УЧЕБНЫЕ ИЗДАНИЯ ДЛЯ БАКАЛАВРОВ

А. И. Новиков

ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИ

УЧЕБНИК

«Дашков и К°»

Серия «Учебные издания для бакалавров»

А. И. Новиков

Экономико-математические методы и модели

Учебник

3-е издание

Рекомендовано ФГБОУ ВПО
«Государственный университет управления»
в качестве учебника для студентов
высших учебных заведений, обучающихся
по направлениям подготовки «Экономика» и «Менеджмент»
(уровень бакалавриата)

Министерство образования и науки Российской Федерации
ФГАУ «Федеральный институт развития образования»
Регистрационный номер рецензии 518 от 30 июня 2015 г.

Москва
Издательско-торговая корпорация «Дашков и К°»
2020

УДК 330.4
ББК 22.18
Н73

Автор:
А. И. Новиков — доктор физико-математических наук, профессор.

Рецензенты:
В. А. Волочиенко — доктор экономических наук, профессор;
И. И. Постников — доктор технических наук, профессор.

Новиков А. И.
Н73 Экономико-математические методы и модели: Учебник для бакалавров / А. И. Новиков. — 3-е изд. — М.: Издательско-торговая корпорация «Дашков и К°», 2020. — 532 с.

ISBN 978-5-394-03782-5

Учебник содержит основные разделы теории экономико-математических методов и моделей, подготовлен в соответствии с требованиями Федерального государственного образовательного стандарта.

Во всех главах учебника приведено большое число примеров с подробным решением.

Для студентов бакалавриата, обучающихся по направлениям подготовки «Экономика» и «Менеджмент», аспирантов, а также практических работников в области финансовой и экономической деятельности.

ISBN 978-5-394-03782-5

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Для изучения различных экономических явлений экономисты используют их упрощенные формальные описания, называемые экономическими моделями. При построении экономических моделей выявляются существенные факторы и отбрасываются несущественные для решения поставленной задачи детали.

В общем виде **модель** можно определить как условный образ (отражение) реального объекта (процесса), который создается для более глубокого изучения действительности. Такое отражение объекта может быть представлено схемой, эскизом, фотографией, моделью описательного характера в виде графиков и таблиц и т. д.

Метод исследования, базирующийся на разработке и использовании моделей, называется **моделированием**. Необходимость моделирования обусловлена сложностью, а порой и невозможностью прямого изучения реального объекта (процесса). Значительно доступнее создавать и изучать прообразы реальных объектов (процессов), т. е. модели. Можно сказать, что теоретическое знание о чем-либо, как правило, представляет собой совокупность различных моделей. Эти модели отражают существенные свойства реального объекта (процесса), хотя на самом деле действительность значительно содержательнее и богаче.

В управлении хозяйственными процессами наибольшее значение имеют прежде всего экономико-математические модели, часто объединяемые в системы моделей.

Экономико-математическая модель — это математическое описание экономического объекта или процесса с целью их исследования и управления ими. Она представляет собой концентрированное выражение общих взаимосвязей и закономерностей экономического явления в математической форме.

В настоящие время все большее применение находят математические методы исследования. Это способствует совершенствованию экономического анализа, его углублению и повышению его действенности.

Экономико-математические методы — обобщающее название комплекса экономических и математических научных дисциплин, объединенных для изучения экономики. В процессе использования экономико-математических методов в экономическом анализе осуществляется построение и изучение экономико-математических моделей, описывающих влияние отдельных факторов на обобщающие экономические показатели деятельности организаций.

Практическими задачами экономико-математического моделирования являются:

• анализ экономических объектов и процессов;

• экономическое прогнозирование;

• выработка управленческих решений.

Единой **классификации экономико-математических моделей** не существует, хотя можно выделить наиболее значимые их группы в зависимости от признака классификации.

По *целевому назначению* модели подразделяются:

• на теоретико-аналитические (используются в исследовании общих свойств и закономерностей экономических процессов);

• прикладные (применяются в решении конкретных экономических задач, таких как задачи экономического анализа, прогнозирования, управления).

По *учету фактора времени* различают модели:

• статические;

• динамические.

В статических моделях экономическая система описана в статике, применительно к одному определенному моменту времени. Динамические модели описывают экономическую систему в развитии.

По *цели создания и применения* различают модели:

• балансовые. В таких моделях отражается требование соответствия наличия ресурсов и их использования;

• эконометрические. Параметры таких моделей оцениваются с помощью методов математической статистики. Эконометрические модели используются для анализа и прогнозирования конкретных экономических процессов с использованием реальной статистической информации. Наиболее распространены модели, представляющие собой системы регрессионных уравнений;

• оптимизационные. Эти модели позволяют найти из множества возможных (альтернативных) вариантов наилучший вариант производства, распределения или потребления. Ограниченные ресурсы при этом будут использованы наиболее эффективным образом для достижения поставленной цели;

• сетевые. Такие модели наиболее широко используются в управлении проектами. Сетевая модель отображает комплекс работ (операций) и событий и их взаимосвязь во времени. Обычно сетевая модель предназначена для выполнения работ в такой последовательности, чтобы сроки выполнения проекта были минимальными. В этом случае ставится задача нахождения критического пути. Однако существуют и такие сетевые модели, которые ориентированы не на критерий времени, а, например, на минимизацию стоимости работ;

• систем массового обслуживания. Данные модели создаются для минимизации затрат времени на ожидание в очереди и времени простоев каналов обслуживания;

• имитационные (экспертные). Такие модели наряду с машинными решениями содержат блоки, где решения принимаются человеком (экспертом). Вместо непосредственного участия человека в принятии решений может использоваться база знаний. В этом случае ЭВМ, специализированное программное обеспечение, база данных и база знаний образуют экспертную систему. Экспертная система предназначена для решения одной или ряда задач методом имитации действий человека, эксперта в данной области.

Классификация экономико-математических методов сводится к классификации научных дисциплин, входящих в их состав. В составе экономико-математических методов можно выделить следующие разделы:

• *экономическая кибернетика*: системный анализ экономики, теория экономической информации и теория управляющих систем;

• *математическая статистика*. Экономические приложения данной дисциплины — выборочный метод, дисперсионный анализ, корреляционный анализ, регрессионный анализ, многомерный статистический анализ, факторный анализ, теория индексов и др.;

• *математическая экономика* и изучающая те же вопросы с количественной стороны *эконометрия*: теория экономического роста, теория производственных функций, межотраслевые балансы, национальные счета, анализ спроса и потребления, региональный и пространственный анализ, глобальное моделирование и др.;

• *методы принятия оптимальных решений*, в том числе исследование операций в экономике;

• *методы экспериментального изучения экономических явлений*: математические методы анализа и планирования экономических экспериментов, методы машинной имитации (имитационное моделирование), деловые игры.

Процесс экономико-математического моделирования можно разделить на несколько основных **этапов**.

1. Постановка экономической проблемы и ее качественный анализ. Этот этап включает выделение важнейших черт и свойств моделируемого объекта; изучение структуры объекта и основных зависимостей, связывающих его элементы; формулирование гипотез, объясняющих поведение и развитие объекта.

2. Построение математической модели. Это этап формализации экономической проблемы, формулирования ее в виде конкретных математических зависимостей и отношений (функций, уравнений, неравенств и т. д.). Как правило, сначала определяется тип математической модели, а затем уточняются детали.

3. Математический анализ модели. Целью этого этапа является выяснение общих свойств модели. Здесь применяются чисто математические приемы исследования. Если удается доказать, что математическая задача не имеет решения, то необходимость в последующей работе по первоначальному варианту модели отпадает, и следует скорректировать либо постановку экономической задачи, либо способы ее математической формализации. При аналитическом исследовании модели выясняются вопросы: единственно ли решение, какие переменные (неизвестные) могут входить в решение, каковы будут соотношения между ними, в каких пределах и в зависимости от каких исходных условий они изменяются, каковы тенденции их изменения и т. д.

4. Подготовка исходной информации. Математическое моделирование предъявляет жесткие требования к системе информации; при этом надо принимать во внимание не только принципиальную возможность подготовки информации требуемого качества, но и затраты на подготовку информационных массивов. В процессе подготовки информации используются методы теории вероятностей, теоретической и математической статистики для организации выборочных обследований, оценки достоверности данных и т.д. При системном экономико-математическом моделировании результаты функционирования одних моделей служат исходной информацией для других.

5. Численное решение. Этот этап включает разработку алгоритмов для численного решения задачи и непосредственное проведение расчетов. Исследование, проводимое численными методами, может существенно дополнить результаты аналитического исследования, а для многих моделей оно является единственно осуществимым.

6. Анализ численных результатов и их применение. На этом заключительном этапе встает вопрос о правильности и полноте результатов моделирования, о степени практической применимости последних. Математические методы проверки могут выявлять некорректные построения модели и тем самым сужать класс потенциально правильных моделей. Неформальный анализ теоретических выводов и численных результатов, получаемых с помощью модели, сопоставление их с имеющимися знаниями и фактами действительности также позволяют обнаруживать недостатки постановки экономической задачи, сконструированной математической модели, ее информационного и математического обеспечения.

Глава 1.
ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

1.1. Задачи линейного программирования

Линейное программирование (ЛП) — наука о методах исследования и отыскания наибольших и наименьших значений линейной функции, на неизвестные которой наложены линейные ограничения. Функция, наибольшее и наименьшее значения которой отыскиваются, называется *целевой функцией.*

Построим математические модели простейших экономических задач.

Пример 1.1 (задача об использовании ресурсов). Для изготовления двух видов продукции P_1, P_2 используются три вида ресурсов S_1, S_2, S_3. Запасы ресурсов, затраты ресурсов на единицу продукции, а также цены единицы продукции приведены в табл. 1.1.

Таблица 1.1

Условия примера 1.1

Ресурсы	Затраты ресурсов на ед. продукции, ед.		Запасы ресурсов (b), ед.
	P_1	P_2	
S_1	2	4	2000
S_2	4	1	1400
S_3	2	1	800
Цена ед. продукции (c), усл. ед.	40	60	

Требуется построить план производства, максимизирующий доход.

Построим математическую модель задачи, определив в ней переменные, ограничения и целевую функцию.

Переменные: x_1, x_2 — количество единиц выпускаемой продукции P_1, P_2 соответственно (объемы производства).

Ограничения. Ограничение на расход ресурсов можно записать в следующем виде

$$\begin{pmatrix}\text{Расход}\\ \text{ресурсов}\end{pmatrix} \le \begin{pmatrix}\text{Запас}\\ \text{ресурсов}\end{pmatrix}.$$

Это приводит к следующей системе ограничений:

$$\begin{cases} 2x_1 + 4x_2 \le 2000, \\ 4x_1 + x_2 \le 1400, \\ 2x_1 + x_2 \le 800. \end{cases}$$

Кроме того, переменные должны быть неотрицательными.

$x_1 \ge 0$, $x_2 \ge 0$ — условие неотрицательности.

Целевая функция — суммарный доход от реализации всей продукции:

$$F(x) = 40x_1 + 60x_2, \; x = (x_1, x_2).$$

Требуется найти такие неотрицательные переменные x_1, x_2, удовлетворяющие ограничениям, при которых суммарный доход максимален, т. е. $\max F(x)$.

Экономико-математическая модель задачи кратко записывается в виде

$$F(x) = 40x_1 + 60x_2 \to \max$$

при ограничениях

$$\begin{cases} 2x_1 + 4x_2 \le 2000, \\ 4x_1 + x_2 \le 1400, \\ 2x_1 + x_2 \le 800; \end{cases}$$

$x_1 \ge 0$, $x_2 \ge 0$ — условие неотрицательности.

Пример 1.2 (задача составления рациона). Пищевой рацион должен содержать не менее b_1, b_2, b_3 питательных веществ S_1, S_2, S_3. Для составления пищевого рациона используются два вида продуктов питания P_1, P_2. Содержание питательных веществ в единице каждого продукта и стоимость единицы продукта (цена) приведены в табл. 1.2.

Таблица 1.2

Условия примера 1.2

Питательные вещества	**Содержание питательных веществ в продукте, ед.**		**Норма (b), ед.**
	P_1	P_2	
S_1	3	1	9
S_2	1	2	8
S_3	1	6	12
Цена ед. продукта (c), усл. ед.	4	6	

Требуется так составить пищевой рацион, чтобы обеспечить норму содержания питательных веществ при его минимальной стоимости.

Построим математическую модель задачи, определив в ней переменные, ограничения и целевую функцию.

Переменные: x_1, x_2 — количество единиц соответствующего вида продукта P_1, P_2.

Ограничения. Ограничение на содержание питательных веществ в рационе можно записать в следующем виде

$$\begin{pmatrix} \text{Содержание} \\ \text{питательного} \\ \text{вещества} \end{pmatrix} \geq \begin{pmatrix} \text{Норма} \\ \text{вещества} \end{pmatrix}.$$

Это приводит к следующей системе ограничений:

$$\begin{cases} 3x_1 + x_2 \geq 9, \\ x_1 + 2x_2 \geq 8, \\ x_1 + 6x_2 \geq 12. \end{cases}$$

Кроме того, переменные должны быть неотрицательными.

$x_1 \geq 0$, $x_2 \geq 0$ — условие неотрицательности.

Целевая функция — общая стоимость рациона:

$$F(x) = 4x_1 + 6x_2, \ x = (x_1, x_2).$$

Требуется минимизировать стоимость рациона при заданных ограничениях, т. е. $\min F(x)$.

Экономико-математическая модель задачи кратко записывается в виде

$$F(x) = 4x_1 + 6x_2 \to \min$$

при ограничениях

$$\begin{cases} 3x_1 + \ x_2 \geq 9, \\ \ x_1 + 2x_2 \geq 8, \\ \ x_1 + 6x_2 \geq 12; \end{cases}$$

$x_1 \geq 0$, $x_2 \geq 0$ — условие неотрицательности.

1.2. Общая задача линейного программирования

Общая задача линейного программирования (ОЗЛП) ставится следующим образом. Необходимо найти экстремум (максимум или минимум) линейной целевой функции

$$F(x) = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max(\min)$$

при ограничениях

$$\begin{cases} a_{11}x_1 + \ a_{12}x_2 + \ldots + a_{1n}x_n \leq b_1, \\ a_{21}x_1 + \ a_{22}x_2 + \ldots + a_{2n}x_n \leq b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \leq b_m; \end{cases}$$

$x_1, x_2, \ldots, x_n \geq 0$ — условие неотрицательности,

где x_j — переменные;

a_{ij}, b_i, c_j ($i = 1, 2 \ldots m$; $j = 1, 2 \ldots n$) — заданные постоянные величины.

В системе ограничений неравенства могут быть направлены в ту или иную сторону ($\leq, \geq$).

Допустимым решением ОЗЛП называется вектор $x = (x_1, x_2, \ldots, x_n)$, удовлетворяющий системе ограничений и условию неотрицательности.

Областью допустимых решений (ОДР) называется множество всех допустимых решений ОЗЛП.

Оптимальным решением ОЗЛП называется допустимое решение, при котором целевая функция достигает своего экстремального значения.

Геометрическая интерпретация ОЗЛП

Рассмотрим ОЗЛП с *двумя переменными* (на плоскости).

$$F(x) = c_1x_1 + c_2x_2 \to \max(\min)$$

при ограничениях

$$\begin{cases} a_{11}x_1 + a_{12}x_2 \leq b_1, \\ a_{21}x_1 + a_{22}x_2 \leq b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 \leq b_m; \end{cases}$$

$x_1, x_2 \geq 0$ — условие неотрицательности.

Свойства решений ОЗЛП тесно связаны со свойствами выпуклых множеств. Рассмотрим на плоскости множество точек (x_1, x_2).

Множество точек на плоскости называется *выпуклым*, если оно вместе с любыми своими двумя точками содержит и весь отрезок, соединяющий их; в противном случае оно называется *невыпуклым*. Примеры выпуклых и невыпуклых множеств показаны на рис 1.1.

Точка A называется *внутренней точкой* выпуклого множества, если в сколь угодно малой окрестности этой точки содержатся только точки этого множества.

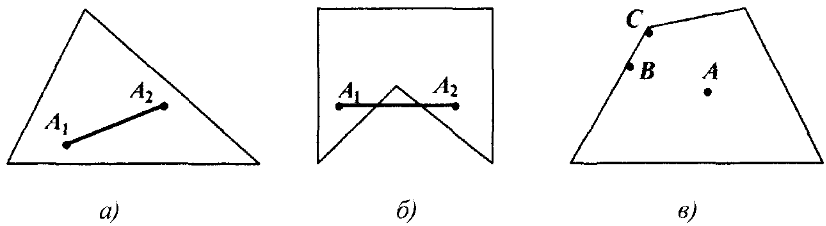


Рис. 1.1. Выпуклые (*а*, *в*) и невыпуклые (*б*) множества

Точка B называется *граничной точкой* выпуклого множества, если в сколь угодно малой окрестности этой точки содержатся как точки данного множества, так и не принадлежащие ему.

Точка C называется *угловой точкой* выпуклого множества, если она является граничной и не лежит внутри отрезка, соединяющего две другие точки этого множества.

Множество называется *замкнутым*, если оно включает все свои граничные точки.

Множество называется *ограниченным*, если существует окружность радиуса конечной длины с центром в любой точке множества, которая полностью содержит в себе данное множество; в противном случае оно называется *неограниченным*.

Пересечением выпуклых множеств называется множество, представляющее общую часть данных множеств.

Свойство. Пересечение выпуклых множеств есть выпуклое множество.

Выпуклым многоугольником называется выпуклое замкнутое ограниченное множество на плоскости, имеющее конечное число угловых точек.

Полуплоскостью называется множество точек на плоскости, координаты которых удовлетворяют неравенству

$$a_1x_1 + a_2x_2 \le b .$$

Уравнение $a_1x_1 + a_2x_2 = b$ является *граничной прямой* полуплоскости.

Граничная прямая делит плоскость на две полуплоскости. Для того чтобы определить, на какую сторону от граничной прямой расположена данная полуплоскость, надо взять произвольную точку на плоскости и подставить координаты этой точки в неравенство. Если неравенство справедливо, то полуплоскость обращена в сторону этой точки, иначе — в противоположную. Направление полуплоскости на рисунках штрихуется.

Полуплоскости являются выпуклыми множествами. Каждое из неравенств системы ограничений определяет полуплоскость. Система ограничений в виде неравенств (пересечение) образует выпуклое множество, которое называется *многоугольником решения задачи*. Стороны этого многоугольника лежат на граничных прямых, а угловые точки определяются как точки пересечения смежных граничных прямых.

Геометрически задача ЛП представляет собой отыскание такой точки многоугольника решений, координаты которой доставляют линейной функции $F(x)$ экстремум, причем допустимыми решениями служат все точки многоугольника решений.

Теорема. Если задача ЛП имеет оптимальное решение, то оно достигается в одной из угловых точек многоугольника решений.

Графическое решение задач ЛП

Наиболее простым и наглядным методом ЛП является графический метод. Он применяется для решения задач ЛП с двумя переменными (на плоскости).

Рассмотрим задачу об использовании ресурсов.

$$F(x) = 40x_1 + 60x_2 \to \max$$

при ограничениях

$$\begin{cases} 2x_1 + 4x_2 \le 2000, \\ 4x_1 + x_2 \le 1400, \\ 2x_1 + x_2 \le 800; \end{cases}$$

$x_1 \ge 0$, $x_2 \ge 0$ — условие неотрицательности.

Решим данную задачу графическим методом.

▼Каждое неравенство системы ограничений определяет полуплоскость с граничной прямой:

$$L_1\colon 2x_1 + 4x_2 = 2000;$$
$$L_2\colon 4x_1 + x_2 = 1400;$$
$$L_3\colon 2x_1 + x_2 = 800.$$

Условие неотрицательности определяет полуплоскости с граничными прямыми $x_1 = 0$, $x_2 = 0$.

На рис. 1.2 показаны граничные прямые и ОДР (заштриховано).

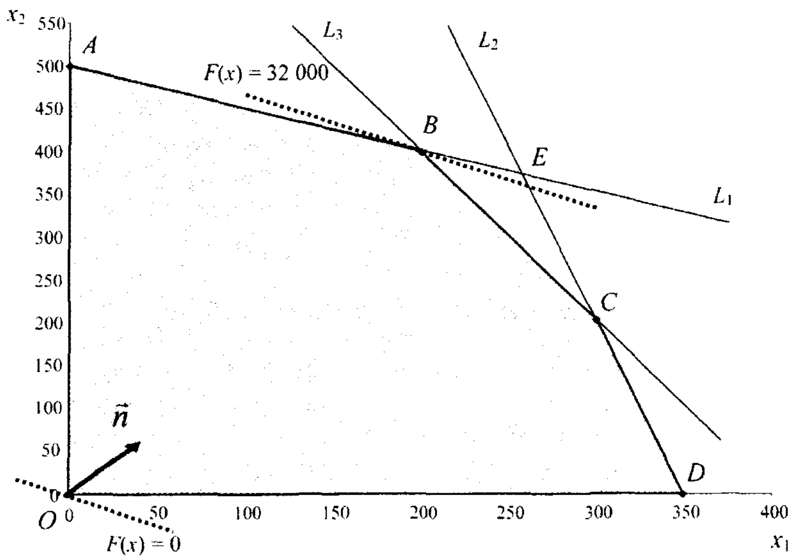


Рис. 1.2. Оптимальное решение модели (задача об использовании ресурсов)

Среди точек этого многоугольника нужно найти такую точку, в которой линейная функция $F = 40x_1 + 60x_2$ принимает максимальное значение.

Линией уровня функции $F(x)$ называется множество точек (x_1, x_2) на плоскости, в которых функция принимает одно и то же значение, т. е. $F(x)= C$.

Уравнение линии уровня целевой функции есть $40x_1 + 60x_2 = C$ (семейство параллельных прямых).

Построим вектор $\vec{n} = (40, 60)$. Вектор $\vec{n}$ указывает направление наибольшего возрастания целевой функции $F(x)$ и перпендикулярен линиям уровня $F(x)= C$ (из-за особенностей масштаба на рисунке не видно). Проведем нулевую линию уровня $F(x) = 0$, или $40x_1 + 60x_2 = 0$, перпендикулярно вектору $\vec{n}$.

Перемещая нулевую линию уровня параллельно самой себе в направлении вектора $\vec{n}$ до тех пор, пока у нее не окажется только одна общая точка с многоугольником решения (угловая точка B), получим оптимальное решение задачи ЛП, соответствующее максимальному значению целевой функции.

Координаты точки B определяются из решения системы

$$\begin{cases} 2x_1 + 4x_2 = 2000, \\ 2x_1 + \quad x_2 = 800. \end{cases} \Rightarrow \quad B = (200;\, 400).$$

Таким образом, графический способ решения задачи дает оптимальное решение:

$$x = (200; 400)\,, \; F_{\max} = F(B) = 40 \cdot 200 + 60 \cdot 400 = 32\,000.$$

Значит, чтобы получить максимальный доход в размере 32 000 усл. ед., необходимо запланировать производство 200 единиц продукции P_1 и 400 единиц продукции P_2.▲

Статус ресурсов

Ограничения линейной модели классифицируются на связывающие и несвязывающие.

Граничная прямая, представляющая связывающее ограничение, проходит через оптимальную точку; в противном случае ограничение является не связывающим. На рис. 1.2 связывающими ограничениями являются ограничения, представленными прямыми L_1, L_3, а несвязывающее ограничение представлено прямой L_2.

Статус ресурсов (дефицитный или недефицитный) устанавливается в зависимости от того, полное или частичное их использование предусматривает оптимальное решение задачи. Если ограничение является связывающим, то этот ресурс относится к дефицитному (используется полностью). Если ограничение является несвязывающим, то ресурс относится к недефицитному, следовательно, ресурсы 1, 3 являются дефицитными, а ресурс 2 — недефицитным.

Графический анализ чувствительности

На практике параметры модели (коэффициенты целевой функции и неравенств-ограничений) с течением времени меняют свои значения. Рассмотрим влияние изменения параметров модели на полученное оптимальное решение задачи ЛП. Такое исследование называется анализом чувствительности.

Изменение коэффициентов целевой функции. Изменение значений коэффициентов c_1 и c_2 приводит к изменению угла наклона прямой $F(x)$. Графический способ решения задачи ЛП показывает, что это может привести к изменению оптимального решения: оно будет достигаться в другой угловой точке пространства решений. Вместе с тем, очевидно, существуют интервалы изменения коэффициентов c_1 и c_2, при которых текущее оптимальное решение сохраняется.

Задача анализа чувствительности и состоит в получении такой информации. В частности, представляет интерес определение интервала оптимальности для отношения c_1/c_2 (или, что то же самое, для c_2/c_1); если значение отношения c_1/c_2 не выходит за пределы этого интервала, то оптимальное решение в данной модели сохраняется неизменным.

Применим процедуру анализа чувствительности к задаче об использовании ресурсов. На рис. 1.3 видно, что функция $F(x) = 40x_1 + 60x_2$ достигает максимального значения в угловой точке B.

При изменении коэффициентов целевой функции $F(x) = c_1x_1 + c_2x_2$ точка B останется точкой оптимального решения до тех пор, пока угол наклона прямой $F(x)$ будет лежать между углами наклона двух прямых, пересечением которых является точка B. Этими прямыми являются L_1: $2x_1 + 4x_2 = 2000$ (ограничение на сырье b_1) и L_3: $2x_1 + x_2 = 800$ (ограничение на сырье b_3). Алгебраически это можно записать следующим образом:

$$\frac{1}{2} < \frac{c_2}{c_1} < \frac{4}{2},\ c_1 \neq 0, \text{ или } \frac{2}{4} < \frac{c_1}{c_2} < \frac{2}{1},\ c_2 \neq 0.$$

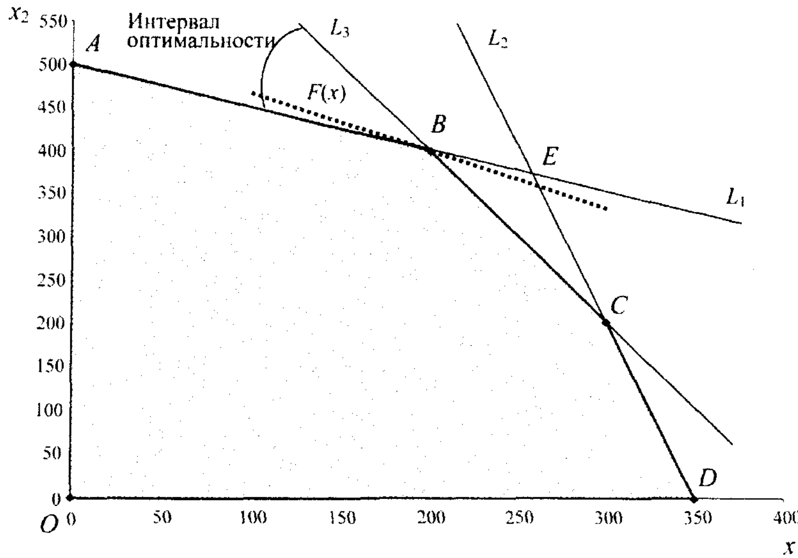


Рис. 1.3. Интервал оптимальности (задача об использовании ресурсов)

Если зафиксировать значение $c_1 = 40$, то из первого неравенства получим интервал оптимальности для коэффициента c_2: $20 \leq c_2 \leq 80$. Если зафиксировать значение $c_2 = 60$, то из второго неравенства получим интервал оптимальности для коэффициента c_1: $30 \leq c_1 \leq 120$.

Изменение ограничений, накладываемых на ресурсы. Изучим чувствительность оптимального решения к изменению ограничений, налагаемых на ресурсы.

Рассмотрим ограничение на третий ресурс. В данной задаче оптимальное решение достигается в угловой точке B, являющейся точкой пересечения прямых, соответствующих ограничениям на ресурсы b_1 и b_3 (рис. 1.4).

При изменении уровня доступности ресурса b_3 (увеличение или уменьшение текущего уровня, равного 800 усл. ед.) точка B оптимального решения «плывет» вдоль отрезка AE. Любое изменение уровня доступности ресурса b_3, приводящее к выходу точки пересечения B из этого отрезка, ведет к неосуществимости оптимального решения в точке B. Поэтому можно сказать, что

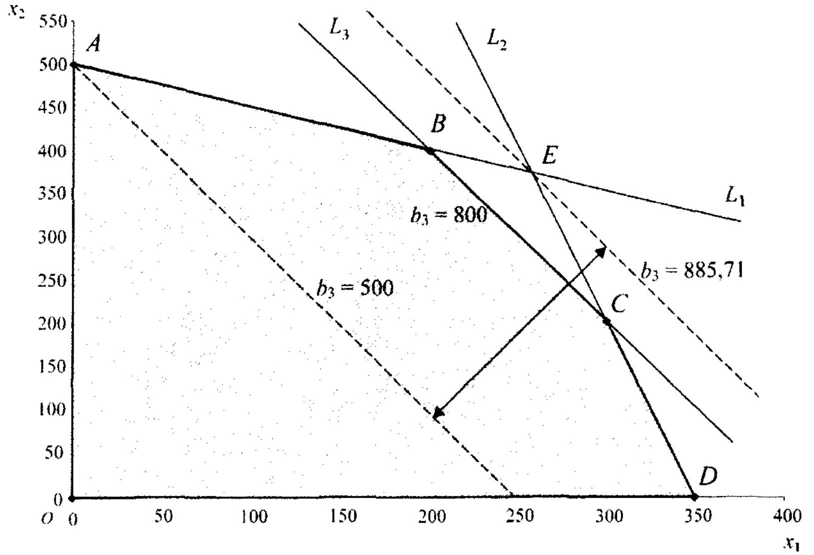


Рис. 1.4. Интервал осуществимости для ресурса b_3 (задача об использовании ресурсов)

конечные точки $A = (0;\ 500)$ и $E = (257{,}1;\ 371{,}4)$ отрезка AE определяют интервал осуществимости для ресурса b_3. Количество ресурса b_3, соответствующее точке A, равно $2x_1 + x_2 = 2 \cdot 0 + 500 = 500$ усл. ед. Аналогично количество ресурса, соответствующее точке E, равно $2 \cdot 257{,}1 + 371{,}4 = 885{,}7$ усл. ед. Таким образом, интервал осуществимости для ресурса b_3 составляет $500 < b_3 < 885{,}7$ усл. ед. Если мы определим b_3 как $b_3 = 800 + D_3$, где D_3 — отклонение количества ресурса b_3 от текущего уровня 800 усл. ед., то последние неравенства можно переписать как $-300 < D_3 < 85{,}7$. Это означает, что текущий уровень ресурса b_3 может быть уменьшен не более чем на 300 усл. ед. и увеличен не более чем на 85,7 усл. ед. В этом случае гарантируется, что оптимальное решение будет достигаться в точке B.

Определим стоимость единицы ресурса как отношение изменения значения целевой функции к изменению доступного количества ресурсов.

Вычислим стоимость единицы ресурса b_3. При изменении количества третьего ресурса от 500 до 885,7 усл. ед. (интервал осуществимости для этого ресурса) значения целевой функции

$F(x)$ будут соответствовать положению точки B на отрезке AE. Обозначив через y_3 стоимость единицы третьего ресурса, получим следующую формулу:

$$y_3 = \frac{\text{изменение значения } F(x) \text{ при перемещении } B \text{ от } A \text{ до } E}{\text{изменение количества } b_3 \text{ при перемещении } B \text{ от } A \text{ до } E}.$$

Если точка B совпадает с точкой $A = (0;\ 500)$, то $F(A) = 40 \cdot 0 + 60 \cdot 500 = 30\ 000$, если же точка B совпадает с точкой $E = (257{,}1;\ 371{,}4)$, тогда $F(E) = 40 \cdot 257{,}1 + 60 \cdot 371{,}4 = 32\ 568$. Отсюда следует, что

$$y_3 = \frac{32568 - 30000}{885{,}7 - 500} = 6{,}66 \text{ — стоимость единицы ресурса } b_3.$$

Подобное рассмотрение можно провести и для первых двух ресурсов. В результате получим: интервал осуществимости для первого дефицитного ресурса — $1400 < b_1 < 3200$ усл. ед., стоимость единицы ресурса b_1 — 13,3; интервал осуществимости для второго недефицитного ресурса — $1200 < b_2 < \infty$ усл. ед., стоимость единицы ресурса b_2 — 0.

Упражнение 1.1. Решить графическим методом задачу составления рациона, дать экономическую интерпретацию полученного решения.

Каноническая форма ОЗЛП

Существуют различные канонические формы ОЗЛП, отличающиеся между собой только типом экстремума (max, min). Для простоты изложения в учебнике общая задача ЛП в канонической (стандартной) форме ставится следующим образом:

$$F(x) = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \min$$

при ограничениях

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m; \end{cases}$$

$x_1, x_2, \ldots, x_n \geq 0$ — условие неотрицательности.

При канонической форме ОЗЛП:

• значения всех переменных модели неотрицательные;

• все ограничения записываются в виде равенств с неотрицательной правой частью, т. е. $b_i \geq 0$;

• целевая функция подлежит минимизации.

Любую линейную модель можно привести к канонической. От задачи ЛП с ограничениями-неравенствами можно перейти к задаче ЛП с ограничениями-равенствами с помощью введения дополнительных переменных:

• остаточной переменной (со знаком «+» в случае ограничений типа $\leq$);

• избыточной переменной (со знаком «–» в случае ограничений типа $\geq$).

Например,

$$\begin{cases} x_1 + 2x_2 \leq 6, \\ 3x_1 - 4x_2 \geq 2 \end{cases} \sim \begin{cases} x_1 + 2x_2 + x_3 = 6, \\ 3x_1 - 4x_2 - x_4 = 2, \\ x_3, x_4 \geq 0, \end{cases}$$

где ~ — знак эквивалентности.

Кроме того, правую часть в системе ограничений-равенств всегда можно сделать неотрицательной, умножая обе части уравнения на (–1).

Например, $2x_1 - 4x_2 = -5 \sim -2x_1 + 4x_2 = 5$.

Максимизация функции эквивалентна минимизации этой же функции с противоположным знаком, т. е. $\max F(x) = \min(-F(x))$.

Например,

$F(x) = 40x_1 + 60x_2 \to \max \sim F(x) = -40x_1 - 60x_2 \to \to \min$.

1.3. Симплексный метод

При анализе графического метода решения задач ЛП получено, что оптимальному решению всегда соответствует одна из угловых точек ОДР. Этот результат положен в основу построения симплекс-метода.

Переход от геометрического способа решения задачи ЛП к симплекс-методу лежит через алгебраическое описание угловых точек пространства решений. Для реализации этого перехода сначала надо привести задачу ЛП к канонической (стандартной) форме.

Каноническая форма записи задачи ЛП необходима, потому что она предоставляет возможность получить базисное решение. Это базисное решение полностью определяет все (геометрические) угловые точки пространства решений. Симплексный метод позволяет эффективно найти оптимальное решение среди всех базисных.

Определение базисных решений

Ограничения ОЗЛП в канонической форме представляют собой систему m линейных уравнений с n неизвестными:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m. \end{cases}$$

Будем считать, что система совместна (имеет хотя бы одно решение) и ее уравнения линейно независимые.

В задачах ЛП обычно $m < n$, т. е. число уравнений системы меньше числа неизвестных. Таким образом, имеем неопределенную систему уравнений.

Разделим n переменных на две группы: $(n - m)$ переменных, которые положим равными нулю, и оставшиеся m переменных, значения которых определяются как решение системы из m линейных уравнений с m неизвестными.

Если это решение *единственное*, тогда соответствующие m переменных называются *базисными*, а остальные $(n - m)$ нулевых переменных — *свободными*. В этом случае результирующие значения переменных составляют *базисное решение* (план).

Каждому разбиению на базисные и свободные переменные соответствует одно базисное решение, а количество способов разбиения не превышает величины $C_n^m = \dfrac{n!}{m!(n-m)!}$.

Пример 1.3. Найти все базисные решения системы

$$\begin{cases} 2x_1 - 3x_2 + 4x_3 + 6x_4 = 3, \\ 3x_1 - 4x_2 + 6x_3 + 8x_4 = 5. \end{cases}$$

▼Поскольку $m = 2 < n = 4$, то существуют две базисные и две свободные переменные.

Найдем все базисные решения путем перебора кандидатов в базисные переменные.

Пусть x_1, x_2 — базисные переменные, x_3, x_4 — свободные.

Положим $x_3 = 0$, $x_4 = 0$, тогда

$\begin{cases} 2x_1 - 3x_2 = 3, \\ 3x_1 - 4x_2 = 5; \end{cases}$ $\begin{vmatrix} 2 & -3 \\ 3 & -4 \end{vmatrix} = 1 \neq 0 \Rightarrow$ единственное решение $x_1 = 3$, $x_2 = 1$.

Базисное решение $x = (3; 1; 0; 0)$.

Пусть x_1, x_3 — базисные переменные, x_2, x_4 — свободные.

Положим $x_2 = 0$, $x_4 = 0$, тогда

$\begin{cases} 2x_1 + 4x_3 = 3, \\ 3x_1 + 6x_3 = 5; \end{cases}$ $\begin{vmatrix} 2 & 4 \\ 3 & 6 \end{vmatrix} = 0 \Rightarrow$ единственного решения нет $\Rightarrow$ $\Rightarrow$ x_1, x_3 — не базисные переменные.

Пусть x_1, x_4 — базисные переменные, x_2, x_3 — свободные.

Положим $x_2 = 0$, $x_3 = 0$, тогда

$\begin{cases} 2x_1 + 6x_4 = 3, \\ 3x_1 + 8x_4 = 5; \end{cases}$ $\begin{vmatrix} 2 & 6 \\ 3 & 8 \end{vmatrix} = -2 \neq 0 \Rightarrow$ единственное решение $x_1 = 3$, $x_4 = -\dfrac{1}{2}$

Базисное решение $x = \left(3; 0; 0; -\dfrac{1}{2}\right)$.

Пусть x_2, x_3 — базисные переменные, x_1, x_4 — свободные.

Положим $x_1 = 0$, $x_4 = 0$, тогда

$$\begin{cases} -3x_2 + 4x_3 = 3, \\ -4x_2 + 6x_3 = 5; \end{cases} \quad \begin{vmatrix} -3 & 4 \\ -4 & 6 \end{vmatrix} = -2 \neq 0 \Rightarrow \text{ единственное решение}$$

$x_2 = 1$, $x_3 = \dfrac{3}{2}$

Базисное решение $x = (0; 1; \frac{3}{2}; 0)$.

Пусть x_2, x_4 — базисные переменные, x_1, x_3 — свободные.

Положим $x_1 = 0$, $x_3 = 0$, тогда

$$\begin{cases} -3x_2 + 6x_4 = 3, \\ -4x_2 + 8x_4 = 5; \end{cases} \quad \begin{vmatrix} -3 & 6 \\ -4 & 8 \end{vmatrix} = 0 \Rightarrow \text{единственного решения нет} \Rightarrow$$

$\Rightarrow x_2, x_4$ — не базисные переменные.

