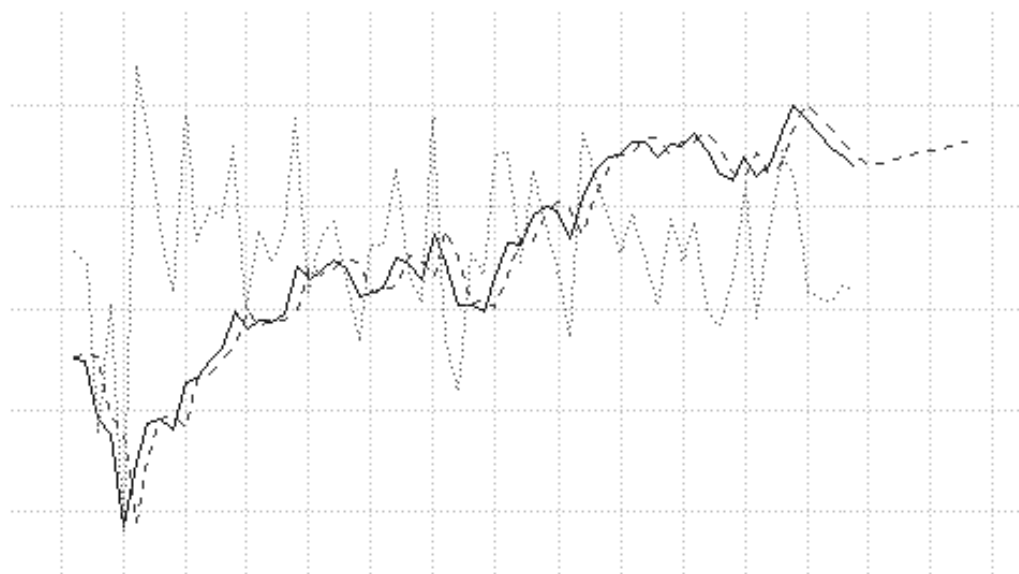


МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ,  
МОЛОДЕЖИ И СПОРТА УКРАИНЫ  
ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ  
ФАКУЛЬТЕТ МАТЕМАТИКИ И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ  
КАФЕДРА ПРИКЛАДНОЙ МАТЕМАТИКИ  
И ТЕОРИИ СИСТЕМ УПРАВЛЕНИЯ

С. А. Мельник, Н. В. Вайсруб, В. В. Грибинюкова

# Моделирование экономических, экологических и социальных процессов



Донецк  
ДонНУ  
2013

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ,  
МОЛОДЕЖИ И СПОРТА УКРАИНЫ  
ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ  
ФАКУЛЬТЕТ МАТЕМАТИКИ И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ  
КАФЕДРА ПРИКЛАДНОЙ МАТЕМАТИКИ  
И ТЕОРИИ СИСТЕМ УПРАВЛЕНИЯ

С. А. Мельник, Н. В. Вайсруб, В. В. Грибинюкова

# Моделирование экономических, экологических и социальных процессов

*Учебное пособие*

Утверждено  
на заседании ученого совета  
факультета математики и  
информационных технологий.  
Протокол № 125 от 14.02.2013 г.

Донецк  
ДонНУ  
2013

УДК 001(075.8)  
ББК (В)22,172

Авторы:

*С. А. Мельник*, кандидат физико-математических наук, доцент, старший научный сотрудник отделения теории вероятностей и математической статистики Института прикладной математики и механики НАН Украины,  
*Н. В. Вайсруб*, кандидат технических наук, доцент кафедры прикладной математики и теории систем управления факультета математики и информационных технологий,  
*В. В. Грибинюкова*, ассистент кафедры прикладной математики и теории систем управления факультета математики и информационных технологий.

Рецензент:

*Ю. В. Шамарин*, кандидат экономических наук, доцент кафедры прикладной математики и теории систем управления факультета математики и информационных технологий.

Ответственный за выпуск:

*С. В. Мышко*, доктор технических наук, профессор, заведующий кафедрой прикладной математики и теории систем управления факультета математики и информационных технологий.

Моделирование экономических, экологических и социальных процессов: учебное пособие / С. А. Мельник, Н. В. Вайсруб, В. В. Грибинюкова. – Донецк: ДонНУ, 2013. – 82 с.

В учебном пособии рассмотрены основные теоретические положения статистических методов моделирования рядов динамики, некоторые задачи математической экономики, связанные с оптимизацией производства. Особое внимание уделено методике построения эконометрических моделей для прогнозирования рядов динамики.

Для студентов специальностей «Прикладная математика» и «Информатика», математиков-аналитиков и разработчиков информационных систем моделирования и прогнозирования.

УДК 001(075.8)  
ББК (В)22,172  
© Мельник С. А., 2013  
© Вайсруб Н. В., 2013  
© Грибинюкова В. В., 2013  
© ДонНУ, 2013

## Содержание

Введение.....	6
Раздел 1 Некоторые эконометрические модели.....	7
1 Статистические методы моделирования	
экономических процессов .....	7
1.1 Постановка эконометрической задачи.....	7
1.2 Типы статистических данных .....	8
1.3 Методы сбора статистической информации.....	8
1.4 Принципы сбора и подготовка данных	
для использования.....	9
1.5 Методы обеспечения репрезентативности .....	9
1.6 Группировка данных.....	10
2 Некоторые сведения из теории вероятностей.....	11
2.1 Числовые характеристики случайной величины.....	11
2.2 Выборочные числовые характеристики .....	13
2.3 Поправки Шеппарда .....	14
2.4 Коэффициент корреляции .....	15
2.5 Свойства выборочных числовых характеристик.....	15
2.6 Основные законы распределения случайных величин,	
используемые в математической статистике .....	16
3 Основы теории динамических рядов .....	19
3.1 Обобщенная аддитивная модель динамического ряда .....	19
3.2 Построение корректирующей компоненты.....	21
3.3 Методы отбора функций тренда.....	22
3.3.1 Критерий Фостера–Стюарта.....	22
3.3.2 Критерий тренда .....	24
3.3.3 Критерий серий .....	24
3.3.4 Механические методы выделения тренда.	
Сглаживание экспериментальных данных.....	25
3.3.5 Типичные формы тренда.....	27
3.3.6 Предварительный отбор функций тренда .....	27
3.3.7 Идея метода наименьших квадратов .....	30
3.3.8 Свойства МНК-оценок .....	30
3.3.9 Применение МНК-оценок для построения тренда.....	32
3.3.10 Оценки точности модели тренда.....	34
3.4 Построение циклической компоненты .....	34
3.4.1 Механический метод построения	
циклической компоненты .....	37

3.4.2 Аналитический метод построения циклической компоненты .....	39
3.4.3 Методы отбора значимых гармоник .....	40
3.5 Проверка остаточной компоненты на адекватность условиям Гаусса–Маркова .....	41
3.5.1 Проверка ряда остатков на случайность .....	42
3.5.2 Проверка ряда остатков на соответствие нормальному распределению .....	43
3.5.3 Проверка равенства нулю математического ожидания остаточной компоненты .....	44
3.5.4 Исследование остаточной компоненты на автокорреляцию .....	45
3.5.5 Причины возникновения автокорреляции в ряду остатков .....	48
3.5.6 Устранение автокорреляции остатков методом авторегрессионного преобразования .....	49
3.5.7 Робастность методов анализа временных рядов .....	51
3.5.8 Свойства оценок при нарушении нормальности.....	52
3.6 Построение прогнозов временного ряда .....	54
Контрольные вопросы к разделу 1 .....	60
Раздел 2 Некоторые задачи математической экономики.....	62
1 Производственные функции .....	62
2 Основные виды производственных функций .....	64
3 Основные типы задач оптимизации производства.....	65
3.1 Задача максимизации прибыли на долговременном промежутке функционирования .....	65
3.2 Задача максимизации прибыли на кратковременном промежутке функционирования.....	66
3.3 Задача максимизации объема выпуска продукции при ограниченных финансовых ресурсах.....	70
3.4 Задача минимизации издержек производства при фиксированном объеме выпуска продукции .....	72
Контрольные вопросы к разделу 2 .....	73
Список рекомендуемой литературы.....	74
Статистические таблицы .....	75

## **Введение**

В данном учебном пособии кратко рассмотрен теоретический материал, излагаемый студентам факультета математики и информационных технологий специальностей «Прикладная математика» и «Информатика» в пределах общего курса «Моделирование экономических, экологических и социальных процессов». Материал состоит из двух самостоятельных разделов: эконометрические модели рядов динамики и задачи оптимизации производства.

Эконометрия – это наука, которая исследует количественные закономерности и взаимосвязи, присущие экономическим, экологическим, социальным процессам и системам, при помощи методов математической статистики. Задачи оптимизации производства являются составной частью математической экономики.

В первом разделе учебного пособия изложена методика построения эконометрических моделей, описывающих динамику экономических показателей. Во втором разделе рассмотрены основные экономические показатели и задачи оптимизации производства. Третий раздел содержит статистические таблицы, необходимые для реализации статистических методов при решении эконометрических задач.

В данном учебном пособии внимание сосредоточено на эконометрических моделях, однако универсальность математических моделей позволяет без труда адаптировать излагаемые методы к задачам, решаемым в экологических и социологических науках.

## Раздел 1 Некоторые эконометрические модели

Перечень задач, решаемых эконометрическими методами, весьма обширен. Среди них можно выделить два класса: задачи многомерного статистического анализа и задачи статистики временных рядов. В задачах первого типа изучается зависимость некоторого суммарного показателя от группы первичных показателей при наличии случайных воздействий на изучаемый объект. В задачах второго типа – зависимость некоторого суммарного показателя от единственного первичного показателя – времени. При внешней схожести этих задач каждая из них имеет свои особенности, и методы их решения существенно отличаются. В данном учебном пособии основное внимание уделено изучению задач моделирования динамики экономических показателей с течением времени.

### 1 Статистические методы моделирования экономических процессов

В реальной действительности на состояние экономических объектов существенное влияние оказывают всевозможные случайные факторы: погодные условия, природные катаклизмы, поведение поставщиков и потребителей, поведение политиков, конкурентная борьба и так далее. Как правило, количество таких факторов велико и природа их неясна. Поэтому основным инструментом построения математических моделей процессов, протекающих в экономике, являются методы математической статистики.

Особенностью задач, которые будем рассматривать, является зависимость изучаемых показателей от времени, т. е. значения показателей не только случайны, но еще и меняются с течением времени согласно некоторой закономерности. Кроме того, суммарный показатель будет один. В такой ситуации исследователь располагает возможностью измерять конечное число раз реальные значения суммарного показателя. Основываясь на этих данных, он должен максимально точно определить характер зависимости суммарного показателя от времени. Если из экономической теории известен принципиальный вид зависимости от времени, то задача несколько упрощается и может быть решена методами статистического анализа временных рядов. Если же экономическая теория не дает такой подсказки, то исследователь вынужден вначале отобрать функции-кандидаты в модели, а затем по результатам исследований выбрать одну из них. В данном учебном пособии излагается методика построения модели во втором случае, т. е. когда вид искомой зависимости заранее не известен.

#### 1.1 Постановка эконометрической задачи

Вначале рассмотрим общие понятия, используемые в прикладном статистическом анализе, а затем перейдем к изучению методов построения математических моделей экономических процессов.

##### *Общая постановка эконометрической задачи*

Рассматривается некоторый экономический объект, состояние которого характеризуется первичными факторами  $x_1, \dots, x_r$  и вторичным

суммарным фактором  $y$ . Предполагается, что  $y$  некоторым образом зависит от  $x_1, \dots, x_r$ . Экспериментальным путем в результате  $n$ -кратных измерений получены данные о значениях факторов.

$$y_1, x_{11}, \dots, x_{r1};$$

$$y_2, x_{12}, \dots, x_{r2};$$

.....

$$y_n, x_{1n}, \dots, x_{rn}.$$

Необходимо

1) выбрать некоторую функцию  $y = f(x_1, \dots, x_r, a_1, \dots, a_m)$ , где  $a_1, \dots, a_m$  – параметры, которая, по мнению исследователя, задает связь  $y_i$  с  $x_{ji}$ ,  $j = \overline{1, r}$ ,  $i = \overline{1, n}$ ;

2) построить формулы, позволяющие находить приближенные значения параметров  $a_k \approx a_k^*(x_{ij}, y_j)$ ,  $k = \overline{1, m}$ ;

3) оценить степень точности построенной модели  $y \approx f(x_1, \dots, x_r, a_1^*, \dots, a_m^*)$ ;

4) оценить уровень адекватности построенной модели рассматриваемому экономическому процессу.

## 1.2 Типы статистических данных

Первым этапом решения эконометрической задачи является сбор статистических данных об изучаемом показателе. Для того, чтобы исследователь имел возможность применять методы математической статистики, сбор данных должен производиться по определенной методике. Тип собранных данных определяет набор математических инструментов, которые будут использоваться. Различают два основных типа статистических данных: **cross-section** – перекрестные данные и **time-series** – временные ряды.

**Определение.** *Перекрестные данные* – это данные, полученные для разных однотипных объектов по одному и тому же показателю в один и тот же момент времени, либо не зависящие от времени.

Например, уровень средней зарплаты в различных регионах страны.

**Определение.** *Временные ряды* характеризуют изменение одного и того же показателя, измеренного у одного и того же объекта в разные моменты времени.

Например, ежемесячные значения уровня инфляции в течение календарного года.

## 1.3 Методы сбора статистической информации

1. Получение информации из официальных статистических справочников.

2. Получение информации путем прямого измерения показателей объекта.



3. Получение информации с помощью опросов.
4. Анкетирование (обычно для показателей, которые не могут быть непосредственно измерены).

#### 1.4 Принципы сбора и подготовка данных для использования

1. При сборе данных необходимо произвести первичный анализ создаваемой модели и составить список показателей  $x_1, \dots, x_r$ , которые действительно влияют на изучаемый показатель  $y$ . Этот список должен быть исчерпывающим.

2. Необходимо заранее определить, каков характер случайных факторов, влияющих на данные.

3. Желательно, чтобы при сборе статистических данных результаты одного измерения не влияли на результаты других измерений.

4. При сборе статистических данных должна быть обеспечена репрезентативность данных.

5. Перед использованием показатели следует привести к одной единице измерения.

Основные виды представления исходных данных: *таблицы, диаграммы, точечные графики.*

При сборе статистических данных важно знать множество возможных значений измеряемых показателей. Такое множество принято называть *генеральной совокупностью* значений измеряемых показателей.

В результате измерения каждого показателя исследователь получает набор чисел, называемый выборкой измеряемых значений.

**Определение.** *Выборка* объема  $n$  – это численные результаты  $n$ -кратно повторенных измерений изучаемого показателя  $y$ , записанные в виде числового вектора  $(y_1, \dots, y_n)$ .

Одним из важнейших понятий прикладной статистики является репрезентативность выборки.

**Определение.** Выборку называют *репрезентативной*, если она обладает следующими свойствами:

- 1) обеспечен случайный выбор данных из генеральной совокупности;
- 2) выборка охватывает все классы возможных значений показателя;
- 3) количество представителей каждого класса достаточно велико, чтобы отражать значимость класса в составе генеральной совокупности.

#### 1.5 Методы обеспечения репрезентативности

##### 1. Метод пропорциональности

Выбор объектов для получения выборки производится пропорционально численности класса в составе генеральной совокупности.

##### 2. Метод случайности

Обеспечивается с помощью генераторов случайных величин.

## 1.6 Группировка данных

В ряде случаев перед использованием данных производят их группировку. Группировка данных может производиться, если количество данных достаточно велико ( $>100$ ) и известно, что данные представляют собой значения абсолютно непрерывной случайной величины.

### *Порядок выполнения группировки данных*

1. Если теоретически заранее известно, что изучаемый показатель  $y \in [a, b]$  (т. е. отрезок  $[a, b]$  – генеральная совокупность), то тогда рассматривают весь отрезок и работают с ним. Если отрезок бесконечен или заранее неизвестен, то в выборке находят левый и правый крайние члены вариационного ряда

$$y_{(1)} = \min_{i=1, n} y_i, \quad y_{(n)} = \max_{i=1, n} y_i$$

и далее рассматривается выборочный отрезок  $[y_{(1)}, y_{(n)}]$ .

2. Разбивают отрезок на  $m$  непересекающихся полуинтервалов, имеющих одинаковую длину. В каждый полуинтервал должно попасть как минимум 8–10 значений выборки.

3. Определяют середины полуинтервалов (или же медианы)  $z_1, \dots, z_m$ .

4. Исходная случайная величина  $\eta$  с выборочными значениями  $y_1, \dots, y_n$  заменяется искусственно созданной вторичной случайной величиной  $\zeta$  с выборочными значениями  $z_1, \dots, z_m$ .

Полученная таким образом выборка  $(z_1, \dots, z_m)$  называется *вторичной*. Каждому построенному в пункте 2 полуинтервалу ставится в соответствие частота  $\nu_k$  попадания значения случайной величины в этот полуинтервал,  $k = \overline{1, m}$ . Дальнейшая работа выполняется на основе вторичной выборки.

## 2 Некоторые сведения из теории вероятностей

Рассмотрим некоторые сведения из теории вероятностей и математической статистики, необходимые для понимания дальнейшего теоретического материала.

**Определение.** *Случайная величина  $\xi(\omega)$  – это величина, которая принимает то или иное значение из множества допустимых значений в зависимости от исхода  $\omega$  случайного эксперимента.*

Правила нахождения вероятностей принятия случайного величиной того или иного значения изучались в курсе теории вероятностей.

**Пример 2.1.** Рассмотрим эксперимент по подбрасыванию игрального кубика. Возможно определить следующие случайные величины.

$\xi(\omega)$  – количество очков, выпавших на верхней грани кубика; множество допустимых значений  $\{1, \dots, 6\}$ . Найдем, например, вероятность события  $P\{\xi(\omega) < 4\}$ .

$$P\{\xi(\omega) < 4\} = P\{\xi(\omega) = 1\} + P\{\xi(\omega) = 2\} + P\{\xi(\omega) = 3\} = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

$\eta(\omega)$  – сумма очков, выпавших на верхней грани и противоположной к ней грани; множество допустимых значений состоит из одного числа 7. Таким образом,  $P\{\eta(\omega) = 7\} = 1$ .

**Пример 2.2.** Рассмотрим случайную величину  $\zeta(\omega)$ , равную расстоянию, которое пролетит снаряд при выстреле из орудия. В данном случае множество значений случайной величины  $\zeta(\omega)$  представляет собой некоторый числовой интервал  $(a; b)$ , а вероятности событий, связанных с данной случайной величиной, следует вычислять с помощью функции распределения.

**Определение.** *Функцией распределения вероятностей случайной величины  $\xi(\omega)$  называется функция  $F_\xi(x): R \mapsto [0; 1]$ , определяемая правилом  $F_\xi(x) = P\{\xi < x\}$ .*

**Определение.** *Плотностью распределения вероятностей непрерывной случайной величины  $\xi(\omega)$  называется функция  $f_\xi(x): R \mapsto R^+$  такая, что  $f_\xi(x) = F'_\xi(x)$ .*

### 2.1 Числовые характеристики случайной величины

К понятию «числовые характеристики» случайной величины  $\xi(\omega)$  относят следующие величины.

1. *Не центрированный момент  $k$ -го порядка  $M^{\xi^k}(\omega) = \int_{-\infty}^{+\infty} x^k f_\xi(x) dx$ .*

Особо выделяют *математическое ожидание*, которое для абсолютно непрерывных случайных величин задается формулой

$$M\xi(\omega) = \int_{-\infty}^{+\infty} x f_{\xi}(x) dx.$$

Математическое ожидание – это теоретически ожидаемое среднее значение случайной величины  $\xi(\omega)$ .

2. *Центрированный момент  $k$ -го порядка*  $M(\xi(\omega) - M\xi(\omega))^k$ .

Особо выделяют *дисперсию*

$$D\xi(\omega) = M|\xi(\omega) - M\xi(\omega)|^2.$$

$\sqrt{D\xi(\omega)}$  – теоретически ожидаемый среднеквадратический разброс значений случайной величины  $\xi(\omega)$ .

3. *Асимметрия* случайной величины  $A_{\xi} = \frac{M(\xi - M\xi(\omega))^3}{(D\xi(\omega))^{3/2}}$ .

Асимметрия случайной величины определяет уровень несимметричности плотности распределения  $\xi(\omega)$ .

На рис. 2.1 сплошной линией изображен график плотности нормально распределенной случайной величины  $y = f_{\xi}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$ . Этот график считается эталоном асимметрии. Асимметрия нормально распределенной случайной величины равна нулю. Пунктирными линиями схематично изображены графики плотностей распределения случайных величин, имеющих положительную и отрицательную асимметрию.

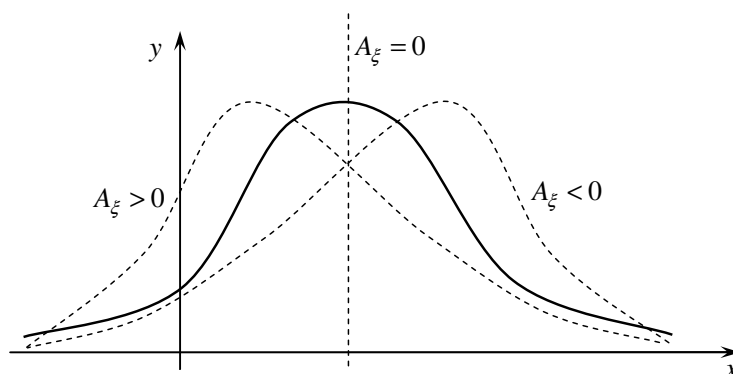


Рис. 2.1 – Графики плотностей с различными значениями асимметрии

4. *Эксцесс* случайной величины  $E_{\xi} = \frac{M|\xi - M\xi|^4}{(D\xi)^2} - 3$ .

Эксцесс случайной величины определяет степень «остроты» плотности распределения величины  $\xi(\omega)$ .

На рис. 2.2 сплошной линией изображен график плотности нормально распределенной случайной величины  $y = f_{\xi}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$ . Этот график считается эталоном эксцесса. Эксцесс нормально распределенной случайной величины равен нулю. Пунктирными линиями схематично

изображены графики плотностей распределения случайных величин, имеющих положительный и отрицательный эксцесс.

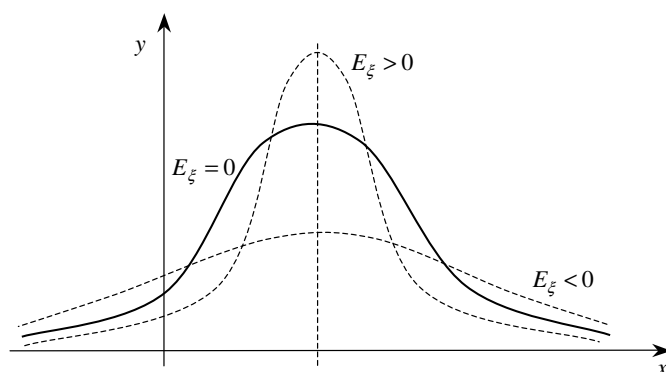


Рис. 2.2 – Графики плотностей с различными значениями эксцесса

5. Медиана случайной величины  $\xi(\omega)$  – это такое число  $\mu$ , что

$$P\{\xi(\omega) < \mu\} = P\{\xi(\omega) > \mu\} = \frac{1}{2}.$$

На рис. 2.3 схематично изображен график некоторой плотности распределения случайной величины и ее медиана  $\mu$ . Площади левой и правой частей заштрихованной области одинаковы и равны 0,5.

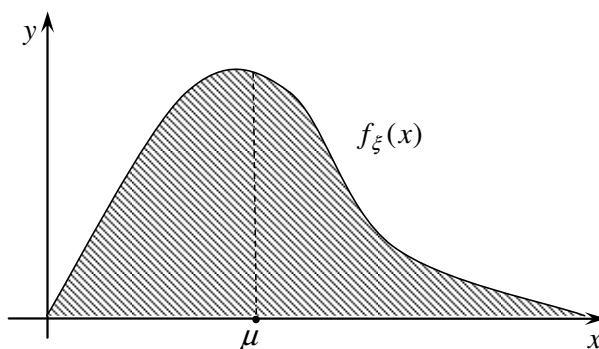


Рис. 2.3 – Определение медианы  $\mu$  случайной величины  $\xi(\omega)$

## 2.2 Выборочные числовые характеристики

К понятию «числовые характеристики выборки значений случайной величины  $\xi(\omega)$ » относят величины, которые являются выборочными аналогами соответствующих теоретических числовых характеристик самой случайной величины. Вследствие усиленного закона больших чисел, для широкого класса случайных величин выборочные числовые характеристики с вероятностью единица к своим теоретическим аналогам. Это позволяет использовать выборочные числовые характеристики для приближенного вычисления значений их теоретических аналогов.

1. Нецентрированный выборочный момент  $k$ -го порядка  $\bar{x}^k = \frac{1}{n} \sum_{i=1}^n x_i^k$ ,

где  $\bar{x} = (x_1, \dots, x_n)$  – выборка измерений случайной величины  $\xi(\omega)$ .

