерционные макроэкономические характеристики становятся более подвижными. Повышается быстрота реакции на конъюнктуру внешнего и внутреннего рынка, на правительственные решения, на новые социально-экономические условия. В связи с этим для прогнозирования таких процессов оправданно применение адаптивных методов прогнозирования.

3. Оценивание коэффициентов адаптивной модели обычно осуществляется на основе рекуррентного метода, который отличается от метода наименьших квадратов и других тем, что не требует повторения всего объема вычислений при появлении новых данных.

Сущность адаптивных методов заключается в возможности построения корректирующихся моделей, способных учитывать результат

прогноза, сделанного на предыдущем шаге. Быстроту реакции модели на изменение в динамике процесса характеризует так называемый параметр адаптации (сглаживания). Он должен быть выбран таким образом, чтобы обеспечивалось адекватное отображение тенденции при одновременной фильтрации случайных отклонений. В качестве критерия оптимальности при выборе параметра адаптации обычно принимают критерий минимума среднего квадрата ошибок прогнозирования [23; 30].

Адаптивными называются методы прогнозирования, позволяющие строить самокорректирующиеся (самонастраивающиеся) экономико-математические модели, которые способны оперативно реагировать на изменение условий путем учета результата прогноза, сделанного на предыдущем шаге, и учета различной информационной ценности уровней ряда. Благодаря указанным свойствам адаптивные методы особенно удачно используются при оперативном и краткосрочном прогнозировании. Наиболее известными являются три модели данного типа: однопараметрическая Р. Брауна, двухпараметрическая Ч. Хольта и трехпараметрическая Дж. Бокса и Г. Дженкинса, отличающиеся рекуррентными выражениями для пересчета текущих оценок коэффициентов. Наиболее простым из многочисленного класса адаптивных методов является метод Брауна, который основан на экспоненциальном сглаживании, поэтому его называют методом экспоненциального сглаживания.

5.2. МЕТОД ЭКСПОНЕНЦИАЛЬНОГО СГЛАЖИВАНИЯ

Сущность метода экспоненциального сглаживания заключается в том, что временной ряд сглаживается с помощью взвешенной скользящей средней, в которой веса подчиняются экспоненциальному закону, т. е. более ранним наблюдениям придаются меньшие веса, чем более поздним.

Экспоненциальное сглаживание предусматривает постоянное обновление модели за счет наиболее свежих данных. В сглаженном виде новый прогноз (для периода $t + 1$) можно представить как взвешенное среднее последнего наблюдения величины в момент времени t и ее прежнего прогноза на этот же период. Причем вес α присваивается последнему наблюдаемому значению, а вес $(1 - \alpha)$ — прогнозу. Формально выражение прогноза можно записать как

$$y_{t+1}^* = \alpha y_t + (1 - \alpha) y_t^*,$$

где y_{t+1}^* — прогнозируемое значение на период $t + 1$;

α — параметр адаптации (сглаживания) ($0 < \alpha < 1$);

y_t — фактическое значение уровня ряда в момент времени t ;

y_t^* — прежний прогноз на момент времени t .

Экспоненциальное сглаживание есть не что иное, как старый прогноз с некоторым уточнением α , умноженным на ошибку последнего прогноза: $y_{t+1} = y_t + \alpha(y_t - y_t^*)$. Таким образом, экспоненциальное сглаживание — это процедура постоянного пересмотра результатов прогнозирования в свете самых последних событий. Параметр сглаживания α является взвешивающим фактором, его реальное значение определяется тем, в какой мере текущее наблюдение должно влиять на прогнозируемую величину. Рекуррентная формула для экспоненциального сглаживания ряда определяется как

$$S_t = \alpha y_t + (1 - \alpha)S_{t-1}, \quad (5.1)$$

где S_t — значение экспоненциальной средней в момент времени t ;
 S_{t-1} — значение экспоненциальной средней в момент времени $t - 1$.

При расчете экспоненциальной средней в момент времени t всегда требуется ее значение в момент $t - 1$. В качестве начального значения S_0 на практике часто используется среднее арифметическое из всех имеющихся уровней временного ряда или его части.

Р. Браунм было доказано, что математические ожидания экспоненциальной средней и исходного ряда совпадают, а дисперсия экспоненциальной средней меньше, чем временного ряда. При высоком значении α дисперсия экспоненциальной средней незначительно отличается от дисперсии исходного ряда. С уменьшением значения параметра адаптации дисперсия экспоненциальной средней сокращается, т. е. возрастает ее отличие от дисперсии ряда.

Величина параметра сглаживания α является ключом к анализу данных. Если требуется, чтобы спрогнозированные значения были стабильны и случайные отклонения сглаживались, необходимо выбрать малое значение параметра (например, при большом сроке прогнозирования). Большое его значение имеет смысл, если нужна быстрая реакция на изменения, происходящие в динамике явления (в частности, при оперативных конъюнктурных прогнозах).

В отдельных случаях Браун предлагает определять величину α по формуле

$$\alpha = \frac{2}{m+1}, \quad (5.2)$$

где m — число наблюдений, входящих в интервал сглаживания. Для коротких временных рядов можно за m брать длину ряда n [17]. Как отмечалось ранее, за критерий оптимальности при выборе параметра адаптации можно принимать критерий минимума среднего квадрата ошибок прогнозирования. В этом случае при построении прогнозов по методу экспоненциального сглаживания последовательно вычисляются прогнозы с различными значениями параметра α : сначала с наименьшим, близким к 0, и далее через определенный шаг до значения, близкого к 1 (например, 0,1; 0,2; 0,3; ...; 0,9). Для каждого

из них рассчитывается величина среднего квадрата ошибки прогноза. Для окончательного прогнозирования выбирается то значение α , для которого величина ошибки будет минимальной.

По теореме, доказанной Брауном и Майером, если предполагается, что некоторый процесс Y_t может быть описан полиномом степени p , т. е.

$$y_t = a_0 + a_1 t + \frac{1}{2} a_2 t^2 + \dots + \frac{1}{p!} a_p t^p + \varepsilon_t, \quad (5.3)$$

то коэффициенты предсказывающего полинома

$$y_{t+l}^* = \hat{a}_0 + \hat{a}_1 l + \frac{1}{2} \hat{a}_2 l^2 + \dots + \frac{1}{p!} \hat{a}_p l^p \quad (5.4)$$

могут быть определены через экспоненциальные средние соответствующих порядков.

Экспоненциальной средней первого порядка ряда Y_t называется выражение (5.1). Брауном выведена следующая рекуррентная формула для определения экспоненциальной средней k -го порядка в точке t :

$$S_t^k = \alpha S_t^{k-1} + (1-\alpha) S_{t-1}^k. \quad (5.5)$$

На практике для сглаживания обычно используются полиномы не выше 2-го порядка. Формулы, необходимые для расчета экспоненциальных средних и оценок коэффициентов для адаптивных моделей нулевого, первого и второго порядков, приведены в табл. 5.1.

Процедура прогнозирования временных рядов по методу экспоненциального сглаживания с учетом выбора оптимального параметра адаптации состоит из следующих этапов.

1. Определяется вид модели экспоненциального сглаживания. Для этого могут быть использованы такие подходы, как графический анализ, метод разностей, метод аналитического выравнивания и др.

2. Задается начальное значение параметра адаптации α (наименьшее, но большее нуля).

3. Определяются начальные условия для выбранной модели. Чаще всего в качестве оценок a_0, a_1, a_2 берут коэффициенты соответствующих полиномов, полученные методом наименьших квадратов.

4. Рассчитываются значения соответствующих экспоненциальных средних.

5. Находятся оценки коэффициентов.

6. Осуществляется прогноз на одну точку вперед y_{t+1}^* по соответствующей модели и определяется отклонение фактического значения временного ряда от прогнозируемого. Шаги с 4-го по 6-й повторяются для всех $t \leq n$, где n — длина временного ряда.

Основные формулы для прогнозирования с помощью метода экспоненциального сглаживания

Вид модели	Начальные условия	Экспоненциальные средние	Оценка коэффициентов	Модель прогноза	Ошибка прогноза
1. Постоянная $y_t = a_0 + \varepsilon_t$	$S_0^1 = \frac{\sum y_t}{n}$	$S_t^1 = \alpha y_t + (1-\alpha)S_{t-1}^1$	$\hat{a}_0 = S_t^1$	$y_{t+1}^* = \hat{a}_0$	$\sigma_{y_{t+1}}^* = \sigma_{\varepsilon_t} \times \sqrt{\frac{\alpha}{2-\alpha}}$
2. Линейная $y_t = a_0 + a_1 t + \varepsilon_t$	$S_0^1 = a_0 - \frac{1-\alpha}{\alpha} a_1$ $S_0^2 = a_0 - \frac{2(1-\alpha)}{\alpha} a_1$	$S_t^1 = \alpha y_t + (1-\alpha)S_{t-1}^1$ $S_t^2 = \alpha S_t^1 + (1-\alpha)S_{t-1}^2$	$\hat{a}_0 = 2S_t^1 - S_t^2$ $\hat{a}_1 = \frac{\alpha}{1-\alpha} (S_t^1 - S_t^2)$	$y_{t+1}^* = \hat{a}_0 + \hat{a}_1 t$	$\sigma_{y_{t+1}}^* = \sigma_{\varepsilon_t} \sqrt{\frac{\alpha}{(2-\alpha)^3}} \times \sqrt{1+4(1-\alpha)-5(1-\alpha)^2+2\alpha(4-3\alpha)+2\alpha^2} t^2$
3. Квадратичная $y_t = a_0 + a_1 t + \frac{1}{2} a_2 t^2 + \varepsilon_t$	$S_0^1 = a_0 - \frac{1-\alpha}{\alpha} a_1 + \frac{(1-\alpha)(2-\alpha)}{2\alpha^2} a_2$ $S_0^2 = a_0 - \frac{2(1-\alpha)}{\alpha} a_1 + \frac{(1-\alpha)(3-2\alpha)}{\alpha^2} a_2$ $S_0^3 = a_0 - \frac{3(1-\alpha)}{\alpha} a_1 + \frac{3(1-\alpha)(4-3\alpha)}{2\alpha^2} a_2$	$S_t^1 = \alpha y_t + (1-\alpha)S_{t-1}^1$ $S_t^2 = \alpha S_t^1 + (1-\alpha)S_{t-1}^2$ $S_t^3 = \alpha S_t^2 + (1-\alpha)S_{t-1}^3$	$\hat{a}_0 = 3(S_t^1 - S_t^2) + S_t^3$ $\hat{a}_1 = \frac{\alpha}{2(1-\alpha)^2} [(6-5\alpha)S_t^1 - 2(5-4\alpha)S_t^2 + (4-3\alpha)S_t^3]$ $\hat{a}_2 = \frac{\alpha^2}{(1-\alpha)^2} (S_t^1 - 2S_t^2 + S_t^3)$	$y_{t+1}^* = \hat{a}_0 + \hat{a}_1 t + \frac{1}{2} t^2 \hat{a}_2$	$\sigma_{y_{t+1}}^* = \sigma_{\varepsilon_t} \times \sqrt{2\alpha + 3\alpha^2 + 3\alpha t^2}$

7. Рассчитывается величина среднего квадрата ошибки по формуле

$$S^2 = \frac{\sum (y_t - y_{\text{расч}})^2}{n - 1}.$$

Задается новое значение параметра адаптации α . Вся процедура с 3-го по 7-й шаг повторяется до тех пор, пока параметр адаптации не достигнет наибольшего (но меньшего единицы) значения.

8. Определяется оптимальный параметр адаптации на основе минимальности величины ошибки S .

9. Окончательная модель прогноза формируется на последнем шаге в момент времени $t = n$ при заданном оптимальном параметре адаптации α . Прогнозное значение вычисляется на базе выражения (5.4) путем подстановки в него значений (оценок) коэффициентов, полученных на последнем шаге $t = n$, и времени упреждения l .

Пример 5.1. В качестве примера для иллюстрации применения метода экспоненциального сглаживания рассмотрим временной ряд потребления электроэнергии, представленный в п. 3.2.2 на рис. 3.3. Уравнение линейного тренда для этого ряда

$$\hat{y}_t = 3782,32 - 64,2t.$$

Для построения модели прогноза использовались формулы, взятые из табл. 5.1, для линейной модели. Изначально задается значение параметра адаптации, близкое к нулю (0,1). Далее по формулам табл. 5.1 определяются начальные условия $S_0^1 = 3841,8$; $S_0^2 = 3901,3$, где за a_0 , a_1 взяты оценки линейного тренда. Затем для $t = 1$ были рассчитаны экспоненциальные средние S_1^1 , S_1^2 и вычислены оценки коэффициентов \hat{a}_0 , \hat{a}_1 , осуществлен прогноз на один шаг вперед (на следующий 1994 г.). Далее для всех $t \leq n$ по рекуррентной формуле вычислялись новые S_t^1 , S_t^2 , по ним определялись соответствующие \hat{a}_0 и \hat{a}_1 и строился прогноз на один шаг.

При прогнозировании потребления электроэнергии на 2013–2014 гг. использовались оценки коэффициентов \hat{a}_0 и \hat{a}_1 , полученных на последнем шаге прогнозирования при $t = n$, и оптимальная величина параметра сглаживания $\alpha = 0,52$ (табл. 5.2), которая была получена при реализации метода экспоненциального сглаживания в табличном процессоре MS Excel по этапной схеме, предложенной ранее.

В табл. 5.3 представлен процесс построения модели прогноза по методу экспоненциального сглаживания при $\alpha = 0,52$.

Таблица 5.2

Значения параметра адаптации α и соответствующие им значения среднеквадратических ошибок S

α	0,10	0,11	0,12	...	0,51	0,52	0,53	...	0,88	0,89	0,90
S	186,51	185,53	184,38	...	146,40	146,38	146,41	...	175,66	177,17	178,72

Таблица 5.3

Расчеты по методу простого экспоненциального сглаживания

Год	Порядковый номер года t	Исходные уровни ряда y_t	Экспоненциальные средние		Оценки коэффициентов		$y_{расч}$	Отклонения фактических значений от расчетных $e_t = y_t - y_{расч}$
			S_t^1	S_t^2	\hat{a}_0	\hat{a}_1		
1993	1	4 252	4 054,7	3 980,9	4 128,5	79,7	—	—
1994	2	4 167	4 113,0	4 049,5	4 176,5	68,6	4 208,2	-41,2
1995	3	3 884	3 994,1	4 020,8	3 967,5	-28,7	4 245,1	-361,1
1996	4	3 730	3 857,0	3 935,8	3 778,3	-85,0	3 938,8	-208,8
1997	5	3 227	3 530,0	3 725,1	3 334,8	-210,6	3 693,3	-466,3
1998	6	2 999	3 254,3	3 480,7	3 027,9	-244,4	3 124,2	-125,2
1999	7	2 912	3 076,6	3 271,0	2 882,3	-209,8	2 783,6	128,4
2000	8	2 889	2 979,2	3 119,5	2 838,9	-151,4	2 672,5	216,5
2001	9	2 829	2 901,2	3 006,2	2 796,3	-113,3	2 687,5	141,5
2002	10	2 840	2 869,5	2 935,2	2 803,7	-71,0	2 683,0	157,0
2003	11	2 909	2 890,0	2 911,7	2 868,2	-23,5	2 732,7	176,3
2004	12	2 852	2 870,3	2 890,2	2 850,3	-21,5	2 844,7	7,3
2005	13	2 903	2 887,3	2 888,7	2 885,8	-1,5	2 828,8	74,2
2006	14	2 999	2 945,3	2 918,1	2 972,5	29,4	2 884,3	-114,7
2007	15	2 812	2 876,1	2 896,3	2 855,9	-21,8	3 001,9	-189,9
2008	16	2 768	2 820,0	2 856,7	2 783,3	-39,6	2 834,1	-66,1
2009	17	2 836	2 828,3	2 841,9	2 814,7	-14,7	2 743,7	92,3
2010	18	2 933	2 882,6	2 863,1	2 902,2	21,1	2 799,9	133,1
2011	19	2 492	2 679,9	2 768,0	2 591,8	-95,1	2 923,4	-431,4
2012	20	2 930	2 809,7	2 789,6	2 829,8	21,7	2 496,6	433,4

Точный прогноз на следующие 3 года рассчитывается таким образом:

$$2013 \text{ год } (l = 1): y(20 + 1) = y(21) = 2\,829,8 + 21,7 \cdot 1 = 2\,851,5;$$

$$2014 \text{ год } (l = 2): y(20 + 2) = y(22) = 2\,829,8 + 21,7 \cdot 2 = 2\,873,2;$$

$$2015 \text{ год } (l = 3): y(20 + 3) = y(23) = 2\,829,8 + 21,7 \cdot 3 = 2\,895,7.$$

В табл. 5.4 представлены результаты прогноза потребления электроэнергии на 2013–2015 гг., а также доверительные интервалы

прогноза для 5% уровня значимости, полученные как $y_{t+l}^* \pm t_{\alpha}^* \sigma_{y_{t+l}^*}$.

Таблица 5.4

Прогноз потребления электроэнергии на 2013–2015 гг.
по методу экспоненциального сглаживания

Год	Прогнозное значение	Доверительный интервал прогноза при $p = 0,95$	
		нижняя граница	верхняя граница
2013	2 851,9	2 523,8	3 180,0
2014	2 873,8	2 462,5	3 285,1
2015	2 895,7	2 399,7	3 391,8

Фактические, расчетные и прогнозные значения потребления электроэнергии в регионе представлены на рис. 5.1. Для проверки точности полученных прогнозных оценок необходимо сравнить их с фактическими данными, рассчитав ошибки прогноза.

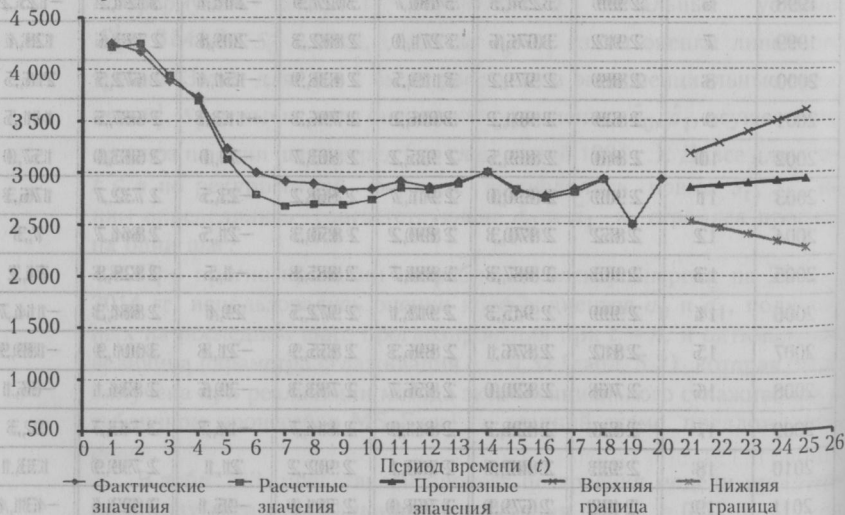


Рис. 5.1. Фактические, расчетные и прогнозные значения потребления электроэнергии в регионе

5.3. МЕТОД ЭКСПОНЕНЦИАЛЬНОГО СГЛАЖИВАНИЯ С ТРЭКИНГ-СИГНАЛОМ

В том случае, если процесс подвержен внезапным изменениям и нужно выявить момент их возникновения, а параметры модифицировать надлежащим образом и на необходимое время, следует использовать так называемый трэкинг-сигнал (следящий сигнал). Он улавливает момент, когда необходимо вмешательство исследователя. Формула трэкинг-сигнала имеет следующий вид:

$$TRS_t = \frac{(\varepsilon_t)_{\text{сгл}}}{|\varepsilon_t|_{\text{сгл}}}, \quad (5.6)$$

где $(\varepsilon_t)_{\text{сгл}}$ — сглаженная ошибка, вычисляемая по формуле

$$(\varepsilon_t)_{\text{сгл}} = \gamma \varepsilon_t + (1 - \gamma)(\varepsilon_{t-1})_{\text{сгл}},$$

где γ — параметр сглаживания;

$|\varepsilon_t|_{\text{сгл}}$ — абсолютные значения ошибки прогноза, получаемые из выражения

$$|\varepsilon_t|_{\text{сгл}} = \gamma |\varepsilon_t| + (1 - \gamma) |\varepsilon_{t-1}|_{\text{сгл}}.$$

Для коротких временных рядов предлагается за начальное значение сглаженной ошибки $(\varepsilon_{t-1})_{\text{сгл}}$ выбирать нуль, а ее модуля $|\varepsilon_{t-1}|_{\text{сгл}}$ — первое значение модуля ошибки $|\varepsilon_t|$ [30].

Схема прогнозирования временных рядов по методу экспоненциального сглаживания с использованием трэкинг-сигнала состоит из следующих этапов.

1. Выбирают вид модели экспоненциального сглаживания.
2. Параметр α определяют из формулы (5.5) либо задают.
3. Задают начальное значение γ .
4. Вычисляют начальные условия.
5. Находят экспоненциальные средние.
6. Определяют оценки коэффициентов модели прогноза.
7. Осуществляют прогноз на одну точку вперед.
8. Находят отклонения фактического значения временного ряда от прогнозируемого.
9. Рассчитывают величину трэкинг-сигнала, которая сравнивается с пороговым значением с учетом выбранной вероятности [30]. Пороговые значения при уровне вероятности 0,95 даны в табл. 5.5.
10. Устанавливают, выходит ли трэкинг-сигнал за пороговое значение:
 - а) если нет, переходят к 5-му этапу схемы прогнозирования;
 - б) если да, параметр α увеличивают на какую-либо величину и

переходят к 5-му этапу, производя расчеты с новым α для периода $t + 2$.

Таблица 5.5

Пороговые значения трэкинг-сигнала при уровне вероятности 0,95

γ	Постоянная модель	Линейная модель	Квадратичная модель
0,01	0,13	0,11	0,10
0,05	0,28	0,24	0,22
0,10	0,39	0,35	0,32
0,15	0,50	0,45	0,41
0,20	0,58	0,52	0,47
0,25	0,64	0,58	0,54
0,30	0,70	0,64	0,60
0,35	0,76	0,69	0,66
0,40	0,80	0,74	0,70
0,45	0,83	0,78	0,75
0,50	0,87	0,82	0,79
0,55	0,90	0,85	0,83
0,60	0,93	0,89	0,86
0,65	0,95	0,91	0,89
0,70	0,96	0,94	0,92
0,75	0,98	0,96	0,94
0,80	0,99	0,97	0,96

Такая последовательность осуществляется до тех пор, пока не будет проанализирован весь временной ряд. Если для последнего уровня ряда величина трэкинг-сигнала не выходит за пороговое значение, то вычисляют оценки параметров прогноза и осуществляют прогноз на заданный период упреждения. Если трэкинг-сигнал выходит за пороговое значение, то параметр следует увеличить на какую-либо величину, а затем перейти к этапам 2, 4–10. Прогноз осуществляется по формуле (5.2) путем подстановки в нее последних значений коэффициентов и времени упреждения l .

Пример 5.2. В качестве примера для иллюстрации применения метода экспоненциального сглаживания с использованием трэкинг-сигнала рассмотрим тот же временной ряд потребления электроэнергии, что и в примере 5.1.

Для построения модели прогноза применялись формулы, взятые из табл. 5.1, для квадратичной модели. В наших расчетах использовалась величина параметра сглаживания $\alpha = 0,52$, которую

получили с помощью критерия минимума средней квадратической ошибки.

В табл. 5.6 представлен процесс построения модели прогноза по методу экспоненциального сглаживания с использованием трэкинг-сигнала.

Таблица 5.6

Схема прогнозирования по методу экспоненциального сглаживания с применением трэкинг-сигнала

Порядковый номер года t	Исходные уровни y_t	Расчетные уровни $y_{расч}$	Отклонения e_t	$ e_t $	$(e_t)_{сгл}$	$ e_t _{сгл}$	TRS
1	4252	—	—	—	—	—	—
2	4167	4208,2	-41,2	41,2	-4,1	41,2	-0,10
3	3884	4245,1	-361,1	361,1	-39,8	73,2	-0,54
4	3730	3938,8	-208,8	208,8	-56,7	86,7	-0,65
5	3227	3693,3	-466,3	466,3	-97,7	124,7	-0,78
6	2999	3124,2	-125,2	125,2	-100,4	124,7	-0,80
7	2912	2783,6	128,4	128,4	-77,5	125,1	-0,62
8	2889	2672,5	216,5	216,5	-48,1	134,2	-0,36
9	2829	2687,5	141,5	141,5	-29,2	135,0	-0,22
10	2840	2683,0	157,0	157,0	-10,5	137,2	-0,08
11	2909	2732,7	176,3	176,3	8,1	141,1	0,06
12	2852	2844,7	7,3	7,3	8,0	127,7	0,06
13	2903	2828,8	74,2	74,2	14,7	122,4	0,12
14	2999	2884,3	114,7	114,7	24,7	121,6	0,20
15	2812	3001,9	-189,9	189,9	3,2	128,4	0,03
16	2768	2834,1	-66,1	66,1	-3,7	122,2	-0,03
17	2836	2743,7	92,3	92,3	5,9	119,2	0,05
18	2933	2799,9	133,1	133,1	18,6	120,6	0,15
19	2492	2923,4	-431,4	431,4	-26,4	151,7	-0,17
20	2930	2496,6	433,4	433,4	19,6	179,8	0,11

При $\gamma = 0,1$ для линейной модели экспоненциального сглаживания критическое значение $TRS = 0,35$. Абсолютные фактические значения трэкинг-сигнала для t со 2-го по 8-й превышают этот уровень, поэтому можно сделать вывод о том, что изучаемый процесс подвержен внезапным изменениям. При необходимости модель может быть скорректирована путем изменения значения γ и/или параметра сглаживания α .

Тест и задания для самоконтроля

Тест

1. Значение параметра адаптации может быть определено:

- а) методом экспертных оценок;
- б) методом проб или аналитическим способом;
- в) это табулированное значение.

2. Какое значение может принимать параметр адаптации α :

- а) $\alpha < 1$;
- б) $0 < \alpha < 1$;
- в) нет ограничений?

3. При каком значении параметра адаптации модель экспоненциального сглаживания будет иметь более «гладкий» характер:

- а) $\alpha = 0,5$;
- б) $\alpha = 0,9$;
- в) $\alpha = 0,1$?

4. Какой вид имеет формула экспоненциальной средней первого порядка:

а) $S_t = \alpha y_t + (1 - \alpha)y_{t-1}$;

б) $S_t = \alpha y_t + (1 - \alpha)S_{t-1}$;

в) $S_t = \alpha y_t + \alpha S_{t-1}$?

5. Дисперсия экспоненциальной средней и дисперсия временного ряда:

- а) совпадают;
- б) дисперсия экспоненциальной средней больше дисперсии временного ряда;
- в) дисперсия экспоненциальной средней меньше дисперсии временного ряда.

6. В качестве критерия оптимальности при выборе параметра адаптации обычно используют:

- а) критерий минимума среднего квадрата ошибок прогнозирования;
- б) критерий максимума среднего квадрата ошибок прогнозирования;
- в) критерий Дарбина — Уотсона.

7. Трэйкинг-сигнал применяется для:

- а) проверки точности модели;
- б) оценки параметров модели;
- в) выявления момента внезапных изменений в динамике изучаемого процесса.

Практические задания

1. Рассчитайте экспоненциальную среднюю для временного ряда курса акций некоторой фирмы. В качестве начального значения экспоненциальной средней возьмите среднее значение из пяти первых уровней ряда. Значение параметра адаптации примите равным 0,1.

t	y_t	t	y_t	t	y_t
1	510	11	494	21	523
2	497	12	499	22	527
3	504	13	502	23	523
4	510	14	509	24	528
5	509	15	525	25	529
6	503	16	512	26	538
7	500	17	510	27	539
8	500	18	506	28	541
9	500	19	515	29	543
10	495	20	522	30	541

2. По данным предыдущего задания рассчитайте экспоненциальную среднюю при значении параметра адаптации, равном 0,5. Сравните графически исходный временной ряд и ряды экспоненциальных средних, полученных при разных параметрах адаптации. Укажите, какой ряд имеет более «гладкий» характер.

3. Прогнозирование курса акций осуществляется на основе адаптивной полиномиальной модели первого порядка. На последнем шаге, соответствующем 16 апреля 2012 г., получены следующие оценки коэффициентов: $a_0 = 405$; $a_1 = 1,87$. Рассчитайте прогноз курса акций на 17 и 18 апреля 2012 г.

4. Прогнозирование курса акций осуществлялось на основе адаптивной квадратичной модели (полиномиальной модели второго порядка). На последнем шаге получены следующие оценки коэффициентов:

$$a_0 = 541; \quad a_1 = 1,7; \quad a_2 = 0,52.$$

Рассчитайте прогноз курса акций на три дня вперед.

5. Имеются данные об изменении некоторого показателя.

t	2007	2008	2009	2010	2011	2012
y_t	734	537	495	504	485	379

Постройте прогноз на 2013–2014 гг., используя метод экспоненциального сглаживания, если известно, что уравнение тренда выглядит следующим образом:

$$y_t = 713,6 - 54,74t.$$

Параметр адаптации возьмите равным 0,3; 0,5; 0,7.

Определите, какая адаптивная модель (с каким параметром) лучше описывает временной ряд. Установите точность наилучшей адаптивной модели и линейного тренда. Сделайте вывод.

6. АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ С ПЕРИОДИЧЕСКИМИ КОЛЕБАНИЯМИ

- 6.1. Методы выявления периодической компоненты.
- 6.2. Статистический анализ сезонности.
- 6.3. Прогнозирование на основе моделей с учетом индексов сезонности.
- 6.4. Моделирование сезонных колебаний при помощи гармоник Фурье.
- 6.5. Адаптивные модели сезонных колебаний.
- 6.6. Фиктивные переменные в моделировании сезонных колебаний.

