

Н. В. Мамонова

**ВРЕМЕННЫЕ РЯДЫ:
АНАЛИЗ И ПРОГНОЗИРОВАНИЕ**

Учебное пособие

Министерство науки и высшего образования Российской Федерации
Байкальский государственный университет

Н. В. Мамонова

**ВРЕМЕННЫЕ РЯДЫ:
АНАЛИЗ И ПРОГНОЗИРОВАНИЕ**

Учебное пособие

Иркутск
Издательство БГУ
2019

УДК 512.64(075.8)
ББК 22.143я7
М22

Печатается по решению редакционно-издательского совета
Байкальского государственного университета

Рецензенты канд. физ.-мат. наук, доц. О. В. Леонова
 канд. техн. наук, доц. О. А. Рогачева

Мамонова Н. В.

М22 Временные ряды: анализ и прогнозирование [Электронный ресурс] :
учеб. пособие / Н. В. Мамонова. – Иркутск : Изд-во БГУ, 2019. – 100 с. –
Режим доступа: <http://lib-catalog.bgu.ru>.

Содержатся теоретические сведения, задачи, упражнения, контрольные вопросы и задания по анализу временных рядов. Излагаемый материал охватывает основные разделы программы курса «Анализ временных рядов и прогнозирование», который изучается после курса «Эконометрика» на экономических специальностях бакалавриата. Кроме необходимых теоретических сведений, приводятся примеры решения стандартных задач.

Рекомендуется для студентов всех экономических специальностей бакалавриата БГУ.

УДК 512.64(075.8)
ББК 22.143я7

© Мамонова Н. В., 2019
© Издательство БГУ, 2019

Оглавление

Введение	4
1. Моделирование тенденции временного ряда	7
1.1. Стационарность, автокорреляция	9
1.2. Модели тенденции развития	11
1.3. Моделирование тенденций с помощью сглаживания	17
1.4. Оценка параметров уравнения тренда	19
1.5. Верификация моделей тенденции	24
1.6. Тесты на наличие автокорреляции остатков	26
Тестовые задания	31
Контрольные вопросы	34
2. Моделирование периодических колебаний	35
2.1. Метод скользящей средней для моделирования сезонной компоненты	35
2.2. Моделирование сезонных колебаний с помощью фиктивных переменных	45
2.3. Моделирование сезонных колебаний с помощью рядов Фурье	47
Тестовые задания	55
Контрольные вопросы	57
3. Регрессионные модели временных рядов	59
3.1. Учет тенденции при построении модели регрессии	59
3.1.1. Методы исключения тенденции	60
3.1.2. Включение в модель фактора времени	68
3.2. Обобщенный метод наименьших квадратов	72
3.3. Учет сезонности в регрессионных моделях	77
Контрольные вопросы	80
Контрольные задания	80
Приложения	95
Список рекомендуемой литературы	99

Введение

В любой сфере экономической деятельности человека встречаются явления, которые интересно и важно изучать в некоторой динамике, так как они изменяются во времени. С течением времени изменяются цены на товары и услуги, изменяются экономические условия любого производственного процесса и т.п. *Совокупность измерений подобного рода показателей в течение некоторого периода времени называется временным рядом.*

Интерес возникает при изучении поведения временного ряда. К примеру, как предсказать будущее на основании знаний прошлого, пытаться выяснить механизм, лежащий в основе процесса, и управлять им. Важно при этом уметь освобождать временной ряд от компонент, которые затемняют его динамику. Часто требуется сжато представлять характерные особенности ряда.

Временным рядом y_t называют последовательность наблюдений, упорядоченную во времени t , хотя возможно упорядочение и по какому-либо другому параметру. Основной чертой, выделяющей анализ временных рядов среди других видов статистического анализа, является существенность порядка, в котором производятся наблюдения.

Различают два вида временных рядов. Измерение некоторых величин (температуры, напряжения и т.д.) производится непрерывно. При этом наблюдения можно фиксировать в виде графика (рис. 1).

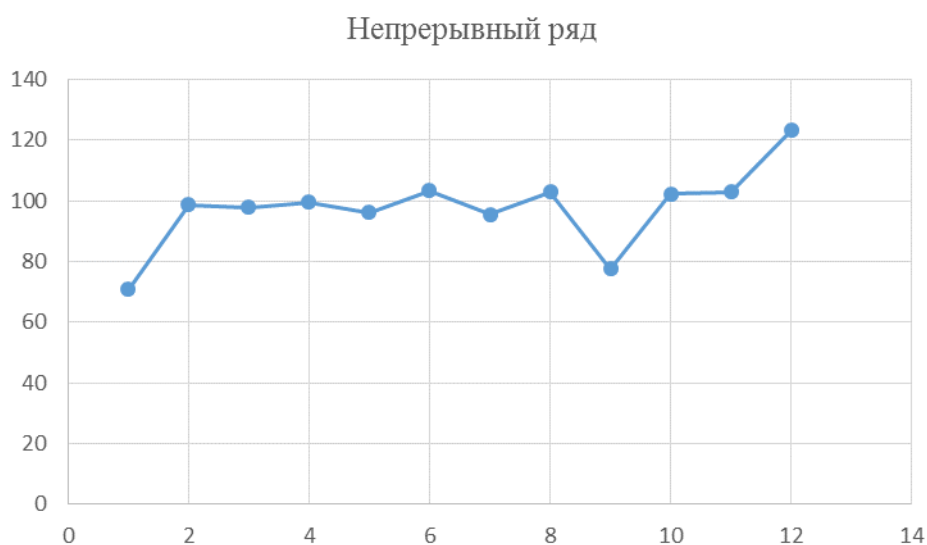


Рис. 1. Непрерывный временной ряд

Но даже в том случае, когда изучаемые величины регистрируются непрерывно, практически при их обработке используются только те значения, которые соответствуют дискретному множеству моментов времени. Следовательно, если время измеряется непрерывно, временной ряд называется *непрерывным*, если же время фиксируется дискретно, т.е. через фиксированный интервал времени, то временной ряд дискретен.

В дальнейшем будем иметь дело только с дискретными временными рядами (табл. 1).

Таблица 1

Дискретный временной ряд	
Месяц	Оборот розничной торговли, у
Январь	70,8
Февраль	98,7
Март	97,9
Апрель	99,6
Май	96,1
Июнь	103,4
Июль	95,5
Август	102,9
Сентябрь	77,6
Октябрь	102,3
Ноябрь	102,9
Декабрь	123,1

Дискретные временные ряды получаются двумя способами:

– выборкой из непрерывных временных рядов через регулярные промежутки времени (например, численность населения, величина собственного капитала фирмы, объем денежной массы, курс акции и т.п.), – такие временные ряды называются *моментными*;

– накоплением переменной в течение некоторого периода времени (например, объем производства какого-либо вида продукции, количество осадков, объем импорта и т.п.), – в этом случае временные ряды называются *интервальными*.

Под анализом временных рядов понимается совокупность математико-статистических методов анализа, предназначенных для выявления структуры данного ряда и для его прогнозирования в будущем. Выявленные структуры временного ряда используются для строительства математической модели исследуемого явления. Прогноз будущих значений временного ряда необходим для эффективного принятия решений.

Временные ряды классифицируются по следующим признакам:

1) по форме представления уровней ряда y_t : абсолютные показатели, относительные показатели, средние величины;

2) по количеству показателей, для которых определяются уровни y_t в каждый момент времени: одномерные и многомерные временные ряды;

3) по характеру временного параметра: дискретные и интервальные временные ряды;

4) по расстоянию между датами и интервалами времени: равностоящие и не равностоящие;

5) по наличию пропущенных значений: полные и неполные временные ряды;

6) детерминированные (найденные по неслучайной функции) или случайные (найденные как результат действий случайной величины) временные ряды;

7) стационарные (основные показатели – средние и дисперсии – постоянны) и нестационарные ряды динамики (содержат тенденцию развития).

Методика прогноза данных рядов разнообразна. Однако прогнозные оценки всегда рассчитываются в несколько этапов.

1. Проверяют базовую линию прогноза.

2. Выявляют закономерности прошлого развития явления.

3. Находят оценки степени достоверности выявленной закономерности развития явления в прошлом (подбор трендовой функции).

4. Найденную закономерность переносят на некоторый период будущего, так называемая *экстраполяция*.

5. Корректируют полученный прогноз с учетом содержательного анализа текущего состояния.

Для получения объективного прогноза развития изучаемого явления данные базовые значения должны иметь одинаковый шаг во времени, фиксироваться в одно и то же время суток и не иметь пропущенных показателей.

Если в наблюдениях отсутствуют результаты за незначительный отрезок времени, то для полноты исследования необходимо их восполнить приближенными данными, например, использовать среднее значения соседних отрезков. Корректировка полученного прогноза выполняется для уточнения полученных долгосрочных прогнозов с учетом влияния сезонности или скачкообразности развития изучаемого явления.

1. Моделирование тенденции временного ряда

В эконометрическом исследовании принято моделировать временной ряд как случайный процесс, который называется стохастическим процессом. *Стохастический процесс* – это статистическое явление, развивающееся во времени согласно законам теории вероятностей. *Случайный процесс* – это случайная последовательность простых случайных явлений: $\{..., Y_{-1}, Y_0, Y_1, \dots\}$, протекающая от минус до плюс бесконечности. Временной ряд – это лишь одна частная реализация такого теоретического стохастического процесса: y_1, y_2, \dots, y_t , которую принято называть выборкой.

Возможные значения временного ряда в данный момент времени t описываются с помощью случайной величины y_t и связанного с ней распределения вероятностей $p(y_t)$. Тогда наблюдаемое значение временного ряда y_t в момент t рассматривается как одно из множества значений, которые могла бы принять эта случайная величина.

Случайные процессы классифицируются на детерминированные и случайные процессы (временные ряды). *Детерминированным называется процесс*, который принимает значение с вероятностью, равной единице. *Случайный процесс* – случайный временной ряд – не является детерминированным.

Стохастические процессы делятся на стационарные и нестационарные. Процесс называется *стационарным*, если он находится в определенном смысле в статистическом равновесии, его вероятностные свойства не зависят от времени. Если эти условия нарушаются, то процесс будет *нестационарным*.

Гауссовские процессы – это процессы, в которых любой набор наблюдений имеет совместное нормальное распределение. Временной ряд, как правило, является одномерной величиной. Однако, встречаются модели, в которых надо рассмотреть совместную динамику набора временных рядов $y_t = (y_{1t}, y_{2t}, \dots, y_{kt})$. Такой набор называют *многомерным временным рядом*, или *векторным рядом*.

При анализе экономических временных рядов традиционно различают различные виды эволюции – динамики. Они могут быть локальными – изолированными или комбинированными. Таким образом, временной ряд может быть представлен в виде нескольких составляющих, которые с экономической точки зрения несут различную содержательную нагрузку:

- *тенденция* – это медленное изменение, происходящее в некотором направлении в течение времени t . Тенденцию называют еще *трендом* или *долговременным движением*;

- *циклические колебания* – это более быстрая, чем тренд, периодическая динамика, в которой наблюдаются фазы возрастания и убывания. Как правило, цикл связан с флуктуациями экономической активности;

- *сезонные колебания* – это изменения, которые происходят регулярно в течение года, недели, суток. Они связаны с сезонами и ритмами человеческой активности;

– *календарные эффекты* – это отклонения, связанные с определенными предсказуемыми календарными событиями, например, праздничные дни, високосность года и т.п.;

– *случайные флуктуации* – это беспорядочные движения относительно большой частоты. Они порождаются влиянием разнородных событий на изучаемую величину (случайный эффект). Зачастую такую величину называют *шумом*;

– *выбросы* – это аномальные движения временного ряда, связанные с редко происходящими событиями, которые резко, но кратковременно отклоняют ряд от общего закона, по которому он движется;

– *структурные сдвиги* – это аномальные движения временного ряда, связанные с редко происходящими событиями, имеющие скачкообразный характер и меняющие тенденцию.

В зависимости от рассматриваемых задач, временные ряды могут состоять из конкретных составляющих в чистом виде, либо иметь очень сложную структуру сочетания разных составляющих. Часто для анализа временного ряда оказывается полезным изолированное рассмотрение отдельных компонент.

Для того чтобы можно было разложить конкретный ряд на эти составляющие, требуется сделать какие-то допущения о том, какими свойствами они должны обладать. Желательно построить сначала формальную статистическую модель, которая бы включала в себя в каком-то виде эти составляющие, затем оценить ее, а после этого на основании полученных оценок вычленить составляющие. Однако построение формальной модели является сложной задачей. В частности, из содержательного описания не всегда ясно, как моделировать те или иные компоненты. Например, тренд может быть детерминированным или стохастическим. Аналогично, сезонные колебания можно комбинировать с помощью детерминированных переменных или с помощью стохастического процесса определенного вида. Компоненты временного ряда могут входить в него аддитивно или мультипликативно. Более того, далеко не все временные ряды имеют достаточно простую структуру, чтобы можно было разложить их на указанные составляющие.

Основная задача эконометрического исследования временного ряда заключается в выявлении и придании количественного выражения составляющим его отдельным компонентам. Простые временные ряды, как правило, легко анализируются визуально по графику и наличие той или иной составляющей не сложно определить: явный рост или спад значений, наличие или отсутствие циклических колебаний и др.

Перед построением модели исходные данные проверяются на сопоставимость (применение одинаковой методики получения или расчета данных), однородность (отсутствие случайных выбросов), устойчивость (наличие закономерности в изменении уровней ряда) и достаточность (число наблюдений должно в 7–10 превосходить число параметров модели).

1.1. Стационарность, автокорреляция

Статистический процесс называется *строго стационарным*, если взаимное распределение вероятностей n наблюдений инвариантно (неизменно) по отношению к общему сдвигу временного аргумента, т.е. общая плотность распределения случайных величин y_{t_1}, \dots, y_{t_n} такая же, как для величин $y_{t_1+\kappa}, \dots, y_{t_n+\kappa}$ при любом целом значении сдвига κ .

Требование стационарности, определенное этими условиями, является достаточно жестким. Как правило, при практическом изучении случайных процессов ограничиваются моментами первого и второго порядка, и тогда говорят о *слабой стационарности*. В этом случае процесс имеет постоянные для всех t моменты первого и второго порядков: *среднее* значение $M(y_t) = \mu$, *дисперсию* $\sigma^2 = M(y_t - \mu)^2$, а также *автоковариацию* $\text{cov}_\kappa = M[(y_t - \mu) \cdot (y_{t+\kappa} - \mu)]$. Ковариация между y_t и $y_{t+\kappa}$ зависит только от величины сдвига κ и не зависит от времени t .

Важной особенностью временных рядов по сравнению с обычными наблюдениями, является наличие связи между последовательными уровнями ряда, вызванные действием каких-либо долговременных причин, что приводит к наличию таких составляющих ряда, как долговременная тенденция и периодическая составляющая.

Корреляционная зависимость между последовательными уровнями временного ряда называется *автокорреляцией уровней* временного ряда.

Степень тесноты автокорреляционной связи между уровнями ряда может быть определена с помощью коэффициентов автокорреляции, т.е. коэффициентов линейной корреляции между уровнями исходного временного ряда и уровнями ряда, сдвинутыми на несколько шагов во времени.

Коэффициент автокорреляции порядка κ стационарного процесса с ненулевой дисперсией имеет вид

$$\rho_\kappa = \frac{M[(y_t - \mu) \cdot (y_{t+\kappa} - \mu)]}{\sqrt{M(y_t - \mu)^2 M(y_{t+\kappa} - \mu)^2}}. \quad (1.1)$$

Следует помнить, что два процесса, имеющие одинаковые моменты первого и второго порядка, могут иметь разный характер распределения.

Автокорреляционной функцией (АКФ) называют последовательность коэффициентов автокорреляции $\{\rho_\kappa\}$, где $\kappa \in (-\infty; +\infty)$. Данная функция симметрична относительно 0, причем $\rho_0 = 1$. Поэтому при рассмотрении поведения АКФ ограничиваются лишь областью положительных значений κ . График автокорреляционной функции называется *коррелограммой*. Очевидно, что с увеличением лага κ число пар наблюдений $(n - \kappa)$ уменьшается. Поэтому в практических подсчетах рекомендуется поддерживать соотношение $\kappa \leq n/4$. Для стационарного временного ряда с увеличением κ автокорреляционная функция должна демонстрировать свойство монотонного убывания по абсолютной величине, так как взаимосвязь между уровнями ряда с ростом κ ослабевает.

Анализ автокорреляционной функции и коррелограммы позволяет выявить структуру ряда, т.е. определить присутствие в ряде той или иной компоненты. Так, например, если наиболее высоким оказался коэффициент автокорреляции первого порядка, то исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка m , то ряд содержит циклические колебания с периодичностью в m моментов времени. Если ни один из коэффициентов автокорреляции не является значимым, то ряд не содержит тенденции и циклических колебаний.

Необходимо подчеркнуть, что линейные коэффициенты автокорреляции характеризуют тесноту только линейной связи текущего и предыдущих уровней ряда. Поэтому по коэффициентам автокорреляции можно судить только о наличии или отсутствии линейной (или близкой к линейной) зависимости. Для проверки ряда на наличие нелинейной тенденции рекомендуется вычислить линейные коэффициенты автокорреляции для временного ряда, состоящего из логарифмов исходных уровней. Отличные от нуля значения коэффициентов автокорреляции будут свидетельствовать о наличии нелинейной тенденции.

Пример 1.1. Имеются данные об уровне безработицы y_t (%) за 18 месяцев (табл. 2):

Таблица 2

Месяц	1	2	3	4	5	6	7	8	9
Уровень y_t	1,5	3,2	4,4	2,4	3,8	5,2	3,2	5,4	7,2
Месяц	10	11	12	13	14	15	16	17	18
Уровень y_t	5,2	7,5	10,2	8,3	11,8	13,8	10,5	13,5	14,8

Определить коэффициенты автокорреляции уровней этого ряда, построить коррелограмму. Интерпретировать полученные результаты.

Решение.

Диаграмма рассеивания данного ряда представлена на рис. 2. График показывает наличие тенденции, а также периодических колебаний. Найдем коэффициенты автокорреляции разных порядков. Преобразуя формулу (1.1), получим формулу, удобную для расчетов:

$$\rho_k = \frac{\sum_{t=k+1}^n (y_t - \bar{y}_1) \cdot (y_{t-k} - \bar{y}_2)}{\sqrt{\sum_{t=k+1}^n (y_t - \bar{y}_1)^2 \cdot \sum_{t=k+1}^n (y_{t-k} - \bar{y}_2)^2}}, \quad (1.2)$$

где k – величина сдвига, лага, определяет порядок коэффициента автокорреляции, \bar{y}_1 и \bar{y}_2 – соответствующие средние исходного ряда и ряда со смещением:

$$\bar{y}_1 = \frac{1}{n-k} \sum_{t=k+1}^n y_t, \quad \bar{y}_2 = \frac{1}{n-k} \sum_{t=k+1}^n y_{t-k}.$$

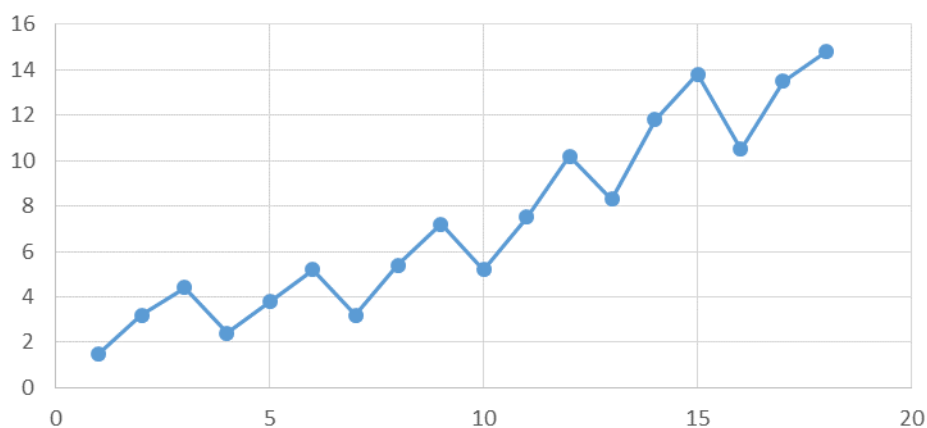


Рис. 2. Диаграмма рассеивания уровня безработицы, %

В результате получаем АКФ: $\rho_1 = 0,859$; $\rho_2 = 0,815$; $\rho_3 = 0,97$; $\rho_4 = 0,725$; $\rho_5 = 0,755$; $\rho_6 = 0,94$; $\rho_7 = 0,646$; $\rho_8 = 0,544$; $\rho_9 = 0,919$. Высокое значение коэффициента автокорреляции первого порядка $\rho_1 = 0,859$ определяет наличие тенденции в данном временном ряду. Вместе с тем, максимальное значение коэффициента автокорреляции при лаге 3 и кратных ему лаге 6 и 9, свидетельствует о регулярном колебании уровней ряда через 3 месяца, уровень безработицы хоть и растет, но каждые три месяца имеет волнообразное изменение.

График коррелограммы представлен на рис. 3.

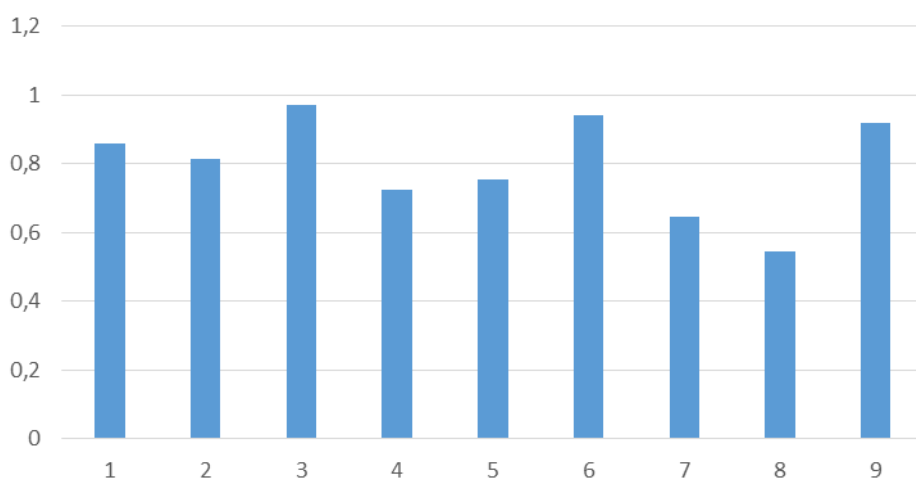


Рис. 3. Коррелограмма уровня безработицы, ρ_k

1.2. Модели тенденции развития

Сравнительно простой моделью временного ряда может служить линейная модель вида:

$$y_t = \mu_t + \varepsilon_t, \quad t \in [1; T], \quad (1.3)$$

где μ_t — полностью детерминированная последовательность или систематическая составляющая, ε_t — последовательность случайных величин, являющаяся

белым шумом¹. Если μ_t зависит от вектора неизвестных параметров θ : $\mu_t = \mu_t(\theta)$, то модель (1.3) является моделью регрессии, и ее параметры можно оценить с помощью метода наименьших квадратов.

Временные ряды, образующие белый шум, удовлетворяют следующим условиям:

$$y_t = \varepsilon_t, \text{ где } M\varepsilon_t = 0; M(\varepsilon_t \cdot \varepsilon_{t+\kappa}) = \begin{cases} \sigma^2, \kappa = 0 \\ 0, \kappa \neq 0 \end{cases}.$$

В качестве примера белого шума можно привести остатки из классической линейной регрессионной модели. В случае нормального распределения остатков они образуют гауссовский белый шум.

Детерминированная компонента μ_t , как правило, сама моделируется как состоящая из нескольких компонент. Например, можно рассмотреть аддитивную модель, в которой временной ряд содержит три компоненты: тренд τ_t , сезонные движения ζ_t и случайные флуктуации ε_t :

$$y_t = \tau_t + \zeta_t + \varepsilon_t. \quad (1.4)$$

Аддитивный временной ряд удобнее всего рассматривать в случае постоянной амплитуды колебаний сезонной компоненты ζ_t .

Однако встречаются экономические временные ряды, в которых лучше использовать мультипликативную схему, как правило, это связано с постепенным возрастанием или затуханием сезонной составляющей:

$$y_t = \tau_t \cdot \zeta_t \cdot \exp(\varepsilon_t). \quad (1.5)$$

Если это выражение прологарифмировать, то получится аддитивный вариант:

$$\ln(y_t) = \ln(\tau_t) + \ln(\zeta_t) + \varepsilon_t,$$

что позволяет оставаться в рамках линейной регрессии и значительно упрощает моделирование.

Рассмотрим основные типы трендовой составляющей τ_t .

Самым простым и очевидным типом тренда является *тренд среднего*, когда временной ряд выглядит как колебания около медленно возрастающей или убывающей величины. Практическая значимость модели тенденции будет тем выше, чем меньше будут остаточные колебания $\varepsilon_t = y_t - \hat{y}_t$.

Второй тип тренда – это *тренд дисперсии*. В этом случае во времени меняется амплитуда колебаний переменной. Иными словами, процесс гетероскедастичен. Встречаются экономические процессы, в которых с ростом среднего, растет и дисперсия.

Третий тип тренда, визуально не всегда наблюдаемый, – имеет изменение величины корреляции между текущим и предшествующим значениями ряда, и называется *тренд автокорреляции*.

¹ Название «белый шум» связано с тем, что спектральная плотность такого процесса постоянная величина, т.е. процесс содержит в одинаковом количестве все частоты, подобно тому как белый цвет содержит в себе все остальные цвета.

Проводя стандартное разложение ряда на компоненты, как правило, под понятием тренда подразумевают изменение среднего уровня переменной, т.е. тренд среднего. В экономике чаще всего используют:

– *полиномиальный* тренд

$$\tau_t = a_0 + a_1 t + \dots + a_p t^p; \quad (1.6)$$

– *экспоненциальный* тренд

$$\tau_t = e^{a_0 + a_1 t + \dots + a_p t^p}; \quad (1.7)$$

– *гиперболический* тренд

$$\tau_t = a + \frac{b}{t}; \quad (1.8)$$

– *гармонический* тренд

$$\tau_t = R \cos(\omega \cdot t + \varphi), \quad (1.9)$$

здесь R – амплитуда колебаний, ω – угловая частота, φ – фаза;

– *логистический* тренд

$$\tau_t = \frac{k}{1 + b e^{-at}}. \quad (1.10)$$

Оценивание параметров полиномиального и экспоненциального трендов несложно производится с помощью обычного МНК.

Гармонический тренд оправдан, когда в составе временного ряда отчетливо прослеживаются периодические колебания. При этом если частота ω известна, то функцию (1.9) можно представить в виде линейной комбинации синуса и косинуса: $\tau_t = \alpha \cos(\omega \cdot t) + \beta \sin(\omega \cdot t)$.

Логистическая кривая нуждается в особом рассмотрении. Функция (1.10) ограничена и имеет горизонтальную асимптоту: $\lim_{t \rightarrow \infty} \tau_t = k$ (рис. 4). В этом преимущество логистической функции перед полиномиальной или экспоненциальной функциями, которые по мере роста $t \rightarrow \infty$ не всегда годятся для прогнозирования. Логистическая кривая наиболее часто используется при изучении социальных и демографических процессов.

Особенностью логистической кривой является нелинейность по оцениваемым параметрам a, b, k , поэтому система уравнений, получаемая с помощью МНК, нелинейна относительно неизвестных параметров и для ее решения могут применяться только численные методы.

Построение модели тенденции включает в себя следующие этапы работы:

- выбор математической функции, описывающей тенденцию;
- оценку параметров модели;
- проверку адекватности выбранной функции и оценку точности модели;
- расчет точечного и интервального прогнозов.

Все многообразие математических функций можно свести в три группы:

1. Функции с монотонным характером возрастания (убывания) и отсутствием пределов роста (снижения).

2. Кривые с насыщением, т.е. устанавливается нижняя или верхняя граница изменения уровней ряда.

3. S – образные кривые, т.е. кривые с насыщением, имеющие точку перегиба.

К первой группе относятся полиномиальные и экспоненциальные тренды (1.6)–(1.7). К кривым с насыщением относят гиперболы вида (1.8). Логистический тренд (1.10) относится к третьей группе кривых.

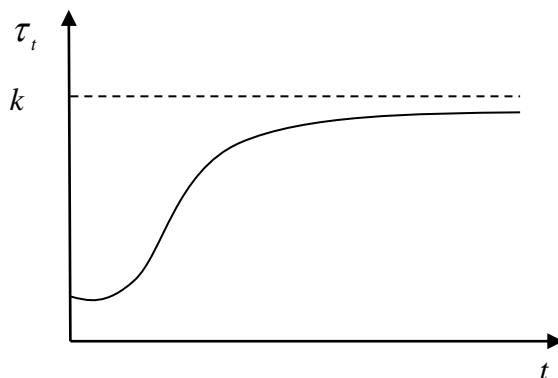


Рис. 4. Логистический тренд

Рассмотрим *линейный* тренд, когда $p = 1$: $\tau_t = a_0 + a_1 t$. По содержанию линейный тренд означает, что уровни временного ряда изменяются с одинаковой скоростью, т.е. с равным абсолютным приростом a_1 . Например, уравнение тренда для индексов потребительских цен за 12 месяцев составило: $\hat{\tau}_t = 99,9 + 1,9t$. Из уравнения очевидно, что за рассматриваемый год цены ежемесячно возрастали в среднем на 1,9 процентных пункта.

Рассмотрим *параболу*, когда $p = 2$: $\tau_t = a_0 + a_1 t + a_2 t^2$. Данная функция рекомендуется для моделирования тенденции, если временной ряд характеризуется постоянным абсолютным ускорением, т.е. постоянными величинами являются вторые разности (приросты абсолютных приростов). В этом случае скорость изменяется линейно (табл. 3).