Пусть x_3, x_4 — базисные переменные, x_1, x_2 — свободные.

Положим $x_1 = 0$, $x_2 = 0$, тогда

$$\begin{cases} 4x_3 + 6x_4 = 3, \\ 6x_3 + 8x_4 = 5; \end{cases} \quad \begin{vmatrix} 4 & 6 \\ 6 & 8 \end{vmatrix} = -4 \neq 0 \Rightarrow \text{единственное решение } x_3 = \frac{3}{2},$$

$x_4 = -\dfrac{1}{2}$.

Базисное решение $x = \left(0; 0; \frac{3}{2}; -\frac{1}{2}\right)$. ▲

Процедура простого перебора всех базисных решений для задачи ЛП не эффективна. Алгоритм симплекс-метода находит оптимальное решение, рассматривая ограниченное количество допустимых базисных решений.

Симплексное преобразование

Базисное решение можно получить с помощью симплексного преобразования, основанного на методе последовательного исключения переменных (метод Жордана — Гаусса). В отличие от метода Гаусса, в методе Жордана — Гаусса на каждом шаге одна переменная исключается из всех уравнений, кроме одного, т. е. не только из всех последующих уравнений, но и из всех предыдущих.

Исходные данные системы ограничений-равенств записывают в виде симплексной таблицы, в которой указаны коэффициенты при неизвестных и свободные члены (табл. 1.3).

Таблица 1.3

Симплексная таблица

Базис	x_1	x_2	**...**	x_n	b	**Базисное решение**
	a_{11} a_{21} ... a_{m1}	a_{12} a_{22} ... a_{m2}	... ...	a_{1n} a_{2n} ... a_{mn}	b_1 b_2 ... b_m	

Базисным переменным соответствуют единичные вектора в столбцах таблицы при этих переменных; полный набор базисных переменных определяет базисное решение. В левом столбце табл. 1.3 «Базис» записываются базисные переменные; в правом — базисное решение.

1. Выделяют разрешающий элемент $a_{ij} \neq 0$. Строку и столбец, где стоит разрешающий элемент, называют *разрешающей строкой* и *разрешающим столбцом*.

2. Элементы разрешающей строки в новой симплексной таблице получаются делением их на разрешающий элемент.

3. Все элементы разрешающего столбца, кроме пересчитанного элемента, равного единице, становятся равны нулю.

4. Все остальные элементы новой таблицы вычисляются **по правилу прямоугольника**: мысленно выделяют прямоугольник, в котором элемент, подлежащий пересчету, и разрешающий элемент образуют главную диагональ. Значение пересчитанного элемента равно разности произведений элементов, стоящих на главной и побочной диагонали, деленной на разрешающий элемент (рис. 1.5):

$$a'_{lk} = \frac{a_{ij}a_{lk} - a_{lj}a_{ik}}{a_{ij}}.$$

$$\begin{tikzcd} a_{ij} \arrow[r, dash] \arrow[d, dash] & a_{ik} \arrow[d, dash] \\ a_{lj} \arrow[r, dash] & \boxed{a_{lk}} \end{tikzcd}$$

Рис. 1.5. Схема пересчета элемента a_{lk}

Процедуры (1–4) повторяются до тех пор, пока все строки не перебывали по одному разу в качестве разрешающей.

Пример 1.4. С помощью симплексных преобразований найти какое-либо базисное решение системы уравнений

$$\begin{cases} 2x_1 - 3x_2 + 4x_3 + 6x_4 = 3, \\ 3x_1 - 4x_2 + 6x_3 + 8x_4 = 5. \end{cases}$$

▼Составляем симплексную таблицу и последовательно производим симплексные преобразования (табл. 1.4).

Таблица 1.4

Симплексные таблицы для примера 1.4

Базис	x_1	x_2	x_3	x_4	b	Базисное решение
	(2)	–3	4	6	3	
	3	–4	6	8	5	
x_1	1	–3/2	2	3	3/2	
	0	(1/2)	0	–1	1/2	
x_1	1	0	2	0	3	(3; 1; 0; 0)
x_2	0	1	0	–2	1	

Пояснение.

В начальной таблице нет базисных переменных, поскольку ни один столбец при переменных не является единичным.

В первой строке выбираем разрешающий элемент (2). Строка и столбец, где стоит разрешающий элемент (2), являются разрешающей строкой и разрешающим столбцом.

Элементы разрешающей строки в новой симплексной таблице получаются делением их на разрешающий элемент (2). Все элементы разрешающего столбца, кроме пересчитанного элемента, равного единице, становятся равны нулю. Все остальные элементы матрицы определяются по правилу прямоугольника:

$$\frac{2\cdot(-4)-3\cdot(-3)}{2}=1/2;\ \frac{2\cdot 6-3\cdot 4}{2}=0;$$
$$\frac{2\cdot 8-3\cdot 6}{2}=-1;\ \frac{2\cdot 5-3\cdot 3}{2}=1/2.$$

В новой таблице столбец x_1 — единичный, следовательно, переменная x_1 — базисная. Запишем ее в базис.

Производим симплексное преобразование полученной таблицы.

В этой таблице разрешающей может быть только вторая строка, в которой выбираем разрешающий элемент (1/2).

Элементы разрешающей строки в новой симплексной таблице получаются делением их на разрешающий элемент (1/2). Все элементы разрешающего столбца, кроме пересчитанного элемента, равного единице, становятся равны нулю. Все остальные элементы матрицы определяются по правилу прямоугольника.

$$\frac{1/2\cdot 1-(-3/2)\cdot 0}{1/2}=1;\ \frac{1/2\cdot 2-(-3/2)\cdot 0}{1/2}=2;$$
$$\frac{1/2\cdot 3-(-3/2)\cdot(-1)}{1/2}=0;\ \frac{1/2\cdot 3/2-(-3/2)\cdot 1/2}{1/2}=3.$$

В результате преобразования получили новую таблицу, в которой столбцы x_1, x_2 — единичные, следовательно, переменные x_1, x_2 — базисные. Запишем их в базис. Получили полный набор базисных переменных, которые определяют базисное решение $x=(3;\ 1;\ 0;\ 0)$. ▲

Нахождение базисного решения системы удобно производить в *Microsoft Excel* (рис. 1.6).

Поясним ввод формул в соответствующие ячейки таблицы.

B2:F3 — размещены данные начальной таблицы.

B2 — выбран разрешающий элемент.

B4 — вводим формулу =B2/$B$2. «Протягиваем» ее до F4 включительно.

	A	B	C	D	E	F	G
1	Базис	x_1	x_2	x_3	x_4	b	Базисное решение
2		**2**	-3	4	6	3	
3		3	**-4**	6	8	**5**	
4	x_1	1	-1,5	2	3	1,5	
5		0	**0,5**	0	-1	0,5	
6	x_1	1	0	2	0	3	
7	x_2	0	1	0	**-2**	1	(3; 1; 0; 0)

Рис. 1.6. Вид листа *Excel* при расчете симплексных преобразований примера 1.4

B5 — вводим формулу =0.

C5 — вводим формулу =($B$2*C3-$B$3*C2)/$B$2. «Протягиваем» ее до F5 включительно.

A4 — вводим базисную переменную x_1.

C5 — выбран разрешающий элемент.

B7 — вводим формулу =B5/$C$5. «Протягиваем» ее до F7 включительно.

B6 — вводим формулу =(B4*$C$5-B5*$C$4)/$C$5. «Протягиваем» ее до F6 включительно.

A6 и A7 — вводим базисные переменные x_1 и x_2 соответственно.

G6 — записываем базисное решение (3; 1; 0; 0).

Метод последовательного однократного замещения

Если одно базисное решение найдено, то для нахождения нового базисного решения поступают следующим образом. Одну из свободных переменных переводят в базисную (включаемая переменная), а соответствующую базисную переменную переводят в свободную (исключаемая переменная).

Столбец симплексной таблицы, связанный с *включаемой переменной*, является *разрешающим столбцом*, а строка, соответствующая *исключаемой переменной*, является *разрешающей строкой*.

Производится симплексное преобразование, которое приводит к замещению только одной базисной переменной. Последовательно выполняя преобразования однократного замещения, находим все базисные решения.

Пример 1.5. Зная одно базисное решение (пример 1.4), найти методом однократного замещения все базисные решения.

▼Нахождение всех базисных решений показано в такой последовательности симплексных таблиц (табл. 1.5).

Таблица 1.5

Симплексные таблицы для примера 1.5

Базис	x_1	x_2	x_3	x_4	b	**Базисное решение**
x_1 x_2	1 0	0 1	2 0	0 (–2)	3 1	(3; 1; 0; 0)
x_1 x_4	1 0	0 –1/2	(2) 0	0 1	3 –1/2	(3; 0; 0; –1/2)
x_3 x_4	1/2 0	0 (–1/2)	1 0	0 1	3/2 –1/2	(0; 0; 3/2; –1/2)
x_3 x_2	1/2 0	0 1	1 0	0 –2	3/2 1	(0; 1; 3/2; 0)

▲

В экономических задачах существенны только неотрицательные переменные, т. е. $x_1, x_2, ..., x_n \geq 0$ — условие неотрицательности.

Базисное решение системы, все координаты которого неотрицательны, называется *допустимым базисным решением* (опорным планом); а если хотя бы одна координата отрицательна, то такое решение называется *недопустимым*.

В рассматриваемом примере допустимыми базисными решениями являются: (3; 1; 0; 0), (0; 1; 3/2; 0).

Допустимые базисные решения соответствуют координатам угловых точек пространства решений.

В вычислительной схеме симплекс-метода реализуется упорядоченный процесс, при котором, начиная с некоторой начальной угловой точки, осуществляются последовательные переходы от одной угловой точки к другой до тех пор, пока не будет найдена точка, соответствующая оптимальному решению.

Допустимое симплексное преобразование

Считаем, что в исходной системе все свободные члены неотрицательны. Дополним симплексное преобразование таким условием, при котором все неотрицательные свободные члены уравнения после преобразования остаются неотрицательными.

Условия допустимости:

• за разрешающий столбец примем такой столбец, в котором имеется хотя бы один положительный элемент;

• если в разрешающем столбце несколько положительных элементов, то за разрешающий элемент возьмем тот из них, для которого минимально отношение свободного члена к соответствующему элементу.

В результате такого допустимого симплексного преобразования неотрицательные свободные члены преобразуются снова в неотрицательные свободные члены.

Повторив допустимое симплексное преобразование несколько раз, переходя каждый раз к новому набору базисных переменных, получим все допустимые базисные решения.

Пример 1.6. Найти все допустимые базисные решения системы

$$\begin{cases} x_1 - 2x_2 + x_3 = 2, \\ x_1 + \quad x_2 + x_4 = 8. \end{cases}$$

▼Расчет допустимых базисных решений представим в виде последовательных симплексных таблиц (табл. 1.6).

Таблица 1.6

Симплексные таблицы для примера 1.6

Базис	x_1	x_2	x_3	x_4	b	**Базисное решение**	
x_3	(1)	–2	1	0	2	(0; 0; 2; 8)	2/1 = 2
x_4	1	1	0	1	8		8/1 = 8
x_1	1	–2	1	0	2	(2; 0; 0; 6)	
x_4	0	(3)	–1	1	6		
x_1	1	0	(1/3)	2/3	6	(6; 2; 0; 0)	
x_2	0	1	–1/3	1/3	2		
x_3	3	0	1	2	18	(0; 8; 18; 0)	
x_2	1	1	0	1	8		

Других допустимых базисных решений нет. ▲

Оптимальное симплексное преобразование

Впишем в симплексную таблицу коэффициенты целевой функции, представленной в виде

$$F - c_1 x_1 - c_2 x_2 - \ldots - c_n x_n = 0 .$$

Пусть с помощью допустимого симплексного преобразования получено какое-то базисное решение.

Целевая функция при таком преобразовании следующим образом выражается через свободные переменные:

$$F - c^*_{r+1} x_{r+1} - \ldots - c^*_n x_n = c^*_o ,$$

где r — число базовых переменных.

При допустимом базисном решении целевая функция принимает значение, равное значению свободного элемента $F = c^*_o$. Возможность уменьшить значение целевой функции F определяется значениями ее коэффициентов при свободных переменных. Направление перехода от одного допустимого базисного решения к другому определяется на основе условий оптимальности.

Условия оптимальности:

• если в строке F симплексной таблицы нет положительных коэффициентов, то базисное решение является оптимальным и задача решена;

• если в строке F симплексной таблицы есть хотя бы один положительный коэффициент, а в столбце при соответствующей переменной нет ни одного положительного элемента, то оптимального решения нет (целевая функция F не ограничена снизу);

• если в строке F симплексной таблицы имеется несколько положительных коэффициентов, то за разрешающий столбец вводимой переменной принимают тот, который соответствует наибольшему положительному коэффициенту строки F, а разрешающая строка выводимой переменной определяется условием допустимости.

Пример 1.7. Решим ОЗЛП симплексным методом.

$$F(x) = -x_1 + x_2 \to \min$$

при ограничениях

$$\begin{cases} -2x_1 + \ \ x_2 \le 2, \\ \ \ \ x_1 - 2x_2 \le 2, \\ \ \ \ x_1 + \ \ x_2 \le 5; \end{cases}$$

$$x_1 \ge 0\,,\ x_2 \ge 0\,.$$

▼Запишем ОЗЛП в канонической форме:

$$F(x) = -x_1 + x_2 \to \min$$

при ограничениях

$$\begin{cases} -2x_1 + \ \ x_2 + x_3 = 2, \\ \ \ \ x_1 - 2x_2 + x_4 = 2, \\ \ \ \ x_1 + \ \ x_2 + x_5 = 5; \end{cases}$$

$$x_1, \ldots, x_5 \ge 0\,.$$

Записываем целевую функцию в виде

$$F + x_1 - x_2 = 0\,.$$

Составляем симплексную таблицу и производим симплексные преобразования (табл. 1.7).

Таблица 1.7

Симплексные таблицы для примера 1.7

Базис	x_1	x_2	x_3	x_4	x_5	b	**Базисное решение**	
x_3	–2	1	1	0	0	2	(0; 0; 2; 2; 5)	
x_4	(1)	–2	0	1	0	2		2/1 = 2
x_5	1	1	0	0	1	5		5/1 = 5
F	1	–1	0	0	0	0		
x_3	0	–3	1	2	0	6	(2; 0; 6; 0; 3)	
x_1	1	–2	0	1	0	2		
x_5	0	(3)	0	–1	1	3		
F	0	1	0	–1	0	–2		

Окончание табл. 1.7

Базис	x_1	x_2	x_3	x_4	x_5	b	**Базисное решение**
x_3	0	0	1	1	1	9	(4; 1; 9; 0; 0)
x_1	1	0	0	1/3	2/3	4	
x_2	0	1	0	–1/3	1/3	1	
F	0	0	0	–2/3	–1/3	–3	

Оптимальное решение $x = (4;\ 1;\ 9;\ 0;\ 0)$, $\min F = -3$. ▲

Пример 1.8. Решим ОЗЛП симплексным методом.

$$F(x) = 3x_1 + 2x_2 \to \max$$

при ограничениях

$$\begin{cases} x_1 + 2x_2 \le 6, \\ 2x_1 + x_2 \le 8, \\ -x_1 + x_2 \le 1, \\ x_2 \le 2; \end{cases}$$

$$x_1, x_2 \ge 0 .$$

▼Запишем ОЗЛП в канонической форме:

$$F(x) = -3x_1 - 2x_2 \to \min$$

при ограничениях

$$\begin{cases} x_1 + 2x_2 + x_3 = 6, \\ 2x_1 + x_2 + x_4 = 8, \\ -x_1 + x_2 + x_5 = 1, \\ x_2 + x_6 = 2; \end{cases}$$

$$x_1, \dots, x_6 \ge 0 .$$

Записываем целевую функцию в виде

$$F + 3x_1 + 2x_2 = 0 .$$

Составляем симплексную таблицу и производим симплексные преобразования (табл. 1.8).

Таблица 1.8

Симплексные таблицы для примера 1.8

Базис	x_1	x_2	x_3	x_4	x_5	x_6	b	Базисное решение	
x_3	1	2	1	0	0	0	6	(0; 0; 6; 8; 1; 2)	6/1=6
x_4	(2)	1	0	1	0	0	8		8/2=4
x_5	–1	1	0	0	1	0	1		
x_6	0	1	0	0	0	1	2		
F	3	2	0	0	0	0	0		
x_3	0	(3/2)	1	–1/2	0	0	2	(4; 0; 2; 0; 5; 2)	4/3
x_1	1	1/2	0	1/2	0	0	4		8
x_5	0	3/2	0	1/2	1	0	5		10/3
x_6	0	1	0	0	0	1	2		2
F	0	1/2	0	–3/2	0	0	–12		
x_2	0	1	2/3	–1/3	0	0	4/3	(10/3; 4/3; 0; 0; 3; 2/3)	
x_1	1	0	–1/3	2/3	0	0	10/3		
x_5	0	0	–1	1	1	0	3		
x_6	0	0	–2/3	1/3	0	1	2/3		
F	0	0	–1/3	–4/3	0	0	–38/3		

Оптимальное решение $x = (10/3;\ 4/3;\ 0;\ 0;\ 3;\ 2/3)$, $\max F = 38/3$. ▲

Пример 1.9. Решим задачу о ресурсах симплексным методом.

$$F(x) = 40x_1 + 60x_2 \to \max$$

при ограничениях

$$\begin{cases} 2x_1 + 4x_2 \le 2000, \\ 4x_1 + x_2 \le 1400, \\ 2x_1 + x_2 \le 800; \end{cases}$$

$$x_1, x_2 \ge 0 .$$

▼Запишем ОЗЛП в канонической форме:

$$F(x) = -40x_1 - 60x_2 \to \min$$

при ограничениях

$$\begin{cases} 2x_1 + 4x_2 + x_3 = 2000, \\ 4x_1 + x_2 + x_4 = 1400, \\ 2x_1 + x_2 + x_5 = 800; \end{cases}$$

$$x_1, \ldots, x_5 \geq 0 .$$

Записываем целевую функцию в виде

$$F + 40x_1 + 60x_2 = 0 .$$

Составляем симплексную таблицу и производим симплексные преобразования (табл. 1.9).

Таблица 1.9

Симплексные таблицы для примера 1.9

Базис	x_1	x_2	x_3	x_4	x_5	b	**Базисное решение**	
x_3	2	(4)	1	0	0	2000	(0; 0; 200; 1400; 800)	500
x_4	4	1	0	1	0	1400		1400
x_5	2	1	0	0	1	800		800
F	40	60	0	0	0	0		
x_2	1/2	1	1/4	0	0	500	(0; 500; 0; 900; 300)	1000
x_4	7/2	0	–1/4	1	0	900		1800/7
x_5	(3/2)	0	–1/4	0	1	300		200
F	10	0	–15	0	0	–30 000		
x_2	0	1	1/3	0	–1/3	400	(200; 400; 0; 200; 0)	
x_4	0	0	1/3	1	–7/3	200		
x_1	1	0	–1/6	0	2/3	200		
F	0	0	–40/3	0	–20/3	–32 000		

Оптимальное решение $x = (200; 400; 0; 200; 0)$, $\max F = 32\ 000$. ▲

Метод искусственного базиса (*М*-метод)

Симплексный метод искусственного базиса применяется при отсутствии первоначального опорного плана исходной задачи ЛП, записанной в канонической форме. Такая ситуация возникает при наличии в исходном ограничении знаков «равно» либо «больше или равно».

М-метод заключается в применении правил симплекс-метода к так называемой *М*-задаче. Она получается из исходной задачи добавлением к левой части системы уравнений в канонической форме таких искусственных единичных векторов с соответствующими неотрицательными искусственными переменными, чтобы вновь полученная матрица содержала систему единичных линейно независимых векторов. В целевую функцию исходной задачи добавляется в случае ее минимизации слагаемое, представляющее собой произведение числа M на сумму искусственных переменных, где M — достаточно большое положительное число.

В полученной таким образом задаче первоначальный опорный план очевиден.

Если в процессе решения *М*-задачи все искусственные векторы вышли из базиса, то получаем исходную задачу. Если оптимальное решение *М*-задачи содержит искусственные векторы или *М*-задача неразрешима, то исходная задача также неразрешима.

Пример 1.10. Решить задачу ЛП.

$$F(x) = 2x_1 + 10x_2 + 4x_3 + 2x_4 \to \min$$

при ограничениях

$$\begin{cases} x_1 + 2x_2 + 2x_3 - 2x_4 = 2, \\ -x_1 + x_2 + x_3 + x_4 = 3. \end{cases}$$

▼Поскольку матрица ограничений не содержит единичных векторов, то введем две искусственные переменные y_1, y_2, а в целевую функцию добавим штраф.

$$F = 2x_1 + 10x_2 + 4x_3 + 2x_4 + M(y_1 + y_2) \to \min$$

при ограничениях

$$\begin{cases} x_1 + 2x_2 + 2x_3 - 2x_4 + y_1 = 2, \\ -x_1 + x_2 + x_3 + x_4 + y_2 = 3. \end{cases}$$

Для M-задачи составляем симплексную таблицу и производим симплексное преобразование (табл. 1.10). Элементы G-строки начальной таблицы вычисляются по формуле G-строка = F-строка + $M \cdot y_1$-строка + $M \cdot y_2$-строка.

Таблица 1.10

Симплексные таблицы для примера 1.10

F		–2	–10	–4	–2	$-M$	$-M$	0
Базис		x_1	x_2	x_3	x_4	y_1	y_2	b
M	y_1	1	2	(2)	–2	1	0	2
M	y_2	–1	1	1	1	0	1	3
G		–2	$3M-10$	$3M-4$	$-M-2$	0	0	$5M$
x_3		1/2	1	1	–1	1/2	0	1
y_2		–3/2	0	0	(2)	–1/2	1	2
G		$-3/2M$	–6	0	$2M-6$	$-3/2M+2$	0	$2M+4$
x_3		–1/4	1	1	0	1/4	1/2	2
x_4		–3/4	0	0	1	–1/4	1/2	1
G		–9/2	–6	0	0	$-M+1/2$	$-M+3$	10

Среди чисел последней строки, кроме свободного члена, при большом M нет положительных, значит, достигнуто оптимальное решение M-задачи.

Оптимальное решение исходной задачи есть

$$x = (0;\ 0;\ 2;\ 1),\ \min F(x) = 10.\ \blacktriangle$$

Пример 1.11. Решить задачу ЛП.

$$F(x) = 3x_1 + 2x_2 + 4x_3 \to \max; \qquad F(x) = -3x_1 - 2x_2 - 4x_3 \to \min;$$

$$\begin{cases} 2x_1 + x_2 = 8, \\ x_1 + x_2 + x_3 = 6; \end{cases} \Rightarrow \begin{cases} 2x_1 + x_2 = 8, \\ x_1 + x_2 + x_3 = 6; \end{cases}$$

$$x_1, x_2, x_3 \geq 0. \qquad x_1, x_2, x_3 \geq 0.$$

▼Поскольку матрица ограничений уже содержит один единичный вектор, то вводим одну искусственную переменную y_1, а в целевую функцию добавим штраф.

Получаем M-задачу

$$F = -3x_1 - 2x_2 - 4x_3 + My_1 \to \min;$$

$$\begin{cases} 2x_1 + x_2 + y_1 = 8, \\ x_1 + x_2 + x_3 = 6; \end{cases}$$

$$x_1, x_2, x_3 \ge 0, y_1 \ge 0.$$

Для M-задачи составляем симплексную таблицу и производим симплексные преобразования (табл. 1.11). Элементы G-строки начальной таблицы вычисляются по формуле G-строка = = F-строка + $M \cdot y_1$-строка + (–4) $\cdot$ x_3-строка.

Таблица 1.11

Симплексные таблицы для примера 1.11

F		3	2	4	$-M$	0
Базис		x_1	x_2	x_3	y_1	b
M	y_1	(2)	1	0	1	8
–4	x_3	1	1	1	0	6
G		$2M-1$	$M-2$	0	0	$8M-24$
x_1		1	1/2	0	1/2	4
x_3		0	1/2	1	–1/2	2
G		0	–3/2	0	$-M+1/2$	–20

Среди чисел последней строки при большом М нет положительных, значит, достигнуто оптимальное решение M-задачи.

Оптимальное решение исходной задачи есть

$$x = (4;\ 0;\ 2),\ \max F(x) = 20.\ \blacktriangle$$

Пример 1.12. Решить задачу ЛП.

$$F(x) = 10x_1 - 5x_2 \to \min$$

при ограничениях

$$\begin{cases} 2x_1 - x_2 \ge 3, \\ x_1 + x_2 \ge 2, \\ -x_1 - 2x_2 \le 1; \end{cases}$$

$$x_1, x_2 \ge 0.$$

▼Запишем ОЗЛП в канонической форме:

$$F(x) = 10x_1 - 5x_2 \to \min$$

при ограничениях

$$\begin{cases} 2x_1 - \ x_2 - x_3 = 3, \\ \ x_1 + \ x_2 - x_4 = 2, \\ -x_1 - 2x_2 + x_5 = 1; \end{cases}$$

$$x_1, x_2, x_3, x_4, x_5 \geq 0.$$

Поскольку матрица ограничений уже содержит один единичный вектор, то вводим две искусственные переменные y_1, y_2, а в целевую функцию добавим штраф.

Получаем M-задачу

$$F(x) = 10x_1 - 5x_2 + M(y_1 + y_2) \to \min$$

при ограничениях

$$\begin{cases} 2x_1 - \ x_2 - x_3 + y_1 = 3, \\ \ x_1 + \ x_2 - x_4 + y_2 = 2, \\ -x_1 - 2x_2 + x_5 \qquad = 1; \end{cases}$$

$$x_1, x_2, x_3, x_4, x_5, y_1, y_2 \geq 0.$$

Для M-задачи составляем симплексную таблицу и производим симплексные преобразования (табл. 1.12). Элементы G-строки начальной таблицы вычисляются по формуле G-строка = F-строка + $M \cdot y1$-строка + $M \cdot y2$-строка.

Таблица 1.12

Симплексные таблицы для примера 1.12

F		-10	5	0	0	0	$-M$	$-M$	0
Базис		x_1	x_2	x_3	x_4	x_5	y_1	y_2	b
M	y_1	(2)	-1	-1	0	0	1	0	3
M	y_2	1	1	0	-1	0	0	1	2
0	x_5	-1	-2	0	0	1	0	0	1
G		$3M - 10$	5	$-M$	$-M$	0	0	0	$5M$

Базис	x_1	x_2	x_3	x_4	x_5	y_1	y_2	b
x_1	1	–1/2	–1/2	0	0	1/2	0	3/2
y_2	0	(3/2)	1/2	–1	0	–1/2	1	1/2
x_5	0	–5/2	–1/2	0	1	1/2	0	5/2
G	0	$3M/2$	$M/2-5$	$-M$	0	$-3M/2+5$	0	$M/2+15$
x_1	1	0	–1/3	–1/3	0	1/3	1/3	5/3
x_2	0	1	1/3	–2/3	0	–1/3	2/3	1/3
x_5	0	0	1/3	–5/3	1	–1/3	5/3	10/3
G	0	0	–5	0	0	$-M+5$	$-M$	15

Среди чисел последней строки, кроме свободного члена, при большом M нет положительных, значит, достигнуто оптимальное решение M-задачи.

Оптимальное решение исходной задачи есть

$$x = (5/3;\ 1/3),\ \min F(x) = 15.\ \blacktriangle$$

1.4. Двойственные задачи

С каждой задачей ЛП можно связать некоторую другую задачу, называемую *двойственной* или *обратной*. При этом первоначальную задачу называют *исходной* или *прямой*. Решение двойственной задачи может быть получено из решения прямой задачи, и наоборот.

Переменные двойственной задачи (y) называются *двойственными оценками*, или «ценами» ресурсов, или теневыми ценами.

Двойственная задача формулируется из условий прямой задачи по следующим правилам.

1. Каждому из m ограничений прямой задачи соответствует переменная двойственной задачи, а каждой из n переменных прямой задачи соответствует ограничение двойственной задачи.

2. Каждый столбец из коэффициентов при переменных ограничения прямой задачи становится соответствующей строкой коэффициентов при переменных ограничения двойственной задачи.

3. Коэффициенты при переменных целевой функции прямой задачи становятся свободными членами ограничения двойственной задачи, а свободные члены ограничения прямой задачи становятся коэффициентами при переменных целевой функции двойственной задачи.

4. Тип экстремума целевой функции и тип ограничений-неравенств меняются (табл. 1.13).

Таблица 1.13

Соответствие типов оптимизации и ограничений-неравенств в прямой и двойственной задачах

Прямая задача		Двойственная задача	
целевая функция	**ограничения**	**целевая функция**	**ограничения**
max	$\le$	min	$\ge$
min	$\ge$	max	$\le$

Соотношение прямой и двойственной задач ЛП показано ниже.

Прямая задача	Двойственная задача
$F(x)=c_1x_1+c_2x_2+\ldots+c_nx_n \to \max;$	$G(y)=b_1y_1+b_2y_2+\ldots+b_my_m \to \min;$
$\begin{cases} a_{11}x_1+a_{12}x_2+\ldots+a_{1n}x_n \le b_1, & \mid y_1 \\ a_{21}x_1+a_{22}x_2+\ldots+a_{2n}x_n \le b_2, & \mid y_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ a_{m1}x_1+a_{m2}x_2+\ldots+a_{mn}x_n \le b_m; & \mid y_m \end{cases}$	$\begin{cases} a_{11}y_1+a_{21}y_2+\ldots+a_{m1}y_m \ge c_1, & \mid x_1 \\ a_{12}y_1+a_{22}y_2+\ldots+a_{m2}y_m \ge c_2, & \mid x_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ a_{1n}y_1+a_{2n}y_2+\ldots+a_{mn}y_m \ge c_n; & \mid x_n \end{cases}$
$x_1,x_2,\ldots,x_n \ge 0.$	$y_1,y_2,\ldots,y_m \ge 0.$

Пример 1.13. Составим двойственную задачу по исходной задаче.

Прямая задача	Двойственная задача
$F(x)=7x_1+6x_2+18x_3 \to \max;$	$G(y)=35y_1+20y_2 \to \min;$
$\begin{cases} x_1+x_2+5x_3 \le 35, & \mid y_1 \\ 2x_1+x_2+2x_3 \le 20, & \mid y_2 \end{cases}$	$\begin{cases} y_1+2y_2 \ge 7, \\ y_1+\ \ y_2 \ge 6, \\ 5y_1+2y_2 \ge 18; \end{cases}$
$x_1,x_2,x_3 \ge 0.$	$y_1,y_2 \ge 0.$

Пример 1.14. Составим двойственную задачу по исходной задаче.

Прямая задача	**Двойственная задача**
$F(x)=x_1+2x_2+3x_3 \to \min;$	$G(y)=9y_1+4y_2 \to \max;$
$\begin{cases} x_1+2x_2+x_3 \ge 9, & \vert y_1 \\ 2x_1+\quad\quad x_3 \ge 4; & \vert y_2 \end{cases}$	$\begin{cases} y_1+2y_2 \le 1, \\ 2y_1 \quad\quad \le 2, \\ y_1+\ y_2 \le 3; \end{cases}$
$x_1,x_2,x_3 \ge 0.$	$y_1,y_2 \ge 0.$

Прежде чем записывать двойственную задачу по исходной, систему ограничений и целевую функцию необходимо привести к соответствующему виду.

Пример 1.15. Составим двойственную задачу по исходной задаче.

Прямая задача:

$$F(x)=2x_1+x_2+5x_3 \to \min; \qquad F(x)=2x_1+x_2+5x_3 \to \min;$$

$$\begin{cases} x_1-\ x_2-\ x_3 \le 4, \\ x_1-5x_2+\ x_3 \ge 5, \\ 2x_1-\ x_2+3x_3 \ge 6; \end{cases} \Rightarrow \begin{cases} -x_1+\ x_2+\ x_3 \ge -4, & \vert y_1 \\ x_1-5x_2+x_3 \ge \ 5, & \vert y_2 \\ 2x_1-\ x_2+3x_3 \ge \ 6; & \vert y_3 \end{cases}$$

$$x_1,x_2,x_3 \ge 0. \qquad\qquad x_1,x_2,x_3 \ge 0.$$

Двойственная задача:

$$G(y)=-4y_1+5y_2+6y_3 \to \max;$$

$$\begin{cases} -y_1+\ y_2+2y_3 \le 2, \\ y_1-5y_2-\ y_3 \le 1, \\ y_1+\ y_2+3y_3 \le 5; \end{cases}$$

$$y_1,y_2,y_3 \ge 0.$$

Упражнение 1.2. Составить двойственную задачу по исходной задаче.

$$F(x)=5x_1+12x_2+4x_3 \to \max;$$

$$\begin{cases} x_1+2x_2+\ x_3 \le 10, & \vert y_1 \\ 2x_1-\ x_2+3x_3 = 8; & \vert y_2 \end{cases}$$

$$x_1,x_2,x_3 \ge 0.$$

Соответствие между переменными

Рассмотрим прямую и двойственную задачи в канонической форме. Прямая задача содержит n основных переменных $x_1, x_2, ..., x_n$. Для сведения системы ограничений-неравенств в эквивалентную ей систему ограничений-равенств нужно ввести m добавочных переменных $x_{n+1}, x_{n+2}, ..., x_{n+m}$ (по числу ограничений).

Таким образом, каждой основной переменной одной задачи можно поставить в соответствие добавочную переменную другой задачи (рис. 1.7).

<table>
<tr><th colspan="8">Переменные исходной задачи</th></tr>
<tr><th colspan="4">Основные</th><th colspan="4">Добавочные</th></tr>
<tr><td>x_1
$\updownarrow$
y_{m+1}</td><td>x_2
$\updownarrow$
y_{m+2}</td><td>…
…</td><td>x_n
$\updownarrow$
y_{m+n}</td><td>x_{n+1}
$\updownarrow$
y_1</td><td>x_{n+2}
$\updownarrow$
y_2</td><td>…
…</td><td>x_{n+m}
$\updownarrow$
y_m</td></tr>
<tr><th colspan="4">Добавочные</th><th colspan="4">Основные</th></tr>
<tr><th colspan="8">Переменные двойственной задачи</th></tr>
</table>

Рис. 1.7. Соответствие между переменными в прямой и двойственной задачах

Соответствия между переменными:

- основные переменные прямой задачи соответствуют добавочным переменным двойственной задачи, и наоборот;
- основные переменные оптимального решения одной из задач по абсолютной величине совпадают с добавочными переменными целевой функции другой задачи при достижении ее оптимума.

Фактически, имея оптимальную симплекс-таблицу одной задачи, можно получить оптимальное решение другой, не производя громоздких вычислений и не используя симплекс-метод.

Пример 1.16. Рассмотрим две взаимно двойственные задачи.

Прямая задача	Двойственная задача
$F(x) = 4x_1 + 5x_2 \to \max;$	$G(y) = 9y_1 + 2y_2 \to \min;$
$\begin{cases} 6x_1 + 3x_2 \le 9, \\ x_1 + x_2 \le 2; \end{cases}$	$\begin{cases} 6y_1 + y_2 \ge 4, \\ 3y_1 + y_2 \ge 5; \end{cases}$
$x_1, x_2 \ge 0.$	$y_1, y_2 \ge 0.$

Требуется по решению прямой задачи найти оптимальное решение двойственной задачи.

▼Решим прямую задачу симплексным методом.

Приведем прямую задачу к канонической форме:

$$F(x) = -4x_1 - 5x_2 \to \min;$$

$$\begin{cases} 6x_1 + 3x_2 + x_3 = 9, \\ x_1 + x_2 + x_4 = 2; \end{cases}$$

$$x_1, x_2, x_3, x_4 \geq 0.$$

Записываем целевую функцию в виде

$$F + 4x_1 + 5x_2 = 0.$$

Составляем симплексную таблицу и производим симплексное преобразование (табл. 1.14).

Таблица 1.14

Симплексные таблицы для примера 1.16

Базис	x_1	x_2	x_3	x_4	b	**Базисное решение**	
x_3	6	3	1	0	9		9/3 = 3
x_4	1	(1)	0	1	2	(0; 0; 9; 2)	2/1 = 2
F	4	5	0	0	0		
x_3	3	0	1	–3	3		
x_2	1	1	0	1	2	(0; 2; 3; 0)	
F	–1	0	0	–5	–10		

Оптимальное решение $x = (0; 2; 3; 0)$, $\max F(x) = 10$.

Составляем схему соответствия переменных:

$$\begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow \\ y_3 & y_4 & y_1 & y_2. \end{array}$$

Основные переменные y_1, y_2 оптимального решения двойственной задачи по абсолютной величине совпадают с добавочными переменными F-строки целевой функции прямой задачи.

Оптимальное решение двойственной задачи:

$$y = (0;\ 5),\ \min G(y) = \max F(x) = 10.\ \blacktriangle$$

Пример 1.17. Решить задачу ОЗЛП.

$$F(x) = x_1 - x_2 \to \max;$$

$$\begin{cases} -2x_1 + \ \ x_2 \le 2, \\ \ \ \ x_1 - 2x_2 \le 2, \\ \ \ \ x_1 + \ \ x_2 \le 5; \end{cases}$$

$$x_1, x_2 \ge 0.$$

$\blacktriangledown$Составим двойственную задачу и по ее решению найдем оптимальное решение исходной задачи.

Решим двойственную задачу симплексным методом.

Двойственная задача	**Каноническая форма**
$G(y) = 2y_1 + 2y_2 + 5y_3 \to \min;$	$G(y) = 2y_1 + 2y_2 + 5y_3 \to \min;$
$\begin{cases} -2y_1 + \ \ y_2 + y_3 \ge \ \ 1, \\ \ \ \ y_1 - 2y_2 + y_3 \ge -1; \end{cases}$	$\begin{cases} -2y_1 + \ \ y_2 + y_3 - y_4 = \ \ 1, \\ \ \ \ y_1 - 2y_2 + y_3 - y_5 = -1; \end{cases}$
$y_1, y_2, y_3 \ge 0.$	$y_1, y_2, y_3, y_4, y_5 \ge 0.$

Записываем целевую функцию в виде $G - 2y_1 - 2y_2 - 5y_3 = 0$.

Составим симплексную таблицу и произведем симплексные преобразования (табл. 1.15).