Потребности экономической практики в различных сферах (в строительстве, сельском хозяйстве, на фондовых рынках, в инвестировании и др.) послужили мощным стимулом к совершенствованию статистической методологии в области выявления, измерения, моделирования и прогнозирования периодических колебаний. В настоящее время при описании и прогнозировании тренд-сезонных процессов используются подходы, связанные с применением индексов сезонности в сочетании с кривыми роста; процедуры, опирающиеся на широкий спектр адаптивных моделей; сезонный вариант модели *ARIMA* и др., а также разрабатываются специализированные подходы, учитывающие специфику конкретных временных рядов.

6.1. МЕТОДЫ ВЫЯВЛЕНИЯ ПЕРИОДИЧЕСКОЙ КОМПОНЕНТЫ

Для выявления периодической компоненты во временном ряду можно воспользоваться различными способами.

1. Наличие периодических колебаний во временном ряду можно определить по виду графика. Так, по графическому изображению ряда, представленного на рис. 6.1, отчетливо проявляется рост показателя к концу каждого года (т. е. каждый год прослеживается одна и та же закономерность).

2. Для определения периодичности можно использовать автокорреляционную функцию (см. п. 2.4). Наличие выбросов (пиковых значений) в ней указывает на временной лаг. Если выбросы повторяются, то можно предположить наличие периодической составляющей в исследуемом ряду. Так, для ряда, изображенного на рис. 6.1, график автокорреляционной функции $r(\tau)$ представлен на рис. 6.2, по которому отчетливо видно наличие пиковых значений, повторяющихся через определенный интервал τ ($\tau = 4, 8, 12$).

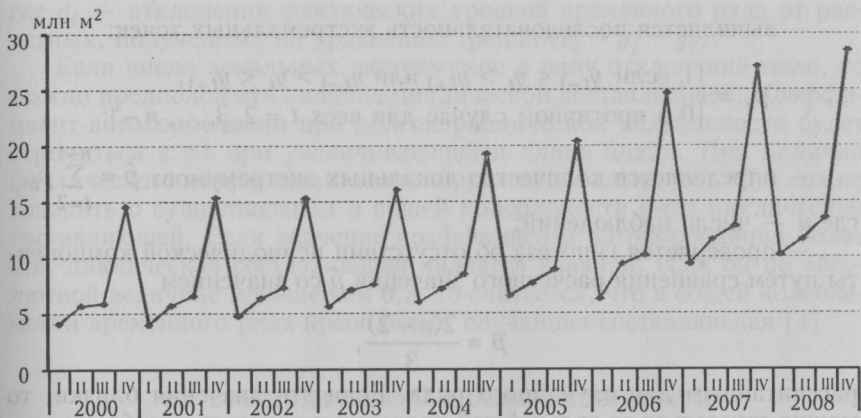


Рис. 6.1. Временной ряд с периодической составляющей

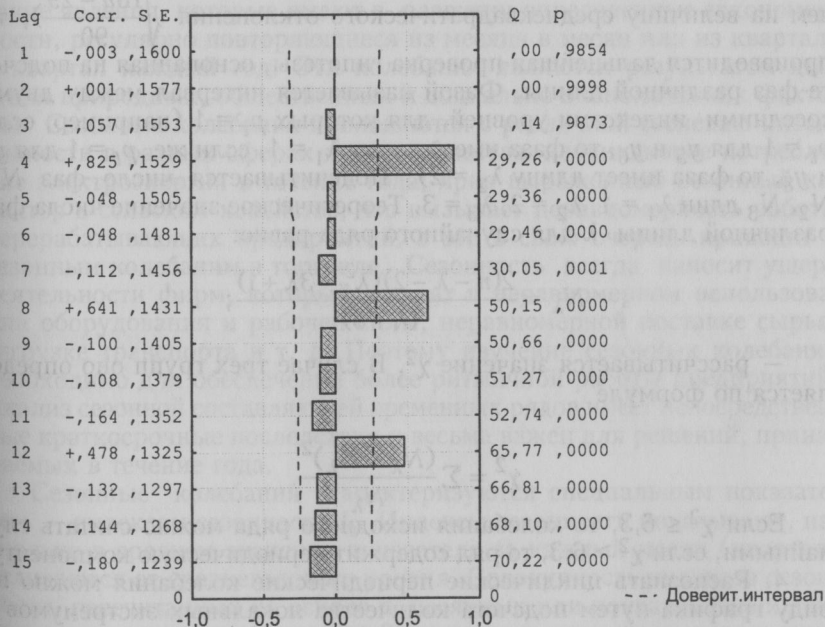


Рис. 6.2. График автокорреляционной функции временного ряда

3. Для длинных рядов также можно использовать критерий, в основе которого лежит подсчет числа экстремальных точек. Этот расчет осуществляется следующим образом:

– выявляется последовательность экстремальных точек:

$$p_t = \begin{cases} 1, & \text{если } y_{t-1} < y_t > y_{t+1} \text{ или } y_{t-1} > y_t < y_{t+1}; \\ 0 & \text{в противном случае для всех } t = 2, 3, \dots, n-1; \end{cases}$$

– определяется количество локальных экстремумов: $\hat{p} = \sum_{t=2}^{n-1} p_t$,
 где n – число наблюдений;
 – проверяется гипотеза об отсутствии периодической компоненты путем сравнения расчетного значения \hat{p} со значением

$$\bar{p} = \frac{2(n-2)}{3},$$

рассчитанным для случайного ряда. Если эти значения близки, то можно отказаться от дальнейшей проверки и признать колебания случайными. Если \hat{p} и \bar{p} значительно отличаются друг от друга (более

чем на величину среднеквадратического отклонения $\sqrt{\frac{16n-29}{90}}$), то

производится дальнейшая проверка гипотезы, основанная на подсчете фаз различной длины. Фазой называется интервал между двумя соседними индексами уровней, для которых $p_t = 1$ (например, если $p_t = 1$ для y_5 и y_6 , то фаза имеет длину $\lambda_1 = 1$, если же $p_t = 1$ для y_6 и y_8 , то фаза имеет длину $\lambda_2 = 2$). Подсчитывается число фаз N_1, N_2, N_3 длин $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 3$. Теоретическое значение числа фаз различной длины λ для случайного ряда равно:

$$N_\lambda = \frac{2(n-\lambda-2)(\lambda^2 - 3\lambda + 1)}{(\lambda + 3)!};$$

– рассчитывается значение χ^2 . В случае трех групп оно определяется по формуле

$$\chi^2 = \sum \frac{(N_\lambda - N_\lambda)^2}{N_\lambda}.$$

Если $\chi^2 \leq 6,3$, то колебания исходного ряда можно считать случайными, если $\chi^2 > 6,3$, то ряд содержит периодическую компоненту.

4. Распознать циклические периодические колебания можно по виду графика путем подсчета количества локальных экстремумов в ряду отклонений от тренда и по коэффициенту автокорреляции отклонений 1-го порядка:

$$r_1 = \frac{\sum_{t=1}^{n-1} e_t e_{t+1}}{\frac{e_1^2}{2} + \sum_{t=2}^{n-1} e_t^2 + \frac{e_n^2}{2}}, \quad (6.1)$$

где e_t — отклонения фактических уровней временного ряда от расчетных, полученных по уравнению тренда ($e_t = y_t - \hat{y}_t$).

Если число локальных экстремумов в ряду отклонений мало, то можно предположить наличие циклической составляющей. Коэффициент автокорреляции при долгопериодической колеблемости будет стремиться к +1 при увеличивающейся длине цикла. При наличии фактического коэффициента автокорреляции больше чем +0,3 можно говорить о существовании в общей колеблемости ряда циклической составляющей. Если значение коэффициента автокорреляции более 0,6, циклическая составляющая является главной; если он по абсолютной величине меньше чем 0,3, то считается, что в общей колеблемости временного ряда преобладает случайная составляющая [4].

6.2. СТАТИСТИЧЕСКИЙ АНАЛИЗ СЕЗОННОСТИ

Одна из основных задач анализа временного ряда состоит в выделении сезонной компоненты, или сезонности. К сезонным относят такие явления, которые имеют в развитии определенные закономерности, регулярно повторяющиеся из месяца в месяц или из квартала в квартал каждый год. Эти колебания являются результатом действия природных, общественных и социально-экономических факторов. Сезонные колебания проявляются с различной степенью интенсивности в разных сферах жизни общества: неодинаковое потребление электроэнергии в течение года; ярко выраженный сезонный характер в сельском хозяйстве, что вызывает неравномерность работы перерабатывающих предприятий, а это в свою очередь приводит к сезонным колебаниям в торговле. Сезонность всегда наносит ущерб деятельности фирм, который состоит в неравномерном использовании оборудования и рабочей силы, неравномерной поставке сырья, загрузке транспорта и т. д. Поэтому изучение сезонных колебаний необходимо для обеспечения более ритмичной работы предприятий. Анализ сезонной составляющей временных рядов имеет непосредственные краткосрочные последствия и весьма важен для решений, принимаемых в течение года.

Сезонные колебания характеризуются специальным показателем — индексом сезонности I_s . Индексы сезонности показывают, насколько в среднем значения показателя за каждый месяц (квартал) отличаются от среднегодового уровня. Совокупность индексов сезонности, рассчитанных на основе помесечных или поквартальных данных за несколько (не менее 3) лет, отражает сезонную волну. В общем виде эти показатели определяются как процентное соотношение фактического и некоторого теоретического, принятого за базу уровней. Для вычисления индексов необходимо учитывать наличие основной тенденции во временном ряду.

1. Если временной ряд не содержит ярко выраженной тенденции

в развитии, то индексы сезонности рассчитываются непосредственно по фактическим данным без предварительного выравнивания (способ постоянной средней). Для каждого месяца (квартала) определяется средняя величина уровня \bar{y}_i , затем вычисляется средний уровень для всего ряда \bar{y} и находится процентное отношение средних для каждого месяца к общему среднему уровню ряда, т. е. индексы сезонности определяются как

$$I_{s_i} = \frac{\bar{y}_i}{\bar{y}} \cdot 100 \%, \quad (6.2)$$

где $i = 1, 2, \dots, 12$, если данные помесечные, или $i = 1, 2, 3, 4$, если поквартальные.

2. Если временной ряд содержит определенную тенденцию в развитии, то, прежде чем вычислить индексы сезонности, по фактическим уровням выявляется основная тенденция, которая может быть выражена с помощью аналитического выравнивания в виде математической функции (способ переменной средней). Ход вычислений индексов сезонности следующий:

1) находится подходящая функция выравнивания (уравнение тренда) (п. 3.2.2);

2) вычисляются теоретические уровни \hat{y}_t , рассчитываемые по уравнению тренда для всех $t = 1, 2, \dots, n$;

3) определяются показатели сезонности как процентное отношение фактических уровней y_t к соответствующим расчетным уровням \hat{y}_t (для всех $t = 1, \dots, n$):

$$f_t = \frac{y_t}{\hat{y}_t} \cdot 100 \%; \quad (6.3)$$

4) индексы сезонности находятся как средние арифметические из показателей сезонности f_t по одноименным периодам (либо по месяцам, либо по кварталам). Например, индекс сезонности для первого квартала определяется так:

$$I_{s_1} = \frac{f_1 + f_5 + f_9 + \dots}{m}, \quad (6.4)$$

где m — число одноименных периодов (лет);

5) расчет индексов заканчивается проверкой, которая заключается в определении среднего индекса сезонности \bar{I}_s . Он должен быть равен 100 % (или 1, если вычисление производится в коэффициентах). Если $\bar{I}_s \neq 100\%$, то следует произвести выравнивание индексов, которое заключается в их умножении на величину, обратную среднему индексу сезонности, $1/\bar{I}_s$.

Часто временные ряды с ярко выраженной сезонностью исполь-

зуют для регрессионного анализа. В этом случае желательно элиминировать (устранить) влияние сезонности. Если необходимо найти уровни временного ряда, в которых элиминировано влияние сезонности, то нужно фактические уровни ряда поделить на соответствующие индексы сезонности (в коэффициентах).

При анализе устойчивости сезонных колебаний в динамике следует рассчитывать показатель колеблемости временного ряда за счет сезонности, вычисляемый по формуле

$$\sigma = \sqrt{\frac{\sum (I_s - 100)^2}{k}},$$

где $k = 12$ (если данные помесечные) либо $k = 4$ (поквартальные).

Снижение значения показателя говорит об уменьшении влияния сезонного фактора на исследуемый показатель.

Пример 6.1. По данным табл. 6.1 о продаже яиц в магазинах города за три года вычислить индексы сезонных колебаний.

Таблица 6.1

Среднедневная реализация яиц, тыс. шт.

Месяц	2008 г.	2009 г.	2010 г.	Среднее значение $\bar{y}_i = \frac{\sum y_i}{m}$	Индексы сезонности	
					мультипликативные, %	аддитивные
1	2	3	4	5	6	7
Январь	94,2	98,8	91,0	94,7	94,8	-5,1
Февраль	95,1	99,4	92,4	95,6	95,8	-4,2
Март	96,7	99,5	100,3	98,8	99,0	-1,0
Апрель	97,6	101,4	99,5	99,5	99,7	-0,3
Май	90,1	89,2	93,1	90,8	91,0	-9,0
Июнь	118,7	124,4	125,9	123,0	123,2	23,2
Июль	116,8	108,4	116,7	114,0	114,2	14,2
Август	100,1	91,0	98,5	96,5	96,7	-3,3
Сентябрь	101,5	101,9	94,8	99,4	99,6	-0,4
Октябрь	92,5	94,2	96,3	94,3	94,5	-5,5
Ноябрь	88,9	89,8	92,2	90,3	90,5	-9,5
Декабрь	99,1	100,0	103,1	100,7	100,9	0,9
Средний уровень	99,3	99,8	100,3	99,8	100,0	0

Рассматривая средние уровни ряда за каждый год, видим, что ряд не имеет ярко выраженной тенденции (в среднем каждый год дневная реализация яиц составляет около 100 тыс. шт.). Поэтому

индексы сезонности будем вычислять без предварительного выравнивания, применяя способ постоянной средней.

1. По данным табл. 6.1 найдем усредненные значения уровней по одноименным периодам по формуле средней арифметической (см. графу 5):

январь: $(94,2 + 98,8 + 91,0) : 3 = 94,7$ и т. д.

2. Затем определим средний уровень ряда по вычисленным среднемесячным уровням \bar{y}_t или по среднегодовым уровням либо по исходным данным:

$$\bar{y} = 99,8.$$

3. Далее для каждого месяца рассчитаем индексы сезонности (см. графу 6):

январь: $I_s = 94,7 : 99,8 = 0,948$ (94,8 %) и т. д.

Совокупность вычисленных индексов характеризует волну продажи во внутригодовой динамике. Полученные данные можно изобразить в виде линейной диаграммы (рис. 6.3).

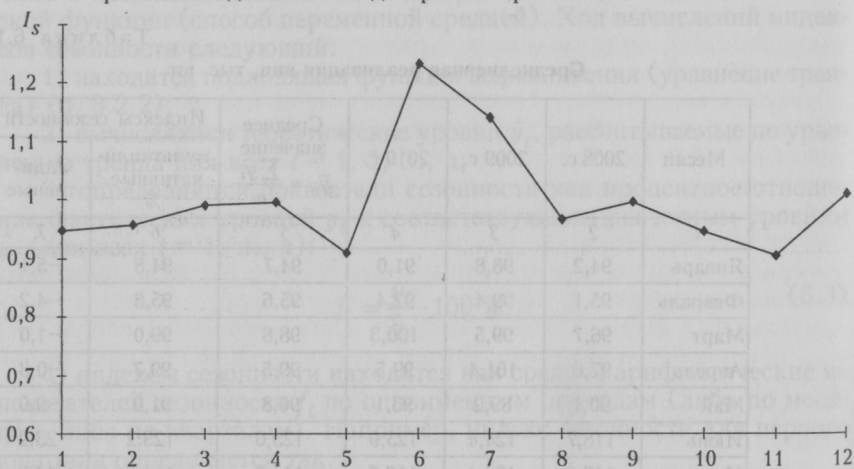


Рис. 6.3. Сезонная волна реализации яиц по месяцам

Из табл. 6.1 (см. графу 6) и графика видно, что сезонные колебания характеризуются значительным повышением в июне (+23,2 %), июле (+14,2 %) и менее процента в декабре. В другие месяцы отмечается снижение реализации.

Пример 6.2. В качестве примера ряда, имеющего тенденцию к росту, рассмотрим поквартальный ряд динамики ввода жилья в РФ за 2000—2008 гг. Фактические значения ряда представлены на графике (см. рис. 6.1) и в табл. 6.2. Данные за 2009—2011 гг. в дальнейшем будут рассмотрены при оценке точности построенных прогнозов.

Таблица 6.2

Динамика ввода жилых домов в РФ, млн м²

Год	Квар-тал	Исходные уровни ряда y_t	Рачетные уровни по уравнению тренда \hat{y}_t	Показатели сезонности		Расчетные значения по индексной сезонной модели	
				мультипликативные f_t	аддитивные g_t	с мультипликативными индексами	с аддитивными индексами
1	2	3	4	5 = 3 : 4	6 = 3 - 4	7	8
2000	I	4,1	7,22	0,568	-3,12	4,37	3,54
	II	5,7	7,21	0,791	-1,51	5,48	4,80
	III	5,9	7,22	0,817	-1,32	5,99	5,41
	IV	14,6	7,25	2,014	7,35	13,14	15,15
2001	I	4,0	7,30	0,548	-3,30	4,42	3,62
	II	5,7	7,37	0,774	-1,67	5,60	4,96
	III	6,6	7,45	0,885	-0,85	6,18	5,65
	IV	15,4	7,56	2,037	7,84	13,71	15,46
2002	I	4,8	7,69	0,624	-2,89	4,66	4,01
	II	6,4	7,83	0,817	-1,43	5,95	5,42
	III	7,3	8,00	0,913	-0,70	6,63	6,19
	IV	15,3	8,18	1,870	7,12	14,83	16,08
2003	I	5,6	8,39	0,668	-2,79	5,08	4,71
	II	7,1	8,61	0,825	-1,51	6,54	6,20
	III	7,6	8,85	0,859	-1,25	7,34	7,04
	IV	16,1	9,11	1,767	6,99	16,52	17,01
2004	I	6,0	9,39	0,639	-3,39	5,69	5,71
	II	7,2	9,69	0,743	-2,49	7,37	7,28
	III	8,5	10,01	0,849	-1,51	8,30	8,20
	IV	19,3	10,35	1,865	8,95	18,76	18,25
2005	I	6,6	10,71	0,616	-4,11	6,49	7,03
	II	7,7	11,09	0,695	-3,39	8,43	8,68
	III	8,9	11,48	0,775	-2,58	9,52	9,68
	IV	20,4	11,90	1,715	8,50	21,57	19,80
2006	I	6,3	12,33	0,511	-6,03	7,47	8,65
	II	9,4	12,79	0,735	-3,39	9,72	10,38
	III	10,2	13,26	0,769	-3,06	11,00	11,45
	IV	24,7	13,75	1,796	10,95	24,93	21,65
2007	I	9,4	14,27	0,659	-4,87	8,64	10,59
	II	11,6	14,80	0,784	-3,20	11,25	12,39
	III	12,8	15,35	0,834	-2,55	12,73	13,54
	IV	27,1	15,92	1,702	11,18	28,86	23,82
2008	I	10,2	16,51	0,618	-6,31	10,00	12,83
	II	11,6	17,12	0,678	-5,52	13,01	14,71
	III	13,5	17,75	0,761	-4,25	14,71	15,94
	IV	28,5	18,39	1,549	10,11	33,34	26,29

1. Для определения индексов сезонности было найдено уравнение тренда. Наилучшим образом тенденцию развития отображает уравнение полинома 2-го порядка $\hat{y}_t = 7,25 - 0,038t + 0,01t^2$. В табл. 6.2 (см. графу 4) приведены уровни, сглаженные по уравнению тренда.

2. Далее для каждого квартала определяются показатели сезонности как отношение фактических уровней к выравненным по уравнению тренда (см. графу 5).

3. Индексы сезонности рассчитываются как среднее арифметическое из показателей сезонности, взятых по одноименным периодам (табл. 6.3):

$$\text{I квартал: } I_s = (0,568 + 0,548 + 0,624 + \dots + 0,618)/9 = 0,606 \text{ (60,6 \%)} \text{ и т. д.}$$

4. Для проверки правильности вычислений индексов определяется средний индекс сезонности:

$$\frac{0,606 + 0,760 + 0,829 + 1,813}{4} = 1,002.$$

Если средний индекс сильно отличается от 1 (или 100 %), то следует произвести выравнивание путем умножения на величину, обратную среднему индексу сезонности. Индексы сезонности в относительном и абсолютном выражении представлены в табл. 6.3.

Таблица 6.3

Индексы сезонности

Квартал	Индексы сезонности		Скорректированные индексы сезонности	
	%	млн м ²	%	млн м ²
I	60,6	-4,09	60,48	-4,09
II	76,0	-2,68	75,85	-2,68
III	82,9	-2,01	82,73	-2,01
IV	181,3	8,78	180,94	8,78
Среднее значение	100,2	0,00	100,0	0,00

Выравненные (скорректированные) индексы сезонности (см. табл. 6.3) характеризуют модель сезонной волны ввода жилой площади во внутригодовом цикле. Наибольший ввод приходится на четвертый квартал со значительным превышением среднегодового уровня на 81,3 %, что составляет 8,78 млн м².

6.3. ПРОГНОЗИРОВАНИЕ НА ОСНОВЕ МОДЕЛЕЙ С УЧЕТОМ ИНДЕКСОВ СЕЗОННОСТИ

При наличии сезонных колебаний во временном ряду модель прогноза должна их учитывать. Очевидно, что процедуры расчета зависят от принятой структурной схемы временного ряда, содержащей сезонность в мультипликативной или аддитивной форме. В случае аддитивной связи сезонные колебания должны учитываться и вычисляться в абсолютных величинах, а мультипликативной — в относительных. Поэтому, прежде чем приступить к прогнозированию явлений, содержащих сезонные колебания, необходимо установить вид связи сезонной и трендовой компонент. Предварительно тип связи можно выявить, основываясь на графическом изображении ряда (см. п. 2.4, рис. 2.2, 2.3).

Тип связи можно определить по динамике отклонений фактических уровней от тренда $y_t - \hat{y}_t$ или $\frac{y_t - \hat{y}_t}{y_t}$. Если абсолютные отклонения имеют тенденцию к росту, а относительные варьируют приблизительно на одном уровне, то это свидетельствует о мультипликативной связи тренда и сезонной компоненты. Тип связи можно определить и по нормальному распределению отклонений фактических уровней от тренда. Если по нормальной кривой распределяются абсолютные сезонные отклонения, то связь между компонентами ряда аддитивна, если относительные — мультипликативна.

Определив вид связи и влияние сезонного фактора, можно использовать найденные закономерности для прогнозирования дальнейшего развития изучаемого явления. При этом можно оценить значения параметров тренда и краткосрочных сезонных колебаний. Случайные колебания можно оценить вероятностным путем. При экстраполяции следует учитывать, что чем продолжительнее период предсказания, тем вероятнее сделать большую ошибку, поэтому следует по мере получения новых фактических данных пересматривать модели.

Модель прогноза с помощью индексов сезонности на любой месяц (квартал) можно представить следующим образом:

$$y_t = I_{s_i} \hat{y}_t + \varepsilon_t, \quad (6.5)$$

где I_{s_i} — индекс сезонности i -го месяца или квартала ($i = 1, 2, \dots, 12$ или $i = 1, 2, 3, 4$), выраженный в коэффициентах;

\hat{y}_t — оценка исследуемого показателя, вычисленная по уравнению тренда;

$y_t^* = I_{s_i} \hat{y}_t$ — прогнозируемое значение показателя в момент времени $t = n + 1, n + 2, \dots, n + l$.

Если использовать аддитивную структурную схему временного ряда, то характеристики сезонности исчисляются в абсолютных величинах, т. е. показатели сезонности определяются как $g_t = y_t - \hat{y}_t$ (и т. д. аналогично рассмотренным в п. 6.1 индексам). Корректировку индексов проводят, если величина среднего индекса не равна нулю, путем вычитания из полученных аддитивных индексов их среднего значения. В табл. 6.3 представлены индексы сезонности, выраженные как в относительных (мультипликативные индексы), так и в абсолютных величинах (аддитивные индексы). В случае если выяснено, что связь между компонентами аддитивна, прогнозные значения получают путем прибавления аддитивных индексов сезонности к уровням, вычисленным по уравнению тренда.

Пример 6.3. На основе полученных в примере 6.2 индексов сезонности построим модели расчета исследуемого показателя для соответствующих кварталов:

$$\text{для I квартала } y_t = 0,6048 \cdot (7,25 - 0,038t + 0,01t^2) + \varepsilon_t;$$

$$\text{для II квартала } y_t = 0,7585 \cdot (7,25 - 0,038t + 0,01t^2) + \varepsilon_t;$$

$$\text{для III квартала } y_t = 0,8273 \cdot (7,25 - 0,038t + 0,01t^2) + \varepsilon_t;$$

$$\text{для IV квартала } y_t = 1,8094 \cdot (7,25 - 0,038t + 0,01t^2) + \varepsilon_t.$$

Расчетные значения по индексной сезонной модели, представленные на рис. 6.4, получают при $t = 1, 2, \dots, n$ (см. графу 7 табл. 6.2).

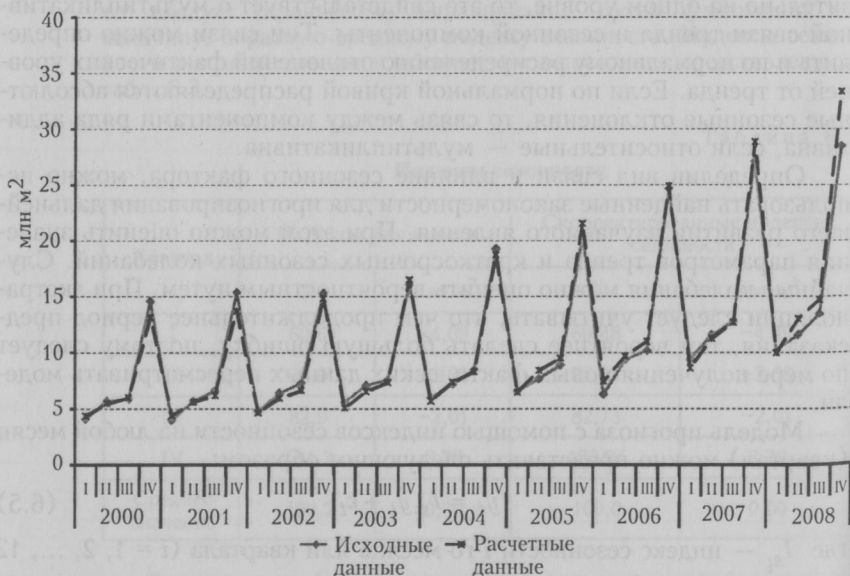


Рис. 6.4. Динамика ввода в эксплуатацию жилых домов в РФ и модель параболического тренда с мультипликативными индексами сезонности

Если при прогнозировании используется процедура сглаживания при помощи простой скользящей средней, то применяется следующий алгоритм.

Алгоритм построения тренд-сезонной модели для аддитивной структуры временного ряда ($Y_t = F_t + S_t + E_t$) [3; 36]

I этап. Производят сглаживание исходного временного ряда с помощью процедуры простой скользящей средней для четной длины интервала сглаживания (период скольжения равен 4 или 12) по формуле (3.3).

II этап. Рассчитывают абсолютные показатели сезонности $g_t = y_t - \tilde{y}_t$, где \tilde{y}_t — выравненный уровень по скользящей средней.

III этап. Вычисляют средние арифметические из одноименных абсолютных показателей сезонности:

$$G_{s_i} = \frac{1}{m} \sum g_i. \quad (6.6)$$

IV этап. Если $\sum G_{s_i} \neq 0$, то проводят корректировку первоначальных значений сезонной составляющей:

$$\tilde{G}_{s_i} = G_{s_i} - \frac{1}{m} \sum G_{s_i}. \quad (6.7)$$

V этап. Из исходных данных вычитают значения скорректированных показателей сезонности с целью элиминирования влияния сезонной составляющей: $y_i - \tilde{G}_{s_i}$.

VI этап. По десеASONализированному временному ряду проводят аналитическое выравнивание, т. е. определяют уравнение тренда \hat{y}_t .

VII этап. Осуществляют точечный прогноз по формуле $y_t^* = \tilde{G}_{s_i} + \hat{y}_t$. Рассчитывают доверительные интервалы.

Алгоритм построения тренд-сезонной модели для мультипликативной структуры временного ряда ($Y_t = F_t S_t E_t$)

I этап. Для выявления тенденции применяют простую скользящую среднюю при четной длине интервала сглаживания по формуле (3.3).

II этап. Рассчитывают показатели сезонности как отношения фактических значений к уровням сглаженного ряда, полученным на предыдущем шаге:

$$k_t = \frac{y_t}{\tilde{y}_t}.$$

III этап. Определяют средние арифметические из одноименных показателей сезонности:

$$K_{s_i} = \frac{1}{m} \sum k_i.$$

IV этап. Если для данных по кварталам $\sum K_{s_i} \neq 4$, а по месяцам — $\sum K_{s_i} \neq 12$, то корректируют первоначальные значения показателей сезонности:

$$\tilde{K}_{s_i} = K_{s_i} \frac{4(12)}{\sum K_{s_i}}. \quad (6.8)$$

V этап. Исходные данные делят на значения скорректированных показателей сезонности с целью элиминирования влияния сезонной составляющей: $\frac{y_i}{\tilde{K}_{s_i}}$.