Таблица 3

Изменение показателей

Время t	Коэффициенты $\tau_t = a_0 + a_1 t + a_2 t^2$	Скорость $\Delta = \tau_t - \tau_{t-1}$	Ускорение $\Delta^2 = \Delta \tau_t - \Delta \tau_{t-1}$
0	a_0	–	–
1	$a_0 + a_1 + a_2$	$a_1 + a_2$	–
2	$a_0 + 2a_1 + 4a_2$	$a_1 + 3a_2$	$2a_2$
3	$a_0 + 3a_1 + 9a_2$	$a_1 + 5a_2$	$2a_2$
4	$a_0 + 4a_1 + 16a_2$	$a_1 + 7a_2$	$2a_2$
5	$a_0 + 5a_1 + 25a_2$	$a_1 + 9a_2$	$2a_2$

Параметр a_0 означает начальный уровень ряда в момент $t = 0$. Параметр a_1 представляет собой средний абсолютный прирост за весь период времени. Параметр a_2 характеризует половину абсолютного ускорения временного ряда.

Например, динамика численности детей в возрасте 7 лет за последние 15 лет в определенном районе характеризуется следующим трендом в тыс. человек: $\hat{\tau}_t = 323,7 + 10,8t - 1,6t^2$. Следовательно, ежегодно численность детей в этом районе сокращается в среднем с ускорением в 3,2 тыс. человек.

Экспоненциальный тренд имеет широкое применение, так как характеризуется стабильным коэффициентом роста. Например, за ряд лет динамика прибыли характеризуется следующим уравнением: $\hat{\tau}_t = e^{2,603+0,405t}$. Следовательно, ежегодно прибыль возрастает в среднем на 50% ($e^{0,405} = 1,5$, т.е. коэффициент роста равен 1,5). Рост по экспоненте означает геометрическую прогрессию уровней временного ряда, и, как следствие, не очень большой период времени, так как ресурсы могут быть ограничены во времени.

Если уровни ряда имеют разную меру пропорциональности во времени, то в качестве тренда можно использовать *степенную* функцию: $\tau_t = at^b$. При $b > 0$ она характеризует непрерывный рост уровней с падающими темпами роста, а при $b < 0$ – их ускоренное снижение. Величина t^b означает базисный коэффициент роста. Поэтому степенная функция практически гарантирует величину среднего коэффициента роста:

$$\bar{K} = \sqrt[b]{t^b}. \quad (1.11)$$

Например, обеспеченность жильем городского населения нашего региона за последние 10 лет (m^2 общей площади на человека) характеризовалась следующим уравнением $y = 15,876 \cdot t^{0,08}$. Следовательно, за весь период обеспеченность населения жильем выросла в 1,202 раза ($10^{0,08}$), т.е. ежегодно она возрастала в среднем на 2,07 %, так как $\bar{K} = \sqrt[10]{1,202} = 1,027$.

Рассмотрим кривые с насыщением – *гиперболу* вида (1.8). При $b > 0$ уровни рассматриваемого ряда снижаются во времени и асимптотически приближаются к параметру a . Например, индексы потребительских цен за пять лет по России изменялись по гиперболе, R^2 – коэффициент детерминации:

$$\tau_t = 95,56 + \frac{87,126}{t}; R^2 = 0,989.$$

Уравнение характеризует падающую тенденцию индекса потребительских цен (ИПЦ), при которой данный индекс не может быть меньше 95,6 %. Тренд описывает 99 % вариации ИПЦ и лишь 1 % ее связан с действиями неучтенных факторов.

Если $b < 0$, то уравнение тренда $\tau_t = a + \frac{b}{t}$ характеризует тенденцию к росту уровней ряда с асимптотической границей равной параметру a . Так, численность мужчин старше трудоспособного возраста в Иркутске за последние 20 лет характеризовалась повышающейся тенденцией: $\tau_t = 296,92 - \frac{89,9}{t}$, из которой следует, что численность мужчин этой возрастной категории за данный период не превышала 296,9 тыс. человек.

Среди кривых с насыщением может использоваться *модифицированная экспонента*:

$$\tau_t = c \pm ab^t, \quad (1.12)$$

где c – асимптота.

Например, при изучении тенденции роста уровня механизации труда целесообразно учитывать ограничение роста. Если изучается динамика детской смертности, то можно установить нижнюю асимптоту, т.е. минимальное значение детской смертности исходя из достигнутых условий жизни. Так, модифицированная экспонента роста уровня механизации труда $\tau_t = 100 - 12,7 \cdot 0,895^t$ означает, что ежегодно скорость ряда снижается в 0,895 раз или на 10,5 %. Верхняя граница уровня механизации труда 100 %.

Модифицированная экспонента служит базовой кривой для других кривых с насыщением, а именно S – образных кривых: *логистической кривой* и *кривой Гомперца*². Тенденция развития явления в S – образных кривых охватывает три этапа: вначале довольно медленный рост, который постепенно убыстряется, далее сменяется уменьшением роста и приближением уровня ряда к предельному значению, т.е. к уровню насыщения.

Логистический тренд вида $\tau_t = \frac{1}{c + ab^t}$ называют кривой Пирла-Рида³. В ней верхняя асимптота составит величину $\frac{1}{c}$ (рис. 5). Точка перегиба у этой кривой равна $t_p = \frac{1}{\ln b} \ln \frac{c}{a}$. Значение τ в точке перегиба равно $\frac{1}{2c}$. При практических расчетах исследователь может не иметь в полном виде S – образную кривую. Тогда точка перегиба находится за пределами наблюдаемых величин уровней ряда. В этом случае верхняя асимптота является теоретическим максимумом и ориентироваться на него в дальнейшем прогнозе достаточно проблематично.

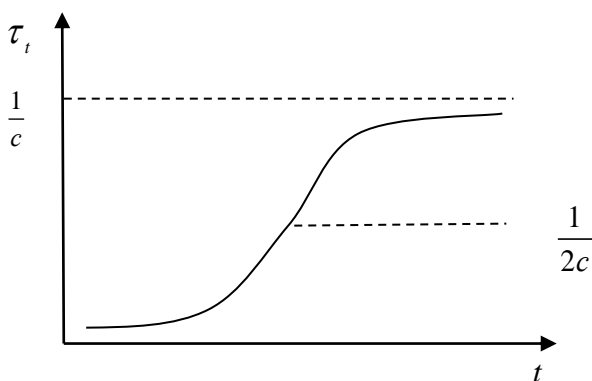


Рис. 5. Логистическая кривая Пирла – Рида

² Бенджамин Гомперц (Гомпертц) (англ. *Benjamin Gompertz*; 1779–1865) – английский математик, статистик и астроном; член Лондонского королевского общества.

³ Лоуренс Рид (англ. *Lawrence Reed*, 1953, Пенсильвания) – американский экономист. Раймонд Пирл (англ. *Raymond Pearl* – *Перл*; 1879–1940, США) – американский биолог, демограф и статистик.

К классу S – образных кривых относится также *кривая Гомперца*

$$\tau_t = c \cdot a^{b^t}. \quad (1.13)$$

Она нашла применение в страховых расчетах и экстраполяции численности населения. Верхняя асимптота соответствует значению параметра c , а нижняя равна нулю, если $\ln a < 0$. Если $\ln a > 0$, то кривая имеет нижнюю асимптоту, равную величине параметра c . Кривая Гомперца основана на модифицированной экспоненте.

1.3. Моделирование тенденций с помощью сглаживания

Цель сглаживания временного ряда заключается в получении ряда с меньшим разбросом уровней, что в ряде случаев позволяет на основе визуального анализа сделать вывод о наличии тенденции и ее характерных особенностях.

Сглаживание динамического ряда по методу скользящей средней заключается в замене исходных уровней ряда y_t сглаженными значениями y'_t , которые получаются как среднее значение определенного числа уровней исходного ряда, симметрично окружающего значение y_t . В результате получается временной ряд y'_t , меньше подверженный колебаниям. Если индивидуальный разброс значений временного ряда около своего среднего значения a характеризуется дисперсией σ^2 , то средняя из m членов ряда $(y_1 + y_2 + \dots + y_m)/m$ будет иметь в m раз меньшую дисперсию σ^2/m .

Для вычисления сглаженных значений y'_t по методу простой скользящей средней используются следующие формулы:

1. *Нечетный интервал сглаживания* $g = 2p + 1$ (интервал сглаживания – количество исходных уровней ряда y_t , используемых для сглаживания):

$$y'_t = \frac{\sum_{i=t-p}^{t+p} y_i}{2p+1} = \frac{y_{t-p} + y_{t-p+1} + \dots + y_{t+p-1} + y_{t+p}}{2p+1}, \quad (1.14)$$

где y_t – фактическое значение уровня исходного ряда в момент t ; y'_t – значение скользящей средней в момент t ; $2p + 1$ – длина интервала сглаживания.

При интервалах сглаживания $g = 3$ и $g = 5$ формула (1.14) принимает вид

$$y'_t = \frac{y_{t-1} + y_t + y_{t+1}}{3}; \quad y'_t = \frac{y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}}{5}.$$

2. *Четный интервал сглаживания* $g = 2p$:

$$y'_t = \frac{\frac{1}{2}y_{t-p} + \sum_{i=t-p+1}^{t+p-1} y_i + \frac{1}{2}y_{t+p}}{2p} = \frac{\frac{1}{2}y_{t-p} + y_{t-p+1} + \dots + y_{t+p-1} + \frac{1}{2}y_{t+p}}{2p}. \quad (1.15)$$

Формула (1.15) при интервалах сглаживания $g = 2$ и $g = 4$ принимает вид

$$y'_t = \frac{\frac{1}{2}y_{t-1} + y_t + \frac{1}{2}y_{t+1}}{2}; \quad y'_t = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4}.$$

При использовании скользящей средней с длиной активного участка $g = 2p + 1$ первые и последние p уровней ряда сгладить нельзя, их значения теряются. Очевидно, что потеря значений последних точек является существенным недостатком, так как для исследования последние «свежие» данные обладают наибольшей информационной ценностью.

Для восстановления потерянных значений временного ряда можно использовать следующий прием:

а) вычисляется средний прирост Δ_y на последнем активном участке (y_{n-g}, \dots, y_n)

$$\Delta_y = \frac{y_n - y_{n-g}}{g - 1},$$

где g — длина активного участка;

б) определяются значения последних $p = (g - 1)/2$ уровней сглаженного временного ряда с помощью последовательного прибавления среднего абсолютного прироста Δ_y к последнему сглаженному значению y'_{n-p}

$$y'_{n-p+1} = y'_{n-p} + \Delta_y, \quad y'_{n-p+2} = y'_{n-p+1} + \Delta_y, \dots, \quad y'_n = y'_{n-1} + \Delta_y.$$

Аналогичная процедура применяется для восстановления первых p уровней временного ряда.

Важным свойством процедуры сглаживания является полное устранение периодических колебаний из временного ряда, если длина интервала сглаживания берется равной или кратной периоду колебаний. Это обстоятельство используется при выделении периодической составляющей динамического ряда.

Пример 1.2. Пусть длина интервала сглаживания $g = 5$, а локальное поведение сглаженного ряда внутри активного участка описывается с помощью параболы $(y_t = b_0 + b_1 t + b_2 t^2)$. Определить весовые коэффициенты для расчета взвешенной скользящей средней.

Перенесем начало координат в середину активного участка, т.е. $t = -2, -1, 0, 1, 2$. По МНК найдем оценки неизвестных коэффициентов квадратного трехчлена, т.е. $R = \sum_{t=-2}^2 (y_t - b_0 - b_1 t - b_2 t^2)^2 \rightarrow \min$.

Сглаженное значение в центральной точке активного участка определится коэффициентом b_0 . После преобразований получим

$$b_0 = \frac{1}{35}(-3y_{-2} + 12y_{-1} + 17y_0 + 12y_1 + 3y_2) \text{ или } b_0 = \frac{1}{35}[-3; 12; 17].$$

Таким образом, оценка сглаженного значения в центральной точке активного участка определится как взвешенная средняя арифметическая из 5 уровней, образующих этот участок. Если в примере задать длины $g = 7, 9, 11$, то весовые коэффициенты получатся равными:

$$g = 7: \quad \frac{1}{21}[-2; 3; 6; 7];$$

$$g = 9: \quad \frac{1}{231}[-21; 14; 39; 54; 59];$$

$$g = 11: \quad \frac{1}{429}[-36; 9; 44; 69; 84; 89].$$

1.4. Оценка параметров уравнения тренда

При использовании полиномов разных степеней оценка параметров уравнения тренда производится методом наименьших квадратов (МНК) по аналогии с оценками параметров уравнений регрессий на основе пространственных данных. В качестве зависимой переменной рассматриваются уровни временного ряда y_t , а в качестве независимой переменной – фактор времени t , который обычно выражается рядом натуральных чисел $1, 2, \dots, n$.

Оценка параметров нелинейных функций проводится МНК после линеаризации, т.е. приведения их к линейному виду.

Например, для оценки параметров показательной кривой $y_t = a \cdot b^t$ линеаризация происходит с помощью логарифмирования функции: $\ln y_t = \ln a + t \cdot \ln b$. Далее строится система нормальных уравнений при решении задачи на минимум:

$$\begin{cases} \sum \ln y_t = n \cdot \ln a + \ln b \cdot \sum t, \\ \sum t \cdot \ln y_t = \ln a \cdot \sum t + \ln b \cdot \sum t^2. \end{cases} \quad (1.16)$$

В результате получается оцененное уравнение:

$$\hat{y}_t = \hat{a} \cdot \hat{b}^t.$$

Пример 1.3. Число зарегистрированных ДТП (на 100 тыс. чел. населения) по Иркутской области за последние 9 лет характеризуются данными (табл. 4).

Таблица 4

2009	2010	2011	2012	2013	2014	2015	2016	2017
105,7	105,3	156,0	158,1	160,1	178,0	191,5	274,6	287,3

Определить тенденцию данного показателя и определить соответствующие параметры данного распределения через два года.

Решение.

Построим диаграмму рассеивания (рис. 6).

Предположим, что точки располагаются вблизи показательной кривой $y_t = a \cdot b^t$. Для построения системы (1.14) рассчитаем вспомогательные величины:

$$\sum \ln y_t = 46,219; \quad \sum t = 45; \quad \sum (t \cdot \ln y_t) = 238,498; \quad \sum t^2 = 285.$$

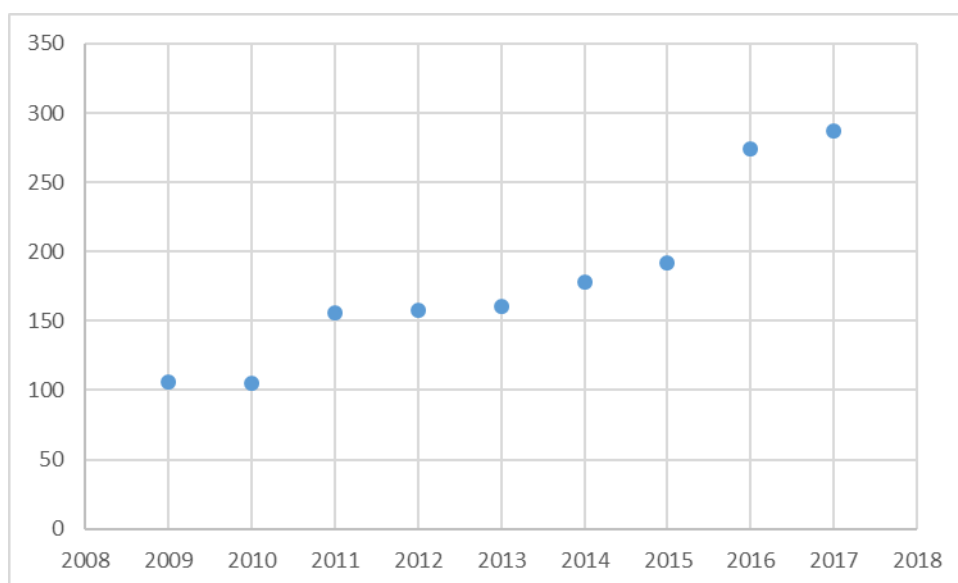


Рис. 6. Диаграмма рассеивания данных

Система нормальных уравнений тогда имеет вид:

$$46,219 = 9 \ln a + 45 \ln b;$$

$$238,498 = 45 \ln a + 285 \ln b.$$

Решая ее, получим значения:

$$\ln a = 4,519, \quad \hat{a} = e^{4,519} = 91,7;$$

$$\ln b = 0,123, \quad \hat{b} = e^{0,123} = 1,131.$$

Соответственно имеем тенденцию данного временного ряда:

$$\hat{y}_t = 91,7 \cdot 1,131^t.$$

За период с 2009 по 2017 г. число дорожно-транспортных происшествий возрастало в среднем ежегодно на 13,1 %. Показательная кривая достаточно хорошо описывает тенденцию исходного временного ряда, так как коэффициент детерминации составил 0,92. Следовательно, данный тренд объясняет 92 % колеблемости уровней ряда и лишь 8 % ее связаны со случайными факторами. Таким образом, через 2 года число ДТП по Иркутской области составит 356,3 единиц на 100 тыс. чел.

Некоторую специфику имеет оценка параметров кривых с насыщением: модифицированной экспоненты, логистической кривой, кривой Гомперца и др. По этим функциям должна быть сначала определена асимптота. Если она может быть задана исследователем на основе анализа временного ряда, то другие параметры могут быть оценены по МНК. В этих случаях данные функции приводятся к линейному виду.

Рассмотрим некоторые примеры.

Пример 1.4. Уровень механизации труда (%) характеризуется динамикой (табл. 5):

Таблица 5

2009	2010	2011	2012	2013	2014	2015	2016
82	85	89	91	92	93	94	96

Построить тенденцию данного ряда в виде модифицированной экспоненты и дать интерпретацию полученным показателям.

Решение.

Рассчитаем параметры модифицированной экспоненты (1.12) $y_t = c - ab^t$.

Так как уровень механизации труда не может превышать 100 %, то имеется объективно заданная верхняя асимптота $c = 100$ (табл. 6).

Таблица 6

Расчет параметров модифицированной экспоненты

Годы	y_t	$Y = c - y_t$	$\ln Y$	t	$t \cdot \ln Y$	t^2	\hat{y}_t
2009	82	18	2,890	1	2,890	1	82,459
2010	85	15	2,708	2	5,416	4	85,616
2011	89	11	2,398	3	7,194	9	88,205
2012	91	9	2,197	4	8,789	16	90,328
2013	92	8	2,079	5	10,397	25	92,069
2014	93	7	1,946	6	11,675	36	93,497
2015	94	6	1,792	7	12,542	49	94,667
2016	96	4	1,386	8	11,090	64	95,627
Σ	722		17,397	36	69,994	204	722,468

Для оценки параметров a и b приведем рассматриваемую функцию к линейному виду $c - y_t = ab^t$; обозначим $(c - y_t)$ через Y и прологарифмируем: $\ln Y = \ln a + t \cdot \ln b$.

Далее применим МНК и получим систему нормальных уравнений вида (1.16):

$$\begin{cases} \sum \ln Y = n \cdot \ln a + \ln b \cdot \sum t, \\ \sum t \cdot \ln Y = \ln a \cdot \sum t + \ln b \cdot \sum t^2. \end{cases}$$

Исходя из расчетов в табл. 6, имеем систему уравнений:

$$17,397 = 8 \ln a + 36 \ln b;$$

$$69,994 = 36 \ln a + 204 \ln b.$$

Решив ее, получим

$$\ln a = 3,063, \quad \hat{a} = e^{3,063} = 21,392;$$

$$\ln b = -0,197, \quad \hat{b} = e^{-0,197} = 0,820.$$

Соответственно оцененное уравнение примет вид:

$$\hat{Y} = 21,392 \cdot 0,82^t.$$

Если вернуться к исходным уровням временного ряда y_t , уравнение модифицированной экспоненты составит $\hat{y}_t = 100 - 21,392 \cdot 0,82^t$, где параметр $\hat{b} = 0,82$ показывает средний коэффициент снижения уровня использования ручного труда за рассматриваемые годы. Расчетные значения \hat{y}_t могут быть найдены путем подстановки в уравнение соответствующих значений времени t .

Для логистической кривой Пирла – Рида $y_t = \frac{1}{c + ab^t}$ (рис. 5) параметры a и b могут быть найдены с помощью МНК, при условии, что асимптота c задана. Тогда данная функция преобразовывается в линейную с помощью логарифмирования: $\frac{1}{y_t} - c = ab^t$; $Y = \frac{1}{y_t} - c$; $\ln Y = \ln a + t \cdot \ln b$. Далее параметры a и b определяются МНК.

При практических расчетах значение верхней асимптоты логистической кривой может быть определено исходя из существования развития явления, различного рода ограничений для его роста (нормативы потребления, законодательные акты), а также графически.

Если верхняя асимптота не задана, то для оценки параметров могут использоваться разные методы: Фишера⁴, Юла⁵ и др. Рассмотрим пример.

Пример 1.5. Производство продукции характеризуется данными (тыс. т) (табл. 7):

Таблица 7

1	2	3	4	5	6	7	8	9
12	28	58	105	182	260	320	360	380

Рассчитать параметры логистической кривой и дать интерпретацию полученным показателям.

Решение.

Рассчитаем параметры логистической прямой, используя метод Фишера. Метод Фишера основан на определении производной для логистической кривой $y_t = \frac{c}{1 + be^{-at}}$. Дифференцируя данную функцию по t , получим уравнение

$$\frac{dy_t}{dt} = a - \frac{a}{c} y_t.$$

Обозначим темп прироста логистической кривой через $z_t = \frac{dy_t}{dt} \cdot \frac{1}{y_t}$. Тогда $z_t = a - \frac{a}{c} y_t$, т.е. для z_t имеем линейную функцию с параметрами a и $\frac{a}{c}$. Чтобы найти решение, необходимо оценить z_t . Предполагая, что интервалы между уровнями в ряду динамики равны, Фишер предложил приближенно оценивать z_t в виде уравнения $z_t = \frac{1}{2} \ln \frac{y_{t+1}}{y_{t-1}}$, $t = 2, 3, \dots, n-1$ (табл. 8). Далее при-

⁴ Рональд Эйлер Фишер (или Рональд, англ. *Sir Ronald Aylmer Fisher*, 1890–1962) – английский статистик, биолог-эволюционист и генетик.

⁵ Юл, Джордж Удни (англ. *George Udny Yule*, 1871–1951) – шотландский статистик, один из основателей анализа временных рядов.

меняем метод наименьших квадратов к уравнению $z_t = a - \frac{a}{c} y_t$, т.е. строим регрессию z_t от y_t , беря данные от $t = 2$ до $t = 8$.

Таблица 8

Расчет параметров по методу Фишера

t	y_t	z_t	Y	$\ln Y$	\hat{y}_t
1	12	–	32,583	3,484	12,155
2	28	0,788	13,393	2,595	26,241
3	58	0,661	5,948	1,783	54,379
4	105	0,572	2,838	1,043	104,336
5	182	0,453	1,214	0,194	176,898
6	260	0,282	0,550	–0,598	256,573
7	320	0,163	0,259	–1,349	321,160
8	360	0,086	0,119	–2,125	361,830
9	380	–	0,061	–2,805	383,516
Σ	1 705			2,222	1 697

Уравнение регрессии запишется в виде $\hat{z}_t = 0,806 - 0,002 y_t$. Исходя из него, находим параметры a и c для логистической кривой. Параметр $a = 0,806$. Данное уравнение статистически значимо: F -критерий равен 687,5, $R^2 = 0,993$. Оценки параметров данного уравнения также значимы: t -критерий для параметра a равен 47,2, для параметра $-\frac{a}{c}$ равен –26,2. Таким образом, верхняя асимптота производства продукции составляет 403 ед.

Параметр b может быть найден через функцию $y_t = \frac{c}{1 + be^{-at}}$. Обозначим через Y выражение $Y = \frac{c}{y_t} - 1$. Тогда имеем уравнение $Y = be^{-at}$. Прологарифмируем его: $\ln Y = \ln b - a \cdot t$. Здесь $\ln b$ является свободным членом. Определим его из первого уравнения системы нормальных уравнений (1.16): $\ln b = \frac{1}{9} \cdot 2,222 + 0,806 \cdot 5 = 4,277$. Соответственно $b = e^{4,277} = 72,016$. Тогда оцененная логистическая кривая имеет вид: $y_t = \frac{403}{1 + 72,016e^{-0,806t}}$. Полученные данные достаточно точно аппроксимируют исходный ряд.

Пример 1.6. Рассматривается динамика товарных запасов (тыс. дол.) некоторого предприятия (табл. 9):

Таблица 9

1	2	3	4	5	6	7	8	9	10	11	12
50	80	114	144	170	190	207	219	228	234	242	244

Рассчитать параметры кривой Гомперца и дать интерпретацию полученным показателям.

Решение.

По экономическим соображениям чрезмерное увеличение запасов не выгодно предпринимателю, поэтому рассмотрим предположение, что верхняя асимптота кривой Гомперца не превысит 250 тыс. дол., т.е. $c = 250$.

Проведем расчеты параметров кривой Гомперца (табл. 10):

Таблица 10

Расчет параметров по кривой Гомперца

t	y	$y/250 = Y$	$\lg Y$	$Y' = \lg[(-1)\lg Y]$	\hat{y}_t
1	2	3	4	5	6
1	50	0,200	-0,6990	-0,1555	44,75145
2	80	0,320	-0,4949	-0,3055	76,87416
3	114	0,456	-0,3410	-0,4672	111,3921
4	144	0,576	-0,2396	-0,6206	143,6385
5	170	0,680	-0,1675	-0,7760	170,9863
6	190	0,760	-0,1192	-0,9238	192,6847
7	207	0,828	-0,0820	-1,0863	209,1293
8	219	0,876	-0,0575	-1,2404	221,2058
9	228	0,912	-0,0400	-1,3979	229,8847
10	234	0,936	-0,0287	-1,5418	236,0301
11	242	0,968	-0,0141	-1,8500	240,3373
12	244	0,976	-0,0106	-1,9767	243,3353

Используя данные пятой графы табл. 10 и применяя метод наименьших квадратов, получим предварительное уравнение

$$\hat{Y}' = 0,0374 - 0,164 \cdot t,$$

в котором параметры выражены через логарифмы. Потенцируя полученные параметры, кривая Гомперца примет следующий вид

$$\hat{y}_t = 250 \cdot 0,0813^{0,6855t}.$$

Расчетные значения для этой кривой приведены в последней графе табл. 10.

1.5. Верификация моделей тенденции

Модель тенденции считается верифицированной, т.е. адекватной реальному процессу, если теоретические уровни рассматриваемого временного ряда достаточно близко подходят к фактическим их значениям: y_t и \hat{y}_t мало отличаются друг от друга. Для оценки адекватности такой модели проводится анализ остатков: $e_t = y_t - \hat{y}_t$.

Модели тенденции также можно сравнить по величине остаточной суммы квадратов рассматриваемых остатков: $S^2 = \sum e_t^2$. Чем меньше эта величина, тем в большей мере уравнение тренда подходит для описания тенденции временного ряда.

Другим показателем при выборе функции тренда является коэффициент детерминации R^2 . Чем выше R^2 , тем выше вероятность того, что данная модель тенденции описывает исходные данные. Величина $1 - R^2$ отражает влия-

ние случайной составляющей, т.е. показывает, какая доля вариации уровней динамического ряда не связана с тенденцией.

Рассмотрим пример 1.3, в котором изучалась динамика числа зарегистрированных ДТП. Тенденцию временного ряда определяло уравнение экспоненты $\hat{y}_t = 91,7 \cdot 1,131^t$, с коэффициентом детерминации $R^2 = 0,92$. Если рассчитать, например, линейный тренд данной динамики: $\hat{y}_t = 69,1 + 22,1 \cdot t$, то коэффициент $R^2 = 0,88$ будет явно ниже экспоненциального. Для линейного тренда остаточная сумма квадратов составила $S^2 = 3874,62$, а для экспоненциального $S^2 = 2617,7$. Следовательно, экспонента лучше описывает тенденцию ряда.

Рассмотренные критерии адекватности модели тенденции не всегда могут приводить к правильным выводам. Необходимо учитывать статистическую значимость параметров уравнения тренда. Если для рассмотренного примера 1.3 построить параболу пятого порядка, то получим уравнение $\hat{y}_t = 288,6 - 356,11 \cdot t + \dots - 0,2609 \cdot t^5$. Если к параметрам данной модели применить критерии Стьюдента, то на 5 % уровне значимости данные параметры будут статистически не значимы и не устойчивы. В результате, имея коэффициент детерминации $R^2 = 0,986$, данную модель нельзя считать верифицированной, адекватной для дальнейших прогнозов.

Уравнение тренда хорошо описывает тенденцию, если отсутствует автокорреляция в остатках $e_t = y_t - \hat{y}_t$, т.е. остатки текущего периода не коррелируют с остатками предыдущего периода.

Измерить автокорреляцию в остатках можно с помощью *коэффициента автокорреляции остатков*:

$$r_{e_t} = \frac{\overline{e_t e_{t-1}} - \bar{e}_t \bar{e}_{t-1}}{\sigma_{e_t} \cdot \sigma_{e_{t-1}}}, \quad (1.17)$$

где $e_t = y_t - \hat{y}_t$, т.е. остатки текущего периода, $e_{t-1} = y_{t-1} - \hat{y}_{t-1}$ остатки предыдущего периода.

Пример 1.7. Изучается динамика численности детей в возрасте от 8 до 13 лет за последние 15 лет (тыс. чел.). В табл. 11 рассмотрены остаточные величины такой динамики.

Таблица 11

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

Рассчитать коэффициент автокорреляции остатков и сделать вывод.

Решение.

Верифицированная модель динамики численности детей характеризуется параболой:

$$\hat{y}_t = 110,5 + 10,8 \cdot t - 0,8 \cdot t^2,$$

где \hat{y}_t численность детей, $R^2 = 0,967$ коэффициент детерминации; все параметры уравнения тренда статистически значимы.

Исходя из формулы (1.17) коэффициент автокорреляции остатков в данном примере окажется равным 0,6278. Величина достаточно немаленькая, чтобы утверждать об отсутствии автокорреляции остатков. Очевидно, данное уравнение тренда не является наилучшим, так как нарушена предпосылка МНК об отсутствии автокорреляции остатков.