Нецентрированный выборочный момент  $k$ -го порядка является выборочным аналогом нецентрированного выборочного момента  $k$ -го

порядка изучаемой случайной величины  $M\xi^k \approx \frac{1}{n} \sum_{i=1}^n x_i^k$  при больших  $n$ .

Особо выделяют *выборочное среднее*  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \approx M\xi(\omega)$ .

2. *Центрированный выборочный момент 2-го порядка*, или *выборочная дисперсия*  $S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \approx M(\xi - M\xi)^2$ .

Поскольку данная формула содержит систематическую погрешность, в численных расчетах для приближенного вычисления дисперсии случайной величины используют *исправленную выборочную дисперсию*

$$\hat{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{n}{n-1} S^2 \approx D\xi(\omega).$$

Эта формула уже не содержит систематической погрешности.

Числа  $\sqrt{\hat{S}^2}$  и  $\sqrt{S^2}$  называются *исправленным* и *неисправленным среднеквадратическим отклонением* выборки от выборочного среднего.

3. Аналогично пунктам 1, 2, заменяя теоретические моменты их выборочными аналогами, получают *выборочные асимметрию* и *эксцесс*.

4. *Медиана* выборки – это такое число  $m$ , что количество элементов выборки, меньших  $m$ , равно количеству элементов выборки, больших  $m$ .

### 2.3 Поправки Шеппарда

При группировке выборки (см. пункт 1.6) возникает вторичная выборка, числовые характеристики которой в общем случае не равны числовым характеристикам исходной выборки. Поскольку вторичная выборка используется для приближенного вычисления числовых характеристик первичной выборки, то для уменьшения погрешности применяют *поправки Шеппарда*.

Пусть  $x_1, \dots, x_n$  – исходная выборка,  $\underbrace{z_1, \dots, z_1}_{n_1}, \underbrace{z_2, \dots, z_2}_{n_2}, \dots, \underbrace{z_m, \dots, z_m}_{n_m}$  – вторичная, где  $n_i$  – количество элементов первичной выборки, попавших в  $i$ -й полуинтервал,  $\sum_{i=1}^m n_i = n$ . Приближенные формулы вычисления выборочных моментов с учетом необходимых поправок имеют следующий вид.

Для моментов нечетного порядка

$$\bar{x} \approx \frac{1}{n} \sum_{j=1}^m z_j \cdot n_j;$$

$$\overline{(x - \bar{x})^3} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 \approx \frac{1}{n} \sum_{j=1}^m (z_j - \bar{z})^3 \cdot n_j;$$

и так далее для всех моментов нечетного порядка.

Для моментов четного порядка

$$S^2 \approx \frac{1}{n} \sum_{j=1}^m (z_j - \bar{z})^2 \cdot n_j - \frac{h^2}{12},$$

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 \approx \frac{1}{n} \sum_{j=1}^m (z_j - \bar{z})^4 - \frac{h^2}{2} \cdot \frac{1}{n} \sum_{j=1}^m (z_j - \bar{z})^4 \cdot n_j + \frac{7}{240} h^4,$$

где  $h$  – длина полуинтервалов разбиения.

## 2.4 Коэффициент корреляции

Коэффициент корреляции двух случайных величин играет важную роль при изучении взаимного влияния случайных величин.

**Определение.** Пусть  $\xi(\omega)$  и  $\eta(\omega)$  – две случайные величины. Число

$$\rho_{\xi, \eta} = \frac{M[(\xi(\omega) - M\xi(\omega))(\eta(\omega) - M\eta(\omega))]}{\sqrt{D\xi(\omega)} \cdot \sqrt{D\eta(\omega)}}$$

называется *коэффициентом корреляции* случайных величин  $\xi(\omega)$  и  $\eta(\omega)$ .

*Свойства коэффициента корреляции*

1.  $-1 \leq \rho_{\xi, \eta} \leq 1$ .
2. Если  $\xi$  и  $\eta$  независимы, то они некоррелированы, т. е.  $\rho_{\xi, \eta} = 0$ .  
Обратное в общем случае неверно.
3. Если  $\xi$  и  $\eta$  имеют совместное нормальное распределение, то их независимость равносильна некоррелированности.
4. Если  $\rho_{\xi, \eta} = -1$ , то существуют числа  $a, b \in R$ ,  $a < 0$ , такие, что  $\eta = a\xi + b$ .
5. Если  $\rho_{\xi, \eta} = 1$ , то существуют числа  $a, b \in R$ ,  $a > 0$ , такие, что  $\eta = a\xi + b$ .

Выборочным аналогом коэффициента корреляции является *выборочный коэффициент корреляции*

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Отметим, что  $-1 \leq r \leq 1$ .

## 2.5 Свойства выборочных числовых характеристик

1. Если ко всем элементам выборки прибавить число  $C$ , то  $\bar{x}$  изменится на  $C$ :  $\overline{x+C} = \bar{x} + C$ .
2. Если все элементам выборки умножить на одно и то же число  $C$ , то  $\bar{x}$  умножится на  $C$ :  $\overline{Cx} = C\bar{x}$ .

$$3. \frac{1}{n} \sum_{i=1}^n (x_i + C - \overline{x+C})^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

$$4. \frac{1}{n} \sum_{i=1}^n (Cx_i - \overline{Cx})^2 = C^2 \cdot \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

## 2.6 Основные законы распределения случайных величин, используемые в математической статистике

1. *Равномерное распределение* на отрезке  $[a; b]$  –  $U[a; b]$

1) эмпирические условия возникновения  $U[a; b]$ :

- значения  $\xi(\omega)$  лежат на  $[a; b]$ ;
- шансы того, что значения  $\xi(\omega)$  будут на интервале  $(\alpha; \beta) \subseteq [a; b]$ , зависят только от длины интервала и не зависят от положения  $(\alpha; \beta)$  на  $[a; b]$ .

2) плотность распределения

$$f_{\xi}(x, a, b) = \begin{cases} \frac{1}{b-a}, & \text{если } x \in [a; b]; \\ 0, & \text{если } x \notin [a; b]. \end{cases}$$

3) числовые характеристики

$$M\xi = \frac{a+b}{2} = \mu; \quad D\xi = \frac{(b-a)^2}{12};$$

$$M\xi^k = \frac{1}{b-a} \cdot \frac{b^{k+1} - a^{k+1}}{k+1}.$$

4) стандартизуемость:  $\eta = \frac{\xi - a}{b - a} \sim U[0; 1]$ .

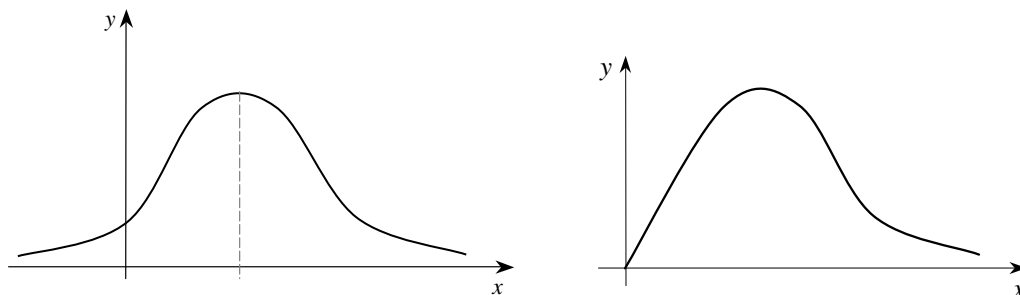
2. *Нормальное распределение* –  $N(a; \sigma^2)$ .

1) эмпирические условия возникновения  $N(a; \sigma^2)$ :

- $\xi(\omega)$  может принимать какие угодно действительные значения;
- распределение имеет ось симметрии;
- изучаемый случайный показатель образуется в результате суммирования многочисленных независимых случайных факторов.

2) плотность распределения (рис. 2.4 а))

$$f_{\xi}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad x \in (-\infty; +\infty).$$





а) б)  
Рис. 2.4 – График плотности: а) нормального распределения;  
б)  $\chi^2$ -распределения

3) числовые характеристики

$$\begin{aligned} M\xi &= a; \quad D\xi = \sigma^2; \\ A\xi &= 0; \quad E\xi = 0; \\ M(\xi(\omega) - M\xi(\omega))^k &= \begin{cases} 0, & \text{если } k \text{ нечетное;} \\ 1 \cdot 2 \cdot \dots \cdot (2m-1) \sigma^{2m}, & \text{если } k \text{ четное.} \end{cases} \end{aligned}$$

4) стандартизуемость: если  $\xi \sim N(a; \sigma^2)$ , то  $\xi = \frac{\xi - a}{\sigma} \sim N(0; 1)$ .

*Основные свойства  $N(a; \sigma^2)$*

1. Если  $\vec{\xi}(\omega) = (\xi_1, \dots, \xi_n)$  имеет многомерное нормальное распределение, то любой подвектор  $(\xi_{i1}, \dots, \xi_{is})$ ,  $1 \leq s \leq n$ , также имеет нормальное распределение.

2. Если вектор  $\vec{\xi}(\omega)$  имеет многомерное нормальное распределение, и его координаты некоррелированы, то есть  $M(\xi_i - a_i)(\xi_j - a_j) = 0$ ,  $i \neq j$ , то  $\xi_i$  и  $\xi_j$  являются независимыми случайными величинами.

3.  $\chi^2$ -распределение с  $n$  степенями свободы –  $\chi^2(n)$ .

1) условия возникновения:

– если  $\xi_1, \dots, \xi_n$  – независимые стандартные нормальные случайные величины, то величина  $\chi^2(n) = \sum_{i=1}^n \xi_i^2$  имеет  $\chi^2(n)$ -распределение.

2) плотность распределения (рис. 2.4 б))

$$f_{\chi^2(n)}(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$$

3) числовые характеристики

$$M\chi^2(n) = n; \quad D\chi^2(n) = 2n.$$

4. *Распределение Стьюдента с  $n$  степенями свободы*

1) условия возникновения:

– случайная величина  $\tau(n)$  имеет распределение Стьюдента с  $n$

степенями свободы, если  $\tau(n) = \frac{\xi_0}{\sqrt{\frac{1}{n} \sum_{i=1}^n \xi_i^2}}$ , где  $\xi_0, \xi_1, \dots, \xi_n$  – независимые

случайные величины с распределением  $N(0;1)$ .

2) плотность распределения (рис. 2.5 а))

$$f_{\tau(n)}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \cdot \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in (-\infty; +\infty).$$

При  $n \rightarrow \infty$   $\tau(n) \approx N(0;1)$ .

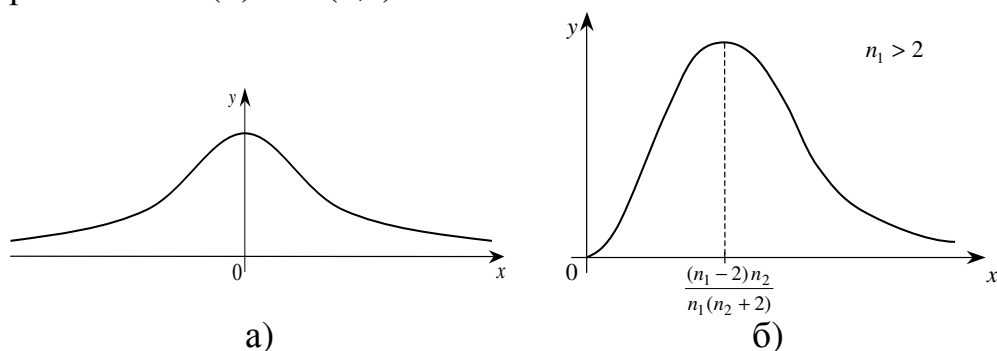


Рис. 2.5 – График плотности: а) распределения Стьюдента; б) распределения Фишера

3) числовые характеристики

$$M\tau(n) = 0;$$

$$D\tau(n) = \begin{cases} \frac{n}{n-2}, & \text{если } n > 2; \\ +\infty, & \text{если } n = 1, 2; \end{cases} \quad \lim_{n \rightarrow \infty} D\tau(n) = 1.$$

## 5. Распределение Фишера

1) условия возникновения:

– случайная величина  $\varphi(n_1; n_2)$  имеет распределение Фишера с

$(n_1; n_2)$  степенями свободы, если  $\varphi(n_1; n_2) = \frac{\frac{1}{n_1} \chi^2(n_1)}{\frac{1}{n_2} \chi^2(n_2)}$ , где  $\chi^2(n_1)$  и

$\chi^2(n_2)$  – независимые случайные величины с распределением  $\chi^2$ .

2) график плотности распределения изображен на рис. 2.5 б).

4) числовые характеристики ( $n_1 > 2$ )

$$M\varphi(n_1, n_2) = \frac{n_2}{n_2 - 2}, \quad D\varphi(n_1, n_2) = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}, \quad n_2 > 4.$$



### 3 Основы теории динамических рядов

Будем изучать поведение некоторой системы, состояния которой меняются с течением времени. Основными факторами, влияющими на систему, являются время и случайные помехи. Обозначим  $\eta(t, \omega)$  – возможное значение измеряемого показателя системы в момент времени  $t$ , где  $\omega$  – элемент некоторого пространства элементарных исходов  $\Omega$ . Если состояния системы замерять в конкретные моменты времени  $t \in \{t_1, t_2, \dots, t_n\}$ , то получим выборку измерений  $\eta(t_i, \omega_0) = y_i, i = 1, \dots, n$ .

Пусть значения изучаемого показателя измерили некоторое количество раз и получили выборку следующего вида:

$y_1, y_2, \dots, y_n$  – значения показателя  $y$ ,

$t_1, t_2, \dots, t_n$  – моменты замеров,

$n$  – количество замеров.

**Определение.** Случайные величины  $\eta(t_1, \omega), \eta(t_2, \omega), \dots, \eta(t_n, \omega)$  называются *динамическим рядом (временным рядом)*, а полученные при измерении данные  $y_1, y_2, \dots, y_n$  – *значениями динамического ряда*, или *выборкой из динамического ряда*.

Обычно считают моменты времени наблюдений равноотстоящими, т. е.  $t_{i+1} - t_i = \Delta t = \text{const}, i = 1, \dots, n-1$ .

Динамические ряды разделяют на моментные и интервальные.

**Определение.** *Моментными* временными рядами называют те ряды, у которых  $y_i$  – значение, отражающее сиюминутное состояние системы. Например, мировая цена нефти в определенный момент времени.

**Определение.** *Интервальными* временными рядами называют ряды, у которых  $y_i$  отражает некоторое суммарное значение показателя за некоторый интервал времени. Например, количество потребленного городом газа за сутки.

#### 3.1 Обобщенная аддитивная модель динамического ряда

Как сказано выше, будем рассматривать только такие системы, поведение которых зависит лишь от времени и случайных факторов. Для таких систем характерным является наличие некоторой закономерности изменения с течением времени. Часто модель имеет вид

$$\eta(t, \omega) = f(t; a_0, \dots, a_m) + \varepsilon(t, \omega),$$

где  $f$  – неслучайная функция,  $a_0, \dots, a_m$  – ее параметры,  $\varepsilon(t, \omega)$  – случайная помеха. Указанная неслучайная закономерность может формироваться из двух компонент: монотонной и периодической. Влияние случайных помех описывает случайная компонента (шум).

**Определение.** *Трендом* называют кусочно монотонно возрастающую или монотонно убывающую непериодическую неслучайную функцию времени  $U(t)$ , которая выражает основную тенденцию развития процесса

на некотором отрезке времени.

На рис. 3.1 схематично изображены график тренда  $U(t)$  и траектория случайного процесса  $\eta(t, \omega)$ . Точками обозначены результаты измерений.

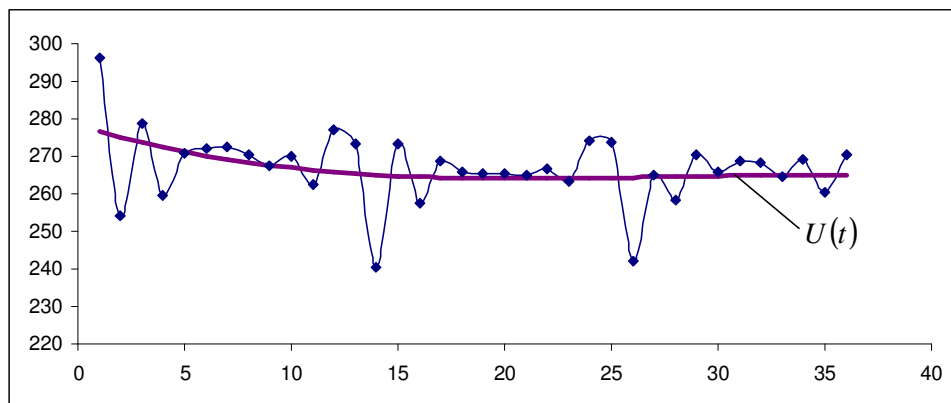


Рис. 3.1 – График тренда  $U(t)$  и траектория случайного процесса  $\eta(t, \omega)$

**Определение.** *Циклической компонентой* называют периодическую неслучайную функцию времени  $V(t)$ , которая характеризует закономерные периодические колебания значений показателя (влияние времени года, суток, недельных и месячных циклов).

**Определение.** *Случайной компонентой* называют последовательность случайных величин  $\varepsilon(t, \omega)$ ,  $t = t_1, t_2, \dots, t_n$ , описывающих случайные отклонения результатов измерений от функции тренда и циклической компоненты.

Одной из распространенных в природе форм моделей динамического ряда считается модель аддитивного типа  $\eta(t, \omega) = U(t) + V(t) + \varepsilon(t, \omega)$ .

На рис. 3.2 схематично изображены график аддитивной модели  $y(t) = U(t) + V(t)$  и траектория случайного процесса  $\eta(t, \omega)$ .

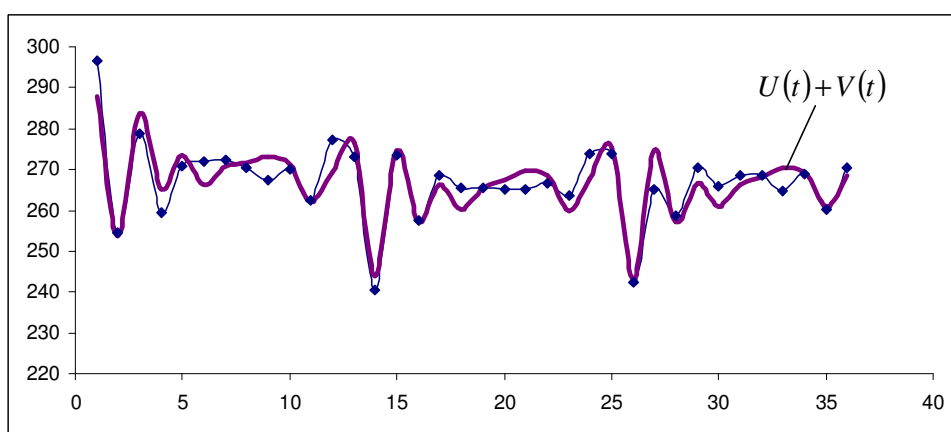


Рис. 3.2 – График аддитивной модели  $y(t) = U(t) + V(t)$  и траектория случайного процесса  $\eta(t, \omega)$

Для того, чтобы значения  $y_1, \dots, y_n$  были сопоставимы между собой (например, при анализе данных, представляющих месячные показатели), в модель добавляют *корректирующую компоненту*  $c(t)$

$$\eta(t, \omega) = c(t) + U(t) + V(t) + \varepsilon(t, \omega).$$

Полученное выражение называется *обобщенной аддитивной моделью динамического ряда*.

Встречаются (но реже) и другие модели, например, мультипликативная модель Самуэльсона

$$\eta(t, \omega) = e^{at} \cdot e^{\varepsilon(t, \omega)}.$$

Подробная информация о разнообразных моделях временных рядов приведена в [1–4]. В данном учебном пособии рассмотрены вопросы построения обобщенной аддитивной модели динамического ряда.

Таким образом, поставлена следующая эконометрическая задача.

Даны результаты экспериментальных измерений

$$y_1, \dots, y_n$$

$$t_1, \dots, t_n$$

Необходимо найти компоненты обобщенной аддитивной модели  $c(t)$ ,  $U(t)$ ,  $V(t)$ , оценить  $\varepsilon(t, \omega)$ .

Процесс решения поставленной задачи состоит из четырех этапов:

- 1) построение корректирующей компоненты  $c(t)$ ;
- 2) построение тренда  $U(t)$ ;
- 3) построение циклической компоненты  $V(t)$ ;
- 4) исследование остаточной компоненты.

**Определение.** Процесс выделения тренда и циклической компоненты называется *фильтрацией* динамического ряда.

Рассмотрим каждый этап построения аддитивной модели.

### 3.2 Построение корректирующей компоненты

Наличие корректирующей компоненты в создаваемой модели не является обязательным. Корректирующая компонента строится тогда, когда исходная выборка является интервальной, временные интервалы между измерениями различны и исходные данные необходимо привести к сопоставимому виду.

Поясним сказанное примерами. Если, например,  $y_i, i = 1, \dots, n$ , – это объемы дневной выручки магазина, то они являются сопоставимыми данными, поскольку продолжительность рабочего дня магазина фиксирована и не изменяется в процессе измерений. В этом случае нет необходимости строить корректирующую компоненту. Если же  $y_i, i = 1, \dots, n$ , – это объемы месячной выручки магазина, то они не являются сопоставимыми данными, поскольку продолжительность рабочего месяца магазина изменяется в процессе измерений. В этом случае данные необходимо привести к сопоставимому виду. Для этого строят корректирующую компоненту.

Рассмотрим построение корректирующей компоненты на примере. Пусть  $y_1, \dots, y_{12}$ , – это ежемесячные данные за некоторый год. В различных

месяцах года различное количество рабочих дней. Однако наиболее часто продолжительность месяца равняется 31 дню, что и будем считать стандартом. Тогда значения  $y_1, y_3, y_5, y_7, y_8, y_{10}, y_{12}$  не требуют корректировки и для них  $c(t) = 0$ . Апрельское значение  $y_4$  требует корректировки, поскольку количество дней в апреле равно 30, что не соответствует выбранному стандарту. Вычислив  $y_4 / 30$ , получим средний объем дневной выручки в апреле. Вычислив  $31 \cdot y_4 / 30$ , получим стандартизованную выручку за апрель. Это данное уже сопоставимо с данными за стандартные месяцы. Корректирующая компонента для апреля имеет вид  $c_4 = 31y_4 / 30 - y_4 = y_4 / 30$ . Прделаав аналогичные вычисления для остальных нестандартных месяцев, получим корректирующую компоненту  $c(t)$  и выборку в сопоставимом виде.

Стандартную продолжительность месяца можно выбрать иначе. Например, можно считать ее равной среднему числу дней в месяце. После приведения данных к сопоставимому виду переходят к выявлению тренда.

### 3.3 Методы отбора функций тренда

Прежде чем приступить к выбору тренда, необходимо провести всесторонний системный анализ изучаемого объекта.

Такой анализ предполагает

- выяснение принципиального математического характера зависимости показателя  $y$  от переменной  $t$ ;
- как, согласно теории, должен вести себя темп роста, темп прироста, эластичность изучаемого показателя;
- присущи ли изучаемому объекту периодические закономерности или монотонные тенденции.

Прежде чем приступить к построению тренда, следует выяснить, имеется ли в ряду измерений какая-либо тенденция. Поскольку ряд измерений представляет собой выборку значений случайных величин, то для решения поставленной проблемы следует использовать методы математической статистики. В данном случае такими методами являются статистические критерии проверки гипотез.

**Определение.** *Статистический критерий* – это решающее правило, согласно которому исследователь принимает то или иное решение, основываясь на экспериментальных данных.

Рассмотрим некоторые критерии, предназначенные для проверки гипотезы о наличии тренда в изучаемом динамическом ряду.

#### 3.3.1 Критерий Фостера–Стюарта

Критерий Фостера–Стюарта [5] является одним из наиболее простых, но и наиболее ненадежных критериев проверки присутствия во временном ряде той или иной тенденции. Критерий успешно работает тогда, когда во временном ряду имеется ярко выраженная тенденция монотонного типа,

сопровождаясь случайными отклонениями со сравнительно малой дисперсией и колебаниями со сравнительно малой амплитудой. Прежде чем применять критерий, рекомендуется провести сглаживание временного ряда методом скользящего среднего.

Рассмотрим следующие величины.

$$z_t = \begin{cases} 1, \text{ если } y_t \geq y_{t-1}, \dots, y_t \geq y_1, \\ 0, \text{ в остальных случаях.} \end{cases}$$

Числа  $z_t$  являются индикаторами *точек неубывания*, или *точек роста* временного ряда.

$$l_t = \begin{cases} 1, \text{ если } y_t \leq y_{t-1}, \dots, y_t \leq y_1, \\ 0, \text{ в остальных случаях.} \end{cases}$$

Числа  $l_t$  являются индикаторами *точек невозрастания*, или *точек убыли* временного ряда. Тогда  $z = \sum_{i=2}^n z_i$  и  $l = \sum_{i=2}^n l_i$  – общее количество точек роста и точек убыли соответственно.