VI этап. По десеонализированному временному ряду проводят аналитическое выравнивание, т. е. определяют уравнение тренда \hat{y}_t .

VII этап. Рассчитывают уровни временного ряда, обусловленные влиянием тенденции и сезонности, и осуществляют прогноз по тренд-сезонной мультипликативной модели $y_t^* = \tilde{K}_{s_i} \hat{y}_t(\text{скор})$.

Пример 6.4. Рассчитаем прогнозные значения ввода в эксплуатацию жилых домов по мультипликативной тренд-сезонной модели. Соответствующие расчеты приведены в табл. 6.4.

1. Сглаживание временного ряда проведем с помощью простой скользящей средней по формуле (3.3):

$$\tilde{y}_3 = \frac{\frac{1}{2} \cdot 4,1 + 5,7 + 5,9 + 14,6 + \frac{1}{2} \cdot 4}{5-1} = 7,56 \approx 7,6 \text{ и т. д.}$$

2. Рассчитываем показатели сезонности как отношение фактических и сглаженных уровней: $k_3 = \frac{5,9}{7,6} \approx 0,78$ и т. д.

3. Определяем средние арифметические из одноименных показателей сезонности. Так, для первого квартала средний показатель сезонности составит:

$$K_{s1} = \frac{0,524 + 0,572 + 0,632 + 0,643 + 0,624 + 0,552 + 0,657 + 0,658}{8} = 0,608,$$

а для третьего —

$$K_{s3} = \frac{0,78 + 0,822 + 0,854 + 0,831 + 0,823 + 0,819 + 0,782 + 0,835}{8} = 0,818.$$

Значения средних показателей сезонности представлены в графе 2 табл. 6.5.

Таблица 6.4

Расчет скорректированных значений по мультипликативной
тренд-сезонной модели

Год	Квар- тал	y_t	\tilde{y}_t	k_t	$y_t(\text{скор}) = \frac{y_i}{K S_i}$	$\hat{y}_t(\text{скор})$	y_t^*
1	2	3	4	5 = 3 : 4	6	7	8
2000	I	4,1	—	—	6,74	7,27	4,42
	II	5,7	—	—	7,58	7,32	5,50
	III	5,9	7,6	0,78	7,20	7,38	6,04
	IV	14,6	7,6	1,934	8,02	7,45	13,58
2001	I	4,0	7,6	0,524	6,58	7,54	4,59
	II	5,7	7,8	0,728	7,58	7,65	5,75
	III	6,6	8,0	0,822	8,06	7,77	6,36
	IV	15,4	8,2	1,875	8,45	7,90	14,38
2002	I	4,8	8,4	0,572	7,89	8,04	4,89
	II	6,4	8,5	0,756	8,51	8,20	6,16
	III	7,3	8,6	0,854	8,91	8,37	6,85
	IV	15,3	8,7	1,751	8,40	8,55	15,58
2003	I	5,6	8,9	0,632	9,21	8,75	5,32
	II	7,1	9,0	0,789	9,45	8,97	6,74
	III	7,6	9,2	0,831	9,28	9,19	7,53
	IV	16,1	9,2	1,748	8,84	9,43	17,18
2004	I	6,0	9,3	0,643	9,87	9,68	5,89
	II	7,2	9,9	0,731	9,58	9,95	7,48
	III	8,5	10,3	0,823	10,38	10,23	8,38
	IV	19,3	10,5	1,845	10,60	10,53	19,17
2005	I	6,6	10,6	0,624	10,85	10,83	6,59
	II	7,7	10,8	0,715	10,24	11,15	8,38
	III	8,9	10,9	0,819	10,87	11,49	9,41
	IV	20,4	11,0	1,848	11,20	11,84	21,56
2006	I	6,3	11,4	0,552	10,36	12,20	7,42
	II	9,4	12,1	0,776	12,51	12,57	9,45
	III	10,2	13,0	0,782	12,45	12,96	10,62
	IV	24,7	13,7	1,803	13,56	13,37	24,34
2007	I	9,4	14,3	0,657	15,46	13,78	8,38
	II	11,6	14,9	0,777	15,43	14,21	10,68
	III	12,8	15,3	0,835	15,63	14,66	12,00
	IV	27,1	15,4	1,757	14,88	15,11	27,53
2008	I	10,2	15,5	0,658	16,77	15,58	9,48
	II	11,6	15,8	0,735	15,43	16,07	12,08
	III	13,5	—	—	16,48	16,57	13,57
	IV	28,5	—	—	15,65	17,08	31,10

4. Так как сумма средних значений показателей сезонности не равна 4, проведем их корректировку по формуле (6.8). Скорректированный коэффициент сезонности для первого квартала

$$\bar{K}_{S1} = 0,608 \cdot \frac{4}{0,608 + 0,751 + 0,818 + 1,820} = 0,608 \text{ и т. д.}$$

Поскольку сумма непамного отличается от 4, то и скорректированные показатели почти равны исходным. Рассчитанные средние показатели сезонности и их скорректированные значения представлены в табл. 6.5.

Таблица 6.5

Оценивание сезонной составляющей в мультипликативной модели

Квартал	K_{S_i}	\bar{K}_{S_i}
I	0,608	0,608
II	0,751	0,752
III	0,818	0,819
IV	1,820	1,821
Итого	3,997	4,000

5. Далее каждый уровень исходного ряда делим на соответствующее значение сезонной компоненты. Тем самым получаем величины, которые содержат только тенденцию и случайную компоненту (см. графа 6 табл. 6.4).

6. По ряду значений $y_t(\text{скор})$, полученных в п. 5, проводим аналитическое выравнивание. В результате получаем уравнение параболы:

$$\hat{y}_t(\text{скор}) = 7,231 + 0,029t + 0,0068t^2 \quad (R^2 = 0,94465).$$

В графе 7 табл. 6.4 приведены расчетные значения по этой модели.

7. Таким образом, расчетные (графа 8 табл. 6.4) и прогнозные значения по мультипликативной тренд-сезонной модели будут определяться из следующих выражений:

для I квартала $y_t^* = 0,608 \cdot (7,231 + 0,029t + 0,0068t^2)$;

для II квартала $y_t^* = 0,752 \cdot (7,231 + 0,029t + 0,0068t^2)$;

для III квартала $y_t^* = 0,819 \cdot (7,231 + 0,029t + 0,0068t^2)$;

для IV квартала $y_t^* = 1,821 \cdot (7,231 + 0,029t + 0,0068t^2)$.

На рис. 6.5 представлены фактические и расчетные по мультипликативной тренд-сезонной модели значения ввода жилых домов по РФ.

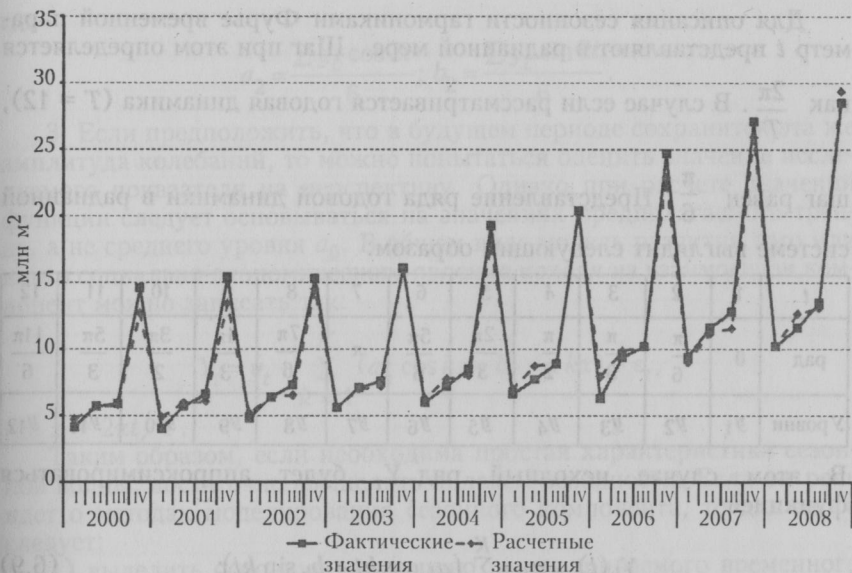


Рис. 6.5. Фактические и расчетные значения ввода в эксплуатацию жилых домов в РФ

6.4. МОДЕЛИРОВАНИЕ СЕЗОННЫХ КОЛЕБАНИЙ ПРИ ПОМОЩИ ГАРМОНИК ФУРЬЕ

При исследовании и для описания периодических явлений целесообразно использовать модели сезонной волны, которые строятся при помощи гармонического анализа. Расчет для построения модели сезонной волны можно провести по следующей схеме [4; 17; 18].

1. На первом этапе выделяется основная тенденция, оцениваются параметры уравнения тренда. Правильность определения тренда является обязательным условием выделения сезонной компоненты.

2. Моделирование ряда, имеющего периодическую составляющую, можно производить при помощи гармонического анализа, суть которого состоит в том, что ряд Y_t преобразовывается в новый $Y_k(t)$, где роль периодической составляющей вычлещивается особенно ярко:

$$Y_k(t) = a_0 + \sum_{k=1}^K \left(a_k \cos \frac{2\pi kt}{T} + b_k \sin \frac{2\pi kt}{T} \right),$$

где a_0, a_k, b_k — неизвестные параметры Фурье, которые находятся по методу наименьших квадратов;

k — гармоника ряда, которая чаще всего берется целым числом от 1 до 4 ($k < T/2$);

T — период колебаний.

Для описания сезонности гармониками Фурье временной параметре t представляют в радианной мере. Шаг при этом определяется как $\frac{2\pi}{T}$. В случае если рассматривается годовая динамика ($T = 12$),

шаг равен $\frac{\pi}{6}$. Представление ряда годовой динамики в радианной системе выглядит следующим образом:

t	1	2	3	4	5	6	7	8	9	10	11	12
рад	0	$\frac{\pi}{6}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{5\pi}{6}$	π	$\frac{7\pi}{6}$	$\frac{4\pi}{3}$	$\frac{3\pi}{2}$	$\frac{5\pi}{3}$	$\frac{11\pi}{6}$
Уровни	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}

В этом случае исходный ряд Y_t будет аппроксимироваться функцией

$$Y_k(t) = a_0 + \sum_{k=1}^K (a_k \cos kt + b_k \sin kt); \quad (6.9)$$

коэффициенты Фурье a_0, a_k, b_k определяются при помощи МНК из следующих соотношений:

$$a_k = \frac{2}{n} \sum_{t=0}^{\frac{11\pi}{6}} y_t \cos kt; \quad b_k = \frac{2}{n} \sum_{t=0}^{\frac{11\pi}{6}} y_t \sin kt; \quad a_0 = \frac{\sum y_t}{n}.$$

Наиболее подходящей считается такая функция $Y_k(t)$, при которой средняя квадратическая ошибка

$$S = \sqrt{\frac{\sum_{t=1, n} (y_t - y_k(t))^2}{T}} \quad (6.10)$$

имеет наименьшее значение.

Первая гармоника при рассмотрении ряда длиной $n = 12$ выглядит следующим образом:

$$y_1(t) = a_0 + a_1 \cos t + b_1 \sin t; \quad (6.11)$$

$$a_0 = \frac{\sum y_t}{12}; \quad a_1 = \frac{\sum y_t \cos t}{6}; \quad b_1 = \frac{\sum y_t \sin t}{6}. \quad (6.12)$$

Ряд Фурье с двумя гармониками имеет вид

$$y_2(t) = a_0 + a_1 \cos t + a_2 \cos 2t + b_1 \sin t + b_2 \sin 2t,$$

где

$$a_2 = \frac{\sum y_t \cos 2t}{6}; b_2 = \frac{\sum y_t \sin 2t}{6}.$$

3. Если предположить, что в будущем периоде сохранится эта же амплитуда колебаний, то можно попытаться оценить значение исследуемого показателя на перспективу. Однако при расчете значений функции следует основываться на значениях предполагаемого тренда, а не среднего уровня a_0 . В общем виде модель развития того или иного социально-экономического явления исходя из взаимосвязи компонент можно записать так:

$$Y_t = \hat{y}_t + \sum_{k=1}^K (a_k \cos kx + b_k \sin kx) + \varepsilon_t,$$

где $x = 2\pi t/T$.

Таким образом, если необходима простая характеристика сезонной волны, достаточно вычислить индексы сезонности. Если же речь идет о методах моделирования сезонного компонента, то для этого следует:

1) выделить основную тенденцию рассматриваемого временного ряда при помощи аналитического выравнивания или простой скользящей средней;

2) оценить сезонные колебания, используя те или иные показатели сезонностей, либо описать эти колебания гармониками Фурье;

3) оценить адекватность полученной прогнозной модели исследуемому явлению (для этого нужно проанализировать отклонения рассматриваемого временного ряда от теоретических значений, полученных по той или иной модели).

Проиллюстрируем расчет первой гармоники Фурье для нашего примера. Поскольку используются квартальные данные, то для представления ряда годовой динамики в радианной системе выбираем

шаг $\frac{\pi}{2}$. Результаты расчетов представлены в табл. 6.6. В исследуемом ряду динамики присутствует тенденция развития, поэтому модель примет вид

$$y_1(t) = \hat{y}_t + a_1 \cos t + b_1 \sin t.$$

В качестве \hat{y}_t примем модель параболы $\hat{y}_t = 7,25 - 0,038t + 0,01t^2$.

Коэффициенты Фурье для первой гармоники определялись по формулам (6.12):

$$a_1 = \frac{2}{36} \cdot (-24,3) = -1,35 \text{ и } b_1 = \frac{2}{36} \cdot (-109) = 6,06.$$

Таким образом, функция прогнозирования примет следующий вид:

$$y_1(t) = 7,25 - 0,038t + 0,01t^2 - 1,35 \cos t - 6,06 \sin t.$$

Результаты расчетов по гармонике Фурье первого порядка

Год	Квар- тал	рад	y_t	$\cos t$	$\sin t$	$y_t \cos t$	$y_t \sin t$	$y_1(t)$
2000	I	0,00	4,1	1,0	0,0	4,1	0,0	5,9
	II	1,57	5,7	0,0	1,0	0,0	5,7	1,2
	III	3,14	5,9	-1,0	0,0	-5,9	0,0	8,6
	IV	4,71	14,6	0,0	-1,0	0,0	-14,6	13,3
2001	I	6,28	4,0	1,0	0,0	4,0	0,0	5,9
	II	7,85	5,7	0,0	1,0	0,0	5,7	1,3
	III	9,42	6,6	-1,0	0,0	-6,6	0,0	8,8
	IV	11,00	15,4	0,0	-1,0	0,0	-15,4	13,6
2002	I	12,57	4,8	1,0	0,0	4,8	0,0	6,3
	II	14,14	6,4	0,0	1,0	0,0	6,4	1,8
	III	15,71	7,3	-1,0	0,0	-7,3	0,0	9,3
	IV	17,28	15,3	0,0	-1,0	0,0	-15,3	14,2
2003	I	18,85	5,6	1,0	0,0	5,6	0,0	7,0
	II	20,42	7,1	0,0	1,0	0,0	7,1	2,6
	III	21,99	7,6	-1,0	0,0	-7,6	0,0	10,2
	IV	23,56	16,1	0,0	-1,0	0,0	-16,1	15,2
2004	I	25,13	6,0	1,0	0,0	6,0	0,0	8,0
	II	26,70	7,2	0,0	1,0	0,0	7,2	3,6
	III	28,27	8,5	-1,0	0,0	-8,5	0,0	11,4
	IV	29,85	19,3	0,0	-1,0	0,0	-19,3	16,4
2005	I	31,42	6,6	1,0	0,0	6,6	0,0	9,4
	II	32,99	7,7	0,0	1,0	0,0	7,7	5,0
	III	34,56	8,9	-1,0	0,0	-8,9	0,0	12,8
	IV	36,13	20,4	0,0	-1,0	0,0	-20,4	18,0
2006	I	37,70	6,3	1,0	0,0	6,3	0,0	11,0
	II	39,27	9,4	0,0	1,0	0,0	9,4	6,7
	III	40,84	10,2	-1,0	0,0	-10,2	0,0	14,6
	IV	42,41	24,7	0,0	-1,0	0,0	-24,7	19,8
2007	I	43,98	9,4	1,0	0,0	9,4	0,0	12,9
	II	45,55	11,6	0,0	1,0	0,0	11,6	8,7
	III	47,12	12,8	-1,0	0,0	-12,8	0,0	16,7
	IV	48,69	27,1	0,0	-1,0	0,0	-27,1	22,0
2008	I	50,27	10,2	1,0	0,0	10,2	0,0	15,2
	II	51,84	11,6	0,0	1,0	0,0	11,6	11,1
	III	53,41	13,5	-1,0	0,0	-13,5	0,0	19,1
	IV	54,98	28,5	0,0	-1,0	0,0	-28,5	24,4
Итого		-	392,1	-	-	-24,3	-109,0	392,1

Расчетные значения получают путем подстановки соответствующего значения t в эту функцию:

$$y_1(1) = 7,25 - 0,038 \cdot 1 + 0,01 \cdot 1^2 - 1,35 \cos 0 - 6,06 \sin 0 = 5,9;$$

$$y_2(1) = 7,25 - 0,038 \cdot 2 + 0,01 \cdot 2^2 - 1,35 \cos \frac{\pi}{2} - 6,06 \sin \frac{\pi}{2} = 1,2;$$

$$y_1(36) = 7,25 - 0,038 \cdot 36 + 0,01 \cdot 36^2 - 1,35 \cos \frac{35\pi}{2} - 6,06 \sin \frac{35\pi}{2} = 24,4.$$

Оценив параметры гармоник Фурье различных порядков ($k = 1, 2, 3, 4$) и рассчитав среднеквадратические ошибки соответствующих моделей, выбираем наилучшую гармонику.

Результаты расчетов для гармоник первого порядка иллюстрирует рис. 6.6.

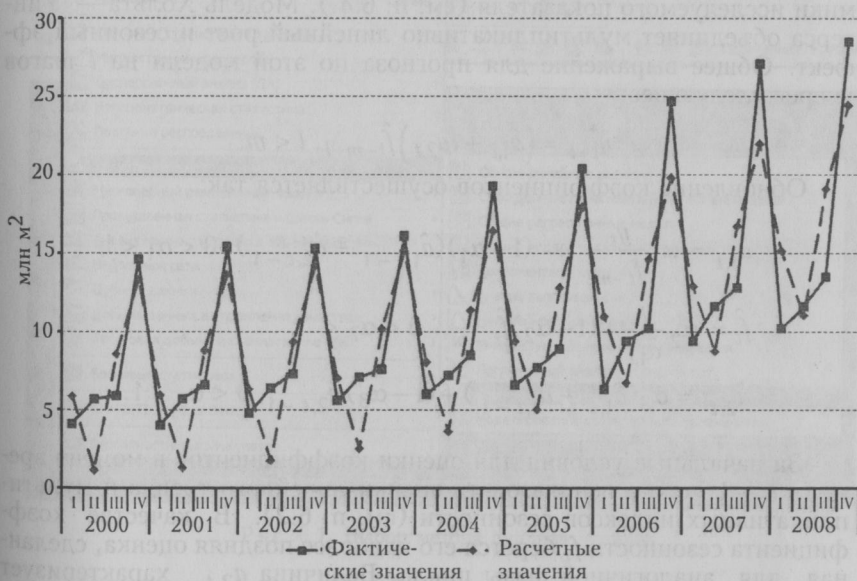


Рис. 6.6. Динамика ввода в действие жилых домов в РФ и гармоника Фурье первого порядка

6.5. АДАПТИВНЫЕ МОДЕЛИ СЕЗОННЫХ КОЛЕБАНИЙ

Временные ряды, содержащие периодические сезонные колебания, могут быть описаны при помощи адаптивных моделей с мультипликативными

$$y_t = a_{1,t} f_t + \varepsilon_t \quad (6.13)$$

и аддитивными

$$y_t = a_{1,t} + g_t + \varepsilon_t \quad (6.14)$$

коэффициентами сезонности.

В формулах (6.13) и (6.14) $a_{1,t}$ — величина, характеризующая тенденцию развития явления;

$f_t, f_{t-1}, f_{t-2}, \dots, f_{t-m+1}$ — мультипликативные коэффициенты сезонности;

$g_t, g_{t-1}, g_{t-2}, \dots, g_{t-m}$ — аддитивные коэффициенты сезонности; m — количество фаз в полном сезонном цикле ($m = 12$ или $m = 4$);

ε_t — случайная компонента с нулевым математическим ожиданием.

Выбор той или иной модели будет продиктован характером динамики исследуемого показателя (см. п. 6.4.). Модель Хольта — Уинтерса объединяет мультипликативно линейный рост и сезонный эффект. Общее выражение для прогноза по этой модели на l шагов вперед следующее:

$$y_{t+l}^* = (\hat{a}_{1,t} + l\hat{a}_{2,t}) \hat{f}_{t-m+l}, \quad l < m.$$

Обновление коэффициентов осуществляется так:

$$\hat{a}_{1,t} = \alpha_1 \frac{y_t}{\hat{f}_{t-m}} + (1 - \alpha_1)(\hat{a}_{1,t-1} + \hat{a}_{2,t-1}), \quad 0 < \alpha_1 < 1;$$

$$\hat{f}_t = \alpha_2 \frac{y_t}{\hat{a}_{1,t}} + (1 - \alpha_2) \hat{f}_{t-m}, \quad 0 < \alpha_2 < 1;$$

$$\hat{a}_{2,t} = \alpha_3 (\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \alpha_3) \hat{a}_{2,t-1}, \quad 0 < \alpha_3 < 1.$$

За начальные условия для оценки коэффициентов в момент времени $t = 1$ можно использовать оценки линейного тренда и мультипликативных индексов сезонности (см. п. 6.2). В качестве коэффициента сезонности f_t берется его наиболее поздняя оценка, сделанная для аналогичной фазы цикла. Величина $a_{2,t}$ характеризует прирост среднего уровня от момента $t - 1$ к моменту t . Если прогноз осуществляется для временного ряда, не имеющего тенденции ($Y_t = S_t + E_t$), то из уравнения прогноза и формул исключается оценка $\hat{a}_{2,t}$.

Прогноз по модели Тейла — Вейджа определяется выражением

$$y_{t+l}^* = \hat{a}_{1,t} + l\hat{a}_{2,t} + \hat{g}_{t-m+l}, \quad l < m.$$

Коэффициенты обновляются следующим образом:

$$\hat{a}_{1,t} = \alpha_1 (y_t - \hat{g}_{t-m}) + (1 - \alpha_1)(\hat{a}_{1,t-1} + \hat{a}_{2,t-1});$$

$$\hat{g}_t = \alpha_2(y_t - \hat{a}_{1,t}) + (1 - \alpha_2)\hat{g}_{t-m};$$

$$\hat{a}_{2,t} = \alpha_3(\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \alpha_3)\hat{a}_{2,t-1}.$$

Оптимальные значения параметров адаптации можно определять экспериментальным путем, взяв за критерий минимум стандартного отклонения ошибки [23].

Адаптивные сезонные модели являются важной составной частью современных пакетов прикладных программ (ППП), ориентированных на решение задач прогнозирования. Алгоритм прогнозирования динамики ввода жилья реализован в ППП «Statistica». После запуска программы выбираем модуль «Временные ряды и прогнозирование» (рис. 6.7).

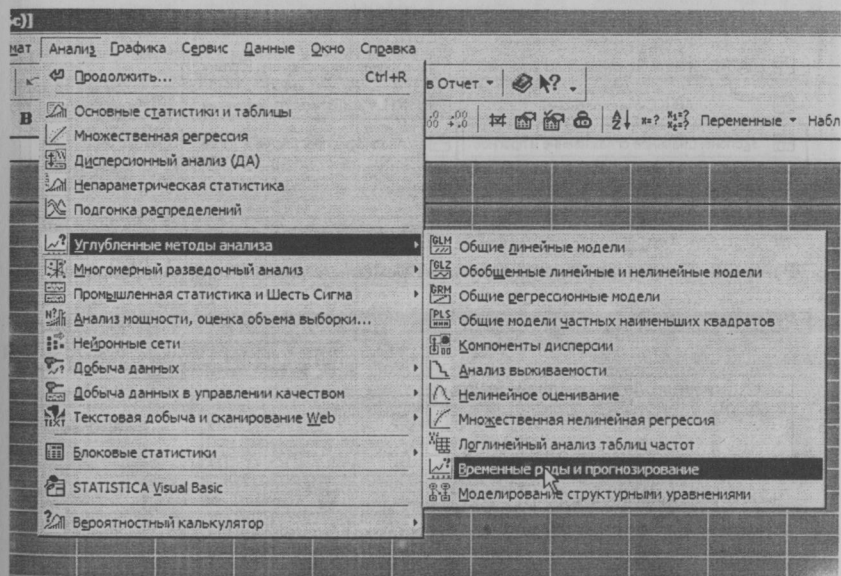


Рис. 6.7. Выбор метода анализа данных

Открыв стартовую панель данного модуля, выберем метод экспоненциального сглаживания (рис. 6.8).

Нажав на соответствующую кнопку, получим варианты выбора моделей экспоненциального сглаживания, реализованные в системе (рис. 6.9). С учетом того, что исходный ряд динамики содержит ярко выраженную квартальную сезонность, зададим значение сезонного лага, равное 4, а также выберем модель линейного тренда с мультипликативной сезонностью.

Значения параметров сглаживания задать в этом же окне можно самостоятельно или при помощи вкладки «Поиск на сетке», либо во

вкладке «Автоматический поиск» подобрать оптимальные параметры на основе одного из критериев точности модели (рис. 6.10).

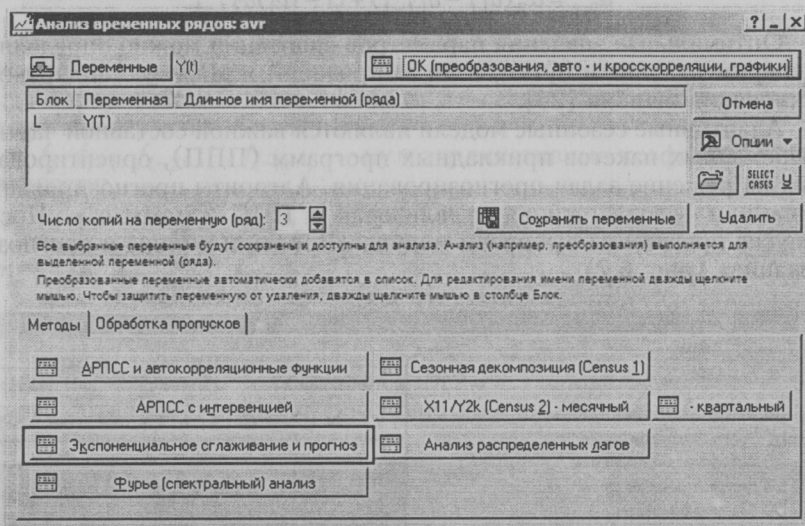


Рис. 6.8. Стартовая панель модуля «Временные ряды и прогнозирование»

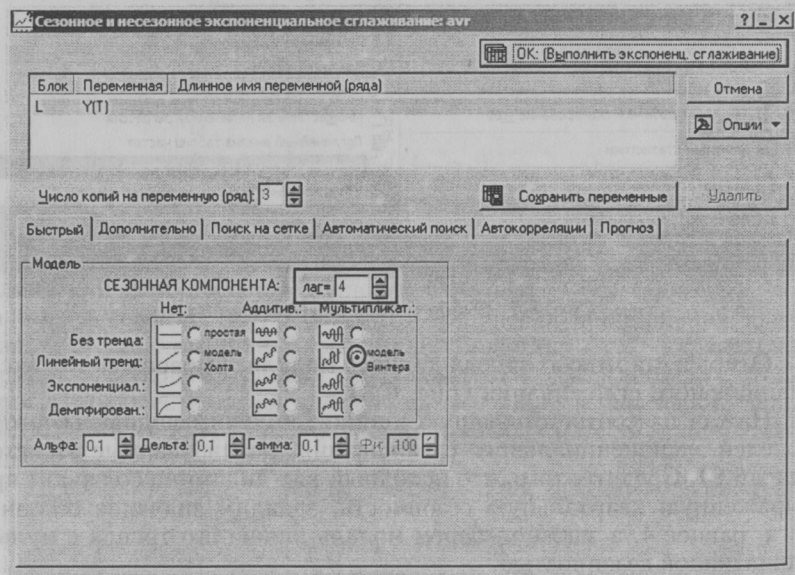


Рис. 6.9. Выбор определенной модели экспоненциального сглаживания

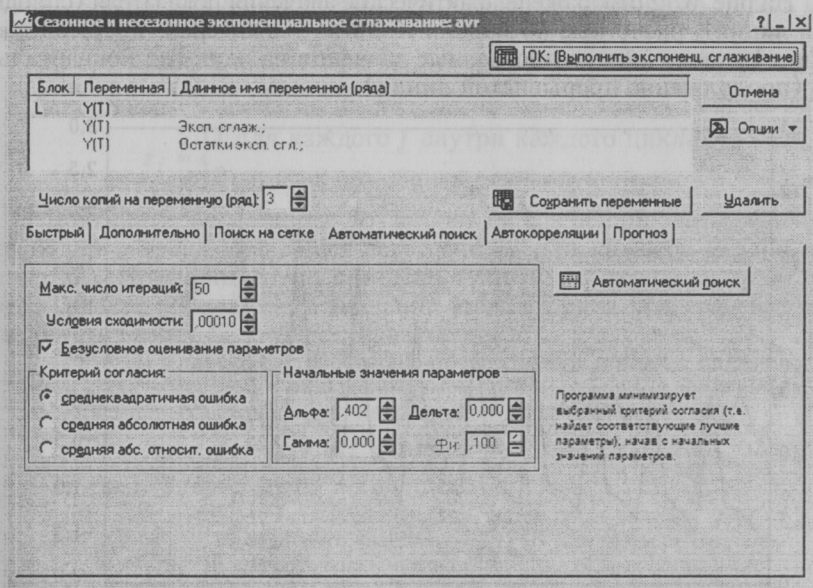


Рис. 6.10. Окно автоматического подбора параметров сглаживания модели

В результате мы получим числовые и графические результаты проведенных расчетов (рис. 6.11 и 6.12).