Таблица 1.15

Симплексные таблицы для примера 1.17

Базис	y_1	y_2	y_3	y_4	y_5	b	Базисное решение	
	−2	1	(1)	−1	0	1		
y_5	−1	2	−1	0	1	1		
G	−2	−2	−5	0	0	0		
y_3	−2	1	1	−1	0	1	(0; 0; 1; 0; 2)	1/1 = 1
y_5	−3	(3)	0	−1	1	2		2/3
G	−12	3	0	−5	0	5		
y_3	−1	0	1	−2/3	−1/3	1/3	(0; 2/3; 1/3; 0; 0)	
y_2	−1	1	0	−1/3	1/3	2/3		
G	−9	0	0	−4	−1	3		

Оптимальное решение $y = (0; 2/3; 1/3; 0; 0)$, $\min G(y) = 3$.

Составляем схему соответствия переменных:

$$\begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow \\ y_4 & y_5 & y_1 & y_2 & y_3. \end{array}$$

Основные переменные x_1, x_2 оптимального решения прямой задачи по абсолютной величине совпадают с добавочными переменными G-строки целевой функции прямой задачи.

Оптимальное решение задачи:

$$x = (4; 1),\ \max F(x) = \min G(y) = 3. \ \blacktriangle$$

Основные теоремы двойственности

Прямая и обратная задачи линейного программирования связаны между собой теоремами двойственности.

Теорема 1. Если одна из двойственных задач имеет оптимальное решение, то и другая тоже имеет оптимальное решение, при этом значения целевых функций на оптимальных решениях совпадают, т. е.

$$\max F(x) = \min G(y), \text{ или } \min F(x) = \max G(y).$$

Экономический смысл 1-й теоремы (задача о ресурсах). План производства (x) и набор оценок ресурсов (y) оказываются оптимальными тогда и только тогда, когда доход от реализации продукции, определяемый по известным заранее ценам $c_1, c_2, \ldots, c_n$ продукции, равен затратам на ресурсы по внутренним, определяемым только из решения задач ценам ресурсов $y_1, y_2, \ldots, y_m$, т. е.

$$F(x) = c_1x_1 + c_2x_2 + \ldots + c_nx_n = b_1y_1 + b_2y_2 + \ldots + b_my_m = G(y).$$

Для всех же других планов обеих задач доход от продукции всегда меньше (или равен) затрат на ресурсы $F(x) \leq G(y)$.

Двойственные оценки обладают тем свойством, что они гарантируют рентабельность оптимального плана, т. е. равенства общей оценки продукции и ресурсов обусловливают убыточность всякого другого плана, отличного от оптимального.

Теорема 2 утверждает, что:

• если в оптимальном решении одной из задач один из компонентов положителен, то для оптимального решения другой задачи соответствующее ограничение выполняется как равенство, т. е.

$$y_i > 0 \text{, то } \sum_j a_{ij} x_j = b_i; \tag{1.1}$$

$$x_j > 0 \text{, то } \sum_i a_{ij} y_i = c_j; \tag{1.2}$$

• если для оптимального решения одной из задач какое-либо ограничение отлично от нуля, то для оптимального решения другой задачи соответствующая компонента равна нулю, т. е.

$$\sum_j a_{ij} x_j < b_i \text{, то } y_i = 0; \tag{1.3}$$

$$\sum_i a_{ij} y_i > c_j \text{, то } x_j = 0. \tag{1.4}$$

Условия (1.1–1.4) интерпретируются следующим образом:

1. Если оценка i-го ресурса $y_i > 0$, то в оптимальном производственном плане (x) этот ресурс используется полностью.

2. Если j-й вид продукции вошел в оптимальный план ($x_j > 0$), то он в оптимальных оценках не убыточен.

3. Если ресурс b_i используется не полностью, то его оценка $y_i = 0$.

4. Если j-й вид продукции убыточен, то он не войдет в план, не будет выпускаться, $x_j = 0$.

Из условий (1.1) и (1.3) следует, что оценки оптимального плана — это мера дефицитности ресурсов. Дефицитный ресурс, полностью использованный по оптимальному плану производства, имеет оценку $y_i > 0$, а недефицитный ресурс, не полностью используемый, имеет оценку $y_i = 0$.

Из условий (1.2) и (1.4) следует, что оценки оптимального плана выступают как инструмент определения эффективности отдельных технологических способов. Данный способ производства используется только в том случае, когда при его реализации оценка затраченных ресурсов и цена полученной продукции совпадают.

Теоремы двойственности позволяют определить оптимальное решение одной из двойственных задач по решению другой.

Теорема об оценках. Значения переменных y_i в оптимальном решении двойственной задачи представляют собой оценки влияния свободных членов b_i в системе ограничений-неравенств прямой задачи на оптимальное значение ее целевой функции, т. е. $\Delta F(x) = y_i \Delta b_i$.

Заметим, что двойственная оценка (теневая цена) различных видов ресурсов рассматривается относительно оптимального решения и ее нельзя отожествить с действительными ценами, по которым осуществляется их закупка. Чем больше ценность y_i ресурса, тем острее его дефицитность. Для недефицитного ресурса его ценность y_i равна нулю.

Решая задачу ЛП симплекс-методом, мы одновременно решаем и исходную, и двойственную задачи.

Решение двойственных задач

Пример 1.18. Запишем взаимно двойственные задачи.

Прямая задача	Двойственная задача
$F(x) = x_1 - x_2 \to \max;$	$G(y) = 2y_1 + 2y_2 + 5y_3 \to \min;$
$\begin{cases} -2x_1 + x_2 \le 2, \\ x_1 - 2x_2 \le 2, \\ x_1 + x_2 \le 5; \end{cases}$	$\begin{cases} -2y_1 + y_2 + y_3 \ge 1, \\ y_1 - 2y_2 + y_3 \ge -1; \end{cases}$
$x_1, x_2 \ge 0$	$y_1, y_2, y_3 \ge 0$

а) Пусть известно решение прямой задачи

$$x = (4;\ 1),\ \max F(x) = 3.$$

Найдем решение двойственной задачи.

▼На основании 1-й теоремы двойственности:

$$\min G(y) = \max F(x) = 3.$$

На основании 2-й теоремы двойственности:

$$\left[\begin{aligned} x_1 &> 0, \\ x_2 &> 0; \end{aligned}\right. \Rightarrow \begin{cases} -2y_1 + y_2 + y_3 = 1, \\ y_1 - 2y_2 + y_3 = -1. \end{cases}$$

Подставляя оптимальное решение $x = (4;\ 1)$ в систему ограничений прямой задачи, получим:

$$\begin{cases} -2\cdot 4 + \phantom{2\cdot{}}1 < 2 \Rightarrow y_1 = 0, \\ \phantom{-2\cdot{}}4 - 2\cdot 1 = 2 \Rightarrow y_2 > 0, \\ \phantom{-2\cdot{}}4 + \phantom{2\cdot{}}1 = 5 \Rightarrow y_3 > 0. \end{cases}$$

Тогда система ограничений двойственной задачи имеет вид

$$\begin{cases} y_2 + y_3 = 1 \\ -2y_2 + y_3 = -1 \end{cases} \Rightarrow \quad y_2 = 2/3,\ y_3 = 1/3,$$

следовательно, оптимальное решение двойственной задачи $y = (0;\ 2/3;\ 1/3)$, $\min G(y) = 3$. ▲

b) Пусть известно решение двойственной задачи:

$$y = (0;\ 2/3;\ 1/3),\ \min G(y) = 3.$$

Найдем решение прямой задачи.

▼На основании 1-й теоремы двойственности:

$$\max F(x) = \min G(y) = 3.$$

По 2-й теореме двойственности:

$$\left[\begin{aligned} y_2 &> 0, \\ y_3 &> 0; \end{aligned}\right. \Rightarrow \begin{cases} x_1 - 2x_2 = 2, \\ x_1 + x_2 = 5; \end{cases} \Rightarrow \quad x_1 = 4,\ x_2 = 1,$$

следовательно, оптимальное решение прямой задачи $x = (4;\ 1)$, $\max F(x) = 3$. ▲

Пример 1.19. Запишем взаимно двойственные задачи.

Прямая задача	Двойственная задача
$F(x) = 4x_1 + 6x_2 \to \min;$ $\begin{cases} 3x_1 + x_2 \geq 9, \\ x_1 + 2x_2 \geq 8, \\ x_1 + 6x_2 \geq 12; \end{cases}$ $x_1, x_2 \geq 0.$	$G(y) = 9y_1 + 8y_2 + 12y_3 \to \max;$ $\begin{cases} 3y_1 + y_2 + y_3 \leq 4, \\ y_1 + 2y_2 + 6y_3 \leq 6; \end{cases}$ $y_1, y_2, y_3 \geq 0.$

а) Пусть известно решение прямой задачи

$$x = (2;\ 3),\ \min F(x) = 26.$$

Найдем решение двойственной задачи.

▼По 1-й теореме двойственности:

$$\max G(y) = \min F(x) = 26.$$

По 2-й теореме двойственности:

$$\left[\begin{array}{l} x_1 > 0, \\ x_2 > 0; \end{array}\right. \Rightarrow \begin{cases} 3y_1 + y_2 + y_3 = 4, \\ y_1 + 2y_2 + 6y_3 = 6. \end{cases}$$

Подставляя оптимальное решение x = (2; 3) в систему ограничений прямой задачи, получим:

$$\begin{cases} 3 \cdot 2 + 3 = 9 \Rightarrow y_1 > 0, \\ 2 + 2 \cdot 3 = 8 \Rightarrow y_2 > 0, \\ 2 + 6 \cdot 3 > 12 \Rightarrow y_3 = 0. \end{cases}$$

Тогда система ограничений двойственной задачи имеет вид:

$$\begin{cases} 3y_1 + y_2 = 4, \\ y_1 + 2y_2 = 6; \end{cases} \Rightarrow \ y_1 = 2/5, y_2 = 14/5,$$

следовательно, оптимальное решение двойственной задачи y = (2/5; 14/5), max $G(y)$ =26. ▲

b) Пусть известно решение двойственной задачи:

$$y = (2/5;\ 14/5),\ \max G(y) = 26.$$

Найдем решение прямой задачи.

▼По 1-й теореме двойственности: $\min F(x) = \max G(y) = 26$.

По 2-й теореме двойственности:

$$\left[\begin{array}{l} y_1 > 0, \\ y_2 > 0; \end{array}\right. \Rightarrow \left\{\begin{array}{l} 3x_1 + \ \ x_2 = 9, \\ \ \ x_1 + 2x_2 = 8; \end{array}\right. \Rightarrow \quad x_1 = 2, x_2 = 3,$$

следовательно, оптимальное решение прямой задачи $x = (2;\ 3)$, $\min F(x) = 26$. ▲

Решение задач ЛП в *Excel*

Решение задач линейного программирования можно произвести с помощью надстройки *Microsoft Excel Поиск решения*. Надстройка становится доступной при установке *Microsoft Excel*. Однако, чтобы использовать эту надстройку в *Excel*, необходимо сначала загрузить ее.

Загрузка надстроек *Поиск решения* и *Анализ данных*:

• в *Microsoft Office* 2007 щелкните значок *Кнопка Microsoft Office*, а затем *Параметры Excel*;

• выберите команду *Надстройки*, а затем в поле *Управление* — пункт *Надстройки Excel*;

• нажмите кнопку *Перейти*;

• в окне *Доступные надстройки* установите флажок *Поиск решения* и нажмите *ОК*.

После загрузки надстройки *Поиск решения* в группе *Анализ* на вкладке *Данные* становится доступна команда *Поиск решения*.

До вызова *Поиск решения* необходимо подготовить данные для решения задачи ЛП на рабочем листе *Excel*.

Пример 1.20. Решим задачу ЛП, используя *Excel* (*Поиск решения*).

$$F(x) = 2x_1 - x_2 \to \min;$$

$$\left\{\begin{array}{l} \ \ x_1 + \ \ x_2 \geq 4, \\ -x_1 + 2x_2 \leq 2, \\ \ \ x_1 + 2x_2 = 10; \end{array}\right.$$

$$x_1, x_2 \geq 0.$$

▼Вид рабочего листа *Excel* с условиями примера представлен на рис. 1.8.

	A	B	C	D	E
1		x_1	x_2		
2					
3	F	2	-1	0	
4	b_1	1	1	0	4
5	b_2	-1	2	0	2
6	b_3	1	2	0	10

Рис. 1.8. Условия примера 1.20 в *Excel*

Данные задачи и формулы в соответствующие ячейки вводятся следующим образом.

B1:C1 — записаны обозначения переменных модели x_1, x_2.

B2:C2 — резервируются для значений переменных модели, которые будут найдены после выполнения процедуры *Поиск решения*.

A3:A6 — записаны обозначения строки целевой функции F и строк ограничений b_1, b_2, b_3.

B3:C3 — записаны коэффициенты при переменных модели в целевой функции.

B4:C6 — заносим матрицу коэффициентов при переменных в системе ограничений модели.

E4:E6 — записаны правые части системы ограничений модели.

D3 — вводим формулу =СУММПРОИЗВ(B3:C3;$B$2:$C$2). «Протягиваем» ее до D6 включительно.

Далее запустим команду *Поиск решения*. В диалоговом окне *Поиск решения*:

• установить целевую ячейку (D3) равной минимальному значению;

• в качестве изменяемых ячеек ввести адрес переменных;

• ввести ограничения (рис. 1.9).

В диалоговом окне *Параметры поиска решения* установить флажки в *Линейная модель* и *Неотрицательные значения* (рис. 1.10).

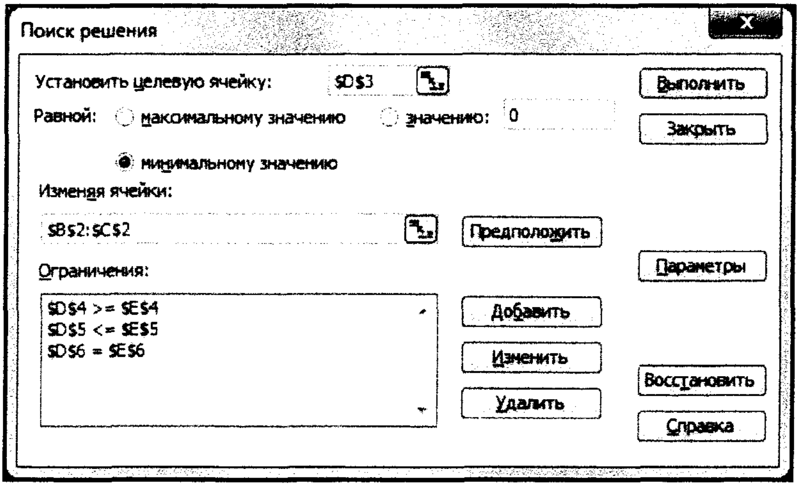


Рис. 1.9. Заполненное диалоговое окно *Поиск решения* для примера 1.20

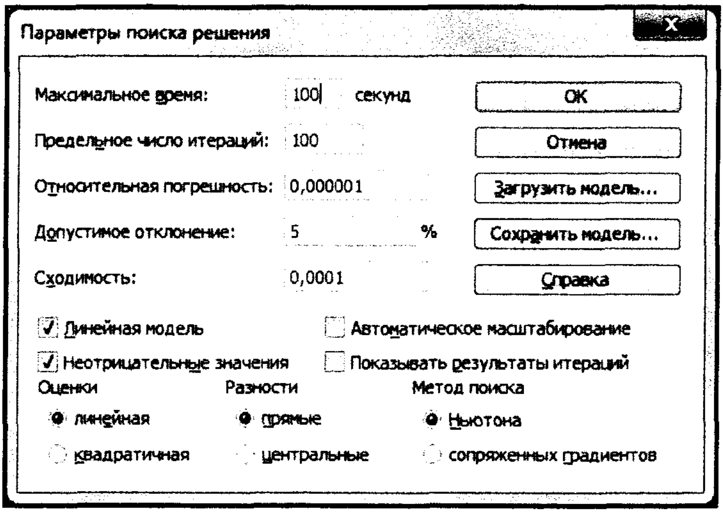


Рис. 1.10. Заполненное диалоговое окно *Параметры поиска решения* для примера 1.20

Вид рабочего листа *Excel* после выполнения команды *Поиск решения* представлен на рис. 1.11.

	A	B	C	D	E
1		x_1	x_2		
2		4	3		
3	F	2	-1	**5**	
4	b_1	1	1	**7**	4
5	b_2	-1	2	**2**	2
6	b_3	1	2	**10**	10

Рис. 1.11. Пример 1.20 в *Excel* после выполнения команды *Поиск решения*

Оптимальное решение: $x = (4; 3)$; $\min F(x) = 5$. ▲

Пример 1.21. Решим задачу составления рациона.

$$F(x) = 4x_1 + 6x_2 \to \min;$$

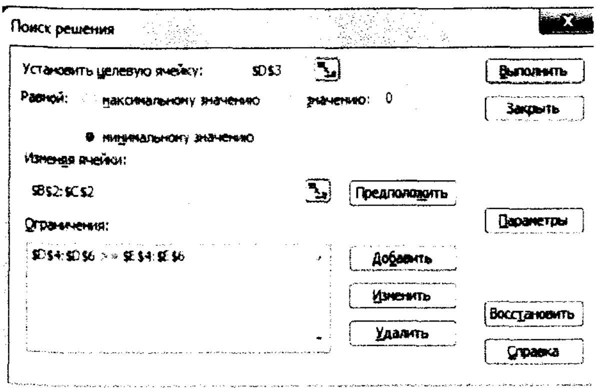
$$\begin{cases} 3x_1 + \ x_2 \geq 9, \\ \ x_1 + 2x_2 \geq 8, \\ \ x_1 + 6x_2 \geq 12; \end{cases}$$

$$x_1, x_2 \geq 0.$$

▼Вид рабочего листа *Excel* с исходными данными примера и заполненного диалогового окна *Поиск решения* представлен на рис. 1.12.

	A	B	C	D	E
1		x_1	x_2		
2					
3	F	4	6	**0**	
4	b_1	3	1	**0**	9
5	b_2	1	2	**0**	8
6	b_3	1	6	**0**	12

а)



б)

Рис. 1.12. Исходные данные в *Excel* (*а*) и заполненное диалоговое окно *Поиск решения* (*б*) для примера 1.21

Вид рабочего листа *Excel* после выполнения команды *Поиск решения* представлен на рис. 1.13.

	A	B	C	D	E
1		x_1	x_2		
2		2	3		
3	F	4	6	**26**	
4	b_1	3	1	**9**	9
5	b_2	1	2	**8**	8
6	b_3	1	6	**20**	12

Рис. 1.13. Пример 1.21 в *Excel* после выполнения команды *Поиск решения*

Оптимальное решение: $x = (2; 3)$; $\min F(x) = 26$. ▲

Пример 1.22. Решим задачу использования ресурсов.

$$F(x) = 40x_1 + 60x_2 \to \max;$$

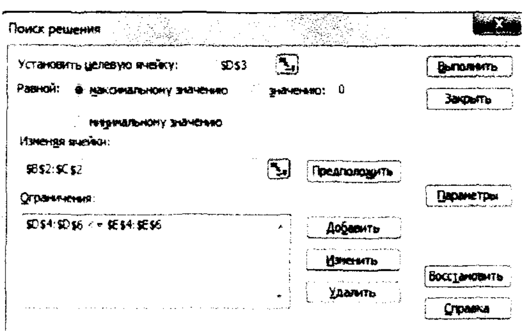
$$\begin{cases} 2x_1 + 4x_2 \le 2000, \\ 4x_1 + \ x_2 \le 1400, \\ 2x_1 + \ x_2 \le 800; \end{cases}$$

$$x_1, x_2 \ge 0.$$

▼Вид рабочего листа *Excel* с исходными данными примера и заполненного диалогового окна *Поиск решения* представлен на рис. 1.14.

	A	B	C	D	E
1		x_1	x_2		
2					
3	F	40	60	**0**	
4	b_1	2	4	**0**	2000
5	b_2	4	1	**0**	1400
6	b_3	2	1	**0**	800

а)



б)

Рис. 1.14. Исходные данные в *Excel* (*а*) и заполненное диалоговое окно *Поиск решения* (*б*) для примера 1.22

Вид рабочего листа *Excel* после выполнения команды *Поиск решения* представлен на рис. 1.15.

	A	B	C	D	E
1		x_1	x_2		
2		**200**	400		
3	F	40	60	**32000**	
4	b_1	2	4	**2000**	2000
5	b_2	4	1	**1200**	1400
6	b_3	2	1	**800**	800

Рис. 1.15. Пример 1.22 в *Excel* после выполнения команды *Поиск решения*

Оптимальное решение: $x = (200; 400)$; $\max F(x) = 32\ 000$.

Результаты решения задачи содержат три отчета (отчет по результатам, отчет по устойчивости и отчет по пределам).

Отчет по результатам (рис. 1.16). В нем приводятся оптимальные значения переменных $x = (200; 400)$, значение целевой функции $\max F(x) = 32\ 000$ и значение статуса ресурса (связанное, не связанное). Это значит, что второй ресурс («не связанное») является недефицитным (используется не полностью), а остальные («связанное») — дефицитными (используются полностью).

Целевая ячейка (Максимум)						
	Ячейка	**Имя**	**Исходное значение**	**Результат**		
	\$D\$3	F	0	32000		
Изменяемые ячейки						
	Ячейка	**Имя**	**Исходное значение**	**Результат**		
	\$B\$2	x_1	0	200		
	\$C\$2	x_2	0	400		
Ограничения						
	Ячейка	**Имя**	**Значение**	**Формула**	**Статус**	**Разница**
	\$D\$4	b_1	2000	\$D\$4<=\$E\$4	Связанное	0
	\$D\$5	b_2	1200	\$D\$5<=\$E\$5	Не связанное	200
	\$D\$6	b_3	800	\$D\$6<=\$E\$6	Связанное	0

Рис. 1.16. Отчет по результатам для примера 1.22

Отчет по устойчивости (рис. 1.17). Этот отчет имеет две части.

Изменяемые ячейки							
	Ячейка	Имя	Результ. Значение	Нормир. Стоимость	Целевой Коэффициент	Допустимое Увеличение	Допустимое Уменьшение
	$B$2	x_1	200	0	40	80	10
	$C$2	x_2	400	0	60	20	40
Ограничения							
	Ячейка	Имя	Результ. Значение	Теневая Цена	Ограничение Правая часть	Допустимое Увеличение	Допустимое Уменьшение
	$D$4	b_1	2000	13,33333	2000	1200	600
	$D$5	b_2	1200	0	1400	1E+30	200
	$D$6	b_3	800	6,66667	800	85,71428	300

Рис. 1.17. Отчет по устойчивости для примера 1.22

Первая часть отчета содержит информацию, относящуюся к переменным. Указан результат x = (200; 400) — оптимальные значения переменных (оптимальный план) прямой задачи.

Для целевых коэффициентов цен c = (40; 60) указаны их допустимое увеличение $\Delta c^+ = (80; 20)$ и допустимое уменьшение $\Delta c^- = (10; 40)$, при которых сохраняется текущий оптимальный план x = (200; 400). Это значит, что изменение коэффициентов c_1, c_2 целевой функции в интервалах

$$(c_1 - \Delta c_1^-; c_1 + \Delta c_1^+) = (40 - 10; 40 + 80) = (30; 120);$$

$$(c_2 - \Delta c_2^-; c_2 + \Delta c_2^+) = (60 - 40; 60 + 20) = (20; 80)$$

не приведет к изменению полученного ранее оптимального плана x.

Вторая часть отчета содержит информацию, относящуюся к ограничениям. Указаны теневые цены ресурсов y = (13,3; 0; 6,7) — оптимальное решение двойственной задачи. Это значит, что наиболее дефицитным является 1-й ресурс, наименее дефицитным — 3-й ресурс, а 2-й — недефицитным ($y_2 = 0$).

Для правых частей ограничений b = (2000; 1400; 800) — запасов ресурсов — указаны их допустимое увеличение Δb^+ и допустимое уменьшение Δb^-, при которых сохраняются первоначальные цены на ресурсы. Это значит, что изменение запасов ресурсов (b) в интервалах

$$(b_1 - \Delta b_1^-; b_1 + \Delta b_1^+) = (2000 - 600; 2000 + 1200) = (1400; 3200);$$

$$(b_2 - \Delta b_2^-; b_2 + \Delta b_2^+) = (1400 - 200; 1400 + \infty) = (1200; \infty);$$

$$(b_3 - \Delta b_3^-; b_3 + \Delta b_3^+) = (800 - 300; 800 + 85,7) = (500; 885,7)$$

не приведет к изменению их теневых цен (y).

Отчет по пределам (рис. 1.18). Он содержит значения целевой функции при отсутствии и присутствии выпуска одного из видов продукции.

	Целевое							
Ячейка	**Имя**	**Значение**						
$D$3	F	32000						
	Изменяемое			**Нижний предел**	**Целевой Результат**		**Верхний Предел**	**Целевой результат**
Ячейка	**Имя**	**Значение**						
$B$2	x_1	200		0	24000		200	32000
$C$2	x_2	400		0	8000		400	32000

Рис. 1.18. Отчет по пределам для примера 1.22

Замечания.

1. Теневые цены y_i указывают, как влияет изменение запасов ресурсов b_i на изменение целевой функции F (дохода). Например, определим изменение дохода от реализации продукции при изменении запасов ресурсов на величины +30; –40; +60 соответственно.

Такое изменение запасов ресурсов приведет к их ограничению до $b_1 = 2030$; $b_2 = 1360$; $b_3 = 860$ ед. Поскольку эти значения находятся в пределах устойчивости оценок (y), то раздельное ΔF_i и суммарное $\Delta F_{\max}$ их влияние на доход есть

$$\Delta F_1 = y_1 \Delta b_1 = 13,33 \cdot 30 = 400,$$

$$\Delta F_2 = y_2 \Delta b_2 = 0 \cdot (-40) = 0,$$

$$\Delta F_3 = y_3 \Delta b_3 = 6,67 \cdot 60 = 400,$$

$$\Delta F_{\max} = \Delta F_1 + \Delta F_2 + \Delta F_3 = 800.$$

2. Теневые цены (y) являются показателями эффективности производства отдельных видов продукции. В оптимальный план может быть включена лишь та продукция j-го вида, для которой затраты на ресурсы при производстве нового вида продукции не превышают при этом доходы (цены c_j единицы нового вида продукции), т. е. $\sum a_{ij} y_i \le c_j$.

Например, оценим целесообразность введения в оптимальный план исходной продукции нового (третьего) вида, нормы затрат которого на производство единицы продукции составляют соответственно 2; 3; 2 ед., а цена — 45 усл. ед.

Поскольку $a_{13}y_1 + a_{23}y_2 + a_{33}y_3 = 2 \cdot 13{,}33 + 3 \cdot 0 + 2 \cdot 6{,}67 = 40 < c_3 = 45$, то введение в план третьего вида продукции выгодно. ▲

Пример 1.23. Фирма выпускает три вида изделия P_1, P_2, P_3, используя при этом четыре вида сырья S_1, S_2, S_3, S_4. Запасы сырья, нормы затрат каждого вида сырья на изделия и цены реализации ед. изделия приведены в табл. 1.16.

Таблица 1.16

Условия примера 1.23

Сырье	**Затраты сырья на ед. изделия, ед.**			**Запасы сырья, ед.**
	P_1	P_2	P_3	
S_1	1	2	1	18
S_2	2	1	1	16
S_3	1	1	0	8
S_4	0	1	1	6
Цена ед. изделия, усл. ед.	3	4	2	

Требуется:

a) составить план производства трех видов изделий, максимизирующий доход;

b) определить дефицитность сырья;

c) установить размеры максимального дохода при изменении запасов сырья на +6; –3; +2; +2 ед. соответственно;

d) определить целесообразность введения в план производства фирмы нового вида изделия (четвертого), нормы затрат которого соответственно равны 1; 2; 2; 0, а цена ед. изделия составляет 15 усл. ед.

▼Обозначим как $x = (x_1, x_2, x_3)$ план производства трех видов изделия.

Экономико-математическая модель задачи имеет вид

$$F(x) = 3x_1 + 4x_2 + 2x_3 \to \max;$$

$$\begin{cases} x_1 + 2x_2 + x_3 \le 18, \\ 2x_1 + \ \ x_2 + x_3 \le 16, \\ x_1 + \ \ x_2 \qquad \le 8, \\ \qquad\quad x_2 + x_3 \le 6; \end{cases}$$

$$x_1, x_2, x_3, x_4 \ge 0.$$

Решение задачи произведем в *Excel.*

a) Оптимальный план производства трех видов изделия составляет $x = (5; 3; 3)$, при этом доход $\max F(x) = 33$ усл. ед. (рис. 1.19).

	A	B	C	D	E	F
1		x_1	x_2	x_3		
2		5	3	3		
3	F	3	4	2	**33**	
4	b_1	1	2	1	**14**	18
5	b_2	2	1	1	**16**	16
6	b_3	1	1	0	**8**	8
7	b_4	0	1	1	**6**	6

Рис. 1.19. Пример 1.23 в *Excel*
после выполнения команды *Поиск решения*

b) Дефицитность сырья можно установить по значению теневых цен $y = (0; 0{,}5; 2; 1{,}5)$ из отчета по устойчивости (рис. 1.20).

Наиболее дефицитным является 3-е сырье, для которого теневая цена $y_3 = 2$. Наименее дефицитным — 2-е сырье, для которого $y_2 = 0{,}5$. Совсем недефицитным — 1-е сырье, $y_1 = 0$.

c) Из отчета по устойчивости найдем интервалы изменений запасов сырья, которые не приведут к изменению их теневых цен.

$$(b_1 - \Delta b_1^-; b_1 + \Delta b_1^+) = (18-4; 18+\infty) = (14; \infty),$$
$$(b_2 - \Delta b_2^-; b_2 + \Delta b_2^+) = (16-6; 16+6) = (10; 22),$$
$$(b_3 - \Delta b_3^-; b_3 + \Delta b_3^+) = (8-3; 8+3) = (5; 11),$$
$$(b_4 - \Delta b_4^-; b_4 + \Delta b_4^+) = (6-6; 6+4) = (0; 10).$$

Изменяемые ячейки							
	Ячейка	**Имя**	**Результ. значение**	**Нормир. стоимость**	**Целевой Коэффициент**	**Допустимое увеличение**	**Допустимое уменьшение**
	\$B\$2	x_1	5	0	3	3	1
	\$C\$2	x_2	3	0	4	1	2
	\$D\$2	x_3	3	0	2	2	1
Ограничения							
	Ячейка	**Имя**	**Результ. значение**	**Теневая цена**	**Ограничение Правая часть**	**Допустимое увеличение**	**Допустимое уменьшение**
	\$E\$4	b_1	14	0	18	1E+30	4
	\$E\$5	b_2	16	0,5	16	6	6
	\$E\$6	b_3	8	2	8	3	3
	\$E\$7	b_4	6	1,5	6	4	6

Рис. 1.20. Отчет по устойчивости для примера 1.23

Изменение запасов сырья, согласно условию задачи, на +6; –3; +2; +2 ед. соответственно приведет к ограничению их запасов до значений $b_1 = 24$, $b_2 = 13$, $b_3 = 10$, $b_4 = 8$, которые находятся в пределах устойчивости теневых цен.

Раздельное ΔF_i и суммарное $\Delta F_{\max}$ влияние изменения запасов сырья на доход есть

$$\Delta F_1 = y_1 \Delta b_1 = 0 \cdot 6 = 0,$$
$$\Delta F_2 = y_2 \Delta b_2 = 0,5 \cdot (-3) = -1,5,$$
$$\Delta F_3 = y_3 \Delta b_3 = 2 \cdot 2 = 4,$$
$$\Delta F_4 = y_4 \Delta b_4 = 1,5 \cdot 2 = 3,$$
$$\Delta F_{\max} = \Delta F_1 + \Delta F_2 + \Delta F_3 + \Delta F_4 = 5,5.$$

d) Оценим целесообразность введения в оптимальный план нового (четвертого) вида изделия, для которого, согласно условию задачи, нормы затрат сырья на ед. нового изделия равны 1; 2; 2; 0 соответственно, а цена ед. изделия составляет 15 усл. ед. Поскольку

$$a_{14}y_1 + a_{24}y_2 + a_{34}y_3 + a_{44}y_4 = 1 \cdot 0 + 2 \cdot 0{,}5 + 2 \cdot 2 + 0 \cdot 1{,}5 = \\ = 5 < c_4 = 15,$$

то введение в план нового вида продукции выгодно. При введении нового вида изделия оптимальный план производства составляет $x = (0;\ 0;\ 6;\ 4)$, при этом доход $\max F(x) = 72$ усл. ед.

Упражнение 1.3. Решить ОЗЛП в *Excel*.

$$F(x) = 10x_1 + 20x_2 \to \max;$$

$$\begin{cases} x_1 + 3{,}5x_2 \le 350, \\ 2x_1 + 0{,}5x_2 \le 240, \\ x_1 + \quad x_2 \le 150, \\ \qquad\quad x_2 \ge 60; \end{cases}$$

$$x_1, x_2, x_3, x_4 \ge 0.$$

Ответ: оптимальное решение $x = (70;\ 80)$, $\max F(x) = 2300$.

Упражнение 1.4. Запишем взаимно двойственные задачи.

Прямая задача	**Двойственная задача**
$F(x) = x_1 - x_2 \to \max;$	$G(y) = 2y_1 + 2y_2 + 5y_3 \to min;$
$\begin{cases} -2x_1 + \quad x_2 \le 2, \\ \quad x_1 - 2x_2 \le 2, \\ \quad x_1 + \quad x_2 \le 5; \end{cases}$	$\begin{cases} -2y_1 + \quad y_2 + y_3 \ge \ 1, \\ \quad y_1 - 2y_2 + y_3 \ge -1; \end{cases}$
$x_1, x_2 \ge 0.$	$y_1, y_2, y_3 \ge 0.$

a) Используя *Excel*, найдите оптимальное решение прямой задачи.

Ответ: $x = (4;\ 1)$, $\max F(x) = 3$.

b) По отчету по устойчивости определите оптимальное решение двойственной задачи.

Ответ: $y = (0;\ 0{,}667;\ 0{,}333)$, $\min G(y) = 3$.

Упражнение 1.5. Запишем взаимно двойственные задачи.

Прямая задача	Двойственная задача
$F(x) = 4x_1 + 6x_2 \to \min;$	$G(y) = 9y_1 + 8y_2 + 12y_3 \to \max;$
$\begin{cases} 3x_1 + x_2 \geq 9, \\ x_1 + 2x_2 \geq 8, \\ x_1 + 6x_2 \geq 12; \end{cases}$	$\begin{cases} 3y_1 + y_2 + y_3 \leq 4, \\ y_1 + 2y_2 + 6y_3 \leq 6; \end{cases}$
$x_1, x_2 \geq 0.$	$y_1, y_2, y_3 \geq 0.$

a) Используя *Excel*, найдите оптимальное решение двойственной задачи.

Ответ: $y = (0{,}4;\ 2{,}8;\ 0)$, $\max G(y) = 26$.

b) По отчету по устойчивости определите оптимальное решение прямой задачи.

Ответ: $x = (2;\ 3)$, $\min F(x) = 26$.

Целочисленное линейное программирование

Под **задачей целочисленного ЛП** понимается задача ЛП, в которой все или некоторые переменные должны принимать целые значения. Для нахождения оптимального решения целочисленных задач используют методы отсечения и метод ветвей и границ.

Сущность *методов отсечения* состоит в том, что сначала задача решается без условия целочисленности. Если решение получается целочисленным, то задача решена. В противном случае к ограничению задачи добавляется новое ограничение, обладающее следующими свойствами. Оно:

• должно быть линейным;

• должно отсекать найденное оптимальное нецелочисленное решение;

• не должно отсекать ни одного целочисленного решения.

Дополнительное ограничение, обладающее указанными свойствами, называется правильным отсечением.

Далее задача решается с учетом нового ограничения. Процесс построения отсечений и решения задачи повторяется до получения целочисленного оптимального решения.

Сущность *метода ветвей и границ* состоит в следующем: множество допустимых решений некоторым способом разбивается на подмножества (ветвления), каждое из которых этим же способом снова разбивается на подмножества. Далее вычисляются оценки (границы), позволяющие отбрасывать подмножества, заведомо не содержащие решения задачи. Процесс продолжается до тех пор, пока не получено оптимальное целочисленное решение задачи.

В надстройке *Excel Поиск решения* для нахождения оптимального целочисленного решения реализован метод ветвей и границ.

Пример 1.24. Найти целочисленное решение.

$$F(x) = 2x_1 + 3x_2 \to \max;$$

$$\begin{cases} 3x_1 + 4x_2 \le 34, \\ \qquad\quad x_2 \le 5; \end{cases}$$

$$x_1, x_2 \ge 0;$$

$$x_1, x_2 \text{ — целые.}$$

▼При решении задачи целочисленного линейного программирования в *Excel Поиск решений* необходимо ввести условия целочисленности. В диалоговом окне *Добавление ограничения* следует выбрать опцию *целое* в раскрывшемся списке *Ограничение* (рис. 1.21).

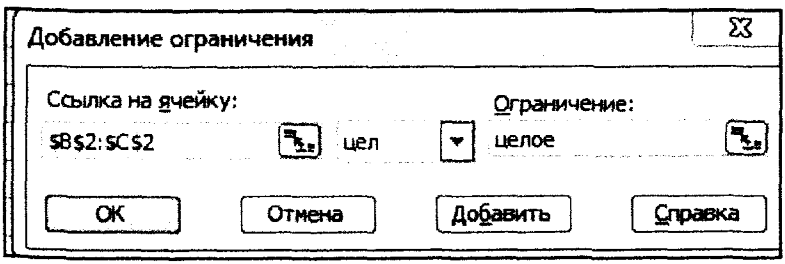
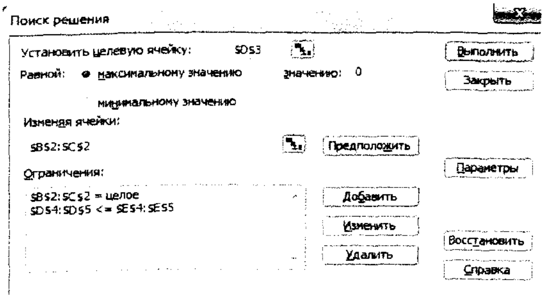


Рис. 1.21. Заполненное диалоговое окно *Добавление ограничения* для примера 1.24

Вид рабочего листа *Excel* с исходными данными примера и заполненного диалогового окна *Поиск решения* представлен на рис. 1.22.

	A	B	C	D	E
1		x_1	x_2		
2		0	0		
3	F	2	3	0	
4	b_1	3	4	0	34
5	b_2		1	0	5

а)



б)

Рис. 1.22. Исходные данные в *Excel* (*а*) и заполненное диалоговое окно *Поиск решения* (*б*) для примера 1.24

Вид рабочего листа *Excel* после выполнения команды *Поиск решения* представлен на рис. 1.23.