Пусть  $d = z - l$ . Очевидно, что  $-(n-1) \leq d \leq n-1$ . Если  $d = n-1$ , то все точки  $y_t$ ,  $t = \overline{2, n}$ , являются точками роста, т. е.  $y_1 \leq y_2 \leq \dots \leq y_n$ , следовательно, в динамическом ряду имеется возрастающий тренд. Если  $d = -(n-1)$ , то  $y_1 \geq y_2 \geq \dots \geq y_n$ , т. е. в динамическом ряду имеется сильный убывающий тренд. Если  $d = 0$ , то  $z = l$ , и тогда ярко выраженной тенденции к росту или убыли нет.

Теоретической основой критерия Фостера–Стюарта служит следующая теорема.

**Теорема.** Случайная величина  $\tau_d = \frac{d}{\sigma}$ , где  $\sigma = \sqrt{2 \sum_{t=2}^n \frac{1}{t}}$ , имеет распределение Стьюдента с  $n$  степенями свободы.

Данная теорема позволяет сформулировать *алгоритм проверки гипотезы о наличии тренда в динамическом ряду*.

1. Выбрать уровень доверия к выводам критерия  $\gamma \in (0; 1)$ .
2. По таблице распределений Стьюдента (Таблица 1) найти критическое число  $t_{n; \alpha}$ , где  $\alpha = 1 - \gamma$  – уровень значимости.
3. По временному ряду вычислить число  $\tau_d$ .
4. Принять решение о наличии или отсутствии тренда согласно следующему правилу:
  - если  $|\tau_d| \leq t_{n; \alpha}$ , то на уровне доверия  $\gamma$  принимаем гипотезу об отсутствии тренда;
  - если  $|\tau_d| > t_{n; \alpha}$ , то на уровне  $\gamma$  отвергаем гипотезу об отсутствии тренда.

При этом, если  $\tau_d > 0$ , то следует предположить, что тренд



возрастающий, если  $\tau_d < 0$  – убывающий.

### 3.3.2 Критерий тренда

Критерий тренда (критерий инверсий) [3, 5] обеспечивает высокую надежность принимаемых решений в случае, когда в динамическом ряду имеется отчетливый монотонный тренд. Его надежность существенно снижается, если динамический ряд содержит закономерные периодические колебания. По сравнению с критерием Фостера–Стюарта, критерий тренда имеет большую устойчивость относительно выбросов данных во временном ряду, особенно если выбросы наблюдались на начальном этапе измерений. При построении критерия тренда используют понятие инверсии данных.

**Определение.** Данные  $y_i$  и  $y_j$  образуют *инверсию*, если  $y_i > y_j$ , но  $i < j$ .

Обозначим  $A(\omega)$  – возможное количество инверсий в динамическом ряду. Если гипотеза об отсутствии тренда в динамическом ряду верна, то случайная величина  $A(\omega)$  имеет известное распределение, представленное Таблицей 2. Это позволяет сформулировать следующий *алгоритм проверки гипотезы о наличии тренда в динамическом ряду*.

1. Выбрать уровень доверия к выводам критерия  $\gamma \in (0;1)$ .
2. По таблице распределения величины  $A(\omega)$  (Таблица 2) найти критические числа  $A_{n;0.5(1+\gamma)}$  и  $A_{n;0.5(1-\gamma)}$ .
3. Вычислить величину  $A$  – количество инверсий во временном ряду.
4. Принять решение о наличии или отсутствии тренда:
  - если  $A_{n;0.5(1+\gamma)} < A < A_{n;0.5(1-\gamma)}$ , то на уровне доверия  $\gamma$  принимаем гипотезу об отсутствии тренда,
  - в противном случае на уровне  $\gamma$  отвергаем гипотезу об отсутствии тренда.

При этом, если  $A \geq A_{n;0.5(1-\gamma)}$ , то, возможно, в динамическом ряду имеются гармонические колебания либо убывающий тренд, если  $A \leq A_{n;0.5(1+\gamma)}$ , то возможен возрастающий тренд.

### 3.3.3 Критерий серий

Критерий серий [3, 5] обеспечивает высокую надежность принимаемых решений в случае, когда в динамическом ряду имеются гармонические колебания, устойчив к выбросам данных во временном ряду.

Прежде чем применять критерий серий, исходные данные необходимо подготовить. Прежде всего, следует учесть, что критерий серий применим только к выборкам с четным количеством измерений. Затем необходимо выбрать опорное число. В качестве опорного числа обычно берут

выборочное среднее либо медиану выборки. Пусть опорным выбрано выборочное среднее  $\bar{y}$ . Поставим в соответствие динамическому ряду последовательность знаков «+» или «-»: если  $y_i < \bar{y}$ , то пишем «-», если  $y_i \geq \bar{y}$ , то «+»,  $i = 1, \dots, n$ .

*Серией* называют последовательность одинаковых знаков, стоящих рядом. Количество знаков в серии называют *длиной серии*. Обозначим  $\rho(\omega)$  – возможное количество серий в динамическом ряду. Если гипотеза об отсутствии тренда в динамическом ряду верна, то случайная величина  $\rho(\omega)$  имеет известное распределение, представленное в Таблице 3. Это позволяет сформулировать *алгоритм проверки гипотезы о наличии тренда в динамическом ряду*.

1. Выбрать уровень доверия к выводам критерия  $\gamma \in (0;1)$ .
2. По таблице распределения величины  $\rho(\omega)$  (Таблица 3) найти критические значения  $r_{0.5n;0.5(1+\gamma)}$  и  $r_{0.5n;0.5(1-\gamma)}$ .
3. Вычислить величину  $r$  – количество серий во временном ряду.
4. Принять решение о наличии или отсутствии тренда:
  - если  $r_{0.5n;0.5(1+\gamma)} < r < r_{0.5n;0.5(1-\gamma)}$ , то на уровне доверия  $\gamma$  принимаем гипотезу об отсутствии тренда,
  - в противном случае на уровне  $\gamma$  отвергаем гипотезу об отсутствии тренда.

### 3.3.4 Механические методы выделения тренда.

#### Сглаживание экспериментальных данных

Если в результате проверки гипотеза о наличии тренда в динамическом ряду подтвердилась, то можно приступать к выбору наиболее подходящей модели тренда. На этом этапе полезно вначале применить механические методы выделения тренда, а затем, проанализировав полученные результаты, перейти к аналитическому построению модели тренда.

Механические методы выделения тренда технически проще и позволяют отчетливее прояснить структуру модели. Из механических методов чаще используют *метод скользящего среднего* (Moving Average, МА(k)) [4, 6]. Суть этого метода состоит в том, что на основе динамического ряда  $y_1, \dots, y_n$  строят новый ряд  $y_1^{(k)}, \dots, y_{n-k+1}^{(k)}$ , где

$$y_i^{(k)} = \frac{1}{k} \sum_{j=0}^{k-1} y_{i+j}, \quad k \in N. \quad \text{В этом случае говорят, что исходный}$$

динамический ряд сглажен с помощью скользящего среднего порядка  $k$ . Величину  $k$  часто называют *величиной окна сглаживания*. Популярность этого метода объясняется тем, что он обладает рядом полезных свойств. Рассмотрим эти свойства.

1. *Влияние сглаживания данных на случайную компоненту*

**Теорема.** Если случайная компонента  $\varepsilon(t, \omega)$  аддитивной модели удовлетворяла условиям Гаусса–Маркова [2–7], то случайная компонента сглаженного динамического ряда  $\varepsilon^{(k)}(t, \omega) = \frac{1}{k} \sum_{j=0}^{k-1} \varepsilon(t+j, \omega)$ ,

$t = 1, 2, \dots, n-k+1$ , представляет собой последовательность зависимых случайных величин, распределенных нормально по закону  $N(0; \sigma^2/k)$  с параметрами распределения, не зависящими от времени.

Таким образом, в результате сглаживания дисперсия динамического ряда уменьшается. Продемонстрируем это.

Пусть  $\eta(1, \omega), \dots, \eta(n, \omega)$  – возможные независимые значения исходного динамического ряда и, согласно условиям Гаусса–Маркова, дисперсия его членов постоянна и равна  $D\eta(i, \omega) = \sigma^2$ . Тогда дисперсия членов сглаженного ряда  $D\eta^{(k)}(i, \omega) = D\left(\frac{1}{k} \sum_{j=0}^{k-1} \eta(i+j, \omega)\right) = \frac{\sigma^2}{k} < \sigma$ . Таким

образом, применение сглаживания снижает случайный разброс данных, выделяя их закономерные тенденции изменения.

## 2. Влияние сглаживания данных на циклическую компоненту

При удачном выборе порядка сглаживания  $k$  в результате сглаживания могут быть полностью удалены закономерные гармонические колебания. Поясним подробнее. Пусть циклические колебания процесса  $\eta(t, \omega)$  задаются единственной элементарной гармоникой с рациональной частотой  $\nu_m = m/N$ ,  $m = 1, \dots, N-1$ , т. е. происходят по гармоническому закону  $A \cos\left(2\pi \frac{m}{N} t\right) + B \sin\left(2\pi \frac{m}{N} t\right)$ . Если выбрать порядок сглаживания

$k = N$ , то значения сумм  $\sum_{i=0}^{N-1} \cos\left(2\pi \frac{m}{N} (i+j)\right)$  и  $\sum_{i=0}^{N-1} \sin\left(2\pi \frac{m}{N} (i+j)\right)$  будут оставаться неизменными при всех целых  $j$ . Это означает, что при таком выборе порядка сглаживания в сглаженном ряду  $\eta^{(k)}(t, \omega)$  закономерные колебания с частотой  $\nu_m$  будут отсутствовать. Динамический ряд, полученный в результате такого сглаживания, будет содержать лишь тренд и случайную компоненту. Полученный на этом этапе результат, к тому же, существенно упрощает процесс построения циклической компоненты, так как явно указывает, на какие подпериоды следует разделить выборку наблюдений.

## 3. Влияние сглаживания данных на форму тренда

Многие из приводимых ниже типичных моделей тренда сохраняют свою структуру при сглаживании скользящим средним.

Продемонстрируем это на некоторых моделях. Пусть  $U(t) = a_0 + a_1 t$ .

Тогда  $U^{(k)}(i) = a_0 + \frac{k-1}{2}a_1 + a_1i$ . Как видим, модель не только осталась линейной, но и сохранила один из коэффициентов  $a_1$ .

Пусть  $U(t) = K + c_0e^{c_1t}$ . Тогда  $U^{(k)}(i) = K + c_0 \frac{1 - e^{kc_1}}{k(1 - e^{c_1})} e^{c_1i}$ . Модель не только осталась экспоненциальной, но и сохранила коэффициенты  $K$  и  $c_1$ .

Итак, если исследователь квалифицированно применил механические методы выделения тренда, он может перейти к аналитическому построению тренда. Предварительно необходимо произвести предварительный отбор функций, которые могут задавать монотонные тенденции исследуемого динамического ряда. Для этого используют признаки наличия тренда определенного типа.

### 3.3.5 Типичные формы тренда

#### 1. Полиномиальный тренд

$$U(t) = a_0 + a_1t + \dots + a_mt^m, \quad m \in N.$$

Особо выделяют *линейный тренд*  $U(t) = a_0 + a_1t$ , типичными для экономических моделей также являются случаи  $m = 2, 3$ .

#### 2. Дробно-линейный тренд

$$U(t) = K + \frac{1}{a_0 + a_1t}.$$

#### 3. Тренд экспоненциального типа

- простая экспонента  $U(t) = c_0e^{c_1t}$ ;
- модифицированная экспонента  $U(t) = K + ab^t$ .

#### 4. Тренд Гомперца

$$U(t) = Ka^{b^t}, \quad a > 0, b > 0, a \neq 1, b \neq 1, K > 0.$$

#### 5. Логистический тренд

$$U(t) = \frac{K}{1 + be^{-at}}, \quad K > 0, a, b > 0.$$

### 3.3.6 Предварительный отбор функций тренда

#### 1. Линейный тренд $U(t) = a_0 + a_1t$

Индикативным признаком линейной функции является равенство  $U'(t) = a_1 = \text{const}$ . Если бы на значения измеряемого показателя не влияли случайные воздействия, то приращения трендовой функции на промежутках времени одинаковой длины оставались бы постоянными. Наличие случайных воздействий не позволяет этим приращениям оставаться постоянными, но ряд значений реальных приращений не должен содержать тренда. Это позволяет сформулировать следующее *правило обнаружения линейного тренда*.

1. Вычисляем величины

$$h_2 = y_2 - y_1, h_3 = y_3 - y_2, \dots, h_n = y_n - y_{n-1}.$$

2. С помощью статистического критерия проверяем гипотезу о наличии тренда в ряду  $h_2, \dots, h_n$ . Если в ряду  $h_2, \dots, h_n$  тренда нет, то ряд  $y_1, \dots, y_n$  имеет линейный тренд, если в ряду  $h_2, \dots, h_n$  тренд есть, то ряд  $y_1, \dots, y_n$  имеет нелинейный тренд.

2. *Полиномиальный тренд*  $U(t) = a_0 + a_1 t + \dots + a_m t^m$

Если проверяется гипотеза о наличии полиномиального тренда  $k$ -го порядка, то следует рассмотреть конечные разности  $k$ -го порядка. Например, при  $k=2$  рассматриваем разности  $q_3 = h_3 - h_2, \dots, q_n = h_n - h_{n-1}$ .

Далее, аналогично с линейным случаем, к ряду  $q_3, \dots, q_n$  применяют статистический критерий для проверки тренда.

3. *Дробно-линейный тренд*

Рассмотрим дробно-линейный тренд вида  $U(t) = \frac{1}{a_0 + a_1 t}$ .

Индикативным признаком указанной функции является равенство  $\left(\frac{1}{U(t)}\right)' = (a_0 + a_1 t)' = a_1 = \text{const}$ .

В этом случае целесообразно проверить на отсутствие тренда ряд

$$h_k = \frac{\frac{1}{y_k} - \frac{1}{y_{k-1}}}{t_k - t_{k-1}}, \quad k = 2, \dots, n.$$

Для проверки гипотезы о наличии дробно-линейного тренда вида  $U(t) = \frac{a_0}{t} + a_1$  целесообразно проверить на отсутствие тренда ряд

$$h_k = \frac{\frac{y_k - y_{k-1}}{1} - \frac{y_{k-1} - y_{k-2}}{1}}{t_k - t_{k-1}}, \quad k = 2, \dots, n.$$

4. *Тренд экспоненциального типа*  $U(t) = e^{a_0 + a_1 t}$

Индикативным признаком указанной функции является равенство  $\frac{U'(t)}{U(t)} = a_1 = \text{const}$ .

В этом случае целесообразно проверить на отсутствие тренда ряд

$$\tau_k = \frac{y_k - y_{k-1}}{y_k(t_k - t_{k-1})}, \quad k = \overline{2, n}.$$

В случае модифицированной экспоненты  $U(t) = K + a b^t$

целесообразно проверить на отсутствие тренда ряд

$$\tau_k = \frac{y_{k+1} - y_k}{y_k - y_{k-1}} \approx \text{const}, k = \overline{2, n-1}.$$

### 5. Кривая Гомперца $U(t) = K a^{b^t}$

Примеры графиков кривой Гомперца при различных сочетаниях значений параметров представлены на рис. 3.3.

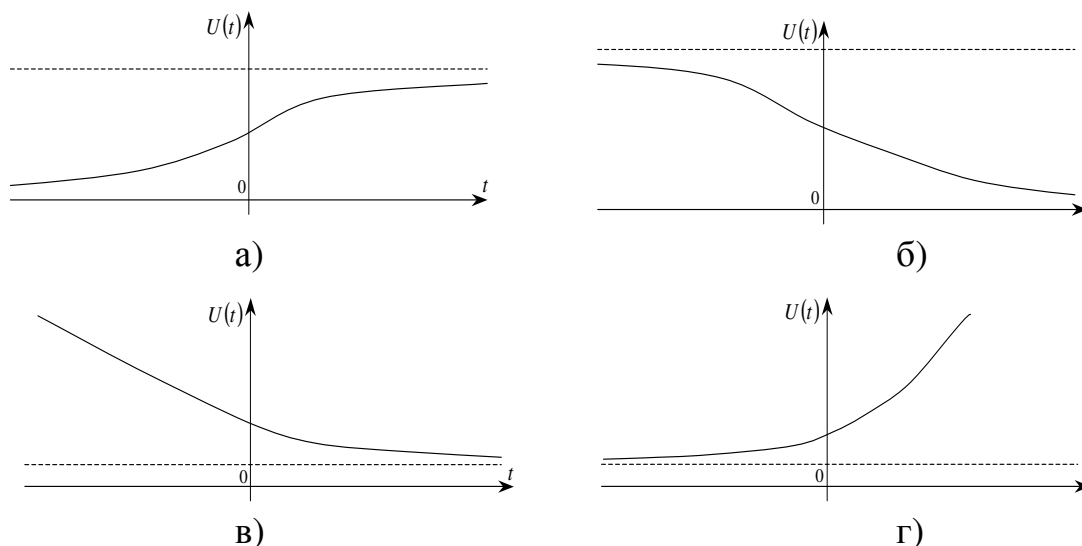


Рис. 3.3 – График кривой Гомперца: а)  $0 < a < 1, 0 < b < 1$ ; б)  $0 < a < 1, b > 1$ ; в)  $a > 1, 0 < b < 1$ ; г)  $a > 1, b > 1$

Признак наличия тренда в форме кривой Гомперца

$$\frac{\ln \frac{y_{k+1}}{y_k}}{\ln \frac{y_k}{y_{k-1}}} = \tau_k \approx \text{const}, k = \overline{2, n-1}.$$

### 6. Логическая кривая $U(t) = \frac{k}{1 + be^{-at}}$

Пример графика этой функции изображен на рис. 3.4.

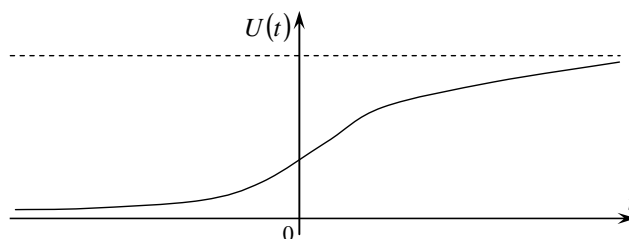


Рис. 3.4 – График логистической кривой

Признак наличия логистического тренда

$$\frac{\frac{1}{y_{k+1}} - \frac{1}{y_k}}{\frac{1}{y_k} - \frac{1}{y_{k-1}}} = \tau_k \approx \text{const}, \quad k = \overline{2, n-1}.$$

Применение указанных методов отбора функций тренда позволяет отобрать несколько функций-претендентов. После этого приступают к аналитическому построению функций.

Основным *аналитическим методом построения тренда* является метод наименьших квадратов.

### 3.3.7 Идея метода наименьших квадратов

Пусть в результате наблюдений получен набор значений  $y_1, \dots, y_n$ . Графически его можно изобразить в виде «облака точек» (рис. 3.5).

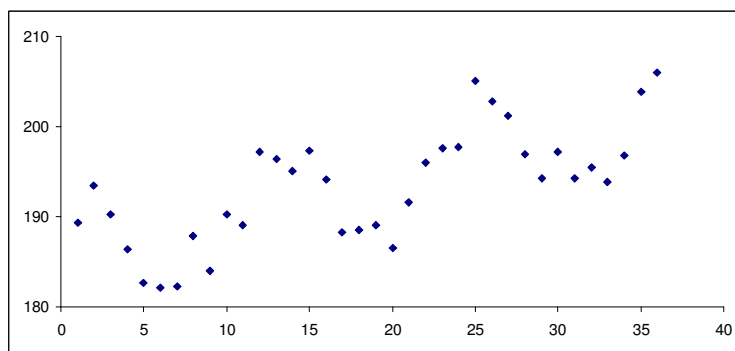


Рис. 3.5 – Точечный график экспериментальных данных

Предполагаемый вид функциональной зависимости исследуемого параметра от времени задается функцией  $y = f(t, a_1, \dots, a_m)$ , которая содержит неизвестные параметры  $a_1, \dots, a_m$ .

Построим функцию  $S^2(a_1, \dots, a_m) = \frac{1}{n} \sum_{i=1}^n (y_i - f(t_i, a_1, \dots, a_m))^2$ .

Найдем точку  $(a_1^*, \dots, a_m^*)$  такую, что  $\min S^2(a_1, \dots, a_m) = S^2(a_1^*, \dots, a_m^*)$ .

Для этого можно воспользоваться необходимым условием экстремума дифференцируемой функции

$$\begin{cases} \frac{\partial S^2}{\partial a_1} = 0, \dots, \frac{\partial S^2}{\partial a_m} = 0, \end{cases}$$

решив эту систему уравнений относительно неизвестных  $a_1, \dots, a_m$ , и отобрать точки минимума из найденных решений системы.

Полученные функции  $a_i^* = a_i^*(t_1, \dots, t_n, y_1, \dots, y_n, n)$ ,  $i = \overline{1, m}$  являются *МНК-оценками* (оценками, полученными методом наименьших квадратов) неизвестных параметров  $a_1, \dots, a_m$ . Эти функции используются для приближенного вычисления значений неизвестных параметров.

### 3.3.8 Свойства МНК-оценок

Основными характеристиками качества статистических оценок являются их несмещенность, состоятельность и эффективность в некотором классе оценок. Напомним определения этих понятий.

**Определение.** Оценка  $a_j^*(y_1, \dots, y_n)$  неизвестного параметра  $a_j$  называется *несмещенной*, если

$$M a_j^*(\eta(t_1, \omega), \dots, \eta(t_n, \omega)) = a_j.$$

Несмещенность оценки означает, что при приближенном вычислении значения неизвестного параметра с помощью данной оценки систематическая ошибка отсутствует.

**Определение.** Оценка  $a_j^*(y_1, \dots, y_n)$  неизвестного параметра  $a_j$  называется *состоятельной*, если

$$\lim_{n \rightarrow \infty} P\{|a_j^*(\eta(t_1, \omega), \dots, \eta(t_n, \omega)) - a_j| > \varepsilon\} = 0, \forall \varepsilon > 0.$$

Состоятельность означает, что точность вычисления  $a_j$  при использовании формулы  $a_j^*(y_1, \dots, y_n)$  возрастает с увеличением  $n$ .

**Определение.** Оценка  $a_j^*(y_1, \dots, y_n)$  неизвестного параметра  $a_j$  называется *эффективной* в некотором классе оценок  $A$ , если она имеет самую маленькую дисперсию в этом классе:

$$D a_j^*(\eta(t_1, \omega), \dots, \eta(t_n, \omega)) \leq D \tilde{a}_j(\eta(t_1, \omega), \dots, \eta(t_n, \omega)), \forall \tilde{a}_j \in A.$$

Обычно классом  $A$  выступает класс линейных оценок, либо класс несмещенных оценок.

В аддитивной модели динамического ряда единственным источником случайных воздействий является случайная компонента. Это означает, что свойства МНК-оценок неизвестных параметров модели напрямую зависят от свойств случайной компоненты. Приводимые ниже условия Гаусса–Маркова обеспечивают высокое качество оценок.

*Условия Гаусса–Маркова*

1.  $\varepsilon(t, \omega)$  является случайной величиной для каждого  $t \geq 0$ .
2.  $M\varepsilon(t, \omega) = 0$  для каждого  $t \geq 0$ .
3.  $D\varepsilon(t, \omega) = \sigma^2 = \text{const}$  для каждого  $t \geq 0$ , т. е. случайные воздействия являются стационарным случайным процессом.
4. Случайные величины  $\varepsilon(t, \omega)$  и  $\varepsilon(s, \omega)$  независимы при всех  $t \neq s$ ,  $t, s \geq 0$ .
5.  $\varepsilon(t, \omega)$  статистически независимы от результатов наблюдений.

**Теорема (о свойствах МНК-оценок).** Если выполнены условия Гаусса–Маркова, то МНК–оценки неизвестных параметров аддитивной модели динамического ряда являются несмещенными, состоятельными, эффективными в классе линейных оценок.

**Замечания**



1. Перечисленные свойства верны независимо от того, каков закон распределения шумов  $\varepsilon(t, \omega)$ .

2. Если дополнительно известно, что шумы имеют нормальное распределение, то оценки эффективны в классе несмещенных оценок.

3. Если из условий Гаусса–Маркова не выполнены 3 и 4, то МНК-оценки остаются несмещенными и состоятельными, но утрачивают эффективность в классе линейных функций.

4. Если не выполнено условие 5, то оценки являются несостоятельными и неэффективными.

### 3.3.9 Применение МНК-оценок для построения тренда

Рассмотрим некоторые аспекты применения МНК-оценок при нахождении коэффициентов отобранных исследователем трендовых моделей.

**Пример.** В качестве примера построения МНК-оценок рассмотрим линейную модель вида  $\eta(t, \omega) = a_0 + a_1 t + \varepsilon(t, \omega)$ .