Эксп. сглажив.: Мульти. сезон. (4) S0=7,052 T0= 2617 (авт)				
Лин. тренд, мульти. сезон.: Альфа= .402 Дельта=0,00 Гамма=0,00				
Y(T)				
Набл.	Y(T)	Сглажен. ряд	Остатки	Сезонные составл.
1	4,10000	4,47188	-0,37188	61,1474
2	5,70000	5,54253	0,15747	75,6089
3	5,90000	6,29457	-0,39457	82,0037
4	14,60000	14,03568	0,56432	181,2400
5	4,00000	4,97198	-0,97198	
6	5,70000	5,86261	-0,16261	
7	6,60000	6,50217	0,09783	
8	15,40000	14,93199	0,46801	
9	4,80000	5,26131	-0,46131	
10	6,40000	6,47421	-0,07421	
11	7,30000	7,20404	0,09596	
12	15,30000	16,48157	-1,18157	
13	5,60000	5,56039	0,03961	
14	7,10000	7,09301	0,00699	
15	7,60000	7,91058	-0,31058	
16	16,10000	17,68192	-1,58192	

Рис. 6.11. Табличные результаты сезонного экспоненциального сглаживания

На рис. 6.12 отображены фактические значения показателя (сплошная линия), расчетные по модели Хольта — Уинтерса (прерывистая линия), также даются прогнозные значения на один год поквартально (продолжение прерывистой линии).

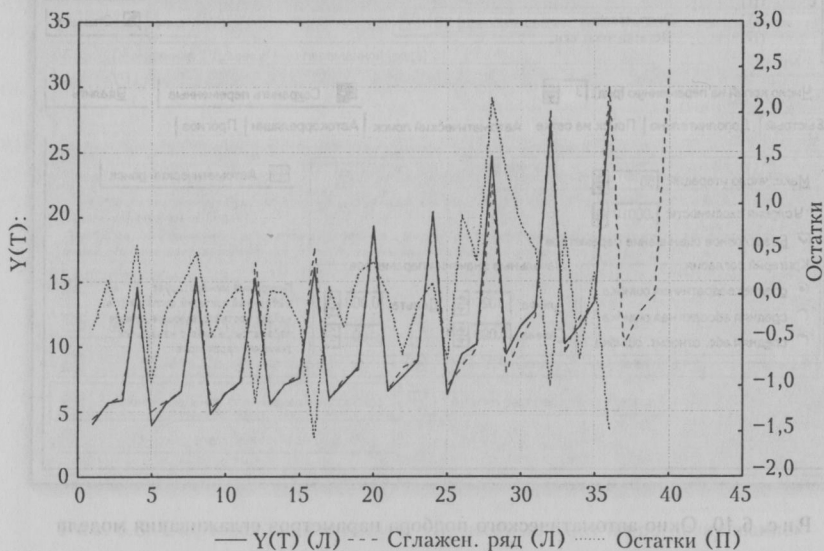


Рис. 6.12. Графические результаты сезонного экспоненциального сглаживания

По ряду остатков можно проверить адекватность построенной модели реальному процессу, задав опцию «Автокорреляции». Значения верхней и нижней границ доверительного интервала необходимо определить самостоятельно, используя формулу (4.4).

6.6. ФИКТИВНЫЕ ПЕРЕМЕННЫЕ В МОДЕЛИРОВАНИИ СЕЗОННЫХ КОЛЕБАНИЙ

При моделировании временных рядов с сезонной составляющей можно использовать модель регрессии с фиктивными переменными. Количество фиктивных переменных в модели регрессии должно быть на единицу меньше, чем число сезонов (т. е. равно трем, если данные поквартальные, или одиннадцати, если дана помесечная динамика). Каждая фиктивная переменная отражает определенный сезон внутри годового цикла и равна единице для этого периода и нулю — для остальных.

Модель регрессии с фиктивными переменными для ряда, имеющего основную тенденцию, выглядит следующим образом:

$$\tilde{y}_t = \hat{y}_t + c_1 z_1 + c_2 z_2 + \dots + c_{k-1} z_{k-1} + \varepsilon_t, \quad (6.12)$$

где \hat{y}_t — функция линии тренда, например для линейного тренда $\hat{y}_t = \hat{a}_0 + a_1t$;

a_j, c_j — коэффициенты регрессии, определяемые по МНК;

z_j — фиктивные переменные,

$$z_j = \begin{cases} 1 & \text{для каждого } j \text{ внутри каждого цикла;} \\ 0 & \text{во всех остальных случаях.} \end{cases}$$

Фиктивные переменные позволяют дифференцировать величину свободного члена уравнения регрессии a_0 для каждого месяца или квартала. Коэффициент a_1 в модели с линейным трендом характеризует среднее абсолютное изменение уровней ряда под воздействием систематически действующих факторов [36].

Пример 6.4. Построим прогноз ввода жилья с помощью модели регрессии с включенным фактором времени и фиктивных переменных. Модель для квартальной динамики имеет вид

$$y_t = a_0 + a_1t + a_2t^2 + c_2z_2 + c_3z_3 + c_4z_4 + \varepsilon_t,$$

где

$$z_2 = \begin{cases} 1 & \text{для второго квартала,} \\ 0 & \text{для всех остальных;} \end{cases} \quad z_3 = \begin{cases} 1 & \text{для третьего квартала,} \\ 0 & \text{для всех остальных;} \end{cases}$$

$$z_4 = \begin{cases} 1 & \text{для четвертого квартала,} \\ 0 & \text{для всех остальных.} \end{cases}$$

В табл. 6.7 занесены значения фактора времени и фиктивных переменных.

Параметры регрессионного уравнения можно оценить с помощью МНК, используя табличный редактор «MS Excel». Результаты расчетов приведены в табл. 6.8.

Уравнение регрессии имеет следующий вид:

$$\tilde{y}_t = 3,88 - 0,07t + 0,01t^2 + 1,46z_2 + 2,17z_3 + 13,0z_4.$$

Коэффициенты c_2, c_3, c_4 характеризуют отклонения уровней временного ряда от уровней, учитывающих сезонные воздействия в I квартале.

Уравнение регрессии для каждого квартала будет выглядеть так:

$$\text{для I квартала} \quad \tilde{y}_t = 3,88 - 0,07t + 0,01t^2;$$

$$\text{для II квартала} \quad \tilde{y}_t = 5,34 - 0,07t + 0,01t^2;$$

$$\text{для III квартала} \quad \tilde{y}_t = 6,05 - 0,07t + 0,01t^2;$$

$$\text{для IV квартала} \quad \tilde{y}_t = 16,88 - 0,07t + 0,01t^2.$$

Чтобы получить прогнозные поквартальные значения показателя ввода жилой площади на следующий год, необходимо в соответствующее уравнение регрессии подставить следующее значение фактора времени t . Так, прогноз на I квартал 2009 г. составит $\tilde{y}_t = 3,88 - 0,07 \cdot 37 + 0,01 \cdot 37^2 = 14,98$, а на II — $\tilde{y}_t = 5,34 - 0,07 \cdot 38 + 0,01 \cdot 38^2 = 17,12$ и т. д.

Таблица 6.7

Исходные данные для расчета параметров уравнения регрессии с фиктивными переменными по временному ряду ввода жилья

Год	Квартал	y_t	t	t^2	z_2	z_3	z_4	Расчетные значения по модели
2000	I	4,1	1	1	0	0	0	3,82
	II	5,7	2	4	1	0	0	5,24
	III	5,9	3	9	0	1	0	5,93
	IV	14,6	4	16	0	0	1	16,76
2001	I	4	5	25	0	0	0	3,78
	II	5,7	6	36	1	0	0	5,28
	III	6,6	7	49	0	1	0	6,05
	IV	15,4	8	64	0	0	1	16,96
2002	I	4,8	9	81	0	0	0	4,06
	II	6,4	10	100	1	0	0	5,64
	III	7,3	11	121	0	1	0	6,49
	IV	15,3	12	144	0	0	1	17,48
2003	I	5,6	13	169	0	0	0	4,66
	II	7,1	14	196	1	0	0	6,32
	III	7,6	15	225	0	1	0	7,25
	IV	16,1	16	256	0	0	1	18,32
2004	I	6,0	17	289	0	0	0	5,58
	II	7,2	18	324	1	0	0	7,32
	III	8,5	19	361	0	1	0	8,33
	IV	19,3	20	400	0	0	1	19,48
2005	I	6,6	21	441	0	0	0	6,82
	II	7,7	22	484	1	0	0	8,64
	III	8,9	23	529	0	1	0	9,73
	IV	20,4	24	576	0	0	1	20,96
2006	I	6,3	25	625	0	0	0	8,38
	II	9,4	26	676	1	0	0	10,28
	III	10,2	27	729	0	1	0	11,45
	IV	24,7	28	784	0	0	1	22,76
2007	I	9,4	29	841	0	0	0	10,26
	II	11,6	30	900	1	0	0	12,24
	III	12,8	31	961	0	1	0	13,49
	IV	27,1	32	1024	0	0	1	24,88
2008	I	10,2	33	1089	0	0	0	12,46
	II	11,6	34	1156	1	0	0	14,52
	III	13,5	35	1225	0	1	0	15,85
	IV	28,5	36	1296	0	0	1	27,32

Результаты оценивания регрессионной модели с фиктивными переменными

Регрессионная статистика						
Множественный коэффициент корреляции R				0,981		
Коэффициент детерминации R^2				0,963		
F-критерий Фишера $F(5,30)$				155,169		
Стандартная ошибка				1,334		
Параметры уравнения регрессии						
	Бета-коэффициент	Стандартная ошибка бета-коэффициента	Оценка коэффициента	Стандартная ошибка	t -статистика	p -значение
Пересечение c y	—	—	3,88	0,788	4,92	0,000
t	-0,123	0,145	-0,07	0,088	-0,85	0,404
t^2	0,590	0,145	0,01	0,002	4,08	0,000
z_2	0,100	0,043	1,46	0,629	2,32	0,028
z_3	0,149	0,043	2,17	0,630	3,45	0,002
z_4	0,892	0,043	13,00	0,632	20,57	0,000

На рис. 6.13 представлены фактические данные о вводе жилой площади и полученные по уравнению регрессии с фиктивными переменными.

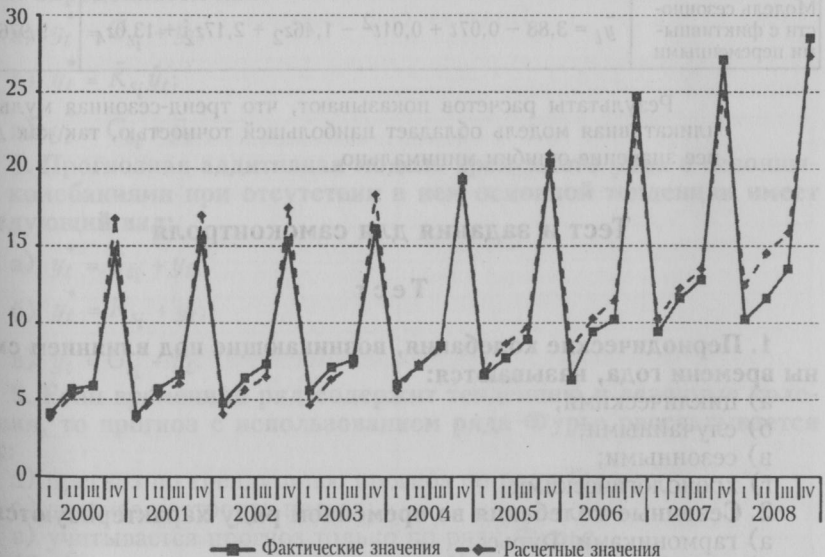


Рис. 6.13. Фактические и расчетные по модели регрессии с фиктивными переменными значения ввода в эксплуатацию жилых домов в РФ

Обобщив результаты расчетов по рассматриваемым моделям, можно сравнить точность полученных результатов. Для этого была рассчитана средняя ошибка аппроксимации для каждой из моделей. Сводные результаты представлены в табл. 6.9.

Таблица 6.9

Оценка точности и адекватности полученных моделей

Модель	Вид модели	Ошибка аппроксимации, %
Мультипликативная индексная модель	I кв.: $y_t^* = 0,6048 \cdot (7,25 - 0,038t + 0,01t^2)$; II кв.: $y_t^* = 0,7585 \cdot (7,25 - 0,038t + 0,01t^2)$; III кв.: $y_t^* = 0,8273 \cdot (7,25 - 0,038t + 0,01t^2)$; IV кв.: $y_t^* = 0,8094 \cdot (7,25 - 0,038t + 0,01t^2)$	6,179
Тренд-сезонная мультипликативная модель	I кв.: $y_t^* = 0,608 \cdot (7,231 - 0,029t + 0,0068t^2)$; II кв.: $y_t^* = 0,752 \cdot (7,231 - 0,029t + 0,0068t^2)$; III кв.: $y_t^* = 0,819 \cdot (7,231 - 0,029t + 0,0068t^2)$; IV кв.: $y_t^* = 1,821 \cdot (7,231 - 0,029t + 0,0068t^2)$	4,927
Гармоника Фурье первого порядка	$y_1(t) = 7,25 - 0,038t + 0,01t^2 - 1,35 \cos t - 6,06 \sin t$	15,158
Модель сезонности с фиктивными переменными	$\tilde{y}_t = 3,88 - 0,07t + 0,01t^2 - 1,46z_2 + 2,17z_2 + 13,0z_4$	9,976

Результаты расчетов показывают, что тренд-сезонная мультипликативная модель обладает наибольшей точностью, так как для нее значение ошибки минимально.

Тест и задания для самоконтроля

Тест

1. Периодические колебания, возникающие под влиянием смены времени года, называются:

- циклическими;
- случайными;
- сезонными;
- хронологическими.

2. Сезонные колебания во временном ряду характеризуются:

- гармониками Фурье;
- коэффициентами автокорреляции;
- индексами сезонности.

3. Индексы сезонности для временного ряда с ярко выраженной тенденцией определяются:

- а) по способу постоянной средней;
- б) по способу переменной средней;
- в) по методу наименьших квадратов.

4. Индексы сезонности (в относительном выражении) характеризуют:

- а) изменение уровня ряда в отчетном периоде относительно уровня в базисном периоде;
- б) изменение значения показателя за определенный квартал (месяц) относительно значения показателя в начале года;
- в) среднее значение уровня в определенном квартале (месяце) относительно среднегодового.

5. Величина среднего индекса сезонности при расчете индексов сезонности в абсолютном выражении равна:

- а) 1;
- б) 0;
- в) 100 %.

6. Для расчета скользящей средней во временных рядах с сезонной составляющей используется формула:

- а) средней арифметической;
- б) средней хронологической;
- в) средней геометрической.

7. Точечный прогноз по мультипликативной тренд-сезонной модели определяется как:

- а) $y_t^* = \tilde{G}_{s_i} + \hat{y}_t$;
- б) $y_t^* = \tilde{K}_{s_i} \hat{y}_t$;
- в) $y_t^* = \tilde{G}_{s_i} + \bar{y}_t$.

8. Прогнозная аддитивная модель временного ряда с сезонными колебаниями при отсутствии в нем основной тенденции имеет следующий вид:

- а) $y_t^* = \tilde{G}_{s_i} + y_t$;
- б) $y_t^* = \tilde{K}_{s_i} + y_t$;
- в) $y_t^* = \tilde{G}_{s_i} + \bar{y}_t$.

9. Если временной ряд содержит тенденцию и сезонные колебания, то прогноз с использованием ряда Фурье рассчитывается так:

- а) прогноз по тренду плюс прогноз по ряду Фурье;
- б) прогноз построить нельзя;
- в) учитывается прогноз только по ряду Фурье.

10. Адаптивная модель Хольта — Уинтерса применяется в анализе временных рядов, имеющих компоненты:

а) F_t, E_t ;

б) E_t ;

в) F_t, S_t, E_t .

11. Прогноз по модели Хольта — Уинтерса на l шагов вперед осуществляется по следующей формуле:

а) $y_t = a_{1,t}f_t + \varepsilon_t$;

б) $y_{t+l}^* = (\hat{a}_{1,t} + l\hat{a}_{2,t})\hat{f}_{t-m+l}$;

в) $y_{t+l}^* = \hat{a}_{1,t} + l\hat{a}_{2,t} + \hat{g}_{t-m+l}$.

12. Если временной ряд не имеет тенденции ($Y_t = S_t + E_t$), то формула прогноза модели Тейла — Вейджа будет выглядеть так:

а) $y_{t+l}^* = \hat{a}_{1,t} + \hat{g}_{t-m+l}$;

б) $y_t = a_{1,t} + g_t + \varepsilon_t$;

в) $y_{t+l}^* = (\hat{a}_{1,t} + l\hat{a}_{2,t})\hat{f}_{t-m+l}$.

13. Коэффициенты c_2, c_3, c_4 в модели регрессии с фиктивными переменными характеризуют:

а) отклонения уровней временного ряда от уровней, учитывающих сезонные воздействия в I квартале;

б) отклонения уровней временного ряда от уровней в предшествующие кварталы;

в) отклонения уровней от среднегодового уровня ряда.

Практические задания

1. По грузовому автотранспортному предприятию имеются следующие данные об объемах перевозок.

Месяц	Среднесуточный объем перевозок, тыс. т		
	2010 г.	2011 г.	2012 г.
Январь	10,2	10,7	10,3
Февраль	10,4	10,4	10,6
Март	10,6	10,8	10,9
Апрель	11,0	11,1	11,3
Май	11,3	11,2	11,2
Июнь	11,5	11,0	11,7
Июль	11,6	11,3	11,8
Август	12,0	11,7	12,4
Сентябрь	11,2	11,6	11,7
Октябрь	10,9	10,7	11,2
Ноябрь	10,2	10,4	10,8
Декабрь	10,0	10,3	10,5

Выявите наличие сезонной неравномерности. Определите индексы сезонности. Сделайте вывод.

2. Имеются данные о производстве угля, тыс. т, в некотором регионе за 2009–2012 гг.

Квартал	2009 г.	2010 г.	2011 г.	2012 г.
I	92,6	83,8	76,4	71,4
II	83,1	75,3	65,6	60,1
III	76,6	68,2	59,2	61,6
IV	84,8	78,4	70,1	69,8

Определите мультипликативные индексы сезонности, если известно, что тенденция описывается параболической функцией

$$\hat{y}_t = 92,803 - 3,42t + 0,105t^2.$$

Сделайте выводы. Постройте сезонную волну.

3. На основе помесечных данных о числе браков в регионе за три года была построена аддитивная сезонная модель. Скорректированные значения сезонной компоненты за соответствующие месяцы представлены в таблице.

Месяц	Значение сезонной компоненты	Месяц	Значение сезонной компоненты
Январь	-1,0	Июль	3,0
Февраль	2,0	Август	1,0
Март	-0,5	Сентябрь	2,5
Апрель	0,3	Октябрь	1,0
Май	-2,0	Ноябрь	-3,0
Июнь	-1,1	Декабрь	?

Определите значение сезонной компоненты за декабрь.

4. Используя значения индексов сезонности, полученные в задании № 2, постройте мультипликативную индексную модель прогноза. Дайте прогноз на следующий год. Изобразите на графике фактические, рассчитанные по мультипликативной модели и прогнозные значения временного ряда. Определите доверительный интервал прогноза.

5. На основании значений сезонной компоненты, представленных в задании № 3, дайте прогноз общего числа браков, заключенных в течение первого квартала следующего года, если известно, что уравнение тренда выглядит следующим образом ($t = 1, 2, \dots, 36$):

$$\hat{y}_t = 2,5 + 0,03t.$$

7. МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ С ПОМОЩЬЮ ПРОИНТЕГРИРОВАННОЙ МОДЕЛИ АВТОРЕГРЕССИИ И СКОЛЬЗЯЩЕГО СРЕДНЕГО (ARIMA-МОДЕЛИ)

1. Модели стационарных временных рядов.
2. ARIMA-модели для описания нестационарных временных рядов.
3. ARIMA-модели для описания временных рядов с сезонной составляющей.
4. Прогнозирование на основе ARIMA-моделей.

Временные ряды, характеризующие экономические явления, как правило, имеют достаточно сложную структуру. Поэтому моделирование их путем построения и объединения моделей тренда, сезонности не всегда приводит к удовлетворительным результатам. Повышение точности прогноза по временным рядам связано с использованием класса моделей, объединяющих в себе тренд, сезонные колебания, зависимость остатков (автокорреляцию, авторегрессию). Это класс ARIMA-моделей (Auto Regressive Integrated Moving Average). В русскоязычной литературе также используются аббревиатуры АРПСС и АРИСС. Объединение различных методов в рамках одной модели позволяет существенно расширить область их применения. Модели этого класса могут хорошо описывать как стационарные, так и нестационарные временные ряды [8; 21; 30; 36].

Практическое использование ARIMA-моделей связывают с появлением методики их построения, разработанной Дж. Боксом и Г. Дженкинсом. В методологии Бокса — Дженкинса применяется итеративный подход к определению допустимой модели среди общего класса ARIMA. Выбранная модель сопоставляется с исходными данными для проверки того, насколько точно она описывает временной ряд. Модель считается приемлемой, если остатки в основном малы, распределены случайно и не содержат полезной информации. Если заданная модель не является удовлетворительной, то процесс повторяется, но уже с использованием новой модели. Подобная итеративная процедура продолжается до тех пор, пока не будет найдена адекватная исследуемому процессу модель. С этого момента она может использоваться для прогнозирования. Схема стратегии выбора модели по методу Бокса — Дженкинса представлена на рис. 7.1.

Если в результате проверки несколько моделей оказываются адекватными исходным данным, то при окончательном выборе следует учитывать имеющую наивысшее качество (более точную) и наименьшее число параметров.

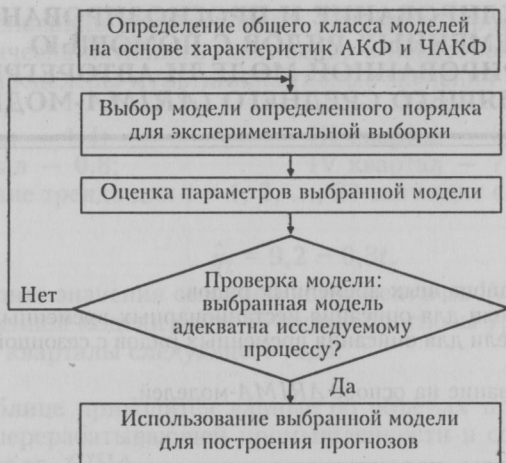


Рис. 7.1. Схема выбора модели по методу Бокса — Дженкинса

Как известно, во временных рядах могут встречаться тенденции среднего уровня, дисперсии и автокорреляции. Стационарность процесса прежде всего означает, что перечисленных тенденций в ряду нет. Под стационарным процессом ξ_t в широком смысле понимается такой, для которого математическое ожидание $M[\xi_t]$, дисперсия $D[\xi_t]$ независимо от рассматриваемого промежутка времени имеют постоянное значение, а автокорреляционная функция r_τ зависит только от длины лага и выполняются следующие соотношения:

$$M[\xi_t] = \text{const} = a;$$

$$D[\xi_t] = \sigma^2;$$

$$r(t_i, t_j) = r(\tau),$$

где $\tau = t_i - t_j$.

Величина автокорреляционной функции зависит не от начала отсчета, а только от величины сдвига τ . Значение a определяет постоянный (средний) уровень, относительно которого варьируют значения временного ряда Y_t , а постоянная величина σ^2 характеризует размах этой вариации.

Экономические процессы, представленные временными рядами, чаще всего являются нестационарными. Это связано прежде всего с наличием основной тенденции (тренда) и/или периодических колебаний. При построении *ARIMA*-модели исходный ряд проверяют на стационарность, преобразовывают его в стационарный при условии нестационарности. Временные ряды могут быть приведены к стационарному виду следующим образом:

— нахождением простых и/или сезонных разностей (подход Бокса — Дженкинса);

— путем выделения тренда и/или фильтрации периодических (сезонных) колебаний.

Как правило, ряд остатков является стационарным. Свидетельством стационарности ряда выступает снижение значений автокорреляционной и частной автокорреляционной функций (ЧАКФ) при возрастании длины лага.

Рассмотрим частные случаи *ARIMA*-моделей.

7.1. МОДЕЛИ СТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ

I. Модель стационарного процесса, выражающая значение показателя y_t в виде линейной комбинации конечного числа предшествующих значений этого показателя и аддитивной случайной составляющей, называется моделью **авторегрессии** (*Auto Regressive*) или **авторегрессионной моделью**.

В общем виде авторегрессионная модель может быть представлена так:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t, \quad (7.1)$$

где a_j — коэффициенты модели, которые оцениваются методом наименьших квадратов или с помощью коэффициентов автокорреляции. Для стационарных процессов $|a_j| < 1$;

p — порядок процесса авторегрессии, определяющий количество периодов, от которого зависит текущее значение y_t ;

ε_t — случайная составляющая, имеющая нормальное распределение с нулевым математическим ожиданием («белый шум»).

Используя оператор сдвига B , смещающий ряд на определенный временной такт ($B y_t = y_{t-1}$, $B^2 y_t = y_{t-2}$, $B^3 y_t = y_{t-3}$ и т. д.), получим модель авторегрессии

$$y_t - a_1 y_{t-1} - a_2 y_{t-2} - \dots - a_p y_{t-p} = \varepsilon_t$$

в следующем виде:

$$y_t - a_1 B y_t - a_2 B^2 y_t - \dots - a_p B^p y_t = \varepsilon_t,$$

или

$$(1 - a_1 B - a_2 B^2 - \dots - a_p B^p) y_t = \varepsilon_t.$$

Выражение в скобках называется оператором авторегрессии и обозначается $\varphi_p(B)$. В неявном виде авторегрессионная модель порядка p записывается как

$$\varphi_p(B) y_t = \varepsilon_t. \quad (7.2)$$

Для оценки параметров авторегрессионной модели порядка p (далее по тексту *AR(p)*) используется метод наименьших квадратов или

метод Юла — Уокера, основанный на расчете коэффициентов автокорреляции.

Метод наименьших квадратов базируется на требовании минимизации дисперсии, и условие выглядит следующим образом:

$$\sum_{t=p+1}^n \left(y_t - \sum_{j=1}^p a_j y_{t-j} \right)^2 \rightarrow \min.$$

Это требование приводит к системе нормальных уравнений

$$\begin{cases} a_1 \sum_{t=p+1}^n y_{t-1}^2 + a_2 \sum y_{t-1} y_{t-2} + \dots + a_p \sum y_{t-1} y_{t-p} = \sum y_t y_{t-1}; \\ a_1 \sum_{t=p+1}^n y_{t-1} y_{t-2} + a_2 \sum y_{t-2}^2 + \dots + a_p \sum y_{t-2} y_{t-p} = \sum y_t y_{t-2}; \\ a_1 \sum_{t=p+1}^n y_{t-1} y_{t-p} + a_2 \sum y_{t-2} y_{t-p} + \dots + a_p \sum y_{t-p}^2 = \sum y_t y_{t-p}. \end{cases}$$

Решая систему относительно a_j , получим коэффициенты авторегрессии $AR(p)$.

Метод Юла — Уокера основан на решении следующей системы:

$$\begin{cases} r_1 = a_1 + a_2 r_1 + \dots + a_p r_{p-1}; \\ r_2 = a_1 r_1 + a_2 + \dots + a_p r_{p-2}; \\ r_p = a_1 r_{p-1} + a_2 r_{p-2} + \dots + a_p. \end{cases}$$

где r_j — коэффициенты автокорреляции различных порядков.

Определение порядка авторегрессии является одним из важнейших этапов при построении авторегрессионных моделей. Предварительная оценка на основе экономического анализа позволяет выделить те значения исследуемого показателя, которые значительно повлияли на его изменения в последующие периоды. Низкий порядок авторегрессии может дать несущественные результаты за счет того, что в модели не использована информация о предыдущем. Повышение порядка в отдельных случаях также может привести к снижению качества модели.

На практике для определения порядка авторегрессии могут использоваться значения АКФ и ЧАКФ. Для стационарных процессов АКФ затухает с ростом сдвига τ . По скорости затухания могут быть сделаны выводы о порядке модели. Они сравниваются с результатами анализа частной автокорреляционной функции. ЧАКФ для процесса

авторегрессии будет иметь ненулевые значения лишь при τ , меньших либо равных p , а начиная с $\tau = p + 1$ теоретическая ЧАКФ равна нулю. Практически это свойство является ключевым при выборе порядка модели авторегрессии.

На практике чаще всего используются два вида авторегрессионных моделей. Процесс авторегрессии 1-го порядка, или *марковский процесс*, описывается моделью

$$y_t = a_1 y_{t-1} + \varepsilon_t.$$

Единственный коэффициент модели равен коэффициенту автокорреляции 1-го порядка: $a_1 = r_1$. Условие стационарности процесса авторегрессии 1-го порядка: $-1 < a_1 < 1$.

Автокорреляционная функция является затухающей экспонентой или синусоидой. Значения ЧАКФ после первого лага равны нулю. Это свойство можно использовать при подборе модели.