	A	B	C	D	E
1		x_1	x_2		
2		6	4		
3	F	2	3	**24**	
4	b_1	3	4	**34**	34
5	b_2	0	1	**4**	5

Рис. 1.23. Пример 1.24 в *Excel* после выполнения команды *Поиск решения*

Оптимальное решение $x = (6; 4)$, $\max F(x) = 24$. ▲

Упражнение 1.6. Найти целочисленное решение.

$$F(x) = 2x_1 + x_2 \to \max;$$

$$\begin{cases} 6x_1 + 4x_2 \ge 14, \\ 2x_1 + 5x_2 \le 10, \\ -x_1 + 3x_2 \ge 1; \end{cases}$$

$$x_1, x_2 \ge 0;$$

$$x_1, x_2 \text{ — целые.}$$

Ответ: оптимальное решение $x = (2; 1)$, $\max F(x) = 5$.

Двоичные (булевы) переменные

Во многих практических случаях переменные принимают не любые целые значения, а лишь одно из двух: либо 0, либо 1. Такие переменные называют двоичными (булевыми).

Пример 1.25 (задача о выборе инвестиционных проектов в условиях ограниченности финансовых ресурсов). У фирмы для выполнения некоторых программ имеется пять инвестиционных проектов, чистая приведенная стоимость (ЧПС) которых указана в табл. 1.17.

Таблица 1.17

Условия примера 1.25

Номер проекта	**ЧПС, усл. ед.**	**Требуемые вложения, усл. ед.**		
		1-й год	**2-й год**	**3-й год**
1	40	12	8	17
2	60	17	17	20
3	38	10	7	21
4	50	7	22	6
5	55	17	14	20
Выделенный объем денежных средств, усл. ед.		54	62	70

Однако фирма не может финансировать все проекты: сумма денег, выделенных на текущий год, составляет 54 усл. ед., а на последующие два года 62 и 70, что меньше необходимых для инвестирования в полном объеме. При этом оставшиеся денежные средства не могут быть перенесены на следующие годы, а также не планируется финансировать более одного раза один и тот же проект.

Требуется распределить выделенные средства в инвестиционные проекты оптимальным способом.

▼Пусть переменные x_1, x_2, x_3, x_4, x_5 — доля вложения в соответствующий проект, причем каждое x_i может принимать только два значения (двоичная переменная):

$$x_i = \begin{cases} 0, \text{ если проект не финансируется,} \\ 1, \text{ если проект финансируется.} \end{cases}$$

Составим экономико-математическую модель задачи.

$$F(x) = 40x_1 + 60x_2 + 38x_3 + 50x_4 + 55x_5 \to \max;$$

$$\begin{cases} 12x_1 + 17x_2 + 10x_3 + 7x_4 + 17x_5 \le 54, \\ 8x_1 + 17x_2 + 7x_3 + 22x_4 + 14x_5 \le 62, \\ 17x_1 + 20x_2 + 21x_3 + 6x_4 + 20x_5 \le 70; \end{cases}$$

x_1, x_2, x_3, x_4, x_5 — двоичные.

Решаем задачу в *Excel* (*Поиск решения*). В диалоговом окне *Поиск решения* следует добавить ограничение, превращающее переменные в двоичные (рис. 1.24).

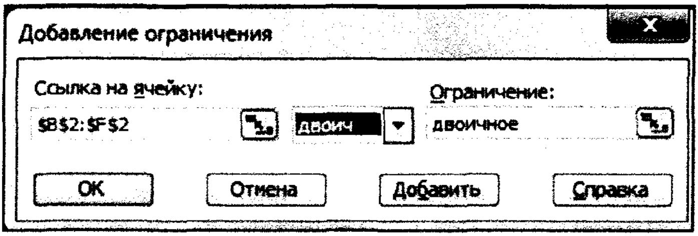
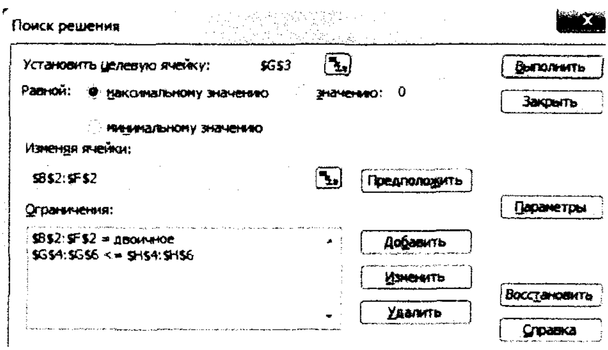


Рис. 1.24. Заполненное диалоговое окно *Добавление ограничения* для примера 1.25

Вид рабочего листа *Excel* с исходными данными примера и заполненного диалогового окна *Поиск решения* представлен на рис. 1.25.

	A	B	C	D	E	F	G	H
1		x_1	x_2	x_3	x_4	x_5		
2								
3	F	40	60	38	50	55	0	
4	b_1	12	17	10	7	17	0	54
5	b_2	8	17	7	22	14	0	62
6	b_3	17	20	21	6	20	0	70

а)



б)

Рис. 1.25. Исходные данные в *Excel* (*а*) и заполненное диалоговое окно *Поиск решения* (*б*) для примера 1.25

В диалоговом окне *Параметры поиска решения* необходимо уменьшить относительную погрешность или допустимое отклонение (рис. 1.26).

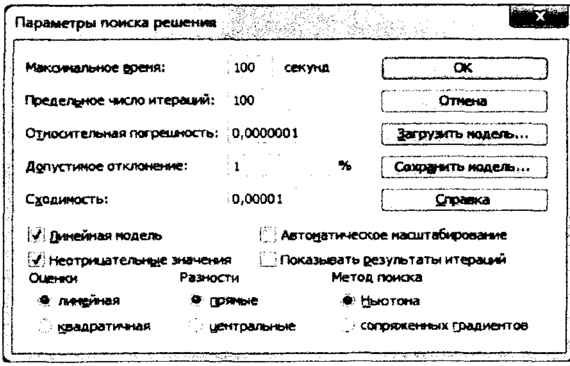


Рис. 1.26. Заполненное диалоговое окно *Параметры поиска решения* для примера 1.25

Вид рабочего листа *Excel* после выполнения команды *Поиск решения* представлен на рис. 1.27.

	A	B	C	D	E	F	G	H
1		x_1	x_2	x_3	x_4	x_5		
2		1	1	0	1	1		
3	F	40	60	38	50	55	**205**	
4	b_1	12	17	10	7	17	**53**	54
5	b_2	8	17	7	22	14	**61**	62
6	b_3	17	20	21	6	20	**63**	70

Рис. 1.27. Пример 1.25 в *Excel* после выполнения команды *Поиск решения*

Оптимальное решение $x = (1;\ 1;\ 0;\ 1;\ 1)$, $\max F(x) = 205$ – оптимальное решение.

Таким образом, необходимо финансировать 1-й, 2-й, 4-й и 5-й проекты, при этом сумма ЧПС проектов максимальна и составляет 205 усл. ед. Для этого потребуются денежные средства в объеме $53 + 61 + 63 = 177$ усл. ед. в течение трех лет при выделенных фирмой $54 + 62 + 70 = 186$ ден. ед. ▲

1.5. Транспортная задача

Постановка транспортной задачи

Транспортная задача (ТЗ) используется при разработке плана перевозок однородного вида продукции, сосредоточенного в нескольких пунктах отправления в пункты назначения.

Пункты отправления (ПО). Имеется m пунктов отправления $A_1, A_2, \ldots, A_m$, в которых сосредоточены грузы в количестве $a_1, a_2, \ldots, a_m$ ед.

Пункты назначения (ПН). Имеется n пунктов назначения $B_1, B_2, \ldots, B_n$, подавших заявки на $b_1, b_2, \ldots, b_n$ ед. товара.

Известны стоимости (тарифы) c_{ij} перевозок единиц товара от каждого ПО в каждый ПН.

Требуется составить такой план перевозок, при котором все заявки на товар были бы выполнены при минимальной стоимости всех перевозок.

В зависимости от соотношения между суммарными запасами груза и суммарными потребностями в нем ТЗ могут быть закрытыми и открытыми.

Если $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$, то задача называется *закрытой.*

Если $\sum_{i=1}^{m} a_i \neq \sum_{j=1}^{n} b_j$, то задача называется *открытой.*

Закрытая транспортная задача

Пусть x_{ij} — количество груза, отправляемого из пункта A_i в пункт B_j, причем $x_{ij} \geq 0$. Запишем условия задачи в виде транспортной таблицы (табл. 1.18).

Строки транспортной таблицы соответствуют ПО (в последней клетке каждой строки указан объем запаса груза), а столбцы — ПН (последняя клетка каждого столбца содержит значение потребности в товаре. Все клетки таблицы (кроме тех, которые расположены в нижней строке и в правом столбце) содержат информацию о перевозках x_{ij} и их стоимости c_{ij}.

Таблица 1.18

Транспортная таблица

ПО	**ПН**				**Запасы a_i**
	B_1	B_2	…	B_n	
A_1	c_{11} x_{11}	c_{12} x_{12}	…	c_{1n} x_{1n}	a_1
A_2	c_{21} x_{21}	c_{22} x_{22}	…	c_{2n} x_{2n}	a_2
…	…	…	…	…	…
A_m	c_{m1} x_{m1}	c_{m2} x_{m2}	…	c_{mn} x_{mn}	a_m
Заявки b_j	b_1	b_2	…	b_n	$\sum_i a_i = \sum_j b_j$

Запишем систему ограничений.

1. Суммарное количество груза, направленного из каждого ПО во все ПН, должно быть равно запасу груза в данном ПО:

$$\begin{cases} x_{11} + x_{12} + \ldots + x_{1n} = a_1, \\ x_{21} + x_{22} + \ldots + x_{2n} = a_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_{m1} + x_{m2} + \ldots + x_{mn} = a_m. \end{cases}$$

2. Суммарное количество груза, доставляемого в каждый ПН из всех ПО, должно быть равно заявке данного ПН:

$$\begin{cases} x_{11} + x_{21} + \ldots + x_{m1} = b_1, \\ x_{12} + x_{22} + \ldots + x_{m2} = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_{1n} + x_{2n} + \ldots + x_{mn} = b_n. \end{cases}$$

Суммарная стоимость всех перевозок (целевая функция) есть

$$L = c_{11}x_{11} + c_{12}x_{12} + \ldots + c_{mn}x_{mn} \to \min.$$

Планом перевозок называется любая совокупность значений переменных x_{ij} — матрица размера $m \times n$.

Допустимым решением называется план x_{ij}, удовлетворяющий системе ограничений ТЗ.

Оптимальным решением называется допустимое решение, доставляющее минимум целевой функции.

Система ограничений содержит mn неизвестных x_{ij} и $(m + n)$ уравнений. В силу соотношения $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$ только $(m + n - 1)$ уравнений из $(m + n)$ являются линейно независимыми. Значит, ранг системы ограничений равен $r = m + n - 1$, следовательно, имеется $(m + n - 1)$ базисных переменных, а остальные $k = mn - (m + n - 1) = (m - 1)(n - 1)$ — свободные.

Опорным решением называется допустимое решение, в котором отличны от нуля не более $(m + n - 1)$ базисных переменных, а остальные равны нулю.

ТЗ как задача ЛП может быть решена симплексным методом, однако благодаря особенностям переменных задачи и системы ограничений разработаны специальные методы. Наиболее применяемым методом является метод потенциалов.

Метод потенциалов включает следующие этапы:

- нахождение исходного опорного решения;
- проверка этого решения на оптимальность;
- переход от одного опорного решения к другому.

Нахождение исходного опорного решения. Решение задачи и ее исходное опорное решение будем записывать в транспортной таблице. В транспортной таблице клетки, которые содержат ненулевые перевозки ($x_{ij} > 0$), называются *занятыми* (им соответствуют базисные переменные опорного решения, число которых не превышает $(m + n - 1)$. Остальные клетки содержат нулевые перевозки ($x_{ij} = 0$) и называются *свободными* (им соответствуют свободные переменные).

Для нахождения исходного опорного решения используется метод северо-западного угла или метод наименьших стоимостей.

Метод северо-западного угла. Суть метода проста: ячейки транспортной таблицы последовательно заполняются максимально возможными объемами перевозок в направлении *сверху вниз* и *слева направо*. То есть сначала заполняется самая верхняя левая ячейка (*«северо-западная» ячейка*), потом следующая справа и т. д. Затем переходят на новую строку и вновь заполняют ее слева направо. И так пока таблица не будет заполнена полностью.

Недостатком данного метода является то, что он не учитывает значения элементов c_{ij} матрицы транспортных расходов, в результате чего полученный этим методом начальный опорный план перевозок может быть достаточно далек от оптимального.

Метод наименьших стоимостей. В различных модификациях метода наименьших стоимостей заполнение клеток матрицы перевозок проводится с учетом значений величин c_{ij}. Так, в модификации двойного предпочтения отмечают клетки с наименьшими стоимостями перевозок сначала по каждой строке, а затем по каждому столбцу. Клетки, имеющие две отметки, заполняют в первую очередь, затем заполняют клетки с одной отметкой, а данные о нераспределенном грузе записывают в неотмеченные клетки с наименьшими стоимостями. При этом из двух клеток с одинаковой стоимостью перевозок предпочтение отдается клетке, через которую осуществляется больший объем перевозок.

Проверка опорного плана на вырожденность. Если окажется, что число занятых клеток меньше, чем $(m + n - 1)$, то задача имеет *вырожденное* решение. Для его возможного исключения в свободную клетку с наименьшей стоимостью вводят нулевую поставку. Нуль помещают в такую клетку, чтобы в каждой строке и каждом столбце было не менее одной занятой клетки.

Проверка опорного решения на оптимальность. Найденное исходное опорное решение проверяется на оптимальность методом потенциалов. В методе потенциалов строке i и столбцу j транспортной таблицы ставятся в соответствие числа u_i, v_j (потенциалы производителей и потребителей).

Потенциалы u_i, v_j можно интерпретировать как плату, которую вносят за перевозку ед. груза каждый из пунктов отправления A_i в каждый из пунктов назначения B_j некоторому третьему лицу (перевозчику).

Обозначим $\overline{c}_{ij} = u_i + v_j$. Величину $\overline{c}_{ij}$ будем называть псевдостоимостью перевозки груза из A_i в B_j.

Потенциалы u_i, v_j находят из равенств $u_i + v_j = c_{ij}$ для $x_{ij} > 0$ (для занятых клеток). Придавая одному из потенциалов произвольное значение (обычно $u_1 = 0$), однозначно определяют остальные потенциалы.

Обозначим $\Delta_{ij} = \overline{c}_{ij} - c_{ij}$ (оценка свободной клетки).

Условие оптимальности: если для всех свободных клеток $\Delta_{ij} \le 0$, то опорное решение является оптимальным; если хотя бы одна из оценок $\Delta_{ij} > 0$, то опорное решение не является оптимальным.

Неоптимальное опорное решение можно улучшить, перейдя из одного опорного решения к другому.

Переход от одного опорного решения к другому. Для свободной клетки с $\Delta_{ij} > 0$ строится цикл. *Циклом* называется замкнутый контур, все вершины которого кроме одной находятся в занятых клетках; при этом направление отдельных отрезков контура могут быть только горизонтальными и вертикальными.

Около свободной клетки цикла ставится знак (+), затем поочередно проставляют знаки (–) и (+). У вершин со знаком (–) выбирают минимальный груз, его прибавляют к грузам, стоящим у вершин со знаком (+), и отнимают от грузов у вершин со знаком (–).

В результате перераспределения груза получают новое опорное решение. Это решение проверяют на оптимальность и т. д. до тех пор, пока не получится оптимальное решение.

Пример 1.26. Решим методом потенциалов ТЗ, заданную табл. 1.19, в которой представим первый опорный план, составленный по способу северо-западного угла.

▼Проверка опорного плана на вырожденность. Число занятых клеток 4 равно $m + n - 1 = 4$, т. е. опорное решение невырожденное.

Стоимость первого плана есть

$$L = 20 \cdot 1 + 3 \cdot 10 + 5 \cdot 10 + 4 \cdot 20 = 180.$$

Вычисление потенциалов для плана перевозки. Приписываем к таблице снизу добавочную строку v_j, а справа — добавочный столбец u_i.

Таблица 1.19

Условия транспортной задачи и первый опорный план для примера 1.26

ПО	**ПН**			a_i	u_i
	B_1	B_2	B_3		
A_1	1 20	3 10	2 5	30	0
A_2	3 2	5 10	4 20	30	2
b_j	20	20	20		
v_j	1	3	2		

Для каждой базисной клетки выполняются условия $u_i + v_j = c_{ij}$.

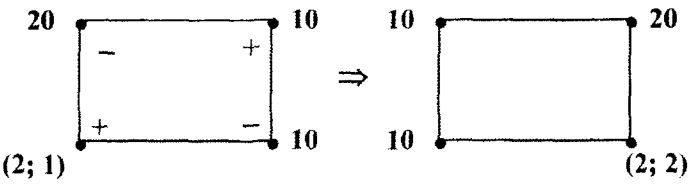
Полагая $u_1 = 0$, находим из этого условия остальные потенциалы:

$$u_1 + v_1 = 1,\ 0 + v_1 = 1 \Rightarrow v_1 = 1;$$
$$u_1 + v_2 = 3, 0 + v_2 = 3 \Rightarrow v_2 = 3;$$
$$u_2 + v_2 = 5, u_2 + 3 = 5 \Rightarrow u_2 = 2;$$
$$u_2 + v_3 = 4, 2 + v_3 = 4 \Rightarrow v_3 = 2.$$

Значения $\bar{c}_{ij} = u_i + v_j$ записываем в левом верхнем углу каждой свободной клетки.

Проверка плана на оптимальность методом потенциалов. Видно, что для свободной клетки (2; 1): $\Delta_{21} = \bar{c}_{21} - c_{21} = 3 - 2 > 0$, следовательно, план не оптимальный.

Перераспределение поставок. Строим для этой клетки цикл пересчета.



Получили новый опорный план (табл. 1.20).

Таблица 1.20

Новый опорный план для примера 1.26

ПО	ПН			a_i	u_i
	B_1	B_2	B_3		
A_1	1 10	3 20	3 5	30	0
A_2	2 10	4 5	4 20	30	1
b_j	20	20	20	60 = 60	
v_j	1	3	3		

Стоимость нового плана $L = 10 \cdot 1 + 2 \cdot 10 + 3 \cdot 20 + 4 \cdot 20 = 170.$

Вычислим для нового плана новые значения потенциалов.

Полагая $u_1 = 0$, находим из этого условия остальные потенциалы:

$$u_1 + v_1 = 1,\ 0 + v_1 = 1 \Rightarrow v_1 = 1;$$

$$u_1 + v_2 = 3,\ 0 + v_2 = 3 \Rightarrow v_2 = 3;$$

$$u_2 + v_1 = 2,\ u_2 + 1 = 2 \Rightarrow u_2 = 1;$$

$$u_2 + v_3 = 4,\ 1 + v_3 = 4 \Rightarrow v_3 = 3.$$

Значения $\overline{c}_{ij} = u_i + v_j$ приведены в левом верхнем углу каждой свободной клетки. Получили, что для всех свободных клеток $\overline{c}_{ij} < c_{ij}$, т. е. $\Delta_{ij} = \overline{c}_{ij} - c_{ij} < 0$, следовательно, новый план оптимальный.

Оптимальное решение:

$$X_{\text{опт}} = \begin{pmatrix} 0 & 45 & 0 \\ 0 & 0 & 30 \\ 20 & 0 & 0 \end{pmatrix},\ L_{\min} = 170 \text{ усл. ед.} \ \blacktriangle$$

Пример 1.27. Решим методом потенциалов ТЗ, заданную табл. 1.21, в которой представлен первый опорный план, составленный по способу минимальной стоимости.

Таблица 1.21

Условия задачи и первый опорный план для примера 1.27

ПО	ПН				a_i	u_i
	B_1	B_2	B_3	B_4		
A_1	3 6	4 3000	3 9	8 3000	6000	0
A_2	2 5	3 2000	2 1000	7 8	3000	–1
A_3	2 4000	3 0	2 6	7 8	4000	–1
b_j	4000	5000	1000	3000		
v_j	3	4	3	8		

▼Число занятых клеток 5 меньше $(m + n - 1) = 6$, следовательно, опорный план вырожденный. Для исключения вырожденности в клетку (3; 2) с наименьшей стоимостью по сравнению с другими свободными клетками включим нулевую поставку. Стоимость первого плана $L = 52\ 000$.

Рассчитаем потенциалы u_i, v_j. Из табл. 1.21 видно, что для всех свободных клеток $\overline{c}_{ij} < c_{ij}$, т. е. $\Delta_{ij} = \overline{c}_{ij} - c_{ij} < 0$ следовательно, исходный план оптимальный.

Оптимальное решение:

$$X_{\text{опт}} = \begin{pmatrix} 0 & 3000 & 0 & 3000 \\ 0 & 2000 & 1000 & 0 \\ 4000 & 0 & 0 & 0 \end{pmatrix}, L_{\min} = 52\ 000 \text{ усл. ед.}\ ▲$$

Открытая транспортная задача

Для открытой модели может быть два случая.

1. Суммарные запасы превышают суммарные потребности, т. е.

$$\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j .$$

В этом случае вводится фиктивный потребитель $B_ф$, потребности которого

$$b_ф = \sum a_i - \sum b_j .$$

2. Суммарные потребности превышают суммарные запасы, т. е.

$$\sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j .$$

В этом случае вводится фиктивный поставщик $A_ф$, запасы которого

$$a_ф = \sum b_j - \sum a_i .$$

Стоимость перевозок единицы груза как для фиктивного потребителя, так и для фиктивного поставщика полагают равной нулю, так как груз в обоих случаях не перевозится. После преобразования задача принимает вид закрытой модели и решается обычным способом.

Пример 1.28. Решим методом потенциалов ТЗ, заданную табл. 1.22.

Таблица 1.22

Условия транспортной задачи для примера 1.28

ПО	ПН			a_i
	B_1	B_2	B_3	
A_1	7	3	6	75
A_2	4	8	2	40
A_3	1	5	3	35
b_j	20	45	30	

▼ $\sum a_i = 150 > \sum b_j = 95$ — открытая модель. Вводим фиктивного потребителя $B_ф$, потребности которого $b_ф = 150 - 95 = 55$, с нулевыми значениями стоимости перевозок единицы товара. В табл. 1.23 представлен первый опорный план, составленный по способу минимальной стоимости.

Таблица 1.23

Первый опорный план для примера 1.28

ПО	ПН				a_i	u_i
	B_1	B_2	B_3	$B_ф$		
A_1	1 7	3 45	2 6	0 30	75	0
A_2	1 4	3 8	2 30	0 10	40	0
A_3	1 20	3 5	2 3	0 15	35	0
b_j	20	45	30	55		
v_j	1	3	2	0		

Число занятых клеток 6 равно $m + n - 1 = 6$, т. е. опорное решение невырожденное. Стоимость первого плана $L = 215$.

Рассчитываем потенциалы u_i, v_j. Из табл. 1.23 видно, что для всех свободных клеток $\bar{c}_{ij} < c_{ij}$, т. е. $\Delta_{ij} = \bar{c}_{ij} - c_{ij} < 0$, следовательно, исходный план оптимальный.

Оптимальное решение:

$$X_{опт} = \begin{pmatrix} 0 & 45 & 0 \\ 0 & 0 & 30 \\ 20 & 0 & 0 \end{pmatrix}, \quad L_{\min} = 215 \text{ усл. ед.} \ \blacktriangle$$

Решение ТЗ в *Excel* (*Поиск решения*)

Надстройка *Поиск решения* в *Microsoft Excel* позволяет напрямую находить оптимальное решение транспортной задачи.

I. Закрытая модель, $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$.

До вызова *Поиск решения* необходимо подготовить данные для решения ТЗ на рабочем листе *Excel*.

Пример 1.29. На трех складах имеется мука в количестве 60, 130, 90 т, которая должна быть доставлена четырем хлебозаводам в количестве 30, 80, 60, 110 т соответственно. Составить оптимальный план перевозок, имеющий минимальные транспортные расходы, если стоимость доставки 1 т муки на хлебозаводы задана матрицей

$$\begin{pmatrix} 6 & 8 & 15 & 4 \\ 9 & 15 & 2 & 3 \\ 6 & 12 & 7 & 10 \end{pmatrix}.$$

▼Вид рабочего листа *Excel* с исходными данными рассматриваемого примера приведен на рис. 1.28.

	A	B	C	D	E	F
1		B_1	B_2	B_3	B_4	a
2	A_1	6	8	15	4	60
3	A_2	9	15	2	3	130
4	A_3	6	12	7	10	90
5	b	30	80	60	110	
6						
7		B_1	B_2	B_3	B_4	Σ
8	A_1	0	0	0	0	0
9	A_2	0	0	0	0	0
10	A_3	0	0	0	0	0
11	Σ	0	0	0	0	
12						
13		min L =	0			

	H	I	J	K
1	В ограничениях Поиска решения			
2	**закрытая**	модель	установить:	
3	Σa =	280	строка Σ	=
4	Σb =	280	столбец Σ	=

Рис. 1.28. Исходные данные в *Excel* для примера 1.29

Приведем последовательность ввода данных задачи и формул в соответствующие ячейки.

A1:F5 — вводится транспортная таблица примера.

I3 — вычисляется сумма по строкам по формуле =СУММ(F2:F4).

I4 — вычисляется сумма по столбцам по формуле =СУММ(B5:E5).

Из сравнения значений делается сообщение о типе модели (закрытая, открытая) и типе ограничений (=, < =) в окне *Поиск решения*.

A7:F11 — составляется начальная матрица перевозок (вначале копируется транспортная таблица, затем редактируется)

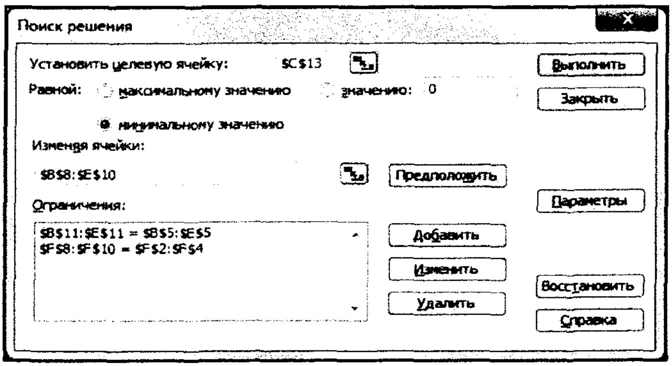
B8:E10 — производится резервирование и обнуление изменяемых переменных (начальная матрица перевозок).

F8:F10 и B11:E11 — подсчитываются суммы Σ по строкам и столбцам начальной матрицы перевозок.

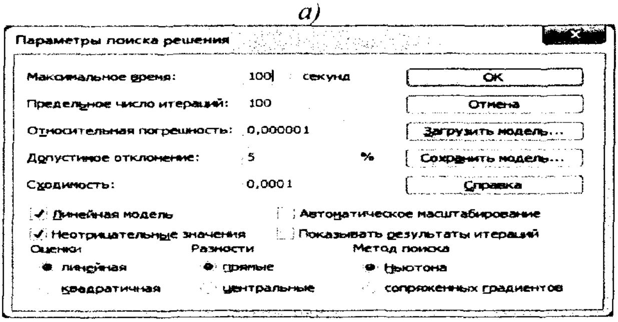
C13 — вычисляется стоимость начального плана L по формуле =СУММПРОИЗВ(B2:E4;B8:E10).

Далее нужно выполнить команду *Поиск решения*. В диалоговом окне *Поиск решения*:

- установить целевую ячейку L равной минимальному значению;
- в качестве изменяемых ячеек ввести матрицу перевозок x_{ij};
- установить ограничение для всех отправителей и всех потребителей;
- щелкнуть по кнопке *Параметры* и в диалоговом окне установить переключатели *Линейная модель* и *Неотрицательные переменные* (рис. 1.29).



а)



б)

Рис. 1.29. Заполненные диалоговые окна *Поиск решения* (*а*) и *Параметры поиска решения* (*б*) для примера 1.29

Вид рабочего листа *Excel* после выполнения команды *Поиск решения* представлен на рис. 1.30.

	B_1	B_2	B_3	B_4	Σ
A_1	0	60	0	0	**60**
A_2	0	0	**20**	110	**130**
A_3	30	20	40	0	**90**
Σ	**30**	**80**	**60**	**110**	

min L =	**1550**

Рис. 1.30. Пример 1.29 в *Excel* после выполнения команды *Поиск решения*

Оптимальное решение:

$$X_{\text{опт}} = \begin{pmatrix} 0 & 60 & 0 & 0 \\ 0 & 0 & 20 & 110 \\ 30 & 20 & 40 & 0 \end{pmatrix}, \ L_{\min} = 1550 \text{ усл. ед.} \ \blacktriangle$$

Пример 1.30. На предприятии имеются три группы станков, каждая из которых может выполнять пять операций по обработке деталей (в любом порядке). Максимальное время работы каждой группы станков соответственно 100; 250; 180 ч. Каждая операция должна выполняться соответственно 100; 120; 70; 110; 130 ч. Продолжительность работы i-й группы станков при выполнении j-й операции задана матрицей

$$\begin{pmatrix} 3 & 5 & 11 & 10 & 5 \\ 5 & 10 & 15 & 3 & 2 \\ 4 & 8 & 6 & 12 & 10 \end{pmatrix}.$$

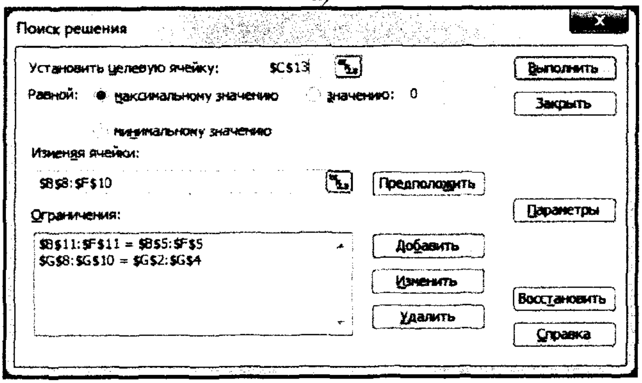
Определить, сколько времени и на какую операцию нужно использовать каждую группу станков, чтобы обработать максимальное количество деталей.

▼Вид рабочего листа *Excel* с исходными данными примера и заполненного диалогового окна *Поиск решения* представлен на рис. 1.31.

	A	B	C	D	E	F	G
1		B_1	B_2	B_3	B_4	B_5	a
2	A_1	3	5	11	10	5	100
3	A_2	5	10	15	3	2	250
4	A_3	4	8	6	12	10	180
5	b	100	120	70	110	130	
6							
7		B_1	B_2	B_3	B_4	B_5	Σ
8	A_1	0	0	0	0	0	**0**
9	A_2	0	0	0	0	0	**0**
10	A_3	0	0	0	0	0	**0**
11	Σ	0	0	0	0	0	
12							
13		max L =	0				

I	J	K	L
В ограничениях Поиска решения			
закрытая	модель	установить:	
Σa =	530	строка Σ	=
Σb =	530	столбец Σ	=

а)



б)

Рис. 1.31. Исходные данные в *Excel* (*а*) и заполненное диалоговое окно *Поиск решения* (*б*) для примера 1.30

Вид рабочего листа *Excel* после выполнения команды *Поиск решения* представлен на рис. 1.32.

	B_1	B_2	B_3	B_4	B_5	Σ
A_1	40	0	0	60	0	**100**
A_2	60	120	70	0	0	**250**
A_3	0	0	0	50	130	**180**
Σ	**100**	**120**	**70**	**110**	**130**	

max L =	**5170**

Рис. 1.32. Пример 1.30 в *Excel* после выполнения команды *Поиск решения*

Оптимальное решение:

$$X_{\text{опт}} = \begin{pmatrix} 40 & 0 & 0 & 60 & 0 \\ 60 & 120 & 70 & 0 & 0 \\ 0 & 0 & 0 & 50 & 130 \end{pmatrix}, \ L_{\max} = 5170 \text{ шт.}$$

Таким образом, по 1-й группе станков целесообразно выполнять операции 1 и 4 продолжительностью 40 и 60 ч соответственно; по 2-й группе — операции 1, 2, 3 продолжительностью 60, 120, 70 ч соответственно; по 3-й группе — операции 4, 5 продолжительностью 50, 130 ч соответственно. При этом максимальное число обработанных деталей составит 5170 шт. ▲

II. Открытая модель, $\sum_{i=1}^{m} a_i \neq \sum_{j=1}^{n} b_j$.

Первый способ. Открытая транспортная задача решается сведением ее к закрытой транспортной задаче.

Для открытой модели может быть два случая

1. Суммарные запасы превышают суммарные потребности, т. е. $\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j$. В этом случае вводится фиктивный потребитель $B_ф$, потребности которого $b_ф = \sum a_i - \sum b_j$.

2. Суммарные потребности превышают суммарные запасы, т. е. $\sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j$. В этом случае вводится фиктивный поставщик $A_ф$, запасы которого $a_ф = \sum b_j - \sum a_i$.

Стоимость перевозок единицы груза как для фиктивного потребителя, так и для фиктивного поставщика полагают равной нулю, так как груз в обоих случаях не перевозится. После преобразования задача принимает вид закрытой модели и решается обычным способом.

Второй способ. Для открытой модели может быть два случая:

- $\sum a_i > \sum b_j$. Неравенство (расчетное $\sum a_i \leq$ заданного $\sum a_i$) означает условие неполного распределения запасов;

- $\sum a_i < \sum b_j$. Неравенство (расчетное $\sum b_j \leq$ заданного $\sum b_j$) означает условие неполного удовлетворения потребностей.

Соответственно с этим необходимо в диалоговом окне *Поиск решения* в группе ограничений по столбцу Σ (по строке Σ) заменить знак «=» на «≤».

Пример 1.31. Три торговых склада могут поставлять некоторое изделие в количестве 9, 4, 8 т. Величина спроса трех магазинов розничной торговли на это изделие равна 3, 5, 6 т соответственно. Стоимость перевозок единицы груза от каждого поставщика к каждому потребителю задана матрицей

$$\begin{pmatrix} 10 & 20 & 5 \\ 2 & 10 & 8 \\ 1 & 20 & 7 \end{pmatrix}.$$

Требуется составить такой план перевозок, при котором спрос магазинов на изделия был бы удовлетворен при минимальной стоимости всех перевозок.

▼*Первый способ.* Составляем транспортную таблицу (рис. 1.33).

	A	B	C	D	E
1		B_1	B_2	B_3	a
2	A_1	10	20	5	9
3	A_2	2	10	8	4
4	A_3	1	20	7	8
5	b	3	5	6	

Рис. 1.33. Исходные данные в *Excel* для примера 1.31

$\sum a_i = 21 > \sum b_j = 14$ — открытая модель. Вводим фиктивного потребителя $B_ф$, потребности которого $b_ф = 21 - 14 = 7$ с нулевыми значениями перевозок единицы товара (рис. 1.34).

Составляем матрицу перевозок (x_{ij}) и формулу для целевой функции.

	A	B	C	D	E	F
1		B_1	B_2	B_3	$B_ф$	a
2	A_1	10	20	5	0	9
3	A_2	2	10	8	0	4
4	A_3	1	20	7	0	8
5	b	3	5	6	7	

Рис. 1.34. Исходные данные в *Excel* с введенным фиктивным потребителем для примера 1.31

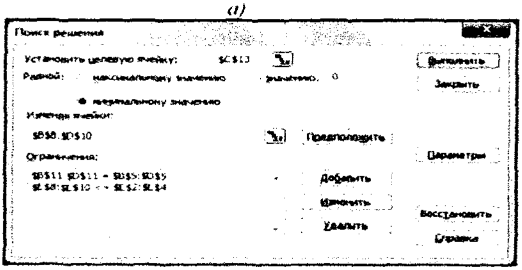
Оптимальное решение:

$$X_{\text{опт}} = \begin{pmatrix} 0 & 1 & 6 \\ 0 & 4 & 0 \\ 3 & 0 & 0 \end{pmatrix}, \ L_{\min} = 93 \text{ . } \blacktriangle$$

▼*Второй способ*. Рабочий лист *Excel* и заполненное диалоговое окно *Поиск решения* имеют вид, представленный на рис. 1.35.

	A	B	C	D	E	F	G	H	I	J
1		B_1	B_2	B_3	a		В ограничениях Поиска решения			
2	A_1	10	20	5	9		**открытая**	модель	установить:	
3	A_2	2	10	8	4		Σa	21	строка Σ	=
4	A_3	1	20	7	8		Σb	14	столбец Σ	<=
5	b	3	5	6						
6										
7		B_1	B_2	B_3	Σ					
8	A_1	0	0	0	0					
9	A_2	0	0	0	0					
10	A_3	0	0	0	0					
11	Σ	0	0	0						
12										
13		min L	0							

а)



б)

Рис. 1.35. Исходные данные в *Excel* (*а*) и заполненное диалоговое окно *Поиск решения* (*б*) для примера 1.31 (второй способ)

Неравенство $E$8:$E10 <= $E$2:$E$4 означает условие неполного распределения запасов.

Вид рабочего листа *Excel* после выполнения команды *Поиск решения* представлен на рис. 1.36.

	B_1	B_2	B_3	Σ
A_1	0	1	6	7
A_2	0	4	0	4
A_3	3	0	0	3
Σ	3	5	6	

min L =	93

Рис. 1.36. Пример 1.31 в *Excel*
после выполнения команды *Поиск решения*

Оптимальное решение $X_{\text{опт}} = \begin{pmatrix} 0 & 1 & 6 \\ 0 & 4 & 0 \\ 3 & 5 & 6 \end{pmatrix}$, $L_{\min} = 93$. ▲

Пример 1.32. Составить оптимальный план перевозки грузов от трех поставщиков с грузами 240, 40, 110 т к четырем потребителям с запросами 90, 190, 40, 130 т. Стоимость перевозок единицы груза от каждого поставщика к каждому потребителю задана матрицей

$$\begin{pmatrix} 7 & 13 & 9 & 8 \\ 14 & 8 & 7 & 10 \\ 3 & 15 & 20 & 6 \end{pmatrix}.$$

▼*Первый способ.* Составляем транспортную таблицу (рис. 1.37).

	A	B	C	D	E	F
1		B_1	B_2	B_3	B_4	a
2	A_1	7	13	9	8	240
3	A_2	14	8	7	10	40
4	A_3	3	15	20	6	110
5	b	90	190	40	130	

Рис. 1.37. Исходные данные в *Excel* для примера 1.32

$\sum a_i = 390 < \sum b_j = 450$ — открытая модель. Вводим фиктивного поставщика $A_ф$, запасы которого $a_ф = 450 - 390 = 60$, с нулевыми значениями стоимости перевозок единицы товара (рис. 1.38).