В этом случае  $S^2(a_0, a_1) = \frac{1}{n} \sum_{i=1}^n (y_i - a_0 - a_1 t_i)^2$ . Найдем точку  $(a_0^*, a_1^*)$  такую, что  $\min S^2(a_0, a_1) = S^2(a_0^*, a_1^*)$ .

В соответствии с необходимым условием экстремума функции, получаем

$$\begin{cases} -\frac{2}{n} \sum_{i=1}^n (y_i - a_0 - a_1 t_i) = 0, \\ -\frac{2}{n} \sum_{i=1}^n (y_i - a_0 - a_1 t_i) t_i = 0. \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^n y_i = a_1 \sum_{i=1}^n t_i + \sum_{i=1}^n a_0, \\ \sum_{i=1}^n y_i t_i = a_1 \sum_{i=1}^n t_i^2 + a_0 \sum_{i=1}^n t_i. \end{cases} \Leftrightarrow \begin{cases} a_0^* = \bar{y} - a_1^* \bar{t}, \\ a_1^* = \frac{\overline{ty} - \bar{t}\bar{y}}{\bar{t}^2 - (\bar{t})^2}. \end{cases}$$

Можно показать, что полученная точка  $(a_0^*, a_1^*)$ , действительно, является точкой минимума функции  $S^2(a_0, a_1)$ .

Для квадратичной модели вида  $\eta(t, \omega) = a_0 + a_1 t + a_2 t^2 + \varepsilon(t, \omega)$  система уравнений для определения оценок  $(a_0^*; a_1^*; a_2^*)$  может быть получена аналогичным образом и имеет вид

$$\begin{cases} a_0 n + a_1 \sum_{i=1}^n t_i + a_2 \sum_{i=1}^n t_i^2 = \sum_{i=1}^n y_i, \\ a_0 \sum_{i=1}^n t_i + a_1 \sum_{i=1}^n t_i^2 + a_2 \sum_{i=1}^n t_i^3 = \sum_{i=1}^n y_i t_i, \\ a_0 \sum_{i=1}^n t_i^2 + a_1 \sum_{i=1}^n t_i^3 + a_2 \sum_{i=1}^n t_i^4 = \sum_{i=1}^n y_i t_i^2. \end{cases}$$

Многие модели (например, дробно-линейные и экспоненциальная), не являющиеся линейными относительно времени, с помощью замены переменных могут быть сведены к линейным (линеаризованы).

Поскольку почти во всех случаях МНК-оценки являются линейными функциями, то получаемые оценки обладают теми же свойствами, что и оценки линейной модели.

Для нелинеаризуемых моделей получены специальные формулы. Одной из таких моделей является модифицированная экспонента  $\eta(t, \omega) = K + ab^t + \varepsilon(t, \omega)$ ,  $K \neq 0$ . МНК-оценки ее параметров  $b, a, K$  задаются формулами Брианта

$$b^* = \frac{(n-1) \sum_{i=1}^{n-1} y_i y_{i+1} - \sum_{i=1}^{n-1} y_i \sum_{i=1}^{n-1} y_{i+1}}{(n-1) \sum_{i=1}^{n-1} y_i^2 - \left( \sum_{i=1}^{n-1} y_i \right)^2},$$

$$a^* = \frac{n \sum_{i=1}^n (b^*)^i y_i - \sum_{i=1}^n (b^*)^i \sum_{i=1}^n y_i}{n \sum_{i=1}^n (b^*)^{2i} - \left( \sum_{i=1}^n (b^*)^i \right)^2},$$

$$K^* = \frac{1}{n} \left( \sum_{i=1}^n y_i - a^* \sum_{i=1}^n (b^*)^i \right).$$

Кривая Гомперца  $\eta(t, \omega) = R e^{-a_0 e^{-a_1 t}} + \varepsilon(t, \omega)$  может быть сведена к модифицированной экспоненте следующим способом:

$$\ln \eta(t, \omega) \approx \ln R - a_0 e^{-a_1 t} + \tilde{\varepsilon}(t, \omega) = K + ab^t + \tilde{\varepsilon}(t, \omega),$$

где  $K = \ln R$ ,  $a = -a_0$ ,  $b = e^{-a_1}$ .

Логистическая кривая  $\eta(t, \omega) = \frac{R}{1 + a_0 e^{-a_1 t}} + \varepsilon(t, \omega)$  также может быть сведена к модифицированной экспоненте.

$$\frac{1}{\eta(t, \omega)} \approx \frac{1}{R} + \frac{a_0}{R} e^{-a_1 t} + \tilde{\varepsilon}(t, \omega) = K + ab^t + \tilde{\varepsilon}(t, \omega),$$

где  $K = \frac{1}{R}$ ,  $a = \frac{a_0}{R}$ ,  $b = e^{-a_1}$ .

В двух последних случаях следует учесть, что возникающие равенства являются приближенными и справедливы лишь тогда, когда случайные помехи имеют малую дисперсию.

При построении полиномиальной модели и модели циклической компоненты полезной оказывается *линейная многофакторная модель регрессии*  $\eta(x_0, x_1, \dots, x_m, \omega) = \sum_{j=0}^m a_j x_j + \varepsilon(\omega)$ .

При построении линейной многофакторной модели проводится  $n$  наблюдений

$$\vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}; X = \begin{pmatrix} x_{10} & x_{11} & \dots & x_{1m} \\ x_{20} & x_{21} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n0} & x_{n1} & \dots & x_{nm} \end{pmatrix}.$$

МНК-оценки  $\vec{a}^* = (a_0^*, a_1^*, \dots, a_m^*)$  параметров этой модели задаются формулой

$$\vec{a}^* = (X^T \cdot X)^{-1} \cdot X^T \cdot \vec{y}.$$

Эта формула задает оценки, которые являются несмещенными, состоятельными, эффективными в классе линейных функций.

Например, модель вида  $\eta(t, \omega) = a_0 + a_1 t + \dots + a_m t^m + \varepsilon(t, \omega)$  сводиться заменой  $z_j = t^j$ ,  $j = \overline{0, m}$ , к линейной многофакторной модели

$$\eta(t, \omega) = \eta(z_0, \dots, z_m, \omega) = a_0 z_0 + a_1 z_1 + \dots + a_m z_m + \varepsilon(t, \omega).$$

### 3.3.10 Оценки точности модели тренда

Для оценки точности модели используют следующие величины.

1. Абсолютная погрешность

$$\Delta = \sqrt{\frac{1}{n-m-1} \sum_{i=1}^n (y_i - f(t_i, a_1^*, \dots, a_m^*))^2}.$$

2. Относительная погрешность

$$\delta = \frac{\Delta}{y} \cdot 100\%.$$

3. Средняя ошибка

$$\varsigma = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - f(t_i, a_1^*, \dots, a_m^*)}{y_i} \right| \cdot 100\%.$$

4. Коэффициент детерминации

$$\hat{R}^2 = 1 - \frac{\Delta^2}{\hat{S}^2(y)}, \text{ где } \hat{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2.$$

### 3.4 Построение циклической компоненты

Напомним, что нашей целью является построений обобщенной аддитивной модели динамического ряда, которая имеет вид

$$\eta(t, \omega) = c(t) + U(t) + V(t) + \varepsilon(t, \omega).$$

На предыдущих этапах рассмотрены методы построения корректирующей компоненты  $c(t)$  и функции тренда  $U(t)$ . Для того, чтобы приступить к построению циклической компоненты  $V(t)$ , необходимо из скорректированных исходных данных удалить тренд. Для этого проделаем следующую операцию – вычтем значения

корректирующей компоненты и тренда из соответствующих значений временного ряда

$$y_1 - Z(t_1) - U(t_1) = \tilde{y}_1,$$

.....

$$y_n - Z(t_n) - U(t_n) = \tilde{y}_n.$$

Если тренд выделен правильно, то новые данные  $\tilde{y}_1, \dots, \tilde{y}_n$  могут содержать лишь закономерные колебания  $V(t)$  и случайные воздействия  $\varepsilon(t, \omega)$ . Наша задача – отделить первые от вторых.

**Замечание.** Для упрощения обозначений далее при использовании новых данных символ волны будет опускаться.

Основным способом аналитического описания закономерных колебаний является представление функции в виде суммы *элементарных гармоник* с различными частотами и амплитудами. Общий вид элементарной гармоник  $V(t) = L \sin(2\pi\nu t + \varphi)$ , где  $\nu$  – положительное число, называемое *частотой* колебания,  $L$  – положительное число, называемое *амплитудой* колебания,  $\varphi \in [0; 2\pi)$  – *начальная фаза* колебания, тогда  $T = \frac{1}{\nu}$  – *период* колебания.

Предложенная форма записи элементарной гармоник удобна тем, что ее параметры имеют ясный физический смысл. Однако, для наших целей удобнее строить элементарную гармоник в виде  $V(t) = A \cos 2\pi\nu t + B \sin 2\pi\nu t$ , а затем приводить ее к указанному стандартному виду, используя формулу  $A \cos \alpha + B \sin \alpha = L \sin(\alpha + \varphi)$ , где  $L = \sqrt{A^2 + B^2}$ ,  $\sin \varphi = \frac{A}{L}$ ,  $\cos \varphi = \frac{B}{L}$ . Таким образом, будем искать  $V(t)$  в виде линейной комбинации элементарных гармоник.

Прежде чем строить циклическую компоненту, необходимо убедиться, что исходные данные подвержены закономерным колебаниям. Обнаружить это визуально, изучая график исходных данных, довольно трудно. Это связано с тем, что зачастую графики экспериментальных данных, зашумленных случайными воздействиями, при наличии закономерных колебаний и их отсутствии плохо различимы. Кроме того, при наличии закономерных колебаний вначале необходимо определить частоты элементарных гармоник, а затем оценивать их амплитуды и начальные фазы. Во многих случаях набор частот элементарных гармоник может быть получен на основе анализа экономического смысла изучаемого процесса.

**Пример.** Пусть  $y_1, y_2, \dots, y_{365}$  представляют собой данные о ежедневных объемах потребления газа населением г. Донецка за весь 2005 год.

Руководствуясь логикой, можно заметить, что для ежедневных данных такого типа характерны колебания с периодом в один год и в одну

неделю. Примем за единицу времени 1 год. Тогда период годовых колебаний  $T=1$  год, а частота  $\nu_1=1$  колеб./год. Соответствующая элементарная гармоника имеет вид  $V_1(t)=L_1 \sin(2\pi t + \varphi_1)$  (рис. 3.6, пунктирная линия). Поскольку максимальное потребление газа приходится на начало и конец года, а минимальное – на середину, то естественно положить  $\varphi_1 = \frac{\pi}{2}$ . График гармоник  $V_1(t)$  изображен на рис. 3.6 сплошной линией. Амплитуда  $L_1$  пока остается неизвестной и ее следует оценить позже.

Теперь учтем недельные колебания. Первым днем недели считаем понедельник. Год содержит 52 полные недели, что составляет 364 дня. Отбросим 365-й день года. Период и частота недельных колебаний равны  $T_2 = \frac{1}{52}$ ,  $\nu_2 = 52$ . Следовательно,  $V_2(t) = L_2 \sin(2\pi \cdot 52t + \varphi_2)$ .

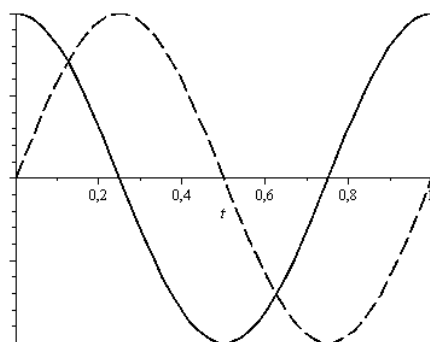


Рис. 3.6 – График гармоник  $V_1(t)$

На рис. 3.7 пунктирной линией изображена одна волна синусоиды с нулевой начальной фазой, т. е.  $V(t) = \sin(2\pi \cdot 52t)$ . Найдем величину начальной фазы  $\varphi_2$ . В недельном цикле максимум потребления приходится на субботу, т. е. на шестой день недели, а максимум пунктирной синусоиды приходится на вторник, т. е. на второй день недели. Значит, чтобы получить синусоиду с максимумом в субботу, нужно пунктирную синусоиду сдвинуть на четыре дня вправо.

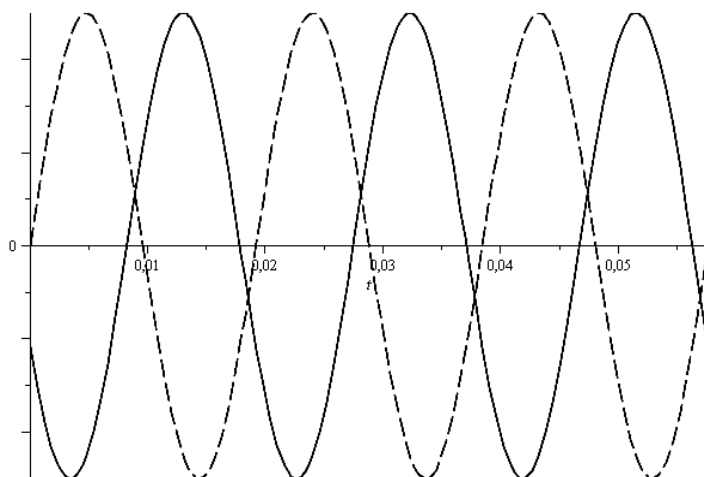


Рис. 3.7 – График гармоник  $V_2(t)$

Однако, первым днем 2005 года было воскресенье, а не понедельник. Значит, преобразованную синусоиду нужно еще сдвинуть на один день влево. В итоге пунктирная синусоида сдвинется на три дня вправо относительно начала года, т. е.  $\varphi_2 = -2\pi \cdot 52 \cdot \frac{3}{364}$  и

$$V_2(t) = L_2 \sin\left(2\pi \cdot 52\left(t - \frac{3}{364}\right)\right) \text{ (см. рис. 3.7).}$$

В итоге модель, которая учитывает обе частоты колебаний, имеет вид  $V(t) = L_1 \sin(2\pi\nu_1 t + \varphi_1) + L_2 \sin(2\pi\nu_2 t + \varphi_2)$ .

Эскиз графика такой функции на временном интервале полгода (для простоты амплитуды выбраны произвольно) изображен на рис. 3.8.

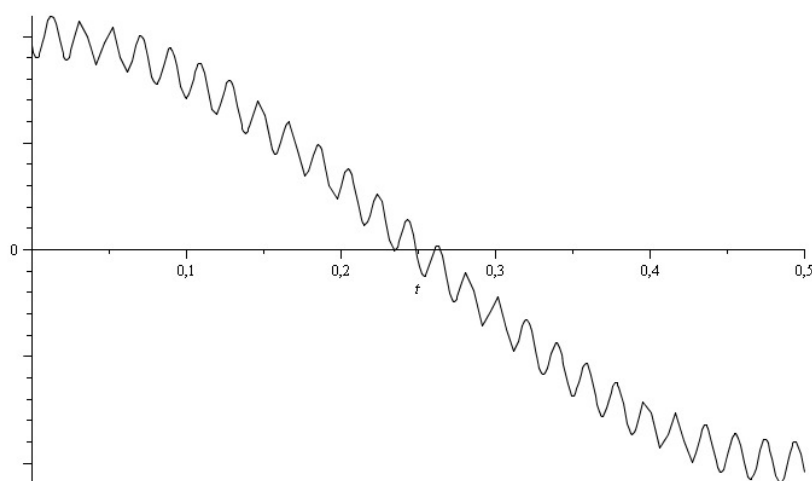


Рис. 3.8 – Эскиз графика циклической компоненты  $V(t)$

Для определения амплитуд  $L_1, L_2$  обозначим  $x_i(t) = \sin(2\pi\nu_i t + \varphi_i)$ ,  $i = 1, 2$ . Тогда мы имеем линейную двухфакторную модель  $V(t) = L_1 x_1(t) + L_2 x_2(t)$ . Коэффициенты этой модели могут быть оценены с помощью МНК-оценок для линейной многофакторной модели.

Как видим, при построении циклической компоненты главную трудность представляет поиск частот элементарных гармоник. Если определить частоты заранее не представляется возможным, то для их выделения используют механические методы.

### 3.4.1 Механический метод построения циклической компоненты

Одним из простейших методов предварительного выделения циклической компоненты является *метод сезонной волны*.

Пусть в наблюдаемом процессе выделены  $M$  временных периодов одинаковой продолжительности (например, годы, кварталы года, месяцы, недели и т. д.). Каждый период состоит из  $N$  измерений. Обозначим  $i$ -тое измерение в  $j$ -том периоде через  $y_i^{(j)}$ ,  $i = 1, \dots, N$ ,  $j = 1, \dots, M$ .

**Определение.** Межпериодными средними динамического ряда  $y_1^{(1)}, \dots, y_N^{(1)}, \dots, y_1^{(M)}, \dots, y_N^{(M)}$  называются величины  $\bar{y}_i = \frac{1}{M} \sum_{j=1}^M y_i^{(j)}$ ,  $i = 1, \dots, N$ .

Сезонной волной называют последовательность межпериодных средних  $\bar{y}_1, \dots, \bar{y}_N$ . Для наглядности сезонную волну изображают графически в виде ломаной, соединяющей точки  $(t_i, \bar{y}_i)$ ,  $i = 1, \dots, N$ .

#### *Порядок построения сезонной волны*

1. В ряду наблюдений выделяем периоды одинаковой длины. Это может быть сделано, например, так, как показано в предыдущем примере. На этом этапе существенную помощь могут оказать результаты механического выделения тренда с помощью скользящего среднего. Напомним, что если порядок сглаживания данных совпадет с периодом колебаний, то колебания исчезают.

2. После выделения периодов временной ряд приобретает вид  $y_1, \dots, y_N, y_{N+1}, \dots, y_{2N}, \dots, y_{N(M-1)+1}, \dots, y_{NM}$  или, что то же самое,  $y_1^{(1)}, \dots, y_N^{(1)}, \dots, y_1^{(M)}, \dots, y_N^{(M)}$ . Усредняем данные из разных периодов  $\bar{y}_i = \frac{1}{M} \sum_{j=0}^{M-1} y_{jN+i} = \frac{1}{M} \sum_{j=1}^M y_i^{(j)}$ ,  $i = 1, \dots, N$ , и получаем сезонную волну.

3. Строим ломаную, соединяющую точки  $(i, \bar{y}_i)$ ,  $i = 1, \dots, N$ . Периодически продолжаем ломаную на все  $M$  периодов. Совмещаем полученный график с графиком исходного динамического ряда. Если графики динамического ряда и сезонной волны ведут себя синхронно, то можно предположить, что исходные данные содержат закономерные периодические колебания с периодом  $N$ . Если же синхронность графиков нарушается хотя бы в одном периоде, то либо в данном ряду периодических колебаний нет, либо неверно выделены базовые периоды.

**Замечание.** Предлагаемый метод построения сезонной волны основан на вычислении межпериодных выборочных средних. Вместо этого можно строить сезонную волну, основываясь на межпериодных медианах исходных данных.

#### *Восполнение пропущенных данных*

При построении циклической компоненты может возникнуть проблема построения сезонной волны, если некоторые измерения пропущены. Поскольку на этом этапе тренд уже удален, то возникает возможность восполнить пропуски во временном ряду. Рассмотрим некоторые методы восполнения пропущенных данных.

1. Пропущенные данные могут быть восполнены средним арифметическим аналогичных значений из других периодов.

2. Восполнение данных можно провести, усредняя предыдущее и следующее значения.

3. Из имеющихся аналогичных данных других периодов наудачу выбирают одно и им восполняют пропуск.

4. Формируют выборку аналогичных данных из других периодов. Согласно условиям Гаусса–Маркова, предположим, что выборка получена из нормального распределения. Вычислим выборочное среднее и исправленную выборочную дисперсию этой выборки. Сгенерируем одно значение нормально распределенной случайной величины с вычисленными параметрами и восполним этим числом пропуск.

Понятно, что искусственное восполнение данных вносит в данные дополнительные помехи и снижает точность результатов. Однако без этого приема выделить циклическую компоненту довольно сложно.

### 3.4.2 Аналитический метод построения циклической компоненты

Если на предыдущем этапе исследования принято решение о наличии циклической компоненты, приступаем к ее построению в аналитической форме. Исходными данными для этого являются числа  $\bar{y}_1, \dots, \bar{y}_N$  – элементы сезонной волны. Будем строить циклическую компоненту в виде частичной суммы ряда Фурье, т. е. в виде суммы элементарных гармоник с неизвестными амплитудами. Дж. Бендат и А. Пирсол предложили следующую форму циклической компоненты:

$$V(t) = A_0 + \sum_{k=1}^{\frac{N}{2}} \left( A_k \cos \frac{2\pi k}{N} t + B_k \sin \frac{2\pi k}{N} t \right).$$

Эта форма легко приводится к линейной многофакторной модели, после чего ее коэффициенты могут быть оценены методом наименьших квадратов. Соответствующие формулы (*формулы Бендата–Пирсола*) имеют вид

$$\begin{aligned} A_0 &= \frac{1}{N} \sum_{i=1}^N \bar{y}_i, \quad A_{\frac{N}{2}} = \frac{1}{N} \sum_{i=1}^N \bar{y}_i \cos \pi i, \\ A_k &= \frac{2}{N} \sum_{i=1}^N \bar{y}_i \cos 2\pi \frac{k}{N} i, \quad k = 1, \dots, \frac{N}{2} - 1, \\ B_k &= \frac{2}{N} \sum_{i=1}^N \bar{y}_i \sin 2\pi \frac{k}{N} i, \quad k = 1, \dots, \frac{N}{2} - 1, \quad B_{\frac{N}{2}} = 0. \end{aligned}$$

Заметим, что эти формулы могут быть применены лишь тогда, когда количество измерений в одном периоде является четным. Если периодом является год или месяц, то это не создает больших трудностей, так как можно считать, что год состоит из 364 дней, а месяц из 30 дней. Такие «урезания» периодов не приводят к заметным потерям информации. Если же периодом является рабочая неделя продолжительностью  $N=5$  дней, то указанные формулы применять нельзя. Преодолеть эту трудность непросто. Кратко опишем два простейших, но не самых надежных пути.



Первый способ предлагает объединить недели в декады, сделав декады новыми периодами, построить новую сезонную волну и к ней применить формулы Бендата–Пирсола. При этом следует помнить, что в действительности периодом является неделя, а не декада. Это означает, что при последующем отборе значимых гармоник следует рассматривать лишь гармоники, период которых не превышает пяти дней.

Второй способ предлагает использовать быстрое преобразование Фурье. В этом случае следует отказаться от использования данных сезонной волны и вернуться к исходным данным  $\tilde{y}_1, \dots, \tilde{y}_n$ . К этим данным применяется быстрое преобразование Фурье. Это можно проделать, например, с помощью пакета прикладных программ MathCad, в котором имеются функции прямого (fft) и обратного (ifft) быстрых преобразований Фурье. В результате такой обработки данных программа сообщает частоты и амплитуды значимых гармоник, содержащихся в исходных данных. Однако описанная процедура требует, чтобы количество измерений равнялось  $2^s$ , где  $s$  – натуральное число. Это создает новые трудности. Например, если имеется 20 данных за четыре пятидневных недели, то придется отбросить либо 4 первых, либо 4 последних измерения, что приводит к потере информации.

### 3.4.3 Методы отбора значимых гармоник

Если при построении циклической компоненты применялись формулы Бендата–Пирсола, то построенное разложение функции  $V(t)$  будет представлять собой сумму элементарных гармоник. Из них нужно отобрать значимые гармоники, т. е. те, которые вносят наиболее существенный вклад в периодические колебания изучаемого показателя. Возможны различные способы осуществления такого отбора. Рассмотрим два из них.

#### 1. Механический (арифметический) метод

Вычисляем амплитуды гармоник  $L_k = \sqrt{A_k^2 + B_k^2}$ ,  $k = 1, \dots, \frac{N}{2}$ . Среди чисел  $L_1, \dots, \frac{L_N}{2}$  выбираем наибольшее и определяем его порядок.

Значимыми будем считать те гармоники, амплитуды которых имеют такой же порядок. Иногда целесообразно считать значимыми и те амплитуды, порядок которых близок к максимальному.

#### 2. Вероятностно-статистический метод

Этот метод основан на проверке гипотезы о значимости гармоник с помощью статистического критерия Фишера.

Обозначим

$$S^2 = \frac{1}{n-N} \left[ \sum_{i=1}^n y_i^2 - nA_0^2 - \frac{n}{2} \sum_{k=1}^{\frac{N}{2}} (A_k^2 + B_k^2) \right],$$

$$F_k = \frac{2S^2}{A_k^2 + B_k^2}, \quad k = 1, \dots, \frac{N}{2}.$$

Теоретической основой для критерия Фишера является теорема, которая гласит: если случайная компонента динамического ряда удовлетворяет условиям Гаусса–Маркова, то распределение возможных значений величины  $F_k$  является распределением Фишера с  $(n-N, 2)$  степенями свободы.