Процессы авторегрессии 2-го порядка, или *процессы Юла*, описываются моделью

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t.$$

Решая систему Юла — Уокера относительно a_1 и a_2 , можно получить выражения для определения параметров:

$$a_1 = -\frac{r_1(1-r_2)}{1-r_1^2}; \quad a_2 = -\frac{r_2-r_1^2}{1-r_1^2}.$$

Условие стационарности накладывает следующие ограничения на коэффициенты a_1 и a_2 : $a_1 + a_2 < 1$; $a_2 - a_1 < 1$; $-1 < a_2 < 1$.

II. Модель скользящего среднего (Moving Average) представляет собой исследуемый показатель в виде линейной комбинации независимых случайных величин, т. е. в ошибках модели в предшествующие периоды сосредоточена информация обо всей предыстории ряда. Процесс скользящего среднего порядка q (обозначается $MA(q)$) имеет вид

$$y_t = \varepsilon_t - b_1 \varepsilon_{t-1} - b_2 \varepsilon_{t-2} - \dots - b_q \varepsilon_{t-q}.$$

При использовании оператора сдвига B модель скользящего среднего в неявном виде будет выглядеть следующим образом:

$$y_t = \psi_q(B) \varepsilon_t, \quad (7.3)$$

где $\psi_q(B)$ — оператор скользящего среднего, $\psi_q(B) = 1 - b_1 B - b_2 B^2 - \dots - b_q B^q$.

Величина q — это количество прошлых ошибок, которые включаются в прогноз последующих наблюдений. Чтобы процесс скользящего среднего был сходящимся, значения b_j по абсолютной величине должны быть меньше единицы. Автокорреляционные функции для процессов $MA(q)$ обращаются в ноль сразу после запаздывания в

q периодов, в то время как частные автокорреляционные приближаются к нулю постепенно.

III. На практике в целях описания стационарного процесса авторегрессии в модель могут быть включены остатки в виде процесса скользящего среднего. Комбинация авторегрессионной модели порядка p и модели скользящего среднего порядка q дает смешанную **авторегрессионную модель скользящего среднего** ($ARMA(p, q)$):

$$y_t - a_1 y_{t-1} - a_2 y_{t-2} - \dots - a_p y_{t-p} = \varepsilon_t - b_1 \varepsilon_{t-1} - b_2 \varepsilon_{t-2} - \dots - b_q \varepsilon_{t-q},$$

или

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t - b_1 \varepsilon_{t-1} - b_2 \varepsilon_{t-2} - \dots - b_q \varepsilon_{t-q}.$$

В неявном виде (через полиномы авторегрессии и скользящего среднего) смешанная авторегрессионная модель скользящего среднего выглядит следующим образом:

$$\varphi_p(B)y_t = \psi_q(B)\varepsilon_t. \quad (7.4)$$

Модели $ARMA(p, q)$ могут описывать широкий спектр поведений стационарных временных рядов. При этом, если $p = 0$, модель упрощается до модели скользящего среднего, а при $q = 0$ переходит в модель авторегрессии порядка p . Как видно, модель $ARMA(p, q)$ будет учитывать при прогнозе предшествующие значения зависимой переменной, а также текущее и прошлые значения величины ошибки.

Порядки p и q в модели $ARMA$ определяются структурой выборочных функций автокорреляции и частной автокорреляции, а также значениями критериев выбора, о которых говорится далее. Для определения порядков можно пользоваться табл. 7.1, в которой представлены поведения коэффициентов АКФ и ЧАКФ.

Таблица 7.1

Поведение коэффициентов АКФ и ЧАКФ в моделях $ARMA(p, q)$

Модель	АКФ	ЧАКФ
$AR(p)$	Затухает	Обрывается на шаге p
$MA(q)$	Обрывается на шаге q	Затухает
$ARMA(p, q)$	Затухает	Затухает

Таким образом, если график АКФ затухает (экспоненциально или синусоидально), а ЧАКФ быстро обрывается после пика на лаге p , то в модели должны присутствовать авторегрессионные слагаемые. Если график АКФ обрывается после пика на лаге q , а ЧАКФ плавно стремится к нулю, то в модели имеются слагаемые скользящего среднего. Если оба графика (и АКФ и ЧАКФ) плавно затухают, то в модели должны присутствовать слагаемые обоих типов. Вывод о значимости

коэффициентов автокорреляции и частной автокорреляции делается путем сравнения их с величиной $\pm \frac{t_\alpha}{\sqrt{n}}$, где n — длина временного ряда (такое сопоставление оправданно только в том случае, если n достаточно велико). В практических расчетах выборочных АКФ и ЧАКФ рекомендуется использовать величину лага $\tau \leq n/4$.

Типичные графики АКФ и ЧАКФ для наиболее часто встречающихся вариантов приведены далее (рис. 7.2—7.4).

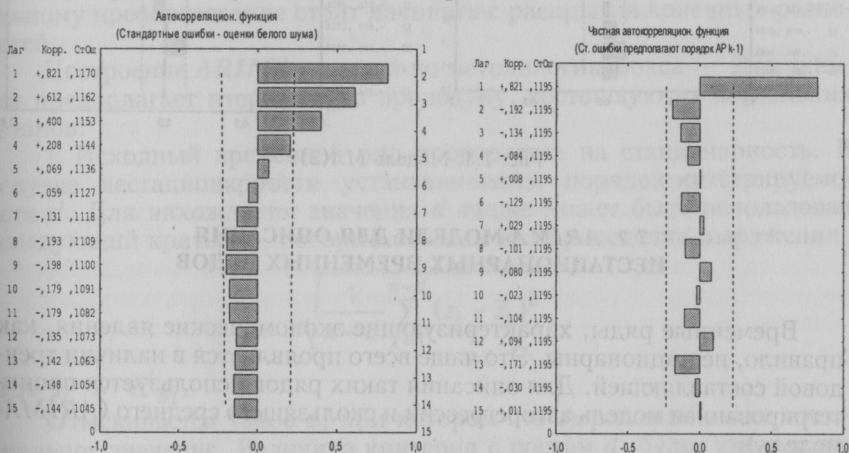


Рис. 7.2. Модель AR(1)

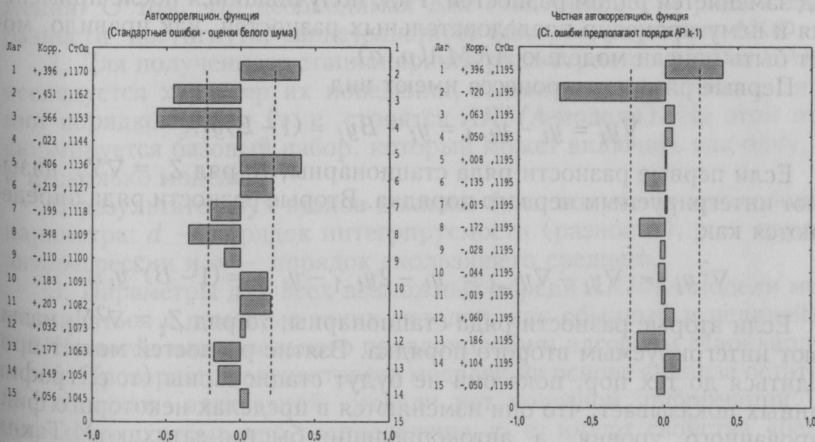


Рис. 7.3. Модель AR(2)

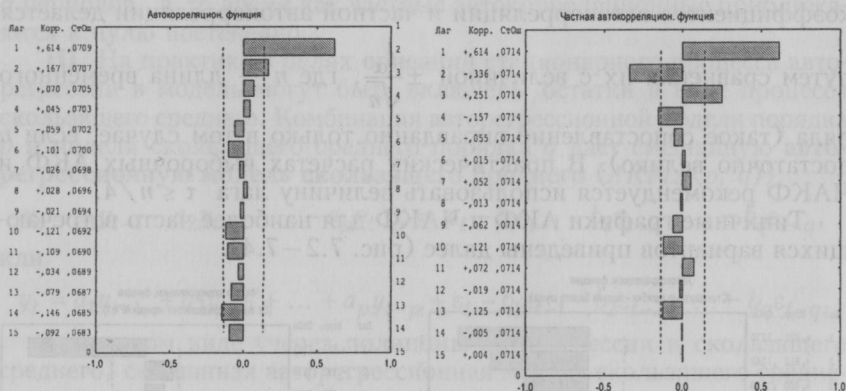


Рис. 7.4. Модель МА(2)

7.2. АРИМА-МОДЕЛИ ДЛЯ ОПИСАНИЯ НЕСТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ

Временные ряды, характеризующие экономические явления, как правило, нестационарны. Это чаще всего проявляется в наличии трендовой составляющей. Для описания таких рядов используется проинтегрированная модель авторегрессии и скользящего среднего (ARIMA-модель).

Используя подход Бокса—Дженкинса, получить стационарные временные ряды возможно путем взятия разностей, т. е. исходный ряд заменяется рядом разностей. Ряд, получившийся после применения к нему процедур последовательных разностей, как правило, может быть описан моделью ARMA(p, q).

Первые разности процесса имеют вид

$$\nabla y_t = y_t - y_{t-1} = y_t - B y_t = (1 - B) y_t.$$

Если первые разности ряда стационарны, то ряд $Z_t = \nabla Y_t$ называют интегрируемым первого порядка. Вторые разности ряда определяются как

$$\nabla^2 y_t = \nabla y_t - \nabla y_{t-1} = y_t - 2y_{t-1} - y_{t-2} = (1 - B)^2 y_t.$$

Если вторые разности ряда стационарны, то ряд $Z_t = \nabla^2 Y_t$ называют интегрируемым второго порядка. Взятие разностей может проводиться до тех пор, пока они не будут стационарны (т. е. график данных показывает, что они изменяются в пределах некоторого фиксированного уровня, а автокорреляции быстро затухают). Таким образом, если получается первый стационарный ряд после d -кратного взятия разностей, говорят, что ряд $Z_t = \nabla^d Y_t$ интегрируем порядка d .

В неявном виде *ARIMA*-модель порядка (p, d, q) выглядит как

$$\varphi_p(B)\nabla^d y_t = \psi_q(B)\varepsilon_t,$$

или

$$\varphi_p(B)(1-B)^d y_t = \psi_q(B)\varepsilon_t. \quad (7.5)$$

Таким образом, *ARIMA*-модель формируется за счет включения в смешанную модель оператора взятия простых разностей порядка d .

При приведении неявного вида модели с заданными порядками к явному преобразование стоит начинать с раскрытия конечных разностей.

Построение *ARIMA*-моделей по методологии Бокса — Дженкинса предполагает итеративную процедуру, состоящую из нескольких этапов.

1. Исходный временной ряд проверяется на стационарность. В случае нестационарности устанавливается порядок интегрируемости d . Для нахождения значения d также может быть использован следующий критерий, который основан на определении выражения

$$\sqrt{\frac{1}{n-d} \sum_{t=1}^{n-d} (z_t - \bar{z}_t)^2},$$

где $z_t = \nabla^d y_t$.

Отыскивается такое d , при котором это выражение имеет минимальное значение. Величина критерия с ростом d будет уменьшаться до тех пор, пока разностный оператор не станет стационарным. Дальнейшее возрастание порядка d ведет к увеличению этого выражения. О том, что необходимая для стационарности ряда степень разности достигнута, свидетельствует быстрое затухание АКФ.

2. Для полученного стационарного ряда строятся АКФ и ЧАКФ, исследуется характер их поведения, выдвигается гипотеза о значении порядков p и q (т. е. строятся *ARMA*-модели). На этом этапе формируется базовый набор, который может включать как одну, так и несколько моделей.

В результате двух этапов должны быть получены три основных параметра: d — порядок интегрируемости (разности), p — порядок авторегрессии и q — порядок скользящего среднего.

3. Параметры для всех возможных версий *ARIMA*-модели могут оцениваться с помощью таких методов, как обычный и нелинейный МНК; метод максимального правдоподобия; алгоритм Марквардта.

4. Проверяется адекватность моделей на основе анализа остатков. В остатках адекватной модели нет полезной информации, они случайно и нормально распределены, т. е. имеют свойства «белого шума».

Вывод о незначимости коэффициентов автокорреляции ряда ос-

татков делается путем сравнения их с величиной $\pm \frac{t_\alpha}{\sqrt{n}}$, т. е. гипотеза

о значимости отвергается, если значения коэффициентов АКФ не выходят за пределы этого интервала.

Проверку незначимости группы первых коэффициентов АКФ также можно осуществлять при помощи теста χ^2 , который базируется на Q -статистике Бокса — Пирса:

$$Q = n \sum_{\tau=1}^k r_\tau^2$$

или Бокса — Льюнга:

$$Q = n(n+2) \sum_{\tau=1}^k \frac{r_\tau^2}{n-\tau}$$

Если $Q > \chi^2$ при ν степенях свободы ($\nu = k - p - q$) и уровне значимости α , то группа первых коэффициентов автокорреляции является значимой (рекомендуется рассматривать $k = n/4$).

5. Если несколько моделей из базового набора оказались адекватными исследуемому процессу, то окончательно выбирается модель с наивысшим качеством и наименьшим числом параметров. Воедино эти требования сведены в информационном критерии Акайка (AIC) или Шварца (SIC).

$$AIC = \frac{p+q}{n} + \ln \left(\frac{\sum e_t^2}{n} \right);$$

$$SIC = \frac{(p+q) \ln n}{n} + \ln \left(\frac{\sum e_t^2}{n} \right).$$

Выбор следует сделать в пользу модели с меньшим значением критериев.

6. Осуществляется прогнозирование на основе выбранной $ARIMA$ -модели порядка (p, d, q) .

7.3. $ARIMA$ -МОДЕЛИ ДЛЯ ОПИСАНИЯ ВРЕМЕННЫХ РЯДОВ С СЕЗОННОЙ СОСТАВЛЯЮЩЕЙ

Нестационарность экономических временных рядов может быть связана с наличием как основной тенденции (тренда), так и периодических колебаний. Ряд имеет сезонность (периодичность) с периодом s , когда сходные особенности ряда повторяются через s интервалов времени (т. е. фактом, относящим временной ряд к сезонным,

является сходство наблюдений, разделенных интервалом s). Сезонность во временных рядах может быть учтена посредством взятия сезонных разностей. Это означает, что для процесса с периодом колебаний s разности вида

$$\nabla_s y_t = y_t - y_{t-s} = y_t - B^s y_t = (1 - B^s) y_t$$

имеют случайный характер, а процесс разностей $Z_t = \nabla_s Y_t$ является стационарным (где $B^s y_t = y_{t-s}$ — сезонный оператор сдвига).

Таким образом, нестационарный ряд называется *сезонно интегрируемым* порядка (d, D) , если он может быть преобразован в стационарный вычислением s -разностей D раз и последующим расчетом конечных разностей d раз. Экономические ряды, как правило, имеют порядок сезонной разности D не более 1, т. е. предполагается, что применение s -разностей один раз позволяет исключить сезонность.

Сезонные *ARIMA*-модели включают обычные авторегрессионные составляющие и составляющие скользящего среднего, отвечающие за корреляции при низких интервалах, а также авторегрессионные составляющие и составляющие скользящего среднего, отвечающие за автокорреляции и частные автокорреляции при сезонных интервалах, а также разности, отвечающие за исключение тенденции, и сезонные разности.

В неявном виде сезонная *ARIMA*-модель выглядит как

$$\Phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D y_t = \Psi_q(B)\Psi_Q(B^s)e_t,$$

или

$$\Phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D y_t = \Psi_q(B)\Psi_Q(B^s)e_t, \quad (7.6)$$

где $\Phi_p(B^s)$ — полином сезонной авторегрессии порядка P ;

$\Psi_Q(B^s)$ — полином сезонного скользящего среднего порядка Q ;

∇_s^D — сезонная разность порядка D .

ARIMA-модель сезонного процесса имеет результирующий порядок $(p, d, q) \times (P, D, Q)_s$.

Пример 7.1. Представим *ARIMA*-модель порядка $(0, 1, 1) \times (0, 1, 1)_{12}$ в явном виде.

Мультипликативная запись модели имеет вид

$$\Phi_0(B)\Phi_0(B^{12})(1-B)^1(1-B^{12})^1 y_t = \Psi_1(B)\Psi_1(B^{12})e_t,$$

или

$$(1-B)(1-B^{12})y_t = (1-b_1B)(1-\beta_1B^{12})e_t.$$

После перемножения операторов получим:

$$(1-B-B^{12}+B^{13})y_t = (1-b_1B-\beta_1B^{12}+b_1\beta_1B^{13})e_t,$$

или

$$y_t - y_{t-1} - y_{t-2} + y_{t-3} = e_t - b_1 e_{t-1} - \beta_1 e_{t-2} + b_1 \beta_1 e_{t-3}.$$

В явном виде эта модель выглядит так:

$$y_t = y_{t-1} + y_{t-2} - y_{t-3} + e_t - b_1 e_{t-1} - \beta_1 e_{t-2} + b_1 \beta_1 e_{t-3},$$

где параметры b_1 и β_1 количественно оценены.

7.4. ПРОГНОЗИРОВАНИЕ НА ОСНОВЕ ARIMA-МОДЕЛЕЙ

Когда модель, адекватно описывающая явление, найдена, можно делать прогнозы на один и несколько периодов вперед с построением доверительных интервалов. Прогнозирование удобнее всего выполнять при помощи выражения ARIMA-модели в явном виде, после того как стали известны оценки параметров a_j , b_j , A_j , β_j .

Прогнозирование ARMA-процессов.

1. После того как произведены необходимые проверки и вычислены оценки параметров авторегрессии порядка p , прогнозные значения y_t на период $t + l$ можно найти по формуле

$$y_{t+l}^* = \hat{a}_1 y_{t+l-1} + \hat{a}_2 y_{t+l-2} + \dots + \hat{a}_p y_{t+l-p} + \varepsilon_{t+l},$$

где неизвестные ε_t заменяются нулями.

Сначала вычисляют прогнозное значение на один шаг вперед (y_{t+1}^*) по формуле

$$y_{t+1}^* = \hat{a}_1 y_t + \hat{a}_2 y_{t+1} + \dots + \hat{a}_p y_{t+1-p}.$$

Затем в модель

$$y_{t+2}^* = \hat{a}_1 y_{t+1}^* + \hat{a}_2 y_t + \dots + \hat{a}_p y_{t+2-p}$$

подставляют вычисленное прогнозное значение y_{t+1}^* и определяют прогнозную величину y_{t+2}^* и т. д.

Так, для AR(1) формулы прогнозных значений на один шаг вперед $y_{t+1}^* = \hat{a}_1 y_t$, на два — $y_{t+2}^* = \hat{a}_1 y_{t+1}^*$ и т. д.

Если исследуемый ряд y_t не является стационарным и может быть представлен в виде суммы долговременной составляющей (тренда) и случайной компоненты (относительно которой выдвигается гипотеза о стационарности), то общий прогноз показателя y_t может быть сделан по обеим составляющим ряда: по тренду — путем простой экстраполяции, а по отклонениям — по модели авторегрессии порядка p . Сумма двух прогнозов даст общий суммарный прогноз для исследуемого ряда. В этом случае прогноз осуществляется по представленным

ранее формулам, где вместо значений y_t берутся значения отклонений от тренда e_t .

2. Для моделей скользящего среднего формула для прогнозирования имеет вид

$$y_{t+l}^* = \varepsilon_{t+l} - \hat{b}_1 \varepsilon_{t+l-1} - \hat{b}_2 \varepsilon_{t+l-2} - \dots - \hat{b}_q \varepsilon_{t+l-q}.$$

Для процесса $MA(1)$ формулы для прогнозирования следующие:

$$y_{t+1}^* = -\hat{b}_1 \varepsilon_t;$$

$$y_{t+l}^* = 0 \text{ для всех } l \geq 2.$$

Для процесса $MA(2)$ формулы для прогнозирования выглядят так:

$$y_{t+1}^* = -\hat{b}_1 \varepsilon_t - \hat{b}_2 \varepsilon_{t-1};$$

$$y_{t+2}^* = -\hat{b}_2 \varepsilon_t,$$

$$y_{t+l}^* = 0 \text{ для всех } l \geq 3.$$

3. Формула для прогнозирования $ARMA$ -процессов в явном виде следующая:

$$y_{t+l}^* = \hat{a}_1 y_{t+l-1} + \hat{a}_2 y_{t+l-2} + \dots + \hat{a}_p y_{t+l-p} + \varepsilon_t - \hat{b}_1 \varepsilon_{t+l-1} - \hat{b}_2 \varepsilon_{t+l-2} - \dots - \hat{b}_q \varepsilon_{t+l-q}.$$

Так, для процесса $ARMA(1,1)$ формулы приобретают вид

$$y_{t+1}^* = \hat{a}_1 y_t - \hat{b}_1 \varepsilon_t,$$

$$y_{t+l}^* = \hat{a}_1 y_{t+l-1} \text{ для всех } l \geq 2.$$

Вероятностные границы предсказанного значения y_t можно получить с помощью следующего интервала:

$$y_t^* - t_\alpha \hat{\sigma}_{\varepsilon_t} \leq y_t \leq y_t^* + t_\alpha \hat{\sigma}_{\varepsilon_t},$$

где y_t^* — предсказываемое значение исследуемого параметра;

$\hat{\sigma}_{\varepsilon_t}^2$ — оценка дисперсии случайной величины ε_t , вычисляемая по формуле

$$\hat{\sigma}_{\varepsilon_t}^2 = \frac{\sum \varepsilon_t^2}{n-p},$$

где n — число наблюдений.

Для прогнозирования можно использовать *ARMA*-модель, в которой будущее значение показателя выражается через бесконечную линейную комбинацию значений $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2} \dots$ случайного компонента

$$y_{t+l}^* = \sum_{j=0}^{\infty} \phi_j \varepsilon_{t+l-j},$$

где веса ϕ_j могут быть определены из системы

$$\begin{cases} \phi_1 = \alpha_1 - \beta_1, \\ \phi_2 = \alpha_1 \phi_1 + \alpha_2 - \beta_2, \\ \phi_3 = \alpha_1 \phi_2 + \alpha_2 \phi_1 + \alpha_3 - \beta_3, \\ \dots \quad \dots \quad \dots \quad \dots \\ \phi_j = \alpha_1 \phi_{j-1} + \alpha_2 \phi_{j-2} + \dots + \alpha_p \phi_{j-p-d} - \beta_j. \end{cases}$$

$\phi_0 = 1, \phi_j = 0$ для $j < 0, \beta_j = 0$ при $j > q$.

Если $j > \max\{p-d-1, q\}$, то $\phi_j = \alpha_1 \phi_{j-1} + \alpha_2 \phi_{j-2} + \dots + \alpha_{p+d} \phi_{j-p+d}$.

Предсказание величины y_{t+l+1} в момент $t+l$ отличается от этой же величины в момент времени t на ошибку предсказания e_{t+1} , умноженную на коэффициент ϕ_l . Поэтому как только становится известна величина y_{t+1} и определена ошибка прогноза на первом шаге

$e_{t+1} = y_{t+1} - y_{t+1}^*$, то можно уточнить все предсказания для периодов $t+2, t+3$ и т. д. по формуле

$$y_{(t+1)+l}^* = y_{t+1}^*(l) = y_t^*(l+1) + \phi_l e_{t+1}, \quad (7.7)$$

где $y_{t+1}^*(l)$ и $y_t^*(l+1)$ — предсказания будущего значения y_{t+l+1} , сделанные в моменты времени $t+1$ и t соответственно.

Дисперсия ошибки прогноза

$$D(e_{t+l}) = \sigma_{\varepsilon_t}^2 \sum_{j=0}^{l-1} \phi_j^2$$

позволяет вычислить границы доверительного интервала прогноза, если сделать предположение, что ε_t имеет характеристики «белого шума» [28; 36].

Прогнозирование *ARIMA*-процессов может быть произведено, как было сказано ранее, при помощи *ARIMA*-модели, выраженной в явном виде:

$$\begin{aligned} y_{t+l}^* = & \hat{a}_1 y_{t+l-1} + \hat{a}_2 y_{t+l-2} + \dots + \hat{a}_p y_{t+l-p} + \dots + \varepsilon_t - \hat{b}_1 \varepsilon_{t+l-1} - \\ & - \hat{b}_2 \varepsilon_{t+l-2} - \dots - \hat{b}_q \varepsilon_{t+l-q}, \end{aligned}$$

где неизвестные значения y_t заменяются прогнозными оценками, неизвестные ε_t — нулями, а известные ε_t — это вычисленные ошибки прогноза на шаг вперед, т. е. $e_t = y_t - y_{t-1}^*$.

Пример 7.2. Для временного ряда длиной $n = 20$ была получена *ARIMA*-модель порядка $(1, 1, 0)$ с оценкой параметра авторегрессии, равной $0,8$. Уточним прогнозные значения, полученные для $t = 21, 22, 23$, учитывая фактические данные $y_{19} = 23,7$, $y_{20} = 23,4$, $y_{21} = 23,1$. Запись модели через разностный оператор имеет следующий вид:

$$(1 - 0,8B)(1 - B)y_t = \varepsilon_t,$$

или

$$(1 - 0,8B - B + 0,8B^2)y_t = \varepsilon_t;$$

$$(1 - 1,8B + 0,8B^2)y_t = \varepsilon_t; \quad (7.8)$$

$$y_t = 1,8y_{t-1} - 0,8y_{t-2} + \varepsilon_t.$$

Формула для прогноза может быть записана в виде

$$y_{t+l}^* = 1,8y_{t+l-1} - 0,8y_{t+l-2} + \varepsilon_{t+l},$$

тогда для $l = 1$ прогнозное значение определяется по формуле

$$y_{t+1}^* = 1,8y_t - 0,8y_{t-1},$$

для $l = 2$

$$y_{t+2}^* = 1,8y_{t+1}^* - 0,8y_t,$$

для $l = 3$

$$y_{t+3}^* = 1,8y_{t+2}^* - 0,8y_{t+1}^*.$$

Подставляя необходимые значения показателя y_t в формулы прогноза, получим прогнозные оценки:

$$y_{21}^* = y_{20+1}^* = 1,8 \cdot 23,4 - 0,8 \cdot 23,7 = 23,16;$$

$$y_{22}^* = y_{20+2}^* = 1,8 \cdot 23,16 - 0,8 \cdot 23,4 = 22,97;$$

$$y_{23}^* = y_{20+3}^* = 1,8 \cdot 22,97 - 0,8 \cdot 23,16 = 22,81;$$

$$y_{24}^* = y_{20+4}^* = 1,8 \cdot 22,81 - 0,8 \cdot 22,97 = 22,69 \text{ и т. д.}$$

Имея фактическое значение показателя при $t = 21$, можно найти ошибку прогноза на первом шаге: $e_{t+1} = y_{t+1} - y_{t+1}^*$. Таким образом, $e_{21} = y_{21} - y_{20+1}^* = 23,1 - 23,16 = -0,06$.

Уточним все предсказания показателя y_t для периодов $t + 2$, $t + 3$, $t + 4$. Для этого определим параметры α_j , β_j из соотношения (7.8). Получим, что $\alpha_1 = 1,8$, $\alpha_2 = -0,8$, $\beta_j = 0$. Коэффициенты ϕ_j определяются из системы: $\phi_1 = \alpha_1 - \beta_1 = 1,8 - 0 = 1,8$; так как $j = 2$ больше, чем $\max\{p-d-1, q\}$, то ϕ_2 определяется по формуле

$$\phi_2 = 1,8\phi_1 - 0,8\phi_0 = 1,8 \cdot 1,8 - 0,8 \cdot 1 = 2,44; \text{ а } \phi_3 = 1,8\phi_2 - 0,8\phi_1 = 1,8 \cdot 2,44 - 0,8 \cdot 1,8 = 2,95 \text{ и т. д.}$$

Следовательно, уточненный прогноз получим по формуле (7.7):

$$y_{22}^* = y_{21}^*(1) = y_{20}^*(2) = \phi_1 e_{21} = 22,97 + 1,8 \cdot (-0,06) = 22,86,$$

$$y_{23}^* = y_{21}^*(2) = y_{20}^*(3) = \phi_2 e_{21} = 22,81 + 2,44 \cdot (-0,06) = 22,67,$$

$$y_{24}^* = y_{21}^*(3) = y_{20}^*(4) = \phi_3 e_{21} = 22,69 + 2,95 \cdot (-0,06) = 22,51.$$

Прогнозирование для сезонной модели также лучше осуществлять по выражению показателя в явном виде. Так, прогноз на l шагов вперед по сезонной *ARIMA*-модели из примера 7.1 осуществляется по формуле

$$y_{t+l}^* = y_{t+l-1} + y_{t+l-12} - y_{t+l-13} + e_{t+l} - b_1 e_{t+l-1} - \beta_1 e_{t+l-12} + b_1 \beta_1 e_{t+l-13}.$$

Например, для получения прогноза на 3 шага вперед имеем:

$$y_{t+3}^* = y_{t+3-1} + y_{t+3-12} - y_{t+3-13} + e_{t+3} - b_1 e_{t+3-1} - \beta_1 e_{t+3-12} - b_1 \beta_1 e_{t+3-13} = y_{t+2} + y_{t-9} - y_{t-10} + e_{t+3} - b_1 e_{t+2} - \beta_1 e_{t-9} - b_1 \beta_1 e_{t-10}.$$

Как было сказано ранее, для получения прогнозов неизвестные y_t заменяются прогнозными оценками, неизвестные e_t — нулями, а известные e_t — это вычисленные ошибки прогноза на шаг вперед, т. е. $e_t = y_t - y_{t-1}^*$.