Для комплексной оценки адекватности модели тенденции можно пользоваться и другими характеристиками, которые обычно рассматриваются при оценке качества регрессионных моделей, например, средней ошибкой аппроксимации, показателями асимметрии и эксцесса для остаточных величин.

1.6. Тесты на наличие автокорреляции остатков

Рассмотрим различные методы проверки модели на автокоррелированность остатков. Самым распространенным является *тест по критерию Дарбина – Уотсона*⁶.

Предположим, что автокорреляция остатков существует, тогда верно следующее соотношение:

$$e_t = \rho \cdot e_{t-1} + v_t, \quad (1.18)$$

где ρ – коэффициент автокорреляции остатков; v_t – независимые случайные величины.

Если $\rho = 0$, то автокорреляция остатков отсутствует и оценка параметров МНК является наилучшей линейной несмещенной оценкой.

Если $\rho \neq 0$, то оценка параметров МНК является смещенной и выводы по модели не корректны. Таким образом, при тестировании на автокорреляцию остатков выдвигается основная нулевая гипотеза об отсутствии автокорреляции остатков $H_0 : \rho = 0$, против альтернативной $H_1 : \rho \neq 0$.

Вводится статистика Дарбина – Уотсона, которая определяется соотношением

$$D - W = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}. \quad (1.19)$$

⁶ Критерий назван в честь Джеймса Дарбина и Джеффри Уотсона, Джеймс Дарбин и Джеффри Уотсон – британские статистики и эконометристы, члены Королевского статистического общества.

Критерий Дарбина – Уотсона и коэффициент автокорреляции остатков связаны между собой соотношением

$$D - W \approx 2(1 - \rho). \quad (1.20)$$

Параметры уравнения тренда обычно оцениваются методом наименьших квадратов, у которого в качестве предпосылок заложены условия, что $\sum_{t=1}^n e_t = 0$ и

$$\bar{e}_t = \frac{1}{n} \sum_{t=1}^n e_t = 0.$$

Поскольку ряд e_{t-1} сдвинут по отношению к ряду e_t на один временной интервал, то вариации в этих рядах мало отличаются друг от друга и можно предположить, что $\sigma_{e_t} \approx \sigma_{e_{t-1}}$.

Тогда коэффициент автокорреляции остатков можно представить в следующем виде

$$\rho = \frac{\sum_{t=2}^n e_t \cdot e_{t-1}}{\sum_{t=1}^n e_t^2}. \quad (1.21)$$

Из соотношения (1.20) очевидно, что при полной положительной автокорреляции остатков ($\rho = 1$) критерий $D - W$ имеет нулевое значение, $D - W = 0$. При полной отрицательной автокорреляции ($\rho = -1$) критерий $D - W$ равен $D - W = 4$. Если же автокорреляция в остатках отсутствует, т.е. $\rho = 0$, то $D - W = 2$. Очевидно, что критерий Дарбина – Уотсона изменяется в пределах $0 \leq D - W \leq 4$.

Дарбин и Уотсон разработали пороговые значения показателя $D - W$, позволяющие принять или отвергнуть гипотезу об отсутствии автокорреляции в остатках $H_0 : \rho = 0$.

При заданном числе наблюдений n (длина динамического ряда) и m параметров при t в уравнении тренда установлены верхняя $D - W_u$ ⁷ и нижняя $D - W_l$ границы критерия, учитывая уровень значимости $\alpha = 0,05$.

Фактическое значение критерия $D - W$ сравнивается с табличными значениями.

Если $D - W < 2$, то возможно следующие варианты:

1) при $D - W < D - W_l$ (нижней границы) нулевая гипотеза об отсутствии автокорреляции отвергается и делается вывод о наличии положительной автокорреляции в остатках;

2) при $D - W > D - W_u$ (верхней границы) нулевая гипотеза об отсутствии автокорреляции не отвергается, т.е. делается вывод об отсутствии корреляционной связи последующих остатков с предыдущими;

3) при $D - W_l \leq D - W \leq D - W_u$ нельзя не отвергнуть, ни принять нулевую гипотезу об отсутствии автокорреляции в остатках, т.е. значение $D - W$ попало

⁷ u – upper, l – lower.

в область неопределённости и необходимы дальнейшие исследования, например, по большому числу наблюдений.

Если фактическое значение $D - W > 2$, что означает отрицательную автокорреляцию, то с пороговыми табличными значениями сравнивается величина $4 - DW$.

При этом возможны следующие варианты:

1) при $4 - DW < D - W_l$ отвергается нулевая гипотеза об отсутствии автокорреляции и делается вывод о наличии отрицательной автокорреляции в остатках;

2) при $4 - DW > D - W_u$ нулевая гипотеза об отсутствии автокорреляции в остатках принимается;

3) при $D - W_l \leq 4 - DW \leq D - W_u$ нельзя сделать определенного вывода о наличии или отсутствии автокорреляции в остатках имеющимся данным.

По величине критерия Дарбина – Уотсона можно определить величину коэффициента автокорреляции остатков исходя из соотношения:

$D - W \approx 2(1 - \rho)$. Отсюда $\frac{D - W}{2} \approx (1 - \rho)$ и соответственно $\rho \approx 1 - \frac{D - W}{2}$. Поэтому, если $D - W > 2$, то $\rho < 0$, а при $D - W < 2$ имеем $\rho > 0$.

Таким образом, если фактическое значение критерия Дарбина – Уотсона не слишком отличается от 2, то можно сделать вывод об отсутствии автокорреляции в остатках.

Графическое изображение критерия Дарбина – Уотсона позволяет наглядно делать вывод о наличии или отсутствии автокорреляции в остатках (рис. 7).

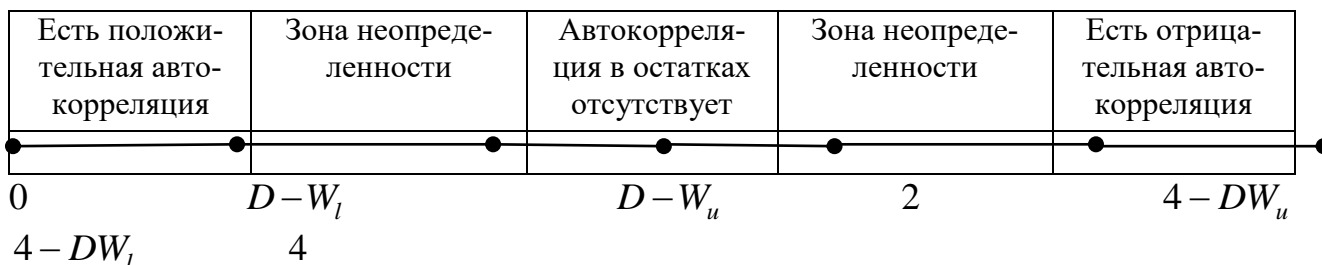


Рис. 7. Схема применения критерия Дарбина – Уотсона

Рассмотрим **пример 1.7**, где изучалась динамика численности детей в возрасте от 8 до 13 лет за последние 15 лет (тыс. чел.). Воспользуемся формулой (1.19) и рассчитаем значение критерия Дарбина – Уотсона $D - W$ для определения наличия автокорреляции в остатках.

В примере $\sum_{t=2}^n (e_t - e_{t-1})^2 = 99,65$; $\sum_{t=1}^n e_t^2 = 146,64$. Следовательно,

$$D - W = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} = \frac{99,65}{146,64} = 0,678.$$

При $n = 15$ и $m = 2$ $D - W_l = 0,95$ (табличные значения критерия⁸). Фактическое значение критерия $0,678$ меньше его нижней границы $0,95$, что подтверждает вывод о наличии в остатках автокорреляции и означает плохой выбор модели тенденции.

Пример 1.8. Изучается динамика просроченной задолженности некоторого товара за 12 месяцев. Предположим, что данная динамика описывается построенным уравнением тренда $\hat{y}_t = 410,1 - 6,5t$ при $t \in [1, \dots, 12]$; $R^2 = 0,921$. Критерий Дарбина – Уотсона $D - W$ составил $2,226$. Требуется оценить автокорреляцию в остатках.

Решение. Фактическое значение критерия $D - W > 2$, то речь идет об оценке отрицательной автокорреляции в остатках. Приблизненно коэффициент автокорреляции остатков $\rho = 1 - \frac{2,226}{2} = -0,113$, что означает слабую автокорреляцию в остатках.

Убедимся в этом через сравнение фактического значения $4 - DW$ с табличными значениями при 5%-ном уровне значимости.

$$4 - DW = 4 - 2,226 = 1,774.$$

При $n = 12$ и $m = 1$ $D - W_l = 0,97$ (нижняя граница) и $D - W_u = 1,33$ (верхняя граница). Так как $4 - DW > D - W_u$, то автокорреляция в остатках отсутствует, и линейный тренд достаточно хорошо описывает исходный динамический ряд.

Графически данная задача также легко решается. Для этого на числовую ось наносят необходимые значения критерия $D - W$ в интервале $[0, \dots, 4]$, а именно:

$$D - W_l = 0,97; D - W_u = 1,33; 4 - DW_u = 2,67; 4 - DW_l = 3,03,$$

и посмотреть, в какой интервал попадает фактическое значение критерия $D - W = 2,226$ (рис. 8).

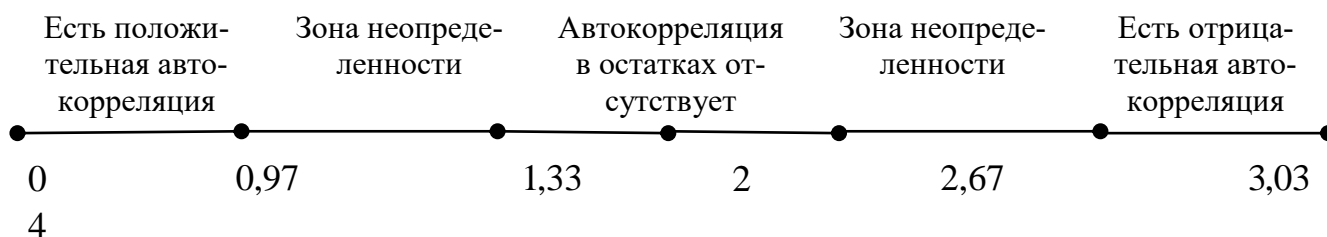


Рис. 8. Интервалы применения критерия Дарбина – Уотсона

Фактическое значение $D - W = 2,226$ попадает в промежуток от $D - W_u$ до $4 - DW_u$. Следовательно, на 5%-ном уровне значимости можно принять нулевую гипотезу об отсутствии автокорреляции в остатках.

⁸ Табличные значения критерия Дарбина – Уотсона см. в прил. 1.

Критерий Дарбина – Уотсона используется для оценки автокорреляции по регрессионным моделям, построенный по временным рядам. Однако его нельзя использовать, если в модели среди объясняющих переменных содержатся лаговые значения результативного признака. Например, инвестиции в основной капитал в текущем периоде рассматриваются не только как функция разных экономических переменных (прибыли, процентной ставки), но и достигнутого в предыдущий период размера инвестиций. Критерий Дарбина – Уотсона является самым распространенным в современной эконометрике.

Другой активно используемый критерий – это тест Бреуша – Годфри⁹. В этом тесте выдвигается гипотеза о независимости случайных величин E_t , разделённых k периодами времени и очищенных от влияния промежуточных значений остатков. Иными словами, модель зависимости остатков имеет вид

$$E_t = a_1 E_{t-1} + a_2 E_{t-2} + \dots + a_k E_{t-k} + v_t,$$

где $E_t, E_{t-1}, E_{t-2}, \dots, E_{t-k}$ – остатки по линейной регрессионной модели, построенной по временным рядам: $y_t = b_0 + b_1 x_{1t} + \dots + b_p x_{pt} + E_t$, v_t – независимые случайные величины.

В данном критерии рассматривается автокорреляция остатков k -го порядка в отличие от автокорреляции первого порядка, оцениваемой критерием Дарбина – Уотсона. В тесте Бреуша – Годфри проверяется нулевая гипотеза:

$$H_0 : a_1 = a_2 = \dots a_k = 0.$$

Если это гипотеза верна, то при большом количестве наблюдений n статистика критерия nR^2 , где R^2 – коэффициент детерминации по модели регрессии для остатков, имеет распределение, близкое к распределению χ^2 ¹⁰ (хи квадрат) с k степенями свободы. Нулевая гипотеза отвергается, если вычислительное значение nR^2 превышает критическое (табличное) значение χ^2 при заданном уровне значимости α , т.е. если имеем соотношение $nR^2 > \chi^2_{1-\alpha}$.

При $k=1$ тест Бреуша – Годфри оценивает автокорреляцию остатков первого порядка и поэтому сопоставим с критерием Дарбина – Уотсона. Выводы по этим критериям почти не расходятся. Несмотря на то, что критерий Бреуша – Годфри является асимптотическим, а критерий Дарбина – Уотсона точным, в ряде случаев целесообразно применения критерия Бреуша – Годфри. Так, его можно использовать и в моделях регрессии с лаговыми значениями результативного признака среди объясняющих переменных. Этот критерий учитывает наличие в остатках автокорреляции не только первого порядка. Поэтому его называют также *тестом серий*.

⁹ Особенность теста Бреуша – Годфри заключается в том, что его можно использовать практически всегда, в отличие от, например, критерия Дарбина – Уотсона. Кроме того, тест позволяет проверить автокорреляцию любого порядка.

¹⁰ Распределение Пирсона – хи квадрат (χ^2). Карл Пирсон – английский математик, статистик, биолог и философ; основатель математической статистики, один из основоположников биометрики. Автор свыше 650 опубликованных научных работ.

Тестовые задания

Задание 1. Структуру временного ряда можно выявить с помощью коэффициента ...

- 1) детерминации уровней ряда;
- 2) регрессии уровней ряда;
- 3) автокорреляции уровней ряда;
- 4) авторегрессии уровней ряда.

Задание 2. Временной ряд – это совокупность значений экономических показателей...

- 1) за несколько последовательных моментов (периодов) времени;
- 2) не зависящих от времени;
- 3) по однотипным объектам;
- 4) за несколько непоследовательных моментов (периодов) времени.

Задание 3. Параметры уравнения тренда определяются _____ методом наименьших квадратов.

- 1) двухшаговым;
- 2) обычным;
- 3) косвенным;
- 4) обобщенным.

Задание 4. Уровнем временного ряда является...

- 1) совокупность значений временного ряда;
- 2) значение временного ряда в конкретный момент (период) времени;
- 3) среднее значение временного ряда;
- 4) значение конкретного момента (периода) времени.

Задание 5. Если наиболее высоким оказался коэффициент автокорреляции первого порядка, то исследуемый ряд содержит...

- 1) тенденцию;
- 2) случайную компоненту;
- 3) циклические колебания;
- 4) сильную нелинейную тенденцию.

Задание 6. Моделирование тенденции осуществляется на основе построения уравнения регрессии зависимости...

- 1) случайной компоненты от времени;
- 2) уровня ряда от времени;
- 3) сезонной компоненты от времени;
- 4) трендовой компоненты от времени.

Задание 7. Под стационарным процессом можно понимать...

- 1) функциональный процесс;

- 2) стохастический процесс, для которого среднее значение и дисперсия независимо от рассматриваемого периода имеют постоянное значение;
- 3) процесс с постоянной тенденцией;
- 4) процесс с возрастающей тенденцией.

Задание 8. Значение коэффициента автокорреляции второго порядка характеризует связь между...

- 1) исходными уровнями и уровнями другого временного ряда;
- 2) двумя временными рядами;
- 3) исходными уровнями и уровнями другого ряда, сдвинутыми на 2 момента времени;
- 4) исходными уровнями и уровнями этого же ряда, сдвинутыми на 2 момента времени.

Задание 9. Стохастическим процессом называется...

- 1) набор случайных переменных $X(t)$, где t – вещественные числа;
- 2) набор случайных переменных $X(t)$, где t – иррациональные числа;
- 3) функциональная связь $X(t)$, где t – вещественные числа;
- 4) набор неслучайных переменных $X(t)$, где t – вещественные числа.

Задание 10. Значение коэффициента автокорреляции рассчитывается по аналогии с...

- 1) линейным коэффициентом корреляции;
- 2) нелинейным коэффициентом корреляции;
- 3) линейным коэффициентом регрессии;
- 4) линейным коэффициентом детерминации.

Задание 11. Проверка, является ли временной ряд «белым шумом» осуществляется с помощью...

- 1) критерия Дарбина – Уотсона;
- 2) величины лага;
- 3) коэффициента автокорреляции;
- 4) Q-статистики Бокса-Пирса.

Задание 12. Стационарность характерна для временного ряда...

- 1) с отрицательной динамикой роста;
- 2) с положительной динамикой роста;
- 3) содержащего сезонные колебаний;
- 4) типа «белый шум».

Задание 13. Автокорреляционной функцией временного ряда называется...

- 1) зависимость приращений значений коэффициентов автокорреляции различных порядков от уровней временного ряда;
- 2) зависимость коэффициентов автокорреляции первого порядка от уровней временного ряда;

3) последовательность отношений коэффициентов автокорреляции к величинам соответствующих лагов;

4) последовательность значений коэффициентов автокорреляции, рассчитанных для разных порядков.

Задание 14. Коррелограммой является...

1) процесс экспериментального нахождения значений автокорреляционной функции;

2) аналитическое выражение для автокорреляционной функции;

3) графическое отображение регрессионной функции;

4) графическое отображение автокорреляционной функции.

Задание 15. Если наиболее высоким оказался коэффициент автокорреляции третьего порядка, то исследуемый ряд содержит...

1) случайную величину, влияющую на каждый третий уровень ряда;

2) нелинейную тенденцию полинома третьего порядка;

3) линейный тренд, появляющийся в каждом третьем уровне ряда;

4) сезонные колебания с периодичностью в три момента времени.

Задание 16. Экономические временные ряды, представляющие собой данные наблюдений за ряд лет, как правило, являются...

1) функционально зависящими от времени временными рядами;

2) нестационарными временными рядами;

3) стационарными временными рядами;

4) строго возрастающими временными рядами.

Задание 17. Значение коэффициента автокорреляции первого порядка равно 0,9, следовательно...

1) линейная связь между последующим и предыдущим уровнями не тесная;

2) линейная связь между последующим и предыдущим уровнями тесная;

3) линейная связь между временными рядами двух экономических показателей тесная;

4) нелинейная связь между последующим и предыдущим уровнями тесная.

Задание 18. Модель временного ряда не предполагает ...

1) независимость значений экономического показателя от времени;

2) учет временных характеристик;

3) зависимость значений экономического показателя от времени;

4) последовательность периодов времени, в течение которых рассматривается поведение экономического показателя.

Задание 19. При моделировании временных рядов экономических показателей необходимо учитывать...

1) независимый от времени уровень исследуемых показателей;

2) конструктивный характер уровней исследуемых показателей;

- 3) функциональный характер уровней исследуемых показателей;
- 4) стохастический характер уровней исследуемых показателей.

Задание 20. Тенденция временного ряда характеризует совокупность факторов...

- 1) оказывающих долговременное влияние и формирующих общую динамику изучаемого показателя;
- 2) оказывающих сезонное воздействие;
- 3) оказывающих единовременное влияние;
- 4) не оказывающих влияния на уровень ряда.

Контрольные вопросы

1. Как используются показатели динамики для обоснования типа модели тренда?
2. Дайте определение автокорреляции уровней и поясните, как она используется при моделировании динамического ряда.
3. Что такое автокорреляционная функция и как рассчитывается выборочная оценка коэффициента автокорреляции?
4. Как интерпретируются параметры модели в виде показательной кривой?
5. Как оцениваются параметры S-образных кривых?
6. Как оценить соответствующие модели характеру тренда?
7. Что такое автокорреляция в остатках и как она измеряется?
8. В чем причины появления автокорреляции в остатках?
9. С какой целью используется критерий Дарбина – Уотсона?
10. В каких пределах принимает значение критерий Дарбина – Уотсона?

2. Моделирование периодических колебаний

При использовании длинных экономических временных рядов может возникать потребность выделять жизненные циклы в развитии явления. Это могут быть циклы в сфере производства, рынка товаров и услуг с длительностью от нескольких месяцев до нескольких лет. Иными словами, в ряду динамики может иметь место циклическая или сезонная составляющая, которая должна быть учтена при построении модели временного ряда. Природа периодических колебаний может быть разной. Простейшим подходом расчета значений сезонной компоненты является *метод скользящей средней*. В экономике также применяются методы, основанные на *рядах Фурье*, так называемый *гармонический анализ* сезонной компоненты. Рассмотрим эти методы.

2.1. Метод скользящей средней для моделирования сезонной компоненты

Простейшим приемом выделения периодической компоненты является использование сглаженного временного ряда по методу простой скользящей средней. Предварительно следует определиться с видом модели временного ряда – аддитивной или мультипликативной. Это можно сделать на основе анализа график временного ряда.

Если амплитуда периодических колебаний примерно постоянна и не зависит от времени, то следует выбирать аддитивную модель:

$$y_t = \tau_t + \zeta_t + \varepsilon_t, \quad (2.1)$$

где y_t – предполагаемое, прогнозное значение уровней исследуемого ряда; τ_t – долговременный тренд; ζ_t – сезонная компонента; ε_t – случайная компонента.

Если амплитуда периодических колебаний возрастает или убывает с ростом уровней ряда, то следует выбирать мультипликативную модель:

$$y_t = \tau_t \cdot \zeta_t \cdot \varepsilon_t. \quad (2.2)$$

Выделение периодической компоненты основывается на том, что если исходный временной ряд содержит периодическую компоненту с периодом g , то сглаженный ряд по методу простой скользящей средней такую компоненту уже не содержит. Таким образом, в случае аддитивной модели периодическая компонента выделяется путем нахождения разности между соответствующими уровнями исходного и сглаженного ряда.

В случае мультипликативной модели периодическая компонента выделяется путем нахождения отношения между соответствующими уровнями исходного и сглаженного ряда. Затем вычисляются средние значения, соответствующие наблюдениям внутри одного периода колебаний.

Алгоритм построения аддитивной модели, с выделением сезонной компоненты:

1. Выравнивание исходного ряда y_t методом скользящей средней.
2. Расчет значений сезонной компоненты ζ_t .

3. Устранение сезонной компоненты из исходных уравнений ряда и получение выравненных данных $\tau_t + \zeta_t$.

4. Аналитическое выравнивание уровней $\tau_t + \varepsilon_t$ и расчет значений τ_t с использованием полученного уравнения тренда.

5. Расчет полученных по модели значений $\tau_t + \zeta_t$.

6. Расчет абсолютной и/или относительной ошибок.

Пример 2.1. Предположим, что имеются условные данные об объемах потребления электроэнергии за 16 кварталов (табл. 12).

Таблица 12

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

Требуется построить аддитивную модель имеющихся данных и провести прогноз на ближайшие три квартала.

Решение. Построим график данного ряда и определим автокорреляционную функцию с целью определения наличия и поведения сезонной составляющей.

Автокорреляционная функция имеет вид (рис. 9): $\rho(k) = \{0,512 (k=1); 0,402 (k=2); 0,450 (k=3); 0,538 (k=4); 0,484 (k=5); 0,280 (k=6); 0,343 (k=7); 0,525 (k=8)\}$.

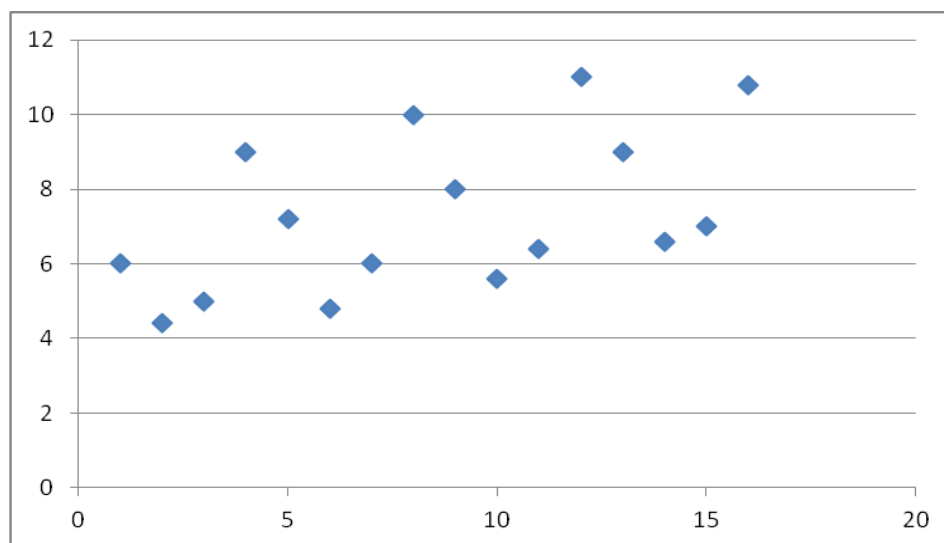


Рис. 9. График исходного ряда

По построенному графику (рис. 9) и по автокорреляционной функции можно сделать вывод, что исследуемый временной ряд содержит сезонные колебания с периодичностью 4. Это объяснимо и с экономической точки зрения: объемы потребления электроэнергии в осенне-зимний период (1 и 4 кварталы) выше, чем в весенне-летний период (2 и 3 кварталы). По графику видно наличие приблизительно равной амплитуды колебаний, что позволяет построить аддитивную модель временного ряда.

1. Проведем выравнивание исходных уровней ряда методом скользящей средней:

1) просуммируем уровни ряда последовательно за каждые 4 квартала со сдвигом на один момент времени и определим условные годовые объемы потребления электроэнергии Σ_i (табл. 13, гр. 3);

2) разделим полученные суммы на 4, тем самым найдем скользящие средние (табл. 13, гр. 4);

3) приведем эти значения в соответствие с фактическими моментами времени, для чего найдем средние значения из двух последовательных скользящих средних – центрированные скользящие средние μ_i (табл. 13, гр. 5).

Таблица 13

Номер квартала, t	Потребление электроэнергии, y_t	Суммы за 4 квартала, Σ_i	Скользящая средняя, $\Sigma_i/4$	Центрированная скользящая средняя, μ_i	Оценка сезонной компоненты
1	2	3	4	5	6
1	6	–	–	–	–
2	4,4	24,4	6,1	–	–
3	5	25,6	6,4	6,25	–1,25
4	9	26	6,5	6,45	2,55
5	7,2	27	6,75	6,625	0,575
6	4,8	28	7	6,875	–2,075
7	6	28,8	7,2	7,1	–1,1
8	10	29,6	7,4	7,3	2,7
9	8	30	7,5	7,45	0,55
10	5,6	31	7,75	7,625	–2,025
11	6,4	32	8	7,875	–1,475
12	11	33	8,25	8,125	2,875
13	9	33,6	8,4	8,325	0,675
14	6,6	33,4	8,35	8,375	–1,775
15	7	–	–	–	–
16	10,8	–	–	–	–

2. Найдем оценки сезонной компоненты как разность между фактическими уровнями ряда и центрированными скользящими средними $y_t - \mu_i$ (табл. 13, гр. 6). Используем эти оценки для расчета значений сезонной компоненты ζ_t : для этого найдем средние значения за каждый квартал оценки сезонной компоненты ζ_t (табл. 14).

В моделях с сезонной компонентой обычно предполагается, что сезонные воздействия за период взаимопогашаются. В аддитивной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равной нулю.

Таблица 14

Показатель	Год	Номер квартала			
		I	II	III	IV
	1	–	–	–1,25	2,55
	2	0,575	–2,075	–1,1	2,7
	3	0,55	–2,025	–1,475	2,875
	4	0,675	–1,775	–	–
Средняя оценка сезонной компоненты для i – квартала, ζ_t		0,6	–1,958	–1,275	2,708
Скорректированная сезонная компонента, $\bar{\zeta}_t$		0,581	–1,977	–1,294	2,690

Для данной модели проверим условие взаимопогашаемости:

$$0,6 - 1,958 - 1,275 + 2,708 = 0,075.$$

Определим корректирующий коэффициент $k = \frac{0,075}{4} = 0,01875$.

Рассчитаем скорректированные значения сезонной компоненты как разность между ее средней оценкой и корректирующим коэффициентом: $\bar{\zeta}_t = \zeta_t - k$. В результате этого, условие взаимопогашаемости выполняется точно: $0,581 - 1,977 - 1,294 + 2,690 = 0$. Таким образом, получены следующие значения сезонной компоненты: для I квартала $\zeta_1 = 0,581$, для II квартала $\zeta_2 = -1,979$, для III $\zeta_3 = -1,294$, для IV $\zeta_4 = 2,690$.

3. Следующим этапом в исследовании – исключим влияние сезонной компоненты ζ_t в рассматриваемом ряду, в результате получим значения, которые будут содержать только тенденцию τ_t и случайную компоненту ε_t : $\tau_t + \varepsilon_t = y_t - \zeta_t$ (табл. 15, гр. 4).

Таблица 15

t	y_t	ζ_t	$\tau_t + \varepsilon_t$	τ_t	$\tau_t + \zeta_t$	ε_t	ε_t^2
1	2	3	4	5	6	7	8
1	6	0,581	5,42	5,90	6,48	–0,48	0,23
2	4,4	–1,977	6,38	6,09	4,11	0,29	0,08
3	5	–1,294	6,29	6,27	4,98	0,02	0,00
4	9	2,69	6,31	6,46	9,15	–0,15	0,02
5	7,2	0,581	6,62	6,65	7,23	–0,03	0,00
6	4,8	–1,977	6,78	6,83	4,86	–0,06	0,00
7	6	–1,294	7,29	7,02	5,73	0,27	0,07
8	10	2,69	7,31	7,21	9,90	0,10	0,01
9	8	0,581	7,42	7,39	7,97	0,03	0,00
10	5,6	–1,977	7,58	7,58	5,60	0,00	0,00
11	6,4	–1,294	7,69	7,77	6,47	–0,07	0,01
12	11	2,69	8,31	7,95	10,64	0,36	0,13

t	y_t	ς_t	$\tau_t + \varepsilon_t$	τ_t	$\tau_t + \varsigma_t$	ε_t	ε_t^2
13	9	0,581	8,42	8,14	8,72	0,28	0,08
14	6,6	-1,977	8,58	8,33	6,35	0,25	0,06
15	7	-1,294	8,29	8,51	7,22	-0,22	0,05
16	10,8	2,69	8,11	8,70	11,39	-0,59	0,35

4. Определим компоненту τ_t данной модели. Для этого проведем аналитическое выравнивание ряда $\tau_t + \varepsilon_t$ с помощью линейного тренда. Используя компьютерную реализацию, получим МНК-оценки линейной регрессии. В результате имеем линейный тренд:

$$\tau_t = 5,716 + 0,186t, \text{ где } t = 1, \dots, 16.$$

(0,015)

Благодаря этому тренду, получаем значения компоненты τ_t (табл. 15, гр. 5). На рис. 10 приведена аппроксимация исходного ряда линейным трендом.