Приведем следующий алгоритм проверки гипотезы о значимости  $k$ -той гармоники.

1. Выбираем уровень доверия  $\gamma \in (0;1)$ .
2. По таблицам распределения Фишера (Таблица 4) с  $(n-N, 2)$  степенями свободы находим критическое число  $F_\gamma(n-N, 2)$ .
3. Вычисляем число  $F_k$ .
4. Принимаем решение:
  - если  $F_k \geq F_\gamma(n-N, 2)$ , то  $k$ -тая гармоника признается незначимой на уровне доверия  $\gamma$ ;
  - если  $F_k < F_\gamma(n-N, 2)$ , то  $k$ -тая гармоника признается значимой на уровне доверия  $\gamma$ .

Таким способом проверяется на значимость каждая гармоника, входящая в разложение Бендата–Пирсола. Незначимые гармоники из состава циклической компоненты удаляются.

На этом процесс построения циклической компоненты завершается.

### 3.5 Проверка остаточной компоненты на адекватность условиям Гаусса–Маркова

После выделения корректирующей компоненты, тренда и циклической компоненты строим *остаточную компоненту*

$$E_1 = y_1 - c(t_1) - U(t_1) - V(t_1), \dots, E_n = y_n - c(t_n) - U(t_n) - V(t_n).$$

Все проведенные выше рассуждения справедливы при условии, что случайная компонента удовлетворяет условиям Гаусса–Маркова. Естественно возникает вопрос: можно ли считать числа  $E_1, \dots, E_n$  значениями случайной величины указанного типа?

Напомним условия Гаусса–Маркова и сформулируем соответствующие вопросы, на которые следует ответить при анализе остаточной компоненты.

1.  $\varepsilon(t, \omega)$  является случайной величиной для каждого  $t \geq 0$ .

Можно ли считать, что полученный ряд остатков  $E_1, \dots, E_n$  является

последовательностью значений некоторой случайной величины, или он ведет себя закономерно?

2. Случайные величины  $\varepsilon(t, \omega)$  при каждом  $t \geq 0$  имеют нормальное распределение.

Можно ли считать, что числа  $E_1, \dots, E_n$  – это результат измерения нормально распределенной случайной величины?

3.  $M(\varepsilon(t, \omega)) = 0$ ,  $D(\varepsilon(t, \omega)) = \sigma^2$ .

Можно ли считать, что  $E_1, \dots, E_n$  – это результаты измерения некоторой случайной величины с нулевым математическим ожиданием и постоянной дисперсией?

4. Случайные величины  $\varepsilon(1, \omega), \dots, \varepsilon(n, \omega)$  – независимы.

Является ли ряд  $E_1, \dots, E_n$  результатом независимых измерений?

Если все перечисленные условия выполнены, то построенная модель считается адекватной условиям Гаусса–Маркова, и применение описанных выше методов является корректным. Если же хотя бы одно из них не подтверждается для выборки  $E_1, \dots, E_n$ , то модель построена не корректно и ее следует скорректировать, либо вовсе от нее отказаться.

Рассмотрим этапы проверки модели на адекватность условиям Гаусса–Маркова.

### 3.5.1 Проверка ряда остатков на случайность

Для проверки ряда остатков  $E_1, \dots, E_n$  на случайность применяют метод Кендела, который также называют *критерием поворотных точек*.

**Определение.** Точка  $E_i$  называется *поворотной*, если  $E_{i-1} < E_i > E_{i+1}$  либо  $E_{i-1} > E_i < E_{i+1}$ ,  $i \in \{2, \dots, n-1\}$ .

Пусть, например, точечный график остатков имеет вид, изображенный на рис. 3.9. Поворотными являются точки с номерами 3, 4, 5. Точка номер 2 не является поворотной. К точкам номер 1 и 6 это понятие не применимо.

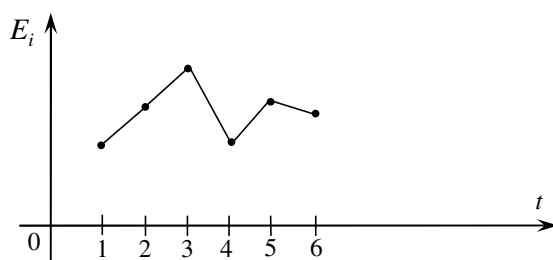


Рис. 3.9 – Точечный график остаточной компоненты

Обозначим через  $\pi_n(\omega)$  случайную величину, равную количеству поворотных точек в ряду возможных значений остаточной компоненты  $E_1(\omega), \dots, E_n(\omega)$ . Р. Кенделом установлено, что

$$M\pi_n(\omega) = \frac{2}{3}(n-2), \quad D\pi_n(\omega) = \frac{16n-29}{90}.$$

Эти равенства позволяют построить доверительный интервал для количества поворотных точек.

Воспользуемся неравенством Чебышева  $P\{|\xi - M\xi| > \varepsilon\} \leq \frac{D\xi}{\varepsilon^2}$ .

Положив  $\xi = \pi_n(\omega)$ , получим

$$P\left\{\left|\pi_n(\omega) - \frac{2}{3}(n-2)\right| > \varepsilon\right\} \leq \frac{1}{\varepsilon^2} \cdot \frac{16n-29}{90};$$

$$P\left\{\left|\pi_n(\omega) - \frac{2}{3}(n-2)\right| \leq \varepsilon \sqrt{D\pi_n(\omega)}\right\} \geq 1 - \frac{1}{\varepsilon^2}.$$

*Алгоритм проверки ряда остатков на случайность*

1. Выбираем уровень доверия  $\gamma \in (0;1)$ .

2. Находим  $\varepsilon$  из равенства  $1 - \frac{1}{\varepsilon^2} = \gamma$ .

3. Строим доверительный интервал

$$\frac{2}{3}(n-2) - \varepsilon \sqrt{\frac{16n-29}{90}} \leq \pi_n(\omega) \leq \frac{2}{3}(n-2) + \varepsilon \sqrt{\frac{16n-29}{90}}.$$

4. Вычисляем  $\hat{\pi}_n$  – количество поворотных точек в ряде  $E_1, \dots, E_n$ .

5. Принимаем решение: если  $\hat{\pi}_n$  принадлежит построенному доверительному интервалу, то ряд  $E_1, \dots, E_n$  признается случайным на уровне доверия  $\gamma$ . В противном случае ряд  $E_1, \dots, E_n$  является неслучайным. В этом случае целесообразно пересмотреть модели тренда и циклической компоненты и повторить проверку для нового ряда остатков.

### 3.5.2 Проверка ряда остатков на соответствие нормальному распределению

Для проверки гипотезы о нормальности распределения ряда остатков может быть использован любой критерий согласия. Мы рассмотрим два из них: критерий асимметрии и эксцесса и RS-критерий.

*Критерий асимметрии и эксцесса*

Выборочные асимметрия и эксцесс определялись нами выше в разделе «Выборочные числовые характеристики». Напомним формулы для их вычисления

$$\bar{A} = \frac{\overline{(x - \bar{x})^3}}{S^3}, \quad \bar{Ex} = \frac{\overline{(x - \bar{x})^4}}{S^4} - 3.$$

Теоретически установлено, что среднеквадратические ошибки выборочной асимметрии и выборочного эксцесса для нормальных выборок равны

$$\varsigma_{\bar{A}} = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}}; \quad \varsigma_{\bar{E}_x} = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}.$$

Доверительный интервал с уровнем доверия  $\gamma = 0,95$  для асимметрии таков:  $|\bar{A}| < 1,5\varsigma_{\bar{A}}$ , для эксцесса:  $|\bar{E}_x| < 1,5\varsigma_{\bar{E}_x}$ . На основе этого получаем следующие правила.

*Правило проверки на значимость асимметрии выборки  $E_1, \dots, E_n$*

- если  $|\bar{A}| < 1,5\varsigma_{\bar{A}}$ , то асимметрия выборки  $E_1, \dots, E_n$  признается незначимой;
- если  $|\bar{A}| > 2\varsigma_{\bar{A}}$ , то асимметрия выборки  $E_1, \dots, E_n$  признается значимой;
- если  $1,5\varsigma_{\bar{A}} \leq |\bar{A}| \leq 2\varsigma_{\bar{A}}$ , то данный критерий однозначного ответа дать не может.

*Правило проверки на значимость эксцесса выборки  $E_1, \dots, E_n$*

- если  $|\bar{E}_x| < 1,5\varsigma_{\bar{E}_x}$ , то эксцесс выборки  $E_1, \dots, E_n$  признается незначимым;
- если  $|\bar{E}_x| > 2\varsigma_{\bar{E}_x}$ , то эксцесс выборки  $E_1, \dots, E_n$  признается значимым;
- если  $1,5\varsigma_{\bar{E}_x} \leq |\bar{E}_x| \leq 2\varsigma_{\bar{E}_x}$ , то данный критерий однозначного ответа дать не может.

Если оба параметра (и асимметрия, и эксцесс) признаются незначимыми, то на уровне доверия  $\gamma = 0,95$  принимается гипотеза о нормальности остаточной компоненты. Если оба параметра признаны значимыми, то следует отказаться от гипотезы о нормальности распределения остаточной компоненты. Если критерий асимметрии и эксцесса не дал ответа, то для уточнения результата можно применить, например, RS-критерий.

#### *RS-критерий*

Статистика критерия имеет вид

$$RS = \frac{E_{\max} - E_{\min}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (E_i - \bar{E})^2}},$$

где  $E_{\max}$ ,  $E_{\min}$  – максимальное и минимальное значения в ряде остатков соответственно,  $\bar{E} = \frac{1}{n} \sum_{i=1}^n E_i$ . Приведем алгоритм проверки

остаточной компоненты на нормальность.

1. Задаем уровень доверия  $\gamma$ .
2. Вычисляем величину  $RS$ .
3. По таблицам критических точек  $RS$ -критерия (Таблица 5) находим величины  $R_n$  и  $R_g$ .
4. Принимаем решение: если  $RS \in (R_n; R_g)$ , то на уровне доверия  $\gamma$  принимается гипотеза о нормальности распределения остаточной компоненты, в противном случае эта гипотеза отвергается.

### 3.5.3 Проверка равенства нулю математического ожидания остаточной компоненты

Для проверки гипотезы о том, что математическое ожидание остаточной компоненты равно нулю, используется тот факт, что величина

$$\hat{t} = \frac{\bar{E} \sqrt{n-1}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (E_i - \bar{E})^2}}$$

имеет распределение Стьюдента с  $n-1$  степенями свободы.

*Алгоритм проверки гипотезы о равенстве нулю математического ожидания остаточной компоненты*

1. Задаем уровень доверия  $\gamma$ .
2. Вычисляем величину  $\hat{t}$ .
3. По таблице критических значений распределения Стьюдента (Таблица 1) находим величину  $t_{n-1; 1-\gamma}$ .
4. Принимаем решение: если  $|\hat{t}| < t_{n-1; 1-\gamma}$ , то на уровне доверия  $\gamma$  принимается гипотеза о равенстве нулю математического ожидания остаточной компоненты, в противном случае эта гипотеза отвергается.

**Замечание.** В случае если указанная гипотеза отвергается, целесообразно ввести величины  $\tilde{E}_i = E_i - \bar{E}$  (новый ряд остатков) и соответственно скорректировать модель тренда.

### 3.5.4 Исследование остаточной компоненты на автокорреляцию

В разделе «Выборочные числовые характеристики» отмечено, что если случайные величины  $\xi$  и  $\eta$  имеют совместное нормальное распределение, то их независимость равносильна некоррелированности. Таким образом, независимость сечений остаточной компоненты равносильна факту отсутствия корреляции между ними. Поскольку имеются измерения  $E_1, \dots, E_n$ , являющиеся значениями одного и того же случайного процесса в различные моменты времени, то следует использовать выборочный коэффициент автокорреляции.

Наряду с выборкой  $E_1, \dots, E_{n-1}$  рассмотрим выборку  $E_2, \dots, E_n$  и определим корреляцию между этими выборками.

**Определение.** Выборочный коэффициент автокорреляции 1-го ранга

для выборки  $E_1, \dots, E_n$  задается формулой

$$k(1) = \frac{\sum_{i=1}^{n-1} E_i E_{i+1} - \frac{1}{n-1} \sum_{i=1}^{n-1} E_i \cdot \sum_{j=1}^{n-1} E_{j+1}}{\left( \left( \sum_{i=1}^{n-1} E_i^2 - \frac{1}{n-1} \left( \sum_{i=1}^{n-1} E_i \right)^2 \right) \left( \sum_{i=1}^{n-1} E_{i+1}^2 - \frac{1}{n-1} \left( \sum_{i=1}^{n-1} E_{i+1} \right)^2 \right) \right)^{1/2}}.$$

Если  $k(1)$  существенно отличен от 0, то это может означать, что каждый последующий член последовательности  $\{E_i\}_{i=1}^n$  зависит от предыдущего.

На практике обычно рассматривается автокорреляция 1-го и 2-го ранга, а высшими рангами пренебрегают.

К сожалению, оценка существенности коэффициента  $k(1)$  затруднена, так как для него построены только асимптотические распределения, т. е. можно получить только приблизительные результаты и при больших объемах выборки.

Избежать этих трудностей можно, если рассматривать выборочный коэффициент циклической автокорреляции, предложенный Андерсеном. Для этого вместе с выборкой  $E_1, E_2, \dots, E_{n-1}, E_n$  рассмотрим выборку  $E_2, E_3, \dots, E_n, E_1$  и определим корреляцию между ними.

**Определение.** Коэффициентом циклической автокорреляции 1-го ранга для выборки  $E_1, \dots, E_n$  называют величину

$$k'(1) = \frac{E_1 E_n + \sum_{i=1}^{n-1} E_i E_{i+1} - n(\bar{E})^2}{\sum_{i=1}^n E_i^2 - n(\bar{E})^2}.$$

Распределение  $k'(1)$  удалось точно построить. Для него составлены статистические таблицы (Таблица 6).

Если в предшествующем исследовании добились того, что  $\bar{E} = 0$ , то коэффициент циклической автокорреляции Андерсена принимает вид

$$k'(1) = \frac{E_1 E_n + \sum_{i=1}^{n-1} E_i E_{i+1}}{\sum_{i=1}^n E_i^2}.$$

*Алгоритм проверки гипотезы об отсутствии автокорреляции с помощью критерия Андерсена*

1. Задаем уровень доверия  $\gamma$ .
2. Вычисляем величину  $k'(1)$ .
3. Если  $k'(1) > 0$ , то берем из статистической таблицы (Таблица 6) значение  $k_+$ , если  $k'(1) < 0$ , то берем значение  $k_-$ .

4. Принимаем решение:

- если  $|k'(1)| < |k_{табл.}|$ , то принимается гипотеза об отсутствии автокорреляции 1-го ранга в ряду  $E_1, \dots, E_n$ ,
- если  $|k'(1)| \geq |k_{табл.}|$ , то считаем, что автокорреляция в ряду  $E_1, \dots, E_n$  присутствует.

Аналогично проверяются гипотезы об отсутствии автокорреляции других рангов.

Наряду с критерием Андерсена рассмотрим еще один критерий – критерий Дарбина–Уотсона. В этом критерии в качестве индикатора наличия автокорреляции предлагается использовать величину

$$DW = \frac{\sum_{i=1}^{n-1} (E_{i+1} - E_i)^2}{\sum_{i=1}^n E_i^2}.$$

При больших  $n$  имеет место соотношение  $\sum_{i=1}^{n-1} E_{i+1}^2 \approx \sum_{i=1}^{n-1} E_i^2$ , что позволяет записать

$$DW \approx 2 \frac{\sum_{i=1}^{n-1} E_i^2 - \sum_{i=1}^{n-1} E_{i+1} E_i}{\sum_{i=1}^n E_i^2} \approx 2(1 - k(1)).$$

Поскольку  $-1 \leq k(1) \leq 1$ , то  $0 \leq DW \leq 4$ . Для величины  $DW$  удалось построить точное распределение.

*Алгоритм проверки гипотезы об отсутствии автокорреляции с помощью критерия Дарбина–Уотсона*

1. Задаем уровень доверия  $\gamma \in (0; 1)$ .
2. Вычисляем выборочное значение  $DW$ .
3. По статистическим таблицам (Таблица 7) находим  $d_n$  и  $d_{в.}$ .
4. Разбиваем отрезок  $[0; 4]$  на подобласти, как показано на рис. 3.10.

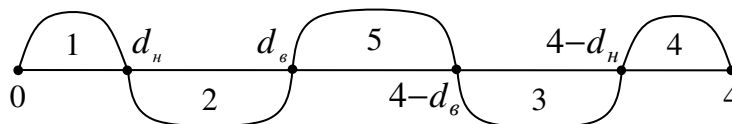


Рис. 3.10 – Критические области критерия Дарбина–Уотсона

5. Наносим точку  $DW$  на отрезок  $[0; 4]$ .
6. Принимаем решение:
  - если  $DW$  попала в область (1), т. е.  $0 \leq DW \leq d_n$ , то признается, что в ряду  $E_1, \dots, E_n$  имеется автокорреляция 1-го ранга и она



положительна;

– если  $DW$  попала в область (4), т. е.  $4 - d_n \leq DW \leq 4$ , то признается, что в ряду  $E_1, \dots, E_n$  имеется автокорреляция 1-го ранга и она отрицательна;

– если  $DW$  попала в область (5), т. е.  $d_g \leq DW \leq 4 - d_g$ , то признается, что в ряду  $E_1, \dots, E_n$  отсутствует автокорреляция 1-го ранга;

– если  $DW$  попала в области (2) или (3), т. е.  $d_n \leq DW \leq d_g$  или  $4 - d_g \leq DW \leq 4 - d_n$ , то критерий Дарбина–Уотсона не может дать ответ; в этом случае для уточнения ответа применяют критерий Андерсена.

Каждый из предложенных критериев имеет свои преимущества и недостатки.

1. Оба критерия являются точными, а не асимптотическими.

2. Вычислительные процедуры критериев достаточно просты.

3. Среди критериев проверки гипотезы об отсутствии автокорреляции критерий  $DW$  обладает наименьшей вероятностью ошибки 2-го рода, при фиксированном уровне доверия. Это преимущество является существенным, поскольку обеспечивает высокую надежность решений, предлагаемых этим критерием. По этой причине критерий Дарбина–Уотсона считается основным критерием в данной области.

4. Критерий Дарбина–Уотсона является рандомизированным, так как имеет область, при попадании в которую критерий не дает ответа. Критерий Андерсена дает однозначные ответы, т. е. является нерандомизированным.

### 3.5.5 Причины возникновения автокорреляции в ряду остатков

Причины возникновения автокорреляции могут быть двух видов: объективные и субъективные. Многим процессам, протекающим в экономике, автокорреляция свойственна от природы. Так, например, цены на биржевых торгах формируются на основе достигнутого уровня цен и тенденции их изменения. Бороться с такой автокорреляцией нет смысла. В этом случае автокорреляцию, как объективное явление, следует учесть при построении модели. Если же автокорреляция возникла вследствие неудачного выбора вида модели, то с ней, как с явлением субъективным, следует бороться. Рассмотрим данную проблему подробнее.

1. Возможно, что неверно сформирован список переменных, влияющих на измеряемый показатель, т. е., кроме зависимости от времени, следует учесть зависимость от еще каких-то факторов. В этом случае задачу следует решать как задачу многофакторного анализа.

2. Возможно, неверно построен тренд. Например, выбрана линейная модель тренда, а в действительности зависимость квадратичная, но с малым старшим коэффициентом. В этом случае необходимо пересмотреть модель тренда.

3. Возможно, при построении циклической компоненты были

отброшены гармоники с малыми амплитудами, которые в действительности процессу присущи. В этом случае следует уточнить список частот закономерных колебаний процесса.

4. Возможно, самому процессу объективно присуща автокорреляционная зависимость, т. е. процесс имеет вид

$$\eta(t+1, \omega) = \rho \eta(t, \omega) + U(t+1) + V(t+1) + \varepsilon(t+1, \omega).$$

В этом случае нужно перестроить модель, приведя ее к указанному виду.

Поскольку условия Гаусса–Маркова требуют независимости значений случайной компоненты, то автокорреляция остатков должна быть устранена, каковы бы ни были причины ее возникновения. Если автокорреляция возникла по субъективным причинам, то следует устранить причины, приведшие к ошибкам при построении модели. Для этого следует построить иной тренд или иную циклическую компоненту. Если из экономической теории следует, что автокорреляция объективно присуща процессу, то ее следует включить в создаваемую модель. Для этого применяют *метод авторегрессионного преобразования*.

### 3.5.6 Устранение автокорреляции остатков методом авторегрессионного преобразования

Многие экономические закономерности описываются уравнениями вида

$$\begin{aligned} h^{(d)}(t) + \rho^{(1)} \cdot h^{(d)}(t-1) + \dots + \rho^{(p)} \cdot h^{(d)}(t-p) = \\ = \varepsilon(t, \omega) + \theta^{(1)} \varepsilon(t-1, \omega) + \dots + \theta^{(q)} \varepsilon(t-q, \omega), \end{aligned}$$

где  $\eta(1, \omega), \dots, \eta(n, \omega)$  – исходный временной ряд;  $h^{(1)}(t) = \eta(t+1) - \eta(t)$  – разности 1-го порядка;  $h^{(d)}(t) = h^{(d-1)}(t+1) - h^{(d-1)}(t)$  – разности  $d$ -го порядка,  $d > 1, d \in N$ ;  $\varepsilon(t, \omega)$  – случайные воздействия;  $q$  – порядок скользящего среднего.

Модели такого типа сокращенно обозначают  $ARIMA(p, d, q)$  – *авторегрессионное интегрированное скользящее среднее* (AutoRegressive Integrated Moving Average).

Подробное описание теории исследования таких моделей изложено в книге Дж. Бокса и Г. Дженкинса [4]. Как видим, модели типа  $ARIMA$  описывают зависимости рекуррентного типа и не предполагают наличия монотонных трендов и циклических компонент. Рассмотрим иную модель, содержащую и авторегрессию первого ранга, и тренд, и циклическую компоненту.

Изучим процесс построения модели авторегрессии в случае, когда закон изменения показателя имеет вид

$$\eta(t+1, \omega) = \rho \eta(t, \omega) + U(t+1) + V(t+1) + \varepsilon(t+1, \omega).$$

Предположим, что в ряду остатков обнаружена автокорреляция ранга 1, т. е.

$$E(t) = \rho^{(1)} E(t-1), \quad t = 2, \dots, n,$$

где  $\rho^{(1)}$  – коэффициент автокорреляции 1-го ранга. Оценка этого коэффициента может быть построена методом наименьших квадратов

$$\begin{aligned} S(\rho^{(1)}) &= \frac{1}{n-1} \sum_{t=2}^n (E(t) - \rho^{(1)} E(t-1))^2 \rightarrow \min, \\ \frac{dS}{d\rho^{(1)}} &= -\frac{2}{n-1} \sum_{t=2}^n E(t-1) (E(t) - \rho^{(1)} E(t-1)) = 0, \\ \sum_{t=2}^n E^2(t-1) \cdot \rho^{(1)} &= \sum_{t=2}^n E(t) \cdot E(t-1), \\ (\rho^{(1)})^* &= \frac{\sum_{t=2}^n E(t) \cdot E(t-1)}{\sum_{t=2}^n E^2(t-1)}. \end{aligned}$$

Полученная оценка является несмещенной и состоятельной. Тогда можно утверждать, что

$$E(t) \approx (\rho^{(1)})^* E(t-1) + \tilde{E}(t).$$

В результате устранения автокорреляции в ряду остатков получим новый ряд остатков  $\tilde{E}_2, \dots, \tilde{E}_n$ , для которого проверяем гипотезу о наличии автокорреляции и в случае ее наличия повторяем описанный процесс для нового ряда остатков. Этот процесс повторяется  $p$  раз.

Пусть авторегрессия имеет первый ранг. Как это учесть в модели? Вначале, для наглядности и простоты, рассмотрим линейную модель.

Предположим, что некоторый показатель обладает линейным трендом  $U(t) = a_0^* + a_1^* t$ , а циклическая компонента отсутствует,  $V(t) \equiv 0$ . Тогда модель изменения показателя имеет вид  $y(t) = a_0^* + a_1^* t + E(t)$ .

Поскольку  $y(t-1) = a_0^* + a_1^* (t-1) + E(t-1)$ , то, вычитая из первого равенства второе, умноженное на  $(\rho^{(1)})^*$ , получаем равенство

$$\begin{aligned} y(t) - (\rho^{(1)})^* y(t-1) &= \\ &= a_0^* \left(1 - (\rho^{(1)})^*\right) + a_1^* \left(t - (\rho^{(1)})^* (t-1)\right) + E(t) - (\rho^{(1)})^* E(t-1) \end{aligned}$$

Обозначим  $E(t) - (\rho^{(1)})^* E(t-1) = \tilde{E}(t)$ , тогда модель принимает вид

$$y(t) = (\rho^{(1)})^* y(t-1) + a_0^* \left(1 - (\rho^{(1)})^*\right) + a_1^* \left(1 - (\rho^{(1)})^*\right) t + a_1^* (\rho^{(1)})^* + \tilde{E}(t),$$

где новые остатки  $\tilde{E}_1, \dots, \tilde{E}_{n-1}$  некоррелированы.