Таким образом, формула для построения точечного прогноза при известных оценках b_1, β_1 следующая:

$$y_{t+3}^* = y_{t+2}^* + y_{t-9} - y_{t-10} - \beta_1 (y_9 - y_{t-10}^*(1)) + b_1 \beta_1 (y_{t-10} - y_{t-11}^*(1)).$$

При прогнозировании *ARIMA*-процессов необходимо начинать работу с минимального количества обоснованно отобранных параметров и добавлять их в модель лишь по мере необходимости. Если некоторые параметры в оцененной *ARIMA*-модели окажутся незначительными (по отношению к t), их следует удалять по одному, и каждый раз заново уточнять модель. Иногда из-за высокой корреляции отдельные параметры, которые давали малый эффект, после очередного удаления могут оказаться значимыми.

Пример 7.3. Используя квартальные данные о вводе жилья в РФ рассмотрим пример построения сезонной *ARIMA*-модели в ППП «Statistica». Фактическая динамика ввода в эксплуатацию жилых домов в РФ представлена на рис. 6.1. В предыдущих параграфах уже было отмечено, что ряд имеет отчетливо возрастающий тренд, а также сезонную составляющую, т. е. необходимо оценивать сезонную *ARIMA*-модель вида (7.6).

Для определения порядка конечных разностей идентификации модели возьмем подходящие разности ряда и рассмотрим соответствующие АКФ и ЧАКФ. Для этого в модуле «Анализ временных рядов» выбираем вкладку «Преобразования, авто- и кросскорреляции, графики». Так как ряд динамики содержит параболический тренд, то порядок простой разности возьмем $d = 2$. Для этого проведем преобразование, выделив сначала переменную $Y(t)$ и отметив разность с лагом 1, а затем переменную с именем $D(-1)$. В результате получим окно, представленное на рис. 7.5.

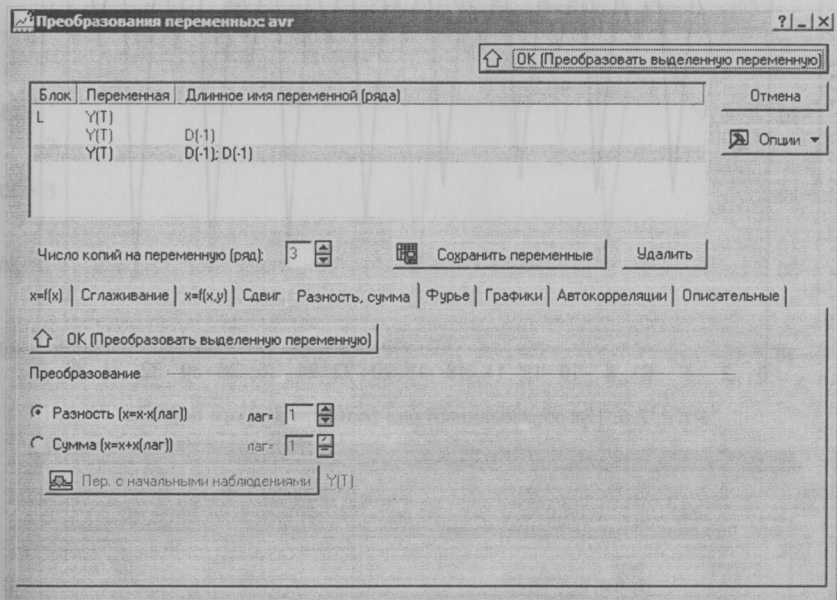


Рис. 7.5. Расчет простой разности

После нахождения простой разности второго порядка получим преобразованный ряд (рис. 7.6).

Для уже преобразованного ряда по аналогии найдем сезонную разность первого порядка. Для этого выделим переменную с именем $D(-1)$; $D(-1)$ и в строке «Разность» укажем лаг, равный 4.

После взятия простых и сезонных разностей графическое изображение преобразованного ряда примет следующий вид (рис. 7.8).

Для определения порядков авторегрессии p и скользящего среднего q используют графики автокорреляционных и частных автокорреляционных функций, которые можно получить, выбрав в окне «Преобразование переменных» вкладку «Автокорреляции» (рис. 7.9).

Длина преобразованного ряда (после взятия разностей) сократилась до 30 ($36 - 2 - 4$). Следовательно, значение лага оцениваемых функций АКФ и ЧАКФ выставим 10. В итоге получим следующие результаты (рис. 7.10, 7.11).

График переменной: Y(T)
D(-1); D(-1)

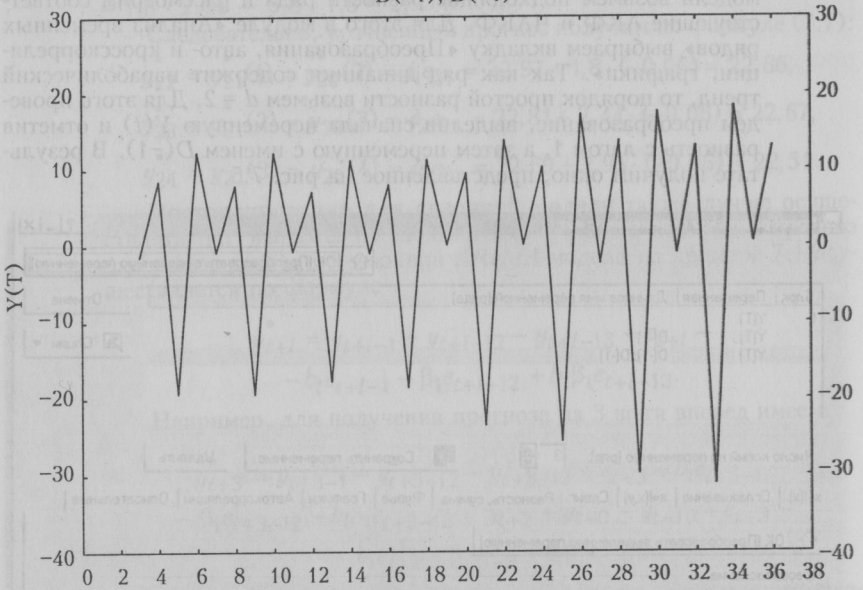


Рис. 7.6. Преобразованный ряд разности второго порядка

Преобразования переменных avg

OK (Преобразовать выделенную переменную)

Блок	Переменная	Длинное имя переменной (ряд)
L	Y(T)	
	Y(T)	D(-1)
	Y(T)	D(-1); D(-1)
	Y(T)	D(-1); D(-1); D(-4)

Отмена

Опции

Число копий на переменную (ряд): 8

Сохранить переменные

Удалить

x=f(x) | Сглаживание | x=f(x,y) | Сдвиг | Разность, сумма | Фурье | Графики | Автокорреляции | Описательные

OK (Преобразовать выделенную переменную)

Преобразование

Разность (x=x-x[лаг]) лаг: 4

Сумма (x=x+x[лаг]) лаг: 1

Пер. с начальными наблюдениями Y(T)

Рис. 7.7. Расчет сезонной разности

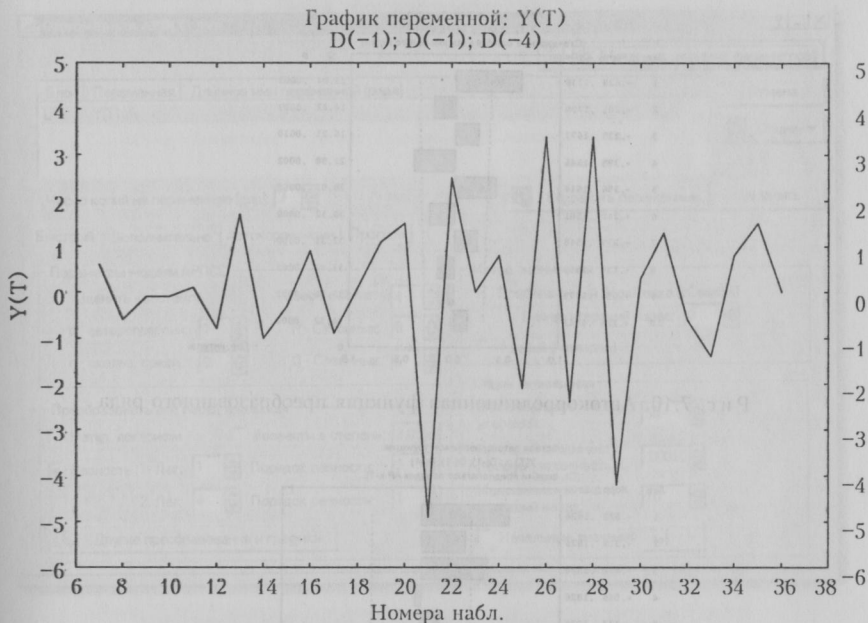


Рис. 7.8. Преобразованный ряд разностей ($d = 2, D = 1$)

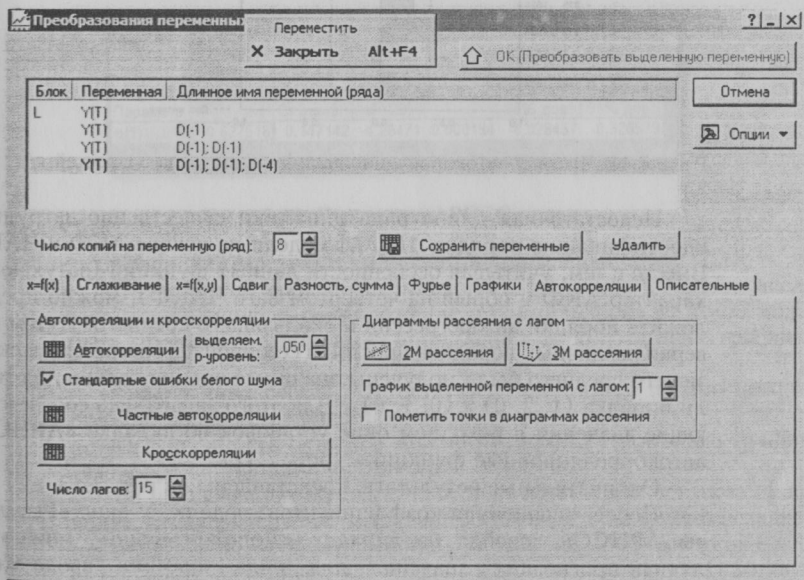
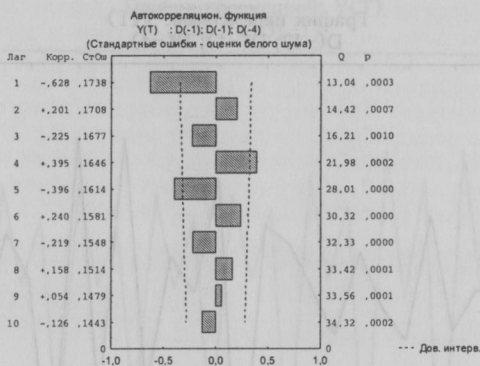
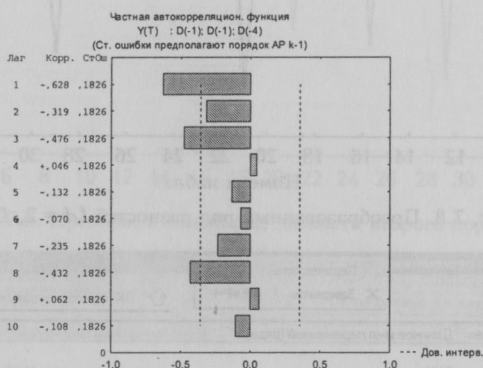


Рис. 7.9. Окно расчета автокорреляционных функций



Р и с. 7.10. Автокорреляционная функция преобразованного ряда



Р и с. 7.11. Частная автокорреляционная функция ряда разностей

Недостаточная длина ряда динамики существенно затрудняет идентификацию порядка *ARIMA*-модели на основе АКФ и ЧАКФ. Вместе с тем, учитывая особенности данных функций (затухающий характер АКФ и обрыв на четвертом лаге ЧАКФ), можно предположить порядки авторегрессии и скользящего среднего. Путем проверки различных порядков простой и сезонной авторегрессии и скользящего среднего были получены значимые оценки только для модели порядка $(1, 2, 0) \times (0, 1, 0)_4$. Для этого задавались соответствующие значения в исходном окне установок во вкладке «АРСС и автокорреляционные функции» (рис. 7.12).

Окончательные результаты представлены на рис. 7.13.

После оценивания коэффициентов модели в окне «Результаты АРСС», перейдя на вкладку «Дополнительно», можно получить прогнозные значения для ряда динамики ввода жилья (рис. 7.14).

Одномерная АРСС: avr

OK (Начать оценивание параметров)

Блок: L Переменная: Y(T) Длинное имя переменной (ряда):

Отмена

Опции

Число копий на переменную (ряд): 3

Сохранить переменные Удалить

Быстрый Дополнительно Автокорреляция Прогноз

Параметры модели АРСС

Оценить константу

Сезонный лаг: 4

р- авторегрессия: 1

Р-Сезонный: 0

q-скользя. средн.: 0

Q-Сезонный: 0

Метод оценивания

Приближенный (МакЛеода и Сейлза)

Кол-во итераций назад: 0

Точный (Меларда)

Преобразовать ряд перед анализом

Natur. логарифм

Возвести в степень: 2.0

Разность 1. Лаг: 1

Порядок разности: 2

2. Лаг: 4

Порядок разности: 1

Другие преобразования и графики

Опции оценивания

Максимальное число итераций: 50

Критерий сходимости (требуемая точность): 0.0001

Максимальное число итераций назад: 10

Начальные значения

Рис. 7.12. Окно установок параметров АРСС

	Исход.: Y(T) (avr)					
	Преобразования: 2*D(1).D(4)					
	Модель(1.2.0)(0.1.0) MS Остаток= 2.0754					
Параметр	Парам.	Асимпт. Ст.ошиб.	Асимпт. t(29)	p	Нижняя 95% дов.	Верхняя 95% дов.
p(1)	-0.627518	0.147142	-4.26471	0.000194	-0.928457	-0.326579

Рис. 7.13. Параметры АРСС-модели

Графически результаты расчетов приведены на рис. 7.15.

Средствами программы оценим адекватность модели исследуемому явлению. Предположение о нормальности остатков может быть проверено с помощью нормальных вероятностных графиков (рис. 7.16).

График показывает, что остатки недостаточно плотно лежат к прямой, соответствующей нормальному распределению.

Независимость остатков можно проверить с помощью графика автокорреляционной функции (рис. 7.17).

За исключением выброса на втором лаге, АКФ остатков не выходит за доверительные интервалы, что свидетельствует о независимости остатков. Особое внимание при анализе АКФ остатков следует обращать на первый лаг и лаги, кратные сезонности (на них не должно быть выбросов).

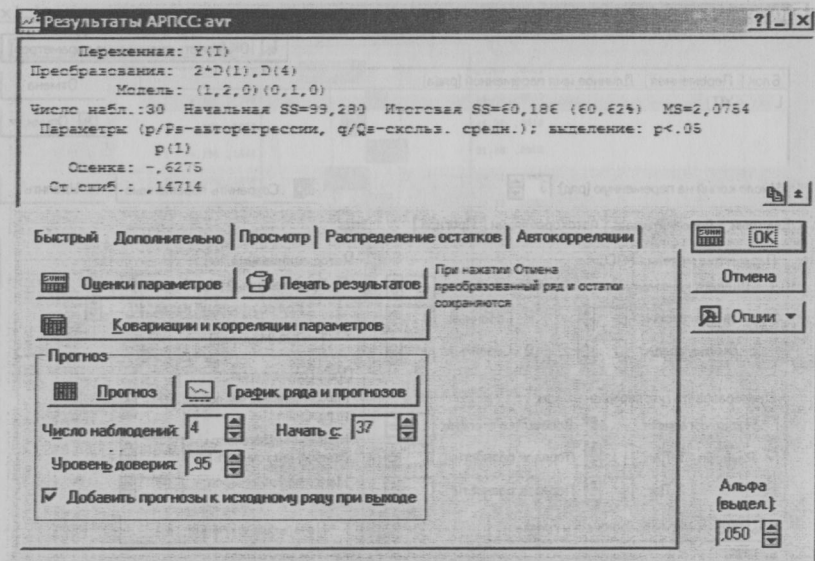


Рис. 7.14. Получение прогноза для АРПСС-модели
 Прогнозы; Модель $(1,2,0)(0,1,0)$ Сезонный лаг: 12
 Исход $Y(T)$
 Начало исходных: 1 Конец исходных: 36

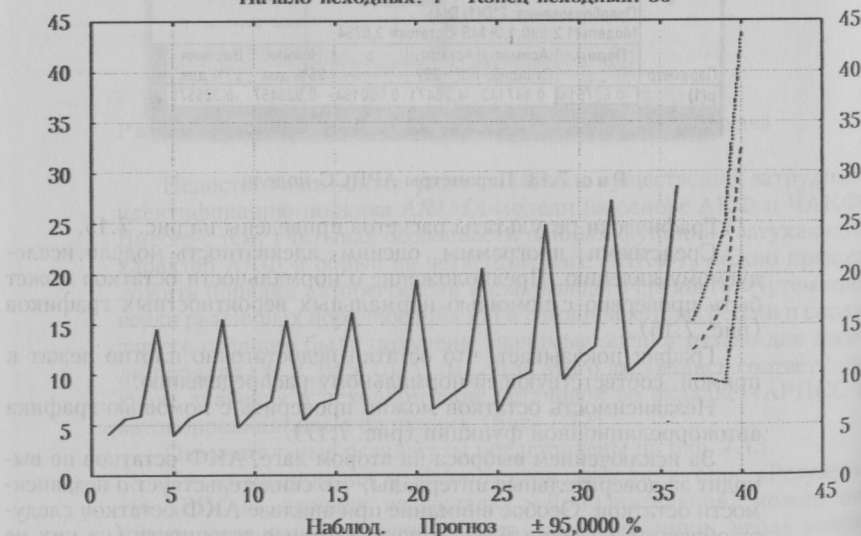


Рис. 7.15. Динамика ввода в эксплуатацию жилых домов в РФ и прогноз по АРПСС-модели, млн m^2

Норм. вероятн. график: $Y(T)$
 АРПСС (1,2,0)(0,1,0) остатки:

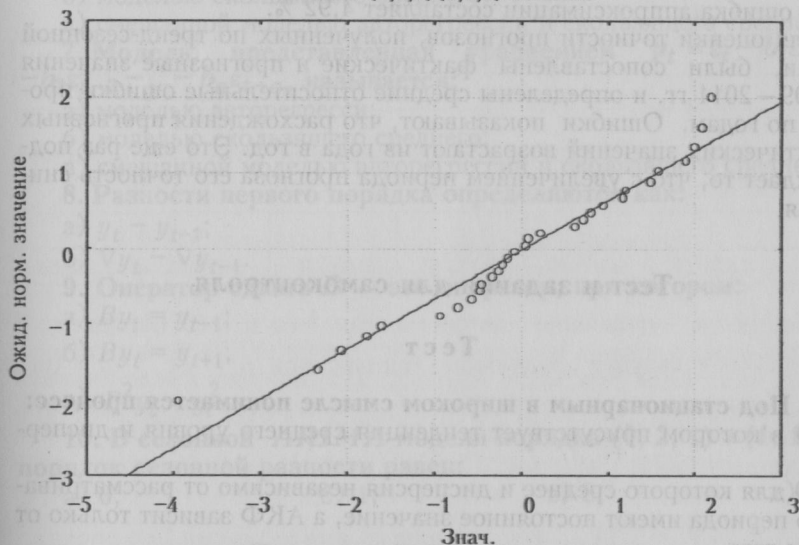


Рис. 7.16. График остатков на нормальной вероятностной бумаге

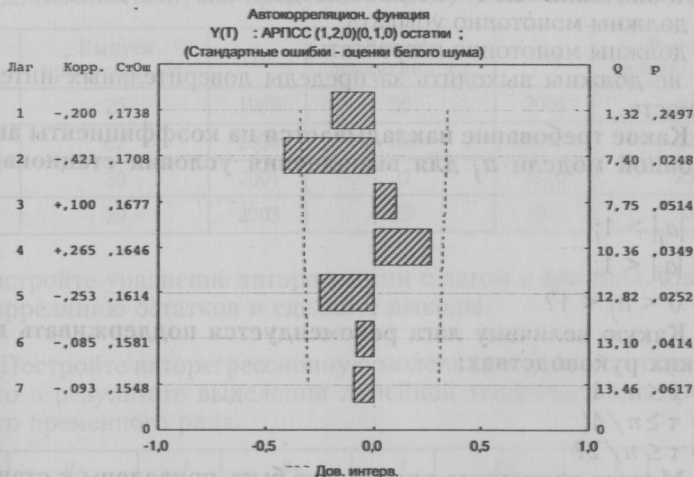


Рис. 7.17. График автокорреляционной функции остатков

Ошибка аппроксимации для сезонной $ARIMA$ -модели порядка $(1, 2, 0) \times (0, 1, 0)_4$ составляет почти 10 %, что говорит о невысокой точности полученной модели. И поэтому, основываясь на расчетах, проведенных в главе 6, предпочтение в выборе модели прогнозирования

ния стоит отдать тренд-сезонной мультипликативной модели, для которой ошибка аппроксимации составляет 4,92 %.

Для оценки точности прогнозов, полученных по тренд-сезонной модели, были сопоставлены фактические и прогнозные значения за 2009—2011 гг. и определены средние относительные ошибки прогноза по годам. Ошибки показывают, что расхождения прогнозных и фактических значений возрастают из года в год. Это еще раз подтверждает то, что с увеличением периода прогноза его точность снижается.

Тест и задания для самоконтроля

Тест

1. Под стационарным в широком смысле понимается процесс:

а) в котором присутствует тенденция среднего уровня и дисперсии;

б) для которого среднее и дисперсия независимо от рассматриваемого периода имеют постоянное значение, а АКФ зависит только от длины лага.

2. Для стационарного временного ряда с увеличением величины лага значения АКФ (по абсолютной величине):

а) должны монотонно убывать;

б) должны монотонно возрастать;

в) не должны выходить за пределы доверительных интервалов значимости.

3. Какое требование накладывается на коэффициенты авторегрессионной модели a_j для выполнения условия стационарности ряда:

а) $|a_j| > 1$;

б) $|a_j| < 1$;

в) $0 < a_j < 1$?

4. Какую величину лага рекомендуется поддерживать в практических руководствах:

а) $\tau \leq n/4$;

б) $\tau \geq n/4$;

в) $\tau \leq n/2$?

5. Многие временные ряды могут быть приведены к стационарному виду после:

а) нахождения простых и/или сезонных разностей;

б) логарифмирования исходного ряда данных.

6. Модель, представленная уравнением $y_t = a_1 y_{t-1} + a_1 y_{t-2} + \dots + a_1 y_{t-2} + \varepsilon_t$, называется:

3. Динамика цен на некоторый товар А характеризуется следующими данными.

t	1	2	3	4	5	...	20	21	22	23	24
y_t	2	3	3	6	4	...	15	12	13	13	14

Получены коэффициенты автокорреляции уровней временного ряда.

τ	1	2	3	4	5	6
r_τ	0,870 25	0,765 79	0,793 43	0,822 78	0,777 90	0,678 33

Постройте уравнения авторегрессии 1-го и 2-го порядков. В расчетах используйте следующие данные: при $n = 24$ $\sum y_t = 186$,

$\sum y_t^2 = 1 794$, при лаге $(t - 4)$ $n = 20$, $\sum y_t y_{t-1} = 1 617$, $\sum y_t y_{t-4} = 1 359$,
 $\sum y_{t-1} y_{t-4} = 1 271$, $\sum y_{t-1}^2 = 1 576$, $\sum y_{t-4}^2 = 1 116$.

4. Пусть $Y_t = 1,1Y_{t-1} + \varepsilon_t$ — процесс авторегрессии 1-го порядка, а ε_t — «белый шум» с нулевым средним. Определите стационарность процесса, используя характеристическое уравнение.

5. Пусть $Y_t = 0,8Y_{t-1} + \varepsilon_t$ — процесс авторегрессии 1-го порядка, а ε_t — «белый шум» с нулевым средним. Определите стационарность процесса, используя характеристическое уравнение.

6. Запишите *ARIMA*-модель порядка $(1, 1, 2)$ в явном виде.

7. Запишите *ARIMA*-модель порядка $(0, 2, 1) \times (1, 1, 1)_{12}$ в явном виде.

1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	3	4	5	6	7
1	2	3	4	5	6	7

8. МНОГОФАКТОРНОЕ МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ

8.1. Типы информации, используемые при построении многофакторных моделей.

8.2. Простая корреляция и регрессия рядов динамики.

8.3. Множественная корреляция и регрессия.

Все явления находятся в постоянном изменении и развитии. Изменения одних явлений, как правило, вызывают изменения других. Это говорит о том, что явления в некоторой степени взаимосвязаны между собой, но эта связь не функциональная (чисто математическая), когда любому значению одного явления соответствует строго определенное значение другого явления. В этом случае говорят о корреляционной связи, при которой одному значению одного явления может соответствовать множество значений другого явления. Корреляционная связь — частный случай стохастической связи, которая проявляется в том, что существуют общие случайные факторы, влияющие на каждое явление неодинаково. Например, стохастически зависимыми можно считать объем произведенной продукции и коэффициент использования рабочего времени на предприятии, так как оба показателя зависят от квалификации персонала, организации труда, ритмичности снабжения и др.

Цель корреляционного анализа — исследование тесноты и направления связи между явлениями. Для аналитического описания формы связи на практике чаще всего используют регрессионный анализ.

Корреляционному и регрессионному анализу должен предшествовать всесторонний теоретический анализ возможности существования связи между явлениями. Необходимо устанавливать все явления, оказывающие влияние на исследуемый процесс и проводить их подробный анализ. Сложность связей между экономическими явлениями часто обуславливает положение, когда корреляционно-регрессионным анализом невозможно охватить весь комплекс взаимосвязей между ними. Следует иметь в виду, что конкретное уравнение связи (модель) всегда строится в некоторой степени абстрагированно от реальности, т. е. построение регрессии выступает как бы экспериментом по выявлению конкретной связи между процессами.

Также нельзя забывать, что исходная информация практически всегда представляет выборочную совокупность и результаты анализа носят вероятностный характер, поэтому необходима их статисти-

ческая оценка с помощью различных критериев. В противном случае исследователь не имеет права распространять результаты анализа на генеральную совокупность.

8.1. ТИПЫ ИНФОРМАЦИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ПОСТРОЕНИИ МНОГОФАКТОРНЫХ МОДЕЛЕЙ

Количественное описание взаимосвязей между экономическими переменными, как правило, осуществляется при помощи построения различных многофакторных регрессионных моделей. Выбор того или иного приема моделирования зависит от поставленных целей и задач исследования. Немаловажную роль при построении таких моделей играет промежуток времени, за который рассматривается информация об изучаемом объекте (изучаемых объектах).

Первый тип информации, который используется для анализа взаимосвязей, — это множество данных, состоящих из наблюдений за однотипными статистическими объектами, такими как страны, регионы, предприятия и т. п., за один период времени (обычно за год). Чаще всего для моделирования на практике используется классическая регрессионная модель. Параметры множественной регрессии, как правило, определяются при помощи метода наименьших квадратов и легко экономически интерпретируются [1; 16; 36; 37]. Основная цель множественной регрессии — построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное воздействие на изменение исследуемого показателя.

Второй тип информации, который используется на практике, — это данные по одному объекту (т. е. по одному предприятию, по одному региону, стране и т. п.) за некоторый длительный промежуток времени. Информация, характеризующая зависимость результирующего признака от факторных, представляется в виде временных рядов данных, которые называют связными рядами динамики. А изучение причинно-следственных зависимостей переменных, представленных в форме временных рядов, считается одной из сложнейших задач в эконометрическом моделировании.

При использовании методов корреляционно-регрессионного анализа не стоит забывать, что исходные совокупности данных должны удовлетворять свойствам случайности, независимости и нормальности распределения. Результаты такого анализа характеризуют изменения явлений через средние величины, а следовательно, корреляционно-регрессионный анализ применим к однородным (типичным) совокупностям. В большинстве случаев временные ряды не удовлетворяют указанным требованиям, и поэтому при проведении корреляционно-регрессионного анализа следует учитывать их специфику.

Применение традиционных методов корреляционного и регресси-

онного анализа может приводить к получению абсурдных результатов (так называемой ложной корреляции), вызванных наличием основной тенденции в каждом ряду или периодических колебаний. Сильная зависимость между уровнями ряда, т. е. автокорреляция уровней, при применении МНК в оценке параметров регрессии приводит к их смещению.

Третий тип информации, который используется при построении многофакторных моделей, — это данные, содержащие сведения об одном и том же множестве объектов за ряд последовательных периодов времени. Множество таких данных называется **панельным**. В качестве панели могут выступать как индивидуумы, так и группы лиц, а также предприятия, домохозяйства, регионы, страны и т. д. Для большинства панельных данных характерно то, что они содержат наблюдения о большом количестве объектов за относительно короткий промежуток времени. Регрессионный анализ по панельным данным дает возможность оценить коэффициенты объединенной модели (*pooled model*), модели с фиксированными эффектами (*fixed effects model*), модели со случайными эффектами (*random effects model*).

8.2. ПРОСТАЯ КОРРЕЛЯЦИЯ И РЕГРЕССИЯ РЯДОВ ДИНАМИКИ

Простой корреляцией принято называть корреляцию между двумя показателями. Для определения тесноты связи на практике часто используется линейный коэффициент корреляции

$$r = \frac{yx - \bar{y}\bar{x}}{\sigma_y\sigma_x} = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}, \quad (8.1)$$

который принимает значения в интервале $-1 < r < 1$. Отрицательное значение коэффициента корреляции указывает на обратную связь между динамикой явлений. Чем ближе его значение по абсолютной величине к единице, тем теснее рассматриваемая связь.

Чтобы использовать корреляционные методы для изучения причинно-следственных связей по временным рядам, необходимо учитывать следующие особенности. **Первая особенность**, возникающая при коррелировании, заключается в наличии автокорреляции уровней, вызванной присутствием тенденции в каждом ряду данных. Наличие сильной зависимости между уровнями может приводить к завышению коэффициента корреляции, т. е. к ложной корреляции.