	A	B	C	D	E	F
1		B_1	B_2	B_3	B_4	a
2	A_1	7	13	9	8	240
3	A_2	14	8	7	10	40
4	A_3	3	15	20	6	110
5	$A_ф$	0	0	0	0	60
6	b	90	190	40	130	

Рис. 1.38. Исходные данные в *Excel* с введенным фиктивным поставщиком для примера 1.32

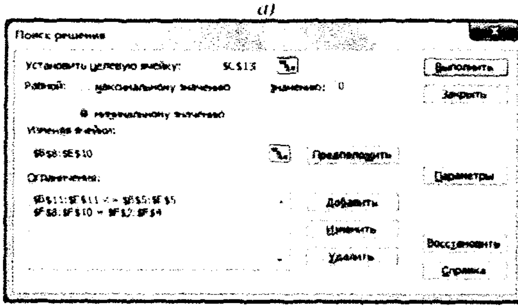
Оптимальное решение:

$$X_{опт} = \begin{pmatrix} 0 & 90 & 40 & 110 \\ 0 & 40 & 0 & 0 \\ 90 & 0 & 0 & 20 \end{pmatrix}, \; L_{\min} = 3120. \; \blacktriangle$$

▼*Второй способ.* Рабочий лист *Excel* и заполненное диалоговое окно *Поиск решения* имеют вид, представленный на рис. *1.39.*

	A	B	C	D	E	F	G	H	I	J	K
1		B_1	B_2	B_3	B_4	a		В ограничениях Поиска решения			
2	A_1	7	13	9	8	240		**открытая**	модель	установить:	
3	A_2	14	8	7	10	40		$\Sigma a =$	390	*строка* Σ	<=
4	A_3	3	15	20	6	110		$\Sigma b =$	450	столбец Σ	=
5	b	90	190	40	130						
6											
7		B_1	B_2	B_3	B_4	Σ					
8	A_1	0	0	0	0	0					
9	A_2	0	0	0	0	0					
10	A_3	0	0	0	0	0					
11	Σ	0	0	0	0						
12											
13		$\min L =$	0								

а)



б)

Рис. 1.39. Исходные данные в *Excel* (*а*) и заполненное диалоговое окно *Поиск решения* (*б*) для примера 1.32 (второй способ)

Неравенство \$B\$11:\$E11 <= \$B\$5:\$E\$5 означает условие неполного удовлетворения потребностей.

Вид рабочего листа *Excel* после выполнения команды *Поиск решения* представлен на рис. 1.40.

	B_1	B_2	B_3	B_4	Σ
A_1	0	90	40	110	**240**
A_2	0	40	0	0	**40**
A_3	90	0	0	20	**110**
Σ	**90**	**130**	**40**	**130**	

min L =	**3120**

Рис. 1.40. Пример 1.32 в *Excel* после выполнения команды *Поиск решения*

Оптимальное решение:

$$X_{\text{опт}} = \begin{pmatrix} 0 & 90 & 40 & 110 \\ 0 & 40 & 0 & 0 \\ 90 & 0 & 0 & 20 \end{pmatrix}, \ L_{\min} = 3120. \ \blacktriangle$$

Задача о назначениях

Задача о назначениях имеет место при назначении людей на должности или работы, автомашин на маршруты, водителей на машины и т. п.

В наиболее общей форме задача о назначениях формулируется следующим образом. Имеется некоторое число работ и некоторое число работников. Любой работник может быть назначен на выполнение любой (но только одной) работы, но с неодинаковыми затратами. Нужно распределить работы так, чтобы выполнить работы с минимальными затратами.

Задача о назначениях является частным случаем транспортной задачи, в которой работники соответствуют пунктам отправления, а работы — пунктам назначения. Особенность лишь в том, что все переменные решения принимают только значения 0 или 1 (двоичные переменные) и в каждом столбце и строке может быть только одно ненулевое значение.

Точно так же, как и транспортная модель, модель назначений может быть несбалансированной, содержать недопустимые назначения, иметь альтернативные решения при одном и том же значении целевой функции. Эти варианты моделей назначения строятся в полной аналогии с соответствующими транспортными моделями.

Для решения задачи о назначениях в *Excel* с использованием надстройки *Поиск решения* на рабочем листе следует выделить ячейки назначений и подсчитать для них суммы по столбцам и по строкам. В ячейку целевой функции следует ввести формулу, вычисляющую сумму произведений затрат работ на план назначений.

В диалоговом окне *Поиск решения* нужно выбрать целевую ячейку, изменяемые ячейки и добавить ограничения: суммы значений изменяемых ячеек для каждой строки и столбца должны быть равны 1; переменные должны быть двоичными.

Пример 1.33. Администрация предприятия приняла на работу пять человек. Каждый из них затрачивает различное время на выполнение определенной работы. Необходимо выполнить пять видов работ. Время выполнения работы каждым работником приведено в табл. 1.24.

Таблица 1.24

Условия примера 1.33

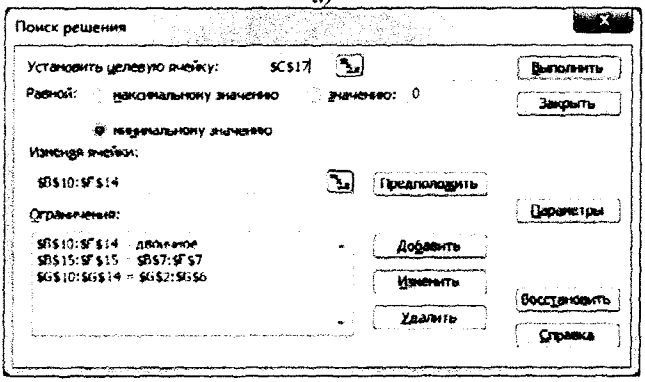
Работник	**Время выполнения работы по видам, ч**				
	1	**2**	**3**	**4**	**5**
A_1	25	16	15	14	13
A_2	25	17	18	23	15
A_3	30	15	20	19	14
A_4	27	20	22	25	12
A_5	29	19	17	32	10

Требуется назначить на каждый вид работы одного из работников, чтобы общее время, необходимое для завершения всех видов работ, было минимальным.

▼Решаем задачу в *Excel*. Рабочий лист *Excel* и заполненное диалоговое окно *Поиск решения* имеют вид, представленный на рис. 1.41.

	A	B	C	D	E	F	G
1		B_1	B_2	B_3	B_4	B_5	a
2	A_1	25	16	15	14	13	1
3	A_2	25	17	18	23	15	1
4	A_3	30	15	20	19	14	1
5	A_4	27	20	22	25	12	1
6	A_5	29	19	17	32	10	1
7	b	1	1	1	1	1	
8							
9		B_1	B_2	B_3	B_4	B_5	Σ
10	A_1	0	0	0	0	0	0
11	A_2	0	0	0	0	0	0
12	A_3	0	0	0	0	0	0
13	A_4	0	0	0	0	0	0
14	A_5	0	0	0	0	0	0
15	Σ	0	0	0	0	0	
16							
17		min L	0				

а)



б)

Рис. 1.41. Исходные данные в *Excel* (*а*) и заполненное диалоговое окно *Поиск решения* (*б*) для примера 1.33

Вид рабочего листа *Excel* после выполнения команды *Поиск решения* представлен на рис. 1.42.

Таблица переменных состоит из единиц и нулей. По единицам определяем, что 1-й работник должен работать на четвертом виде работы, 2-й — на первом, 3-й — на втором, 4-й — на пятом, 5-й — на третьем. Общее время выполнения всех видов работ составляет 83 ч.

	B_1	B_2	B_3	B_4	B_5	Σ
A_1	0	0	0	1	0	**1**
A_2	1	0	0	0	0	**1**
A_3	0	1	0	0	0	**1**
A_4	0	0	0	0	1	**1**
A_5	0	0	1	0	0	**1**
Σ	**1**	**1**	**1**	**1**	**1**	

min L =	**83**

Рис. 1.42. Пример 1.33 в *Excel* после выполнения команды *Поиск решения* ▲

Пример 1.34. В распоряжении некоторой компании имеется шесть торговых точек и шесть продавцов. Из опыта известно, что эффективность работы продавцов в различных торговых точках неодинакова. Коммерческий директор оценил деятельность каждого продавца в каждой торговой точке при условии, что первый продавец назначается во вторую торговую точку. Результаты этой оценки представлены в табл. 1.25.

Таблица 1.25

Условия примера 1.34

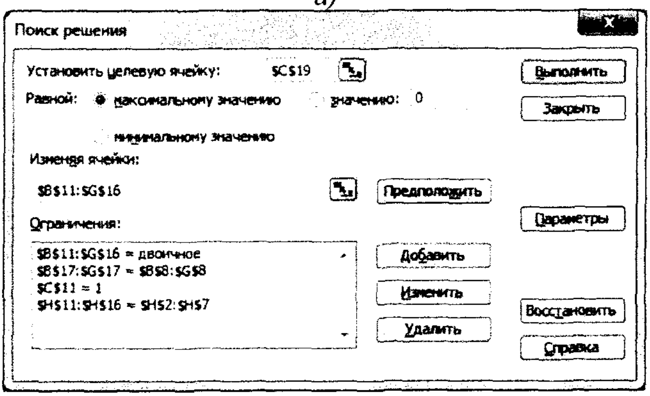
Продавец	**Объем продаж по торговым точкам, ед.**					
	1	**2**	**3**	**4**	**5**	**6**
A_1	68	72	75	83	75	69
A_2	56	60	58	63	61	59
A_3	35	38	40	45	25	27
A_4	40	42	47	45	53	36
A_5	62	70	68	67	69	70
A_6	65	63	69	70	72	68

Как коммерческий директор должен осуществить назначение продавцов, чтобы достичь максимального объема продаж?

▼Решаем задачу в *Excel*. Рабочий лист *Excel* и заполненное диалоговое окно *Поиск решения* имеют вид, представленный на рис. 1.43.

	A	B	C	D	E	F	G	H
1		B_1	B_2	B_3	B_4	B_5	B_6	a
2	A_1	68	72	75	83	75	69	1
3	A_2	56	60	58	63	61	59	1
4	A_3	35	38	40	45	25	27	1
5	A_4	40	42	47	45	53	36	1
6	A_5	62	70	68	67	69	70	1
7	A_6	65	63	69	70	72	68	1
8	b	1	1	1	1	1	1	
9								
10		B_1	B_2	B_3	B_4	B_5	B_6	Σ
11	A_1	0	0	0	0	0	0	**0**
12	A_2	0	0	0	0	0	0	**0**
13	A_3	0	0	0	0	0	0	**0**
14	A_4	0	0	0	0	0	0	**0**
15	A_5	0	0	0	0	0	0	**0**
16	A_6	0	0	0	0	0	0	**0**
17	Σ	**0**	**0**	**0**	**0**	**0**	**0**	
18								
19		max L =	0					

а)



б)

Рис. 1.43. Исходные данные в *Excel* (*а*) и заполненное диалоговое окно *Поиск решения* (*б*) для примера 1.34

Вид рабочего листа *Excel* после выполнения команды *Поиск решения* представлен на рис. 1.44.

По единицам таблицы определяем закрепление продавцов за торговыми точками, при этом объем продаж составляет 365 ед.

	B_1	B_2	B_3	B_4	B_5	B_6	Σ
A_1	0	1	0	0	0	0	**1**
A_2	1	0	0	0	0	0	**1**
A_3	0	0	0	1	0	0	**1**
A_4	0	0	0	0	1	0	**1**
A_5	0	0	0	0	0	1	**1**
A_6	0	0	1	0	0	0	**1**
Σ	**1**	**1**	**1**	**1**	**1**	**1**	

max L =	**365**

Рис. 1.44. Пример 1.34 в *Excel* после выполнения команды *Поиск решения* ▲

Пример 1.35. На предприятии имеется 6 автомобилей разных моделей.

Необходимо в разные районы области перевести 5 грузов. Затраты по перевозке каждого груза каждым автомобилем различны и приведены в табл. 1.26.

Таблица 1.26

Условия примера 1.35

Автомобиль	**Затраты по перевозке груза, ден. ед.**				
	B_1	B_2	B_3	B_4	B_5
A_1	37	17	52	73	72
A_2	11	39	70	20	27
A_3	12	21	25	11	30
A_4	49	35	36	35	74
A_5	40	31	78	66	79
A_6	77	14	59	67	78

Требуется выбрать автомобиль для каждого вида груза так, чтобы затраты на перевозку были минимальными. Определить эти затраты.

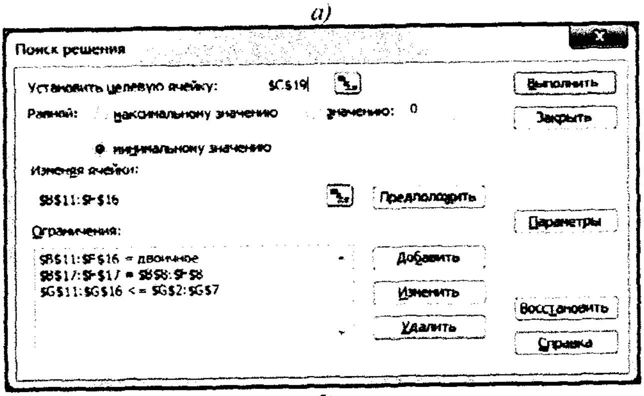
▼Решаем задачу в *Excel*. В данном примере число автомобилей больше, чем грузов, т. е. один автомобиль окажется невостребованным. По этой причине в группе ограничений по столбцу необходимо заменить знак «=» на «≤».

Рабочий лист *Excel* и заполненное диалоговое окно *Поиск решения* имеют вид, представленный на рис. 1.45.

Вид рабочего листа *Excel* после выполнения команды *Поиск решения* представлен на рис. 1.46.

	A	B	C	D	E	F	G
1		B_1	B_2	B_3	B_4	B_5	a
2	A_1	37	17	52	73	72	1
3	A_2	11	39	70	20	27	1
4	A_3	12	21	25	11	30	1
5	A_4	49	35	36	35	74	1
6	A_5	40	31	78	66	79	1
7	A_6	77	14	59	67	78	1
8	b	1	1	1	1	1	
9							
10		B_1	B_2	B_3	B_4	B_5	Σ
11	A_1	0	0	0	0	0	0
12	A_2	0	0	0	0	0	0
13	A_3	0	0	0	0	0	0
14	A_4	0	0	0	0	0	0
15	A_5	0	0	0	0	0	0
16	A_6	0	0	0	0	0	0
17	Σ	0	0	0	0	0	
18							
19		min L =	0				

а)



б)

Рис. 1.45. Исходные данные в *Excel* (*а*) и заполненное диалоговое окно *Поиск решения* (*б*) для примера 1.35

По единицам таблицы определяем закрепление автомобилей за грузами, при этом затраты по перевозке грузов составляют 125 ден. ед.

	B_1	B_2	B_3	B_4	B_5	Σ
A_1	1	0	0	0	0	**1**
A_2	0	0	0	0	1	**1**
A_3	0	0	0	1	0	**1**
A_4	0	0	1	0	0	**1**
A_5	0	0	0	0	0	**0**
A_6	0	1	0	0	0	**1**
Σ	**1**	**1**	**1**	**1**	**1**	

min L =	**125**

Рис. 1.46. Пример 1.35 в *Excel* после выполнения команды *Поиск решения* ▲

Упражнение 1.7. Фирма получила заказы на разработку пяти программных продуктов. Для выполнения этих заказов решено привлечь пятерых наиболее опытных программистов. Каждый из них должен написать одну программу. В табл. 1.27 приведены оценки времени, необходимого программистам для выполнения каждой из этих работ.

Таблица 1.27

Условия упражнения 1.7

Программист	**Время выполнения работ по видам, дней**				
	1	**2**	**3**	**4**	**5**
A_1	46	59	24	62	67
A_2	47	56	32	55	70
A_3	44	52	19	61	60
A_4	47	59	17	64	73
A_5	43	65	20	60	75

Оценки даны самими программистами, и у фирмы нет основания им не доверять. Распределите работы между программистами, чтобы общее количество человеко-дней, затраченное на выполнение всех пяти заказов, было минимальным.

Ответ: 1-й программист должен работать над программой 5, 2-й — над программой 4, 3-й — над программой 2, 4-й — над программой 3, 5-й — над программой 1. Общее время завершения всех видов работ составляет 234 человеко-дня.

Упражнение 1.8. Фирма получила заказы на разработку пяти программных продуктов. В фирме работают шесть квалифицированных программистов, которым можно поручить выполнение этих заказов. Каждый из них дал оценку времени, требуемого для разработки программ. Эти оценки приведены в табл. 1.28.

Таблица 1.28

Условия упражнения 1.8

Программист	**Время разработки программ по видам, дней**				
	1	**2**	**3**	**4**	**5**
A_1	46	59	24	62	67
A_2	47	56	32	55	70
A_3	44	52	19	61	60
A_4	47	59	17	64	73
A_5	43	65	20	60	75
A_6	41	53	28	54	68

Выполнение каждого из пяти заказов фирма решила поручить одному программисту. Ясно, что один из программистов не получит заказа.

Каждому программисту, которому будет поручено выполнять заказ, фирма предложила оплату 1 тыс. руб. в день. Распределите работу между программистами, чтобы общие издержки на разработку программ были минимальными.

Ответ: 1-й программист не получит заказ, 2-й будет работать над программой 4, 3-й — над программой 5, 4-й — над программой 3, 5-й — над программой 1, 6-й — над программой 2. Общие издержки составляют 228 тыс. руб.

Глава 2.
НЕЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

2.1. Задача нелинейного программирования

В общем виде задача нелинейного программирования (НП) может быть записано в виде

$$F(x_1, x_2, ..., x_n) \to \max(\min)$$

при ограничениях

$$\begin{cases} g_1(x_1, x_2, ..., x_n) \le b_1, \\ g_2(x_1, x_2, ..., x_n) \le b_2, \\ \dots\dots\dots\dots\dots\dots\dots \\ g_m(x_1, x_2, ..., x_n) \le b_m, \end{cases}$$

где F, g_i — известные функции;

b_i — заданные числа.

В задачах нелинейного программирования, в отличие от задач линейного программирования, целевая функция F или функции ограничений g_i не являются линейными. При решении задач НП для целевой функции F необходимо определить глобальный экстремум. Наличие локальных экстремумов затрудняет решение задачи, так как большинство существующих методов НП не позволяет установить, является ли найденный экстремум локальным или глобальным. Для задач НП, в отличие от линейных задач, нет единого метода решения.

2.2. Графическое решение задач нелинейного программирования

Рассмотрим задачу НП с двумя переменными (на плоскости).

$$F(x_1, x_2) \to \max(\min)$$

при ограничениях

$$\begin{cases} g_1(x_1, x_2) \le b_1, \\ g_2(x_1, x_2) \le b_2, \\ \dots\dots\dots\dots\dots\dots \\ g_m(x_1, x_2) \le b_m. \end{cases}$$

Множество точек $x = (x_1, x_2)$ на плоскости, удовлетворяющих ограничениям задачи, называется областью допустимых решений (ОДР).

Решить задачу НП графически — значит найти такую точку $x = (x_1, x_2)$ из ОДР, через которую проходит линия уровня $F(x_1, x_2) = C$, соответствующая максимальному (минимальному) значению целевой функции.

Наиболее существенное отличие задач НП от задач ЛП заключается в том, что оптимальное решение может находиться как на границе ОДР, так и являться его внутренней точкой.

Этапы графического решения задач НП следующие:

- построить ОДР задачи, определяемую системой ограничений;
- построить линии уровня целевой функции $F(x_1, x_2) = C$ при различных значениях параметра C;
- определить направление возрастания или убывания линий уровня целевой функции;
- найти точку ОДР, через которую проходит линия уровня целевой функции, соответствующая ее максимальному или минимальному значению.

Решение задачи нелинейного программирования средствами *Excel* отличается от решения задачи ЛП следующим:

- начальные значения переменных не должны быть равны нулю;

• в окне *Параметры поиска решения* необходимо убрать флажки *Линейная модель* и, если это необходимо, *Неотрицательные значения.*

Задача с линейной целевой функцией и нелинейной системой ограничений

Пример 2.1. Найти глобальный максимум функции

$$F(x) = 2x_1 + x_2 \to \max$$

при ограничениях

$$x_1^2 + x_2^2 \le 16\ ;$$

$$x_1, x_2 \ge 0.$$

▼Область допустимых решений (ОДР) — часть окружности радиуса 4 с центром в точке $O(0;\ 0)$ (рис. 2.1).

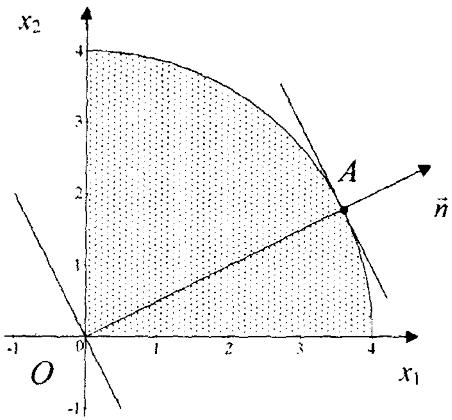


Рис. 2.1. Область допустимых решений примера 2.1

Линии уровня целевой функции $2x_1 + x_2 = C$ есть параллельные прямые.

Построим вектор $\vec{n} = gradF(x) = (2;1)$. Вектор $\vec{n}$ указывает направление наибольшего возрастания целевой функции $F(x)$ и перпендикулярен линиям уровня.

Глобальный минимум, равный нулю, достигается в точке $O(0;\ 0)$, т. е. $F_{\min}(0;\ 0) = 0$. Глобальный максимум достигается в точке A касания линии уровня и окружности. Найдем координаты точки A.

В точке A градиенты целевой функции и окружности $x_1^2 + x_2^2 = 16$ лежат на одной прямой (коллинеарны), причем

$$gradF(x) = (2;1)\,,\ grad(x_1^2 + x_2^2) = (2x_1; 2x_2)\,.$$

Коллинеарность двух векторов предполагает пропорциональность их соответствующих координат, т. е.

$$\frac{2x_1}{2} = \frac{2x_2}{1} \Rightarrow x_1 = 2x_2.$$

Для получения координат точки A решим систему

$$\begin{cases} x_1 = 2x_2, \\ x_1^2 + x_2^2 = 16. \end{cases} \Rightarrow x_1 = \frac{8}{\sqrt{5}},\ x_2 = \frac{4}{\sqrt{5}},\ F(x) = 2\cdot\frac{8}{\sqrt{5}} + \frac{4}{\sqrt{5}} = \frac{20}{\sqrt{5}},$$

т. е. $F_{\max}\left(\frac{8}{\sqrt{5}}; \frac{4}{\sqrt{5}}\right) = \frac{20}{\sqrt{5}}$.

Решим пример 2.1 в *Excel*. Рабочий лист *Excel* имеет вид, представленный на рис. 2.2.

	A	B	C	D	E
1		x_1	x_2		
2		1	1		
3	F	2	1	3	
4	b_1			2	16

Рис. 2.2. Условия примера 2.1 в *Excel*

Данные задачи и формулы в соответствующие ячейки вводятся следующим образом.

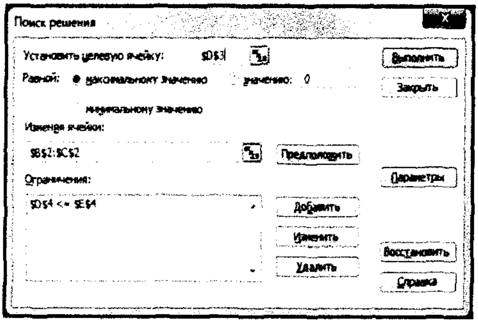
B2:C2 — введены ненулевые значения переменных.

B3:C3 — записаны коэффициенты при переменных модели в целевой функции.

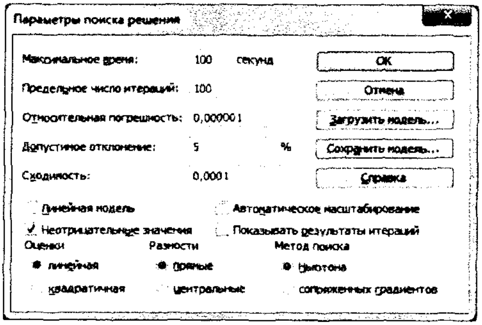
D3 — вводим формулу =СУММПРО-ИЗВ(B3:C3;B2:C2).

D4 — вводим формулу =B2^2+C2^2.

Заполненные диалоговые окна *Поиск решения* и *Параметры поиска решения* изображены на рис. 2.3.



а)



б)

Рис. 2.3. Заполненные диалоговые окна *Поиск решения* (*а*) и *Параметры поиска решения* (*б*) для примера 2.1

Вид рабочего листа *Excel* после выполнения команды *Поиск решения* представлен на рис. 2.4.

	A	B	C	D	E
1		x_1	x_2		
2		3,57758	1,78912		
3	F	2	1	**8,94427**	
4	b_1			**16**	16

Рис. 2.4. Пример 2.1 в *Excel* после выполнения команды *Поиск решения*

Оптимальное решение:

$$\max F(3{,}578;\ 1{,}789) = 8{,}944.\ \blacktriangle$$

Пример 2.2. Найти глобальные экстремумы

$$F(x) = 2x_1 + 3x_2 \to \max\ (\min)$$

при ограничениях

$$\begin{cases} x_1x_2 \geq 3, \\ x_1^2 + x_2^2 \leq 16; \end{cases}$$
$$x_1, x_2 \geq 0.$$

▼Построим область допустимых решений (рис. 2.5). ОДР представляет собой замкнутую область, ограниченную окружностью $x_1^2 + x_2^2 = 16$ и гиперболой $x_2 = 3/x_1$.

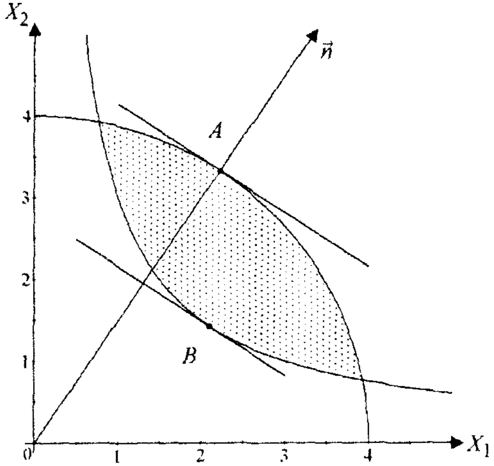


Рис. 2.5. Область допустимых решений примера 2.2

Функция $F(x)$ возрастает в направлении вектора $\vec{n} = gradF(x) = (2;3)$, а ее линии уровня перпендикулярны вектору $\vec{n}$.

Таким образом, максимум достигается в точке A, а минимум — в точке B.

Определим координаты точек A, B из условия коллинеарности соответствующих градиентов:

$$gradF(x) = (2;3), \ grad(x_1^2 + x_2^2) = (2x_1; 2x_2), \ grad(x_1x_2) = (x_2; x_1).$$

Определим координаты точки A:

$$\frac{2x_1}{2} = \frac{2x_2}{3} \Rightarrow x_1 = \frac{2}{3}x_2; \begin{cases} x_1 = \dfrac{2}{3}x_2, \\ x_1^2 + x_2^2 = 16. \end{cases} \Rightarrow x_1 = \frac{8}{\sqrt{13}}, \ x_2 = \frac{12}{\sqrt{13}}.$$

Тогда $F(x)=2\cdot\frac{8}{\sqrt{13}}+3\cdot\frac{12}{\sqrt{13}}=\frac{52}{\sqrt{13}}$, т. е. $F_{\max}\left(\frac{8}{\sqrt{13}};\frac{12}{\sqrt{13}}\right)=\frac{52}{\sqrt{13}}$.

Определим координаты точки B:

$$\frac{x_2}{2}=\frac{x_1}{3} \Rightarrow x_2=\frac{2}{3}x_1; \quad \begin{cases} x_2=\frac{2}{3}x_1, \\ x_1x_2=3. \end{cases} \Rightarrow x_1=\frac{3}{\sqrt{2}},\ x_2=\sqrt{2}.$$

Тогда $F(x)=2\cdot\frac{3}{\sqrt{2}}+3\cdot\sqrt{2}=6\sqrt{2}$, т. е. $F_{\min}\left(\frac{3}{\sqrt{2}};\sqrt{2}\right)=6\sqrt{2}$. ▲

Задача с нелинейной целевой функцией и линейной системой ограничений

Пример 2.3. Найти глобальный максимум функции

$$F(x)=(x_1-1)^2+(x_2-5)^2\to\max$$

при ограничениях

$$\begin{cases} x_1-2x_2\ge -4, \\ 5x_1+2x_2\le 20; \end{cases}$$

$$x_1, x_2\ge 0.$$

▼ОДР задачи является многоугольник $OABC$ (рис. 2.6).

Линии уровня целевой функции есть окружности с центром в точке $M(1;\ 5)$. С увеличением радиуса окружности значение целевой функции F увеличивается. Максимальное значение целевой функции достигается в точке C, минимальное — в точке D.

Координаты точки $C(4;\ 0) \Rightarrow F(C)=34$, т. е. $F_{\max}(4;\ 0)=34$.

Для градиентов $gradF(x)=(2(x_1-1);2(x_2-5))$, $grad(x_1-2x_2)=$ $=(1;-2)$ запишем условие их коллинеарности

$$\frac{2(x_1-1)}{1}=\frac{2(x_2-5)}{-2} \Rightarrow 2x_1+x_2=7,$$

тогда координаты точки D определяются из решения системы

$$\begin{cases} 2x_1+x_2=7 \\ x_1-2x_2=-4 \end{cases} \Rightarrow x_1=2,\ x_2=3,\ F_{\min}(2;3)=5.$$

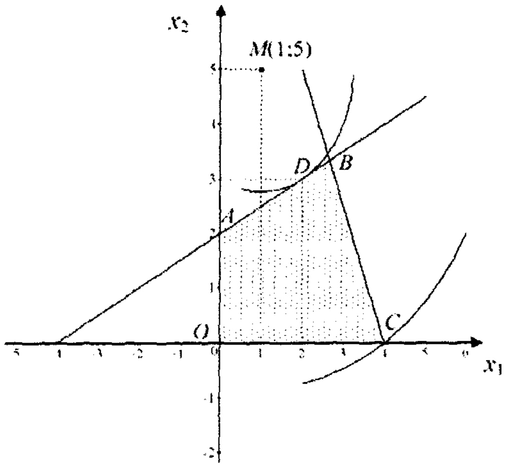
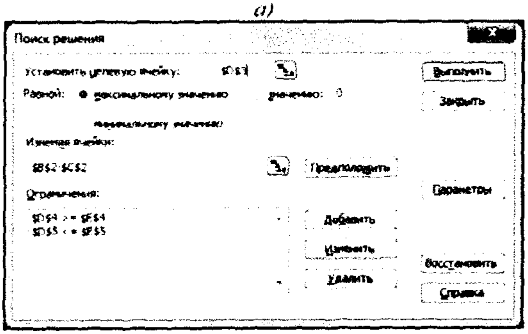


Рис. 2.6. Область допустимых решений примера 2.3

Решим пример 2.3 в *Excel*. Рабочий лист *Excel* и заполненное диалоговое окно *Поиск решения* имеют вид, представленный на рис. 2.7.

	A	B	C	D	E
1		x_1	x_2		
2		1	1		
3	*F*			16	
4	b_1	1	-2	-1	-4
5	b_2	5	2	7	20

а)



б)

Рис. 2.7. Исходные данные в *Excel* (*а*) и заполненное диалоговое окно *Поиск решения* (*б*) для примера 2.3

Вид рабочего листа *Excel* после выполнения команды *Поиск решения* представлен на рис. 2.8.

	A	B	C	D	E
1		x_1	x_2		
2		4	0		
3	F			34	
4	b_1	1	-2	4	-4
5	b_2	5	2	20	20

Рис. 2.8. Пример 2.3 в *Excel* после выполнения команды *Поиск решения*

Оптимальное решение:

$$\max F(4;\,0) = 34. \ \blacktriangle$$

Задача с нелинейной целевой функцией и нелинейной системой ограничений

Пример 2.4. Найти глобальные экстремумы

$$F(x) = (x_1 - 2)^2 + (x_2 - 1)^2 \to \max(\min)$$

при ограничениях

$$x_1^2 + x_2^2 \le 16\,;$$
$$x_1,\, x_2 \ge 0.$$

▼ОДР — часть окружности радиуса 4 с центром в точке $O(0;\,0)$ (рис. 2.9).

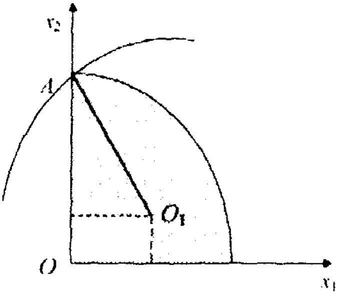


Рис. 2.9. Область допустимых решений примера 2.4

Линии уровня целевой функции $F(x)$ являются окружности с центром в точке $O_1(2;\ 1)$. Глобальный минимум достигается в точке $O_1(2;\ 1)$, при этом $F_{\min}(O_1) = 0$. Глобальный максимум — в точке $A(0;\ 4)$, при этом $F_{\max}(A) = 13$. ▲

2.3. Условный экстремум. Метод множителей Лагранжа

Пусть требуется решить задачу нелинейного программирования следующего вида (задача нахождения условного экстремума):

$$F(x_1, x_2, ..., x_n) \to \max(\min)$$

при ограничениях

$$g_i(x_1, x_2, ..., x_n) = 0\ ,\ i = \overline{1, m}\ ,$$

где функции F, g_i непрерывны вместе со своими первыми частными производными.

Отыскание условного экстремума можно свести к задаче безусловного экстремума с помощью метода Лагранжа.

Метод Лагранжа состоит в построении функции

$$L(x_1, x_2, ..., x_n, \lambda_1, \lambda_2, ..., \lambda_m) = F(x_1, x_2, ..., x_n) + \sum_{i=1}^{m} \lambda_i g_i(x_1, x_2, ..., x_n)\ ,$$

которая называется *функцией Лагранжа,* а λ_i — *множители Лагранжа.*

Запишем необходимые условия экстремума функции Лагранжа

$$\begin{cases} \dfrac{\partial L}{\partial x_j} = \dfrac{\partial F}{\partial x_j} + \sum\limits_{i=1}^{m} \lambda_i \dfrac{\partial g_i}{\partial x_j} = 0, (j = \overline{1, n}), \\ \dfrac{\partial L}{\partial \lambda_i} = g_i(x_1, x_2, ..., x_n) = 0, (i = \overline{1, m}). \end{cases}$$

Эта система уравнений определяет оптимальное решение поставленной задачи. Если система уравнений имеет несколько решений, то выбирают наибольшее и наименьшее значения функции $F(x_1, x_2, ..., x_n)$ в полученных точках.

Пример 2.5. Найти точку условного экстремума функции

$$F(x_1, x_2, x_3) = x_1 x_2 + x_2 x_3$$

при ограничениях

$$\begin{cases} x_1 - x_2 = 2, \\ x_2 + 2x_3 = 4. \end{cases}$$

▼Составляем функцию Лагранжа

$$L(x_1, x_2, x_3, \lambda_1, \lambda_2) = x_1 x_2 + x_2 x_3 + \lambda_1 (x_1 - x_2 - 2) + \lambda_2 (x_2 + 2x_3 - 4).$$

Найдем частные производные функции Лагранжа по переменным x_1, x_2, x_3, λ_1, λ_2. Приравняв к нулю полученные выражения, решим систему

$$\begin{cases} \dfrac{\partial L}{\partial x_1} = x_2 + \lambda_1 = 0, \\ \dfrac{\partial L}{\partial x_2} = x_1 + x_3 - \lambda_1 + \lambda_2 = 0, \\ \dfrac{\partial L}{\partial x_3} = x_2 + 2\lambda_2 = 0, \\ \dfrac{\partial L}{\partial \lambda_1} = x_1 - x_2 - 2 = 0, \\ \dfrac{\partial L}{\partial \lambda_2} = x_2 + 2x_3 - 4 = 0; \end{cases} \Rightarrow \lambda_1 = 4, \lambda_2 = 2, x_1 = -2, x_2 = -4, x_3 = 4.$$

Получили точку $(-2; -4; 4)$ в которой $F(-2; -4; 4) = -8$. Определим характер экстремума, изменяя полученные значения переменных. Измененные значения должны удовлетворять заданной системе ограничений. Возьмем $x_1 = -1 > -2$, тогда из системы ограничений получим $x_2 = -3$, $x_3 = 7/2$, при этом $F(-1; -3; 7/2) = -15/2 > -8$. Возьмем $x_1 = -3 < -2$, тогда из системы ограничений получим $x_2 = -5$, $x_3 = 9/2$, при этом $F(-3; -5; 9/2) = -15/2 > -8$. Следовательно, $F(-2; -4; 4) = -8$ — минимальное значение функции. ▲

Упражнение 2.1. Найти условный экстремум функции

$$F(x_1, x_2) = x_1 x_2$$

при ограничении

$$x_1^2 + x_2^2 = 8 .$$

Ответ: $F_{\max} = F(2; 2) = F(-2; -2) = 4$, $F_{\min} = F(2; -2) = F(-2; 2) = -4$.

Метод множителей Лагранжа с ограничениями в виде неравенств

Метод множителей Лагранжа можно применять и в случае задачи с ограничениями в виде неравенств.

Пусть дана задача, в которой

$$F(x_1, x_2, \dots, x_n) \to \max(\min)$$

при ограничениях

$$g_i(x_1, x_2, \dots, x_n) \le 0, \; i = \overline{1, m} .$$

Если в рассматриваемой задаче имеются условия неотрицательности переменных, то предполагается, что они включены в указанные ограничения.

Вычислительная процедура метода множителей Лагранжа включает следующие шаги.

Шаг 1. Находятся точки безусловного экстремума функции $F(x_1, x_2, \dots, x_n)$ из уравнений $\dfrac{\partial F}{\partial x_j} = 0$, $j = \overline{1, n}$. Из полученных точек отбираются те, координаты которых удовлетворяют ограничениям.

Шаг 2. Последовательно активизируются ограничения задачи (т. е. преобразовываются в равенства) и при наличии ограничения оптимизируется функция $F(x_1, x_2, \dots, x_n)$ методом множителей Лагранжа. Из полученных точек отбираются те, координаты которых удовлетворяют ограничениям.

Шаг 3. Из точек, полученных на первом и втором шаге, выбираются те, которые доставляют наибольшее и наименьшее значение функции $F(x_1, x_2,\dots, x_n)$.

Пример 2.6. Найти наибольшее и наименьшее значения функции

$$F(x_1,x_2)=x_1^2+x_2^2$$

при ограничениях

$$\begin{cases} 2x_1+3x_2\le 13, & (a) \\ x_1+x_2\ge 2, & (b) \\ x_1\ge 0, & (c) \\ x_2\ge 0. & (d) \end{cases}$$

▼Точка безусловного экстремума находится как решение системы уравнений

$$\begin{cases} \dfrac{\partial F}{\partial x_1}=2x_1=0, \\ \dfrac{\partial F}{\partial x_2}=2x_2=0; \end{cases} \Rightarrow\ (x_1,x_2)=(0;0).$$

Так как точка (0; 0) не удовлетворяет условию (b), ограничения поочередно активизируются.