Отметим, что новая модель учитывает авторегрессию исходных

данных и, при этом, сохранила тип тренда – линейный.

Теперь рассмотрим произвольные  $U(t)$  и  $V(t)$ . Пусть модель имеет вид  $y(t) = U(t) + V(t) + E(t)$ , где в ряду  $E_1, \dots, E_n$  имеется автокорреляция 1-го порядка. Умножим равенство  $y(t-1) = U(t-1) + V(t-1) + E(t-1)$  на  $(\rho^{(1)})^*$  и вычтем из первоначального.

$$y(t) - (\rho^{(1)})^* y(t-1) = (U(t) - (\rho^{(1)})^* U(t-1)) + (V(t) - (\rho^{(1)})^* V(t-1)) + \underbrace{(E(t) - (\rho^{(1)})^* E(t-1))}_{\tilde{E}(t)}.$$

Получим общий вид модели

$$y(t) = (\rho^{(1)})^* y(t-1) + U(t) - (\rho^{(1)})^* U(t-1) + V(t) - (\rho^{(1)})^* V(t-1) + \tilde{E}(t),$$

где  $\tilde{E}(t)$  – некоррелированные остатки.

**Замечание.** Рассмотренный метод построения модели авторегрессии отличается простотой. Если модель содержит только авторегрессию и не содержит ни тренда, ни циклической компоненты, то построенная модель обладает хорошими свойствами. Если же модель содержит, например, тренд, то свойства оценок коэффициентов такой модели нам неизвестны. Это является следствием того, что при устранении автокорреляции остатков возникают произведения оценок коэффициентов тренда на оценку коэффициента автокорреляции. Хотя каждая из этих оценок обладает высокими качествами, сказать что-либо об их произведении весьма трудно.

### 3.5.7 Робастность методов анализа временных рядов

При построении обобщенной аддитивной модели динамического ряда предполагалась справедливость условий Гаусса–Маркова. Но эти условия, в свою очередь, представляют собой математические модели реальных природных явлений и описывают их лишь приближенно. Следовательно, перед применением разработанных методов необходимо выяснить, как велики могут быть отклонения от этих условий, чтобы применение методов не приводило к существенно неверным результатам. Таким образом, возникает проблема устойчивости методов относительно невыполнения условий их применения, т. е. проблема *робастности* (robustness).

В реальной практике статистических исследований на стадии подготовки исходных данных может осуществляться устранение выбросов в выборке. На рис. 3.11 изображена выборка, в которой предпоследнее измерение  $y_5$  можно считать выбросом, так как оно заметно отличается от других.

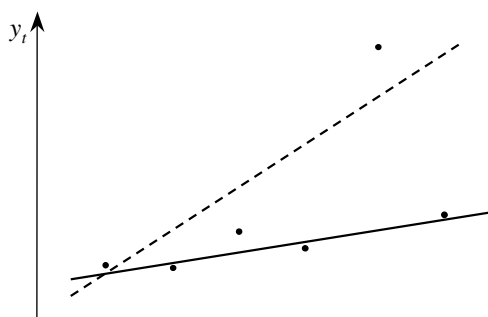


Рис. 3.11 – Пример выборки, имеющей выброс

Поскольку предполагается, что случайная компонента распределена нормально, то наличие выбросов в выборке является естественным. Однако учет выброса может существенно снизить точность модели.

**Пример.** Пусть имеется следующий ряд наблюдений:

–2,9; 2; –1,8; –1,7; 1,9; 1,5; –1,2; 1,1; –0,01; 0,02.

Выборочное среднее ряда равно  $\bar{y} = -0,109$ .

Предположим, что в последнем измерении вместо 0,02 получили бы 6,0, тогда выборка примет вид –2,9; 2; –1,8; –1,7; 1,9; 1,5; –1,2; 1,1; –0,01; 6,0. Тогда выборочное среднее равнялось бы  $\tilde{y} = 0,489 \approx -4,5 \cdot \bar{y}$ . Как видим, наличие всего одного выброса может существенно влиять на числовые характеристики выборки, а следовательно, и на характер модели (рис. 3.11).

Разработаны следующие методы борьбы с выбросами: *усечение выборки* и *винзорирование выборки*. Самым простым методом борьбы с выбросами является усечение – отсечение (удаление) выбросов. Например, из выборки удаляют самое большое и самое маленькое число. При таком подходе объем выборки уменьшается и теряется полезная информация.

При винзорировании выборки предлагается действовать более мягко – не просто удалять наблюдения, а заменять их более подходящими. Если удаляются  $k$  самых больших значений и  $k$  самых маленьких, то вместо них в выборку включают соответственно самое большое из оставшихся и самое маленькое из оставшихся.

Для второй выборки приведенного выше примера заменим соответственно –2,9 на –1,8 и 6 на 2. В результате для винзорированной выборки  $\bar{y} = 0,199$ , что близко к среднему первой выборки. Однако, если первоначальная выборка имела нормальное распределение, то винзорированная выборка уже не обязана быть нормально распределенной. Рассмотрим, к каким последствиям может привести отсутствие нормального распределения.

### 3.5.8 Свойства оценок при нарушении нормальности

Пусть истинная зависимость изучаемого показателя от времени имеет вид  $\eta(t, \omega) = \sum_{k=1}^m a_k x_k(t) + \varepsilon(t, \omega)$ , где  $x_1(t), \dots, x_m(t)$  – известные функции.

Предположим, что помехи  $\varepsilon(t, \omega)$  распределены нормально.

Обозначим  $\vec{x} = (x_1(t), \dots, x_m(t))$ ,  $\vec{y} = (y_1, \dots, y_n)$  – ряд экспериментальных данных,  $\vec{a} = (a_1, \dots, a_m)$  – параметры (коэффициенты) модели. Будем считать  $t_i = i$ ,  $i = \overline{1, n}$ . Сформируем матрицу значений функций  $x_1(t), \dots, x_m(t)$  в моменты времени  $1, 2, \dots, n$ .

$$X = \begin{pmatrix} x_1(1) & x_1(2) & \dots & x_1(n) \\ x_2(1) & x_2(2) & \dots & x_2(n) \\ \dots & \dots & \dots & \dots \\ x_m(1) & x_m(2) & \dots & x_m(n) \end{pmatrix}.$$

Тогда  $\eta(t, \omega) = (\vec{a}, \vec{x}(t)) + \varepsilon(t, \omega)$ . Вектор оценок коэффициентов, согласно методу наименьших квадратов, имеет вид

$$\vec{a}^* = \vec{y} \cdot X^T (X \cdot X^T)^{-1}.$$

Погрешность модели

$$(\sigma^2)^* = \frac{1}{n-m} \sum_{i=1}^n \left( y_i - \sum_{k=1}^m a_k^* x_k(i) \right)^2 = \frac{1}{n-m} \left| \vec{y} - \vec{a}^* \vec{x} \right|^2.$$

Формулы для  $a^*$  и  $(\sigma^2)^*$  получены методом наименьших квадратов, следовательно, оценки являются несмещенными, состоятельными и эффективными в классе несмещенных. Эти свойства гарантированы, если случайная компонента удовлетворяет условиям Гаусса–Маркова. Более того, справедлива следующая теорема [3].

**Теорема.** Если  $\varepsilon(t, \omega)$  есть последовательность независимых случайных величин, имеющих  $N(0; \sigma^2)$ –распределение, то

- 1)  $(\vec{a}^* - \vec{a})\sqrt{D} \in N(\vec{0}, \sigma^2 E)$ , где  $D = X \cdot X^T$ ;
- 2)  $\frac{(n-m)(\sigma^2)^*}{\sigma^2} \sim \chi_{n-m}^2$ ;
- 3) случайные величины  $a_1^*, \dots, a_m^*$  и  $(\sigma^2)^*$  независимы;
- 4) величина  $\frac{|\vec{a}^* - \vec{a}| \cdot X|^2}{|\vec{y} - \vec{a}^* \cdot X|^2}$  имеет распределение Фишера с  $(m; n-m)$

степенями свободы;

- 5) величина  $\frac{(\vec{a}^* - \vec{a})\sqrt{D}}{\sqrt{(\sigma^2)^*}}$  имеет распределение Стюдента со степенью

свободы  $n-m$ ;

- 6) величина  $\frac{y^*(t) - y(t)}{d(t) \cdot (\sigma^2)^*}$  имеет распределение Стюдента со степенью

свободы  $n-m$ , где  $y^*(t) = \sum_{k=1}^m a_k^* x_k(t)$ ,  $y(t) = \sum_{k=1}^m a_k x_k(t)$ ,

$$d^2(t) = \vec{x}(t) \cdot D^{-1} \cdot \vec{x}^T(t).$$

Утверждения теоремы играют важную роль в приложениях, так как позволяют строить доверительные интервалы для параметров модели. Из пункта б), в частности, следует алгоритм построения прогноза временного ряда, который будет изложен ниже.

Утверждения теоремы обладают робастностью. Если требование о нормальности  $\varepsilon(t, \omega)$  не выполнено, то утверждения теоремы в общем случае неверны.

В этом случае оценки  $\vec{a}^*$ ,  $(\sigma^2)^*$  остаются несмещенными, но не будут эффективными в классе несмещенных оценок.

В утверждении 3) оценки будут некоррелированными, но их независимость не гарантируется.

Пункты 1) и 2) теоремы могут быть заменены следующими асимптотическими соотношениями

$$1') \frac{(\vec{a}^* - \vec{a})\sqrt{D}}{\sigma} \xrightarrow{n \rightarrow \infty} N(\vec{0}, E);$$

$$2') \frac{\sqrt{n-m} \cdot ((\sigma^2)^* - 1)}{\sqrt{2} \cdot \sigma^2} \xrightarrow{n \rightarrow \infty} N(0; 1).$$

Таким образом, при отсутствии нормальности выборки для выборок большого объема выборки утверждения теоремы могут быть аппроксимированы приведенными асимптотическими соотношениями.

### 3.6 Построение прогнозов временного ряда

Если построенная модель динамического ряда является высокоточной и адекватна условиям Гаусса–Маркова, то ее можно использовать не только для изучения формы зависимости данных, но и для построения прогноза поведения ряда.

**Замечание.** Использование моделей временных рядов, построенных описанными выше методами, для построения прогнозов, возможно лишь при выполнении следующих условий:

1) характер зависимости показателя  $\eta(t, \omega)$  от времени и случая не будет меняться, по крайней мере, до того момента времени, в который мы хотим построить прогноз;

2) построенная модель  $y(t)$  является высокоточной и высокоадекватной;

3) изучаемая закономерность изменения показателя является случайной функцией только от времени и не зависит от других переменных.

Прогнозы временных рядов могут быть отнесены к двум классам: *интерполяционные* и *экстраполяционные* прогнозы.

**Определение.** Пусть  $[t_1; t_n]$  – отрезок наблюдения показателя  $\eta(t, \omega)$ ,  $t$  – момент времени, для которого необходимо оценить возможное

значение  $\eta(t, \omega)$ ,  $t \neq t_i$ ,  $\forall i = \overline{1, n}$ . Если  $t \in [t_1, t_n]$ , то прогноз называется *интерполяционным*; если  $t \notin [t_1, t_n]$  – *экстраполяционным*.

В большинстве случаев исследователя интересует экстраполяция в случае  $t > t_n$ .

*Постановка задачи прогнозирования поведения динамического ряда*

Предполагается, что значения измеряемого показателя изменяются по закону  $\eta(t, \omega) = U(t) + V(t) + \varepsilon(t, \omega)$ . Имеются измерения значений динамического ряда  $y_1, \dots, y_n$ . Выбран некоторый уровень доверия  $\gamma \in (0; 1)$ . Для заданного момента времени  $t$  необходимо построить доверительный интервал уровня  $\gamma$ , центром которого является  $M\eta(t, \omega)$  – ожидаемое среднее значение показателя  $\eta(t, \omega)$ .

**Определение.** Под *прогнозом* уровня  $\gamma \in (0; 1)$  значения временного ряда в момент времени  $t$ , не совпадающий с моментами измерений, будем понимать доверительный интервал  $(\check{\theta}(t, \omega); \hat{\theta}(t, \omega))$  такой, что

$$P\{\check{\theta}(t, \omega) < M\eta(t, \omega) < \hat{\theta}(t, \omega)\} = \gamma.$$

**Замечание.** Границами доверительного интервала являются случайные величины  $\check{\theta}(t, \omega) = \check{\theta}(t, \eta(1, \omega), \dots, \eta(n, \omega))$  и  $\hat{\theta}(t, \omega) = \hat{\theta}(t, \eta(1, \omega), \dots, \eta(n, \omega))$ . Чтобы построить прогноз в виде числового интервала, необходимо вычислить их значения по имеющейся выборке измерений  $\check{\theta}(t) = \check{\theta}(t, y_1, \dots, y_n)$  и  $\hat{\theta}(t) = \hat{\theta}(t, y_1, \dots, y_n)$ .

**Определение.** Величина  $\Delta(t) = 0.5(\hat{\theta}(t) - \check{\theta}(t))$ , равная половине ширины прогнозного интервала, называется *абсолютной погрешностью прогноза*, а величина  $\delta = \frac{\Delta(t)}{y(t)} \cdot 100\%$  – *относительной погрешностью прогноза*.

*Построение прогноза для модели вида* 
$$\eta(t, \omega) = \sum_{k=1}^m a_k x_k(t) + \varepsilon(t, \omega)$$

В данном случае  $x_k(t)$ ,  $k = \overline{1, m}$ , – точно известные функции,  $t = 1, \dots, n$ ,  $\varepsilon(1, \omega), \dots, \varepsilon(n, \omega)$  – случайные величины, удовлетворяющие условиям Гаусса–Маркова.

Произведя измерения исследуемого показателя, получили выборку измерений  $\vec{y} = (y_1, \dots, y_n)$ , где  $y_1 = \eta(1, \omega_0)$  – результат измерения в момент  $t = 1$ , ...,  $y_n = \eta(n, \omega_0)$  – результат измерения в момент  $t = n$ . Поскольку рассматриваемая модель является линейной функцией от регрессоров, то оценки ее параметров могут быть получены с помощью МНК-формул для многофакторной линейной модели. Вектор оценок коэффициентов модели имеет вид

$$\vec{a}^*(y_1, \dots, y_n) = (a_1^*(y_1, \dots, y_n), \dots, a_m^*(y_1, \dots, y_n)) = (y_1, \dots, y_n) \cdot X^T (X X^T)^{-1},$$

где  $X$  – матрица значений регрессоров.



Оценка дисперсии шумов задается формулой

$$(\sigma^*)^2 = \frac{1}{n-m} \sum_{i=1}^n \left( y_i - \sum_{k=1}^m a_k^*(y_1, \dots, y_n) x_k(i) \right)^2.$$

Обе оценки являются несмещенными, состоятельными и эффективными в классе несмещенных. Тогда модельная функция имеет вид

$$y(t) = \sum_{k=1}^m a_k^*(y_1, \dots, y_n) x_k(t).$$

Чтобы использовать эту модельную функцию для аппроксимации поведения изучаемого показателя, следует учесть возможные воздействия случайных помех. В этом случае в качестве имитатора реального процесса  $\eta(t, \omega)$  можно рассмотреть модельный процесс

$$y(t, \omega) = \sum_{k=1}^m a_k^*(\eta(1, \omega), \dots, \eta(n, \omega)) \cdot x_k(t) + \varepsilon(t, \omega) \approx \eta(t, \omega).$$

1) в силу несмещенности оценок  $a_k^*$  справедливо равенство

$$\begin{aligned} My(t, \omega) &= \sum_{k=1}^m Ma_k^*(\eta(1, \omega), \dots, \eta(n, \omega)) \cdot x_k(t) + M\varepsilon(t, \omega) = \\ &= \sum_{k=1}^m a_k x_k(t) + 0 = M\eta(t, \omega). \end{aligned}$$

Следовательно, модель  $y(t, \omega)$  является несмещенной оценкой  $\eta(t, \omega)$ .

2) в силу сильной состоятельности МНК-оценок

$$a_k^*(\eta(1, \omega), \dots, \eta(n, \omega)) \xrightarrow[n \rightarrow \infty]{\text{с вероятностью 1}} a_k,$$

т. е.  $a_k^* \approx a_k$  при больших  $n$ . Следовательно, модель является состоятельной оценкой  $\eta(t, \omega)$ .

Обратим внимание на то, что функции  $a_k^*(y_1, \dots, y_n)$  являются линейными функциями по переменным  $y_1, \dots, y_n$ . Подставив их в  $y(t, \omega)$ , получим

$$y(t, \omega) = \vec{\eta}(t, \omega) \cdot X^T (XX^T)^{-1} \cdot \vec{x}^T(t) + \varepsilon(t, \omega) = \sum_{i=1}^n \alpha_i(t) \cdot \eta(i, \omega) + \varepsilon(t, \omega),$$

$$\text{где } \vec{\alpha}(t) = X^T (XX^T)^{-1} \cdot \vec{x}^T(t).$$

Поскольку предполагается, что случайные воздействия являются независимыми нормально распределенными случайными величинами, то

$\eta(t, \omega) \in N\left(\sum_{k=1}^m a_k x_k(t); \sigma^2\right)$  и величины  $\eta(1, \omega), \dots, \eta(n, \omega)$  являются

независимыми случайными величинами. Так как момент прогноза  $t$  не совпадает с моментами измерений, то  $\varepsilon(t, \omega)$  не зависит от

$\eta(1, \omega), \dots, \eta(n, \omega)$ . Из сказанного следует, что  $y(t, \omega) \in N\left(\sum_{k=1}^m a_k x_k(t); D\right)$ .

Найдем  $D$  – дисперсию прогнозного значения.

$$\begin{aligned} D = Dy(t, \omega) &= \sum_{i=1}^n \alpha_i^2(t) D\eta(t, \omega) + D\varepsilon(t, \omega) = \\ &= \sum_{i=1}^n \alpha_i^2(t) \sigma^2 + \sigma^2 = \sigma^2 \left(1 + \sum_{i=1}^n \alpha_i^2(t)\right). \end{aligned}$$

Дисперсия шумов  $\sigma^2$  неизвестна, и поэтому вместо нее используем ее оценку  $(\sigma^*)^2$ . В этом случае для построения доверительного интервала прогноза применяют следующую теорему [3].

**Теорема.** Величина  $\tau_{n-m}(\omega)$  имеет распределение Стьюдента с  $n-m$

степенями свободы, где  $\tau_{n-m}(\omega) = \frac{y(t, \omega) - \sum_{k=1}^m a_k x_k(t)}{\sqrt{(\sigma^*)^2 \cdot \left(1 + \sum_{i=1}^n \alpha_i^2(t)\right)}}$ .

Эта теорема позволяет сформулировать следующий алгоритм построения интервала прогноза.

1. Задаем уровень доверия  $\gamma \in (0; 1)$ .
2. По таблицам распределения Стьюдента (Таблица 1) находим  $t_{n-m, \gamma}$ .
3. Строим интервал прогноза

$$\tilde{\theta}(t) = \sum_{k=1}^m a_k^*(y_1, \dots, y_n) x_k(t) - t_{\gamma, n-m} \cdot z,$$

$$\hat{\theta}(t) = \sum_{k=1}^m a_k^*(y_1, \dots, y_n) x_k(t) + t_{\gamma, n-m} \cdot z,$$

где  $z = \sigma^*(y_1, \dots, y_n) \sqrt{1 + \sum_{i=1}^n \alpha_i^2(t)}$ .

Работу алгоритма продемонстрируем на примере линейной модели.

**Пример.** Пусть  $\eta(t, \omega) = a_0 + a_1 t + \varepsilon(t, \omega)$ . Здесь  $x_1(t) \equiv 1$ ,  $x_2(t) = t$ .

Построив МНК-оценки коэффициентов  $a_0$  и  $a_1$ , получим модельную

функцию  $y(t) = \bar{y} + \frac{\sum_{i=1}^n (t_i - \bar{t})(y_i - \bar{y})}{\sum_{i=1}^n (t_i - \bar{t})^2} \cdot (t - \bar{t})$ . Перегруппируем слагаемые и

представим эту функцию в виде линейной комбинации значений  $y_1, \dots, y_n$ .

В результате получим  $y(t) = \sum_{i=1}^n \alpha_i(t) y_i$ , где  $\alpha_i(t) = \frac{1}{n} + \frac{(t-\bar{t})(t_i-\bar{t})}{\sum_{i=1}^n (t_i-\bar{t})^2}$ .

Тогда  $1 + \sum_{i=1}^n \alpha_i^2(t) = \frac{n+1}{n} + \frac{(t-\bar{t})^2}{\sum_{i=1}^n (t_i-\bar{t})^2}$ , и интервал прогноза имеет вид

$$\left( y(t) - t_{\gamma, n-m} \sigma^* \sqrt{\frac{n+1}{n} + \frac{(t-\bar{t})^2}{\sum_{i=1}^n (t_i-\bar{t})^2}}; y(t) + t_{\gamma, n-m} \sigma^* \sqrt{\frac{n+1}{n} + \frac{(t-\bar{t})^2}{\sum_{i=1}^n (t_i-\bar{t})^2}} \right).$$

Отметим, что в случае линейной модели половина ширины интервала прогноза равна  $\Delta(t) = t_{\gamma, n-m} \sigma^*(y_1, \dots, y_n) \sqrt{\frac{n+1}{n} + \frac{(t-\bar{t})^2}{\sum_{i=1}^n (t_i-\bar{t})^2}}$  и задает

абсолютную погрешность прогноза. Минимально возможная погрешность достигается при  $t = \bar{t}$ . При отдалении точки прогноза от точки  $\bar{t}$  погрешность растет. На рис. 3.12 пунктирной линией изображены границы доверительных интервалов для различных моментов прогноза.

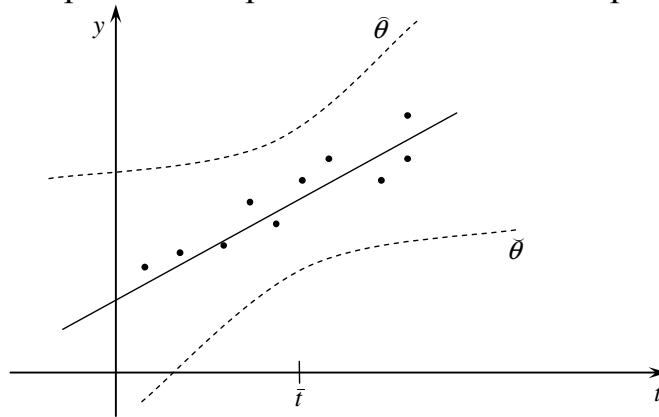


Рис. 3.12 – Точечный прогноз (сплошная линия) и границы доверительных интервалов (пунктир)

Построим интервал прогноза для процесса  $\eta(t, \omega) = U(t) + V(t) + \varepsilon(t, \omega)$  в случае, когда случайные воздействия имеют автокорреляцию 1-го ранга. Из опыта построения аддитивных моделей динамических рядов известно, что в ряде случаев оценки коэффициентов тренда  $U(t)$  и циклической компоненты  $V(t)$  являются линейными функциями от результатов измерений, т. е. модельная функция имеет вид

$$y(t) = \sum_{i=1}^n \alpha_i(t) y_i, \text{ а модельный процесс —}$$

$$y(t, \omega) = \sum_{i=1}^n \alpha_i(t) \eta(i, \omega) + \varepsilon(t, \omega),$$

где  $\alpha_i(t)$ ,  $i = 1, \dots, n$  – это известные функции. Так как мы предполагаем наличие у помех автокорреляции 1-го ранга, то  $\varepsilon(t, \omega) = \rho \varepsilon(t-1, \omega) + \varepsilon^{(1)}(t, \omega)$ , и новые помехи  $\varepsilon^{(1)}(t, \omega)$  уже удовлетворяют условиям Гаусса–Маркова. Из последнего равенства получаем

$$\varepsilon(t, \omega) = \rho^{t-1} \varepsilon(1, \omega) + \sum_{k=0}^{t-2} \rho^k \varepsilon^{(1)}(t-k, \omega).$$

Все слагаемые в правой части независимы. Пусть  $t_i = i$ ,  $t = n + l$ . Тогда

$$\begin{aligned} y(n+l, \omega) &= \alpha_1(n+l) [U(1) + V(1) + \varepsilon(1, \omega)] + \\ &+ \sum_{i=2}^n \alpha_i(n+l) \left[ U(i) + V(i) + \rho^{i-1} \varepsilon(1, \omega) + \sum_{k=0}^{i-2} \rho^k \varepsilon^{(1)}(i-k, \omega) \right] + \\ &+ \rho^{n+l-1} \varepsilon(1, \omega) + \sum_{k=0}^{n+l-2} \rho^k \varepsilon^{(1)}(n+l-k, \omega). \end{aligned}$$

Перегруппировав слагаемые, получаем

$$\begin{aligned} y(n+l, \omega) &= \sum_{i=1}^n \alpha_i(n+l) [U(i) + V(i)] + \left[ \sum_{i=1}^n \alpha_i(n+l) \rho^i + \rho^{n+l-1} \right] \cdot \varepsilon(1, \omega) + \\ &+ \sum_{k=2}^n \left[ \sum_{i=k}^n \rho^{i-k} \alpha_i(n+l) + \rho^{n+l-k} \right] \cdot \varepsilon^{(1)}(k, \omega) + \sum_{k=1}^l \rho^{l-k} \varepsilon^{(1)}(n+k, \omega). \end{aligned}$$

Все слагаемые в правой части этого равенства являются независимыми величинами. Это позволяет нам вычислить дисперсию модельного процесса.