Чтобы получить правильную картину связи между временными рядами, не искаженную автокорреляцией, необходимо исключить и основную тенденцию из уровней рядов. Это можно сделать путем вычитания из фактических значений уровней ряда теоретических, найденных по уравнению тренда ($\varepsilon_t = y_t - \hat{y}_t$, $\eta_t = x_t - \hat{x}_t$).

Корреляцию отклонений от трендов можно найти по следующей формуле:

$$r = \frac{\sum \varepsilon_t \cdot \eta_t}{\sqrt{\sum \varepsilon_t^2 \cdot \sum \eta_t^2}}, \quad (8.2)$$

где ε_t, η_t — отклонения от соответствующих тенденций исследуемых явлений.

По аналогичной формуле рассчитывается линейный коэффициент корреляции, если тенденция исключена взятием разностей (только вместо остаточных величин берутся значения первых разностей).

Если исходные совокупности данных содержат периодические сезонные колебания, то при коррелировании прежде всего должно быть устранено влияние сезонности. Элиминировать это влияние можно делением исходных рядов данных на соответствующие индексы сезонности, выраженные в коэффициентах.

Если существует предположение, что величины связаны между собой нелинейно, то оценка тесноты связи может быть дана по индексу корреляции

$$R = \sqrt{1 - \frac{\sigma_{\varepsilon(\eta_t)}^2}{\sigma_{\varepsilon}^2}}, \quad (8.3)$$

где $\sigma_{\varepsilon(\eta_t)}^2$ — дисперсия, характеризующая отклонения фактических значений остатков результативного признака от теоретических, рассчитанных по уравнению регрессии;

σ_{ε}^2 — дисперсия остаточных величин результативного признака.

Абсолютное значение индекса корреляции находится в пределах $0 < |R| < 1$. Считается, что чем ближе значение индекса корреляции к единице, тем теснее связь между явлениями.

Пример 8.1. Для выяснения тесноты связи между явлениями были взяты следующие данные: X_t — объем промышленной продукции региона; Y_t — объем продукции сельского хозяйства региона (табл. 8.1).

Рассчитав коэффициент корреляции по формуле (8.1), получим $r = 0,877$, что говорит о сильной взаимосвязи между этими явлениями. Однако на тесноту связи между исследуемыми показателями, возможно, влияет зависимость каждого последующего уровня от предыдущего, т. е. наличие автокорреляции в каждом ряду. Проведенная проверка показала, что коэффициенты автокорреляции первого порядка (см. п. 3.2) для ряда X_t $r_1 = 0,996$, а для ряда Y_t $r_1 = 0,88$. Это означает, что в исследуемых рядах существует автокорреляция уровней. Поэтому, прежде чем говорить о тесноте связи между показателями X_t и Y_t , необходимо устранить ее влияние.

Объемы промышленной и сельскохозяйственной продукции региона

Год	X_t	Y_t	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
2003	248,3	70,9	-106,7	-9,8	1 045,6	11 384,9	95,8
2004	285,9	77,0	-69,1	-3,7	255,7	4 774,8	13,6
2005	322,8	78,1	-32,2	-2,6	83,7	1 036,8	6,7
2006	345,0	81,6	-10,0	0,9	-9,0	100,0	0,8
2007	374,3	78,7	19,3	-2,0	-38,6	372,5	4,0
2008	395,7	87,0	40,7	6,3	256,0	1 656,5	39,9
2009	420,0	87,9	65,0	7,2	468,9	4 225,0	52,0
2010	448,0	84,3	93,0	3,6	334,8	8 649,0	13,1
Итого	2 840,0	645,5	—	—	2 395,7	32 199,5	225,9

Одним из способов устранения автокорреляции является исключение из анализируемых рядов их тенденций. Для ряда X_t уравнение тренда выражается уравнением прямой $\hat{x}_t = 231,3 + 27,5t$, а для ряда $Y_t - \hat{y}_t = 71,7 + 2,1t$. Коэффициент корреляции для рядов отклонений, вычисленный по формуле (8.2), равен 0,22. Это говорит о том, что между исходными рядами данных существует ложная корреляция.

Вторая особенность применения корреляционного анализа к рядам динамики состоит в возможном наличии временного лага, т. е. когда влияние одного явления на другое осуществляется с некоторым запаздыванием (например, капитальные вложения отражаются на изменении объема продукции через определенный период). Величину этого запаздывания (временного лага) можно определить на основе анализа данных, вычислив по коэффициентам корреляции исходного ряда Y_t и ряда X_t , сдвинутого во времени на период запаздывания τ , кросскорреляционную функцию. Присутствие пиков в функции $r(\tau)$ указывает на наличие временного лага, а его длина соответствует τ периодам. Использование лаговых зависимостей предполагает большую продолжительность временных рядов в связи со сдвигом. Получаемые уравнения регрессии с учетом временного лага оказываются статистически более надежными, чем обычная регрессия.

Третьей особенностью корреляции временных рядов является возможность наличия переменной корреляции, которая показывает, как меняется значение коэффициента корреляции с течением времени. По серии коэффициентов корреляции, вычисленных в скользящем порядке с вычитанием одного уровня ряда и прибавлением следующего через определенный интервал скользящего, можно получить более полные сведения о том, как изменялась теснота связи во времени между явлениями.

При моделировании связи между временными рядами, если есть предположение о сильной зависимости уровней одного ряда друг от друга (о наличии автокорреляции), можно использовать регрессию:

- по отклонениям от трендов;
- по конечным разностям;
- по исходным уровням ряда с включением фактора времени.

Оценки коэффициентов для этих регрессий можно проводить при помощи обычного метода наименьших квадратов.

1. Уравнение регрессии по отклонениям от трендов имеет вид $\varepsilon_t = a + b\eta_t$. Свободный член для регрессии a на остатках является незначимым, т. е. близок или равен нулю. Коэффициент регрессии b означает, что случайные отклонения по ряду y_t в среднем в b раз выше случайных колебаний в ряду x_t . Можно расписать уравнение регрессии по отклонениям через конкретные уровни временных рядов:

$$y_t - \hat{y}_t = a + b(x_t - \hat{x}_t),$$

или

$$y_t = \hat{y}_t + a + b(x_t - \hat{x}_t). \quad (8.4)$$

Для прогнозирования по этой модели необходимо определить прогнозные значения по трендам \hat{x}_t и \hat{y}_t , каким-либо способом оценить величину предполагаемого отклонения фактического значения от трендового $x_t - \hat{x}_t = \eta_t$. Далее путем подстановки в полученное уравнение этих значений можно получить прогнозное значение результативного признака.

Пример 8.2. В таблице представлены данные об изменении объема продаж (Y_t) и стоимости капитала (X_t) фирмы за десять лет.

Таблица 8.2

Объем продаж и стоимость капитала фирмы

t	Y_t	X_t	Δy_t	Δx_t
1	2	3	4	5
1	5	8	—	—
2	6	9	1	1
3	7	9	1	0
4	7	10	0	1
5	8	10	1	0
6	8	10	0	0
7	8	10	0	0
8	9	11	1	1
9	10	11	1	0
10	12	12	2	1

Предположим, что переменные связаны между собой линейной зависимостью. При этом для ряда Y_t уравнение тренда описывается уравнением $\hat{y}_t = 4,53 + 0,63t$, а для ряда X_t — уравнением $\hat{x}_t = 8,0 + 0,36t$. Уравнение линейной регрессии по рядам отклонений имеет вид $\hat{\epsilon}_t = 0,00015 + 1\eta_t$. Опустив статистическую оценку значимости уравнения регрессии, заметим, что при условии его надежности это уравнение можно использовать для прогнозирования переменной y_t в зависимости от предполагаемого изменения x_t . Подставив в уравнение (8.4) полученные оценки трендов и регрессии на остатках, получим $y_t = -3,46985 + 0,27t + x_t$. Для оценки x_t можно использовать полученное уравнение тренда для $t = 11$ или экспертные оценки. Прогнозное значение $x_{t=11} = 8,0 + 0,36 \cdot 11 = 11,96$. Таким образом, если в следующем году стоимость капитала составит 11,96 млн руб., то объем продаж $y_{t=11} = 11,46015$ млн руб.

2. Уравнение линейной регрессии по первым разностям имеет следующий вид: $\Delta y_t = a + b\Delta x_t$. Коэффициент b показывает, как изменится прирост результативного признака с изменением скорости роста факторного признака. Существенным недостатком этого метода является потеря информации при переходе к конечным разностям.

Если модель будет характеризоваться высоким показателем детерминации и отсутствием автокорреляции в остатках, то можно перейти к прогнозированию конкретных уровней по уравнению $y_{t=p} = y_n + a + b(x_{t=p} - x_n)$. За величину $x_{t=p}$ можно взять планируемое значение показателя или оценить его по тренду ряда x_t .

Пример 8.3. По данным примера 8.2 определим первые разности (ценные абсолютные приросты) для рядов Y_t и X_t (см. столбцы 4, 5 табл. 8.2). Оценим коэффициенты регрессии с помощью МНК, используя MS Excel. Уравнение регрессии по первым разностям имеет следующий вид: $\Delta y_t = 0,6 + 0,4\Delta x_t$. Таким образом, если у фирмы есть финансовая возможность увеличения капитала на полмиллиона рублей ($\Delta x_t = 0,5$), объем ее продаж в следующем году может увеличиться на 800 тыс. руб. ($\Delta y_t = 0,6 + 0,4 \cdot 0,5 = 0,8$).

3. В корреляционно-регрессионном анализе устранить влияние какого-либо фактора можно, если зафиксировать воздействие этого фактора на результат и на другие включенные в модель факторы. Поэтому включение фактора времени в регрессионную модель позволяет уменьшить (устранить) влияние автокорреляции. Этот метод называется методом Фриша—Боу. В том случае, если в ряду остатков от регрессионной модели наблюдается автокорреляция, а явления y_t и x_t развиваются по линейным (одинаковым) тенденциям, можно линейно ввести в уравнение регрессии фактор времени t :

$$y_t = a + bx_t + ct. \quad (8.5)$$

Коэффициент c перед временным фактором показывает воздействие на изменение результативного показателя всех не включенных в модель регрессии факторов. Преимущество модели (8.5) перед моделями на отклонениях от трендов и на разностях состоит в том, что в ней используются исходные уровни временных рядов.

Пример 8.4. По данным примера 8.2 оценим коэффициенты регрессии (8.5), используя MS Excel. В результате имеем: $\hat{y}_t = -3,4667 + x_t + 0,2667t$. Это уравнение регрессии почти совпадает с полученным по отклонениям от трендов, но по выполнению является менее трудоемким. В рассмотренной модели параметр b показывает, что рост стоимости капитала на 1 млн руб. приведет к такому же увеличению объема продаж при условии неизменной тенденции. Параметр c характеризует, насколько увеличится объем продаж независимо от изменения стоимости капитала, т. е. под влиянием других факторов, не включенных в модель регрессии.

В ряде случаев значимый коэффициент корреляции, рассчитанный по исходным уровням временных рядов, может отражать не наличие ложной корреляции между рядами, а истинную причинно-следственную связь между ними. Такая причинно-следственная зависимость в уровнях двух или более временных рядов, выражающаяся в совпадении или противоположной направленности их тенденций и случайной колеблемости, называется **коинтеграцией** [1; 36]. Эта направленность имеет устойчивый характер и должна сохраняться на протяжении длительного периода времени. Считается, что между временными рядами существует коинтеграция, если линейная комбинация этих временных рядов есть стационарный временной ряд. Остатки для простой линейной регрессии могут быть представлены как линейная комбинация рядов y_t и x_t :

$$\varepsilon_t = y_t - a - bx_t.$$

Для тестирования гипотезы об отсутствии коинтеграции между двумя временными рядами можно использовать величину критерия Дарбина – Уотсона, полученную для ряда остатков регрессии: $y_t = a + bx_t + \varepsilon_t$. Только в этом случае проверку гипотезы проводят сравнением с критическими значениями, полученными методом Монте-Карло [36]. Критические значения для разных уровней значимости следующие: для 1%-го уровня значимости – 0,511; для 5%-го – 0,386; для 10%-го – 0,322. Если фактическое значение критерия Дарбина – Уотсона выше критического для заданного уровня значимости, то гипотеза об отсутствии коинтеграции временных рядов отклоняется (и наоборот). Стоит помнить, что коинтеграция означает совпадение динамики временных рядов на длительном промежутке времени. И поэтому при наличии коротких временных рядов данных, даже если формальные критерии показывают наличие коинтеграции, моделирование взаимосвязей может привести к неверным результатам.

Как было сказано ранее, автокорреляция уровней при применении МНК в оценке параметров регрессии приводит к их смещению. При построении модели регрессии, если доказано, что ряды коинтегрированы, это смещение оценок устраняется по следующей схеме.

1. Построить уравнение регрессии $y_t = a + bx_t$ на исходных данных.

2. Определить значение критерия Дарбина — Уотсона и сравнить с критическим значением при определенном уровне значимости. (т. е. подтвердить коинтеграцию временных рядов).

3. Провести перерасчет исходных данных в соответствии со следующими формулами:

$$\begin{aligned}y'_t &= y_t - r_1 y_{t-1}; \\x'_t &= x_t - r_1 x_{t-1}.\end{aligned}\tag{8.6}$$

Здесь r_1 — коэффициент автокорреляции 1-го порядка, рассчитанный для ряда остатков ϵ_t . Для вычисления этого коэффициента можно использовать формулы (2.6) или (3.16).

4. Определить оценки коэффициентов a' и b' уравнения регрессии y'_t от x'_t обычным МНК.

5. Найти коэффициент a исходного уравнения из соотношения

$$a = \frac{a'}{1 - r_1}.$$

6. В итоге уравнение регрессии выглядит так:

$$\tilde{y}_t = a + b'x_t.\tag{8.7}$$

8.3. МНОЖЕСТВЕННАЯ КОРРЕЛЯЦИЯ И РЕГРЕССИЯ

Экономические процессы редко развиваются индивидуально, в большинстве случаев они образуют целые комплексы, где одно самостоятельное явление воздействует на другие или само подчиняется влиянию других явлений. Для исследования влияния ряда явлений на одно резульативное применяется множественный корреляционно-регрессионный анализ, который подробно рассматривается в курсах «Многомерные статистические методы в экономике» и «Эконометрика».

В качестве уравнения регрессии наиболее часто используется линейная функция

$$\tilde{y} = a_0 + a_1x_1 + a_2x_2 + \dots + a_kx_k,\tag{8.8}$$

коэффициенты которой обычно определяются при помощи МНК.

При построении многофакторных моделей, полученных по временным рядам, возникают две проблемы — автокорреляция и мультиколлинеарность. Как отмечалось ранее, для оценки параметров модели (8.8) используется метод наименьших квадратов, который основывается на предпосылке о независимости друг от друга отдельных наблюдений по одной и той же переменной. В большинстве случаев в экономических временных рядах последовательные наблюдения зависят друг от друга, т. е. между ними существует автокорреляция.

В таких условиях МНК дает неоптимальные оценки параметров регрессии. Как было рассмотрено выше, для устранения автокорреляции, в качестве важнейшей меры рекомендуется исключать тенденции из временных рядов. Другой способ уменьшения автокорреляции основан на включении фактора времени в уравнение множественной регрессии. Этот прием приводит к частичному устранению автокорреляции лишь при условии, что все рассматриваемые временные ряды имеют одинаковую тенденцию.

Автокорреляцию в наблюдениях $\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_n$ можно объяснять следующими причинами:

- 1) в изменении изучаемых величин имеется тенденция;
- 2) в модели регрессии учтены не все факторы, оказывающие существенное влияние на величину результативного фактора;
- 3) неправильно выбрана форма связи между зависимой и независимой переменными;
- 4) в модели не учтен ряд факторов, влияние каждого из которых в отдельности несущественно, но, когда их изменения по направлению и фазе совпадают, в отклонениях может возникнуть автокорреляция;
- 5) имеются ошибки измерения в независимых переменных.

Наиболее распространенным приемом выявления наличия автокорреляции является использование критерия Дарбина — Уотсона (см. п. 3.2).

Другим необходимым условием получения оптимальных оценок параметров модели (8.8) является отсутствие среди факторов x_i линейно-зависимых. Нарушение этого условия ведет к возникновению явления мультиколлинеарности. Один из практических методов ее выявления — определение парного коэффициента корреляции

$$r_{x_i x_j} = \frac{\overline{x_i x_j} - \overline{x_i} \overline{x_j}}{\sigma_i \sigma_j}. \quad (8.8)$$

Факторы считаются линейно-зависимыми, если значение коэффициента по абсолютному значению превышает 0,8. Не стоит забывать, что по временным рядам коэффициенты корреляции могут быть завышены, т. е. может наблюдаться ложная корреляция. Поэтому выводы о наличии сильной зависимости между факторами нужно подтверждать вычислением коэффициентов корреляции на остаточных величинах. Для устранения мультиколлинеарности необходимо первоначально выяснить основные причины ее возникновения. Они могут заключаться в следующем.

1. В качестве независимых переменных выбраны показатели, характеризующие один и тот же признак. (Стоимость основных фондов, численность рабочих и т. д. между собой коррелированы, так как являются характеристиками размера предприятия.)

2. Неоднородность совокупности исходных данных. (Если совокупность состоит из двух частей, каждая из которых формировалась при различных условиях, то высокая корреляционная связь между переменными может быть обусловлена случайным положением изменений переменных в одной совокупности по сравнению с другой.)

3. Использование в качестве независимых переменных величин, сумма которых — постоянное число (удельные веса, проценты и т. д.).

4. Наличие исходных данных, представляющих собой последовательности, имеющие в динамике закономерные компоненты.

Устранение мультиколлинеарности в первую очередь связано с исключением из модели регрессии одной или нескольких связанных независимых переменных. Если же имеются сомнения относительно того, какую из линейно связанных переменных оставить в модели, то можно применить процедуру пошаговой регрессии.

Другой путь устранения эффекта мультиколлинеарности состоит в преобразовании независимых переменных в новые переменные, не коррелированные между собой. Для этого можно использовать метод главных компонент, факторный анализ. (Эти методы подробно рассмотрены в курсе «Многомерные статистические методы».)

При изучении связанных рядов динамики и их моделировании необходимы:

1) тщательный предварительный анализ сущности изучаемого явления (для установления коинтегрируемости или исключения ложной корреляции);

2) выявление структуры изучаемых временных рядов, т. е. компонентного состава, и по возможности элиминирование (устранение из данных) сезонной компоненты.

Кроме того, временные ряды должны иметь достаточную для проведения регрессионного анализа длину — минимум в 4 раза больше, чем количество факторных признаков, включенных в окончательную регрессионную модель. Достаточная длина необходима также для выяснения наличия в рядах временного лага (тесной корреляционной связи, сдвинутой во времени). Если имеются достаточно длинные временные ряды данных, то можно при моделировании и прогнозировании использовать уравнения с учетом временного лага.

Тест и задания для самоконтроля

Тест

1. Какой метод используется для количественной оценки тесноты связи между явлениями:

- регрессионный анализ;
- корреляционный анализ;
- дисперсионный анализ?

2. Если во временных рядах присутствует основная тенденция, то парный коэффициент корреляции рассчитывается:

- а) по исходным уровням временных рядов;
- б) по отклонениям от тренда или первым разностям;
- в) используется только индекс корреляции.

3. В каких пределах может принимать значения парный коэффициент корреляции:

- а) $(-1, 1)$;
- б) $(0, 1)$;
- в) без ограничений?

4. Какие особенности существуют при коррелировании временных рядов:

- а) никаких;
- б) невозможно проведение корреляционного анализа;
- в) из рядов должна быть исключена тенденция?

5. Уравнение регрессии по временным рядам можно построить:

- а) только по исходным уровням с включением фактора времени;
- б) только по отклонениям от трендов;
- в) только по первым разностям;
- г) всеми способами, указанными в п. а, б, в.

6. В уравнении регрессии $\Delta y_t = a + b\Delta x_t$ параметр b показывает:

- а) насколько изменится скорость роста результативного признака с изменением скорости роста факторного признака;
- б) случайные колебания по ряду y_t в среднем в b раз сильнее, чем случайные колебания ряда x_t ;
- в) как в среднем изменится значение y с изменением x на единицу своего измерения?

7. В уравнении регрессии по уровням временных рядов с включением фактора времени $y_t = a + bx_t + ct$ параметр b означает:

- а) насколько изменится скорость роста результативного признака с изменением скорости роста факторного признака;
- б) случайные колебания по ряду y_t в среднем в b раз сильнее, чем случайные колебания ряда x_t ;
- в) как в среднем изменится значение y с изменением x на единицу своего измерения.

Практические задания

1. Рассчитав коэффициенты корреляции по исходным уровням рядов динамики и по отклонениям от трендов, сделайте вывод о тесноте связи между переменными Y_t и X_t .

t	1	2	3	4	5	6	7	8	9	10
Y_t	325,69	340,79	349,39	373,59	389,79	399,09	421,49	441,39	458,29	472,33
X_t	15,69	16,69	17,69	19,09	20,79	21,69	23,09	24,09	25,19	26,56

2. Используя исходные данные задания 1, постройте уравнение регрессии по первым разностям, сделайте вывод.

3. Осуществите прогноз на один шаг вперед, используя уравнение регрессии по первым разностям, полученным в предыдущем примере.

4. По данным за 10 месяцев определите тесноту связи между прибылью предприятия (Y_t , тыс. руб.) и затратами на рекламу (X_t , тыс. руб.).

t	1	2	3	4	5	6	7	8	9	10
Y_t	250	305	314	338	354	363	375	376	377	380
X_t	259,7	294,1	316,3	333,1	346,7	358,2	368,2	377,2	385,3	392,6

5. Используя исходные данные задания 4, постройте уравнение регрессии по отклонениям от трендов, осуществите прогноз на один шаг вперед, сделайте вывод.

6. По промышленному предприятию имеются поквартальные данные за три года об уровне производительности труда (Y_t , тыс. руб. на одного работника) и доле активной части основных фондов (X_t , %).

t	1	2	3	4	5	6	7	8	9	10	11	12
Y_t	5	6	6	7	8	10	11	11	13	12	13	15
X_t	19,9	18,9	19,8	27,9	22,2	29,7	38,7	36	46	37,8	45	54

Постройте модель регрессии с включением в нее фактора времени. Сделайте вывод.

8. Имеются данные по предприятию за 12 лет: Y_t — уровень рентабельности предприятия, %, и X_t — коэффициент оборачиваемости активов.

t	1	2	3	4	5	6	7	8	9	10	11	12
Y_t	9	11,5	14	16,8	18	18,3	18,5	19	19,2	20,8	22	24,4
X_t	2	2,6	3	3,4	4	4,1	4,2	4,4	4,6	5	6	7,1

Постройте уравнение регрессии по исходным данным, проверив его статистическую значимость. Выделите ряд остатков, проверив его на наличие (отсутствие) автокорреляции.

Сделайте вывод о коинтеграции временных рядов.

ЗАДАНИЯ ДЛЯ ЛАБОРАТОРНЫХ РАБОТ

Задание 1. Выявление и анализ основной тенденции временного ряда

На основе данных прил. 5 (табл. П5.1) соответствующего варианта:

- 1) определите наличие тенденции в исследуемом временном ряду с помощью метода Фостера — Стюарта и критериев серий;
- 2) выделите основную тенденцию развития исследуемого ряда с помощью простой и взвешенной скользящей средней.

Задание 2. Моделирование и прогнозирование основной тенденции развития ряда динамики

По исследуемому временному ряду соответствующего варианта:

- 1) выберите тип кривой роста на основе графического метода, метода последовательных разностей;
- 2) определите параметры выбранной модели на основе метода наименьших квадратов;
- 3) проверьте правильность выбранного уравнения тренда на основе: минимизации суммы квадратов отклонений фактических данных от теоретических;
- 4) сделайте точечный и интервальный прогнозы на 2—3 периода упреждения на основе выбранного уравнения тренда;
- 5) оцените точность полученной модели при помощи средней ошибки аппроксимации.

Задание 3. Статистический анализ и моделирование случайной компоненты

По данным соответствующего варианта на основе полученного в задании 2 уравнения тренда:

- 1) определите отклонения расчетных значений, полученных по уравнению тренда, от фактических;
- 2) проверьте, адекватна ли выбранная модель исследуемому процессу (проверка гипотез о случайности, независимости, нормальности распределения случайной компоненты);
- 3) постройте автокорреляционную функцию ряда остатков;
- 4) сформулируйте выводы.

Задание 4. Прогнозирование одномерных временных рядов

По данным любого статистического ежегодника выберите ряд динамики не менее 20 уровней и выполните следующие действия:

- 1) проанализируйте характер тенденции исходного временного ряда, используя графический метод: проверьте предпосылки и обоснуйте возможность построения моделей кривых роста;
- 2) для ряда отклонений (после проверки необходимых гипотез) постройте модель авторегрессии 1-го или 2-го порядка;

3) постройте общий точечный и интервальный прогноз на 2—3 периода упреждения.

Задание 5. Прогнозирование с помощью адаптивных методов

По данным любого статистического ежегодника выберите временной ряд, состоящий не менее чем из 20 уровней, и выполните следующие действия:

1) постройте точечный и интервальный прогнозы на 2—3 периода упреждения методом простого экспоненциального сглаживания, предварительно с помощью средней квадратической ошибки выбрав параметр адаптации;

2) постройте точечный и интервальный прогноз на 2—3 периода упреждения методом экспоненциального сглаживания с применением трэкинг-сигнала;

3) оцените точность полученных моделей.

Задание 6. Моделирование и прогнозирование сезонных колебаний временного ряда

По данным прил. 5 (табл. П5.2) постройте модель сезонной волны, выполнив следующие действия:

1) изобразите графически исходные данные соответствующего варианта и произведите их визуальный анализ;

2) проверьте исходный ряд на наличие тенденции;

3) постройте автокорреляционную функцию для ряда исходных данных (и/или для ряда остатков от тренда);

4) определите индексы сезонности и постройте индексные модели прогноза, оценив точность моделей, выберите наилучшую;

5) определите абсолютные отклонения фактических данных от расчетных, полученных по наилучшей модели;

6) проверьте случайность характера этих отклонений на основе критериев серий;

7) определите гармонику Фурье, наилучшим образом отражающую периодичность изменения уровней ряда;

8) постройте адаптивную модель сезонных колебаний;

9) оцените точность моделей, полученных в п. 7 и 8;

10) сделайте прогноз на 2—3 мес. (квартала) следующего года, оценив результаты по различным моделям.

Задание 7. Моделирование и прогнозирование связанных временных рядов

По данным любого статистического ежегодника или сайта www.gks.ru выберите 2 временных ряда, характеризующих взаимосвязанные явления и постройте модель связанных рядов динамики, выполнив следующие действия:

1) определите результативные и факторные признаки, постройте

график зависимости резульативного признака от каждого из факторных (поле корреляции);

2) проверьте временные ряды на наличие автокорреляции в уровнях;

3) проверьте ряды динамики на наличие автокорреляции в остатках (в случае ее наличия в уровнях);

4) постройте регрессионную модель (можно методом Фриша — Боу);

5) проверьте значимость коэффициентов регрессии на основе критерия Стьюдента;

6) сформулируйте выводы.

1. Айвазян С. А. Прикладная статистика и основы эконометрики / С. А. Айвазян, В. С. Мхитарян. — М. : Юнити, 1998. — 1022 с.
2. Андерсен Т. Статистический анализ временных рядов / Т. Андерсен. — М. : Мир, 1976. — 155 с.
3. Афанасьев В. Н. Анализ временных рядов и прогнозирование / В. Н. Афанасьев, М. М. Юзбашев. — М. : Финансы и статистика, 2001. — 228 с.
4. Афанасьев В. Н. Моделирование и прогнозирование временных рядов / В. Н. Афанасьев, Т. В. Лебедева. — М. : Финансы и статистика, 2009. — 292 с.
5. Афанасьев В. Н. Статистические методы прогнозирования в экономике / В. Н. Афанасьев, Т. В. Лебедева. — М. : Финансы и статистика, 2009. — 180 с.
6. Бахметова Г. Ш. Методы демографического прогнозирования / Г. Ш. Бахметова. — М. : Финансы и статистика, 1982. — 159 с.
7. Бестужев-Лада И. В. Окно в будущее. Современные проблемы социального прогнозирования / И. В. Бестужев-Лада. — М. : Мысль, 1970. — 80 с.
8. Бокс Дж. Анализ временных рядов. Прогноз и управление. Вып. 1 / Дж. Бокс, Г. Дженкинс. — М. : Мир, 1974. — 406 с.
9. Боровиков В. П. Прогнозирование в системе Statistica в среде Windows / В. П. Боровиков. — М. : Финансы и статистика, 2006. — 368 с.
10. Бутакова М. М. Экономическое прогнозирование : методы и приемы практических расчетов / М. М. Бутакова. — М. : КноРус, 2010. — 167 с.
11. Вайну Я. Я. Корреляция рядов динамики / Я. Я. Вайну. — М. : Статистика, 1977. — 119 с.
12. Венсель В. В. Интегральная корреляция: статистическое моделирование рядов динамики / В. В. Венсель. — М. : Финансы и статистика, 1983. — 223 с.
13. Дуброва Т. А. Прогнозирование социально-экономических процессов / Т. А. Дуброва. — М. : Маркет ДС, 2010. — 189 с.
14. Дуброва Т. А. Статистические методы прогнозирования в экономике / Т. А. Дуброва, М. Ю. Архипова. — М. : Издат. центр ЕАОИ, 2008. — 136 с.
15. Закс Н. Статистическое оценивание / Н. Закс. — М. : Статистика, 1976. — 598 с.
16. Езекиэл М. Методы корреляции и регрессии / М. Езекиэл, К. Фокс. — М. : Статистика, 1966. — 558 с.
17. Иващенко Т. А. Статистическое изучение основной тенденции и взаимосвязи в рядах динамики / Т. А. Иващенко, Т. С. Кильдишев, Р. А. Шмойлова. — Томск : Изд-во Том. ун-та, 1985. — 168 с.
18. Кильдишев Т. С. Анализ временных рядов и прогнозирование / Т. С. Кильдишев, А. А. Френкель. — М. : Статистика, 1973. — 103 с.
19. Ковалева Л. Н. Многофакторное прогнозирование на основе рядов динамики / Л. Н. Ковалева. — М. : Статистика, 1980. — 103 с.
20. Корхин А. С. Моделирование экономических систем с распределенным лагом / А. С. Корхин. — М. : Финансы и статистика, 1981. — 160 с.