Рассмотрим ограничение $2x_1 + 3x_2 - 13 = 0$. Функция Лагранжа в этом случае примет вид

$$L(x_1,x_2,\lambda)=x_1^2+x_2^2+\lambda(2x_1+3x_2-13).$$

Следовательно,

$$\begin{cases} \dfrac{\partial L}{\partial x_1}=2x_1+2\lambda=0, \\ \dfrac{\partial L}{\partial x_2}=2x_2+3\lambda=0, \\ \dfrac{\partial L}{\partial \lambda}=2x_1+3x_2-13=0; \end{cases} \Rightarrow\ \lambda=-2,\ x_1=2,\ x_2=3.$$

Точка (2; 3) удовлетворяет всем ограничениям задачи.

Рассмотрим ограничение $x_1 + x_2 - 2 = 0$. Функция Лагранжа в этом случае примет вид

$$L(x_1, x_2, \lambda) = x_1^2 + x_2^2 + \lambda(x_1 + x_2 - 2).$$

Следовательно,

$$\begin{cases} \dfrac{\partial L}{\partial x_1} = 2x_1 + \lambda = 0, \\ \dfrac{\partial L}{\partial x_2} = 2x_2 + \lambda = 0, \\ \dfrac{\partial L}{\partial \lambda} = x_1 + x_2 - 2 = 0; \end{cases} \Rightarrow \lambda = -2,\ x_1 = 1,\ x_2 = 1.$$

Точка (1; 1) удовлетворяет всем ограничениям задачи.

Рассмотрим ограничение $x_1 = 0$. Функция Лагранжа в этом случае примет вид

$$L(x_1, x_2, \lambda) = x_1^2 + x_2^2 + \lambda x_1.$$

Следовательно,

$$\begin{cases} \dfrac{\partial L}{\partial x_1} = 2x_1 + \lambda = 0, \\ \dfrac{\partial L}{\partial x_2} = 2x_2 = 0, \\ \dfrac{\partial L}{\partial \lambda} = x_1 = 0; \end{cases} \Rightarrow \lambda = 0,\ x_1 = 0,\ x_2 = 0.$$

Точка (0; 0) не удовлетворяет ограничению (b).

Рассмотрим ограничение $x_2 = 0$. Функция Лагранжа в этом случае примет вид

$$L(x_1, x_2, \lambda) = x_1^2 + x_2^2 + \lambda x_2.$$

Следовательно,

$$\begin{cases} \dfrac{\partial L}{\partial x_1} = 2x_1 = 0, \\ \dfrac{\partial L}{\partial x_2} = 2x_2 + \lambda = 0, \\ \dfrac{\partial L}{\partial \lambda} = x_2 = 0; \end{cases} \Rightarrow \lambda = 0, x_1 = 0, x_2 = 0.$$

Точка (0; 0) не удовлетворяет ограничению (b).

Сравнивая значения функции $F(x_1, x_2)$ в точках (2; 3), (1; 1) получаем $F_{\min}(1;1) = 1^2 + 1^2 = 2$ и $F_{\max}(2;3) = 2^2 + 3^2 = 13$. ▲

Упражнение 2.2. Найти наибольшее и наименьшее значения функции

$$F(x_1, x_2) = x_1^2 + x_2^2 - 2x_2 + 3$$

при ограничениях

$$\begin{cases} x_1^2 + x_2 \le 4, \\ x_2 \ge 0. \end{cases}$$

Ответ: $F_{\min}(0;1) = 2$, $F_{\max}(0;4) = 11$.

Глава 3.
ЭЛЕМЕНТЫ ТЕОРИИ ИГР

3.1. Основные понятия теории игр

Конфликтными называются ситуации, в которых участвуют стороны, преследующие различные цели.

Теория игр есть математическая теория конфликтных ситуаций. Стороны, участвующие в конфликте, называются *игроками*. Каждый игрок имеет некоторое множество выборов, называемых *стратегиями*.

Игра называется *конечной*, если у каждого игрока имеется конечное число стратегий, и *бесконечной* в противном случае.

Исход (результат) игры определяется *функцией выигрыша*, задаваемой аналитически или таблично (платежной матрицей).

Игра с двумя игроками, в которой выигрыш одного из них равен проигрышу другого, называется *игрой двух лиц с нулевой суммой* или *антагонистической*. В такой игре достаточно определить результат в виде выигрыша одного из игроков.

Предполагается, что функция выигрыша и множество стратегий, доступных каждому из игроков, общеизвестны.

Оптимальной стратегией игрока называется такая стратегия, которая при многократном повторении игры обеспечивает одному игроку максимально возможный средний выигрыш, а другому — минимальный средний проигрыш.

Задачей теории игр является определение оптимальной стратегии для каждого игрока. При этом предполагается, что лицу, принимающему решение, противостоит разумный противник.

Платежная матрица

Рассмотрим конечную парную игру с нулевой суммой, где у игрока A имеется m стратегий A_1, A_2, ..., A_m, а у игрока B — n стратегий $B_1, B_2,\ldots,B_n$.

В результате выбора игроками любой пары стратегий $(A_i; B_j)$ однозначно определяется исход игры, т. е. выигрыш a_{ij} игрока A (положительный или отрицательный) и проигрыш $(-a_{ij})$ игрока В.

Набор выигрышей a_{ij} для разных значений i, j располагают в виде матрицы, строки которой отвечают стратегиям игрока A, а столбцы — стратегиям игрока B. Такая матрица называется *платежной матрицей* или матрицей игры размера $m \times n$. Общий вид матрицы представлен в табл. 3.1.

Таблица 3.1

Общий вид платежной матрицы

Стратегия игрока A	Стратегия игрока B			
	B_1	B_2	...	B_n
A_1	a_{11}	a_{12}	...	a_{1n}
A_2	a_{21}	a_{22}	...	a_{2n}
...	...	...	...	...
A_m	a_{m1}	a_{m2}	...	a_{mn}

Задать игру — это задать m стратегий игрока A, n стратегий игрока B и платежную матрицу.

Пример 3.1. Пусть каждый из двух игроков A, B может записать независимо от другого цифры 1, 2, 3. Если разность между цифрами положительна, то первый игрок выигрывает количество очков, равное разности между цифрами, и наоборот, если разность отрицательна, то выигрывает второй игрок. Если разность равна нулю, то игра заканчивается вничью.

В этой игре каждый из игроков имеет три стратегии, которые составляют матрицу игры 3×3, представляющую выигрыш игрока A. Составим платежную матрицу (табл. 3.2).

В такой матричной игре целью игрока A является максимизация своего выигрыша, целью игрока B — минимизация своего проигрыша.

Таблица 3.2

Платежная матрица для примера 3.1

Стратегия игрока A	Стратегия игрока B		
	$B_1 = 1$	$B_2 = 2$	$B_3 = 3$
$A_1 = 1$	0	–1	–2
$A_2 = 2$	1	0	–1
$A_3 = 3$	2	1	0

3.2. Решение матричной игры в чистых стратегиях

Для решения игры двух лиц с нулевой суммой предлагается критерий *минимакса-максимина*. Этот критерий основывается на выборе наилучшей из наихудших возможностей и является наиболее осторожным.

Рассмотрим игру с матрицей, представленной в табл. 3.3.

Таблица 3.3

Платежная матрица

Стратегия игрока A	Стратегия игрока B				α_i
	B_1	B_2	...	B_n	
A_1	a_{11}	a_{12}	...	a_{1n}	α_1
A_2	a_{21}	a_{22}	...	a_{2n}	α_2
...	...	...	...	...	...
A_m	a_{m1}	a_{m2}	...	a_{mn}	α_m
β_j	β_1	β_2	...	β_n	

Каждый игрок стремится обеспечить себе максимальный выигрыш при любых действиях противника. Найдем оптимальные стратегии для каждого из игроков.

Игрок А считает, что какую бы стратегию он ни выбрал, игрок B выберет стратегию, минимизирующую его выигрыш, т. е.

$$\min_j a_{ij} = \alpha_i \text{ (по строкам).}$$

За оптимальную игрок A возьмет стратегию, для которой α_i максимален, т. е.

$$\alpha = \max_i \alpha_i = \max_i \min_j a_{ij}.$$

Выбранная игроком A стратегия называется *максиминной стратегией*, а соответствующее ей значение выигрыша α называют *нижней ценой игры*. Это гарантированный выигрыш игрока A при любой стратегии игрока B.

В итоге, если игрок A придерживается своей максиминной стратегии, его выигрыш будет не меньше нижней цены игры, т. е. $a_{ij} \geq \alpha$.

Игрок B считает, что какую бы стратегию он ни выбрал, игрок A выберет стратегию, максимизирующую его выигрыш, т. е.

$$\max_i a_{ij} = \beta_j \text{ (по столбцам).}$$

За оптимальную игрок B возьмет стратегию, для которой β_j минимально, т. е.

$$\beta = \min_j \beta_j = \min_j \max_i a_{ij} .$$

Выбранная игроком B стратегия называется *минимаксной стратегией*, а соответствующее ей значение проигрыша β называют *верхней ценой игры*. Это гарантированный проигрыш игрока B при любых стратегий игрока A.

В итоге, если игрок B придерживается своей минимаксной стратегии, его проигрыш в любом случае будет не больше верхней цены игры, т. е. $a_{ij} \leq \beta$.

Из условий, определяющих критерий минимакса-максимина, следует:

$$\alpha \leq a_{ij} \leq \beta .$$

Игра, для которой $\alpha = \beta$, называется *игрой с седловой точкой.*

Решением игры называется пара $(A_i^*; B_j^*)$ оптимальных стратегий, соответствующих седловой точке. Элемент $a_{ij} = \nu$, соответствующий решению игры, называется *ценой игры*, причем $\nu = \alpha = \beta$.

Решение игры обладает свойством устойчивости: если один из игроков придерживается своей оптимальной стратегии, то для другого не выгодно отклоняться от своей оптимальной стратегии.

Если игра имеет седловую точку ($\alpha = \beta$), то говорят, что она решается в *чистых стратегиях*.

Пример 3.2. Найдем решение игры, заданной платежной матрицей, представленной в табл. 3.4.

Таблица 3.4

Платежная матрица для примера 3.2

Стратегия игрока A	Стратегия игрока B			α_i
	B_1	B_2	B_3	
A_1	0	–1	–2	–2
A_2	1	0	–1	–1
A_3	2	1	0	0
β_j	2	1	0	

▼Нижняя и верхняя цена игры $\alpha = \max\limits_i \alpha_i = 0$ и $\beta = \min\limits_j \beta_j = 0$.

Поскольку $\alpha = \beta$, то игра имеет седловую точку.

Решение игры — (A_3, B_3), цена игры $v = 0$. ▲

Пример 3.3. Найти решение игры, заданной платежной матрицей, представленной в табл. 3.5.

Таблица 3.5

Платежная матрица для примера 3.3

Стратегия игрока A	Стратегия игрока B			α_i
	B_1	B_2	B_3	
A_1	1	2	3	1
A_2	4	5	6	4
β_j	4	5	6	

▼Нижняя и верхняя цена игры $\alpha = \max\limits_i \alpha_i = 4$ и $\beta = \min\limits_j \beta_j = 4$.

Поскольку $\alpha = \beta$, то игра имеет седловую точку.

Решение игры — (A_2, B_1), цена игры $v = 4$. ▲

Пример 3.4. Найдем решение игры, заданной платежной матрицей, представленной в табл. 3.6.

▼Нижняя и верхняя цена игры $\alpha = \max\limits_i \alpha_i = 0,7$ и $\beta = \min\limits_j \beta_j = 0,7$.

Таблица 3.6

Платежная матрица для примера 3.4

Стратегия игрока A	**Стратегия игрока B**			α_i
	B_1	B_2	B_3	
A_1	0,5	0,6	0,8	0,5
A_2	0,9	0,7	0,8	0,7
A_3	0,7	0,6	0,6	0,6
β_j	0,9	0,7	0,8	

Поскольку $\alpha = \beta$, то игра имеет седловую точку.

Решение игры — (A_2, B_2), цена игры $\nu = 0{,}7$. ▲

Упражнение 3.1. Найти решение игры, заданной платежной матрицей, представленной в табл. 3.7.

Таблица 3.7

Платежная матрица для упражнения 3.1

Стратегия игрока A	**Стратегия игрока B**			α_i
	B_1	B_2	B_3	
A_1	2	0	–1	–1
A_2	3	4	2	2
A_3	–2	1	0	–2
A_4	5	1	5	1
β_j	5	4	5	

Ответ: игра не имеет седловой точки.

3.3. Решение матричной игры в смешанных стратегиях

Пример 3.5. Пусть игра задана платежной матрицей, представленной в табл. 3.8.

Найдем решение игры.

▼Нижняя и верхняя цена игры $\alpha = \max\limits_i \alpha_i = 4$ и $\beta = \min\limits_j \beta_j = 7$.

Поскольку $\alpha \neq \beta$, то игра не имеет седловой точки. ▲

Таблица 3.8

Платежная матрица для примера 3.5

Стратегия игрока A	Стратегия игрока B				α_i
	B_1	B_2	B_3	B_4	
A_1	2	7	6	10	2
A_2	8	4	9	5	4
β_j	8	7	9	10	

Если игра не имеет седловой точки ($\alpha < \beta$), то применение чистых стратегий не дает оптимального решения игры. Поиск решения таких игр приводит к применению смешанных стратегий.

Смешанными называются стратегии, состоящие в случайном чередовании чистых стратегий.

Пусть $x = (x_1, x_2, \ldots, x_m)$, $y = (y_1, y_2, \ldots, y_n)$ наборы вероятностей, с которыми игроки A, B выбирают свои чистые стратегии, причем

$$\sum_i x_i = \sum_j y_j = 1,$$

где $x_i, y_j \geq 0$ для всех i, j.

Платежную матрицу при смешанных стратегиях представим в виде табл. 3.9.

Таблица 3.9

Платежная матрица при смешанных стратегиях

Стратегия	y_1	y_2	...	y_n
x_1	a_{11}	a_{12}	...	a_{1n}
x_2	a_{21}	a_{22}	...	a_{2n}
...	...	...	...	...
x_m	a_{m1}	a_{m2}	...	a_{mn}

Основная теорема теории игр утверждает, что каждая матричная игра с нулевой суммой имеет по крайней мере одно решение, возможно, в области смешанных стратегий.

Применение оптимальной стратегии позволяет получить выигрыш, равный цене игры ν, причем $\alpha \le \nu \le \beta$, где α, β — нижняя и верхняя цена игры.

Оптимальная смешанная стратегия игрока *A* обеспечивает ему средний выигрыш (математическое ожидание), равный цене игры ν, если игрок *B* применяет свою оптимальную стратегию, или больше цены игры ν, если игрок *B* применяет неоптимальную стратегию.

Это приводит к системе неравенств

$$\begin{cases} a_{11}x_1 + a_{21}x_2 + \ldots + a_{m1}x_m \ge \nu, & (B_1) \\ a_{12}x_1 + a_{22}x_2 + \ldots + a_{m2}x_m \ge \nu, & (B_2) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ a_{1n}x_1 + a_{2n}x_2 + \ldots + a_{mn}x_m \ge \nu; & (B_n) \end{cases} \tag{3.1}$$
$$x_1 + x_2 + \ldots + x_m = 1.$$

Левая часть каждого из неравенств представляет собой математическое ожидание выигрыша игрока *A* при использовании игроком *B* против него той или иной стратегии.

Оптимальная смешанная стратегия игрока *B* обеспечивает ему средний проигрыш (математическое ожидание), равный цене игры ν, если игрок *A* применяет свою оптимальную стратегию, или меньше цены игры ν, если игрок *A* применяет неоптимальную стратегию.

Это приводит к системе неравенств

$$\begin{cases} a_{11}y_1 + a_{12}y_2 + \ldots + a_{1n}y_n \le \nu, & (A_1) \\ a_{21}y_1 + a_{22}y_2 + \ldots + a_{2n}x_n \le \nu, & (A_2) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ a_{m1}y_1 + a_{m2}y_2 + \ldots + a_{mn}y_n \le \nu; & (A_m) \end{cases} \tag{3.2}$$
$$y_1 + y_2 + \ldots + y_n = 1.$$

Левая часть каждого из неравенств представляет собой математическое ожидание проигрыша игрока *B* при использовании игроком *A* против него той или иной стратегии.

Известны несколько методов нахождения оптимальных стратегий в играх двух лиц с нулевой суммой. Задача нахождения оптимального решения усложняется с ростом числа стратегий.

Решение игр ($m \times n$) с помощью ЛП

Каждая конечная игра двух лиц с нулевой суммой может быть представлена как задача ЛП, и наоборот, каждая задача ЛП может быть представлена как игра. Способ нахождения решения игры методом ЛП особенно эффективен для игр большой размерности.

Оптимальная смешанная стратегия игрока A определяется условием

$$v \to \max$$

при ограничениях (3.1).

Полученная задача ЛП может быть упрощена делением всех $(n + 1)$ ограничений на v. Эта операция возможна при $v > 0$. В противном случае, если $v < 0$, необходимо поменять знаки неравенств.

Таким образом, полагая $v > 0$, можем записать ограничения задачи в виде

$$\begin{cases} a_{11}(x_1 / v) + a_{21}(x_2 / v) + \ldots + a_{m1}(x_m / v) \geq 1, \\ a_{12}(x_1 / v) + a_{22}(x_2 / v) + \ldots + a_{m2}(x_m / v) \geq 1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1n}(x_1 / v) + a_{2n}(x_2 / v) + \ldots + a_{mn}(x_m / v) \geq 1; \end{cases}$$
$$(x_1 / v) + (x_2 / v) + \ldots + (x_m / v) = 1 / v.$$

Полагая $t_i = (x_i/v)$, $i = \overline{1, m}$, и в силу того, что

$$\max v = \min 1 / v = \min(t_1 + t_2 + \ldots + t_m),$$

запишем задачу в виде

$$F(t) = t_1 + t_2 + \ldots + t_m \to \min$$

при ограничениях

$$\begin{cases} a_{11}t_1 + a_{21}t_2 + \ldots + a_{m1}t_m \geq 1, \\ a_{12}t_1 + a_{22}t_2 + \ldots + a_{m2}t_m \geq 1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1n}t_1 + a_{2n}t_2 + \ldots + a_{mn}t_m \geq 1; \end{cases}$$

$$t_1, t_2, \ldots, t_m \geq 0.$$

Оптимальное решение игрока A будет

$$x^* = (x_1^*, x_2^*, \ldots, x_m^*)\text{, где } x_i^* = t_i \nu^*,\ \nu^* = 1/F_{\min}.$$

Оптимальная смешанная стратегия игрока B определяется условием

$$\nu \to \min$$

при ограничениях (3.2).

Эта задача также может быть записана как задача ЛП

$$G(s) = s_1 + s_2 + \ldots + s_n \to \max$$

при ограничениях

$$\begin{cases} a_{11}s_1 + a_{12}s_2 + \ldots + a_{1n}s_n \leq 1, \\ a_{21}s_1 + a_{22}s_2 + \ldots + a_{2n}s_n \leq 1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}s_1 + a_{m2}s_2 + \ldots + a_{mn}s_n \leq 1; \end{cases}$$

$$s_1, s_2, \ldots s_n \geq 0.$$

где $s_j = (y_j/\nu)$, $j = \overline{1,n}$, поскольку $\min \nu = \max 1/\nu = \max(s_1 + + s_2 + \ldots + s_n)$.

Задача игрока B является двойственной к задаче игрока A, тем самым оптимальное решение одной из задач дает оптимальное решение другой задачи.

Пример 3.6. Торговая фирма разработала несколько вариантов плана продаж товаров на предстоящей ярмарке с учетом меняющейся конъюнктуры рынка и спроса покупателей. Получающиеся от их возможных сочетаний показатели дохода представлены в табл. 3.10.

Таблица 3.10

Платежная матрица для примера 3.6

План продаж	Доход, усл. ед.		
	R_1	R_2	R_3
Q_1	8	4	2
Q_2	2	8	4
Q_3	1	2	8

Определим оптимальный план продаж.

▼Обозначим через x_1, x_2, x_3 — вероятности применения торговых стратегий Q_1, Q_2, Q_3, а вероятности реализации вариантов конъюнктуры рынка и спроса покупателей R_1, R_2, R_3 — через y_1, y_2, y_3.

Для первого игрока (торговая фирма) математическая модель задачи имеет вид

$$F(t) = t_1 + t_2 + t_3 \to \min$$

при ограничениях

$$\begin{cases} 8t_1 + 2t_2 + t_3 \geq 1, \\ 4t_1 + 8t_2 + 2t_3 \geq 1, \\ 2t_1 + 4t_2 + 8t_3 \geq 1; \end{cases}$$
$$t_1, t_2, t_3 \geq 0.$$

Решаем задачу в *Excel* (*Поиск решения*) (рис. 3.1).

	A	B	C	D	E	F
1		t_1	t_2	t_3		
2		0.1020	0.0561	0.0714		
3	F	1	1	1	**0,2296**	
4	B_1	8	2	1	**1**	1
5	B_2	4	8	2	**1**	1
6	B_3	2	4	8	**1**	1

Рис. 3.1. Пример 3.1 в *Excel* после выполнения команды *Поиск решения* (с точки зрения торговой фирмы)

Оптимальное решение:

$$t = (0{,}1020;\ 0{,}0561;\ 0{,}0714),\ F_{\min} = 0{,}2296,$$

тогда цена игры

$$v^* = 1/F_{\min} = 1/0{,}2296 = 4{,}3554,$$

а оптимальное решение

$$x^* = tv^* = (0{,}1020;\ 0{,}0561;\ 0{,}0714) \cdot 4{,}3554 = \\ = (0{,}4443;\ 0{,}2443;\ 0{,}3110).$$

Таким образом, торговая фирма на ярмарке должна придерживаться стратегии $x^* = (0{,}4443;\ 0{,}2443;\ 0{,}3110)$, при этом она получит доход $v^* = 4{,}3554$.

Для второго игрока (конъюнктура рынка и спрос покупателей) математическая модель задачи имеет вид:

$$G(s) = s_1 + s_2 + s_3 \to \max$$

при ограничениях

$$\begin{cases} 8s_1 + 4s_2 + 2s_3 \le 1, \\ 2s_1 + 8s_2 + 4s_3 \le 1, \\ s_1 + 2s_2 + 8s_3 \le 1; \end{cases}$$
$$s_1, s_2, s_3 \ge 0.$$

Решаем задачу в *Excel* (*Поиск решения*) (рис. 3.2).

	A	B	C	D	E	F
1		s_1	s_2	s_3		
2		0.0714	0.0561	0,1020		
3	F	1	1	1	**0,2296**	
4	B_1	8	4	2	**1**	1
5	B_2	2	8	4	**1**	1
6	B_3	1	2	8	**1**	1

Рис. 3.2. Пример 3.1 в *Excel* после выполнения команды *Поиск решения* (с точки зрения покупателей)

Оптимальное решение:

$$s = (0{,}0714;\ 0{,}0561;\ 0{,}1020),\ G_{max} = 0{,}2296,$$

тогда цена игры

$$v^* = 1/G_{max} = 1/0{,}2296 = 4{,}3554,$$

а оптимальное решение

$$y^* = sv^* = (0{,}0714;\ 0{,}0561;\ 0{,}1020) \cdot 4{,}3554 = \\ = (0{,}3110;\ 0{,}2443;\ 0{,}4443),$$

при этом $v^* = 4{,}3554$ — расходы покупателей. ▲

Пример 3.7. Предприятием разработан ряд хозяйственных стратегий по продаже товаров A_1, A_2, A_3 (стратегии предприятия) с учетом трех вариантов поведения покупателей B_1, B_2, B_3 (стратегии покупателей). Платежная матрица представляет собой оценки прибыли (табл. 3.11).

Таблица 3.11

Платежная матрица для примера 3.7

План продаж	**Прибыль, млн руб.**			α_i
	B_1	B_2	B_3	
A_1	8	13	10	8
A_2	4	10	2	2
A_3	9	6	0	0
β_j	9	13	10	

Определим: *a*) существует ли решение в чистых стратегиях; *b*) если нет, то найдем оптимальную смешанную стратегию игрока A (предприятие) и игрока B (покупатели).

▼*a*) Нижняя и верхняя цена игры $\alpha = \max\limits_i \alpha_i = 8$ и $\beta = \min\limits_j \beta_j = 5$.

Поскольку $\alpha \neq \beta$, то игра не имеет седловой точки, следовательно, решения игры в чистых стратегиях не существует.

b) Для игрока A (предприятие) математическая модель задачи имеет вид

$$F(t) = t_1 + t_2 + t_3 \to \min$$

при ограничениях

$$\begin{cases} 8t_1 + 4t_2 + 9t_3 \ge 1, \\ 13t_1 + 10t_2 + 6t_3 \ge 1, \\ 10t_1 + 2t_2 \ge 1; \end{cases}$$

$$t_1, t_2, t_3 \ge 0.$$

Решая задачу в *Excel* (*Поиск решения*), получим оптимальное решение

$$t = (0{,}1000;\ 0;\ 0{,}0222),\ F_{\min} = 0{,}1222,$$

тогда цена игры

$$v^* = 1/F_{\min} = 1/0{,}1222 = 8{,}1833,$$

а оптимальное решение

$$x^* = tv^* = (0{,}1000;\ 0;\ 0{,}0222) \cdot 8{,}1833 = (0{,}8183;\ 0;\ 0{,}1817),$$

при этом $v^* = 8{,}1833$ — доход предприятия.

Для игрока B (покупатели) математическая модель задачи имеет вид

$$G(s) = s_1 + s_2 + s_3 \to \max$$

при ограничениях

$$\begin{cases} 8s_1 + 13s_2 + 10s_3 \le 1, \\ 4s_1 + 10s_2 + 2s_3 \le 1, \\ 9s_1 + 6s_2 \le 1; \end{cases}$$

$$s_1, s_2, s_3 \ge 0.$$

Решая задачу в *Excel* (*Поиск решения*), получим оптимальное решение

$$s = (0{,}1111;\ 0;\ 0{,}0111),\ G_{\max} = 0{,}1222,$$

тогда цена игры

$$v^* = 1/G_{\max} = 1/0{,}1222 = 8{,}1833,$$

а оптимальное решение

$$y^* = s\nu^* = (0{,}1111;\ 0;\ 0{,}0111) \cdot 8{,}1833 = (0{,}9092;\ 0;\ 0{,}0908),$$

при этом $\nu^* = 8{,}1833$ — расходы покупателей. ▲

Аналитическое решение игры (2 × 2)

Пусть игра задана платежной матрицей $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$.

Если такая игра имеет седловую точку, то оптимальное решение — это пара чистых стратегий, соответствующих этой точке. Для игры, в которой отсутствует седловая точка, в соответствии с основной теоремой теории игр, оптимальное решение существует и определяется парой смешанных стратегий $(x_1^*; x_2^*)$ и $(y_1^*; y_2^*)$.

Платежную матрицу при смешанных стратегиях представим в виде табл. 3.12, где $(x_1; x_2), (y_1; y_2)$ — наборы вероятностей, с которыми игроки A, B выбирают свои чистые стратегии.

Таблица 3.12

Платежная матрица при смешанных стратегиях для игры (2 × 2)

Стратегия	y_1	y_2
x_1	a_{11}	a_{12}
x_2	a_{21}	a_{22}

Средний выигрыш первого игрока, если он использует оптимальную смешанную стратегию $(x_1^*; x_2^*)$ при любых чистых стратегиях второго игрока, равен цене игры ν. Это приводит к системе уравнений:

$$\begin{cases} a_{11}x_1 + a_{21}x_2 = \nu, & \text{(при стратегии } B_1 \text{ игрока } B) \\ a_{12}x_1 + a_{22}x_2 = \nu, & \text{(при стратегии } B_2 \text{ игрока } B) \\ x_1 + x_2 = 1. \end{cases}$$

Решая эту систему, получим оптимальное решение

$$x_1^* = \frac{a_{22} - a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}}; \; x_2^* = \frac{a_{11} - a_{12}}{a_{11} + a_{22} - a_{12} - a_{21}}$$

и цену игры

$$\nu = \frac{a_{22}a_{11} - a_{12}a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}}.$$

Средний проигрыш второго игрока, если он использует оптимальную смешанную стратегию $(y_1^*; y_2^*)$ при любых чистых стратегиях первого игрока, равен цене игры ν. Это приводит к системе уравнений:

$$\begin{cases} a_{11}y_1 + a_{12}y_2 = \nu, & \text{(при стратегии } A_1 \text{ игрока } A) \\ a_{21}y_1 + a_{22}y_2 = \nu, & \text{(при стратегии } A_2 \text{ игрока } A) \\ y_1 + y_2 = 1. \end{cases}$$

Оптимальное решение для второго игрока будет

$$y_1^* = \frac{a_{22} - a_{12}}{a_{11} + a_{22} - a_{12} - a_{21}}; \; y_2^* = \frac{a_{11} - a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}}.$$

Пример 3.8. Найдем решение игры, заданной платежной матрицей, представленной в табл. 3.13.

Таблица 3.13

Платежная матрица для примера 3.8

Стратегия игрока A	**Стратегия игрока B**		α_i
	B_1	B_2	
A_1	5	2	2
A_2	3	4	3
β_j	5	4	

▼Нижняя и верхняя цена игры $\alpha = \max\limits_i \alpha_i = 3$ и $\beta = \min\limits_j \beta_j = 4$.

Поскольку $\alpha \neq \beta$, то игра не имеет седловой точки, следовательно, решения игры в чистых стратегиях не существует.

Платежную матрицу при смешанных стратегиях представим в виде табл. 3.14.

Таблица 3.14

Платежная матрица при смешанных стратегиях для игры (2 × 2) для примера 3.8

Стратегия	y_1	y_2
x_1	5	2
x_2	3	4

Для игрока *A* оптимальное решение при смешанных стратегиях определяется из системы уравнений

$$\begin{cases} 5x_1 + 3x_2 = \nu, \\ 2x_1 + 4x_2 = \nu, \\ x_1 + x_2 = 1. \end{cases} \Rightarrow \; x_1^* = 1/4;\; x_2^* = 3/4;\; \nu = 7/2.$$

Оптимальное решение игрока *A* — (1/4; 3/4), $\nu = 7/2$.

Для игрока *B* оптимальное решение в смешанных стратегиях определяется из системы уравнений

$$\begin{cases} 5y_1 + 2y_2 = \nu, \\ 3y_1 + 4y_2 = \nu, \\ y_1 + y_2 = 1. \end{cases} \Rightarrow \; y_1^* = 1/2;\; y_2^* = 1/2;\; \nu = 7/2.$$

Оптимальное решение игрока *B* — (1/2; 1/2), $\nu = 7/2$.

Таким образом, оптимальное решение есть

$$x^* = (1/4;\ 3/4),\ y^* = (1/2;\ 1/2),\ \nu = 7/2. \ ▲$$

Графическое решение игр вида (2 × *n*) и (*m* × 2)

Этот метод применим только к играм, в которых хотя бы один игрок имеет только две стратегии. Рассмотрим игру вида (2 × *n*), в которой игрок *A* имеет две стратегии (табл. 3.15).

Таблица 3.15

Платежная матрица при смешанных стратегиях для игры ($2 \times n$)

Стратегия	y_1	y_2	...	y_n
x_1	a_{11}	a_{12}	...	a_{1n}
$x_2 = 1 - x_1$	a_{21}	a_{22}	...	a_{2n}

Игра предполагает, что игрок A смешивает стратегии A_1 и A_2 с соответствующими вероятностями x_1 и $(1 - x_1)$, $0 < x_1 < 1$. Игрок B смешивает стратегии $B_1, B_2, ..., B_n$ с вероятностями $y_1, y_2, ..., y_n$, где $y_j \geq 0, j = 1, 2, ..., n$, и $y_1 + y_2 + ... + y_n = 1$.

Ожидаемые выигрыши первого игрока, соответствующие чистым стратегиям второго игрока, представлены в табл. 3.16.

Таблица 3.16

Ожидаемые выигрыши первого игрока, соответствующие чистым стратегиям второго игрока, для игры ($2 \times n$)

Чистая стратегия второго игрока	Ожидаемый выигрыш первого игрока
1	$x_1a_{11} + (1 - x_1)a_{21} = (a_{11} - a_{21})x_1 + a_{21}$
2	$x_1a_{12} + (1 - x_1)a_{22} = (a_{12} - a_{22})x_1 + a_{22}$
...	...
n	$x_1a_{1n} + (1 - x_1)a_{2n} = (a_{1n} - a_{2n})x_1 + a_{2n}$

Следовательно, игрок A ищет величину x_1, которая максимизирует минимум ожидаемых выигрышей.

Пример 3.9. Найдем решение игры, заданной платежной матрицей, представленной в табл. 3.17.

Таблица 3.17

Платежная матрица для примера 3.9

Стратегия игрока A	Стратегия игрока B			
	B_1	B_2	B_3	B_4
A_1	2	2	3	–1
A_2	4	3	2	6

▼Игра не имеет решения в чистых стратегиях, следовательно, стратегии должны быть смешанными.

Оптимальное решение игрока А. Ожидаемые выигрыши игрока A, соответствующие чистым стратегиям игрока B, приведены в табл. 3.18.

Таблица 3.18

Ожидаемые выигрыши игрока *A*, соответствующие чистым стратегиям игрока *B*, для примера 3.9

Чистая стратегия игрока B	Ожидаемый выигрыш игрока A
1	$-2x_1 + 4$ (1)
2	$-x_1 + 3$ (2)
3	$x_1 + 2$ (3)
4	$-7x_1 + 6$ (4)

На рис. 3.3 изображены четыре прямые линии, соответствующие чистым стратегиям игрока B. Чтобы определить наилучший результат из наихудших, построена нижняя огибающая четырех указанных прямых (изображенная толстыми линиями), которая представляет минимальный (наихудший) выигрыш для игрока A независимо от того, что делает игрок B.

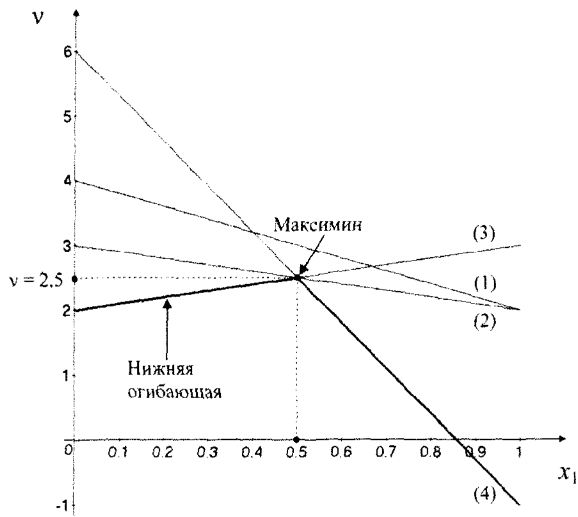


Рис. 3.3. Ожидаемые выигрыши игрока A, соответствующие чистым стратегиям игрока B, для примера 3.9

Максимум нижней огибающей соответствует максиминному решению в точке $x_1^* = 0{,}5$. Эта точка определяется пересечением прямых 3 и 4. Следовательно, оптимальным решением для игрока A является ($x_1^* = 0{,}5; x_2^* = 0{,}5$). Соответствующая цена игры $v = 2{,}5$ определяется подстановкой $x_1 = 0{,}5$ в уравнение либо прямой 3, либо прямой 4.

Оптимальное решение игрока B. Оптимальная смешанная стратегия игрока B определяется двумя стратегиями, которые формируют нижнюю огибающую графика игрока A. Это значит, что игрок B может смешивать стратегии B_3 и B_4, в этом случае $y_1 = y_2 = 0$ и $y_4 = 1 - y_3$. Следовательно, ожидаемые платежи игрока B, соответствующие чистым стратегиям игрока A, имеют такой вид (табл. 3.19).

Таблица 3.19

Ожидаемые платежи игрока *B*, соответствующие чистым стратегиям игрока *A*, для примера 3.9

Чистые стратегии игрока *A*	Ожидаемые платежи игрока *B*
1	$4y_3 - 1$ (3)
2	$-4y_3 + 6$ (4)

Наилучшее решение из наихудших для игрока B представляет собой точку минимума верхней огибающей двух заданных прямых (рис. 3.4).

Эта процедура эквивалентна решению уравнения $4y_3 - 1 = -4y_3 + 6 \Rightarrow y_3 = 7/8$, что определяет цену игры $v = 4 \cdot (7/8) - 1 = 2{,}5$.

Таким образом, решением игры для игрока A является смешивание стратегий A_1 и A_2 с равными вероятностями 0,5 и 0,5, а для игрока B — смешивание стратегий B_3 и B_4 с вероятностями 7/8 и 1/8. (В действительности игра имеет альтернативное решение для игрока B, так как максиминная точка на рис. 3.3 определяется более чем двумя прямыми.) ▲

Для игры, в которой игрок A имеет m стратегий, а игрок B — только две, решение находится аналогично. Главное отличие состоит в том, что здесь строятся графики функций, представляющих ожидаемые платежи второго игрока, соответствующие чистым стратегиям игрока A. В результате ведется поиск минимаксной точки верхней огибающей построенных прямых.

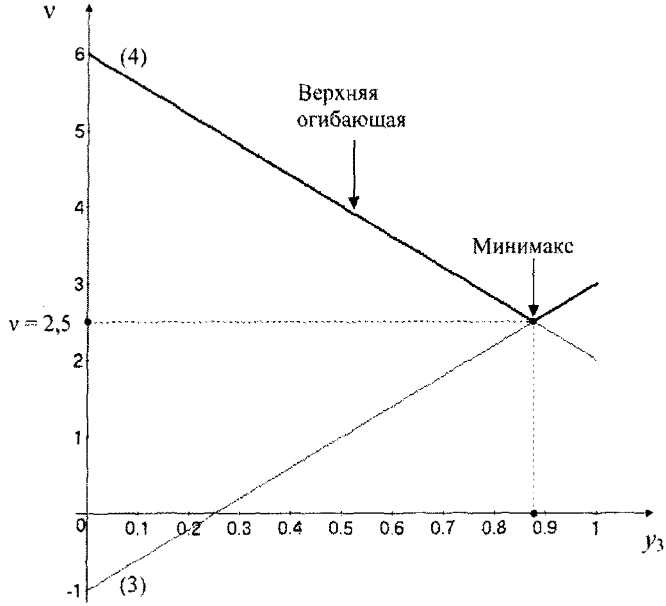


Рис. 3.4. Ожидаемые платежи игрока *B*, соответствующие чистым стратегиям игрока *A*, для примера 3.9

Упражнение 3.2. Найти решение игры, заданной платежной матрицей, представленной в табл. 3.20.