По условию  $D\varepsilon(1, \omega) = \sigma^2$ . Из авторегрессионного соотношения получаем  $D\varepsilon^{(1)}(t, \omega) = (1 - \rho^2) \sigma^2$ . Так как модельный процесс  $y(t, \omega)$  представлен в виде суммы независимых случайных величин, то дисперсия величины  $y(n+l, \omega)$  равна

$$D(\rho, \sigma) = Dy(n+l, \omega) = \sigma^2 D_1^2 + \sigma^2 (1 - \rho^2) \sum_{k=2}^n D_k^2 + \sigma^2 (1 - \rho^{2l}),$$

$$\text{где } D_k = \sum_{i=k}^n \alpha_i(n+l) \rho^{i-k} + \rho^{n+l-1}, \quad k = 1, \dots, n.$$

Обозначим  $D^* = D(\rho^*, \sigma^*)$ . Поскольку в этом случае модельный процесс содержит неизвестный коэффициент авторегрессии  $\rho$ , то заменим его оценкой  $\rho^*$ . В этом случае построить точные границы прогнозного интервала не удастся. Поэтому для построения прогнозного интервала в этом случае используем неравенство Чебышева

$$P\{|\xi(\omega) - M\xi(\omega)| < \beta\} \geq 1 - \frac{D\xi(\omega)}{\beta^2}.$$

В нашем случае

$$P\{|y(t, \omega) - My(t, \omega)| < \beta\} = P\{|y(t, \omega) - M\eta(t, \omega)| < \beta\} = \gamma.$$

Найдем  $\beta$  из равенства  $1 - \frac{D^*}{\beta^2} = \gamma$ ;  $\beta = \sqrt{\frac{D^*}{1-\gamma}}$ .

Тогда интервал прогноза имеет вид

$$\left( y(n+l) - \sqrt{\frac{D^*}{1-\gamma}}; y(n+l) + \sqrt{\frac{D^*}{1-\gamma}} \right).$$

Следует заметить, что прогнозы, построенные с помощью неравенства Чебышева, как правило, не отличаются высокой точностью. Поэтому такая методика прогнозирования применяется лишь при краткосрочном прогнозировании.

Прогнозирование поведения временных рядов – это весьма сложный раздел теории случайных процессов. Научой разработаны разнообразные методы построения прогнозов показателей, меняющихся во времени. В данном учебном пособии изложены лишь наиболее простые из них. Для построения высокоточных прогнозов применяют специальные методы, учитывающие индивидуальные особенности поведения прогнозируемого показателя. Детальное изложение теории таких методов и практики их применения можно найти в литературе, указанной в конце данного пособия.

## Контрольные вопросы к разделу 1

### Основной уровень

1. Сформулируйте постановку эконометрической задачи.
2. Дайте определение временного ряда.
3. Опишите обобщенную аддитивную модель временного ряда и ее составляющие.
4. Объясните, в каком случае необходимо осуществлять приведение членов временного ряда к сопоставимому виду.
5. Дайте определение тренда временного ряда.
6. Укажите основные этапы процедуры построения тренда.
7. Объясните предназначение критерия Фостера–Стюарта.
8. Опишите, как проверить наличие во временном ряду основных форм трендовых моделей.
9. Укажите основной метод оценивания коэффициентов трендовых моделей.
10. Объясните, как сравнивают различные трендовые модели между собой.
11. Дайте определение циклической компоненты временного ряда.
12. Укажите основные этапы процедуры построения циклической компоненты.
13. Дайте определение сезонной волны. Объясните ее предназначение.

14. Укажите, в каком аналитическом виде следует строить циклическую компоненту временного ряда.
15. Дайте определение элементарной гармоник.
16. Перечислите способы проверки гармоник на значимость.
17. Укажите, в каких единицах измеряются амплитуды гармоник, входящих в состав циклической компоненты.
18. Дайте определение остаточной компоненты.
19. Опишите, какими свойствами должна обладать остаточная компонента временного ряда.
20. Укажите, какой критерий используется для проверки гипотезы о случайности остаточной компоненты.
21. Укажите, какие критерии можно применять для проверки гипотезы о нормальности остаточной компоненты.
22. Опишите алгоритм проверки гипотезы об отсутствии автокорреляции остатков.
23. Укажите величины, которые используются для оценки точности построенной аддитивной модели.

### Углубленный уровень

1. Опишите, как влияет сглаживание временного ряда на слагаемые аддитивной модели.
2. Докажите, что сглаживание временного ряда уменьшает его дисперсию.
3. Докажите, что сглаживание временного ряда сохраняет структуру тренда.
4. Укажите, как изменяются коэффициенты линейного тренда при сглаживании.
5. Перечислите известные вам критерии для проверки гипотезы о наличии тренда во временном ряду.
6. Опишите принцип проверки наличия во временном ряду основных форм трендовых моделей.
7. Перечислите свойства, которыми обладают МНК-оценки коэффициентов тренда. Укажите, при каких условиях эти свойства справедливы.
8. Перечислите возможные причины того, что временной ряд и сезонная волна ведут себя не синхронно.
9. Укажите основной метод оценки коэффициентов циклической компоненты.
10. Объясните, к чему приводит отбрасывание незначимых гармоник и как с этим бороться.
11. Перечислите условия Гаусса–Маркова. Опишите, какие свойства МНК-оценок утрачиваются при нарушении этих условий.
12. Объясните, что показывают выборочные значения асимметрии и эксцесса.

13. Укажите, в каком случае для проверки нормальности используется RS-критерий.

14. Объясните, почему на основании факта отсутствия автокорреляции остатков делается вывод об их независимости.

15. Укажите возможные причины возникновения автокорреляции остатков.

16. Опишите принцип построения прогнозов временного ряда.

## Раздел 2 Некоторые задачи математической экономики

Одним из распространенных типов экономических задач являются задачи оптимизации производства. Объектом исследования в них является фирма, производящая некоторую продукцию. Результаты деятельности фирмы определяются производственной функцией, которая задает объем выпуска продукции при определенных сочетаниях производственных факторов. Перед исследователем ставится задача поиска такого сочетания факторов, которое оптимизирует производственный процесс. Таким образом, задачи оптимизации производства относятся к классу оптимизационных задач. При решении таких задач необходимо задать критерий цели (или целевую функцию) оптимизации.

Критерий цели – это правило, позволяющее оценивать состояние изучаемого объекта и сравнивать его с желаемым, а также устанавливать достаточность и недостаточность точности имитации объекта.

*Основные принципы построения целевых функций*

1) однозначность: для поставленной задачи целевая функция единственна;

2) управляемость: функция зависит только от тех аргументов и параметров, которые доступны для измерения и управления;

3) подходящая форма: функция должна иметь экономический смысл и обладать необходимыми аналитическими свойствами.

*Примеры целевых функций*

1) доход от реализации продукции  $R(x_1, \dots, x_n) = \sum_{i=1}^n p_i x_i$ ;

2) издержки производства  $c(x_1, \dots, x_n) = \sum_{i=1}^n p_i c_i$ ;

3) прибыль  $\dot{R}(x_1, \dots, x_n) = R(x_1, \dots, x_n) - C(x_1, \dots, x_n)$ ;

4) качество решения  $F(y_1, \dots, y_n) = \sum_{i=1}^n \alpha_j (y_j - y_i^*)^2$ ;

5) целевая функция времени  $T = \min(t_f - t_0)$ .

### 1 Производственные функции

В ряде задач в качестве целевой функции рассматривают производственную функцию. Рассмотрим работу некоторого производственного объекта (завода, фабрики, фирмы, цеха и т. п.). Предположим, что система и показатели работы объекта с течением времени не меняются. Для организации выпуска продукции необходимо  $n$  видов ресурсов. Обозначим  $x = (x_1, \dots, x_n)$  – набор, элементами которого являются количества каждого из ресурсов, затраченных в производстве;  $y$  – объем выпуска продукции, производимой с помощью этих ресурсов за некоторый фиксированный промежуток времени.



**Определение.** Функция  $y = f(x_1, \dots, x_n)$ , указывающая связь между выпуском  $y$  и ресурсами  $x_i$ ,  $i = \overline{1, n}$ , называется *производственной функцией* (ПФ).

**Замечание 1.** Переменные  $x_i$ ,  $i = \overline{1, n}$ , могут измеряться как в натуральном, так и в денежном выражении. Предпочтительнее использовать одну и ту же денежную единицу.

**Замечание 2.** Ввиду того, что каждая производственная единица (фабрика, фирма, цех, человек) является индивидуальной даже в рамках стандартного производственного процесса, то, говоря о ПФ, следует иметь в виду, что она отражает лишь принципиальный вид связи, т. е. она зависит от набора неизвестных параметров:  $y = f(x_1, \dots, x_n, a_1, \dots, a_m)$ , значения которых  $(a_1, \dots, a_m)$  должны индивидуально подбираться для каждого производственного объекта. Такого рода проблемы решаются методами эконометрии.

Согласно экономической теории, ПФ должна отражать закон убывающей эффективности производства, который гласит: «с увеличением объемов используемых ресурсов объем выпуска продукции должен расти, но скорость его роста должна снижаться». Это свойство проиллюстрировано на рис. 4.1.

К производственным функциям предъявляют ряд требований, обусловленных положениями экономической теории.

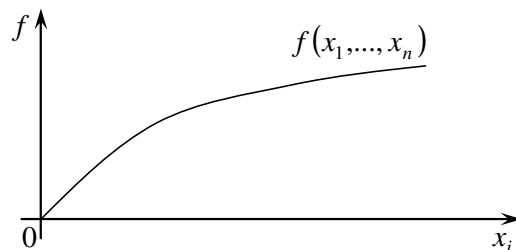


Рис. 4.1 – Иллюстрация закона убывающей эффективности производства

#### Основные требования, предъявляемые к ПФ

- 1\*. Без ресурсов нет выпуска:  $f(0) = 0$ .
- 2\*. Все ресурсы необходимы в производстве:  $f(x_1, \dots, 0, \dots, x_n) = 0$ .
3. С ростом объема использования каждого ресурса объем выпуска возрастает:

$$x_i'' > x_i' \Rightarrow f(x_1, \dots, x_i'', \dots, x_n) > f(x_1, \dots, x_i', \dots, x_n), \quad i = \overline{1, n}.$$

4. Закон убывающей эффективности производства: с ростом затрат одного ресурса при неизменных затратах других ресурсов величина прироста выпуска не растет:

$$\frac{\partial^2 f}{\partial x_i^2} \leq 0, \quad i = \overline{1, n}.$$

5. С ростом использования объема одного из ресурсов предельная эффективность другого ресурса возрастает:

$$\frac{\partial}{\partial x_j} \left( \frac{\partial}{\partial x_i} f(x_1, \dots, x_n) \right) \geq 0, \quad i, j = \overline{1, n}, \quad i \neq j.$$

**Замечание.** Требования, помеченные символом \*, не являются обязательными.

## 2 Основные виды производственных функций

**Определение.** Производственной функцией Кобба–Дугласа (ПФКД) называется функция

$$\text{ПФКД}(K, L) = a_0 K^{a_1} L^{a_2}, \quad a_0, a_1, a_2 > 0, \quad a_1 + a_2 = 1, \quad K \geq 0, \quad L \geq 0,$$

где  $K$  – объем капиталовложений в производство,  $L$  – трудозатраты.

**Пример.** С 1960 по 1985 гг. по результатам исследований экономики СССР было построено 2 модели, задающие объем ВВП, в виде ПФКД:

$$y = 1,022 K^{0,538} L^{0,462} \quad \text{– без учета научно-технического прогресса;}$$

$$y = 1,038 e^{0,03t} K^{0,97} L^{0,23} \quad \text{– с учетом научно-технического прогресса.}$$

Легко показать, что ПФКД обладает всеми свойствами ПФ.

**Определение.** Производственной функцией Леонтьева называется функция

$$Y(K, L) = \min\{aK, bL\} = \begin{cases} aK, & \text{если } K \leq \frac{b}{a}L, \\ bL, & \text{если } K > \frac{b}{a}L, \end{cases} \quad a, b > 0.$$

ПФ Леонтьева имеет особенность на прямой  $K = \frac{b}{a}L$ : она не дифференцируема в точках этой прямой. Однако вне этой прямой все необходимые свойства ПФ выполняются.

**Определение.** Линейной производственной функцией называется функция

$$Y(K, L) = a_0 + a_1 K + a_2 L, \quad a_0 \in R, \quad a_1, a_2 > 0.$$

Для линейной производственной функции требования ПФ 1 и 2 в общем случае не выполняются.

Легко заметить, что между ПФКД и линейной ПФ имеется следующая взаимосвязь:

$$\begin{aligned} \ln(\text{ПФКД}(K; L)) &= \ln(a_0 K^{a_1} L^{a_2}) = \ln a_0 + a_1 \ln K + a_2 \ln L = \\ &= \text{ПФЛинейная}(\ln K; \ln L). \end{aligned}$$

**Определение.** Производственной функцией с постоянной эластичностью замещения ресурсов называется функция

$$\text{ПФПЭЗ}(K; L) = a_0 \left( u K^{-\rho} + (1-u) L^{-\rho} \right)^{-\frac{n}{\rho}},$$

$$a_0 > 0, \quad u \in (0; 1), \quad \rho \in [-1; 0) \cup (0; +\infty).$$

Взаимосвязь ПФПЭЗ с ПФКД, ПФ Леонтьева и линейной ПФ

1. При  $\rho = -1$  справедливо

$$\text{ПФПЭЗ}(K, L) = a_0(uK + (1-u)L)^n = (\text{ПФЛинейная}(K, L))^n.$$

2. При  $\rho \rightarrow 0$  справедливо

$$\begin{aligned} \lim_{\rho \rightarrow 0} \ln(uK^{-\rho} + (1-u)L^{-\rho})^{-\frac{1}{\rho}} &= - \lim_{\rho \rightarrow 0} \frac{\ln(uK^{-\rho} + (1-u)L^{-\rho})}{\rho} = \left\{ \begin{array}{l} \text{правило} \\ \text{Лопиталя} \end{array} \right\} = \\ &= \lim_{\rho \rightarrow 0} \frac{uK^{-\rho} \ln K + (1-u)L^{-\rho} \ln L}{uK^{-\rho} + (1-u)L^{-\rho}} = \frac{u \ln K + (1-u) \ln L}{u + 1 - u} = \ln(K^u L^{1-u}). \end{aligned}$$

Следовательно,  $\lim_{\rho \rightarrow 0} \text{ПФПЭЗ}(K, L) = \text{ПФКД}(K^n, L^n)$ .

3. При  $\rho \rightarrow +\infty$  справедливо

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} (uK^{-\rho} + (1-u)L^{-\rho})^{-\frac{1}{\rho}} &= \left\{ \begin{array}{l} \text{при } K < L : \\ K/L < 1 \end{array} \right\} = \\ &= \lim_{\rho \rightarrow +\infty} \left( K \left( u + (1-u) \left( \frac{K}{L} \right)^{\rho} \right)^{-\frac{1}{\rho}} \right) = K. \end{aligned}$$

Аналогично, при  $K > L$  получаем  $L$ .

Пусть  $K = L$ , тогда  $\lim_{\rho \rightarrow +\infty} (uK^{-\rho} + (1-u)L^{-\rho})^{-\frac{1}{\rho}} = K = L$ .

Получаем  $\lim_{\rho \rightarrow +\infty} \text{ПФПЭЗ}(K, L) = \text{ПФЛеонтьева}(K^n, L^n)$ .

### 3 Основные типы задач оптимизации производства

Рассмотрим следующие типы задач оптимизации производства и методы их решения:

- 1) задача максимизации прибыли на долговременном промежутке функционирования фирмы;
- 2) задача максимизации прибыли на кратковременном промежутке функционирования фирмы;
- 3) задача максимизации объема выпуска продукции при ограниченных финансовых ресурсах;
- 4) задача минимизации издержек производства при фиксированном объеме выпуска продукции.

#### 3.1 Задача максимизации прибыли на долговременном промежутке функционирования

*Эмпирическая формулировка задачи*

Рассматривается работа некоторого производства в стационарном режиме, т. е. характеристики производства, цены на ресурсы и продукцию считаются неизменными.

Пусть процесс производства описывается ПФ известного вида  $y = f(x_1, \dots, x_n)$ , где  $x_i$  – объем  $i$ -го привлекаемого ресурса. Известны также  $p_0$  – цена за единицу выпускаемой продукции и  $p_1, \dots, p_n$  – цены за единицу привлекаемых ресурсов соответственно.

Работа фирмы рассматривается на неограниченном временном отрезке.

Необходимо определить такие объемы ресурсов  $(x_1, \dots, x_n)$ , используя которые, фирма получит наибольшую прибыль на рассматриваемом промежутке времени.

#### *Математическая постановка задачи*

Дано:  $x_i \geq 0$ ,  $i = 1, \dots, n$ ;  $p_i > 0$ ,  $i = 0, 1, \dots, n$ ;  $y = f(x_1, \dots, x_n)$  – обладает свойствами ПФ.

Найти: точку  $(x_1^*, \dots, x_n^*)$  такую, что

$$\max_{x_i \geq 0, i=1, \dots, n} (p_0 f(x_1, \dots, x_n) - p_1 x_1 - \dots - p_n x_n) = p_0 f(x_1^*, \dots, x_n^*) - p_1 x_1^* - \dots - p_n x_n^*.$$

#### *План решения задачи в случае $n = 2$*

Рассмотрим прибыль от реализации продукции

$$PR(x_1, x_2) = p_0 f(x_1, x_2) - p_1 x_1 - p_2 x_2.$$

Поскольку рассматривается долговременный промежуток функционирования производства, то оборотные средства производителя могут быть использованы многократно. Это позволяет считать, что финансовые возможности фирмы практически неограниченны. В силу сказанного, поставленную задачу можно рассматривать как задачу поиска глобального максимума, т. е. необходимо найти точки глобального максимума функции  $PR(x_1, x_2)$  на множестве  $x_1 \geq 0$ ,  $x_2 \geq 0$ .

Если  $f(x_1, x_2)$  дифференцируема, то находим точки  $(x_1^*, x_2^*)$ , подозрительные на экстремум, как решения системы

$$\begin{cases} \frac{\partial PR}{\partial x_1} = 0, \\ \frac{\partial PR}{\partial x_2} = 0. \end{cases} \Leftrightarrow \begin{cases} \frac{\partial f}{\partial x_1} = \frac{p_1}{p_0}, \\ \frac{\partial f}{\partial x_2} = \frac{p_2}{p_0}. \end{cases}$$

Дополнительно вычисляем  $PR(0,0)$ ,  $PR(0, x_1)$ ,  $PR(x_2, 0)$ ,  $PR(+\infty, x_2)$ ,  $PR(x_1, +\infty)$ ,  $PR(+\infty, +\infty)$  и сравниваем полученные значения с  $PR(x_1^*, x_2^*)$ .

### **3.2 Задача максимизации прибыли на кратковременном промежутке функционирования**

Рассматривается та же проблема максимизации прибыли, но работа фирмы рассматривается на кратковременном промежутке функционирования производства. Поскольку теперь временной отрезок ограничен, то ограничены и финансовые возможности производства. Пусть  $I$  – объем финансовых ресурсов, которыми располагает фирма для

приобретения ресурсов. Тогда должно выполняться неравенство  $p_1x_1 + \dots + p_nx_n \leq I$ .

Будем рассматривать случай  $n = 2$ . Множество возможных планов производства  $M$  изображено на рис. 4.2.

Функция прибыли  $\Pi R$  непрерывна, следовательно, на ограниченном замкнутом множестве  $M$  функция  $\Pi R$  всегда достигает наибольшего значения. Следовательно, поставленная задача заведомо имеет решение.

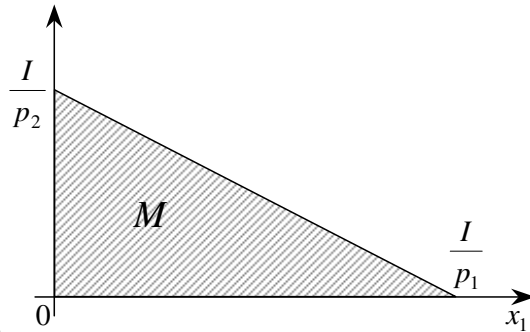


Рис. 4.2 – Множество возможных планов производства  $M$

*Возможный путь решения задачи*

1. Для функции  $\Pi R$  найти точки, подозрительные на экстремум.
  2. Выбрать из них те, которые принадлежат  $M$ , и вычислить в них значения  $\Pi R$ .
  3. Найти значения  $\Pi R$  на осях  $Ox_1$ ,  $Ox_2$ .
  4. Найти значения  $\Pi R$  на прямой  $p_1x_1 + p_2x_2 = I$ .
  5. Выбрать наибольшее значение  $\Pi R$  из полученных.
- Соответствующая точка  $(x_1^*, x_2^*)$  есть оптимальный план производства.

Заметим, что если точка экстремума прибыли существует, то для нее всегда выполняется равенство

$$\frac{\frac{\partial f(x_1^*, x_2^*)}{\partial x_1}}{\frac{\partial f(x_1^*, x_2^*)}{\partial x_2}} = \frac{p_1}{p_2}.$$

Эта точка называется *точкой рыночного равновесия производителя*.

Условие того, что полученная точка является точкой максимума (или линией максимумов), согласно критерию Сильвестра, имеет вид

$$\left\{ \begin{array}{l} \frac{\partial^2 f}{\partial x_1^2}(x_1^*, x_2^*) < 0, \\ \begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{vmatrix} \geq 0. \end{array} \right.$$

**Определение.** Пусть  $y^*$  – оптимальный фиксированный объем производства. Уравнение  $y^* = f(x_1, x_2)$  задает функцию  $x_2 = x_2(x_1)$ , которая называется *изоквантой*.

**Теорема.** Изокванта является убывающей функцией, либо выпуклой вниз, либо линейной.

Покажем, что изокванта убывает. Продифференцировав равенство  $y^* = f(x_1, x_2) = f(x_1, x_2(x_1))$ , получаем

$$0 = \frac{\partial f(x_1, x_2(x_1))}{\partial x_1} + \frac{\partial f}{\partial x_2} \cdot \frac{dx_2(x_1)}{dx_1}.$$

Тогда

$$\frac{dx_2(x_1)}{dx_1} = - \frac{\frac{\partial f(x_1, x_2(x_1))}{\partial x_1}}{\frac{\partial f(x_1, x_2(x_1))}{\partial x_2}} = - \frac{p_1}{p_2} < 0.$$

Следовательно, функция  $x_2(x_1)$  убывает.

Рассмотрим выпуклость изокванты.

$$\begin{aligned} \frac{dx_2(x_1)}{dx_1} &= - \frac{\frac{\partial f}{\partial x_1}}{\frac{\partial f}{\partial x_2}}; \\ \frac{d^2 x_2(x_1)}{dx_1^2} &= - \frac{\left[ \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_1 \partial x_2} \cdot \frac{dx_2}{dx_1} \right] \cdot \frac{\partial f}{\partial x_2} - \frac{\partial f}{\partial x_1} \cdot \left[ \frac{\partial^2 f}{\partial x_1 \partial x_2} + \frac{\partial^2 f}{\partial x_2^2} \cdot \frac{dx_2}{dx_1} \right]}{\left( \frac{\partial f}{\partial x_2} \right)^2}, \\ \frac{d^2 x_2(x_1)}{dx_1^2} &= - \frac{\frac{\partial^2 f}{\partial x_1^2} \cdot \frac{\partial f}{\partial x_2} - 2 \frac{\partial^2 f}{\partial x_1 \partial x_2} \cdot \frac{\partial f}{\partial x_1} + \frac{\partial^2 f}{\partial x_2^2} \cdot \left( \frac{\partial f}{\partial x_1} \right)^2}{\left( \frac{\partial f}{\partial x_2} \right)^2} \leq 0. \end{aligned}$$

Последнее неравенство верно в силу свойств ПФ  $f(x_1, x_2)$ . Теорема доказана.

Таким образом, в точке оптимума  $(x_1^*, x_2^*)$  связь между ресурсами имеет вид, изображенный на рис. 4.3.