21. Левин В. С. Инвестиции в основной капитал в России : статистический анализ динамики и прогнозирование / В. С. Левин. — М. : Финансы и статистика, 2007. — 304 с.
22. Лизер Р. Эконометрические методы краткосрочного прогнозирования / Р. Лизер. — М. : Статистика, 1979. — 153 с.
23. Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования / Ю. П. Лукашин. — М. : Статистика, 2002. — 254 с.
24. Льюис Х. Д. Методы прогнозирования экономических показателей / Х. Д. Льюис. — М. : Статистика, 1986. — 240 с.
25. Маленво Э. Статистические методы эконометрии. Вып. 1 / Э. Маленво. — М. : Статистика, 1975. — 423 с.
26. Маленво Э. Статистические методы эконометрии. Вып. 2 / Э. Маленво. — М. : Статистика, 1976. — 325 с.
27. Статистическое моделирование и прогнозирование / под ред. А. Г. Гранберга. — М. : Финансы и статистика, 1990. — 382 с.
28. Статистический анализ экономических временных рядов и прогнозирования. — М. : Наука, 1973. — 296 с. — (Учен. зап. по статистике ; т. 22—23).
29. Теория прогнозирования и принятия решений / под ред. С. Л. Саркисяна. — М. : Высш. шк., 1977. — 158 с.
30. Френкель А. А. Прогнозирование производительности труда: методы и модели / А. А. Френкель. — М. : Экономика, 1989. — 214 с.
31. Ханк Д. Э. Бизнес-прогнозирование / Д. Э. Ханк. — М. : Вильямс, 2003. — 656 с.
32. Четыркин Е. М. Статистические методы прогнозирования / Е. М. Четыркин. — М. : Статистика, 1977. — 200 с.
33. Чураков Е. П. Прогнозирование эконометрических рядов / Е. П. Чураков. — М. : Финансы и статистика, 2008. — 208 с.
34. Швырков В. В. Моделирование внутригодичных колебаний и спроса / В. В. Швырков, Т. С. Швыркова. — М. : Статистика, 1973. — 175 с.
35. Юзбашев М. М. Статистический анализ тенденций и колеблемости / М. М. Юзбашев, А. И. Манелля. — М. : Финансы и статистика, 1983. — 207 с.
36. Эконометрика : учебник / под ред. И. И. Елисеева. — М. : Финансы и статистика, 2006. — 576 с.
37. Эконометрика : учебник / под ред. В. С. Мхитаряна. — М. : Проспект, 2008. — 384 с.

Экономико-статистические ресурсы Интернета

1. www.gks.ru — Федеральная служба государственной статистики.
2. www.cbr.ru — Центральный Банк Российской Федерации.
3. www.minfin.ru — Министерство финансов Российской Федерации.
4. ac.gov.ru — Аналитический центр при Правительстве Российской Федерации.
5. www.rbk.ru — информационное агентство «РосБизнесКонсалтинг».
6. www.akm.ru — информационное агентство АК&М.
7. www.cemi.rssi.ru — Центральный экономико-математический институт РАН.
8. www.akdi.ru — Агентство консультаций и деловой информации (АКДИ).
9. www.forecast.ru — Центр макроэкономического анализа и прогнозирования.
10. rts.micex.ru — Московская биржа.
11. www.fedstat.ru — Единая межведомственная информационно-статистическая система (ЕМИСС).

Приложение 1. Этапы проведения экспертного опроса с применением методики эвристического прогнозирования.

Приложение 2. Список тем.

Приложение 3. Значения нижней и верхней границ критерия Дарбина—Уотсона при 5% уровне значимости.

Приложение 4. Значения K^* для оценки доверительных интервалов прогноза при доверительной вероятности 0,997.

Приложение 5. Варианты заданий для лабораторный работ.

ПРИЛОЖЕНИЕ 1

Этапы проведения экспертного опроса с применением методики эвристического прогнозирования

1. *Составляется анкета экспертного опроса.* Вопросы анкеты содержат:

- анкетные данные эксперта (год рождения, образование, специальность, занимаемая должность, стаж работы в занимаемой должности, общий стаж работы);

- индивидуальный уровень осведомленности эксперта по предложенным разделам (дисциплинам) по десятибалльной шкале;

- перечень строго сформулированных вопросов, которые должны быть даны в общепринятых терминах; исключать всякую смысловую неоднозначность; должны логически соответствовать структуре объекта прогноза.

Информационным массивом для разработки прогнозов методом эвристического прогнозирования является набор заполненных экспертами таблиц и анкет.

2. *По результатам опроса экспертов формируется матрица ответов экспертов*, в которой первая строка соответствует ответам первого эксперта на все вопросы по обследуемому объекту (вопросы, на которые эксперт не отвечал, также записаны в этой строке); вторая — ответам на вопросы второго эксперта и т. д. (табл. П1.1). Внутри каждого вопроса может быть несколько подвопросов или альтернатив, которые должен оценить эксперт.

3. Определяется матрица специализации экспертов, в которой отображается уровень осведомленности эксперта (табл. П1.2).

В табл. П1.2 введены следующие обозначения: n — число экспертов; S — число рассматриваемых специализаций; Q_{fd} — условный балл, характеризующий уровень осведомленности эксперта с номером $d = 1, 2, \dots, n$ специальности $f = 1, 2, \dots, S$.

Матрица ответов экспертов

Номер эксперта	Номер вопроса						
	1	2			...	m	
	1	2	3	4	j	$M-1$	M
1	a_{11}	a_{12}	a_{13}	a_{14}	a_{1j}	a_{1M-1}	a_{1M}
2							
3							
...
n	a_{n1}	a_{n2}	a_{n3}	a_{n4}	a_{nj}	a_{nM-1}	a_{nM}

Таблица П1.2

Матрица специализации эксперта

	1	2	...	n
1	Q_{11}	Q_{12}		Q_{1n}
2	Q_{21}	Q_{22}		Q_{2n}
...				
S	Q_{S1}	Q_{S2}		Q_{Sn}

4. Для установления наибольшей компетентности в конкретных вопросах строится матрица предпочтительности специализации эксперта (табл. П1.3).

Таблица П1.3

Матрица предпочтительности специализации эксперта

	H_1	H_2	...	H_m
1	h_{11}	h_{12}		h_{1m}
2	h_{21}	h_{22}		h_{2m}
...
S	h_{S1}	h_{S2}		h_{Sm}

В табл. П1.3 введены следующие обозначения: m – число вопросов; S – число рассматриваемых специализаций; h_{kl} – степень предпочтения специализации номера $k = 1, 2, \dots, S$ для вопроса с номером $l = 1, 2, \dots, m$. Для оценки степени предпочтительности h_{kl} каждого вопроса можно воспользоваться шкалой, в соответствии с которой, если для данного вопроса эта специализация эксперта:

- весьма существенна, то $h_{kl} = 2$;
- полезна, то $h_{kl} = 1$;
- незначительна, то $h_{kl} = 0$.

5. Далее строится матрица компетентности экспертов (табл. П1.4).

Матрица компетентности экспертов

	H_1	H_2	...	H_m
1	p_{11}	p_{12}		p_{1m}
2	p_{21}	p_{22}		p_{2m}
...
n	p_{S1}	p_{S2}		p_{Sm}

При этом коэффициенты компетентности экспертов p вычисляются по формуле

$$p_{zl} = \frac{Q_{1z}h_{1l} + Q_{2z}h_{2l} + \dots + Q_{Sz}h_{Sl}}{h_{1l} + h_{2l} + \dots + h_{Sl}}, \quad (9.2)$$

где Q_{Sl} – условный балл, характеризующий уровень осведомленности эксперта с номером z ($z = 1, 2, \dots, n$) в специализации с номером d ($d = 1, 2, \dots, S$);

h_{kl} – степень предпочтительности специализации с номером k ($k = 1, 2, \dots, S$) для вопроса с номером l ($l = 1, 2, \dots, m$).

В матрице компетентности, если эксперт не отвечал на поставленный вопрос, то его вес должен быть равен нулю.

6. Для анализа объекта в целом необходимо рассчитывать по каждому подвопросу средневзвешенные оценки по ответам экспертов. По каждому вопросу желательно знать распределение мнений экспертов. Его можно характеризовать различными величинами: дисперсией, коэффициентом асимметрии, эксцессом и т. д., для чего можно построить гистограммы. В качестве меры точности экспертных методов используется усредненное значение коэффициента вариации.

7. Из дальнейшего анализа исключаются те факторы, которые на основе средневзвешенной оценки экспертов не являются существенными, и добавляются в дальнейший анализ факторы, которые были предложены экспертами и не учтены разработчиками.

ПРИЛОЖЕНИЕ 2

Список тем

1. Методы прогнозирования производства сельскохозяйственной продукции (молочной, животноводческой и др.)
2. Моделирование и прогнозирование спроса в розничной торговле.
3. Моделирование и прогнозирование в социальной сфере.
4. Прогнозирование развития региона.
5. Моделирование и прогнозирование спроса населения.
6. Внутрифирменное прогнозирование.
7. Методы прогнозирования возможного банкротства предприятия.
8. Методы прогнозирования финансовых результатов в отраслях экономики (промышленности, сельском хозяйстве, транспорте, связи и т. д.).
9. Методы прогнозирования в исследовании фондового рынка.

10. Методы прогнозирования конъюнктуры рынка ценных бумаг.
11. Методы прогнозирования развития страхового рынка региона.
12. Методы прогнозирования спроса на страховые услуги.
13. Моделирование и прогнозирование деловой активности биржевых структур.
14. Прогнозирование заболеваемости населения.
15. Методы прогнозирования развития рынка жилья.
16. Моделирование и прогнозирование миграционных процессов.
17. Модели прогноза деловой активности рынка ценных бумаг.
18. Моделирование и прогнозирование демографических процессов (численности населения, рождаемости, смертности, продолжительности жизни и т. д.).
19. Прогнозирование и планирование в агропромышленном производстве.
20. Методы моделирования и прогнозирования уровня жизни населения.
21. Методы моделирования и прогнозирования качества жизни населения.
22. Методы прогнозирования основных показателей правовой статистики.
23. Моделирование и прогнозирование уровня занятости населения региона.
24. Моделирование и прогнозирование обеспеченности жильем населения региона.
25. Методы прогнозирования урожайности зерновых культур.
26. Прогнозирование спроса на рынке продовольственных товаров.
27. Прогнозирование налоговых поступлений в бюджет региона.
28. Моделирование и прогнозирование розничного товарооборота региона.
29. Прогнозирование выбросов в атмосферный воздух и сбросов в водную среду предприятиями региона.
30. Моделирование и прогнозирование развития рынка продукции.
31. Моделирование и прогнозирование эффективности промышленного производства в регионе.
32. Прогнозирование и планирование закупок сельскохозяйственной продукции.
33. Прогнозирование и планирование капитального строительства на уровне региона (муниципалитета).
34. Моделирование и прогнозирование развития малого бизнеса в регионе.
35. Моделирование и прогнозирование развития рынка ипотечного страхования.
36. Статистический анализ и прогнозирование производительности труда.
37. Прогнозирование и планирование объемов производства продукции на предприятии.
38. Прогнозирование и планирование объемов продаж на предприятии.
39. Статистический анализ и прогнозирование рентабельности предприятия (банка, страховой компании).
40. Методы прогнозирования сбыта сельскохозяйственной продукции.
41. Прогнозирование и планирование прибыли на предприятии.
42. Планирование и прогнозирование потребности в кадрах предприятия (местного самоуправления).

Значения нижней и верхней границ критерия Дарбина — Уотсона
при 5% уровне значимости

Длина ряда (n)	$k = 1$		$k = 2$		$k = 3$	
	$d_{Н}$	$d_{В}$	$d_{Н}$	$d_{В}$	$d_{Н}$	$d_{В}$
10	0,88	1,32	0,70	1,64	0,53	2,02
11	0,93	1,32	0,66	1,60	0,60	1,93
12	0,97	1,33	0,81	1,58	0,66	1,86
13	1,01	1,34	0,86	1,56	0,72	1,82
14	1,05	1,35	0,91	1,55	0,77	1,78
15	1,08	1,36	0,95	1,54	0,82	1,75
16	1,10	1,37	0,98	1,54	0,86	1,73
17	1,13	1,38	1,02	1,54	0,90	1,71
18	1,16	1,39	1,05	1,53	0,93	1,69
19	1,18	1,40	1,08	1,53	0,97	1,68
20	1,20	1,41	1,10	1,54	1,00	1,68
21	1,22	1,42	1,13	1,54	1,03	1,67
22	1,24	1,43	1,15	1,54	1,05	1,66
23	1,26	1,44	1,17	1,54	1,08	1,66
24	1,27	1,45	1,19	1,55	1,10	1,66
25	1,29	1,45	1,21	1,55	1,12	1,66
26	1,30	1,46	1,22	1,55	1,14	1,65
27	1,32	1,47	1,24	1,56	1,16	1,65
28	1,33	1,48	1,26	1,56	1,18	1,65
29	1,34	1,48	1,27	1,56	1,20	1,65
30	1,35	1,49	1,28	1,57	1,21	1,65
31	1,36	1,50	1,30	1,57	1,23	1,65
32	1,37	1,50	1,31	1,57	1,24	1,65
33	1,38	1,51	1,32	1,58	1,26	1,65
34	1,39	1,51	1,33	1,58	1,27	1,65
35	1,40	1,52	1,34	1,58	1,28	1,65
36	1,41	1,52	1,35	1,59	1,29	1,65

Примечание. $d_{Н}$ и $d_{В}$ — нижняя и верхняя границы критерия d ; k — число переменных в модели.

Значения K^* для оценки доверительных интервалов прогноза при доверительной вероятности 0,997

Длина ряда n	Линейный тренд			Параболический тренд		
	Период упреждения l			Период упреждения l		
	1	2	3	1	2	3
7	2,638	2,874	3,139	3,948	5,755	8,152
8	2,463	2,639	2,836	3,459	4,754	6,461
9	2,342	2,478	2,631	3,144	4,124	5,408
10	2,252	2,361	2,482	2,926	3,695	4,698
11	2,182	2,271	2,370	2,763	3,384	4,189
12	2,127	2,201	2,283	2,636	3,148	3,808
13	2,083	2,146	2,215	2,536	2,965	3,516
14	2,046	2,100	2,159	2,455	2,830	3,286
15	2,015	2,062	2,113	2,386	2,701	3,100
16	1,988	2,029	2,073	2,330	2,604	2,950
17	1,965	2,001	2,040	2,280	2,521	2,823
18	1,945	1,977	2,012	2,238	2,451	2,717
19	1,928	1,965	1,987	2,201	2,391	2,627
20	1,911	1,937	1,965	2,169	2,339	2,549
21	1,897	1,921	1,946	2,139	2,293	2,481
22	1,885	1,906	1,929	2,113	2,252	2,422
23	1,873	1,893	1,914	2,090	2,217	2,371
24	1,863	1,880	1,899	2,060	2,185	2,325
25	1,853	1,870	1,887	2,049	2,156	2,284

ВАРИАНТЫ ЗАДАНИЙ ДЛЯ ЛАБОРАТОРНЫХ РАБОТ

Таблица П1

Исходные данные для анализа рядов, не содержащих сезонную компоненту

Год	Вариант задания																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1989	1,65	23,46	0,54	20,86	12,6	3,54	15,48	12,19	23,75	6,5	512	255	170	508	13,4	551	3,4	500	255	6,5	12,3
1990	2,59	14,86	2,16	20,11	18,92	7,81	9,29	8,41	28	7,1	511	255	170	507	14,6	553	3,7	499	255	7,1	14,3
1991	6,18	20,14	5,39	16,41	17,08	12,83	8,26	14,68	33,01	7,4	505	252	168	501	15,2	551	3,8	493	252	7,4	13,0
1992	6,26	21,59	3,48	18,95	15,51	6,73	5,45	8,64	16,78	7,7	502	250	167	498	15,8	549	4,0	490	250	7,7	12,8
1993	6,44	18,98	4,54	21,43	8,97	6,29	10,49	32,94	18,16	7,2	502	250	167	498	14,8	548	3,7	490	250	7,3	13,2
1994	7,16	21,77	7,99	16,54	14,52	15,88	14,47	22,61	20,05	8,4	502	248	167	498	17,2	539	4,3	490	249	8,0	13,7
1995	10,56	20,27	7,95	11,55	12,77	12,27	9,46	45,92	3,18	7,6	502	247	165	493	15,6	538	3,9	485	248	7,6	14,0
1996	10,93	16,86	7,01	14,39	12,96	7,84	8,79	23,63	16,11	7,5	497	250	165	492	15,4	533	3,9	484	247	7,5	15,2
1997	9,53	16,23	9,89	20,66	5,55	10,71	12,96	18,52	21,66	7,7	496	251	166	497	15,8	537	4,0	489	251	7,7	13,8
1998	10,64	18,55	12,35	15,31	11,09	14,6	15,37	36,22	20,16	8,0	501	255	167	500	16,4	533	4,1	492	255	8,0	14,5
1999	17,43	14,87	12,91	9,34	9,23	17,48	11,82	50,1	24,71	8,2	504	263	170	507	16,8	532	4,2	499	263	8,2	14,8
2000	14,72	11,96	14,42	11,39	5,03	12,97	11,34	46,22	15,63	8,9	511	256	175	523	18,2	525	4,6	515	256	8,9	15,8
2001	15,5	14,41	14,13	11,34	2,15	11,34	20,84	23,63	16,27	8,1	522	255	169	510	16,6	516	4,2	502	255	8,1	17,3
2002	15,01	13,42	18,67	10,07	8,95	23,82	16,58	47,3	18,99	8,5	514	253	172	508	17,4	520	4,4	500	255	8,5	17,6
2003	17,83	10,44	16,95	5,95	8,04	19,97	12,47	40,03	21,12	8,7	512	258	174	504	17,8	522	4,5	496	258	8,7	17,0
2004	18,43	8,26	15,84	4,59	5,68	11,51	7,05	56,53	8,34	8,9	508	261	174	513	18,2	535	4,6	505	263	8,9	15,2
2005	17,69	8,86	19,23	8,74	0,14	18,07	15,08	38,41	14,96	9,3	517	262	176	520	19,0	519	4,8	512	264	9,3	15,8
2006	19,8	9,53	22,05	9,96	5,85	22,11	16,97	51,47	17,7	10,2	524	264	174	521	20,8	512	5,2	513	262	10,2	16,0
2007	22,64	6,88	22,59	3,03	4,21	23,12	13,51	46,29	20,21	10,3	525	262	176	525	21,0	509	5,3	517	260	10,3	15,2
2008	22,86	4,1	21,15	3,17	2,56	15,52	13,45	35,41	8,31	10,0	526	265	179	521	20,4	504	5,1	513	266	10,0	15,6
2009	21,56	7,61	23,98	4,45	0,08	20,02	16,55	67,79	12,36	9,1	529	269	180	526	18,6	505	4,7	515	268	9,5	16,2
2010	22,16	4,92	26,45	4,06	3,87	24,36	18,47	74,21	14,59	9,3	540	270	180	536	19,0	510	4,8	505	260	9,4	16,4
2011	25,82	1,79	29,8	0,16	1,1	27,02	21,76	79,12	21,72	9,2	545	272	182	537	18,8	510	4,7	507	258	9,6	16,1
2012	26,5	0,1	27,41	1,52	0,85	21,31	14,04	45,1	28,69	8,9	543	271	181	541	18,2	519	4,6	510	263	9,9	16,7

Таблица П2

Исходные данные для анализа сезонных временных рядов

Год	Квар- тал	Вариант задания																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2009	I	28,0	336,0	11,2	35,0	92,6	16,7	8,1	345	61,0	31,1	67,5	17,6	25,6	4,1	9,3	9,7	97,3	90,8	40,0	11,2
	II	27,3	327,0	10,9	34,1	83,1	15,7	10,5	356	60,2	30,3	63,2	14,3	19,2	3,1	7,4	5,7	77,6	91,0	32,0	7,2
	III	27,6	331,2	11,0	34,5	76,6	14,9	10,4	334	60,9	30,7	63,1	11,4	24,0	3,8	5,9	8,7	62,6	90,0	25,6	10,2
	IV	29,6	355,8	11,9	37,1	84,8	16,9	13,1	311	65,0	32,9	64,2	15,7	32,0	5,1	8,1	13,7	85,6	97,9	35,2	15,2
2010	I	29,7	343,2	11,4	35,8	83,8	16,6	8,6	253	62,9	31,8	66,2	17,1	28,8	4,6	8,9	11,7	93,6	64,1	38,4	13,2
	II	28,6	333,6	11,1	34,8	75,3	15,2	10,6	257	61,3	30,9	62,0	13,6	22,4	3,6	7,0	7,7	73,8	66,3	30,4	9,2
	III	27,8	337,2	11,2	35,1	68,2	13,6	11,1	264	61,9	31,2	61,8	10,7	27,2	4,4	5,6	10,7	58,4	70,9	24,0	12,2
	IV	28,1	362,4	12,1	37,8	78,4	16,2	13,8	251	66,1	33,6	63,3	14,3	35,2	5,6	7,4	15,7	77,8	101,1	32,0	17,2
2011	I	30,2	350,4	11,7	36,5	76,4	16,5	10,3	267	64,1	32,4	65,3	15,7	32,0	5,1	8,1	13,7	85,6	88,2	35,2	15,2
	II	29,2	341,4	11,4	35,6	65,6	14,7	13,1	267	62,6	31,6	61,1	12,1	24,0	3,8	6,3	8,7	66,1	87,5	27,2	10,2
	III	28,5	342,6	11,4	35,7	59,2	13,4	14,1	258	62,8	31,7	60,5	10,0	30,4	4,9	5,2	12,7	54,5	86,5	22,4	14,2
	IV	28,6	369,0	12,3	38,4	70,1	16,0	15,4	262	67,0	34,2	62,1	12,9	38,4	6,2	6,7	17,7	70,7	107,1	28,8	19,1
2012	I	30,8	355,8	11,9	37,1	71,4	16,3	15,1	354	65,0	32,9	64,2	14,6	35,2	5,6	7,4	15,7	77,8	100,1	32,0	17,2
	II	29,7	349,2	11,6	36,4	61,0	14,1	17,4	365	63,9	32,3	60,1	10,7	25,6	4,1	5,6	9,7	58,3	106,4	24,0	11,2
	III	29,1	349,8	11,7	36,4	61,6	13,1	16,4	383	64,0	32,4	59,4	8,6	32,0	5,1	4,4	13,7	46,0	104,3	19,2	12,2
	IV	29,2	375,6	12,5	39,1	69,8	15,9	17,9	368	68,3	34,8	60,9	11,4	20,0	6,4	5,9	18,7	62,3	130,4	25,6	20,2

ПРЕДИСЛОВИЕ	3
1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРОГНОЗИРОВАНИЯ В ЭКОНОМИКЕ	4
1.1. Понятие категорий «прогноз» и «прогнозирование». Классификация прогнозов и методов прогнозирования	4
1.2. Этапы прогнозирования социально-экономических явлений и процессов	13
1.3. Оценка точности статистических моделей прогнозирования. Верификация прогнозов	15
Тест и задания для самоконтроля	18
2. ВРЕМЕННЫЕ РЯДЫ И ИХ ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ	22
2.1. Понятие и виды временных рядов	22
2.2. Требования, предъявляемые к временным рядам	24
2.3. Основные показатели динамики экономических явлений	26
2.4. Компонентный состав временного ряда	29
Тест и задания для самоконтроля	34
3. АНАЛИЗ И МОДЕЛИРОВАНИЕ ОСНОВНОЙ ТЕНДЕНЦИИ ВРЕМЕННОГО РЯДА	39
3.1. Виды тенденций и проверка гипотезы о существовании основной тенденции	39
3.2. Методы выявления основной тенденции временного ряда	45
3.2.1. Механическое сглаживание	46
3.2.2. Аналитическое выравнивание временных рядов	49
3.3. Критерии адекватности моделей тренда	61
Тест и задания для самоконтроля	65
4. ПРОГНОЗИРОВАНИЕ ДИНАМИКИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ	73
4.1. Простейшие приемы экстраполяции	73
4.2. Прогнозирование на основе моделей кривых роста	76
4.3. Прогнозирование рядов динамики, не имеющих тенденции	84
Тест и задания для самоконтроля	86
5. АДАПТИВНЫЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ	91
5.1. Сущность адаптивных методов	91
5.2. Метод экспоненциального сглаживания	92
5.3. Метод экспоненциального сглаживания с трэкинг-сигналом	99
Тест и задания для самоконтроля	102
6. АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ С ПЕРИОДИЧЕСКИМИ КОЛЕБАНИЯМИ	104
6.1. Методы выявления периодической компоненты	104
6.2. Статистический анализ сезонности	107
6.3. Прогнозирование на основе моделей с учетом индексов сезонности	113
6.4. Моделирование сезонных колебаний при помощи гармоник Фурье	119
6.5. Адаптивные модели сезонных колебаний	123
6.6. Фиктивные переменные в моделировании сезонных колебаний	128
Тест и задания для самоконтроля	132

7. МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ С ПОМОЩЬЮ ПРОИНТЕГРИРОВАННОЙ МОДЕЛИ АВТОРЕГРЕССИИ И СКОЛЬЗЯЩЕГО СРЕДНЕГО (ARIMA-МОДЕЛИ)	137
7.1. Модели стационарных временных рядов	139
7.2. ARIMA-модели для описания нестационарных временных рядов	144
7.3. ARIMA-модели для описания временных рядов с сезонной составляющей	146
7.4. Прогнозирование на основе ARIMA-моделей	148
Тест и задания для самоконтроля	160
8. МНОГОФАКТОРНОЕ МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ	163
8.1. Типы информации, используемые при построении многофакторных моделей	164
8.2. Простая корреляция и регрессия рядов динамики	165
8.3. Множественная корреляция и регрессия	171
Тест и задания для самоконтроля	173
ЗАДАНИЯ ДЛЯ ЛАБОРАТОРНЫХ РАБОТ	176
БИБЛИОГРАФИЧЕСКИЙ СПИСОК	179
ПРИЛОЖЕНИЯ	181
1. АНАЛИЗ ВРЕМЕННЫХ РЯДОВ	181
1.1. Введение	181
1.2. Методы анализа временных рядов	181
1.2.1. Методы анализа временных рядов	181
1.2.2. Методы анализа временных рядов	181
1.2.3. Методы анализа временных рядов	181
1.3. Критерии качества модели	181
Тест и задания для самоконтроля	181
2. ПРОГНОЗИРОВАНИЕ ДИНАМИКИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ	181
2.1. Прогнозирование динамики социально-экономических процессов	181
2.2. Прогнозирование динамики социально-экономических процессов	181
2.3. Прогнозирование динамики социально-экономических процессов	181
Тест и задания для самоконтроля	181
3. АДАПТИВНЫЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ	181
3.1. Основы адаптивных методов	181
3.2. Метод адаптивных временных рядов	181
3.3. Метод адаптивных временных рядов с учетом сезонности	181
Тест и задания для самоконтроля	181
4. АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ С ПЕРИОДИЧЕСКИМИ КОМПОНЕНТАМИ	181
4.1. Методы анализа временных рядов с периодическими компонентами	181
4.2. Статистические методы анализа временных рядов с периодическими компонентами	181
4.3. Прогнозирование временных рядов с периодическими компонентами	181
Тест и задания для самоконтроля	181
5. АДАПТИВНЫЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ	181
5.1. Основы адаптивных методов	181
5.2. Метод адаптивных временных рядов	181
5.3. Метод адаптивных временных рядов с учетом сезонности	181
Тест и задания для самоконтроля	181
6. АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ С ПЕРИОДИЧЕСКИМИ КОМПОНЕНТАМИ	181
6.1. Методы анализа временных рядов с периодическими компонентами	181
6.2. Статистические методы анализа временных рядов с периодическими компонентами	181
6.3. Прогнозирование временных рядов с периодическими компонентами	181
6.4. Методы анализа временных рядов с периодическими компонентами	181
6.5. Адаптивные методы анализа временных рядов с периодическими компонентами	181
6.6. Функциональные методы анализа временных рядов с периодическими компонентами	181
Тест и задания для самоконтроля	181

Учебное издание

САЖИН Юрий Владимирович
КАТЫНЬ Альбина Валерьевна
САРАЙКИН Юрий Владимирович

АНАЛИЗ ВРЕМЕННЫХ РЯДОВ И ПРОГНОЗИРОВАНИЕ

Учебник

Редактор *Л. Н. Вечканова*
Технический редактор *Т. А. Сальникова*
Дизайн обложки *О. В. Ребрушкиной*
Компьютерная верстка *Т. Н. Аношкиной*
Корректор *С. П. Славкина*

Подписано в печать 25.11.13. Формат 60×84 1/16.
Тираж 200 экз. Заказ № 1878.
Издательство «Мир»
Тел.: (495) 796-0841, 796-0842, 796-0843
www.mir.ru