Таблица 3.20

Платежная матрица для упражнения 3.2

Стратегия игрока *A*	Стратегия игрока *B*	
	B_1	B_2
A_1	4	3
A_2	2	4
A_3	0	5
A_4	–1	6

Ответ: оптимальное решение есть

$x^* = (0{,}875;\ 0;\ 0;\ 0{,}125)$, $y^* = (0{,}375;\ 0{,}625)$, $\nu = 3{,}375$.

3.4. Игра с природой

В некоторых ситуациях лицу, принимающему решение, противостоит не разумный противник, а природа, которая действует случайно.

Принятие решений в условиях полной неопределенности

Пусть рассматривается игра с природой с четырьмя стратегиями игрока A и тремя состояниями природы П. Матрица выигрышей задана табл. 3.21.

Таблица 3.21

Матрица выигрышей

Стратегия игрока A	Состояние природы			M_i	α_i	ω_i	γ_i
	$П_1$	$П_2$	$П_3$				
A_1	20	30	15	21,7	15	30	21
A_2	75	20	35	43,3	20	75	42
A_3	25	80	25	43,3	25*	80	47*
A_4	85	5	45	45*	5	85*	37

Если данных о вероятностях состояний среды (природы) не имеется, то лицо, принимающее решения (ЛПР), находится в условиях неопределенности.

Основной метод, позволяющий найти оптимальное решение в условиях неопределенности, состоит в формулировке некоторой гипотезы о поведении среды, позволяющей дать каждому альтернативному решению числовую оценку.

Рассмотрим некоторые критерии, используемые при выборе оптимальной стратегии игрока A в условиях неопределенности.

1. Критерий Байеса — Лапласа. В качестве оптимальной выбирается та стратегия, которая дает максимум математического ожидания выигрыша, т. е.

$$\max_i \left\{ \sum_j a_{ij} p_j \right\},$$

где p_j — вероятность реализации состояния $П_j$.

Поскольку в нашем примере вероятности неизвестны, то предполагается равновероятность состояний природы (*критерий Лапласа*).

В столбце M_i таблицы указаны средние арифметические

$$M_i = \frac{1}{n}\sum_{j=1}^{n} a_{ij} .$$

Из величин M_i максимальное значение равно 45, следовательно, оптимальной является стратегия A_4.

2. Максиминный критерий Вальда (критерий пессимиста). В качестве оптимальной выбирается та стратегия, при которой минимальный выигрыш максимален, т. е.

$$\max_i \min_j a_{ij} .$$

Критерий является *пессимистическим*, поскольку считается, что природа будет действовать наихудшим образом для человека.

В столбце α_i табл. 3.21 указаны $\alpha_i = \min_j a_{ij}$. Из величин α_i максимальная величина есть 25, следовательно, оптимальной является стратегия A_3.

3. Критерий максимума (критерий оптимиста). В качестве оптимальной выбирается та стратегия, при которой максимальный выигрыш максимален, т. е.

$$\max_i \max_j a_{ij} .$$

Критерий является *оптимистическим*, поскольку считается, что природа будет наиболее благоприятна для человека.

В столбце ω_i табл. 3.21 указаны $\omega_i = \max_j a_{ij}$. Из величин ω_i максимальная равная равна 85, следовательно, оптимальной является стратегия A_4.

4. Критерий Гурвица. В качестве оптимальной выбирается та стратегия, при которой максимальна линейная комбинация минимального и максимального выигрышей, т. е.

$$\max_i \left\{ \lambda \min_j a_{ij} + (1-\lambda)\max_j a_{ij} \right\}, \quad 0 \le \lambda \le 1 ,$$

где λ — показатель пессимизма.

Если $\lambda = 1$, критерий Гурвица превращается в пессимистический критерий Вальда, а при $\lambda = 0$ — в критерий крайнего оптимизма. Обычно показатель λ принимается в пределах от 0,5 до 0,7. Пусть $\lambda = 0{,}6$.

В столбце γ_i табл. 3.21 указаны $\gamma_i = 0{,}6\ \alpha_i + 0{,}4\ \omega_i$. Из величин γ_i максимальная равна 47, следовательно, оптимальной является стратегия A_3.

5. Критерий Сэвиджа (критерий сожалеющего пессимиста). В качестве оптимальной выбирается та стратегия, при которой минимален максимальный риск, т. е.

$$\min_i \left\{ \max_j r_{ij} \right\}.$$

Риском называют разность между выигрышем, который можно получить, если знать действительное состояние природы, и выигрышем, который будет получен при отсутствии этой информации, т. е.

$$r_{ij} = \max_j a_{ij} - a_{ij}.$$

Матрица риска для нашего случая представлена в табл. 3.22.

Таблица 3.22

Матрица риска

Стратегия игрока A	**Состояние природы**			**δ_i**
	П$_1$	**П$_2$**	**П$_3$**	
A_1	65	50	30	65
A_2	10	60	10	60*
A_3	60	0	20	60*
A_4	0	75	0	75

В столбце δ_i табл. 3.22 указаны $\delta_i = \max_j r_{ij}$. Из величин δ_i минимальная равна 60, следовательно, оптимальной является любая из стратегий A_2, A_3.

Каждый из рассмотренных критериев не может быть признан вполне удовлетворительным для окончательного выбора решений, однако их совместный анализ позволяет более наглядно представить последствия принятия тех или иных управленческих решений.

Пример 3.10. Фирма производит детские платья и костюмы, реализация которых зависит от состояния погоды. Затраты фирмы в течение августа — сентября на единицу продукции составили: платья — 7 ден. ед., костюма — 28 ден. ед. Цена реализации составляет 15 и 50 ден. ед. соответственно.

По данным наблюдений за несколько предыдущих лет, фирма может реализовать в условиях теплой погоды 1950 платьев и 610 костюмов, а при прохладной погоде 630 платьев и 1050 костюмов.

В связи с возможными изменениями погоды определим стратегию фирмы при выпуске продукции, обеспечивающую ей максимальный доход от ее реализации. Задачу решим с использованием критериев природы, приняв $\lambda = 0{,}5$.

▼Фирма располагает двумя стратегиями:

- A_1 — в этом году будет теплая погода;
- A_2 — погода будет прохладная.

Возможные состояния природы:

- B_1 — будет теплая погода;
- B_2 — будет прохладная погода.

Доход фирмы (ден. ед.) составит:

- при выборе стратегии A_1 и состоянии погоды B_1

$$1950 \cdot (15 - 7) + 610 \cdot (50 - 28) = 29\,020;$$

- при выборе стратегии A_1 и состоянии погоды B_2

$$630 \cdot (15 - 7) + 610 \cdot (50 - 28) - 7 \cdot (1950 - 630) = 9220;$$

- при выборе стратегии A_2 и состоянии погоды B_1

$$630 \cdot (15 - 7) + 610 \cdot (50 - 28) - 28 \cdot (1050 - 610) = 6140;$$

- при выборе стратегии A_2 и состоянии погоды B_2

$$630 \cdot (15 - 7) + 1050 \cdot (50 - 28) = 28\,140.$$

Рассматривая фирму и погоду в качестве двух игроков, запишем платежную матрицу со столбцами M_i, α_i, ω_i, γ_i (табл. 3.23).

Таблица 3.23

Платежная матрица для примера 3.10, ден. ед.

Стратегия фирмы	Состояние природы		M_i	α_i	ω_i	γ_i
	B_1	B_2				
A_1	29 020	9220	19 120*	9220*	29 020*	19 120*
A_2	6140	28 140	17 140	6140	28 140	17 140

Рассмотрим использование различных критериев природы.

Критерий Лапласа: max M_i = 19 120, т. е. фирме целесообразно использовать стратегию A_1.

Критерий Вальда: $\max\limits_i(\alpha_i) = 9220$, т. е. фирме целесообразно использовать стратегию A_1.

Критерий максимума: $\max\limits_i(\omega_i) = 29\ 020$, т. е. фирме целесообразно использовать стратегию A_1.

Критерий Гурвица: $\max\limits_i(\gamma_i) = 19\ 120$, т. е. фирме целесообразно использовать стратегию A_1.

Критерий Сэвиджа. Запишем матрицу риска r_{ij} со столбцом δ_i (табл. 3.24).

Таблица 3.24

Матрица риска для примера 3.10, ден. ед.

Стратегия фирмы	Состояние природы		δ_i
	B_1	B_2	
A_1	0	18 920	18 920*
A_2	22 880	0	22 880

Так как min δ_i = 18 920, то фирме целесообразно использовать стратегию A_1.

Оптимальной по всем критериям является стратегия A_1.

Таким образом, фирме целесообразно производить 1950 платьев и 610 костюмов, тогда при любой погоде она получит доход не менее 9220 ден. ед. ▲

Принятие решений в условиях риска

Математическая модель задачи принятия решений в условиях риска предполагает задание дополнительной информации о поведении природы в виде вероятностей ее различных состояний.

Когда состояниям природы поставлены в соответствие вероятности, заданные экспертно либо вычисленные, решение обычно принимается на основе критерия *максимума ожидаемого среднего выигрыша* или *минимума ожидаемого среднего риска.*

Если для некоторой игры с природой, задаваемой *платежной матрицей* $(a_{ij})_{m\times n}$, стратегиям природы П соответствует вектор вероятности $p = (p_1, p_2, \dots, p_n)$ состояний среды, то лучшей стратегией игрока A будет та, которая обеспечит ему максимальный средний выигрыш, т. е.

$$\max_i M_i = \max_i \sum_j a_{ij} p_j .$$

Применительно к *матрице рисков* лучшей будет та стратегия игрока, которая обеспечивает ему минимальный средний риск, т. е.

$$\min_i R_i = \min_i \sum_j r_{ij} p_j .$$

Эти критерии эквивалентны в том смысле, что оптимальные значения для них обеспечивает одна и та же стратегия игрока A.

Пример 3.11. Пусть платежная матрица и вероятности состояния среды представлены табл. 3.25.

Таблица 3.25

Платежная матрица для примера 3.11

Стратегия игрока A	Вероятность различных состояний среды				M_i
	0,4	0,2	0,2	0,2	
A_1	5	2	8	4	4,8
A_2	2	3	4	12	4,6
A_3	8	5	3	10	6,8
A_4	1	4	2	8	3,2

▼В столбце M_i табл. 3.25 указаны $M_i = \sum_j a_{ij} p_j$.

Поскольку $\max M_i = 6,8$, то A_3 — лучшая стратегия игрока A.

Матрица риска, соответствующая данной платежной матрице, представлена в табл. 3.26.

Таблица 3.26

Матрица риска для примера 3.11

Стратегия игрока A	**Вероятность различных состояний среды**				**R_i**
	0,4	**0,2**	**0,2**	**0,2**	
A_1	3	3	0	8	3,4
A_2	6	2	4	0	3,6
A_3	0	0	5	2	1,4
A_4	7	1	6	4	5

В столбце R_i табл. 3.26 указаны $R_i = \sum_j r_{ij} p_j$.

Поскольку $\min R_i = 1,4$, то A_3 — лучшая стратегия игрока A. ▲

Оптимальность по Парето. В примере 3.11 была получена оптимизационная двухкритериальная задача по выбору наилучшего решения, так как каждое решение имеет две характеристики — средний ожидаемый доход M и средний ожидаемый риск R.

Существует несколько способов постановки таких оптимизационных задач. Рассмотрим один из них в общем виде.

Пусть A — некоторое множество операций (решений). Каждая операция a имеет две числовые характеристики $M(a)$ и $R(a)$ (например, доход и риск), и разные операции обязательно различаются хотя бы одной характеристикой. При выборе наилучшей операции желательно, чтобы M было больше, а R — меньше.

Будем говорить, что *операция a доминирует операцию b* и обозначать $a \succ b$, где $\succ$ — знак предпочтения, если $M(a) \geq M(b)$ и $R(a) \leq R(b)$ и хотя бы одно из этих неравенств строгое. При этом операция a называется *доминирующей*, а операция b — *доминируемой*.

Доминируемая операция не может быть наилучшей, следовательно, наилучшую операцию надо искать среди недоминируемых операций.

Множество недоминируемых операций называют *множеством Парето* или *множеством оптимальности по Парето.*

На множестве Парето каждая из характеристик M, R — однозначная функция другой, т. е. по характеристике M можно определить характеристику R, и наоборот.

Пример 3.12. По условиям примера 3.11 найдем множество Парето-оптимальных операций (стратегий).

▼Каждую операцию (M_i, R_i) отметим как точку на плоскости (M, R), получим четыре точки (рис. 3.5).

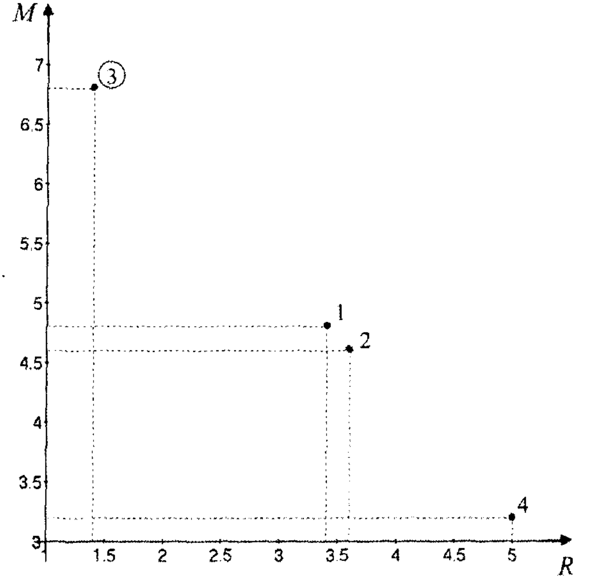


Рис. 3.5. Изображение операций примера 3.11 на плоскости (M, R)

Чем выше точка (M_i, R_i), тем операция более доходная, чем точка правее, тем операция более рискованная. Значит, нужно выбирать точку выше и левее.

В нашем примере множество Парето состоит только из одной третьей операции.

Для нахождения лучшей операции иногда применяют подходящую *взвешивающую формулу* (обобщенный критерий), которая для операции i с характеристиками (M_i, R_i) дает одно число, например, $q(i) = M_i - 2R_i$, по которому и определяют лучшую операцию.

Для операций рассматриваемого примера имеем:

$$q(1) = 4{,}8 - 2 \cdot 3{,}4 = -2;\ q(2) = -2{,}6;\ q(3) = 4;\ q(4) = -6{,}8.$$

Видно, что третья операция — лучшая. ▲

Взвешивающая формула выражает отношение лица, принимающего решение, к доходу и риску. Если ЛПР применяет данную формулу, то он согласен на увеличение риска операции на единицу, если доход операции при этом увеличится не менее чем на две единицы, поскольку $q(M + 2, R + 1) = q(M, R)$.

Пример 3.13. Фирма может выпускать продукцию одного из следующих шести видов: A, B, C, D, E, F, спрос на которые зависит от погоды. Глава фирмы должен принять решение, какой из этих видов продукции выпускать в течение предстоящего летнего сезона.

Прибыль фирмы зависит от того, каким будет лето — дождливым (Д), жарким (Ж) или умеренным (У), и определяется табл. 3.27.

Таблица 3.27

Прибыль от продажи продукции в зависимости от погоды (пример 3.13), ден. ед.

Вид продукции	Погода		
	Д	Ж	У
A	80	60	40
B	70	40	80
C	70	50	60
D	50	50	70
E	75	50	50
F	35	75	60

Выбор какого варианта является оптимальным?

▼Рассмотрим сначала **задачу принятия решения в условиях неопределенности**. Для определения стратегии фирмы используем несколько критериев, приняв $\lambda = 0{,}5$. Составляем платежную матрицу со столбцами M_i, α_i, ω_i, γ_i (табл. 3.28).

Таблица 3.28

Платежная матрица для примера 3.13, ден. ед.

Вид продукции	Погода			M_i	α_i	ω_i	γ_i
	Д	Ж	У				
A	80	60	40	60	40	80*	60
B	70	40	80	63,33*	40	80*	60
C	70	50	60	60	50*	70	60
D	50	50	70	56,67	50*	70	60
E	75	50	50	58,33	50*	75	62,5*
F	35	75	60	56,67	35	75	55

Критерий Лапласа: $\max M_i = 63{,}33 \Rightarrow$ продукция *B*.

Критерий Вальда: $\max \alpha_i = 50 \Rightarrow$ продукция *C*, *D* или *F*.

Критерий максимума: $\max \omega_i = 80 \Rightarrow$ продукция *A* или *B*.

Критерий Гурвица: $\max \gamma_i = 62{,}5 \Rightarrow$ продукция *E*.

Критерий Сэвиджа. Запишем матрицу риска (табл. 3.29).

Таблица 3.29

Матрица риска для примера 3.13, ден. ед.

Вид продукции	Погода			δ_i
	Д	Ж	У	
A	0	15	40	40
B	10	35	0	35
C	10	25	20	25*
D	30	25	10	30
E	5	25	30	30
F	45	0	20	45

Так как $\min \delta_i = 25$, то оптимально будет производить продукцию *C*.

Рассмотренные критерии не позволяют сделать однозначный выбор оптимальной стратегии.

Рассмотрим теперь **задачу принятия решения в условиях риска**. Пусть ЛПР имеет информацию о состоянии среды в виде вероятностей наступления дождливого, жаркого и умеренного лета, равных соответственно $p_1 = 0{,}2$, $p_2 = 0{,}5$, $p_3 = 0{,}3$.

Для каждой стратегии фирмы и различных состояний среды найдем математическое ожидание выигрыша M и среднее квадратическое отклонение выигрыша σ, рассматриваемое в качестве показателя риска, т. е.

$$M_i = \sum_j a_{ij} p_j,\ \sigma_i^2 = \sum_j (a_{ij} - M)^2 p_j,\ \sigma_i = \sqrt{\sigma_i^2}\ .$$

Результаты представим в виде табл. 3.30.

Таблица 3.30

Математическое ожидание выигрыша и среднее квадратическое отклонение выигрыша для примера 3.13, ден. ед.

Вид продукции	M	σ
A	58	14,0
B	58	18,3
C	57	7,8
D	56	9,2
E	55	10,0
F	62,5	15,2

Фактически здесь рассматривается задача двухкритериальной оптимизации, где в качестве решения (операции) выступают M и σ. Операция является рискованной, поскольку она может иметь несколько исходов, не равноценных для ЛПР. При выборе наилучшей операции желательно, чтобы M было больше, а σ меньше.

Найдем множество Парето-оптимальных альтернатив для данной задачи. Каждую операцию (M_i, σ_i) отметим как точку на плоскости (M, σ), получим шесть точек (рис. 3.6).

Из рис. 3.6 очевидно, что Парето-оптимальное множество состоит из трех элементов: $\{A, C, F\}$.

Парето-оптимальные операции нужно сравнить и произвести их ранжирование. В качестве обобщенного критерия сравнения операций используем формулу

$$q(M, \sigma) = M - \lambda\sigma\ ,$$

где $\lambda > 0$ — мера несклонности к риску.

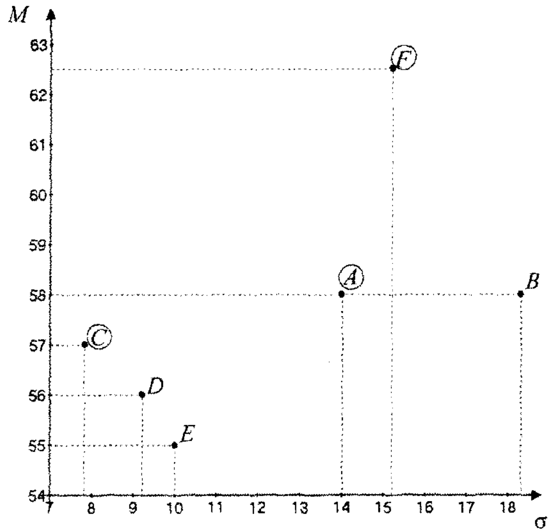


Рис. 3.6. Изображение операций примера 3.13 на плоскости (M, σ)

В данной задаче

$$q(A) = 58 - 14\lambda,\ q(C) = 57 - 7{,}8\lambda,\ q(F) = 62{,}5 - 15{,}2\lambda.$$

Для ранжирования Парето-оптимального множества $\{A, C, F\}$ по обобщенному критерию q для каждой пары операций a, b из условия $q(a) = q(b)$ найдем граничное значение λ, отделяющее предпочтение этой пары:

$$q(A) = q(C) \Rightarrow \lambda = 0{,}16;\ q(A) = q(F) \Rightarrow \lambda = 3{,}75;$$
$$q(C) = q(F) \Rightarrow \lambda = 0{,}74.$$

Из полученных граничных значений λ найдем минимальное и максимальное значения:

$$\min(0{,}16;\ 3{,}75;\ 0{,}74) = 0{,}16,\ \max(0{,}16;\ 3{,}75;\ 0{,}74) = 3{,}75.$$

В результате интервал $(0; \infty)$ изменения параметра λ разбивается на три интервала: $(0; 0{,}16)$, $(0{,}16; 3{,}75)$, $(3{,}75; \infty)$.

Если для ЛПР его мера несклонности к риску $0 \le \lambda < 0{,}16$, то для него ранжирование множества Парето-оптимальных операций совпадает с их ранжированием по величине математического ожидания дохода: $F \succ A \succ C$, оптимальной будет операция F.

При $\lambda > 3{,}75$ ранжирование множества Парето-оптимальных операций совпадает с их ранжированием по показателю риска: $C \succ A \succ F$, оптимальной будет операция C.

В случае $0{,}16 \le \lambda \le 3{,}75$ мера несклонности к риску находится в зоне неопределенности:

• если взять $\lambda = 0{,}5 < 0{,}74$, то

$$q(A) = 58 - 14 \cdot 0{,}5 = 51;$$
$$q(C) = 57 - 7{,}8 \cdot 0{,}5 = 53{,}1;$$
$$q(F) = 62{,}5 - 15{,}2 \cdot 0{,}5 = 54{,}9.$$

Получили следующее ранжирование операций: $F \succ C \succ A$, оптимальной операцией будет F;

• если взять $\lambda = 1 > 0{,}74$, то

$$q(A) = 58 - 14 = 44;$$
$$q(C) = 57 - 7{,}8 = 49{,}2;$$
$$q(F) = 62{,}5 - 15{,}2 = 47{,}3.$$

Получили следующее ранжирование операций: $C \succ F \succ A$, оптимальной операцией будет C. ▲

Глава 4.
МОДЕЛЬ МЕЖОТРАСЛЕВОГО БАЛАНСА

4.1. Структура и содержание таблицы межотраслевого баланса

Эффективное функционирование экономики предполагает наличие *баланса* между отдельными отраслями. Каждая отрасль при этом выступает двояко: с одной стороны, как *производитель* некоторой продукции, а с другой — как *потребитель* продуктов, вырабатываемых другими отраслями.

Межотраслевой баланс (МОБ) представляет собой таблицу, в которой отражены взаимные связи между отраслями.

Различают отчетный и плановый межотраслевые балансы. Такие балансы могут составляться для страны, региона и предприятия. *Отчетный межотраслевой баланс* отражает структуру производства и потребления продукции, произведенной в стране за отчетный год. *Плановый межотраслевой баланс* предназначен для планирования производства валового внутреннего продукта.

В зависимости от того, в каких единицах измеряются межотраслевые потоки, различают балансы *натуральные* и *стоимостные*.

Рассмотрим наиболее простой вариант модели баланса.

Пусть имеется n различных отраслей промышленности, каждая из которых производит свою продукцию. Часть производимой продукции идет на потребление этой и другими отраслями в качестве сырья, а другая часть идет на непроизводственное потребление.

Рассмотрим процесс производства за некоторый период времени, например год.

Введем обозначения:

x_i — общий (валовой) выпуск i-й отрасли;

x_{ij} — объем продукции i-й отрасли, потребляемой j-й отраслью в процессе производства (производственное потребление);

y_i — объем продукции i-й отрасли для непроизводственного потребления (конечное потребление).

Общий вид межотраслевого баланса представлен в табл. 4.1.

Таблица 4.1

Общий вид межотраслевого баланса

Номер отрасли	**Производственное потребление**				**Конечное потребление**	**Валовой выпуск**
	1	**2**	**...**	***n***		
1	x_{11}	x_{12}	...	x_{1n}	y_1	x_1
2	x_{21}	x_{22}	...	x_{2n}	y_2	x_2
...	...	...	...	...	...	...
n	x_{n1}	x_{n2}	...	x_{nn}	y_n	x_n

Валовой объем продукции равен суммарному объему продукции, потребляемой всеми отраслями, и конечного продукта.

Соотношениями межотраслевого баланса называются уравнения

$$\begin{cases} x_1 = x_{11} + x_{12} + \ldots + x_{1n} + y_1, \\ x_2 = x_{21} + x_{22} + \ldots + x_{2n} + y_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_n = x_{n1} + x_{n2} + \ldots + x_{nn} + y_n, \end{cases} \quad \text{или} \quad x_i = \sum_{j=1}^{n} x_{ij} + y_i,\ i = \overline{1,n}.$$

Будем рассматривать стоимостный МОБ, когда все величины, входящие в уравнения, имеют стоимостное выражение.

4.2. Коэффициенты прямых и полных затрат

Коэффициентами прямых затрат (технологическими коэффициентами) называются величины

$$a_{ij} = x_{ij}/x_j,\ 0 \le a_{ij} < 1,\ i, j = \overline{1,n},$$

показывающие затраты продукции i-й отрасли на выпуск одной единицы продукции j-й отрасли.

Предполагается, что коэффициенты a_{ij} остаются постоянными в течение ряда лет. Это означает линейную зависимость материальных затрат от валового выпуска, т. е.

$$x_{ij} = a_{ij} x_j ,\ i, j = \overline{1, n}$$

В предположении линейности соотношения МОБ примут вид

$$\begin{cases} x_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + y_1, \\ x_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + y_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_n = a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n + y_n. \end{cases}$$

Обозначим $X = \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix}$, $Y = \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix}$, $A = \begin{pmatrix} a_{11} \ldots a_{1n} \\ a_{21} \ldots a_{2n} \\ \ldots\ldots\ldots\ldots \\ a_{n1} \ldots a_{nn} \end{pmatrix}$,

где X — вектор валового выпуска;

Y — вектор конечного продукта;

A — матрица прямых затрат.

Тогда систему МОБ можно записать в матричном виде:

$$X = AX + Y.$$

Это уравнение называется уравнением линейного МОБ.

Основная задача МОБ состоит в отыскании такого вектора валового выпуска X, который при известной матрице прямых затрат A обеспечивает вектор конечного продукта Y.

В соответствии с экономическим смыслом задачи имеем:

1) все компоненты матрицы A и вектора Y являются положительными (для краткости это записывают как $A \geq 0$, $Y \geq 0$);

2) все компоненты вектора X также должны быть неотрицательными ($X \geq 0$).

Матрица A называется *продуктивной*, если для любого вектора $Y \geq 0$ существует решение $X \geq 0$ уравнения $X = AX + Y$. В этом случае и модель МОБ называется продуктивной.

Перепишем уравнение $X = AX + Y$ в виде $(E - A)X = Y$, тогда если $\det(E - A) \neq 0$, то $X = (E - A)^{-1}Y$.

Критерий продуктивности. Матрица A продуктивна тогда и только тогда, когда матрица $S = (E - A)^{-1}$ существует и неотрицательна.

Используя матрицу S, можно записать

$$X = SY.$$

Матрица S называется *матрицей полных затрат*. Элементы s_{ij} матрицы S называются *коэффициентами полных затрат*. Они показывают величину валового выпуска продукции i-й отрасли, необходимую для обеспечения выпуска единицы конечного продукта j-й отрасли.

Пример 4.1. В табл. 4.2 приведены данные по балансу между двумя отраслями за отчетный период.

Таблица 4.2

Межотраслевой баланс для примера 4.1, усл. ед.

Отрасль	Потребление		Конечный продукт	Валовой выпуск	Расчет
	1	2			
1	100	160	240	500	100 + 160 + 240 = 500
2	275	40	85	400	275 + 40 + 85 = 400

Вычислим необходимый объем валового выпуска каждой отрасли, если конечный продукт первой отрасли должен увеличиться в два раза, а конечный продукт второй отрасли — на 20 %.

▼По условию $x_1 = 500$, $x_2 = 400$, $y_1 = 240$, $y_2 = 85$, $x_{11} = 100$, $x_{12} = 160$, $x_{21} = 275$, $x_{22} = 40$.

Найдем коэффициенты прямых затрат.

$$a_{11} = \frac{x_{11}}{x_1} = \frac{100}{500} = 0,2\,;\; a_{12} = \frac{x_{12}}{x_2} = \frac{160}{400} = 0,4\,;$$

$$a_{21} = \frac{x_{21}}{x_1} = \frac{275}{500} = 0,55\,;\; a_{22} = \frac{x_{22}}{x_2} = \frac{40}{400} = 0,1\,,$$

т. е. $A = \begin{pmatrix} 0,2 & 0,4 \\ 0,55 & 0,1 \end{pmatrix}$.

Тогда $(E-A)=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}-\begin{pmatrix}0,2 & 0,4\\ 0,55 & 0,1\end{pmatrix}=\begin{pmatrix}0,8 & -0,4\\ -0,55 & 0,9\end{pmatrix}$.

Поскольку $\det(E-A)=\begin{vmatrix}0,8 & -0,4\\ -0,55 & 0,9\end{vmatrix}=0,5\neq 0$, то существует обратная матрица $(E-A)^{-1}=S$, равная $S=(E-A)^{-1}=\dfrac{1}{0,5}\times$ $\times\begin{pmatrix}0,9 & 0,4\\ 0,55 & 0,8\end{pmatrix}=\begin{pmatrix}1,8 & 0,8\\ 1,1 & 1,6\end{pmatrix}$.

Все элементы матрицы S неотрицательны, следовательно, матрица A продуктивна.

По условию задачи вектор конечного продукта $Y=\begin{pmatrix}2\cdot 240\\ 1,2\cdot 85\end{pmatrix}=$ $=\begin{pmatrix}480\\ 102\end{pmatrix}$, тогда $X=SY=\begin{pmatrix}1,8 & 0,8\\ 1,1 & 1,6\end{pmatrix}\cdot\begin{pmatrix}480\\ 102\end{pmatrix}=\begin{pmatrix}945,6\\ 691,2\end{pmatrix}$, т. е. валовой выпуск в 1-й отрасли надо увеличить до 945,6 усл. ед., а во 2-й отрасли до 691,2 усл. ед. ▲

Замечание. При расчете обратной матрицы и умножении матриц можно воспользоваться функциями *Excel*: МОБР и МУМНОЖ.

При использовании МОБР необходимо:

- выделить диапазон ячеек для размещения обратной матрицы;
- выбрать функцию МОБР;
- ввести диапазон ячеек, где содержится матрица $(E-A)$;
- нажать клавиши CTRL + SHIFT + ENTER.

При использовании МУМНОЖ необходимо:

- выделить диапазон ячеек для размещения результата;
- выбрать функцию МУМНОЖ;
- ввести диапазон ячеек, где содержатся матрицы S, Y;
- нажать клавиши CTRL + SHIFT + ENTER.

Кроме того, при расчете формул $a_{ij}=x_{ij}/x_j$ или $x_{ij}=a_{ij}x_j$ удобно предварительно вектор-столбец валового выпуска X преобразовать в вектор-строку (транспонировать) с помощью функции *Excel* ТРАНСП.

При использовании ТРАНСП необходимо:

• выделить диапазон ячеек для размещения вектора-строки X;

• выбрать функцию ТРАНСП;

• ввести диапазон ячеек, где содержится вектор-столбец X;

• нажать клавиши CTRL + SHIFT + ENTER.

Пример 4.2. В табл. 4.3 приведены данные баланса трех отраслей промышленности за отчетный период.

Таблица 4.3

Межотраслевой баланс отчетного периода для примера 4.2, усл. ед.

Отрасль	**Потребление (x_{ij})**			**Конечный продукт (Y)**	**Валовой выпуск (X)**
	1	**2**	**3**		
1	5	35	20	40	100
2	10	10	20	50	100
3	20	10	10	10	50

Составим МОБ планового периода, если каждый конечный продукт по отрасли увеличится соответственно до 60, 70 и 30 усл. ед. Коэффициенты прямых затрат те же, что и в отчетном периоде.

▼Матрица прямых затрат $A = \begin{pmatrix} 0{,}05 & 0{,}35 & 0{,}40 \\ 0{,}10 & 0{,}10 & 0{,}40 \\ 0{,}20 & 0{,}10 & 0{,}20 \end{pmatrix}$ следует из формулы $a_{ij} = x_{ij}/x_j$.

Матрица $(E - A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0{,}05 & 0{,}35 & 0{,}40 \\ 0{,}10 & 0{,}10 & 0{,}40 \\ 0{,}20 & 0{,}10 & 0{,}20 \end{pmatrix} =$

$$= \begin{pmatrix} 0{,}95 & -0{,}35 & -0{,}40 \\ -0{,}10 & 0{,}90 & -0{,}40 \\ -0{,}20 & -0{,}10 & 0{,}80 \end{pmatrix}.$$

Матрица полных затрат $S = (E - A)^{-1} = \begin{pmatrix} 1{,}322 & 0{,}623 & 0{,}973 \\ 0{,}311 & 1{,}323 & 0{,}817 \\ 0{,}370 & 0{,}321 & 1{,}595 \end{pmatrix}$.

Вектор конечного продукта планового периода $Y = \begin{pmatrix} 60 \\ 70 \\ 30 \end{pmatrix}$, тогда валовой выпуск $X = SY = \begin{pmatrix} 1,322 & 0,623 & 0,973 \\ 0,311 & 1,323 & 0,817 \\ 0,370 & 0,321 & 1,595 \end{pmatrix} \cdot \begin{pmatrix} 60 \\ 70 \\ 30 \end{pmatrix} =$ $= \begin{pmatrix} 152,1 \\ 135,8 \\ 92,5 \end{pmatrix}$. Межотраслевые поставки $x_{ij} = \begin{pmatrix} 7,6 & 47,5 & 37,0 \\ 15,2 & 13,6 & 37,0 \\ 30,4 & 13,6 & 18,5 \end{pmatrix}$ следуют из формулы $x_{ij} = a_{ij}x_j$.

Приведем результаты расчета на рабочем листе *Excel* (рис. 4.1).

G4 fx {=ТРАНСП(E3:E5)}

	A	B	C	D	E	F	G	H	I	J	K
1	МОБ отчетного периода										
2		x_{ij}		Y	X						
3	5	35	20	40	100		X (транспонирован)				
4	10	10	20	50	100		100	100	50		
5	20	10	10	10	50						
6											
7		A				E				$E - A$	
8	0.05	0.35	0.40		1	0	0		0.95	-0.35	-0.40
9	0,10	0.10	0.40		0	1	0		-0.10	0.90	-0.40
10	0,20	0,10	0,20		0	0	1		-0.20	-0.10	0.80
11											
12		$S = (E - A)^{-1}$			Y		$X = SY$				
13	1.32	0.62	0.97		60		152.14		X (транспонирован)		
14	0.31	1.32	0.82		70		135.80		152.14	135.80	92,51
15	0.37	0,32	1,60		30		92.51				
16											
17							МОБ планового периода				
18		x_{ij}				x_{ij}		Y	X		
19	7,61	47.53	37.00		7.61	47.53	37.00	60	152.14		
20	15,21	13.58	37,00		15,21	13.58	37.00	70	135.80		
21	30,43	13,58	18.50		30.43	13.58	18.50	30	92.51		

Рис. 4.1. Пример 4.2 в *Excel*

Поясним ввод формул $a_{ij} = x_{ij}/x_j$ и $x_{ij} = a_{ij}x_j$ в соответствующие ячейки таблицы. В ячейку A8 вводим формулу =A3/G$4 и «протягиваем» ее вначале по горизонтали, затем по вертикали до С10 включительно. В ячейку А19 вводим формулу =А8*I$14 и «протягиваем» ее вначале по горизонтали, затем по вертикали до С21 включительно.

В результате МОБ планового периода имеет вид, представленный в табл. 4.4.

Таблица 4.4

Межотраслевой баланс планового периода для примера 4.2, усл. ед.

Отрасль	Потребление (x_{ij})			Конечный продукт (Y)	Валовой выпуск (X)
	1	2	3		
1	7,61	47,53	37,00	60	152,14
2	15,21	13,58	37,00	70	135,80
3	30,43	13,58	18,50	30	92,51

▲

Пример 4.3. Рассматривается трехотраслевой МОБ. Известна матрица коэффициентов прямых затрат и задан вектор конечного продукта:

$$A=\begin{pmatrix}0,3 & 0,1 & 0,0\\ 0,2 & 0,4 & 0,1\\ 0,1 & 0,0 & 0,2\end{pmatrix},\ Y=\begin{pmatrix}100\\ 200\\ 300\end{pmatrix}.$$

Определим валовое производство X, обеспечивающее заданный конечный продукт, межотраслевые поставки продукции, и приведем схему межотраслевого баланса.

▼Матрица $(E - A) = \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} - \begin{pmatrix}0,3 & 0,1 & 0,0\\ 0,2 & 0,4 & 0,1\\ 0,1 & 0,0 & 0,2\end{pmatrix} =$

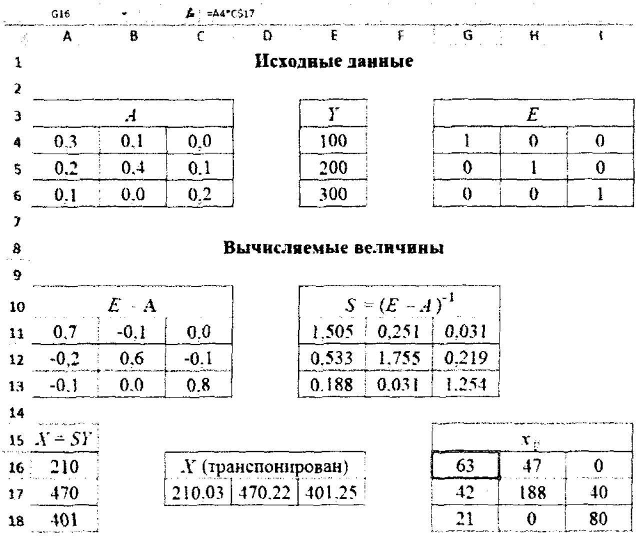
$= \begin{pmatrix}0,7 & -0,1 & 0,0\\ -0,2 & 0,6 & -0,1\\ -0,1 & 0,0 & 0,8\end{pmatrix}$. Матрица полных затрат $S = (E - A)^{-1} =$

$= \begin{pmatrix}1,505 & 0,251 & 0,031\\ 0,533 & 1,755 & 0,219\\ 0,188 & 0,031 & 1,254\end{pmatrix}$. Валовое производство $X = SY =$

$$= \begin{pmatrix} 1{,}505 & 0{,}251 & 0{,}031 \\ 0{,}533 & 1{,}755 & 0{,}219 \\ 0{,}188 & 0{,}031 & 1{,}254 \end{pmatrix} \cdot \begin{pmatrix} 100 \\ 200 \\ 300 \end{pmatrix} = \begin{pmatrix} 210 \\ 470 \\ 401 \end{pmatrix}.$$ Межотраслевые постав-

ки $x_{ij} = \begin{pmatrix} 63 & 47 & 0 \\ 42 & 188 & 40 \\ 21 & 0 & 80 \end{pmatrix}$ следуют из формулы $x_{ij} = a_{ij}x_j$.