**Пример.** Производство функционирует по закону Кобба–Дугласа  $Y = a_0 K^{a_1} L^{a_2}$ ,  $p_0$  – цена произведенного товара,  $p_1$ ,  $p_2$  – цены ресурсов,  $I$  – бюджет производства.

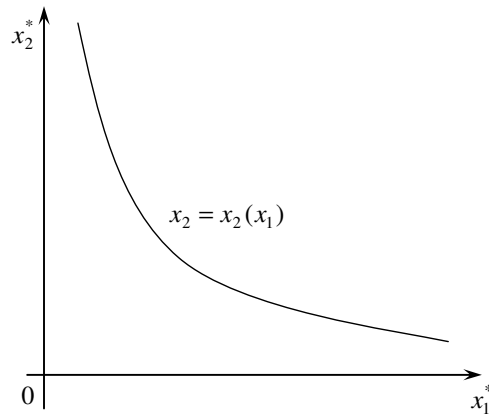


Рис. 4.3 – Форма зависимости  $x_2 = x_2(x_1)$

Необходимо

1) построить план производства, максимизирующий прибыль на кратковременном промежутке;

2) изучить зависимость производственных показателей от колебания цен на ресурсы.

В данном случае  $\Pi R = a_0 p_0 K^{a_1} L^{a_2} - p_1 K - p_2 L$ .

1. Находим точки, подозрительные на экстремум

$$\begin{cases} \frac{\partial \Pi R}{\partial K} = a_0 p_0 a_1 K^{-a_2} L^{a_2} - p_1 = 0, \\ \frac{\partial \Pi R}{\partial L} = a_0 p_0 a_2 K^{a_1} L^{-a_1} - p_2 = 0. \end{cases}$$

Из системы определяем  $L = \frac{p_1 a_2}{p_2 a_1} K$ . Найдя точку  $K^* = \frac{a_1}{p_1} I$ ,

$L^* = \frac{a_2}{p_2} I$ , являющуюся точкой пересечения прямой  $L = \frac{p_1 a_2}{p_2 a_1} K$  и

бюджетной прямой  $x_1 p_1 + x_2 p_2 = I$ , получаем оптимальный план производства (рис. 4.4).

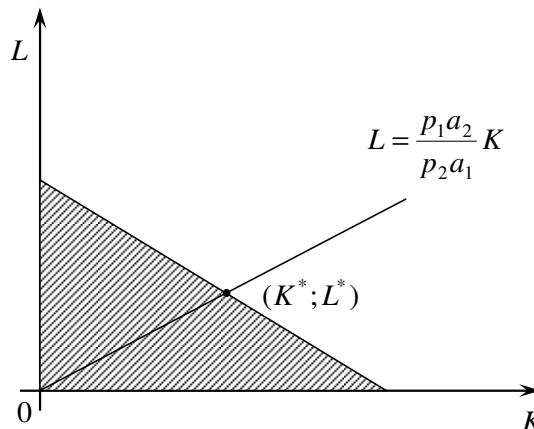


Рис. 4.4 – Оптимальный план производства  $(K^*; L^*)$

Таким образом, оптимальный план производства  $K^* = \frac{a_1}{p_1} I$ ,  $L^* = \frac{a_2}{p_2} I$ ,  
 максимальная прибыль  $PR(K^*, L^*) = I \left[ p_0 a_0 \left( \frac{a_1}{p_1} \right)^{a_1} \left( \frac{a_2}{p_2} \right)^{a_2} - 1 \right]$ .

2. Пусть цена второго ресурса остается неизменной, а цена первого ресурса меняется, т. е.  $p_2 = const$ ,  $\bar{p}_1$  – новая цена, причем  $\bar{p}_1 > p_1$ . Тогда из результатов выполнения пункта 1 следует, что  $\bar{K}^* < K^*$  и  $\bar{PR}^* < PR^*$ . Таким образом, рост цен на ресурсы снижает прибыльность производства.

### 3.3 Задача максимизации объема выпуска продукции при ограниченных финансовых ресурсах

#### Эмпирическая формулировка задачи

Рассматривается работа некоторого производства в стационарном режиме, т. е. характеристики производства, цены на ресурсы и продукцию считаются неизменными.

Пусть процесс производства описывается ПФ вида  $y = f(x_1, \dots, x_n)$ , где  $x_i$  – объем  $i$ -го привлекаемого ресурса. Известны также  $p_0$  – цена за единицу выпускаемой продукции и  $p_1, \dots, p_n$  – цены за единицу привлекаемых ресурсов соответственно.

Предполагается, что финансовые возможности производителя по закупке ресурсов ограничены величиной  $I$  грн. Необходимо определить такие объемы ресурсов  $(x_1, \dots, x_n)$ , используя которые, фирма получит наибольший объем выпуска продукции в рамках имеющихся финансовых ресурсов.

#### Математическая постановка задачи

Дано:  $x_i \geq 0$ ,  $i = 1, \dots, n$ ;  $p_i > 0$ ,  $i = 0, 1, \dots, n$ ;  $y = f(x_1, \dots, x_n)$  – обладает свойствами ПФ,  $I > 0$ .

Найти: точку  $(x_1^*, \dots, x_n^*)$  такую, что

$$\max_{p_1 x_1 + \dots + p_n x_n \leq I} f(x_1, \dots, x_n) = f(x_1^*, \dots, x_n^*).$$

#### План решения задачи в случае $n = 2$

Множество возможных планов производства изображено на рис. 4.5.

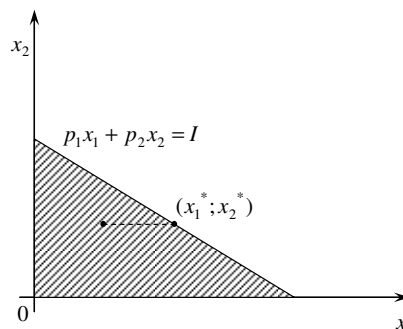


Рис. 4.5 – Множество возможных планов производства



Рассматриваемая задача является задачей нелинейного программирования. Сведем ее к задаче на условный экстремум. Очевидно, что решение задачи существует всегда. Покажем, что решение лежит на бюджетной прямой  $p_1x_1 + p_2x_2 = I$ .

Пусть оптимальная точка  $(x_1^*, x_2^*)$  лежит внутри треугольника возможных планов производства (рис. 4.5). ПФ обладает следующими свойствами:

$$\frac{\partial f}{\partial x_1} > 0, \frac{\partial f}{\partial x_2} > 0.$$

Двигаясь вдоль одной переменной (параллельно одной из координатных осей) и учитывая указанные свойства ПФ, получаем, что при таком движении значения ПФ увеличиваются. Но этого не может быть, так как в точке  $(x_1^*, x_2^*)$  ПФ достигает наибольшего значения. Значит, точка  $(x_1^*, x_2^*)$  не может лежать внутри треугольника, то есть она лежит на бюджетной прямой.

Таким образом, исходная задача сводится к следующей: найти  $(x_1^*, x_2^*)$  такую, что  $\max_{p_1x_1 + p_2x_2 = I} f(x_1, x_2) = f(x_1^*, x_2^*)$ . Это задача уже является задачей на условный экстремум и может быть решена, например, с помощью функции Лагранжа или методом понижения размерности задачи.

#### I. Решение с помощью функции Лагранжа

Вводим функцию Лагранжа  $L(x_1, x_2) = f(x_1, x_2) - \lambda(p_1x_1 + p_2x_2 - I)$  и составляем систему уравнений для поиска точек, подозрительных на экстремум.

$$\begin{cases} \frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} - \lambda p_1 = 0, \\ \frac{\partial L}{\partial x_2} = \frac{\partial f}{\partial x_2} - \lambda p_2 = 0, \\ p_1x_1 + p_2x_2 = I. \end{cases} \Rightarrow \begin{cases} \lambda = \frac{1}{p_1} \cdot \frac{\partial f}{\partial x_1}, \\ \lambda = \frac{1}{p_2} \cdot \frac{\partial f}{\partial x_2}. \end{cases} \Rightarrow \begin{cases} \frac{\frac{\partial f}{\partial x_1}}{\frac{\partial f}{\partial x_2}} = \frac{p_1}{p_2}, \\ p_1x_1 + p_2x_2 = I. \end{cases}$$

Решив записанную систему уравнений, найдем  $(x_1^*, x_2^*)$ .

#### II. Решение методом понижения размерности задачи

Выразим  $x_2$  из бюджетного уравнения,  $x_2 = \frac{I - p_1x_1}{p_2}$  и подставим в

ПФ. Теперь задача сведена к классической задаче поиска наибольшего значения функции на отрезке:

$$y = f\left(x_1, \frac{I - p_1x_1}{p_2}\right) \rightarrow \max, x_1 \in \left[0; \frac{I}{p_1}\right],$$

которая может быть решена методами математического анализа.

### 3.4 Задача минимизации издержек производства при фиксированном объеме выпуска продукции

#### *Эмпирическая формулировка задачи*

Рассматривается работа некоторого производства в стационарном режиме, т. е. характеристики производства, цены на ресурсы и продукцию считаются неизменными. Пусть процесс производства описывается ПФ вида  $y = f(x_1, \dots, x_n)$ , где  $x_i$  – объем  $i$ -го привлекаемого ресурса. Известны также  $p_0$  – цена за единицу выпускаемой продукции и  $p_1, \dots, p_n$  – цены за единицу привлекаемых ресурсов соответственно.

Предполагается, что производитель работает в установившемся режиме и объем производства за единицу времени остается равным фиксированному значению  $y^0$ . Необходимо определить такие объемы ресурсов  $(x_1, \dots, x_n)$ , используя которые, фирма получит наименьшие издержки производства при фиксированном объеме выпуска продукции.

#### *Математическая постановка задачи*

Дано:  $x_i \geq 0$ ,  $i = 1, \dots, n$ ;  $p_i > 0$ ,  $i = 0, 1, \dots, n$ ;  $y = f(x_1, \dots, x_n)$  – обладает свойствами ПФ.

Найти: точку  $(x_1^*, \dots, x_n^*)$  такую, что

$$\min_{y=y^0} (p_1 x_1 + \dots + p_n x_n) = p_1 x_1^* + \dots + p_n x_n^*.$$

#### *План решения задачи в случае $n = 2$*

Данная задача является задачей на условный экстремум и может быть решена с помощью стандартных методов математического анализа.

I. Решение с помощью функции Лагранжа

В данном случае функция Лагранжа имеет вид  $L(x_1, x_2) = p_1 x_1 + p_2 x_2 - \lambda(f(x_1, x_2) - y^0)$ . Далее, следуя стандартному алгоритму, составляем систему уравнений

$$\begin{cases} \frac{\partial L}{\partial x_1} = 0, \\ \frac{\partial L}{\partial x_2} = 0, \\ \frac{\partial L}{\partial \lambda} = 0. \end{cases}$$

Решив ее, получим точки, подозрительные на экстремум. Выбрав точку наименьшего значения функции издержек на множестве  $y^0 = f(x_1, x_2)$ , получим решение задачи.

II. Решение методом понижения размерности задачи

Можно свести задачу к одномерной. Так как  $f$  монотонна (как ПФ),

то можно из равенства  $y^0 = f(x_1, x_2)$  выразить  $x_2$  через  $x_1$ :  $x_2 = \varphi(x_1)$ . Теперь задача сведена к классической задаче поиска наименьшего значения функции на полуоси:

$$c(x_1, \varphi(x_1)) = p_1 x_1 + p_2 \varphi(x_1) \rightarrow \min, \quad x_1 \geq 0,$$

которая может быть решена стандартными методами математического анализа.

## Контрольные вопросы к разделу 2

### Основной уровень

1. Дайте определение производственной функции.
2. Перечислите основные свойства производственной функции.
3. Проиллюстрируйте графически закон убывающей эффективности производства.
4. Перечислите основные виды производственных функций.
5. Перечислите рассмотренные задачи оптимизации производства.
6. Сформулируйте эмпирическую и математическую постановки задачи максимизации прибыли на долговременном промежутке функционирования.
7. Сформулируйте эмпирическую и математическую постановки задачи максимизации прибыли на кратковременном промежутке функционирования.
8. Сформулируйте эмпирическую и математическую постановки задачи максимизации объема выпуска продукции при ограниченных финансовых ресурсах.
9. Сформулируйте эмпирическую и математическую постановки задачи минимизации издержек производства при фиксированном объеме выпуска продукции.

### Углубленный уровень

1. Объясните смысл требований 4 и 5 производственной функции.
2. Укажите, все ли свойства производственной функции выполняются для линейной производственной функции.
3. Приведите алгоритм решения задачи максимизации прибыли на долговременном промежутке функционирования.
4. Приведите алгоритм решения задачи максимизации прибыли на кратковременном промежутке функционирования.
5. Приведите алгоритм решения задачи максимизации объема выпуска продукции при ограниченных финансовых ресурсах.
6. Приведите алгоритм решения задачи минимизации издержек производства при фиксированном объеме выпуска продукции.

### Список рекомендуемой литературы

1. Замков О. О. Математические методы в экономике / О. О. Замков, А. В. Толстопятенко, Ю. В. Черемных. – М.: Дело и сервис, 2001. – 365 с.
2. Четыркин Е. М. Статистические методы прогнозирования / Е. М. Четыркин. – М.: Статистика, 1977. – 232 с.
3. Бендат Дж. Измерение и анализ временных рядов / Дж. Бендат, А. Пирсол. – М.: Мир, 1971. – 408 с.
4. Бокс Дж. Анализ временных рядов. Прогноз и управление / Дж. Бокс, Г. Дженкинс. – М.: Мир, 1974. – Вып. 1. – 406 с.
5. Кобзарь А. И. Прикладная математическая статистика / А. И. Кобзарь. – М.: Физматлит, 2006. – 816 с.
6. Кендэл М. Временные ряды / М. Кендэл. – М.: Финансы и статистика, 1981. – 199 с.
7. Андерсон Т. Статистический анализ временных рядов / Т. Андерсон. – М.: Мир, 1976. – 756 с.
8. Боровков А. А. Математическая статистика. Оценка параметров. Проверка гипотез / А. А. Боровков. – М.: Наука, 1984. – 472 с.
9. Боровков А. А. Математическая статистика. Дополнительные главы / А. А. Боровков. – М.: Наука, 1984. – 143 с.
10. Афифи А. Статистический анализ / А. Афифи, С. Эйзен. – М.: Мир, 1982. – 488 с.
11. Кобелев Н. Б. Практика применения экономико-математических методов и моделей / Н. Б. Кобелев. – М.: Финстатинформ, 2001. – 246 с.

## Статистические таблицы

Таблица 1 – Процентные точки распределения Стьюдента

Значение  $t_{n,\alpha}$  такое, что  $\text{Prob} \left[ |\hat{t}| > t_{n,\alpha} \right] = \alpha$

Число степеней свободы, $n$	Уровень значимости, $\alpha$ (%)					
	20	10	5	2	1,0	0,05
1	3,078	6,314	12,706	31,821	63,657	127,3
2	1,886	2,920	4,303	6,965	9,925	14,089
3	1,638	2,353	3,152	4,541	5,841	7,453
4	1,533	2,132	2,776	3,747	4,604	5,598
5	1,476	2,015	2,571	3,365	4,032	4,773
6	1,444	1,943	2,447	3,143	3,307	4,317
7	1,415	1,895	2,365	2,999	3,499	4,290
8	1,397	1,860	2,306	2,890	3,355	3,833
9	1,383	1,833	2,262	2,821	3,250	3,690
10	1,372	1,812	2,226	2,764	3,169	3,581
11	1,363	1,796	2,201	2,718	3,108	3,497
12	1,356	1,872	2,179	2,681	3,055	3,428
13	1,350	1,771	2,160	2,650	3,012	3,372
14	1,345	1,716	2,145	2,624	2,977	3,326
15	1,341	1,553	2,131	2,602	2,947	3,286
16	1,337	1,746	2,120	2,583	2,921	3,252
17	1,333	1,740	2,111	2,567	2,898	3,222
18	1,330	1,734	2,101	2,552	2,878	3,197
19	1,328	1,729	2,930	2,539	2,861	3,174
20	1,325	1,725	2,860	2,528	2,845	3,153
21	1,323	1,721	2,800	2,518	2,831	3,135
22	1,321	1,717	2,740	2,503	2,819	3,119
23	1,319	1,714	2,690	2,500	2,807	3,104
24	1,318	1,711	2,640	2,492	2,800	3,910
25	1,316	1,706	2,600	2,485	2,787	3,078
26	1,215	1,706	2,560	2,479	2,779	3,670
27	1,314	1,703	2,520	2,473	2,771	3,570
28	1,313	1,701	2,430	2,467	2,763	3,470
29	1,311	1,699	2,450	2,467	2,756	3,380
30	1,310	1,697	2,420	2,467	2,750	3,200
40	1,303	1,684	2,210	2,423	2,704	2,971
60	1,296	1,671	2,000	2,390	2,660	2,915
120	1,289	1,658	1,980	2,358	2,617	2,860
$\infty$	1,282	1,645	1,960	2,326	2,576	2,807

Таблица 2 – Процентные точки распределения числа инверсий

Значение  $A_{n,\alpha}$  такое, что  $\text{Prob}[A_N > A_{N,\alpha}] = \alpha$ , где  $N$  – общее число наблюдаемых значений

$n = \frac{N}{2}$	Уровень значимости, $\alpha$					
	0,99	0,975	0,95	0,05	0,025	0,01
10	9	11	13	31	33	35
12	16	18	21	44	47	49
14	24	27	30	60	63	66
16	34	38	41	78	81	85
18	45	50	54	98	102	107
20	59	64	69	120	125	130
30	152	162	171	263	272	282
40	290	305	319	460	474	489
50	473	495	514	710	729	751
60	702	731	756	1013	1038	1067
70	977	1014	1045	1369	1400	1437
80	1299	1344	1382	1777	1815	1860
90	1668	1721	1766	2238	2283	2336
100	2083	2145	2198	2751	2804	2866

Таблица 3 – Процентные точки распределения количества серий

Значения  $r_{n,\alpha}$  такие, что  $\text{Prob}[r_n > r_{n,\alpha}] = \alpha$ , где  $n = N_1 = N_2 = N$

$n = \frac{N}{2}$	Уровень значимости, $\alpha$					
	0,99	0,975	0,95	0,05	0,025	0,01
5	2	2	3	8	9	9
6	2	3	3	10	10	11
7	3	3	4	11	12	12
8	4	4	5	12	13	13
9	4	5	6	13	14	15
10	5	6	6	15	15	16
11	6	7	7	16	16	17
12	7	7	8	17	18	18
13	7	8	9	18	19	20
14	8	9	10	19	20	21
15	9	10	11	20	210	22
16	10	11	11	22	22	23
18	11	12	13	24	25	26
20	13	14	15	26	27	28
25	17	18	19	32	33	34
30	21	22	24	37	39	40
35	25	27	28	43	44	46
40	30	31	33	48	50	51
45	34	36	37	54	55	57
50	38	40	42	59	61	63
60	47	49	51	70	72	74
70	56	58	60	81	83	85
80	65	68	70	91	93	96
90	74	77	79	102	104	107
100	84	86	88	113	115	117

Таблица 4 – Процентные точки распределения Фишера

Уровень значимости  $\alpha = 0,05$ 

Число степ. своб. $n_2$	Число степеней свободы $n_1$										
	1	2	3	4	5	6	7	8	9	10	11
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,0
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,39	19,40
3	10,13	9,55	9,25	9,12	9,05	8,94	8,89	8,85	8,81	8,79	8,76
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,99	5,93
5	6,65	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70
6	5,99	5,14	4,76	4,53	4,39	4,26	4,21	4,15	4,10	4,06	4,03
7	5,59	4,74	4,35	4,13	3,97	3,87	3,79	3,73	3,68	3,64	3,60
8	5,32	4,46	4,07	3,84	3,69	3,53	3,50	3,44	3,39	3,35	3,31
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94
11	4,84	3,89	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63
14	4,6	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,56
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45
17	4,45	3,59	3,20	2,96	2,81	2,7	2,61	2,55	2,49	2,43	2,41
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31
21	4,32	3,47	3,07	2,84	2,63	2,57	2,49	2,42	2,37	2,32	2,28
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26
23	4,23	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,24
24	4,26	3,40	3,01	2,78	2,62	2,61	2,42	2,36	2,30	2,25	2,22
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20

Таблица 4 – Процентные точки распределения Фишера (продолжение)

Уровень значимости  $\alpha = 0,05$ 

Число степеней свободы $n_2$	Число степеней свободы $n_1$										
	12	14	15	16	20	24	30	40	60	120	$\infty$
1	244,0	245,0	246,0	246,0	248,0	249,1	250,1	251,1	252,2	253,3	254,3
2	19,41	19,42	19,43	19,44	19,45	19,45	19,46	19,47	19,48	19,49	19,50
3	8,74	8,71	8,70	8,69	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	5,91	5,87	5,86	5,84	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	4,68	4,64	4,62	4,60	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	4,00	3,96	3,94	3,92	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	3,57	3,52	3,51	3,49	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	3,28	3,23	3,23	3,20	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	3,07	3,02	3,01	2,93	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	2,91	2,86	2,85	2,82	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	2,79	2,74	2,74	2,70	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	2,69	2,64	2,62	2,60	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	2,60	2,55	2,53	2,51	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	2,53	2,48	2,46	2,44	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	2,48	2,43	2,40	2,39	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	2,42	2,37	2,35	2,33	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	2,38	2,33	2,31	2,29	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	2,34	2,29	2,27	2,25	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	2,31	2,26	2,23	2,21	2,16	–	–	–	–	–	–
20	2,38	2,23	2,20	2,18	2,12	–	–	–	–	–	–
21	2,25	2,20	2,13	2,15	2,10	–	–	–	–	–	–
22	2,28	2,18	2,15	2,13	2,07	–	–	–	–	–	–
23	2,20	2,14	2,13	2,10	2,05	–	–	–	–	–	–
24	2,18	2,13	2,11	2,09	2,03	–	–	–	–	–	–
25	2,16	2,11	2,09	2,06	2,01	–	–	–	–	–	–



Таблица 5 – Критические значения RS-критерия

Уровень значимости  $\alpha = 0,05$ 

Количество наблюдений, $n$	Граница RS-критерия	
	нижняя	верхняя
10	2,67	3,69
15	2,96	4,14
20	3,18	4,49
25	3,34	4,71
30	3,47	4,89
35	3,59	5,03
40	3,67	5,17
45	3,74	5,28
50	3,82	5,37
55	3,88	5,46
60	3,93	5,54
65	3,97	5,62
70	4,02	5,70
75	4,06	5,77
80	4,09	5,83
85	4,12	5,88
90	4,14	5,92
95	4,15	5,94
100	4,16	5,96

Таблица 6 – Критические значения коэффициента циклической автокорреляции

Количество наблюдений, <i>n</i>	Положительные значения		Отрицательные значения	
	5 %	1 %	5 %	1 %
5	0,253	0,297	–0,753	–0,798
6	0,345	0,447	–0,708	–0,863
7	0,370	0,510	–0,674	–0,799
8	0,371	0,531	–0,625	–0,764
9	0,366	0,533	–0,593	–0,737
10	0,360	0,525	–0,564	–0,705
11	0,353	0,515	–0,539	–0,679
12	0,348	0,505	–0,516	–0,655
13	0,341	0,495	–0,497	–0,634
14	0,335	0,485	–0,479	–0,615
15	0,328	0,475	–0,462	–0,597
20	0,299	0,432	–0,399	–0,524
25	0,276	0,398	–0,356	–0,473
30	0,257	0,370	–0,325	–0,433
35	0,242	0,347	–0,300	–0,401
40	0,229	0,326	–0,279	–0,376
45	0,218	0,314	–0,262	–0,356
50	0,208	0,301	–0,248	–0,339
55	0,199	0,289	–0,236	–0,324
60	0,191	0,278	–0,225	–0,310
65	0,184	0,268	–0,216	–0,298
70	0,178	0,259	–0,207	–0,287
75	0,173	0,250	–0,199	–0,276

Таблица 7 – Критические значения статистики Дарбина–Уотсона

Уровень значимости  $\alpha = 0,05$ 

Количество наблюдений, $n$	Ранг автокорреляции									
	1		2		3		4		5	
	$d_n$	$d_e$	$d_n$	$d_e$	$d_n$	$d_e$	$d_n$	$d_e$	$d_n$	$d_e$
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	1,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	1,90	1,71	0,78	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	0,86	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	1,90
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	1,66	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	1,24	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	1,72	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

Навчальне видання

*Мельник Сергій Анатолійович  
Вайсруб Наталя Володимирівна  
Грібінюкова Вікторія Вікторівна*

## **Моделювання економічних, екологічних і соціальних процесів**

(Рос. мовою)

Коректор  
Технічний редактор

Н. М. Гіюк

План вид. 2013 р., поз. № 73.

Донецький національний університет  
83001, м. Донецьк, вул. Університетська, 24.  
Свідоцтво про внесення суб'єкта видавничої справи  
до Державного реєстру  
серія ДК № 1854 від 24.06.2004 р.