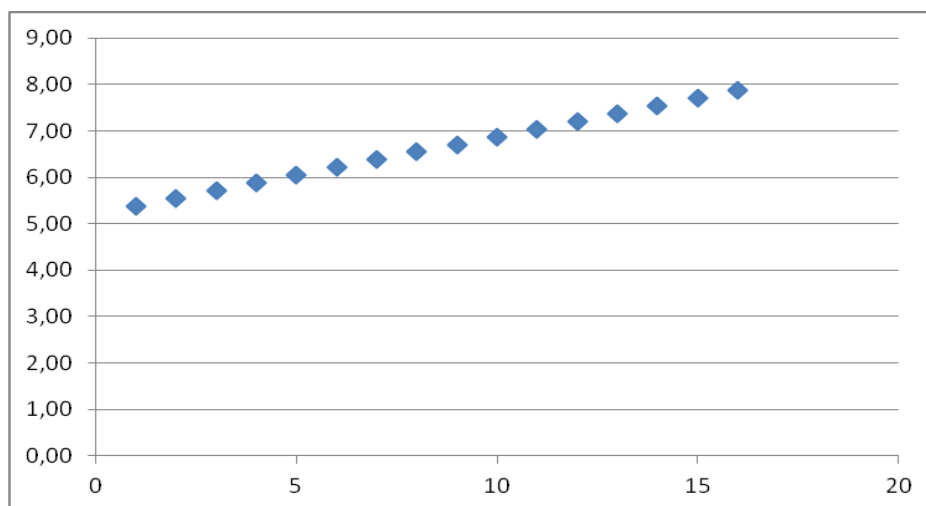


Рис. 10. Аппроксимация линейным трендом

5. Найдем значения уровней ряда, полученные по аддитивной модели. Для этого вычислим значения $\tau_t + \varsigma_t$ (табл. 15, гр. 6).

6. В соответствии с методикой построения аддитивной модели расчет ошибки проводится по формуле: $\varepsilon_t = y_t - (\tau_t + \varsigma_t)$. Это абсолютная ошибка (табл. 15, гр. 7).

По аналогии с регрессионным анализом для оценки качества построенной модели, а также для выбора наилучшей модели, найдем сумму квадратов остатков ε_t^2 , которая используется для подсчета коэффициента детерминации.

$$R^2 = \left(1 - \frac{RSS}{TSS} \right) \cdot 100\%, \text{ где } RSS = \varepsilon_t^2 - \text{сумма квадратов остатков}$$

данного ряда, TSS – общая сумма квадратов.

Для данной модели: $RSS = 1,1$, $TSS = 67,12$, в результате коэффициент детерминации равен:

$$R^2 = \left(1 - \frac{1,1}{67,12}\right) \cdot 100\% = 0,984 \cdot 100\% = 98,4\% .$$

Таким образом, можно сделать вывод, что аддитивная модель объясняет 98,4 % общей вариации уровней временного ряда потребления электроэнергии за последние 16 кварталов.

Проведем прогноз на ближайшие 3 квартала ($t = 17, 18, 19$).

$$\tau_{17} = 5,716 + 0,186 \cdot 17 = 8,878; \tau_{18} = 9,064; \tau_{19} = 9,25.$$

$$\varsigma_{17} = \varsigma_1 = 0,581; \varsigma_{18} = \varsigma_2 = -1,977; \varsigma_{19} = \varsigma_3 = -1,294.$$

В результате прогнозное значение y_t показателей электроэнергии имеют значения:

$$y_{17} = \tau_{17} + \varsigma_1 = 9,459; y_{18} = \tau_{18} + \varsigma_2 = 7,087; y_{19} = \tau_{19} + \varsigma_3 = 7,956.$$

Отсюда можно сделать вывод, что за три квартала следующего пятого года будет потрачено примерно 24,5 у.е. электроэнергии.

Алгоритм построения мультипликативной модели, с выделением сезонной компоненты:

1. Выравнивание исходного ряда y_t методом скользящей средней.
2. Расчет значений сезонной компоненты ς_t .
3. Устранение сезонной компоненты из исходных уравнений ряда и получение выравненных данных $\tau_t \cdot \varsigma_t$.
4. Аналитическое выравнивание уровней $\tau_t \cdot \varepsilon_t$ и расчет значений τ_t с использованием полученного уравнения тренда.
5. Расчет полученных по модели значений $\tau_t \cdot \varsigma_t$.
6. Расчет абсолютной и/или относительной ошибок.

Пример 2.2. Пусть имеются поквартальные данные о прибыли компании за последние четыре года (табл. 16).

Таблица 16

Год	Квартал			
	I	II	III	IV
1	72	100	90	64
2	70	92	80	58
3	62	80	68	48
4	52	60	50	30

Требуется построить мультипликативную модель имеющихся данных и провести прогноз на ближайшие полгода.

Решение. Построим график данного ряда и определим автокорреляционную функцию с целью определения наличия и поведения сезонной составляющей.

Автокорреляционная функция имеет вид (рис. 11): $\rho(k) = \{-0,761 (k = 1); -0,699 (k = 2); -0,650 (k = 3); -0,637 (k = 4); -0,562 (k = 5); -0,413 (k = 6); -0,372 (k = 7)\}$.

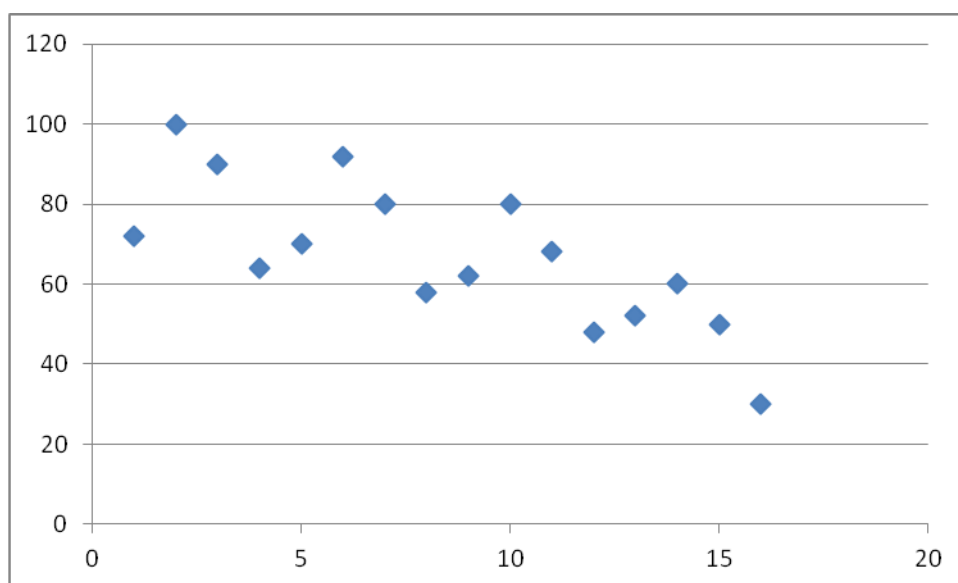


Рис. 11. График исходного ряда

По построенному графику (рис. 11) и по автокорреляционной функции можно сделать вывод, что исследуемый временной ряд содержит сезонные колебания с периодичностью 4. Прибыль компании в весенне-летний период выше, чем в осенне-зимний период. По графику видно наличие убывающей амплитуды колебаний, о чем свидетельствует и автокорреляционная функция, это позволяет построить мультипликативную модель временного ряда.

1. Проведем выравнивание исходных уровней ряда методом скользящей средней:

1) просуммируем уровни ряда последовательно за каждые 4 квартала со сдвигом на один момент времени и определим условные годовые объемы потребления электроэнергии Σ_i (табл. 17, гр. 3);

2) разделим полученные суммы на 4, тем самым найдем скользящие средние (табл. 17, гр. 4);

3) приведем эти значения в соответствие с фактическими моментами времени, для чего найдем средние значения из двух последовательных скользящих средних – центрированные скользящие средние μ_i (табл. 17, гр. 5).

Таблица 17

Номер квартала, t	Потребление электроэнергии, y_t	Суммы за 4 квартала, Σ_i	Скользящая средняя, $\Sigma_i/4$	Центрированная скользящая средняя, μ_i	Оценка сезонной компоненты
1	2	3	4	5	6
1	72	—	—	—	—
2	100	326	81,5	—	—
3	90	324	81	81,25	1,108
4	64	316	79	80	0,800
5	70	306	76,5	77,75	0,900
6	92	300	75	75,75	1,215
7	80	292	73	74	1,081
8	58	280	70	71,5	0,811
9	62	268	67	68,5	0,905

Номер квартала, t	Потребление электроэнергии, y_t	Суммы за 4 квартала, Σ_i	Скользящая средняя, $\Sigma_i/4$	Центрированная скользящая средняя, μ_i	Оценка сезонной компоненты
10	80	258	64,5	65,75	1,217
11	68	248	62	63,25	1,075
12	48	228	57	59,5	0,807
13	52	210	52,5	54,75	0,950
14	60	192	48	50,25	1,194
15	50	—	—	—	—
16	30	—	—	—	—

2. Найдем оценки сезонной компоненты как частное от деления фактических уровней ряда на центрированные скользящие средние y_t / μ_i (табл. 17, гр. 6).

Используем эти оценки для расчета значений сезонной компоненты ζ_i : для этого найдем средние значения за каждый квартал оценки сезонной компоненты ζ_{ii} (табл. 18).

В моделях с сезонной компонентой обычно предполагается, что сезонные воздействия за период взаимопогашаются. В мультипликативной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равной числу периодов в цикле, т.е. в задаче – четырем, так как в условии задачи число периодов одного цикла (год) равно 4 кварталам.

Таблица 18

Показатель	Год	Номер квартала			
		I	II	III	IV
	1	—	—	1,108	0,800
	2	0,900	1,215	1,081	0,817
	3	0,905	1,217	1,075	0,807
	4	0,950	1,194	—	—
Средняя оценка сезонной компоненты для i –квартала, ζ_{ii}		0,918	1,209	1,088	0,808
Скорректированная сезонная компонента, $\bar{\zeta}_{ii}$		0,913	1,202	1,082	0,803

Для данной модели проверим условие взаимопогашаемости:

$$0,918 + 1,209 + 1,088 + 0,808 = 4,023.$$

Определим корректирующий коэффициент $k = \frac{4}{4,023} = 0,994$.

Рассчитаем скорректированные значения сезонной компоненты как произведение ее средней оценки на корректирующий коэффициент: $\bar{\zeta}_{ii} = \zeta_{ii} \cdot k$. В результате этого, условие взаимопогашаемости выполняется точно: $0,913 + 1,202 + 1,082 + 0,803 = 4$.

Таким образом, получены следующие значения сезонной компоненты: для I квартала $\varsigma_1 = 0,913$, для II квартала $\varsigma_2 = 1,202$, для III $\varsigma_3 = 1,082$, для IV $\varsigma_4 = 0,803$.

3. Следующим этапом в исследовании – рассчитаем выровненные значения тенденции τ_t и случайной ошибки ε_t по формуле: $\tau_t \cdot \varepsilon_t = y_t / \varsigma_t$ (табл. 19, гр. 4).

Таблица 19

t	y_t	ς_i	$\tau_t \cdot \varepsilon_t$	τ_t	$\tau_t \cdot \varsigma_t$	ε_t	ε'_t	$(\varepsilon'_t)^2$
1	2	3	4	5	6	7	8	9
1	72	0,913	78,86	87,80	80,16	0,90	-8,16	66,66
2	100	1,202	83,19	85,03	102,20	0,98	-2,20	4,86
3	90	1,082	83,18	82,25	89,00	1,01	1,00	1,00
4	64	0,803	79,70	79,48	63,82	1,00	0,18	0,03
5	70	0,913	76,67	76,70	70,03	1,00	-0,03	0,00
6	92	1,202	76,54	73,93	88,86	1,04	3,14	9,85
7	80	1,082	73,94	71,15	76,99	1,04	3,01	9,08
8	58	0,803	72,23	68,38	54,91	1,06	3,09	9,57
9	62	0,913	67,91	65,60	59,90	1,04	2,10	4,43
10	80	1,202	66,56	62,83	75,52	1,06	4,48	20,08
11	68	1,082	62,85	60,05	64,98	1,05	3,02	9,14
12	48	0,803	59,78	57,28	45,99	1,04	2,01	4,03
13	52	0,913	56,96	54,50	49,76	1,05	2,24	5,02
14	60	1,202	49,92	51,73	62,18	0,97	-2,18	4,73
15	50	1,082	46,21	48,95	52,97	0,94	-2,97	8,79
16	30	0,803	37,36	46,18	37,08	0,81	-7,08	50,12

4. Определим компоненту τ_t данной модели. Для этого проведем аналитическое выравнивание ряда $\tau_t \cdot \varepsilon_t$ с помощью линейного тренда. Используя компьютерную реализацию, получим МНК-оценки линейной регрессии. В результате имеем линейный тренд:

$$\tau_t = 90,58 - 2,775t, \text{ где } t = 1, \dots, 16.$$

(0,225)

Благодаря этому тренду, получаем значения компоненты τ_t (табл. 19, гр. 5). На рис. 12 приведена аппроксимация исходного ряда линейным трендом.

5. Найдем значения уровней ряда, полученные по мультипликативной модели. Для этого вычислим значения $\tau_t \cdot \varsigma_t$ (табл. 19, гр. 6).

6. В соответствии с методикой построения мультипликативной модели расчет ошибки проводится по формуле: $\varepsilon_t = y_t / (\tau_t \cdot \varsigma_t)$ (табл. 19, гр. 7).

Если временной ряд ошибок ε_t не содержит автокорреляции, его можно использовать вместо исходного ряда, изучая взаимосвязи с другими временными рядами.

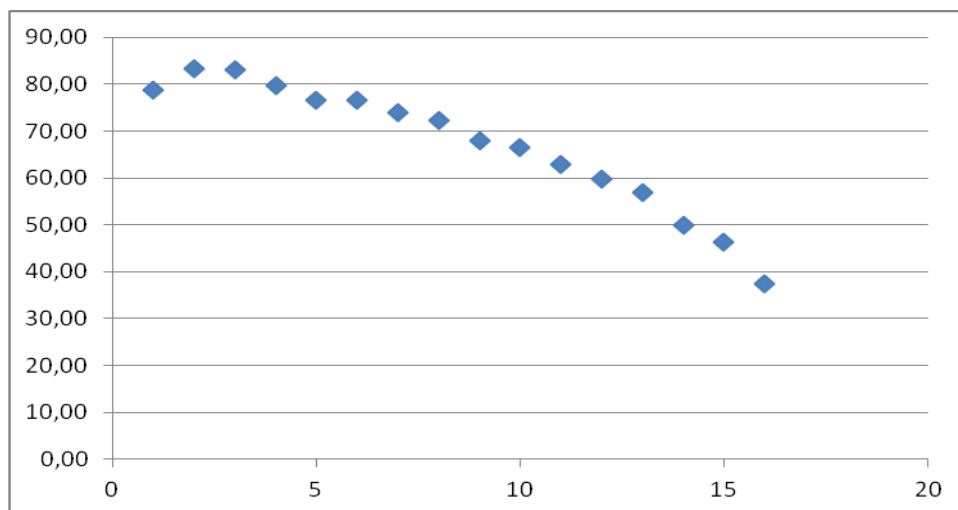


Рис. 12. Аппроксимация линейным трендом

Для того чтобы сравнить мультипликативную модель с другими моделями временного ряда, можно по аналогии с аддитивной моделью использовать сумму квадратов абсолютных ошибок. Абсолютная ошибка в мультипликативной модели определяется по формуле: $\varepsilon'_t = y_t - (\tau_t \cdot \varsigma_t)$ (табл. 19, гр. 8).

По аналогии с регрессионным анализом для оценки качества построенной модели, а также для выбора наилучшей модели, найдем сумму квадратов остатков $(\varepsilon'_t)^2$, которая используется для подсчета коэффициента детерминации.

$$R^2 = \left(1 - \frac{RSS}{TSS}\right) \cdot 100\%, \text{ где } RSS = (E')^2 - \text{сумма квадратов остатков}$$

данного ряда, TSS – общая сумма квадратов.

Для данной модели: $RSS = 207,4$, $TSS = 5023$, в результате коэффициент детерминации равен:

$$R^2 = \left(1 - \frac{207,4}{5023}\right) \cdot 100\% = 0,959 \cdot 100\% \approx 96\%.$$

Таким образом, можно сделать вывод, что мультипликативная модель объясняет 96 % общей вариации уровней временного ряда прибыли компании за последние 16 кварталов.

Проведем прогноз на ближайшие полгода, т.е. два квартала ($t = 17, 18$).

$$\tau_{17} = 90,58 - 2,775 \cdot 17 = 43,41; \tau_{18} = 40,63.$$

$$\varsigma_{17} = 0,913; \varsigma_{18} = 1,202.$$

В результате, прогнозное значение y_t показателей прибыли компании за следующие полгода имеют значения:

$$y_{17} = \tau_{17} + \varsigma_{17} = 44,323; y_{18} = \tau_{18} + \varsigma_{18} = 41,832;$$

$$44,323 + 41,832 = 86,155.$$

2.2. Моделирование сезонных колебаний с помощью фиктивных переменных

Рассмотрим метод моделирования временного ряда, содержащего сезонные колебания, основанный на включении в модель фиктивных переменных. Количество фиктивных переменных принимается равным числу наблюдений в пределах одного цикла колебаний без единицы. Например, при моделировании поквартальных данных необходимо ввести три дополнительные переменные

$$z_1 = \begin{cases} 1, & \text{весна} \\ 0, & \text{не весна} \end{cases}; z_2 = \begin{cases} 1, & \text{лето} \\ 0, & \text{не лето} \end{cases}; z_3 = \begin{cases} 1, & \text{осень} \\ 0, & \text{не осень} \end{cases}. \quad (2.3)$$

Зиме в этом случае соответствуют нулевые значения всех фиктивных переменных. Уравнение регрессии с учетом фиктивных переменных принимает вид

$$y_t = a + b \cdot t + c_1 \cdot z_1 + c_2 \cdot z_2 + c_3 \cdot z_3 + \varepsilon_t. \quad (2.4)$$

Коэффициенты c_i характеризуют отклонение уровней первых трех сезонов по отношению к последнему. Фактор времени в этой модели позволит учесть влияние тенденции. Поэтому модель с фиктивными переменными может рассматриваться как частный случай аддитивной модели временного ряда.

Поскольку фиктивные переменные принимают только два возможных значения 1 или 0, то практически имеем модель тенденции для каждого квартала.

Иными словами, параметры c_i при фиктивных переменных отражают изменение уровня ряда соответствующего квартала под воздействием сезонности по сравнению с кварталом.

Пример 2.3. Предположим, что по данным за десять лет о численности безработных некоторого района (табл. 20) (тыс. чел.) была предложена линейная модель. Используя фиктивные переменные, построить аддитивную модель, содержащую сезонные составляющие.

Таблица 20

t	1	2	3	4	5	6	7	8	9	10
y_t	120	115	95	90	100	98	90	88	89	70

Решение. Введем в рассмотрение фиктивные переменные, которые будут учитывать сезонность данных о численности безработных:

$$z_1 = \begin{cases} 1, & \text{весна} \\ 0, & \text{не весна} \end{cases}; z_2 = \begin{cases} 1, & \text{лето} \\ 0, & \text{не лето} \end{cases}; z_3 = \begin{cases} 1, & \text{осень} \\ 0, & \text{не осень} \end{cases}.$$

По имеющимся данным была построена следующая аддитивная модель:

$$y_t = a + b \cdot t + c_1 \cdot z_1 + c_2 \cdot z_2 + c_3 \cdot z_3 + \varepsilon_t,$$

где коэффициенты определяются обычным МНК:

$$y_t = 112,85 - 3,975 \cdot t + 10,025 \cdot z_1 + 5,33 \cdot z_2 - 0,475 \cdot z_3$$

$$(7,26) \quad (0,83) \quad (6,85) \quad (6,8) \quad (7,49)$$

$$R^2 = 0,85; F = 6,9; p = 0,028.$$

Параметр $b = -3,975$ указывает на тенденцию снижения уровней ряда при элиминировании сезонности, т.е. ежеквартально независимо от влияния сезонности численность безработных снижалась в среднем на 3 975 человек.

Параметры c_1, c_2 показывают, что весной и летом уровни ряда независимо от влияния тенденции были в среднем выше, чем осенью и зимой.

Параметр $a = 112,85$ характеризует уровень безработных в зимние месяцы. В рассматриваемой аддитивной модели параметры c_i при фиктивных переменных показывают не уровни сезонности соответствующего месяца, а их отличие от воздействия сезонности в зимние месяцы.

Вместе с тем, зная показатели c_i , можно определить показатели сезонности. Исходя из содержания параметров при фиксированных переменных, имеем

$$c_1 = S_1 - S_4;$$

$$c_2 = S_2 - S_4;$$

$$c_3 = S_3 - S_4,$$

где S_1, S_2, S_3 и S_4 – показатели сезонности соответствующих кварталов.

Тогда $c_1 + c_2 + c_3 = S_1 + S_2 + S_3 - 3S_4$.

Суммарное значение показателей сезонности для аддитивной модели $\sum_{i=1}^4 S_i = 0$, поэтому можно записать, что $c_1 + c_2 + c_3 = 0 - S_4 - 3S_4$ или $c_1 + c_2 + c_3 = 0 - 4S_4$. Отсюда получим значение сезонной компоненты для IV квартала

$$S_4 = -\frac{1}{4}(c_1 + c_2 + c_3).$$

Для нашего примера, используя значения параметров модели, получим

$$S_4 = -\frac{1}{4}(10,025 + 5,33 - 0,475) = -3,72.$$

Далее определяем сезонные компоненты I, II и III кварталов

$$S_1 = c_1 + S_4;$$

$$S_2 = c_2 + S_4;$$

$$S_3 = c_3 + S_4.$$

По данным примера имеем значения компонент

$$S_1 = 10,025 - 3,72 = 6,305;$$

$$S_2 = 5,33 - 3,72 = 1,61;$$

$$S_3 = -0,475 - 3,72 = -4,195;$$

$$S_4 = -3,72.$$

Условие $\sum_{i=1}^4 S_i = 0$ выполнилось полностью.

Часто в практических исследованиях сезонность изучают по месячным данным, ибо сезонность может проявлять себя и внутри квартала. В этом случае применение модели с фиктивными переменными потребует информации не менее, чем за 7–8 лет, чтобы на каждый параметр модели приходилось доста-

точное число степеней свободы и можно было получить надежные оценки параметров. При ограниченной по числу лет информации изучение сезонности по месячным данным целесообразно вести по аддитивной модели, основанной на разложении уровней временного ряда по компонентам.

2.3. Моделирование сезонных колебаний с помощью рядов Фурье

Ряды Фурье представляют собой декомпозицию динамического ряда на составляющие, которые связаны с частотой колебаний уровней. Построение таких рядов зависит от наличия или отсутствия тенденции в ряду динамики. При отсутствии тенденции, т.е. при стационарном динамическом ряде, методика построения ряда Фурье применяется непосредственно к уровням временного ряда. Если же в изучаемом ряде динамики наблюдается тенденция, то ряд Фурье применяется к отклонениям от тенденции. Соответственно эти различия учитываются и при прогнозировании: по стационарному временному ряду прогноз дается по ряду Фурье, а по ряду с тенденцией – производится суммарный прогноз, исходя из развитий уровней ряда и отклонений от тренда.

Рассмотрим два вида таких рядов.

Ряд Фурье по стационарному ряду

Стационарный ряд с периодическими колебаниями имеет вид (рис. 13).

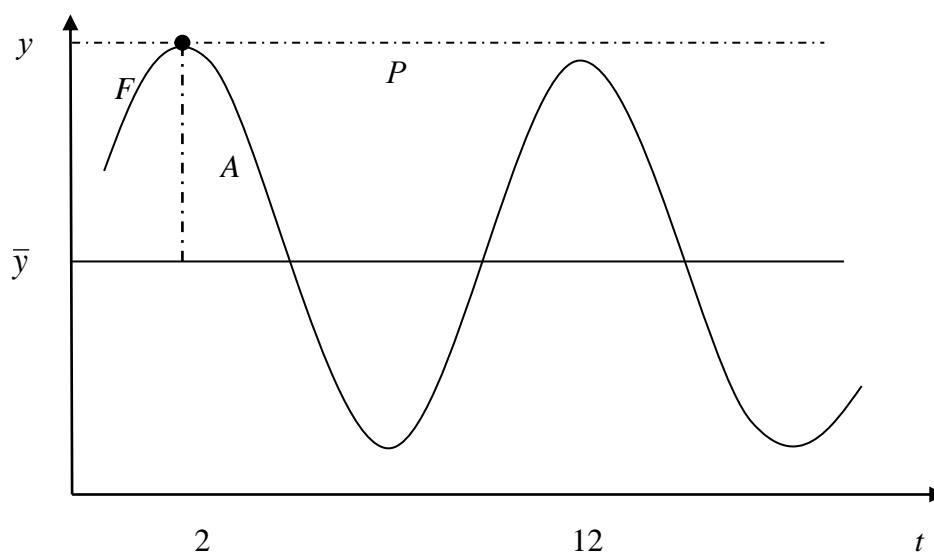


Рис. 13. Стационарный ряд с периодическими колебаниями

Уровни ряда варьируются вокруг среднего значения \bar{y} , а их колебания (волны) повторяются. Интервал времени, необходимый, чтобы динамический ряд начал повторяться, называется *периодом* и обозначен на графике P . Его величина (расстояние между пиками или впадинами) составляет на графике $12 - 2 = 10$ месяцев. Если ряд имеет период P , то он, как правило, имеет также период $2P$, $3P$ и т.д. в общем случае для стационарного периодического временного ряда справедливо равенство:

$$y_t = y_t + cP, \text{ где } c = 1, 2, \dots$$

Величина, обратная периоду, называется *частотой динамического ряда* (f): $f = 1/P$. Частота указывает на число повторений цикла в единицу времени: $f = 1/10$ в месяц (по графику рис. 13).

Отклонение от среднего уровня до пика (или впадины) называется *амплитудой временного ряда* (на рис. 13 величина A).

Расстояние между началом отсчета времени ($t = 0$) и ближайшим пиковым значением называется *фазой* (F).

Стационарный периодический временной ряд можно задать четырьмя параметрами: периодом P или частотой f , амплитудой A , фазой F и средним значением \bar{y} , что может быть представлено в виде

$$y_t = \bar{y} + A \cos W(t - F), \quad (2.5)$$

где W – угловая частота, измеряемая в радианах в единицу времени и равная $W = 2\pi \cdot f$; $0 \leq W \leq 2\pi$; F – фаза.

Рассмотренное выражение (2.5) называется *гармоническим представлением ряда* и часто записывается через синусы и косинусы без упоминания о фазе:

$$y_t = \bar{y} + a \cos Wt + b \sin Wt, \quad (2.6)$$

где $a = A \cos F$ и $b = A \sin F$.

Ввиду того, что $\cos^2 x + \sin^2 x = 1$, то $a^2 + b^2 = A^2$, т.е. существует взаимосвязь между амплитудой колебаний и параметрами гармоник, параметры гармоник также связаны с фазой ряда

$$\operatorname{tg} F = \frac{\sin F}{\cos F} = \frac{b}{a}, \quad \operatorname{arctg} \frac{b}{a} = F.$$

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

$$y_t = \bar{y} + \sum_{i=1}^{\infty} a_i \cos W_i t + \sum_{i=1}^{\infty} b_i \sin W_i t. \quad (2.7)$$

Анализируемые ряды динамики обычно имеют конечную длину N . Поэтому ряд Фурье приобретает вид

$$y_t = \bar{y} + \sum_{i=1}^n a_i \cos W_i t + \sum_{i=1}^n b_i \sin W_i t. \quad (2.8)$$

где $n = N/2$, N – длина временного ряда.

При замене \bar{y} параметром a_0 ряд Фурье принимает вид

$$y_t = a_0 + \sum_{i=1}^n a_i \cos W_i t + \sum_{i=1}^n b_i \sin W_i t. \quad (2.9)$$

Оценка параметров данного уравнения обычно дается МНК. Покажем его применение для случая одной гармоник:

$$y_t = a_0 + a_1 \cos t + b_1 \sin t, \quad (2.10)$$

где t принимает значения от нуля с последующим увеличением на $\frac{2\pi}{N}$.

Система нормальных уравнений примет вид

$$\begin{cases} Na_0 + a_1 \sum \cos t + b_1 \sum \sin t = \sum y_t, \\ a_0 \sum \cos t + a_1 \sum \cos^2 t + b_1 \sum \sin t \cos t = \sum y_t \cos t, \\ a_0 \sum \sin t + a_1 \sum \cos t \sin t + b_1 \sum \sin^2 t = \sum y_t \sin t. \end{cases} \quad (2.11)$$

В этой системе $\sum \cos t = \sum \sin t = 0$. Поэтому из первого уравнения системы получаем, что $a_0 = \frac{\sum y_t}{N} = \bar{y}$. Так как $\sum \sin t \cos t = 0$, то из второго уравнения системы получим оценку параметра a_1 , а из третьего – параметра b_1 .

$$a_1 = \frac{\sum y_t \cos t}{\sum \cos^2 t}, \quad (2.12)$$

$$b_1 = \frac{\sum y_t \sin t}{\sum \sin^2 t}. \quad (2.13)$$

Так как $\cos^2 t = \frac{1 + \cos 2t}{2}$, то

$$\sum \cos^2 t = \frac{\sum (1 + \cos 2t)}{2} = \frac{1}{2} (N + \sum \cos 2t) = \frac{1}{2} (N + 0) = \frac{N}{2}.$$

Аналогично $\sum \sin^2 t = \frac{N}{2}$, т.к. $\sum \sin^2 t = \sum (1 - \cos^2 t) = N - \frac{N}{2} = \frac{N}{2}$.

Следовательно, параметры гармонии определяются как

$$a_1 = \frac{2 \sum y_t \cos t}{N} \text{ и } b_1 = \frac{2 \sum y_t \sin t}{N}.$$

Ряд Фурье с двумя гармониками имеет вид

$$y_t = a_0 + a_1 \cos t + b_1 \sin t + a_2 \cos 2t + b_2 \sin 2t. \quad (2.14)$$

При этом параметры a_0 , a_1 и b_1 соответствуют тем значениям, которые были найдены при рассмотрении одной гармоник. Параметры a_2 и b_2 найдем аналогично:

$$a_2 = \frac{2 \sum y_t \cos 2t}{N} \text{ и } b_2 = \frac{2 \sum y_t \sin 2t}{N}.$$

В общем виде для i гармоник параметры ряда Фурье определяются по формулам

$$a_i = \frac{2 \sum y_t \cos W_i t}{N} \text{ и } b_i = \frac{2 \sum y_t \sin W_i t}{N}. \quad (2.15)$$

Чаще всего описание временного ряда не превышает четырех гармоник.

Пример 2.4. Рассматривается производство товара «Х» по месяцам. Динамика объема представлена в табл. 21.

Таблица 21

Номер месяца	1	2	3	4	5	6	7	8	9	10
y_t	22	24	23	14	6	5	6	8	15	17
Номер месяца	11	12	13	14	15	16	17	18	19	20
y_t	24	25	24	18	8	5	9	14	19	23

Описать модель динамики данного товара, используя ряд Фурье.

Решение. Графическое изображение данного ряда (рис. 14) позволяет сделать вывод, что рассматриваемый ряд является стационарным.

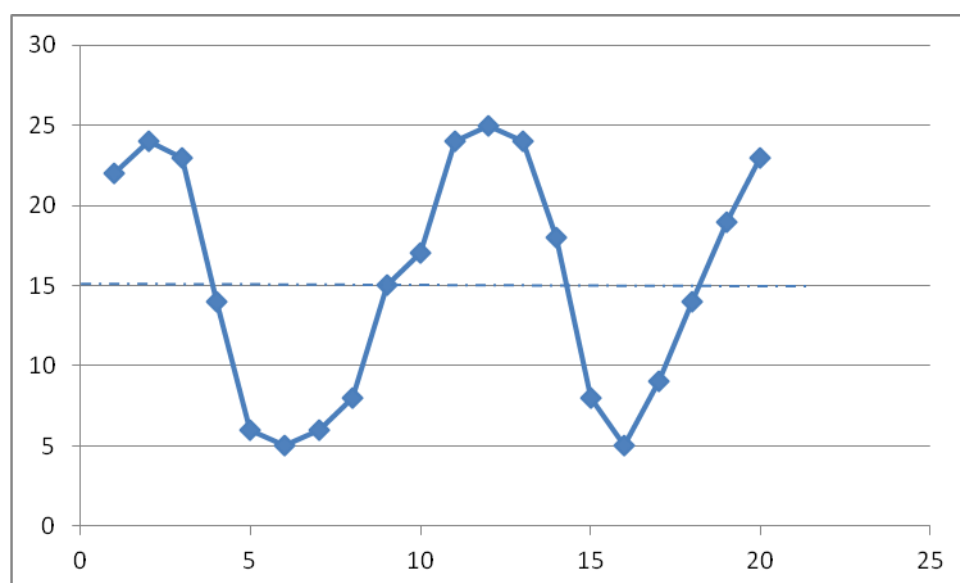


Рис. 14. Периодический ряд динамики производство товара X

Не сложно найти среднее и дисперсию данного ряда: $\bar{y} = 15,45$; $\sigma^2 = 52,15$.

Построим таблицу расчетов для определения параметров ряда Фурье (табл. 22).

Таблица 22

Расчет по ряду Фурье

N	y_t	t	$\cos t$	$\sin t$	$\cos 2t$	$\sin 2t$	$\cos 3t$	$\sin 3t$	$\cos 4t$	$\sin 4t$
1	22	0	1	0	1	0	1	0	1	0
2	24	$0,1\pi$	0,95	0,31	0,81	0,59	0,59	0,81	0,31	0,95
3	23	$0,2\pi$	0,81	0,59	0,31	0,95	-0,31	0,95	-0,81	0,59
4	14	$0,3\pi$	0,59	0,81	-0,31	0,95	-0,95	0,31	-0,81	-0,59
5	6	$0,4\pi$	0,31	0,95	-0,81	0,59	-0,81	-0,59	0,31	-0,95
6	5	$0,5\pi$	0	1	-1	0	0	-1	1	0
7	6	$0,6\pi$	-0,31	0,95	-0,81	-0,59	0,81	-0,59	0,31	0,95
8	8	$0,7\pi$	-0,59	0,81	-0,31	-0,95	0,95	0,31	-0,81	0,59
9	15	$0,8\pi$	-0,81	0,59	0,31	-0,95	0,31	0,95	-0,81	-0,59
10	17	$0,9\pi$	-0,95	0,31	0,81	-0,59	-0,59	0,81	0,31	-0,95
11	24	π	-1	0	1	0	-1	0	1	0
12	25	$1,1\pi$	-0,95	-0,31	0,81	0,59	-0,59	-0,81	0,31	0,95
13	24	$1,2\pi$	-0,81	-0,59	0,31	0,95	0,31	-0,95	-0,81	0,59
14	18	$1,3\pi$	-0,59	-0,81	-0,31	0,95	0,95	-0,31	-0,81	-0,59
15	8	$1,4\pi$	-0,31	-0,95	-0,81	0,59	0,81	0,59	0,31	-0,95
16	5	$1,5\pi$	0	-1	-1	0	0	1	1	0
17	9	$1,6\pi$	0,31	-0,95	-0,81	-0,59	-0,81	0,59	0,31	0,95

N	y_t	t	$\cos t$	$\sin t$	$\cos 2t$	$\sin 2t$	$\cos 3t$	$\sin 3t$	$\cos 4t$	$\sin 4t$
18	14	$1,7\pi$	0,59	-0,81	-0,31	-0,95	-0,95	-0,31	-0,81	0,59
19	19	$1,8\pi$	0,81	-0,59	0,31	-0,95	-0,31	-0,95	-0,81	-0,59
20	23	$1,9\pi$	0,95	-0,31	0,81	-0,59	0,59	-0,81	0,31	-0,95
Σ	309									

Отсчет t ведется с нуля, прибавляя каждый раз величину $\frac{2\pi}{N}$, т.е. в этом случае $0,1\pi$ (графа t). Таблица содержит значения $\cos t$, $\sin t$, ..., $\sin 4t$ для расчета параметров уравнения с четырьмя гармониками

$$y_t = a_0 + a_1 \cos t + b_1 \sin t + a_2 \cos 2t + b_2 \sin 2t + a_3 \cos 3t + b_3 \sin 3t + a_4 \cos 4t + b_4 \sin 4t.$$

Чтобы воспользоваться ранее приведенными формулами a_i и b_i , были найдены по данным табл. 22 следующие значения:

$$\begin{aligned} \Sigma y \cos t &= 6,667; & \Sigma y \sin t &= -17,948; \\ \Sigma y \cos 2t &= 92,883; & \Sigma y \sin 2t &= 26,577; \\ \Sigma y \cos 3t &= -2,698; & \Sigma y \sin 3t &= -10,568; \\ \Sigma y \cos 4t &= -16,753; & \Sigma y \sin 4t &= 11,274. \end{aligned}$$

Ввиду того, что в примере $\frac{2}{N} = 0,1$, то a_i и b_i составят

$$\begin{aligned} a_1 &= 0,6667; & b_1 &= -1,7948; \\ a_2 &= 9,2883; & b_2 &= -2,6577; \\ a_3 &= -0,2698; & b_3 &= -1,0568; \\ a_4 &= -1,6753; & b_4 &= 1,1274. \end{aligned}$$

Соответственно ряд Фурье представит собой следующее выражение:

$$y_t = \bar{y} + \sum_{i=1}^4 C_{it},$$

где C_{it} – гармоники вида

$$C_{it} = a_i \cos W_i t + b_i \sin W_i t.$$

Для примера соответствующие гармоники приведены в табл. 23.

Таблица 23

Периодические составляющие динамического ряда

Номер гармоники	Гармоническая функция	R^2
1	$0,6667 \cos t - 1,7948 \sin t$	0,035
2	$9,2883 \cos 2t - 2,6577 \sin 2t$	0,930
3	$-0,2698 \cos 3t - 1,0568 \sin 3t$	0,942
4	$-1,6753 \cos 4t + 1,1274 \sin 4t$	0,976

Ряд Фурье с одной гармоникой имеет вид

$$\hat{y}_t = 15,45 + 0,6667 \cos t - 1,7948 \sin t,$$

а с двумя:

$$\hat{y}_t = 15,45 + 0,6667 \cos t - 1,7948 \sin t + 9,2883 \cos 2t - 2,6577 \sin 2t.$$

Аналогично записывается модель с тремя и четырьмя гармониками.

Как очевидно из последней графы табл. 23, модель с двумя гармониками хорошо описывает исходный динамический ряд. Хотя коэффициент детерминации R^2 при увеличении числа гармоник и возрастает, но параметры модели для третьей гармоники по t -критерию Стьюдента оказываются статистически незначимыми. Поэтому при выборе модели лучше предпочесть модель с двумя гармониками.

Как очевидно из графика (рис. 14), для рассматриваемого временного ряда амплитуда колебаний A приближается к 10, что и имеет место для уравнения с двумя гармониками

$$A_2 = \sqrt{a_2^2 + b_2^2} = \sqrt{9,2883^2 + 2,6577^2} = 9,66,$$

т.е. интервал, через который ряд начинает повторяться, равен 10 месяцев.

Для уравнения с одной гармоникой период повторения составит 20 месяцев и, естественно, выровненный динамический ряд (ряд \hat{y}_t) плохо аппроксимирует исходные данные (рис. 15).

Для прогноза, используя ряд Фурье с двумя гармониками, в уравнение подставляем следующее по порядку значение t : в примере на 21-й месяц $t = 2\pi$. С учетом того, что $\cos 2\pi = \cos 4\pi = 1$, а $\sin 2\pi = \sin 4\pi = 0$, прогноз окажется следующим:

$$\hat{y}_p = 15,45 + 0,67 \cos 2\pi - 1,79 \sin 2\pi + 9,29 \cos 4\pi - 2,66 \sin 4\pi = 25,4 \text{ ед.}$$

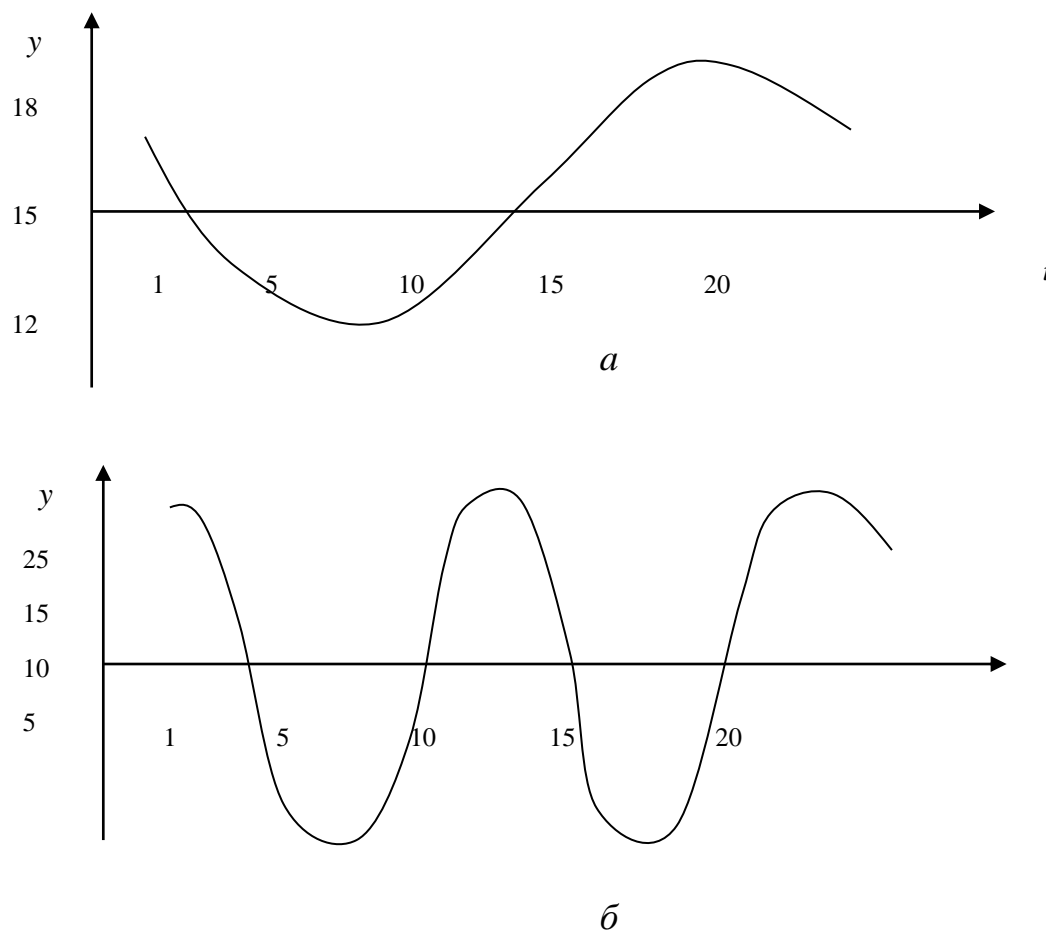


Рис. 15. Ряды с одной (а) и двумя (б) гармониками

Ряд Фурье с тенденцией

В экономике чаще всего встречаются динамические ряды с наличием тенденции. В этом случае при существовании периодических колебаний ряд Фурье может быть использован, если привести ряд к стационарному виду. Для этой цели можно найти линейный тренд $\hat{y}_t = a + bt$ и применить ряд Фурье к остаточным величинам $e_t = y_t - \hat{y}_t$. Возможен и другой путь решения: ряд Фурье строится по первым разностям, что равносильно учету линейного тренда. Иными словами, по ряду динамики определяются цепные абсолютные приросты: $\Delta_t = y_t - y_{t-1}$, которые далее используются как информационная база для построения ряда Фурье.

Пример 2.5. Динамика средних цен на товар по месяцам года характеризуется табл. 24.

Таблица 24

t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	6,5	8,5	8,6	8	7,6	7,8	8,9	11	11,3	10,6	10,7	11

Построить ряд Фурье для дальнейшего прогноза динамики.

Решение. Графическое изображение данного ряда изображено на рис. 16.

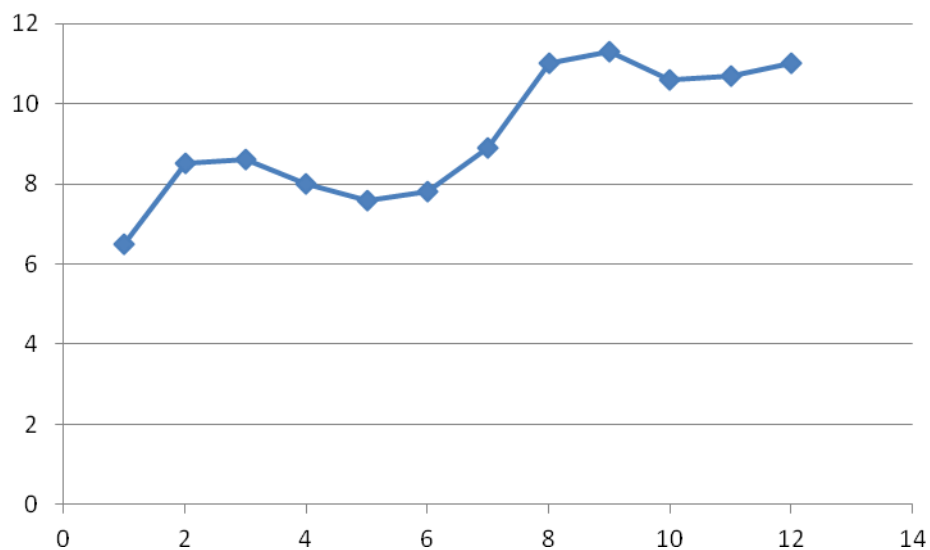


Рис. 16. Графическое изображение данного ряда

График позволяет сделать вывод, что ряд имеет тенденцию. Уравнение линейного тренда имеет вид

$$\hat{y}_t = 6,683 + 0,388t, R^2 = 0,737.$$

Оно описывает 73,7 % вариации средних цен и статистически значимо, так как $F = 28$ при табличном значении $F = 4,84$ (для 5%-ного уровня значимости). Подставив в данное уравнение соответствующие значения t , получим расчетные величины средних цен \hat{y}_t и остатки $e_t = y_t - \hat{y}_t$. График остатков представлен на рис. 17.

Остатки e_t представляют собой стационарный ряд и хорошо описываются рядом Фурье с двумя гармониками

$$e_t = 0,123 \cos t - 0,296 \sin t - 0,137 \cos 2t + 1,005 \sin 2t; R^2 = 0,88.$$

В данном уравнении свободный член a_0 отсутствует, т.к. $\sum e_t = 0$.

Модель рассматриваемого динамического ряда представит собой систему уравнений

$$\begin{cases} y_t = a + bt + e_t, \\ e_t = a_1 \cos t + b_1 \sin t + a_2 \cos 2t + b_2 \sin 2t. \end{cases}$$

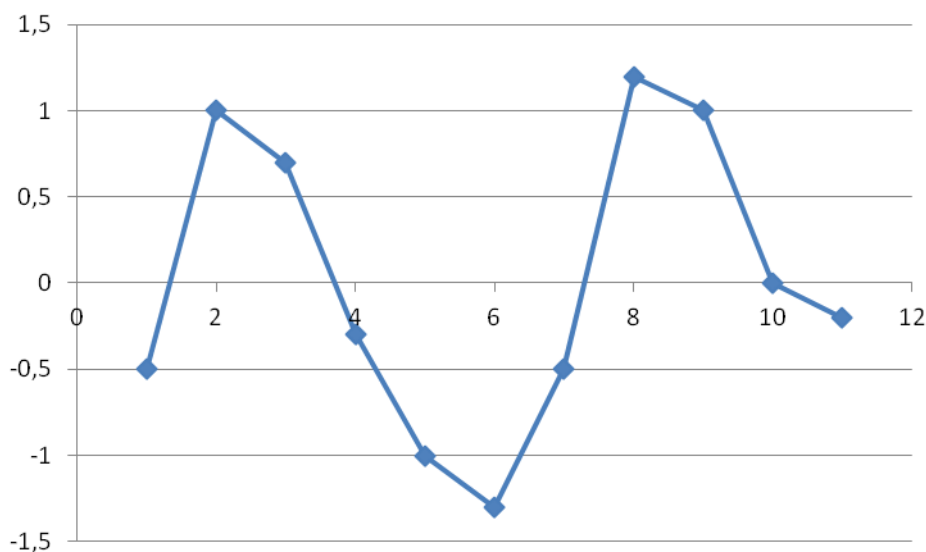


Рис. 17. График остатков от линейного тренда

В примере модель ряда имеет вид

$$\begin{cases} y_t = 6,683 + 0,388t + e_t, \\ e_t = 0,123 \cos t - 0,296 \sin t - 0,137 \cos 2t + 1,005 \sin 2t. \end{cases}$$

Подставив в данную систему значения t (для тренда: 1, 2, ..., 12; для остатков $0, \frac{\pi}{6}, \frac{2\pi}{6}, \dots, \frac{11\pi}{6}$), найдем теоретические значения уровней ряда динамики, которые тесно коррелируют с исходными данными $R^2 = 0,9855$.

Аналогично поступим и для прогноза на следующий месяц года, т.е. строится суммарный прогноз: прогноз по тренду плюс прогноз по ряду Фурье для остаточных величин.

В рассмотренном примере прогноз на 14-й месяц составит:

а) по тренду: $6,683 + 0,388 \cdot 14 = 12,122$;

б) по остаткам: $0,123 \cos \frac{\pi}{6} - 0,296 \sin \frac{13\pi}{6} - 0,137 \cos \frac{13\pi}{6} + 1,005 \sin \frac{13\pi}{6} = 0,123 \cdot 0,866 - 0,296 \cdot 0,5 - 0,137 \cdot 0,5 + 1,005 \cdot 0,866 = 0,76$;

в) общий суммарный итог: $12,122 + 0,76 = 12,882$.

Таким образом, очевидно, что ряд Фурье может использоваться для отображения и прогнозирования динамики с сезонными колебаниями. При этом амплитуда колебаний не должна превышать четыре квартала или 12 месяцев, так как сезонные колебания – это внутригодовые колебания. Вместе с тем сезонные колебания могут изучаться и с помощью иных моделей, позволяющих не только учесть сезонность, но и измерить ее количественно, что имеет, несомненно, практическое значение.

Тестовые задания

Задание 1. Стационарность временного ряда означает отсутствие...

- 1) значений уровней ряда;
- 2) тренда;
- 3) временной характеристики;
- 4) наблюдений по уровням временного ряда.

Задание 2. Известны значения аддитивной модели временного ряда: Y_t – значение уровня ряда, $Y_t = 30$, T – значение тренда, $T = 15$, E – значение компоненты случайных факторов $E = 2$, определите значение сезонной компоненты S .

- 1) $S = 13$;
- 2) $S = 0$;
- 3) $S = 1$;
- 4) $S = -1$.

Задание 3. Модель временного ряда предполагает...

- 1) пренебрежение временными характеристиками ряда;
- 2) отсутствие последовательности моментов (периодов) времени, в течение которых рассматривается поведение экономического показателя;
- 3) независимость значений экономического показателя от времени;
- 4) зависимость значений экономического показателя от времени.

Задание 4. Уровень временного ряда может формироваться под воздействием тенденции, сезонных колебаний и...

- 1) случайных воздействий;
- 2) циклических колебаний;
- 3) динамической составляющей;
- 4) тренда.

Задание 5. При построении модели временного ряда проводится...

- 1) расчет значений компонент для каждого уровня временного ряда;
- 2) расчет последующих и предыдущих значений уровней временного ряда;
- 3) расчет каждого уровня временного ряда;
- 4) расчет средних значений компонент для временного ряда в целом.

Задание 6. Циклические колебания (с периодичностью более 3 лет) связаны с ...

- 1) сезонностью некоторых видов экономической деятельности (сельское хозяйство, туризм и т.д.);
- 2) общей динамикой конъюнктуры рынка;
- 3) трендовыми взаимодействиями между экономическими показателями;
- 4) воздействиями аномальных факторов.

Задание 7. Построена аддитивная модель временного ряда, где Y_t – значение уровня ряда, $T = 10$, T – значение тренда, S – значение сезонной компоненты, E – значение случайной компоненты. Определите вариант правильно найденных значений компонент уровня ряда.

- 1) $T = 5$, $S = 2$, $E = 0$;
- 2) $T = 5$, $S = 2$, $E = 1$;
- 3) $T = 5$, $S = 2$, $E = 3$;
- 4) $T = 7$, $S = 5$, $E = 2$.

Задание 8. Временной ряд называется стационарным, если он является реализацией _____ процесса.

- 1) неслучайного;
- 2) нестационарного стохастического;
- 3) функционального;
- 4) стационарного стохастического.

Задание 9. Под лагом подразумевается число...

- 1) уровней исходного временного ряда;
- 2) пар значений, по которым рассчитывается коэффициент автокорреляции;
- 3) периодов, по которым рассчитывается коэффициент автокорреляции;
- 4) временных рядов, по которым осуществляется расчет коэффициента автокорреляции.

Задание 10. Построена мультипликативная модель временного ряда, где Y_t – значение уровня ряда, $Y_t = 10$, T – значение тренда, S – значение сезонной компоненты, E – значение случайной компоненты. Определите вариант правильно найденных значений компонент уровня ряда.

- 1) $T = 5$, $S = 2$, $E = 3$;
- 2) $T = 5$, $S = 2$, $E = 0$;
- 3) $T = 5$, $S = 2$, $E = 1$;
- 4) $T = 5$, $S = 2$, $E = -1$.

Задание 11. Основной задачей моделирования временных рядов является...

- 1) исключение уровней из совокупности значений временного ряда;
- 2) исключение значений каждой из трех компонент из уровней временного ряда;

3) выявление и придание количественного значения каждой из трех компонент;

4) добавление новых уровней к совокупности значений временного ряда.

Задание 12. Известны значения мультипликативной модели временного ряда: Y_t – значение уровня ряда, $Y_t = 15$, T – значение тренда, $T = 5$, S – значение сезонной компоненты, $S = 3$, определите значение компоненты E (случайные факторы).

1) $E = 1$;

2) $E = -1$;

3) $E = 0$;

4) $E = -3$.

Задание 13. Может ли ряд содержать только одну из компонент?

1) не может, так как временной ряд не содержит компонент, влияющих на его уровни;

2) не может, так как уровень ряда должен формироваться под воздействием всех трех компонент;

3) может, если он представлен данными, описывающими совокупность различных объектов в определенный момент (период) времени;

4) может, если другие две компоненты не участвуют в формировании уровня ряда.

Задание 14. Временной ряд характеризует...

1) данные, описывающие совокупность различных объектов в определенный момент времени;

2) зависимость последовательных моментов (периодов) времени;

3) совокупность последовательных моментов (периодов) времени;

4) данные, описывающие один объект за ряд последовательных моментов (периодов) времени.

Задание 15. Укажите справедливые утверждения по поводу критерия Дарбина – Уотсона:

1) равен 0 в случае отсутствия автокорреляции;

2) изменяется в пределах от 0 до 4;

3) позволяет проверить гипотезу о наличии автокорреляции первого порядка;

4) применяется для проверки гипотезы о наличии гетероскедастичности остатков.

Контрольные вопросы

1. Опишите алгоритм построения аддитивной модели временного ряда.

2. Дайте определение сезонной декомпозиции и корректировки временных рядов.

3. В чем состоит различие аддитивных моделей при наличии и отсутствии тренда в ряду динамики?

4. В чем отличие подходов к оцениванию сезонности в аддитивной и мультипликативной моделях?

5. Как используются фиктивные переменные при моделировании сезонных колебаний?

6. Как изменяется ряд Фурье для оценки периодических колебаний?

7. Как оцениваются параметры модели в виде ряда Фурье?

8. Как используется ряд Фурье при наличии тенденции в ряду динамики?

3. Регрессионные модели временных рядов

Временные ряды как источник информации имеют определенную специфику взаимосвязей, которые учитываются при построении регрессионных моделей. Это связано с возможной ложной корреляцией и ложной регрессией. Высокая корреляция между уровнями временных рядов может иметь место и при отсутствии реальной связи между явлениями. Например, если проконтролировать динамические ряды заработной платы и дорожно-транспортных нарушений, то можно обнаружить коэффициент детерминации, близкий к единице, хотя совершенно очевидно, что такие модели не обязаны быть наилучшими. Установление связи там, где ее на самом деле нет, означает, что имеет место ложная корреляция. Соответственно и уравнение связи в этом случае будет представлять собой ложную регрессию.

Наличие ложной корреляции и ложной регрессии связано с повышающейся тенденцией каждого из рядов динамики, с автокорреляцией их уровней. Поэтому при построении регрессионных моделей по рядам динамики требуется их предварительная специальная обработка.

Если ряды динамики характеризуются наличием тренда, то при построении модели регрессии надо учесть тренд, например, исключить его. В противном случае корреляция уровней ряда динамики будет преувеличена. Предположим, что строится регрессия личных сбережений граждан от доходов населения по данным за ряд лет. Коэффициент детерминации при этом составил 0,95. Может показаться, что получен хороший результат и уравнение регрессии пригодно для прогноза. Однако анализируя остатки ($y - \hat{y}_x$), можно обнаружить наличие а них автокорреляции. Следовательно, полученное уравнение регрессии содержит систематическую погрешность, так как не учитывает влияние тенденции. высокое значение коэффициента детерминации указывает лишь на то, что обоим рядам свойственна тенденция к повышению уровней. Наличие в двух рядах динамики детерминированного тренда приведет, естественно, к ложной регрессии, которой нельзя воспользоваться для прогнозирования, так как остатки будут автокоррелированы и оценка параметра регрессии окажется неэффективной. Модель регрессии по временным рядам пригодна для прогноза, только если остатки $y - \hat{y}_x$ представляют собой стационарный ряд.

Если ряды динамики характеризуются еще и периодическими колебаниями, то при построении модели регрессии следует учесть обе компоненты динамических рядов. В этом случае можно из первоначальных данных исключить как тенденцию, так и периодическую составляющую. Модель регрессии может быть построена либо по остаточным величинам, либо с включением в нее обоих компонент динамического ряда наряду с экономическими переменными.

3.1. Учет тенденции при построении модели регрессии

Методы учета тенденции при построении модели регрессии по временным рядам делятся на две группы:

- методы исключения тенденции из уровней динамического ряда и построение модели по остаточным величинам;
 - включение в модель регрессии фактора времени.
- Рассмотрим эти методы поподробнее.

3.1.1. Методы исключения тенденции

Возможны два подхода для исключения тенденции из уровней временного ряда: метод последовательных разностей и метод отклонений от тренда. Наиболее точным из них является **метод отклонений от тренда**, так как тенденция учитывается в виде уравнения тренда, описывающего закономерность изменения уровней ряда во времени. **Метод последовательных разностей** учитывает тенденцию, представленную полиномом соответствующей степени. Так, если тенденция линейная, то регрессия строится по первым разностям, т.е. абсолютным приростам; если же тенденция характеризуется параболой второй степени, то для модели регрессии используются вторые разности, т.е. абсолютные ускорения.

Метод последовательных разностей

Если в ряде динамики имеется четко выраженная линейная тенденция, то ее можно устранить, перейдя от исходных уровней ряда y_t к цепным абсолютным приростам Δ_t , т.е. первым разностям. Объясняется это тем, что линейный тренд характеризуется постоянным абсолютным приростом. Его величина в уравнении $\hat{y}_t = a + b \cdot t$ соответствует параметру b . Первые разности в линейном тренде будут варьировать за счет случайной составляющей ε_t вокруг своей константы – параметра b . Тенденция в уровнях временного ряда будет устранена. Так, цепной прирост можно представить как

$$\Delta_t = y_t - y_{t-1},$$

и если $y_t = a + b \cdot t + \varepsilon_t$, то

$$\Delta_t = (a + b \cdot t + \varepsilon_t) - (a + b \cdot (t-1) + \varepsilon_{t-1}) = b + (\varepsilon_t - \varepsilon_{t-1}). \quad (3.1)$$

Если ряд динамики характеризуется тенденцией в виде параболы второго порядка, то для ее устранения можно заменить исходные уровни ряда на вторые разности Δ_t'' , т.е. на величину абсолютных ускорений. Как известно, парабола второй степени характеризуется постоянным абсолютным ускорением (вторыми разностями), а первые разности имеют линейную тенденцию. Поэтому для динамического ряда с тенденцией в виде параболы $\hat{y}_t = a + b \cdot t + c \cdot t^2$ вторые разности Δ_t'' будут колебаться вокруг величины $2c$ за счет случайной ошибки ε_t , так как $y_t = \hat{y}_t + \varepsilon_t$ всегда. Соответственно тенденция из исходных данных временного ряда будет устранена.

Представим абсолютное ускорение как

$$\Delta_t'' = \Delta_t - \Delta_{t-1},$$

и если $y_t = a + b \cdot t + c \cdot t^2 + \varepsilon_t$, то

$$\begin{aligned}\Delta_t'' &= (a + bt + ct^2 + \varepsilon_t) - (a + b(t-1) + c(t-1)^2 + \varepsilon_{t-1}) = \\ &= (b - c) + 2ct + (\varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}),\end{aligned}\quad (3.2)$$

т.е. первые разности Δ_t являются линейной функцией от времени t . Вторые разности Δ_t'' окажутся равными

$$\begin{aligned}\Delta_t'' &= (b - c) + 2ct + (\varepsilon_t - \varepsilon_{t-1}) - [(b - c) + 2c(t-1) + (\varepsilon_{t-1} - \varepsilon_{t-2})] = \\ &= 2c + (\varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}).\end{aligned}\quad (3.3)$$

Они не зависят от фактора времени t и могут быть использованы для построения регрессии по временным рядам.

Аналогично можно показать, что если тенденция характеризуется полиномом третьей степени, то для модели регрессии следует использовать третьи разности, чтобы исключить тенденцию из уровней ряда. Однако модели регрессии по вторым, третьим разностям мало информативны с точки зрения их интерпретации и последующего использования в прогнозировании. Поэтому ограничимся рассмотрением регрессии по первым разностям.

При исследовании двух динамических рядов с линейными тенденциями модель линейной регрессии примет вид

$$\Delta y_t = a + b \cdot \Delta x_t + \xi_t, \quad (3.4)$$

где Δy_t , Δx_t – первые разности, ξ_t – случайная ошибка.

Модель (3.4) по существу является моделью скорости роста. Она строится как обычная модель регрессии, но не по уровням динамических рядов, а по их приростам, т.е. по продифференцированным рядам.

Параметр b в модели характеризует среднее изменение скорости ряда y_t с изменением абсолютного прироста x_t ряда на единицу.

Следует заметить, что если модель будет характеризоваться высоким показателем R^2 и отсутствием автокорреляции в остатках, то для прогнозирования конкретных значений y_t можно перейти к уравнению вида

$$y_p = y_n + a + b(x_p - x_n), \quad (3.5)$$

где y_p – прогнозное значение временного уровня ряда y_t ; y_n – конечный уровень ряда y_t ; x_p – прогнозное значение уровня ряда x_t , x_n – конечный уровень ряда x_t .

В данном уравнении величина $x_p - x_n = \Delta x_p$ оценивает прогнозное значение скорости ряда x_t , а $y_p - y_n = \Delta y_p$ – прогнозное значение скорости ряда y_t .

Прогнозное значение фактора x_p может быть дано либо по модели $x_t = f(z_t)$, где z_t – объясняющая переменная, либо по тренду $\hat{x}_p = a + b \cdot t$. От того, насколько хорошо спрогнозировано значение фактора x_p , зависит качество прогноза y_p .

Пример 3.1. Затраты электроэнергии (y_t – тыс. кВт·ч) и объем выпущенной продукции K (x_t – тыс. ед.) характеризуются по предприятию за девять кварталов следующими данными (табл. 25):

Таблица 25

t	1	2	3	4	5	6	7	8	9
x_t	3,4	3,9	4,2	4,6	5,0	5,5	5,8	6,1	6,5
y_t	1,7	2,0	2,2	2,4	2,7	3,1	3,3	3,5	3,9

Построить возможные прогнозные значения для затрат электроэнергии.

Решение. Применим к имеющимся данным МНК и получим следующие уравнения регрессии

$$y_t = -0,748 + 0,701x_t + \varepsilon_t; R^2 = 0,9948.$$

Значение коэффициента детерминации R^2 , близкое к единице, обусловлено наличием линейной тенденции в рассматриваемых временных рядах. Модель регрессии по критерию Дарбина – Уотсона ($D - W = 1,15$) не позволяет отклонить гипотезу об отсутствии автокорреляции в остатках: табличные значения его при 5%-ном уровне значимости составило 0,82 – нижнее и 1,32 – верхнее значения.

Чтобы устранить из данных тенденцию, найдем первые разности (табл. 26).

Таблица 26

Первые разности временных рядов

t	1	2	3	4	5	6	7	8	9
Δx_t	–	0,5	0,3	0,4	0,4	0,5	0,3	0,3	0,4
Δy_t	–	0,3	0,2	0,2	0,3	0,4	0,2	0,2	0,4

Используя МНК, получим уравнение регрессии по первым разностям:

$$\Delta y_t = -0,023 + 0,769\Delta x_t + \xi_t; R^2 = 0,525; D - W = 1,401.$$

Коэффициент регрессии $b = 0,769$ показывает, что увеличение скорости роста объема продукции на 1 тыс. ед. приводит в среднем к увеличению абсолютного прироста затрат электроэнергии на 769 кВт·ч.

В данном уравнении регрессии отсутствует автокорреляция в остатках: критерий Дарбина – Уотсона превышает верхнее табличное значение 1,33.

Прогноз на 10-й год выполним по уравнению

$$y_p = y_n + a + b(x_p - x_n), \text{ где } y_n = 3,9, x_n = 6,5;$$

x_p найдем по уравнению линейного тренда

$$\hat{x}_t = 3,075 + 0,385 \cdot t; R^2 = 0,997;$$

при $t = 10$, $x_p = 6,925$.

Соответственно $y_p = 3,9 - 0,023 + 0,769(6,925 - 6,5) = 4,2$ тыс. кВт·ч.

Пример 3.2. Администрация торговой сети интересуется, есть ли взаимосвязь между объемом продаж и средним возрастом сотрудников компании. Для этого были собраны данные за последние десять лет (табл. 27).

Таблица 27

t	1	2	3	4	5	6	7	8	9	10
Объем продаж, y_t	300	310	325	340	350	370	385	425	435	450
Средний возраст, x_t	20	22	25	30	32	29	31	33	32	28

Известно также: $\sum y_t = 3690$, $\sum y_t^2 = 1387300$, $\sum x_t = 282$, $\sum x_t^2 = 8132$, $\sum x_t y_t = 105555$.

Уравнения трендов для каждого из рядов составили:

а) для ряда x_t : $\hat{x}_t = 22,13 + 1,103 \cdot t$;

б) для ряда y_t : $\hat{y}_t = 288,42 + 9,75 \cdot t + 0,7 \cdot t^2$.

Задание.

1. Определить коэффициент корреляции между изучаемыми рядами по их уровням.

2. Определить коэффициент корреляции между изучаемыми рядами по отклонениям от указанных выше линейного и параболического трендов соответственно.

3. Выбрав одну из полученных мер, охарактеризовать тесноту связи между временными рядами объемов продаж и долей женщин среди работников компании. Обоснуйте ваш выбор.

Решение.

1) Определим коэффициент корреляции по имеющимся уровням временных рядов:

$$r_{xy} = \frac{n \sum x_t y_t - \sum x_t \sum y_t}{\sqrt{(n \sum x_t^2 - (\sum x_t)^2)(n \sum y_t^2 - (\sum y_t)^2)}} = \frac{10 \cdot 105555 - 282 \cdot 3690}{\sqrt{(10 \cdot 8132 - (282)^2)(10 \cdot 1387300 - (3690)^2)}} = 0,697.$$

2) Построим оцененные значения для \hat{x}_t и \hat{y}_t , используя имеющиеся модели (табл. 28):

Таблица 28

t	1	2	3	4	5	6	7	8	9	10
\hat{y}_t	298,87	310,72	323,97	338,62	354,67	372,12	390,97	411,22	432,87	455,92
\hat{x}_t	23,23	24,33	25,44	26,54	27,65	28,75	29,85	30,95	32,06	33,16

Разности между этими рядами представлены в табл. 29.

Таблица 29

t	1	2	3	4	5	6	7	8	9	10
Δy_t	1,13	-0,72	1,03	1,38	-4,67	-2,12	-5,97	13,78	2,13	-5,92
Δx_t	-3,23	-2,34	-0,44	3,46	4,36	0,25	1,15	2,05	-0,06	-5,16

Тогда коэффициент корреляции по первым разностям равен $r_{\Delta x \Delta y} = 0,22$.

3) Данное значение свидетельствует, что отклонения между изучаемыми рядами несущественны и модели содержат только тенденции.

Метод отклонений от тренда

Рассматриваемые для модели регрессии ряды динамики могут иметь разные тенденции, так как уравнение тренда записывается любой математической функцией. Например, ряд x_t описывается гиперболой, а ряд y_t – параболой. В этом случае метод отклонений от тренда позволяет исключить из каждого временного ряда соответствующую ему тенденцию.

Алгоритм построения регрессии при применении метода отклонений следующий.

1. Для каждого временного ряда определяются уравнение тренда и теоретические значения \hat{y}_t ; \hat{x}_t .

2. По каждому из рядов находятся остаточные величины $dy = y_t - \hat{y}_t$; $dx = x_t - \hat{x}_t$.

3. Строится модель регрессии

$$dy = f(dx). \quad (3.5)$$

В линейной регрессии $dy = a + b \cdot dx + \xi_t$ параметр b показывает как в среднем изменяется величина случайных отклонений по ряду y_t с изменением случайных колебаний ряда x_t на единицу. Если при этом оба ряда характеризуются линейной тенденцией, то параметр $a = 0$, так как $\sum dy = \sum dx = 0$.

Тогда модель линейной регрессии примет вид $dy = b \cdot dx + \xi_t$ и параметр b будет выступать коэффициентом пропорциональности. Его величина будет показывать, во сколько раз случайные отклонения по ряду y_t в среднем выше (ниже) случайных отклонений по ряду x_t .

Для прогноза конкретных значений y_t можно перейти к уравнению, связывающему между собой уровни временных рядов. С этой целью в модель регрессии $dy = a + b \cdot dx$ подставим значения dy и dx , раскрыв их содержание, т.е. $dy = y_t - \hat{y}_t$ и $dx = x_t - \hat{x}_t$.

Тогда имеем, например, для линейной регрессии $d\hat{y} = a + b \cdot dx$, т.е. $y_t - \hat{y}_t = a + b(x_t - \hat{x}_t)$, или $y_t = \hat{y}_t + a + b(x_t - \hat{x}_t)$.

Данную модель можно использовать для прогноза

$$y_p = \hat{y}_{t=p} + a + b(x_p - \hat{x}_{t=p}), \quad (3.6)$$

где y_p – прогнозное значение; $\hat{y}_{t=p}$ – прогноз по тренду при $t = p$; x_p – прогнозное значение x , найденное либо по модели регрессии, либо как $x_n + \Delta x_n$; $\hat{x}_{t=p}$ – прогноз x исходя из уравнения тренда при $t = p$.

Результат прогноза зависит от качества прогноза фактора x и от качества трендовых моделей, используемых в прогнозировании.

Пример 3.3. По данным за 10 мес. рассматривается зависимость прибыли предприятия (y_t – тыс. р.) от затрат на рекламу (x_t – тыс. р.) (табл. 30).

Таблица 30

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	14	20	18	23	24	25	27	24	25	26

Применить к рассматриваемым динамическим рядам метод отклонений от тренда.

Решение. Каждый из представленных рядов имеет повышающуюся тенденцию, которая достаточно хорошо описывается степенной функцией (рис. 18а и 18б).

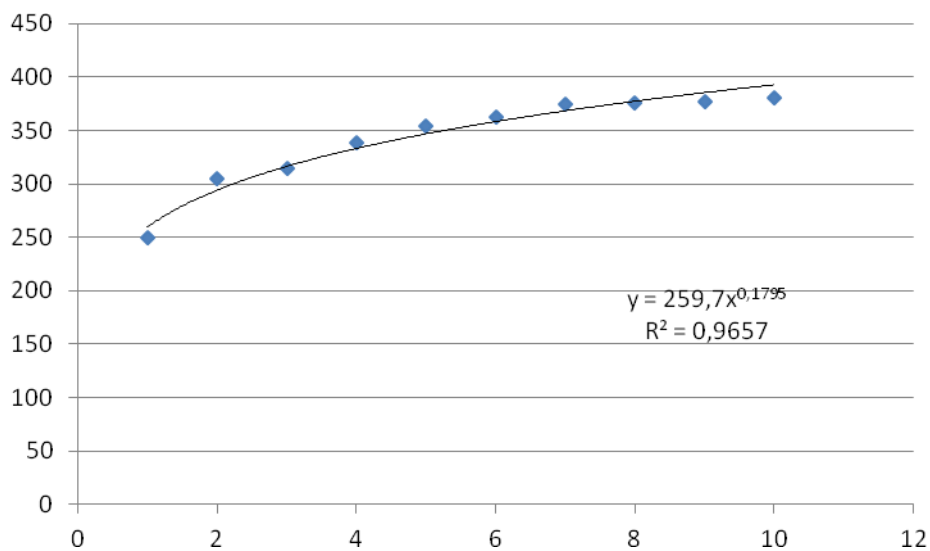


Рис. 18а. Линия тренда прибыли предприятия y_t

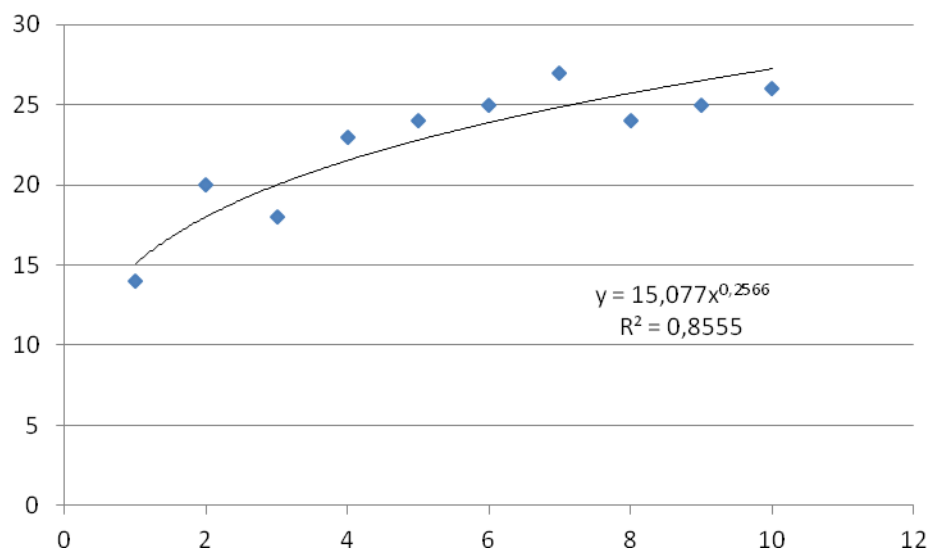


Рис. 18б. Линия тренда затрат на рекламу x_t

Для ряда прибыли уравнение тренда составило $\hat{y}_t = 259,7 \cdot t^{0,1795}$; $R^2 = 0,9657$, а для ряда затрат на рекламу – $\hat{x}_t = 15,077 \cdot t^{0,2566}$; $R^2 = 0,8555$. Автокорреляция в остатках отсутствует: $r_{a_1} = 0,116$ для ряда y_t и $r_{a_1} = -0,129$ для ряда x_t .

Если коррелировать исходные уровни рядов динамики x_t и y_t , то коэффициент детерминации R^2 составит 0,925, а уравнение регрессии запишется в виде $\hat{y}_t = 116,621 + 10,026 \cdot x_t$. Однако ввиду наличия в каждом из рядов четкой тенденции можно предположить, что результаты регрессионно-корреляционного анализа завышены. Поэтому используем метод устранения тенденции, найдя отклонения от тренда $dy = y_t - \hat{y}_t$ и $dx = x_t - \hat{x}_t$ (табл. 31).

Таблица 31

Расчет остаточных величин для построения модели регрессии

t	y_t	\hat{y}_t	dy	x_t	\hat{x}_t	dx
1	250	259,7	-9,7	14	15,1	-1,1
2	305	294,1	10,9	20	18	2
3	314	316,3	-2,3	18	20	-2
4	338	333,1	4,9	23	21,5	1,5
5	354	346,7	7,3	24	22,8	1,2
6	363	358,2	4,8	25	23,9	1,1
7	375	368,3	6,7	27	24,8	2,2
8	376	377,2	-1,2	24	25,7	-1,7
9	377	385,3	-8,3	25	26,5	-1,5
10	380	392,6	-12,6	26	27,2	-1,2

Применяя к рядам отклонений dy и dx МНК, получим уравнение линейной регрессии $dy = -0,1398 + 4,0589dx$; $R^2 = 0,7022$. Полученное уравнение регрессии показывает, что при устранении из исходных уровней временных рядов тенденции имеет место связь между остаточными величинами. Поэтому данную модель можно использовать в прогнозировании. На это указывает и отсутствие в модели автокорреляции остатков. Коэффициент автокорреляции равен $-0,1372$, а критерий Дарбина – Уотсона равен 1,679. Сравнивая с табличными значениями величину $(4 - DW)$ при $\alpha = 0,05$ и числе степеней свободы, равном 10, увидим, что фактическое значение критерия превышает верхнюю границу 1,32.

Для прогноза на 11-й мес. воспользуемся уравнением вида (3.6)

$$y_p = \hat{y}_{t=p} + a + b(x_p - \hat{x}_{t=p}),$$

где $\hat{y}_{t=p} = 259,7 \cdot t^{0,1795} = 259,7 \cdot 11^{0,1795} = 399,39$; $x_p = 27$ (взято как $x_n = 26$ и плюс $\Delta_n = 1$); $\hat{x}_{t=p} = 15,077 \cdot t^{0,2566} = 15,077 \cdot 11^{0,2566} = 27,9$.

Тогда прогнозное значение y_p составит

$$y_p = 399,39 - 0,1398 + 4,0589(27 - 27,9) = 395,6 \text{ тыс. р.}$$

Пример 3.4. Имеются данные за десять лет об объеме производства и мощностях на одном из предприятий промышленности области.

Таблица 32

Год	1	2	3	4	5	6	7	8	9	10
Объем производства, y_t	6,79	6,88	7,07	7,17	7,33	7,52	7,62	7,72	7,8	7,98
Мощности x_t	65,6	68,1	70,4	73,3	76,5	78,6	81	83	85,4	85,9

Результаты аналитического выравнивания привели к следующим уравнениям трендов для каждого из рядов:

а) для временного ряда объема производства:

$$\hat{y}_t = 6,61 + 0,15 \cdot t - 0,0015 \cdot t^2, (t = -3, -1, 1, 3);$$

б) для временного ряда мощностей:

$$\hat{x}_t = 63,7 + 2,38 \cdot t, (t = -3, -1, 1, 3).$$

Задание.

1. Определить коэффициент корреляции между временными рядами, используя непосредственно исходные уровни, первые разности для мощностей и вторые разности для объема производства, отклонения от основной тенденции.

2. Объяснить различия полученных результатов.

3. Рассчитать коэффициент автокорреляции 1 и 2 порядков внутри каждого временного ряда.

Решение.

1) Найдем коэффициент корреляции между непосредственными уровнями исходных рядов:

$$r_{x_t y_t} = \frac{n \sum x_t y_t - \sum x_t \sum y_t}{\sqrt{(n \sum x_t^2 - (\sum x_t)^2)(n \sum y_t^2 - (\sum y_t)^2)}} = 0,994.$$

Очевидно, что между рассматриваемыми уровнями рядов x_t и y_t имеется высокая корреляция.

Рассчитаем последовательные разности для конкретных временных промежутков (табл. 33), используя формулы (3.1) $\Delta_t = y_t - y_{t-1}$ и (3.2) $\Delta_t'' = \Delta_t - \Delta_{t-1}$:

Таблица 33

Расчет последовательных разностей

t	\hat{y}_t	\hat{y}_{t-1}	Δ_t	Δ_{t-1}	Δ_t''	\hat{x}_t	\hat{x}_{t-1}	Δ_t
-3	6,15	—	—	—	—	56,56	—	—
-1	6,46	6,15	0,31	—	—	61,32	56,56	4,76
1	6,76	6,46	0,3	0,31	-0,01	66,08	61,32	4,76
3	7,05	6,76	0,29	0,3	-0,01	70,84	66,08	4,76

Равные разности Δ_t'' и Δ_t для рассматриваемых периодов t свидетельствуют о том, что рассматриваемые временные ряды стабильны и характеризуются высокой корреляцией $r_{\hat{x}_t \hat{y}_t} = 0,9996$.

2) Рассчитаем отклонения от трендов, используя метод устранения тенденции $dy = y_t - \hat{y}_t$ и $dx = x_t - \hat{x}_t$ (табл. 34).

Таблица 34

Расчет отклонений от трендов

t	y_t	\hat{y}_t	dy	x_t	\hat{x}_t	dx
1	6,79	6,759	0,031	65,6	66,08	-0,48
2	6,88	6,904	-0,024	68,1	68,46	-0,36
3	7,07	7,047	0,023	70,4	70,84	-0,44
4	7,17	7,186	-0,016	73,3	73,22	0,08
5	7,33	7,323	0,007	76,5	75,60	0,90
6	7,52	7,456	0,064	78,6	77,98	0,62
7	7,62	7,587	0,034	81,0	80,36	0,64
8	7,72	7,714	0,006	83,0	82,74	0,26
9	7,80	7,839	-0,039	85,4	85,12	0,28
10	7,98	7,960	0,020	85,9	87,50	-1,60

Коэффициент корреляции по построенным отклонениям равен $r_{x_t - \hat{x}_t, y_t - \hat{y}_t} = 0,0476$. Низкое значение коэффициента свидетельствует о том, что изучаемые ряды не зависят от фактора времени.

Метод отклонений показывает, что разброс значений невелик и все отклонения концентрируются около нуля. Высокое значение $r_{\hat{x}_t, \hat{y}_t} = 0,9996$ свидетельствует о том, что оцененная регрессия показывает себя значительно лучше, чем рассматриваемые уровни.

3) Рассчитаем коэффициенты автокорреляции по каждому из временных рядов:

по ряду y_t : $r_1 = 0,9919$; $r_2 = 0,9899$;

по ряду x_t : $r_1 = 0,9957$; $r_2 = 0,9935$.

Высокие значения коэффициентов автокорреляции 1 и 2 порядков подтверждают наличие высокой корреляции и линейной тенденции.

3.1.2. Включение в модель фактора времени

Модель регрессии по двум временным рядам *с включением* в нее как отдельной независимой переменной **фактора времени** t имеет вид

$$y = a + b \cdot x + c \cdot t + \xi \quad (3.7)$$

где $t = 1, 2, \dots, n$.

Включая в регрессию фактор времени t , мы устраняем линейную тенденцию из уровней временных рядов. Это объясняется спецификой множественной регрессии: коэффициенты регрессии показывают изолированное влияние на результат соответствующего фактора при неизменном уровне других факторов. В рассматриваемом двухфакторном уравнении регрессии (3.7) коэффициент регрессии b характеризует «чистое» воздействие переменной x на результат y в условиях неизменной тенденции, т.е. при ее устранении.

Математически доказано, что если временные ряды характеризуются линейной тенденцией, то включение в модель фактора времени t равносильно построению модели регрессии по отклонениям от трендов с последующим переходом от нее к исходным уровням временного ряда зависимой переменной y . Это связано с тем, что уравнение регрессии $\hat{y} = a + b \cdot x + c \cdot t$ может быть построено двумя путями:

- применяя метод наименьших квадратов, получаем оценки параметров a, b, c (именно так строится данная модель при компьютерной обработке);
- последовательно включаем в модель линейную тенденцию ряда y и линейную регрессию остаточных величин $dy = b \cdot dx + \xi$, где dy, dx – остаточные величины от линейных тенденций.

Существует и другой подход построения линейной модели регрессии с включением фактора времени. Алгоритм построения модели следующий.

1. Строится линейное уравнение тренда для ряда y_t

$$\hat{y}_t = \tilde{a} + \tilde{c} \cdot t.$$

2. Строится линейное уравнение тренда для ряда x_t

$$\hat{x}_t = A + B \cdot t.$$

3. Находятся остаточные величины dy и dx

$$dy = y_t - (\tilde{a} + \tilde{c} \cdot t);$$

$$dx = x_t - (A + B \cdot t).$$

4. Строится регрессия по отклонениям от трендов

$$dy = b \cdot dx + \xi.$$

5. Определяется модель ряда y_t

$$y_t = \hat{y}_t + dy \text{ или } y_t = \tilde{a} + \tilde{c} \cdot t + b(x_t - A - B \cdot t) + \xi.$$

Откуда имеем уравнение

$$y_t = (\tilde{a} - bA) + (\tilde{c} - bB) \cdot t + b \cdot x_t + \xi. \quad (3.8)$$

Данное уравнение соответствует уравнению регрессии (3.7):

$$y_t = a + b \cdot x_t + c \cdot t + \xi,$$

где

$$a = \tilde{a} - bA; \quad (3.9)$$

$$c = \tilde{c} - bB. \quad (3.10)$$

Рассмотренный подход к построению модели регрессии позволяет понять, что уравнение регрессии $y_t = a + b \cdot x_t + c \cdot t + \xi$ с включением фактора времени t учитывает линейные тенденции для временных рядов y_t и x_t . Кроме того, строя регрессию по отклонениям от линейных трендов, мы получаем остатки ξ_t те же, что и в регрессии с включением линейного фактора времени t . Поэтому при наличии в рядах линейных тенденций целесообразно строить модель регрессии по исходным уровням рядов динамики, включая в нее фактор времени t . В этом случае модель регрессии по отклонениям от трендов не информативна. Тем более надо учесть, что регрессия по отклонениям от линейных

трендов является составной частью регрессии с включением фактора времени t (шаг 4 при втором подходе).

В регрессии $\hat{y} = a + b \cdot x + c \cdot t$ параметр b показывает, на сколько единиц изменяется в среднем y при изменении x на одну единицу в условиях неизменной тенденции; параметр c показывает средний абсолютный прирост y в условиях неизменного уровня объясняющей переменной x .

Пример 3.5. Имеются данные за 3 года некоторого промышленного предприятия в поквартальном разрезе уровня производительности труда (y – в тыс. р. на одного работника) и доли активной части основных фондов (x – в %) (табл. 35).

Таблица 35

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	9,9	18,9	19,8	27,9	22,2	29,7	38,7	36	46	37,8	45	54

Рассмотреть модель, включая в нее фактор времени.

Решение. Модель регрессии с включением в нее фактора времени t имеет вид:

$$\hat{y}_t = 2,94 + 0,104 \cdot x_t + 0,53 \cdot t; R^2 = 0,977; F = 188,3.$$

(5,47) (2,43) (3,44)

В целом уравнение регрессии значимо, как и его параметры (табличное значение F - критерия при $\alpha = 0,05$ и числе степеней свободы 2 и 9 равно 4,26, а табличное значение t - критерия Стьюдента при $\alpha = 0,05$ и $df = 9$ равно 2,26). Эти же результаты получим, применяя метод последовательного включения в модель регрессии линейной тенденции ряда y_t и линейной регрессии отклонений от трендов.

Так, в рассматриваемом примере уравнение линейного тренда для ряда y_t составило:

$$\hat{y}_t = 3,954 + 0,892 \cdot t,$$

где $t = 1, 2, \dots, 12$.

Уравнение хорошо описывает тенденцию: $R^2 = 0,961$, коэффициент автокорреляции в остатках равен 0,0083.

Линейное уравнение тренда для ряда x_t составило:

$$\hat{x}_t = 9,729 + 3,451 \cdot t; R^2 = 0,9104.$$

Для каждого ряда были рассчитаны отклонения от трендов

$$dy = y_t - \hat{y}_t \text{ и } dx = x_t - \hat{x}_t.$$

Получено уравнение регрессии по отклонению трендов:

$$dy = 0,104dx + \xi; R^2 = 0,396.$$

Далее, с использованием формул (3.9), (3.10), были получены оценки параметров регрессии $y = a + b \cdot x + c \cdot t + \xi$

$$a = 3,95 - 0,104 \cdot 9,729 = 2,943;$$

$b = 0,104$ (совпадает с величиной коэффициента регрессии по уравнению $dy = b \cdot dx + \xi$);

$$c = 0,892 - 0,104 \cdot 3,451 = 0,533.$$

Данные оценки совпадают с тем, что было получено ранее по модели с включением в нее фактора времени как объясняющей переменной.

Остаточные величины ξ_t по модели для отклонений от трендов совпадают с остатками для регрессии с включением фактора времени. Автокорреляция в остатках небольшая: $r_{a_1} = 0,296$.

В рассматриваемой модели параметр b показывает, что рост доли активной части основных фондов на 1%-ный пункт в условиях неизменной тенденции способствует росту уровня производительности труда на 0,104 тыс. р. Параметр c характеризует среднеквартальный прирост производительности труда независимо от изменения доли активной части основных фондов, т.е. обусловленный влиянием других факторов, не учитываемых в регрессии.

Принцип введения в модель фактора времени сохраняется и при изучении трех и более связанных рядов динамики. Так, если строится регрессия $y = f(x_1, x_2, x_3)$, то включение в нее фактора времени t приводит чаще всего к модели вида:

$$\hat{y}_t = a + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + c \cdot t. \quad (3.11)$$

В ней параметры b_1, b_2 и b_3 показывают изолированное воздействие каждой объясняющей переменной на результат y , а параметр c – средний абсолютный прирост y в условиях неизменности значений переменных x_1, x_2 и x_3 .

Время в качестве независимой переменной часто вводится в виде линейного члена даже если другие переменные подвергаются логарифмированию или иному преобразованию. Например, производственная функция с включением фактора времени часто записывается:

$$P = a \cdot K^{b_1} \cdot L^{b_2} \cdot e^{ct}, \quad (3.12)$$

где P – объем продукции; K – основной капитал; L – занятость; e – экспонента; t – фактор времени, взятый как натуральный ряд.

Линеаризуем данную зависимость:

$$\ln P = \ln a + b_1 \ln K + b_2 \ln L + ct. \quad (3.13)$$

Здесь фактор времени t введен в модель линейно. Рассматриваемая производственная функция не линейна относительно оцениваемых параметров. В ней параметры b_1 и b_2 являются коэффициентами эластичности, показывая, на сколько процентов повышается объем продукции при увеличении соответствующего фактора (K и L) на 1 % в условиях неизменности тенденции.

Параметр c обычно интерпретируется как автономный рост объема продукции в условиях неизменности факторов производства K и L . Так, если $c = 0,0175$, то $e^{0,0175} = 1,0177$ и, следовательно, ежегодно (если t – годы) объем продукции возрастает в среднем при неизменных уровнях затрат капитала и труда в 1,0177 раза, или на 1,77 %.

Если тенденция в рядах динамики характеризуется полиномом второй и более высоких степеней, то в модель регрессии вводятся t и t^2 , а иногда t в более высокой степени.

В этом случае рассматривается регрессия вида

$$\hat{y} = a + b \cdot x + c \cdot t + d \cdot t^2 \quad (3.14)$$

при двух временных рядах

$$\text{или } \hat{y} = a + \sum_{i=1}^p b_i \cdot x_i + c \cdot t + d \cdot t^2 \quad (3.15)$$

при p временных рядах.

Вводя в модель регрессии фактор времени в виде t, t^2, \dots, t^k , предполагаем, что коэффициенты при переменных остаются во времени неизменными и характеризуют силу связи результата y с соответствующей объясняющей переменной x .

Если предполагается, что в регрессии коэффициенты при независимой переменной подвержены изменению во времени, то в модель можно ввести преобразованные переменные $t \cdot x$ (где t – время). Оценка параметров модели дается МНК.

Модель регрессии с включением в нее фактора времени как независимой переменной не всегда эффективна ввиду возможной мультиколлинеарности факторов. Если временные ряды, используемые в регрессии, характеризуются четкой тенденцией ($R^2 > 0,9$), то корреляция t и x_i может превышать корреляцию x_i с y , и параметры регрессии при объясняющих переменных x_i оказываются ненадежными и экономически не интерпретируемыми.

3.2. Обобщенный метод наименьших квадратов

Существуют различные методы устранения автокорреляции в остатках. Они зависят от причин автокорреляции. Автокорреляция в остатках может быть следствием неправильной спецификации модели: не учтена важная объясняющая переменная, неправильно выбрана форма связи. В этом случае стараются изменить математическую модель, например, линейную модель заменяют степенной, или уточнить набор объясняющих переменных. Однако если эти попытки не увенчались успехом и автокорреляция в остатках имеет место быть, то для ее устранения можно применить обобщенный метод наименьших квадратов (ОМНК).

ОМНК можно использовать как для парной, так и для множественной регрессии. Для простоты решения проблемы рассмотрим регрессию двух временных рядов:

$$y_t = a + b \cdot x_t + \xi_t. \quad (3.16)$$

Для периода времени $(t-1)$ справедливо равенство

$$y_{t-1} = a + b \cdot x_{t-1} + \xi_{t-1}. \quad (3.17)$$

Если имеет место автокорреляция в остатках, т.е. последующие во времени остатки зависят от предыдущих, то регрессия остатков может быть представлена как

$$\xi_t = c + d\xi_{t-1} + V_t, \quad (3.18)$$

где V_t – случайная ошибка для линейной регрессии остатков.

Но так как $\xi_t = y_t - \hat{y}_t$, где \hat{y}_t – теоретические уровни линейной регрессии y от x , то $\sum \xi_t = 0$ и $\bar{\xi}_t = 0$. Полагая, что $\bar{\xi}_{t-1} \cong \bar{\xi}_t$, имеем $c = \bar{\xi}_t - d\bar{\xi}_{t-1} = 0$. Тогда регрессия остатков примет вид:

$$\xi_t = d\xi_{t-1} + V_t. \quad (3.19)$$

Параметр d определим по формуле:

$$d = \frac{\text{cov}(\xi_t, \xi_{t-1})}{\sigma_{\xi_{t-1}}^2}, \quad (3.20)$$

где $\text{cov}(\xi_t, \xi_{t-1}) = \overline{\xi_t \xi_{t-1}} - \bar{\xi}_t \cdot \bar{\xi}_{t-1} = \overline{\xi_t \xi_{t-1}}$.

В результате получим, что $d = \frac{\overline{\xi_t \xi_{t-1}}}{\sigma_{\xi_{t-1}}^2}$. Предполагая, что $\sigma_{\xi_t}^2 \cong \sigma_{\xi_{t-1}}^2$, можно

записать, что:

$$d = \frac{\overline{\xi_t \xi_{t-1}}}{\sigma_{\xi_{t-1}}^2} = \frac{\sum \xi_t \xi_{t-1}}{\sum \xi_t^2}, \quad (3.21)$$

т.е. d – коэффициент автокорреляции остатков первого порядка. Обозначим его через ρ . Тогда регрессия остатков примет вид:

$$\xi_t = \rho \xi_{t-1} + V_t, \quad (3.22)$$

где ρ – коэффициент автокорреляции остатков первого порядка; V_t – случайная ошибка, удовлетворяющая всем предпосылкам МНК.

Уравнение (3.22) используется при тестировании моделей на автокорреляцию остатков.

Предполагая, что ρ известен, вычтем из уравнения (3.16) уравнение (3.17), умноженное на ρ :

$$y_t - \rho y_{t-1} = a(1 - \rho) + b(x_t - \rho x_{t-1}) + (\xi_t - \rho \xi_{t-1}). \quad (3.23)$$

В итоге имеем линейное уравнение с новыми переменными $(y_t - \rho y_{t-1})$, $(x_t - \rho x_{t-1})$. Обозначим новую зависимую переменную $y_t - \rho y_{t-1}$ через y_t^* , а объясняющую переменную $x_t - \rho x_{t-1}$ через x_t^* . Примем также, что $a(1 - \rho) = a^*$. Учитывая, что $\xi_t - \rho \xi_{t-1} = V_t$, получим уравнение:

$$y_t^* = a^* + b \cdot x_t^* + V_t, \quad (3.24)$$

где V_t – независимые случайные величины, имеющие нормальное распределение.

Так как ошибки V_t удовлетворяют предпосылкам МНК (они не содержат автокорреляцию), то оценки a^* и b будут обладать свойствами несмещенных оценок и могут быть получены обычным МНК.

Уравнение (3.24) возможно только при $t \geq 1$, так как при $t = 1$ отсутствует лаговая переменная. Чтобы не уменьшать число степеней свободы рекоменду-

ется для первого периода времени ($t=1$) использовать *поправку Прайса – Уинстена*

$$x_1^* = \sqrt{1-\rho^2} \cdot x_1; y_1^* = \sqrt{1-\rho^2} \cdot y_1. \quad (3.25)$$

Таким образом, ОМНК предполагает, что вместо исходных переменных y_t и x_t используются взвешенные переменные $PY = y^*$ и $PX = x^*$, где P – веса. В матричном виде модель регрессии примет вид $PY = PXB + P\xi$. В ней матрица весов P составит:

$$P = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 & 0 \\ 0 & 0 & -\rho & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\rho & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -\rho & 1 \end{bmatrix}.$$

Иными словами, матрица исходных данных трансформируется

$$y^* = \begin{bmatrix} \sqrt{1-\rho^2} \cdot y_1 \\ y_2 - \rho \cdot y_1 \\ y_3 - \rho \cdot y_2 \\ \dots \\ y_n - \rho \cdot y_{n-1} \end{bmatrix}; x^* = \begin{bmatrix} x_1 \sqrt{1-\rho^2} \\ x_2 - \rho \cdot x_1 \\ x_3 - \rho \cdot x_2 \\ \dots \\ x_n - \rho \cdot x_{n-1} \end{bmatrix}.$$

Для длинных динамических рядов поправки Прайса – Уинстена может не применяться. Тогда матрица весов не содержит первую строку рассмотренной матрицы P , и в расчетах используется $(n-1)$ преобразованных наблюдений y_t^* и x_t^* .

К преобразованным переменным y_t^* и x_t^* применяется традиционный МНК и оцениваются параметры a^* и b . Далее из соотношения $a^* = a(1-\rho)$ можно найти параметр a как

$$a = \frac{a^*}{(1-\rho)}. \quad (3.26)$$

ОМНК распространяется и на случай множественной регрессии

$$y_t = a + b_1 \cdot x_{1t} + b_2 \cdot x_{2t} + \dots + b_p \cdot x_{pt} + \xi_t.$$

Если имеет место автокорреляция остатков и $\xi_t = \rho \xi_{t-1} + V_t$, то

$$y_t - \rho y_{t-1} = a(1-\rho) + b_1(x_{1t} - \rho x_{1t-1}) + b_2(x_{2t} - \rho x_{2t-1}) + \dots + b_p(x_{pt} - \rho x_{pt-1}) + (\xi_t - \rho \xi_{t-1}). \quad (3.27)$$

Другой формат этой модели:

$$y_t^* = a^* + b_1 x_{1t}^* + b_2 x_{2t}^* + \dots + b_p x_{pt}^* + V_t. \quad (3.28)$$

Применяя к переменным $y_t^*, x_{1t}^*, x_{2t}^*, \dots, x_{pt}^*$ традиционный МНК, найдем оценки параметров b_1, b_2, \dots, b_p . Свободный член модели определим как $a = a^*(1 - \rho)$. Далее можно написать искомую модель регрессии, в которой устранена автокорреляция остатков.

Иными словами, применение ОМНК к регрессии с автокоррелированными остатками сводится к двухшаговой процедуре:

- преобразование исходных уровней динамических рядов с помощью известного значения коэффициента автокорреляции остатков первого порядка;
- применение к преобразованным данным обычного МНК.

Пример 3.6. По данным за 12 кварталов по предприятию рассматривается зависимость уровня рентабельности (y_t – в %) от оборачиваемости активов (x_t – в разах) (табл. 36).

Таблица 36

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	3,5	4	4,1	4,2	4,4	4,6	5	6	7,1

Провести анализ данного ряда.

Решение. Непосредственное применение регрессионного анализа к исходным данным привело к следующим результатам:

$$\hat{y}_t = 5,2078 + 2,9506 \cdot x_t, \quad (4,3) \quad (10,8),$$

$$R^2 = 0,9216; \quad F = 117,6.$$

Уравнение регрессии статистически значимо, как и его параметры, это видно по статистикам критериев Фишера и Стьюдента. Исследуем автокорреляцию остатков.

Значения остатков ξ_t и ξ_{t-1} приведены в табл. 37.

Таблица 37

Исходные данные и результаты анализа

t	y_t	x_t	ξ_t	ξ_{t-1}	y_t^*	x_t^*	\hat{y}_t
1	9	2	-2,109	–	7,725	1,717	13,5
2	11,5	2,6	-1,379	-2,109	6,883	1,574	14,8
3	14	3	-0,060	-1,379	8,101	1,666	15,7
4	16,8	3,5	1,265	-0,060	9,618	1,961	16,8
5	18	4	0,990	1,265	9,382	2,205	17,8
6	18,3	4,1	0,995	0,990	9,066	2,048	18,1
7	18,5	4,2	0,900	0,995	9,112	2,097	18,3
8	19	4,4	0,809	0,900	9,510	2,245	18,7
9	19,2	4,6	0,419	0,809	9,453	2,343	19,2
10	20,8	5	0,839	0,419	10,95	2,640	20,0
11	22	6	-0,912	0,839	11,33	3,435	22,2
12	24,4	7,1	-1,757	-0,912	13,11	4,022	24,6

Коэффициент автокорреляции остатков найдем по формуле

$$r_{a_1} = \frac{\sum e_t \cdot e_{t-1}}{\sum e_t^2} = \frac{8,3037}{16,1867} = 0,513.$$

Его величина не очень мала, по критерию Дарбина – Уотсона нулевая гипотеза об отсутствии автокорреляции остатков не может быть отвергнута (величина $DW = 0,508$ ниже нижней границы $0,97$ при уровне значимости $0,05$). Для устранения автокорреляции в остатках применим ОМНК.

Определим преобразованные значения зависимой и объясняющей переменных, взвесив их на коэффициент автокорреляции остатков. В качестве весов ρ будем использовать полученное значение $r_{a_1} = 0,513$.

Для $t = 1$ преобразованные значения составят

$$y_1^* = y_1 \sqrt{1 - \rho^2} = 9 \sqrt{1 - (0,513)^2} = 7,725;$$

$$x_1^* = x_1 \sqrt{1 - \rho^2} = 2 \sqrt{1 - (0,513)^2} = 1,717.$$

Остальные переменные со значениями $t = 2, 3, \dots, 12$ преобразуются по формулам $y_t^* = y_t - \rho y_{t-1}$; $x_t^* = x_t - \rho x_{t-1}$. Результаты расчета новых переменных приведены в табл. 37. К преобразованным переменным y_t^* и x_t^* применяем обычный МНК. Уравнение регрессии будет иметь следующий вид:

$$y_t^* = 4,4474 + 2,1778 x_t^*; R^2 = 0,9016; F = 91,66.$$

Хотя в модели с преобразованными переменными коэффициент детерминации ниже, чем в первоначальной модели, зато автокорреляция в остатках отсутствует (коэффициент автокорреляции остатков равен $0,047$). В данном уравнении величина $4,4474$ – параметр a_t^* . Перейдем от него к параметру a :

$$a = \frac{a^*}{(1 - \rho)} = 9,1322.$$

В окончательном виде уравнение регрессии составит

$$\hat{y}_t = 9,1322 + 2,1778 \cdot x_t$$

(рассчитанные по нему теоретические значения представлены в последней графе табл. 37).

В данном уравнении коэффициент регрессии свободен от автокоррелированности ошибок, поэтому он не является результатом ложной регрессии и позволяет делать более корректные выводы относительно связи результативного признака с объясняющей переменной.

Рассмотренный ОМНК базируется на предположении, что коэффициент автокорреляции остатков ρ известен. Однако на практике точное значение ρ неизвестно и используется его оценка $\hat{\rho}$.

В качестве оценки $\hat{\rho}$ может использоваться значение $r_{e_t e_{t-1}}$, полученное на основе критерия Дарбина – Уотсона для модели по исходным данным. Ранее было показано, что $DW \cong 2(1 - r_{e_t e_{t-1}})$. Соответственно имеем уравнение:

$$r_{e_t e_{t-1}} \cong 1 - \frac{DW}{2}. \quad (3.29)$$

В нашем примере $DW = 0,508$ до коррекции выводов величина. Соответственно $\hat{\rho}$ составит 0,746.

Данный метод оценивания дает неплохие результаты при достаточно большом числе наблюдений. Используя $\hat{\rho} = 0,746$ и процедуру ОМНК к табл. 37, получим уравнение с некоррелированными остатками

$$\hat{y}_t = 3,314 + 1,67 \cdot x_t; R^2 = 0,735,$$

но коэффициент автокорреляции остатков несколько выше ($-0,212$), хотя нулевая гипотеза об отсутствии автокорреляции остатков принимается. Практически эта модель по коэффициенту регрессии отличается от предыдущей и исходной моделей.

3.3. Учет сезонности в регрессионных моделях

При отсутствии тенденции в рядах динамики и при наличии сезонной компоненты в модели регрессии зависимая переменная y_t может рассматриваться как функция объясняющих переменных $x_{1t}, x_{2t}, \dots, x_{pt}$ и сезонного фактора z , который можно ввести в модель в виде фиктивных переменных 0 и 1, например, для квартальных данных в модель вводятся фиктивные переменные z_1, z_2 и z_3 , учитывающие информацию I, II и III кварталов:

$z_1 = 1$ для I квартала;

$z_2 = 1$ для II квартала;

$z_3 = 1$ для III квартала;

в остальных случаях $z_1 = z_2 = z_3 = 0$.

В общем виде линейная модель регрессии имеет вид

$$y_t = a + \sum_{i=1}^p b_i \cdot x_{it} + \sum_{j=1}^k c_j \cdot z_j, \quad (3.30)$$

где $k = 1, 2, 3$ (для квартальных данных); $k = 1, 2, \dots, 11$ (для месячных данных). Построение подобного рода моделей требует достаточно длинных рядов динамики, чтобы параметры модели были статистически значимы и надежны.

Коэффициенты регрессии b_i , показывают воздействие соответствующего фактора x_i на y_t при устранении сезонности. Влияние сезонности оценивается через коэффициенты c_j , которые представляют собой разность среднего уровня y_t для j -го квартала по сравнению с базовым, т.е. если за базу сравнения взят IV квартал, то $c_1 = \bar{y}_I - \bar{y}_{IV}$ при условии, что факторы $x_{1t}, x_{2t}, \dots, x_{pt}$ неизменны. Соответственно $c_2 = \bar{y}_{II} - \bar{y}_{IV}$, $c_3 = \bar{y}_{III} - \bar{y}_{IV}$. Чтобы оценить влияние сезонности I квартала по сравнению со II, можно использовать разность $(c_2 - c_1)$, а влияние III квартала по сравнению со II – $(c_3 - c_2)$ и т.п.

Практическое использование рассматриваемой модели затруднено следующими обстоятельствами:

- ограничено число объясняющих переменных ввиду сравнительно коротких динамических рядов в экономических исследованиях. Так, по данным за 5–7 лет в поквартальном разрезе в модели может быть учтена только одна объясняющая переменная;
- сезонность по y_t и по x_t может не совпадать по кварталам (например, пик доходов может приходиться на зиму, а пик расходов – на лето);
- коэффициент регрессии при факторе может подвергаться сезонности, а в модели он предполагается независимым от сезонной компоненты;
- ряды динамики обнаруживают тенденцию.

Рассмотрим содержание линейной модели на следующем примере:

$$y_t = a + bx_t + c_1z_1 + c_2z_2 + c_3z_3 + \xi_t.$$

Пример 3.7. Объем продаж товара фирмой (y_t – в тыс. ед.) исследуется в зависимости от объема продаж его дочерним предприятием (x_t – в тыс. ед.) по данным за 5 лет. В табл. 38 приводится соответствующая информация по кварталам.

Таблица 38

Годы	Кварталы								Итого	
	I		II		III		IV			
	y_t	x_t	y_t	x_t	y_t	x_t	y_t	x_t	y_t	x_t
2014	11	9	15	10	6	8	12	9	44	36
2015	11	10	16	9	4	3	13	11	44	33
2016	10	7	14	10	7	8	12	11	43	36
2017	10	12	16	9	8	11	13	12	47	44
2018	11	8	18	16	7	6	12	12	48	42
Итого	53	46	79	54	32	36	62	55	226	191
В сред- нем	10,6	9,2	15,8	10,8	6,4	7,2	12,4	11	45,2	38,2

Решение. По данной таблице видно, что объем продаж практически стабилен по годам, но сильно варьируется по кварталам. Если строить модель, не обращая внимания на сезонность, то результаты моделирования окажутся плохими:

$$\hat{y}_t = 2,674 + 0,903 \cdot x_t + \xi_t,$$

(1,1) (3,8)

$$R^2 = 0,4466.$$

Они существенно улучшаются, если в модели будет учтена сезонность. Для этого введем в модель сезонную компоненту с помощью фиктивных переменных. Взяв за базу сравнения IV квартал, построим модель зависимости объема продаж фирмой от объема продаж дочерним предприятием с учетом сезонности:

$$\hat{y}_t = a + bx_t + c_1z_1 + c_2z_2 + c_3z_3,$$

где $z_1 = 1$ для I квартала и 0 для остальных; $z_2 = 1$ для II квартала и 0 для остальных; $z_3 = 1$ для III квартала и 0 для остальных.

Результаты регрессии оказались следующими:

$$\hat{y}_t = 8,896 + 0,319x_t - 1,227z_1 + 3,464z_2 - 4,789z_3,$$

(8,3) (3,5) (-2,15) (6,3) (-7,4)

$$R^2 = 0,955; DW = 1,998.$$

Как видим, существенно повысился коэффициент детерминации. Все параметры модели статистически значимы и отсутствует автокорреляция остатков. Параметр при переменной x_t означает, что независимо от сезонного фактора с ростом объема продаж дочерним предприятием на 1 тыс. ед. объем продаж головной организации возрастает в среднем на 0,319 тыс. ед.

Для отдельных кварталов модель регрессии составит:

для I квартала $\hat{y}_t = 7,669 + 0,319 \cdot x_t$;
 для II квартала $\hat{y}_t = 12,36 + 0,319 \cdot x_t$;
 для III квартала $\hat{y}_t = 4,107 + 0,319 \cdot x_t$;
 для IV квартала $\hat{y}_t = 8,896 + 0,319 \cdot x_t$.

Коэффициенты регрессии при z показывают влияние сезонности: изменение объема продаж в головной организации в соответствующем квартале по сравнению с IV кварталом, когда $z_1 = z_2 = z_3 = 0$, в условиях неизменности объема продаж в дочернем предприятии. Все коэффициенты регрессии при сезонном факторе z согласовываются в примере с изменением средних уровней результата по отдельным кварталам в сравнении с IV кварталом:

По модели	По итоговой строке таблицы
$c_1 = -1,227$	$\bar{y}_I - \bar{y}_{IV} = 10,6 - 12,4 = -1,8$
$c_2 = 3,464$	$\bar{y}_{II} - \bar{y}_{IV} = 15,8 - 12,4 = 3,4$
$c_3 = -4,789$	$\bar{y}_{III} - \bar{y}_{IV} = 6,4 - 12,4 = -6,0$

Наибольшие изменения в средних уровнях наблюдаются в III и II кварталах (по сравнению с IV кварталом). Автокорреляция в остатках по модели практически отсутствует: $r_{a_1} = 0,165$.

Если в рядах динамики наблюдается тенденция и сезонная компонента одновременно, то модель регрессии может иметь вид (при двух временных рядах):

$$\hat{y}_t = a + bx_t + ct + d_1z_1 + d_2z_2 + d_3z_3. \quad (3.31)$$

Следует заметить, что при наличии четкой тенденции как в ряду зависимой, так и в рядах объясняющих переменных модель с сезонной составляющей в виде линейно введенных фиктивных переменных не всегда может давать хорошие результаты ввиду мультиколлинеарности факторов, особенно, если корреляция объясняющих переменных с фактором времени выше их корреляции с результативным признаком. В этом случае коэффициенты регрессии часто статистически не значимы или имеют не интерпретируемые знаки.

Предположим, что строится модель регрессии объема продаж фирмой туристических путевок в Австрию от цены на них за последние 3 года (данные характеризуются тенденцией и сезонностью; пользуются спросом горнолыж-

ные маршруты зимой и термальные озера летом). Располагая информацией по месяцам, вводим в модель фактор времени t (1, 2, 3, ..., 36) и сезонную компоненту в виде фиктивных переменных z_1 и z_2 , принимающих значение 1 соответственно для зимних и летних периодов и 0 – для остальных.

Модель имеет вид $y_t = a + bx_t + ct + d_1z_1 + d_2z_2 + \xi_t$. Параметр b показывает, что независимо от тенденции и сезонности рост цен на 1 ден. ед. соответствует изменению числа проданных путевок на b ед. Ежемесячно под воздействием тенденции и независимо от периода времени объем продаж возрастает в среднем на c ед. При абстрагировании от влияния тенденции и фактора цен сезонность проявляет себя в значениях параметров c_1 и c_2 : в зимний период объем продаж возрастает на c_1 , а летом – на c_2 ед. (все коэффициенты регрессии при сезонной компоненте больше нуля и статистически значимы по t -критерию Стьюдента).

Сезонность можно рассматривать и в более крупном плане: например, осенне-зимний период и весенне-летний период. В этом случае фиксированная переменная может принимать значение 1, например, для I и IV кварталов, а для II и III кварталов – значение 0 (или наоборот). Такой подход уменьшает число вводимых в модель фиктивных переменных, что может быть полезным при сравнительно коротких рядах динамики. Этот подход возможен также, если в каком-то из периодов влияние сезонности оказалось по t -критерию Стьюдента статистически незначимым.

Контрольные вопросы

1. Что такое ложная корреляция?
2. Почему регрессия по временным рядам требует специальных методов построения модели?
3. В чем состоит смысл метода первых разностей?
4. Опишите алгоритм построения модели регрессии по отклонениям от тренда.
5. Как учитывается тенденция при встроенности в нее фактора времени?
6. В каком случае модель по отклонениям от тренда совпадает с моделью с включением в нее фактора времени?
7. В чем смысл использования обобщенного метода наименьших квадратов при построении модели регрессии по временным рядам?
8. Какова цель поправки Прайса – Уинстена?
9. Что собой представляет метод Кохрейна – Оркатта?
10. Опишите алгоритм ОМНК для построения модели регрессии по временным рядам.

Контрольные задания

1. Администрация банка изучает динамику депозитов физических лиц за ряд лет (млн дол. в сопоставимых ценах). Исходные данные представлены ниже:

								Сумма
Время, лет, t	1	2	3	4	5	6	7	28
Депозиты физ. лиц, x	2	6	7	3	10	12	13	53
Известно, что: $\sum x^2 = 511$								

Задание.

1) Построить уравнение линейного тренда и дать интерпретацию его параметров.

2) Определить коэффициент детерминации для линейного тренда.

3) Администрация банка предполагает, что среднегодовой абсолютный прирост депозитов физических лиц составляет не менее 2,5 млн. долл. Подтверждается ли это предположение результатами, которые получили?

2. Изучается динамика потребления мяса в регионе. Для этого были собраны данные об объемах среднедушевого потребления мяса y_t (кг) за 7 месяцев. Предварительная обработка данных путем логарифмирования привела к получению следующих результатов:

Месяц	1	2	3	4	5	6	7
$\ln y_t$	2,1	2,11	2,13	2,17	2,17	2,22	2,31

Задание.

1) Построить уравнение экспоненциального тренда и дать интерпретацию его параметров.

2) Определить коэффициент детерминации для экспоненциального тренда.

3. Имеются данные об урожайности зерновых в хозяйствах области:

Год	Урожайность зерновых, ц/га
1	10,2
2	10,7
3	11,7
4	13,1
5	14,9
6	17,2
7	20,0
8	23,2

Задание.

1) Обосновать выбор типа уравнения тренда.

2) Рассчитать параметры уравнения тренда.

3) Дать прогноз урожайности зерновых на следующий год.

4. Имеются следующие данные об уровне безработицы y_t (%) за 8 месяцев:

Месяц	1	2	3	4	5	6	7	8
y_t	8,8	8,6	8,4	8,1	7,9	7,6	7,4	7,0

Задание.

- 1) Определить коэффициенты автокорреляции уровней этого ряда первого и второго порядка.
- 2) Обосновать выбор уравнения тренда и определить его параметры.
- 3) Интерпретировать полученные результаты.

5. Пусть имеется следующий временной ряд:

t	1	2	3	4	5	6	7	8
y_t	20	10

Известно, что $\sum x_t = 150$, $\sum x_t^2 = 8100$, $\sum x_t x_{t-1} = 7350$.

Задание.

- 1) Определить коэффициент автокорреляции уровней этого ряда первого порядка.
- 2) Установить, включает ли исследуемый временной ряд тенденцию.

6. Экспорт, импорт, внешнеторговый оборот Австрии и Бельгии за 1991–2010 гг. характеризуются данными в таблице:

Год	Австрия, млн евро			Бельгия, млн евро		
	экспорт	импорт	внешне- торговый оборот	экспорт	импорт	внешне- торговый оборот
1991	44	43	87	20	21	41
1992	47	46	93	21	22	43
1993	51	51	102	24	24	48
1994	56	56	112	27	28	55
1995	62	63	125	31	30	61
1996	67	71	138	32	33	65
1997	72	74	146	35	35	70
1998	79	80	159	40	40	80
1999	95	91	186	48	47	95
2000	117	131	248	56	53	109
2001	129	126	255	61	58	119
2002	146	144	290	68	63	131
2003	166	164	330	84	81	165
2004	204	206	410	111	110	221
2005	209	205	414	106	106	212
2006	236	247	483	126	125	251
2007	257	278	535	147	149	296
2008	281	280	561	154	157	311
2009	328	332	660	179	186	365
2010	366	386	752	202	212	414

Задание.

- 1) По каждому ряду построить график динамики.
- 2) Провести расчет параметров трендов разной формы.

3) Оценить качество каждого тренда через среднюю ошибку аппроксимации, линейный коэффициент автокорреляции отклонений.

4) Оценить статистическую значимость трендов через F -критерий, значимость параметров тренда – через t -критерий.

5) Выбрать лучшую форму тренда и выполнить по ней точечный прогноз на 2015 г.

6) Оценить ошибку прогноза и построить доверительный интервал прогноза для уровня значимости 0,05.

7. Имеются поквартальные данные по розничному товарообороту России за определенный период.

Номер квартала	Товарооборот, % к предыдущему периоду	Номер квартала	Товарооборот, % к предыдущему периоду
1	100	11	98,8
2	93,9	12	101,9
3	96,5	13	113,1
4	101,8	14	98,4
5	107,8	15	97,3
6	96,3	16	102,1
7	95,7	17	97,6
8	98,2	18	83,7
9	104	19	84,3
10	99	20	88,4

Задание.

1) Построить график временного ряда.

2) Рассчитать трендовую и сезонную компоненты.

3) Построить аддитивную и мультипликативную модели временного ряда.

4) Построить автокорреляционную функцию временного ряда.

5) Оценить качество каждой модели через показатели средней абсолютной ошибки и среднего относительного отклонения. Выбрать лучшую модель.

8. Имеются данные за десять лет о производительности труда и электровооруженности труда на одном из предприятий промышленности области.

Показатель	1	2	3	4	5	6	7	8	9	10
Среднегодовая выработка продукции на 1 рабочего, y	28,7	31,7	31,7	32,6	33,9	31,2	33,3	42,6	46,0	49,9
Электровооруженность, x	3,33	3,39	3,5	3,63	3,81	3,84	3,88	4,07	4,12	4,17

Результаты аналитического выравнивания привели к следующим уравнениям трендов для каждого из рядов:

а) для временного ряда производительности труда:

$$\hat{y}_t = 33,19 + 1,04 \cdot t + 0,09 \cdot t^2, \quad (t = -9, -7, -5, -3, -1, 1, 3, 5, 7, 9);$$

б) для временного ряда электровооруженности:

$$\hat{x}_t = 3,774 + 0,049 \cdot t, \quad (t = -9, -7, -5, -3, -1, 1, 3, 5, 7, 9).$$

Задание.

1) Определить коэффициент корреляции между временными рядами, используя непосредственно исходные уровни, первые разности для электровооруженности и вторые разности для производительности труда, отклонения от основной тенденции.

2) Объяснить различия полученных результатов.

3) Рассчитать коэффициент автокорреляции внутри каждого временного ряда.

9. Экспорт, импорт, внешнеторговый оборот Ирландии и Греции за 1991–2010 гг. характеризуются данными в таблице:

Год	Ирландия, у.е.			Греция, у.е.		
	экспорт	импорт	внешне- торговый оборот	экспорт	импорт	внешне- торговый оборот
1991	236	247	483	1 266	1 261	2 527
1992	257	278	535	1 474	1 499	2 973
1993	281	280	561	1 540	1 570	3 110
1994	328	332	660	1 798	1 866	3 664
1995	366	386	752	2 026	2 125	4 151
1996	405	419	824	2 286	2 357	4 643
1997	431	412	843	2 640	2 694	5 334
1998	450	434	884	2 924	2 864	5 788
1999	498	496	994	3 337	3 277	6 614
2000	549	547	1 096	3 479	3 379	6 858
2001	523	510	1 033	3 367	3 187	6 554
2002	527	520	1 047	3 477	3 334	6 811
2003	590	584	1 174	3 900	3 719	7 619
2004	669	661	1 330	4 498	4 320	8 818
2005	737	720	1 457	4 660	4 506	9 166
2006	775	758	1 533	4 846	4 658	9 504
2007	792	772	1 564	4 980	4 713	9 693
2008	787	773	1 560	5 012	4 674	9 686
2009	835	842	1 677	5 491	5 108	10 599
2010	887	911	1 798	5 764	5 377	11 141

Задание.

1) По каждому ряду построить график динамики.

2) Провести расчет параметров трендов разной формы.

3) Оценить качество каждого тренда через среднюю ошибку аппроксимации, линейный коэффициент автокорреляции отклонений.

4) Оценить статистическую значимость трендов через F -критерий, значимость параметров тренда – через t -критерий.

5) Выбрать лучшую форму тренда и выполнить по ней точечный прогноз на 2015 г.

6) Оценить ошибку прогноза и построить доверительный интервал прогноза для уровня значимости 0,05.

10. Имеются данные об объеме экспорта из РФ (у.е., цены Фондовой Общероссийской биржи) за 24 квартала.

Номер квартала	Экспорт, у.е.	Номер квартала	Экспорт, у.е.
1	4 087	13	6 975
2	4 737	14	6 891
3	5 768	15	7 527
4	6 005	16	7 971
5	5 639	17	5 875
6	6 745	18	6 140
7	6 311	19	6 248
8	7 107	20	6 041
9	5 741	21	4 626
10	7 087	22	6 501
11	7 310	23	6 284
12	8 600	24	6 707

Задание.

- 1) Построить график временного ряда.
- 2) Рассчитать трендовую и сезонную компоненты.
- 3) Построить аддитивную и мультипликативную модели временного ряда.
- 4) Построить автокорреляционную функцию временного ряда.
- 5) Оценить качество каждой модели через показатели средней абсолютной ошибки и среднего относительного отклонения. Выбрать лучшую модель.

11. Администрация торговой фирмы интересуется, есть ли взаимосвязь между объемом продаж и удельным весом женщин среди работников компании. Для этого были собраны данные за последние десять лет:

Год	1	2	3	4	5	6	7	8	9	10
Объем продаж, y	378	385	393	403	414	428	444	462	481	495
Удельный вес женщин, x	25	24	27	30	31	29	31	33	34	28

Известно также: $\sum y_t = 3788$, $\sum y_t^2 = 1604488$, $\sum x_t = 264$, $\sum x_t^2 = 78388$, $\sum x_t y_t = 112001$.

Уравнения трендов для каждого из рядов составили:

а) для ряда x_t : $\hat{x}_t = 23,5 + 1,17 \cdot t$;

б) для ряда y_t : $\hat{y}_t = 374,14 + 3,33 \cdot t + 0,95 \cdot t^2$.

Задание.

1) Определить коэффициент корреляции между изучаемыми рядами по их уровням.

2) Определить коэффициент корреляции между изучаемыми рядами по отклонениям от указанных выше линейного и параболического трендов соответственно.

3) Выбрав одну из полученных мер, охарактеризовать тесноту связи между временными рядами объемов продаж и долей женщин среди работников компании. Обоснуйте ваш выбор.

12. Имеются данные об уровне безработицы y_t (%) за 8 месяцев:

Месяц	1	2	3	4	5	6	7	8
Уровень y_t	3,5	3,2	3,0	2,8	2,5	2,4	2,4	2,2

Задание.

- 1) Определить коэффициенты автокорреляции уровней этого ряда первого и второго порядка.
- 2) Обосновать выбор уравнения тренда и определить его параметры.
- 3) Интерпретировать полученные результаты.

13. Имеются данные о разрешениях на строительство нового частного жилья в процентном соотношении к уровню 2000 г. в определенном регионе:

Месяц	2001	2002	2003	2004	2005
Январь	72,9	61,4	71,2	78,3	86,4
Февраль	113,4	51,0	69,9	76,4	87,5
Март	86,2	55,3	74,3	74,5	80,2
Апрель	80,8	59,1	70,2	68,5	84,3
Май	73,7	59,5	68,4	71,6	86,8
Июнь	69,2	64,3	68,5	72,1	86,9
Июль	71,9	62,5	68,6	73,3	85,2
Август	69,9	63,1	70,6	76,2	85,0
Сентябрь	69,4	61,2	69,7	79,8	87,5
Октябрь	63,3	63,2	72,3	81,2	90,0
Ноябрь	60,0	64,3	73,5	83,5	88,4
Декабрь	61,0	63,9	72,5	88,0	85,7

Задание.

- 1) Рассчитать трендовую и сезонную компоненты.
- 2) Провести расчет параметров трендов разной формы.
- 3) Построить автокорреляционную функцию временного ряда количества разрешений на строительство частного жилья. Охарактеризовать структуру этого ряда.
- 4) Оценить статистическую значимость трендов через F -критерий, значимость параметров тренда – через t -критерий.
- 5) Оценить ошибку прогноза и построить доверительный интервал прогноза для уровня значимости 0,05.

14. Динамика объема платных услуг населению региона по кварталам 2006–2009 гг. характеризуются следующими данными:

Квартал	Объем платных услуг населению	Квартал	Объем платных услуг населению
1	2 428	9	3 528
2	2 010	10	3 838
3	2 981	11	3 916
4	3 074	12	4 142
5	2 893	13	4 441
6	3 198	14	5 583
7	3 250	15	6 230
8	3 495	16	6 497

Задание.

- 1) Построить график временного ряда.
- 2) Рассчитать трендовую и сезонную компоненты.
- 3) Построить аддитивную и мультипликативную модели временного ряда.
- 4) Построить автокорреляционную функцию временного ряда.
- 5) Оценить качество каждой модели через показатели средней абсолютной ошибки и среднего относительного отклонения. Выбрать лучшую модель.

15. В таблице приводятся данные об уровне производительности труда по экономике США (x) и среднечасовой заработной плате в экономике США (y), в сопоставимых ценах за несколько лет:

Год	x	y	Год	x	y	Год	x	y
1	65,6	6,79	11	87,0	8,03	21	98,6	7,78
2	68,1	6,88	12	90,2	8,21	22	99,9	7,69
3	70,4	7,07	13	92,6	8,53	23	100,0	7,68
4	73,3	7,17	14	95,0	8,55	24	102,2	7,79
5	76,5	7,33	15	93,3	8,285	25	104,6	7,80
6	78,6	7,52	16	95,5	8,12	26	106,1	7,77
7	81,0	7,62	17	98,3	8,24	27	108,3	7,81
8	83,0	7,72	18	99,8	8,36	28	109,4	7,73
9	85,4	7,89	19	100,4	8,40	29	110,4	7,69
10	85,9	7,98	20	99,3	8,17	30	109,5	7,64

Задание.

- 1) Оценить обычным МНК параметры модели с распределенным лагом, характеризующей зависимость заработной платы от производительности труда, при величине лага 2, 3 и 4. Проанализировать полученные результаты.
- 2) Оценить параметры этой же модели при величине лага 3 и 4 в предположении полиномиальной структуры лага (в качестве функции, описывающей структуру лага, выбрать полином второго порядка). Проанализировать полученные результаты. Сравнить с результатами в п. 1 и сделать вывод.

16. Имеются данные о потреблении и личных доходах населения за определенный период.

Время, лет	1	2	3	4	5	6	7
Потребление	300	310	325	340	350	370	385
Личные доходы	335	340	360	378	400	417	430

Задание.

- 1) Построить уравнение линейного тренда и дать интерпретацию его параметров.
- 2) Определить коэффициент детерминации для линейного тренда.
- 3) Охарактеризовать тесноту связи между рядами по их уровням, по первым разностям. Сделать вывод.

17. Экспорт, импорт, внешнеторговый оборот Италии и Румынии за 1991–2010 гг. характеризуются данными в таблице:

Год	Италия, млн евро			Румыния, млн евро		
	экспорт	импорт	внешнеторговый оборот	экспорт	импорт	внешнеторговый оборот
1991	44	43	87	20	21	41
1992	47	46	93	21	22	43
1993	51	51	102	24	24	48
1994	56	56	112	27	28	55
1995	62	63	125	31	30	61
1996	67	71	138	32	33	65
1997	72	74	146	35	35	70
1998	79	80	159	40	40	80
1999	95	91	186	48	47	95
2000	117	131	248	56	53	109
2001	129	126	255	61	58	119
2002	146	144	290	68	63	131
2003	166	164	330	84	81	165
2004	204	206	410	111	110	221
2005	209	205	414	106	106	212
2006	236	247	483	126	125	251
2007	257	278	535	147	149	296
2008	281	280	561	154	157	311
2009	328	332	660	179	186	365
2010	366	386	752	202	212	414

Задание.

- 1) По каждому ряду построить график динамики.
- 2) Провести расчет параметров трендов разной формы.
- 3) Оценить качество каждого тренда через среднюю ошибку аппроксимации, линейный коэффициент автокорреляции отклонений.
- 4) Оценить статистическую значимость трендов через F -критерий, значимость параметров тренда – через t -критерий.
- 5) Выбрать лучшую форму тренда и выполнить по ней точечный прогноз на 2015 г.

б) Оценить ошибку прогноза и построить доверительный интервал прогноза для уровня значимости 0,05.

18. Приводятся сведения об уровне среднегодовых цен на какао-бобы из Бразилии.

Год	Цена	Год	Цена	Год	Цена	Год	Цена
1	29,4	8	183,5	15	105,3	22	47,5
2	23,5	9	153,5	16	94,9	23	45,0
3	26,2	10	140,7	17	92,0	24	44,5
4	48,5	11	107,1	18	83,9	25	55,9
5	73,4	12	87,5	19	72,7	26	60,5
6	56,6	13	68,3	20	56,9	27	64,1
7	77,0	14	83,1	21	49,1	28	71,0

Задание.

- 1) Построить график временного ряда.
- 2) Рассчитать трендовую и сезонную компоненты.
- 3) Построить аддитивную и мультипликативную модели временного ряда.
- 4) Построить автокорреляционную функцию временного ряда.
- 5) Оценить качество каждой модели через показатели средней абсолютной ошибки и среднего относительного отклонения. Выбрать лучшую модель.

19. Имеются данные о динамике оборота розничной торговли и потребительских цен региона за определенный период.

Месяц	Оборот розничной торговли	Индекс потребительских цен	Месяц	Оборот розничной торговли	Индекс потребительских цен
Январь	70,8	101,7	Январь	74,3	110,0
Февраль	98,7	101,1	Февраль	92,9	106,4
Март	97,9	100,4	Март	106,0	103,2
Апрель	99,6	100,1	Апрель	99,8	103,2
Май	96,1	100,0	Май	105,2	102,9
Июнь	103,4	100,1	Июнь	99,7	100,8
Июль	95,5	100,0	Июль	99,7	101,6
Август	102,9	105,8	Август	107,9	101,5
Сентябрь	77,6	145,0	Сентябрь	98,8	101,4
Октябрь	102,3	99,8	Октябрь	104,6	101,7
Ноябрь	102,9	102,7	Ноябрь	106,4	101,7
Декабрь	123,1	109,4	Декабрь	122,7	101,2

Задание.

- 1) Построить график временного ряда.
- 2) Рассчитать трендовую и сезонную компоненты.
- 3) Построить аддитивную и мультипликативную модели временного ряда.
- 4) Построить автокорреляционную функцию временного ряда.

5) Оценить качество каждой модели через показатели средней абсолютной ошибки и среднего относительного отклонения. Выбрать лучшую модель.

20. Для прогнозирования объема продаж компании на основе поквартальных данных за 2003–2007 гг. была построена аддитивная модель временного ряда объема продаж. Уравнение, моделирующее динамику трендовой компоненты этой модели, имеет вид: $T = 100 + 2t$. Показатели за 2006 г., полученные в ходе построения аддитивной модели, представлены в таблице:

Время года	Фактический объем продаж в 2006 г.	Компонента, полученная по аддитивной модели		
		трендовая	сезонная	случайная
Зима	100			+4
Весна			10	+5
Лето	150		25	
Осень				

Задание.

Определить недостающие в таблице данные, учитывая, что объем продаж компании за 2006 г. в целом составил 490 млн р.

21. На основе помесечных данных о потреблении электроэнергии в регионе (млн кВт·ч) за последние 3 года была построена аддитивная модель временного ряда. Скорректированные значения сезонной компоненты за соответствующие месяцы приводятся ниже:

Январь	+25	Май	–32	Сентябрь	+2
Февраль	+10	Июнь	–38	Октябрь	+15
Март	+6	Июль	–25	Ноябрь	+27
Апрель	–4	Август	–18	Декабрь	?

Уравнение тренда имеет вид: $T = 300 + 1,5t$, $t = 1...36$.

Задание.

- 1) Определить значение сезонной компоненты за декабрь.
- 2) На основе построенной модели дать точечный прогноз ожидаемого потребления электроэнергии в течение первого квартала следующего года.

22. На основе поквартальных данных об уровне безработицы в летнем курортном городе (% от экономически активного населения) за последние 5 лет была построена мультипликативная модель временного ряда. Скорректированные значения сезонной компоненты за каждый квартал приводятся ниже:

I квартал	1,4	III квартал	0,7
II квартал	0,8	IV квартал	–

Уравнение тренда выглядит следующим образом: $T = 9,2 - 0,3t$ (при расчете параметров тренда для нумерации кварталов использовались натуральные числа $t = 1 + 20$).

Задание.

- 1) Определить значение сезонной компоненты за IV квартал.
- 2) На основе построенной модели дать точечный прогноз уровня безработицы на I и II квартал следующего года.

23. В целях прогнозирования объема экспорта страны на будущие периоды были собраны данные за 30 лет по следующим показателям: x_t – объем экспорта (млрд дол., в сопоставимых ценах); y_t – индекс физического объема промышленного производства (% к предыдущему году). Ниже представлены результаты предварительной обработки исходных данных.

1) Уравнения линейных трендов:

а) для ряда y_t : $y_t = 3,1 + 1,35t + \varepsilon_t$, $R^2 = 0,91$, $d = 2,31$;

б) для ряда x_t : $x_t = -8,4 + 4,8t + \varepsilon_t$, $R^2 = 0,89$, $d = 2,08$.

2) Уравнение регрессии по уровням временных рядов:

$$y_t = -10,5 + 0,5x_t + \varepsilon_t, R^2 = 0,95, d = 2,21.$$

3) Уравнение регрессии по первым разностям уровней временных рядов:

$$\Delta y_t = 1,4 + 0,03\Delta x_t + \varepsilon_t, R^2 = 0,86, d = 2,25.$$

4) Уравнение регрессии по вторым разностям уровней временных рядов:

$$\Delta^2 y_t = 0,7 + 0,012\Delta^2 x_t + \varepsilon_t, R^2 = 0,47, d = 2,69.$$

5) Уравнение регрессии по уровням временных рядов с включением фактора времени:

$$y_t = 4,23 + 0,24x_t + 0,78t + \varepsilon_t, R^2 = 0,97, d = 0,9.$$

Задание.

1) Сформулировать свои предположения относительно величины коэффициента автокорреляции первого порядка в каждом из рядов. Ответ обосновать.

2) Выбрать наилучшее уравнение регрессии, которое можно использовать для прогнозирования объема экспорта, дать интерпретацию параметров.

24. Изучается зависимость объема продаж бензина (y_t) от динамики потребительских цен (x_t). Полученные за последние 6 кварталов данные представлены в таблице:

Показатель	1 кв.	2 кв.	3 кв.	4 кв.	5 кв.	6 кв.
Индекс потребительских цен, % к кварталу 1	100	104	112	117	121	126
Средний за день объем продаж бензина в течение квартала, тыс. л	89	83	80	77	75	72

Известно также: $\sum x_t = 680$, $\sum y_t = 476$, $\sum x_t y_t = 53648$, $\sum x_t^2 = 77566$.

Задание.

- 1) Построить модель зависимости объема продаж бензина от индекса потребительских цен с включением фактора времени.
- 2) Дать интерпретацию параметров полученной модели.

25. Годовое потребление товара и доходы населения (тыс. р.) за 2009–2017 гг. приведены в таблице:

Показатель	2009	2010	2011	2012	2013	2014	2015	2016	2017
Потребление	46	50	54	59	62	67	75	86	100
Доходы	53	57	64	70	73	82	95	110	127

Задание.

- 1) Определить уравнение регрессии, включив в него фактор времени, если известно, что $\sum x_t = 731$, $\sum y_t = 599$, $\sum x_t y_t = 52179$, $\sum x_t^2 = 64361$, $\sum y_t^2 = 42367$.
- 2) Дать интерпретацию параметров полученной модели.

26. В таблице приводятся данные о потреблении и личных доходах населения за 2005–2011 гг.:

Показатель	2005	2006	2007	2008	2009	2010	2011
Потребление	300	310	325	340	350	370	385
Доходы	335	340	360	378	400	417	430

Задание.

- 1) Построить уравнение линейной регрессии, используя метод первых разностей.
- 2) Охарактеризуйте тесноту связи между рядами по их уровням, по первым разностям. Сделать выводы.

27. По данным за 30 лет изучается зависимость рентабельность продукции компании y_t (%) от численности занятых ручным трудом x_t (чел.). Были получены следующие варианты уравнений регрессии:

- а. по уровням временных рядов:

$$\hat{y}_t = 2 - 0,5x_t + \varepsilon_t, R^2 = 0,9025, d = 0,8;$$

- б. по первым разностям уровней:

$$\Delta \hat{y}_t = 3 + 0,1\Delta x_t + \varepsilon_t, R^2 = 0,49, d = 1,2;$$

- с. по вторым разностям уровней:

$$\Delta^2 \hat{y}_t = 15 - 0,062\Delta^2 x_t + \varepsilon_t, R^2 = 0,7225, d = 2,1;$$

d. по уровням рядов с включением фактора времени:

$$\hat{y}_t = -7 - 0,02x_t + 0,3t + \varepsilon_t, \quad R^2 = 0,95, \quad d = 2,2.$$

(−3,1) (3,7)

В таблице приведены известные коэффициенты автокорреляции первого порядка:

Ряд	По уровням ряда	По первым разностям уровней ряда	По вторым разностям уровней ряда
x_t	0,99	0,8	0,05
y_t	0,86	0,86	0,1

Задание.

1) Определить коэффициенты корреляции по уровням временных рядов, по первым разностям временных рядов и по вторым разностям временных рядов. Охарактеризовать тесноту связи между временными рядами рентабельности продукции и численности занятых ручным трудом. Обосновать выбор одной из мер тесноты связи.

2) Исследовать полученные уравнения регрессии на автокорреляцию в остатках.

3) Выбрать наилучшее уравнение регрессии и дать интерпретацию параметрам.

28. Имеются данные за десять лет о производительности труда и электроснабжении труда на одном из предприятий промышленности области:

Показатель	1	2	3	4	5	6	7	8	9	10
Среднегодовая выработка продукции на 1 рабочего, y_t	28,7	31,7	31,7	32,6	33,9	31,2	33,3	42,6	46,0	49,9
Электроснабжение, x_t	3,33	3,39	3,50	3,63	3,81	3,84	3,88	4,07	4,12	4,17

Результаты аналитического выравнивания привели к следующим уравнениям трендов для каждого из рядов:

а) для временного ряда производительности труда:

$$\hat{y}_t = 33,19 + 1,04 \cdot t + 0,95 \cdot t^2; \quad t = -9, -7, -5, -3, -1, 1, 3, 5, 7, 9;$$

б) для временного ряда электроснабжения труда:

$$\hat{x}_t = 3,774 + 0,049 \cdot t; \quad t = -9, -7, -5, -3, -1, 1, 3, 5, 7, 9.$$

Задание.

1) Определить коэффициент корреляции между временными рядами, используя непосредственно исходные уровни, первые разности для электроснабжения и вторые разности для производительности труда, отклонение от основной тенденции.

2) Объяснить различия полученных результатов.

3) Рассчитать коэффициент автокорреляции внутри каждого временного ряда.

29. На основе данных за последние 20 лет изучается зависимость между уровнем дивидендов по обыкновенным акциям y_t (%) от прибыли компании x_t (тыс. дол.). Имеется следующая информация:

1) Результаты аналитического выравнивания рядов:

Тренд в форме параболы:

а) для ряда y_t : $\hat{y}_t = 0,8 + 0,3 \cdot t - 0,05 \cdot t^2$, $R^2 = 0,95$;

б) для ряда x_t : $\hat{x}_t = 3 - 0,65 \cdot t - 0,01 \cdot t^2$, $R^2 = 0,85$.

Линейные тренды:

а) для ряда y_t : $\hat{y}_t = 2 + 0,05 \cdot t$, $R^2 = 0,38$;

б) для ряда x_t : $\hat{x}_t = 0,65 + 0,8 \cdot t$, $R^2 = 0,24$.

2) Коэффициенты корреляции:

по исходным данным уровням рядов – 0,98;

по отклонениям от трендов в форме параболы – 0,78;

по отклонениям от линейных трендов – 0,45;

по первым разностям – 0,42;

по вторым разностям – 0,84.

Задание.

1) Существует ли взаимосвязь между исследуемыми временными рядами? Если есть, то указать ее количественную характеристику. Обосновать ответ.

2) Пояснить причины различий полученных мер тесноты связи.

30. В таблице указаны остатки регрессии:

Год	Остатки	Год	Остатки	Год	Остатки
1	-0,7	5	0	9	0
2	0	6	0,3	10	0,3
3	-0,2	7	-0,1	11	0,3
4	0,9	8	-0,1	12	-0,1

Задание.

1) Оценить автокорреляцию остатков.

2) Используя критерий Дарбина – Уотсона, сделать вывод относительно рассматриваемой регрессии.

Приложения

1. Табличные значения критерия Дарбина – Уотсона

на уровне значимости 5 %

DW_l – нижняя граница, DW_u – верхняя граница,

p – число объясняющих переменных, n – число наблюдений

n	$p = 1$		$p = 2$		$p = 3$		$p = 4$		$p = 5$	
	DW_l	DW_u	DW_l	DW_u	DW_l	DW_u	DW_l	DW_u	DW_l	DW_u
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	1,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,85	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,99
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,34	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
60	1,55	1,62	1,51	1,65	1,58	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,60	1,70	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,60	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

2. Значения функции
$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$

x	Сотые доли								
	0	1	2	3	4	5	6	7	8
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0200	0,0239	0,0279	0,0319
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517
0,7	0,2580	0,2611	0,2642	0,2673	0,2703	0,2734	0,2764	0,2794	0,2823
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810
1,2	0,3849	0,3869	0,3883	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761
2,0	0,4772	0,4779	0,4783	0,4788	0,4793	0,4798	0,4803	0,4807	0,4812
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4879	0,4881	0,4884	0,4887
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980
2,9	0,4981	0,4982	0,4983	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990

3. Квантили распределения Стьюдента $t_p(k)$

k	p	0,900	0,950	0,975
1		3,078	6,314	12,706
2		1,886	2,920	4,303
3		1,638	2,353	3,182
4		1,533	2,132	2,776
5		1,476	2,015	2,571
6		1,440	1,943	2,447
7		1,415	1,895	2,365
8		1,397	1,860	2,306
9		1,383	1,833	2,262
10		1,372	1,812	2,228
11		1,363	1,796	2,201
12		1,356	1,782	2,179
13		1,350	1,771	2,160
14		1,345	1,761	2,145
15		1,341	1,753	2,131
16		1,337	1,746	2,120
17		1,333	1,740	2,110
18		1,330	1,734	2,101
19		1,328	1,729	2,093
20		1,325	1,725	2,086
21		1,323	1,721	2,080
22		1,321	1,717	2,074
23		1,319	1,714	2,069
24		1,318	1,711	2,064
25		1,316	1,708	2,060
26		1,315	1,706	2,056
27		1,314	1,703	2,052
28		1,313	1,701	2,048
29		1,311	1,699	2,045
30		1,310	1,697	2,042
40		1,303	1,684	2,021
60		1,296	1,671	2,000
120		1,289	1,658	1,980
∞		1,282	1,645	1,960

4. Квантили распределения Фишера $F_p(k_1, k_2)$

k_2	k_1	$p = 0,95$							
		1	2	3	4	5	6	7	8
1		161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9
2		18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37
3		10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85
4		7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04
5		6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82
6		5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15
7		5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73
8		5,32	4,66	4,07	3,84	3,69	3,58	3,50	3,44
9		5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23
10		4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07
11		4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95
12		4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85
13		4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77
14		4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70
15		4,54	3,68	3,29	3,05	2,90	2,79	2,71	2,64
16		4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59
17		4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55
18		4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51
19		4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48
20		4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45
21		4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42
22		4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40
23		4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37
24		4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36
25		4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34
26		4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32
27		4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31
28		4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29
29		4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28
30		4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27
40		4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18
60		4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10
120		3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02
∞		3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94

Список рекомендуемой литературы

Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике / В.Е. Гмурман. – М. : Высш. шк., 1975, 1997, 1998, 1999, 2000.

Гмурман В.Е. Теория вероятностей и математическая статистика / В.Е. Гмурман. – М. : Высш. шк., 1978, 1997, 1998, 1999, 2000.

Ежова Л.Н. Теория вероятностей : учеб. пособие / Л.Н. Ежова [и др.]. – Иркутск : Изд-во ИГЭА, 1996.

Ежова Л.Н. Теория вероятностей и математическая статистика / Л.Н. Ежова. – Иркутск : Изд-во ИГЭА, 2000.

Карасев А.И. Курс высшей математики для экономических вузов / А.И. Карасев, З.М. Аксютин, Т.И. Савельева. – М. : Высш. шк., 1982.

Колемаев В.А. Теория вероятностей и математическая статистика / В.А. Колемаев, О.В. Староверов, В.Б. Турундаевский. – М. : Высш. шк., 1991.

Колде Я.К. Практикум по теории вероятностей и математической статистики / Я.К. Колде. – М. : Высш. шк., 1991.

Ниворожкина Л.И. Основы статистики с элементами теории вероятностей для экономистов: руководство для решения задач / Л.И. Ниворожкина, З.А. Морозова, И.А. Герасимова, И.В. Житников. – Ростов н/Д : Феникс, 1999.

Мацкевич И.П. Сборник задач и упражнений по высшей математике: Теория вероятностей и математическая статистика : учеб. пособие / И.П. Мацкевич, Г.П. Свирид, Г.М. Булдык. – Минск : Высш. шк., 1996.

Учебное издание

Мамонова Наталья Вячеславовна

**Временные ряды:
анализ и прогнозирование**

Учебное пособие

ИД № 06318 от 26.11.01.
Подписано в пользование 20.03.19.

Издательство Байкальского государственного университета.
664003, г. Иркутск, ул. Ленина, 11.

<http://bgu.ru>.