Приведем результаты расчета на рабочем листе *Excel* (рис. 4.2).



G16 f_x =A4*C$17

	A	B	C	D	E	F	G	H	I
1				Исходные данные					
2									
3	A				Y		E		
4	0.3	0.1	0.0		100		1	0	0
5	0.2	0.4	0.1		200		0	1	0
6	0.1	0.0	0.2		300		0	0	1
7									
8				Вычисляемые величины					
9									
10	E - A				$S=(E-A)^{-1}$				
11	0.7	-0.1	0.0		1.505	0.251	0.031		
12	-0.2	0.6	-0.1		0.533	1.755	0.219		
13	-0.1	0.0	0.8		0.188	0.031	1.254		
14									
15	$X=SY$						x_{ij}		
16	210		X (транспонирован)				63	47	0
17	470		210.03	470.22	401.25		42	188	40
18	401						21	0	80

Рис. 4.2. Пример 4.3 в *Excel*

В результате схема МОБ имеет вид, представленный в табл. 4.5.

Таблица 4.5

Межотраслевой баланс для примера 4.3

Отрасль	Потребление			Конечный продукт	Валовой выпуск
	1	2	3		
1	63	47	0	100	210
2	42	188	40	200	470
3	21	0	80	300	401

▲

Глава 5.
ЭЛЕМЕНТЫ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

5.1. Структура и классификация систем массового обслуживания

Системы, в которых в случайные моменты времени возникают заявки на обслуживание и имеются устройства для обслуживания этих заявок, называются **системами массового обслуживания (СМО)**.

Для любой СМО характерно многократное выполнение однотипных операций. Примерами СМО могут служить банки, магазины, билетные кассы, телефонные станции.

Каждая СМО состоит из определенного числа обслуживающих устройств, называемых *каналами обслуживания*. В качестве каналов обслуживания могут выступать кассиры, продавцы, линии связи и т. д.

На вход в СМО поступает поток заявок (требований) на обслуживание. Заявки на обслуживание и продолжительность обслуживания носят случайный характер. Это создает нерегулярность в работе СМО, служит причиной его перегрузок и недогрузок. В какие-то периоды времени скапливается очень большое количество заявок, которые либо становятся в очередь, либо покидают СМО необслуженными.

Предметом теории массового обслуживания является установление зависимости между факторами, определяющими функциональные возможности системы массового обслуживания, и эффективностью ее функционирования.

По числу каналов обслуживания различают:

• *одноканальные* СМО;

• *многоканальные* СМО.

В зависимости от характера образования очереди различают:

• *системы с отказами* (без очереди), в которых при занятости всех каналов обслуживания заявка покидает систему необслуженной;

• *системы с неограниченным ожиданием* (очередью), в которых заявка встает в очередь, если в момент ее поступления все каналы обслуживания были заняты;

• *системы с ожиданием и ограниченной очередью*, в которых время ожидания ограничено какими-либо условиями или существуют ограничения на число заявок, стоящих в очереди.

По характеру обслуживания заявок различают:

• *замкнутые* СМО;

• *разомкнутые* СМО.

В замкнутых СМО обслуживается ограниченное число потребителей и обслуженная заявка через некоторое время вновь возвращается на обслуживание. В разомкнутых СМО поток поступающих на обслуживание заявок практически нескончаем.

В качестве *показателей эффективности* СМО используются: среднее число заявок, обслуживаемых в единицу времени; среднее число заявок в очереди; среднее время ожидания обслуживания; вероятность отказа в обслуживании без ожидания и т. п. (средние величины понимаются как математические ожидания соответствующих случайных величин).

Поток событий. Под *потоком событий* понимается последовательность однородных событий, следующих одно за другим в какие-то случайные моменты времени. Например, поток вызовов на телефонной станции, поток покупателей и т. п.

Поток событий называется *стационарным*, если вероятность наступления того или иного числа событий в течение некоторого промежутка времени зависит только от длины этого промежутка.

Поток событий называется *потоком без последствий*, если число событий, попадающих на некоторый участок времени, не зависит от числа событий, попадающих на другие участки.

Поток событий называется *ординарным*, если невозможно наступление двух и более событий.

Поток событий называется *пуассоновским* (простейшим), если он обладает тремя свойствами: стационарен, ординарен и не имеет последствий.

Интенсивностью потока λ называется частота появления события, или среднее число событий в единицу времени. Для стационарного потока интенсивность постоянна.

Для пуассоновского потока вероятность поступления на обслуживание m событий за промежуток времени t находится по формуле Пуассона

$$P_m(t) = \frac{(\lambda t)^m}{m!} e^{-\lambda t}, \; m = 0, 1, 2, \ldots$$

Интервал времени $t_{\text{пост}}$ между двумя соседними произвольными заявками имеет показательное распределение с плотностью распределения

$$f(t) = \lambda e^{-\lambda t},$$

для которого математическое ожидание $M(t_{\text{пост}}) = 1/\lambda$. Величина $M(t_{\text{пост}}) = \bar{t}_{\text{пост}}$ обозначает средний интервал времени между соседними поступающими заявками.

Величина $\lambda = 1/\bar{t}_{\text{пост}}$ называется *интенсивностью потока заявок* и отражает среднее число заявок, поступающих в систему обслуживания в единицу времени.

Интервал $t_{\text{пост}}$ между поступлениями — величина случайная, однако средний интервал $\bar{t}_{\text{пост}}$ предполагается заданным. Пусть $\bar{t}_{\text{пост}} = 6$ мин, тогда интенсивность поступления заявок λ (среднее число заявок в единицу времени) есть 1/6, т. е. на обслуживание за 1 мин в среднем поступает 1/6 заявок.

Кроме характеристик входного потока заявок режим СМО зависит еще от характеристик системы обслуживания: числа каналов обслуживания и быстродействия каждого канала.

Время обслуживания $t_{\text{об}}$, в течение которого какая-либо заявка находится на обслуживании, называется *длительностью обслуживания* и является случайной величиной. Как правило, эта случайная величина подчиняется показательному закону распределения с плотностью распределения

$$g(t) = \mu e^{-\mu t},$$

для которой математическое ожидание $M(t_{об}) = 1/\mu$, а величина $M(t_{об}) = \bar{t}_{об}$ — средняя длительность обслуживания.

Величина $\mu = 1/\bar{t}_{об}$ называется *интенсивностью потока обслуживания* и отражает среднее число заявок, обслуживаемых каналом за единицу времени.

Граф состояний. Процесс работы СМО — это *случайный процесс*, т. е. состояния системы меняются в соответствии с вероятностными закономерностями. Будем рассматривать только *процессы с дискретными состояниями*: все возможные состояния S_0, S_1, S_2, ... системы известны заранее, а переход из одного возможного состояния в другое происходит скачком в момент, когда происходит какое-то случайное событие (появление новой заявки, начало или окончание обслуживания, уход заявки из очереди и т. д.).

Если моменты возможных переходов системы из состояния в состояние не фиксированы заранее, а случайны, то это *процесс с непрерывным временем.*

Таким образом, процесс работы СМО представляет собой случайный процесс с дискретными состояниями и непрерывным временем.

Для СМО с простейшим входным потоком и экспоненциальным временем обслуживания характерно *отсутствие последствия*, т. е. будущее развитие процесса зависит только от текущего состояния и не зависит от того, как происходило развитие в прошлом.

При анализе случайных процессов с дискретными состояниями удобно пользоваться специальной схемой — *графом состояний.* Состояния системы S_i изображаются кружками (прямоугольниками), а возможный переход из состояния S_i в состояние S_j — стрелкой, соединяющей эти состояния.

Очень часто на стрелках указывают *интенсивности* соответствующих переходов (среднее число переходов в единицу времени). Такой граф состояний называют *размеченным графом состояний.*

Пример 5.1. Система состоит из двух узлов, каждый из которых в случайный момент времени может выйти из строя, после чего мгновенно начинается ремонт узла, продолжающийся заранее неизвестное случайное время. Построим граф состояний этой системы.

▼Возможные состояния системы: S_0 (оба узла работают), S_1 (первый узел работает, второй ремонтируется), S_2 (второй узел работает, первый ремонтируется), S_3 (оба узла ремонтируются). Граф системы приведен на рис. 5.1, где λ_{ij} — интенсивность перехода из состояния S_i в состояние S_j.

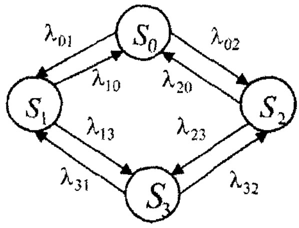


Рис. 5.1. Граф для примера 5.1

На графе рис. 5.1 отсутствуют стрелки из S_0 в S_3 и из S_1 в S_2. Это объясняется тем, что выходы узлов из строя предполагаются независимыми друг от друга, поэтому вероятностью одновременного выхода из строя двух узлов (переход из S_0 в S_3) или одновременного окончания ремонта двух узлов (переход из S_3 в S_0) можно пренебречь. ▲

Уравнения Колмогорова. Пусть $S_0, S_1, \ldots, S_n$ — всевозможные состояния СМО. При исследовании СМО нужно уметь находить вектор вероятностей $(p_0(t), p_1(t), \ldots, p_n(t))$ состояний системы в любой момент времени, где $p_i(t)$ — вероятность того, что в момент времени t система находится в состоянии S_i.

Используя граф состояний, можно составить систему уравнений Колмогорова. При составлении i-го уравнения ($i = 0, \ldots, n$) этой системы нужно рассмотреть состояние S_i на размеченном графе состояний.

Правило составления уравнений Колмогорова. В левой части уравнения находится производная по времени $p_i'(t)$. Каждой входящей в состояние S_i из состояния S_j стрелке (на ней указана интенсивность λ_{ji}) в правой части уравнения соответствует слагаемое $\lambda_{ji} p_j(t)$ со знаком «+». Каждой выходящей из состояния S_i в состояние S_k стрелке (на ней указана интенсивность λ_{ik}) в правой части уравнения соответствует слагаемое $\lambda_{ik}p_i(t)$ со знаком «–». Также надо задать вероятности состояний в момент времени $t = 0$.

Пример 5.2. Запишем уравнения Колмогорова для системы из примера 5.1 (рис. 5.1).

▼Считаем, что в начальный момент времени $t = 0$ система находилась в состоянии S_0, тогда $p_0(0) = 1$, $p_1(0) = p_2(0) = p_3(0) = 0$.

Рассмотрим состояние S_0. В состояние S_0 входят стрелки из состояний S_1 и S_2 с интенсивностями λ_{10} и λ_{20} соответственно. Им соответствуют слагаемые $\lambda_{10}p_1(t) + \lambda_{20}p_2(t)$. Из состояния S_0 выходят стрелки в состояния S_1 и S_2 с интенсивностями λ_{01} и λ_{02} соответственно. Им соответствуют слагаемые $-\lambda_{01}p_0(t) - \lambda_{02}p_0(t) = -(\lambda_{01} + \lambda_{02})p_0(t)$. Тогда уравнение Колмогорова для состояния S_0 будет иметь вид

$$p_0'(t) = \lambda_{10}p_1(t) + \lambda_{20}p_2(t) - (\lambda_{01} + \lambda_{02})p_0(t).$$

Рассмотрим состояние S_1. В состояние S_1 входят стрелки из состояний S_0 и S_3 с интенсивностями λ_{01} и λ_{31} соответственно. Им соответствуют слагаемые $\lambda_{01}p_0(t) + \lambda_{31}p_3(t)$. Из состояния S_1 выходят стрелки в состояния S_0 и S_3 с интенсивностями λ_{10} и λ_{13} соответственно. Им соответствуют слагаемые $-\lambda_{10}p_1(t) - \lambda_{13}p_1(t) = -(\lambda_{10} + \lambda_{13})p_1(t)$. Тогда уравнение Колмогорова для состояния S_1 будет иметь вид

$$p_1'(t) = \lambda_{01}p_0(t) + \lambda_{31}p_3(t) - (\lambda_{10} + \lambda_{13})p_1(t).$$

Аналогично составляем уравнения для остальных состояний. В результате получаем систему уравнений Колмогорова

$$\begin{cases} p_0'(t) = \lambda_{10}p_1(t) + \lambda_{20}p_2(t) - (\lambda_{01} + \lambda_{02})p_0(t), \\ p_1'(t) = \lambda_{01}p_0(t) + \lambda_{31}p_3(t) - (\lambda_{10} + \lambda_{13})p_1(t), \\ p_2'(t) = \lambda_{02}p_0(t) + \lambda_{32}p_3(t) - (\lambda_{20} + \lambda_{23})p_2(t), \\ p_3'(t) = \lambda_{13}p_1(t) + \lambda_{23}p_2(t) - (\lambda_{31} + \lambda_{32})p_3(t), \\ p_0(0) = 1,\ p_i(0) = 0,\ i = 1, 2, 3. \end{cases} \quad ▲$$

Предельные вероятности состояний. Уравнения Колмогорова дают возможность найти все вероятности состояний как *функции времени.* Особый интерес представляют вероятности системы $p_i(t)$ в *предельном стационарном режиме,* т. е. при $t \to \infty$, которые называются *предельными* (или *финальными*) вероятностями состояний.

В теории случайных процессов доказывается, что *если число состояний системы конечно и из каждого из них можно (за конечное число шагов) перейти в любое другое состояние, то предельные вероятности существуют.*

Предельная вероятность состояния S_i имеет четкий смысл: она показывает *среднее относительное время пребывания системы в этом состоянии.* Например, если предельная вероятность состояния S_0, т. е. $p_0 = 0{,}5$, то это означает, что в среднем половину времени система находится в состоянии S_0.

Так как предельные вероятности постоянны, то, заменяя в уравнениях Колмогорова их производные нулевыми значениями, получим систему линейных алгебраических уравнений, описывающих стационарный режим. Для системы S с графом состояний, изображенным на рис. 5.1, такая система уравнений имеет вид

$$\begin{cases} (\lambda_{01} + \lambda_{02})p_0 = \lambda_{10}p_1 + \lambda_{20}p_2, \\ (\lambda_{10} + \lambda_{13})p_1 = \lambda_{01}p_0 + \lambda_{31}p_3, \\ (\lambda_{20} + \lambda_{23})p_2 = \lambda_{02}p_0 + \lambda_{32}p_3, \\ (\lambda_{31} + \lambda_{32})p_3 = \lambda_{13}p_1 + \lambda_{23}p_2. \end{cases}$$

Данную систему можно составить непосредственно по размеченному графу состояний по правилу: *слева в уравнениях стоит предельная вероятность данного состояния p_i, умноженная на суммарную интенсивность всех потоков, ведущих из данного состояния, а справа — сумма произведений интенсивностей всех потоков, входящих в i-е состояние, на вероятности тех состояний, из которых эти потоки исходят.*

Пример 5.3. Найти предельные вероятности для системы S примера 5.1, рис. 5.1, при $\lambda_{01} = 1$, $\lambda_{02} = 2$, $\lambda_{10} = 2$, $\lambda_{13} = 2$, $\lambda_{20} = 3$, $\lambda_{23} = 1$, $\lambda_{31} = 3$, $\lambda_{32} = 2$.

▼Система алгебраических уравнений, описывающих стационарный режим для данной системы, имеет вид

$$\begin{cases} 3p_0 = 2p_1 + 3p_2, \\ 4p_1 = p_0 + 3p_3, \\ 4p_2 = 2p_0 + 2p_3, \\ 5p_3 = 2p_1 + p_2. \end{cases}$$

Последнее уравнение системы есть сумма трех предыдущих уравнений, поэтому вместо него включим нормировочное условие $p_0 + p_1 + p_2 + p_3 = 1$:

$$\begin{cases} 3p_0 = 2p_1 + 3p_2, \\ 4p_1 = p_0 + 3p_3, \\ 4p_2 = 2p_0 + 2p_3, \\ p_0 + p_1 + p_2 + p_3 = 1. \end{cases}$$

Решив систему, получим $p_0 = 0{,}40$, $p_1 = 0{,}20$, $p_2 = 0{,}27$, $p_3 = 0{,}13$. В предельном стационарном режиме система S в среднем 40 % времени будет находиться в состоянии S_0 (оба узла исправны), 20 % — в состоянии S_1 (первый узел ремонтируется, второй работает), 27 % — в состоянии S_2 (второй узел ремонтируется, первый работает) и 13 % времени — в состоянии S_3 (оба узла ремонтируются). ▲

Процесс гибели и размножения. В теории массового обслуживания широкое распространение имеет специальный класс случайных процессов — *процессов гибели и размножения*. Размеченный граф состояний такого процесса представлен на рис. 5.2.

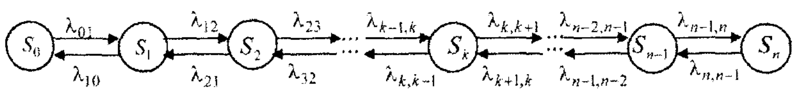


Рис. 5.2. Размеченный граф состояний
для процесса гибели и размножения

Рассмотрим упорядоченное множество состояний системы $S_0, S_1, \ldots, S_n$. Переходы могут осуществляться из любого состояния только в состояния с соседними номерами, т. е. из состояния S_k возможны переходы только либо в состояние S_{k-1}, либо в состояние S_{k+1}. Предположим, что все потоки событий, переводящие систему по стрелкам графа, простейшие с соответствующими интенсивностями $\lambda_{k,\ k+1}$ или $\lambda_{k-1,\ k}$. По графу, представленному на рис. 5.2, составим и решим алгебраические уравнения для предельных вероятностей состояний.

В соответствии с правилом составления таких уравнений получим для состояния S_0: $\lambda_{01}p_0 = \lambda_{10}p_1$, для состояния S_1: $(\lambda_{12} + \lambda_{10})p_1 = \lambda_{01}p_0 + \lambda_{21}p_2$, которое с учетом предыдущего уравнения приводится к виду $\lambda_{12}p_1 = \lambda_{21}p_2$.

Аналогично записывая уравнения для предельных вероятностей других состояний, можно получить следующую систему уравнений:

$$\begin{cases} \lambda_{01}p_0 = \lambda_{10}p_1, \\ \lambda_{12}p_1 = \lambda_{21}p_2, \\ \dots\dots\dots\dots\dots\dots \\ \lambda_{k-1,k}p_{k-1} = \lambda_{k,k-1}p_k, \\ \dots\dots\dots\dots\dots\dots \\ \lambda_{n-1,n}p_{n-1} = \lambda_{n,n-1}p_n, \end{cases}$$

к которой добавляется нормировочное условие

$$p_0 + p_1 + p_2 + \ldots + p_n = 1.$$

Решая полученную систему с добавочным условием, получим

$$p_0 = \left(1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} + \ldots + \frac{\lambda_{n-1,n} \cdots \lambda_{12}\lambda_{01}}{\lambda_{n,n-1} \cdots \lambda_{21}\lambda_{10}}\right)^{-1}, \tag{5.1}$$

$$p_1 = \frac{\lambda_{01}}{\lambda_{10}}p_0, \ p_2 = \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}}p_0, \ldots, \ p_n = \frac{\lambda_{n-1,n} \cdots \lambda_{12}\lambda_{01}}{\lambda_{n,n-1} \cdots \lambda_{21}\lambda_{10}}p_0. \tag{5.2}$$

Легко заметить, что в формулах (5.2) для p_1, p_2, …, p_n коэффициенты при p_0 есть слагаемые, стоящие после единицы в формуле (5.1). Числители этих коэффициентов представляют произведение всех интенсивностей, стоящих у стрелок, ведущих слева направо до данного состояния S_k ($k = 1, 2, \ldots, n$), а знаменатели — произведение всех интенсивностей, стоящих у стрелок, ведущих справа налево (с начала и до состояния S_k).

Пример 5.4. Процесс гибели и размножения представлен графом (рис. 5.3). Найти предельные вероятности состояний.

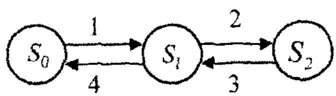


Рис. 5.3. Размеченный граф состояний для примера 5.4

▼По формулам (5.1) и (5.2) найдем

$$p_0 = \left(1 + \frac{1}{4} + \frac{2 \cdot 1}{3 \cdot 4}\right)^{-1} = 0{,}706; \quad p_1 = \frac{1}{4} \cdot 0{,}706 = 0{,}176;$$

$$p_2 = \frac{2 \cdot 1}{3 \cdot 4} \cdot 0{,}706 = 0{,}118,$$

т. е. в установившемся, стационарном режиме в среднем 70,6 % времени система будет находиться в состоянии S_0, 17,6 % — в состоянии S_1 и 11,8 % — в состоянии S_2. ▲

5.2. Средства массового обслуживания с отказами (без очереди)

В качестве показателей эффективности СМО с отказами будем рассматривать:

- A — *абсолютную пропускную способность СМО*, т. е. среднее число заявок, обслуживаемых в единицу времени;
- Q — *относительную пропускную способность*, т. е. среднюю долю пришедших заявок, обслуживаемых системой;
- $P_{\text{отк}}$ — *вероятность отказа*, т. е. того, что заявка покинет СМО необслуженной;
- $\overline{k}$ — *среднее число занятых каналов* (для многоканальной системы).

Одноканальная система с отказами. СМО содержит один обслуживающий канал, на который поступает поток заявок с интенсивностью λ. Образование очереди не допускается. Если заявка застала обслуживающий канал занятым, то она покидает систему.

Время обслуживания заявки есть случайная величина, которая подчиняется экспоненциальному закону распределения с параметром μ. Среднее время обслуживания одной заявки $\overline{t}_{об} = 1/\mu$.

Рассмотрим систему S, которая имеет два состояния: S_0 — канал свободен, S_1 — канал занят. Размеченный граф состояний представлен на рис. 5.4. Найдем предельные вероятности состояний системы и показатели ее эффективности.

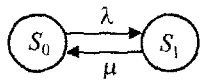


Рис. 5.4. Размеченный граф состояний для одноканальной СМО с отказами

В предельном стационарном режиме система алгебраических уравнений для вероятностей состояний имеет вид

$$\begin{cases} \lambda p_0 = \mu p_1, \\ \mu p_1 = \lambda p_0, \end{cases}$$

т. е. система вырождается в одно уравнение. Учитывая нормировочное условие $p_0 + p_1 = 1$, найдем предельные вероятности состояний

$$p_0 = \frac{\mu}{\lambda + \mu}, \quad p_1 = \frac{\lambda}{\lambda + \mu},$$

которые выражают среднее относительное время пребывания системы в состоянии S_0 (когда канал свободен) и S_1 (когда канал занят), т. е. *определяют* соответственно относительную пропускную способность Q системы и вероятность отказа $P_{отк}$:

$$Q = p_0 = \frac{\mu}{\lambda + \mu}, \quad P_{отк} = p_1 = \frac{\lambda}{\lambda + \mu}.$$

Абсолютную пропускную способность найдем, умножив относительную пропускную способность Q на интенсивность потока отказов:

$$A = \lambda Q = \frac{\lambda\mu}{\lambda + \mu}.$$

Пример 5.5. В городскую справочную службу за необходимыми сведениями по телефонной одноканальной линии ($n = 1$) обращаются в среднем 1 раз за 0,5 мин. Среднее время выполнения одной заявки составляет 2 мин. Определим показатели эффективности справочной службы как СМО, если поток заявок простейший, а при занятости телефонной линии клиент получает отказ в обслуживании.

▼По условию $\bar{t}_{\text{пост}} = 0{,}5$ мин, $\bar{t}_{\text{об}} = 2$ мин, $n = 1$.

Интенсивности поступления и обслуживания заявок:

$$\lambda = \frac{1}{\bar{t}_{\text{пост}}} = \frac{1\ \text{заявка}}{0{,}5\ \text{мин}} = 2\ \text{заяв./мин} = 120\ \text{заяв./ч},$$

$$\mu = \frac{1}{\bar{t}_{\text{об}}} = \frac{1\ \text{заявка}}{2\ \text{мин}} = 0{,}5\ \text{заяв./мин} = 30\ \text{заяв./ч}.$$

Параметр нагрузки $\rho = \lambda / \mu = 120/30 = 4$.

Определим показатели эффективности СМО:

• вероятность, что система свободна (вероятность дозвониться до оператора справочной службы):

$$p_0 = \frac{\mu}{\lambda + \mu} = \frac{30}{120 + 30} = 0{,}2,$$

т. е. 20 % времени в течение часа канал не будет занят;

• вероятность, что система занята (вероятность отказа в обслуживании):

$$p_1 = P_{\text{отк}} = \frac{\lambda}{\lambda + \mu} = \frac{120}{120 + 30} = 0{,}8,$$

т. е. 80 % из числа поступивших заявок не принимаются к обслуживанию;

• относительная пропускная способность:

$$Q = 1 - P_{\text{отк}} = 1 - 0{,}8 = 0{,}2;$$

• абсолютная пропускная способность:

$$A = \lambda Q = 120 \cdot 0,2 = 24 \text{ заяв./ч},$$

т. е. в среднем в час может быть обслужено 24 заявки. ▲

Упражнение 5.1. Известно, что заявки на телефонные переговоры в телевизионном ателье поступают с интенсивностью λ, равной 90 заявок в час, а средняя продолжительность разговора по телефону $\overline{t}_{об}$ = 2 мин. Определить показатели эффективности работы СМО (телефонной связи) при наличии одного телефонного номера.

Ответ: Q = 0,25; $P_{отк}$ = 0,75; A = 22,5 заяв./ч.

Многоканальная система с отказами. Пусть имеется n каналов обслуживания, на которые поступает поток заявок с интенсивностью λ. Образование очереди не допускается. Если заявка застала все обслуживающие каналы занятыми, то она покидает систему. Если в момент поступления требования имеется свободный канал, то он немедленно приступает к обслуживанию поступившего требования. Каждый канал может обслуживать только одно требование. Все каналы функционируют независимо.

Время обслуживания заявки есть случайная величина, которая подчиняется экспоненциальному закону распределения с параметром μ. Среднее время обслуживания одной заявки $\overline{t}_{об} = 1/\mu$. Найдем предельные вероятности состояний системы и показатели ее эффективности.

Возможные состояния СМО: S_0 (все каналы свободны), S_1 (один канал занят, остальные свободны), ..., S_n (все каналы заняты).

Размеченный граф состояний многоканальной СМО соответствует процессу гибели и размножения (рис. 5.5.).

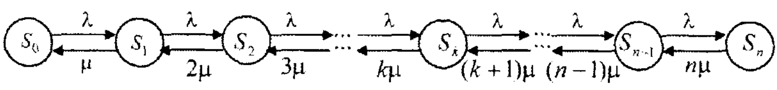


Рис. 5.5. Размеченный граф состояний многоканальной СМО с отказами

Поток заявок последовательно переводит систему из любого левого состояния в соседнее правое с одной и той же интенсивностью λ. Интенсивность же потока обслуживаний, переводящих систему из любого правого состояния в соседнее левое состояние, постоянно меняется в зависимости от состояния.

Действительно, если СМО находится в состоянии S_2 (два канала заняты), то она может перейти в состояние S_1 (один канал занят), когда закончит обслуживание либо первый, либо второй канал, т. е. суммарная интенсивность их потоков обслуживании будет 2μ.

Аналогично суммарный поток обслуживаний, переводящий СМО из состояния S_3 (три канала заняты) в S_2, будет иметь интенсивность 3μ, т. е. может освободиться любой из трех каналов и т. д.

Из формулы (5.1) для схемы гибели и размножения получим предельную вероятность

$$p_0 = \left(1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{2!\mu^2} + \ldots + \frac{\lambda^k}{k!\mu^k} + \ldots + \frac{\lambda^n}{n!\mu^n}\right)^{-1},$$

где члены разложения

$$\frac{\lambda}{\mu},\ \frac{\lambda^2}{2!\mu^2},\ldots,\ \frac{\lambda^n}{n!\mu^n}$$

будут представлять собой коэффициенты при p_0 в выражениях для предельных вероятностей $p_1, p_2, \ldots, p_k, \ldots, p_n$.

Величина

$$\rho = \frac{\lambda}{\mu}$$

называется *приведенной интенсивностью потока заявок* или *интенсивностью нагрузки канала.* Она выражает среднее число заявок, приходящих за среднее время обслуживания одной заявки. Теперь

$$p_0 = \left(1 + \rho + \frac{\rho^2}{2!} + \ldots + \frac{\rho^k}{k!} + \ldots + \frac{\rho^n}{n!}\right)^{-1},\quad p_k = \frac{\rho^k}{k!}p_0,$$

где p_0 — вероятность, что все каналы СМО свободны от обслуживания;

p_k — вероятность, что обслуживанием заняты k каналов ($k = 0, 1, 2, \ldots, n$).

Эти формулы для предельных вероятностей получили название *формул Эрланга.*

Эти формулы позволяют получить следующие показатели эффективности работы СМО.

Вероятность отказа СМО есть предельная вероятность того, что все n каналов системы заняты обслуживанием, т. е.

$$P_{\text{отк}} = p_n = \frac{\rho^n}{n!} p_0.$$

Относительная пропускная способность — вероятность того, что заявка будет обслужена:

$$Q = 1 - P_{\text{отк}} = 1 - \frac{\rho^n}{n!} p_0.$$

Абсолютная пропускная способность — это интенсивность потока обслуженных системой заявок (среднее число заявок, обслуживаемых в единицу времени):

$$A = \lambda Q = \lambda \left(1 - \frac{\rho^n}{n!} p_0\right).$$

Так как каждый занятый канал обслуживает в среднем μ заявок в единицу времени, то *среднее число занятых каналов* есть:

$$\bar{k} = \frac{A}{\mu} = \rho\left(1 - \frac{\rho^n}{n!} p_0\right).$$

Доля каналов, занятых обслуживанием: $q = \dfrac{\bar{k}}{n}$.

Пример 5.6. Стол заказов гастронома принимает заказы по двум телефонам. Среднее число поступающих в течение часа заказов равно 70, среднее время обслуживания заказов — 2 мин. Определим показатели эффективности работы этой СМО.

▼По условию $\lambda = 70$ зак./ч, $\bar{t}_{\text{об}} = 2$ мин, $n = 2$.

Интенсивность обслуживания $\mu = 1/\bar{t}_{об} = 60/2 = 30$ зак./ч.

Параметр нагрузки $\rho = \lambda/\mu = 70/30 = 7/3$.

Вероятность того, что СМО с двумя каналами простаивает:

$$p_0 = \left(1 + \rho + \frac{\rho^2}{2!}\right)^{-1} = \left(1 + \frac{7}{3} + \frac{7^2}{3^2 \cdot 2}\right)^{-1} = \frac{18}{109} = 0{,}165.$$

Вероятность, что обслуживанием занят один канал:

$$p_1 = \frac{\rho^1}{1!} p_0 = \frac{7}{3} \cdot 0{,}165 = 0{,}385.$$

Вероятность, что оба канала заняты обслуживанием (вероятность отказа):

$$p_2 = P_{отк} = \frac{\rho^2}{2!} p_0 = \frac{7^2}{3^2 \cdot 2} \cdot 0{,}165 = 0{,}45.$$

Относительная пропускная способность:

$$Q = 1 - P_{отк} = 1 - 0{,}45 = 0{,}55.$$

Абсолютная пропускная способность:

$$A = \lambda Q = 70 \cdot 0{,}55 = 38{,}5 \text{ зак./ч.}$$

Среднее количество занятых каналов:

$$\bar{k} = \frac{A}{\mu} = \frac{38{,}5}{30} = 1{,}28.$$

Средняя доля каналов, занятых обслуживанием:

$$q = \frac{\bar{k}}{n} = \frac{1{,}28}{2} = 0{,}64. \ \blacktriangle$$

Пример 5.7. По условиям примера 5.6 определим, насколько увеличится относительная пропускная способность стола заказов, если к приему заказов подключить еще один телефон.

▼Для СМО с тремя каналами

$$p_0 = \left(1+\rho+\frac{\rho^2}{2!}+\frac{\rho^3}{3!}\right)^{-1} = \left(1+\frac{7}{3}+\frac{7^2}{3^2\cdot 2}+\frac{7^3}{3^3\cdot 6}\right)^{-1} = \frac{81}{662} = 0,1224 .$$

Вероятность отказа:

$$p_3 = P_{\text{отк}} = \frac{\rho^3}{3!}p_0 = \frac{7^3}{3^3\cdot 6}\cdot 0,1224 = 0,26 .$$

Относительная пропускная способность:

$$Q = 1 - P_{\text{отк}} = 1 - 0,26 = 0,74 .$$

Абсолютная пропускная способность:

$$A = \lambda Q = 70\cdot 0,74 = 51,8 \text{ зак./ч.}$$

Таким образом, подключение к приему заказов еще одного телефона увеличило относительную пропускную способность СМО с 0,55 до 0,74; абсолютная пропускная способность выросла примерно с 39 до 52 зак./ч. ▲

Пример 5.8. Единая телефонная справочная служба железнодорожного вокзала имеет 6 телефонных линий. В дневное время с 8 до 20 часов поступает в среднем 2400 запросов. Оператор на каждой из 6 линий на ответ по одному запросу затрачивает в среднем 1,5 мин. При занятости всех каналов соединения со справочной службой не происходит и клиент получает отказ в обслуживании. Найдем показатели эффективности справочной службы СМО: вероятность отказа в обслуживании, относительную и абсолютную загруженность СМО, среднее число занятых каналов, если поток клиентов простейший. Сколько каналов должно работать одновременно, чтобы из каждых 100 звонивших в справочную службу клиентов отказ в обслуживании получили бы не более 5?

▼По условию $\lambda = 2400$ запросов за 12 часов, $\bar{t}_{\text{об}} = 1,5$ мин, $n = 6$.

Интенсивность потока поступающих в СМО запросов:

$$\lambda = 2400/12 = 200 \text{ запр./ч.}$$

Интенсивность потока запросов, обслуживаемых по каждому из каналов:

$$\mu = 1/\bar{t}_{об} = 60/1{,}5 = 40 \text{ запр./ч.}$$

Параметр загрузки $\rho = \lambda/\mu = 200/40 = 5$.

Вероятность, что СМО с шестью каналами простаивает:

$$p_0 = \left(1 + \rho + \frac{\rho^2}{2!} + \ldots + \frac{\rho^6}{6!}\right)^{-1} = \left(1 + 5 + \frac{5^2}{2!} + \ldots + \frac{5^6}{6!}\right)^{-1} =$$
$$= (113{,}118)^{-1} = 0{,}00884.$$

Вероятность отказа в обслуживании:

$$P_{отк} = p_6 = \frac{\rho^6}{6!} p_0 = \frac{5^6}{6!} \cdot 0{,}00884 = 21{,}701 \cdot 0{,}00884 = 0{,}192.$$

Относительная пропускная способность:

$$Q = 1 - P_{отк} = 1 - 0{,}192 = 0{,}808.$$

Абсолютная пропускная способность:

$$A = \lambda Q = 200 \cdot 0{,}808 = 161{,}6 \text{ запр./ч.}$$

Среднее количество занятых каналов:

$$\bar{k} = \frac{A}{\mu} = \frac{161{,}6}{40} = 4{,}04.$$

Средняя доля каналов, занятых обслуживанием:

$$q = \frac{\bar{k}}{n} = \frac{4{,}04}{6} = 0{,}67.$$

Для того чтобы из каждых 100 звонивших в справочную службу клиентов отказ в обслуживании получили бы не более 5, необходимо, чтобы выполнялось неравенство

$$Q \geq 0{,}95 \ \left(Q \geq \frac{100-5}{100}\right), \text{ а } P_{отк} = 1 - Q \leq 0{,}05.$$

Вычислим значение $P_{\text{отк}}$, увеличивая число каналов n:

• $n = 7;\ p_0 = \left(113{,}118 + \dfrac{5^7}{7!}\right)^{-1} = (113{,}118 + 15{,}501)^{-1} = 0{,}00777;$

$$P_{\text{отк}} = p_7 = \frac{\rho^7}{7!} \cdot p_0 = \frac{5^7}{7!} \cdot 0{,}00777 = 0{,}121,\ Q = 0{,}879;$$

• $n = 8;\ p_0 = \left(128{,}619 + \dfrac{5^8}{8!}\right)^{-1} = (128{,}619 + 9{,}688)^{-1} = 0{,}00728;$

$$P_{\text{отк}} = P_8 = \frac{\rho^8}{8!} \cdot p_0 = \frac{5^8}{8!} \cdot 0{,}00728 = 0{,}071,\ Q = 0{,}929;$$

• $n = 9;\ p_0 = \left(137{,}307 + \dfrac{5^9}{9!}\right)^{-1} = (137{,}307 + 5{,}382)^{-1} = 0{,}00726.$

$$P_{\text{отк}} = p_9 = \frac{\rho^9}{9!} \cdot p_0 = \frac{5^9}{9!} \cdot 0{,}00726 = 0{,}039,\ Q = 0{,}961.$$

Следовательно, для выполнения условия задачи одновременно должно работать 9 каналов. ▲

Упражнение 5.2. Трехканальная телефонная линия. Заявка-вызов, поступившая в момент, когда все $n = 3$ канала заняты, получает отказ. Простейший поток заявок поступает с интенсивностью $\lambda = 60$ звонков/ч. Время обслуживания заявки есть случайная величина, которая подчиняется экспоненциальному закону распределения. Средняя продолжительность разговора $\overline{t}_{\text{об}} = 3$ мин. Определить показатели эффективности работы СМО.

Ответ: $Q = 0{,}654$; $A = 39{,}24$ звонка/ч; $\overline{k} = 1{,}962$; $q = 0{,}654$.

5.3. Средства массового обслуживания с неограниченной очередью

Одноканальная СМО с неограниченной очередью. СМО содержит один обслуживающий канал. На вход поступает простейший поток заявок с интенсивностью λ. Если заявка застала обслуживающий канал занятым, то она встает в очередь и ожидает начала обслуживания.

Время обслуживания заявки есть случайная величина, которая подчиняется экспоненциальному закону распределения с параметром μ. Среднее время обслуживания одной заявки $\overline{t}_{\text{об}} = 1/\mu$. Найдем предельные вероятности состояний системы и показатели ее эффективности.

Возможные состояния СМО: S_0 (канал свободен), S_1 (канал занят, очереди нет), S_2 (канал занят, в очереди одна заявка), S_3 (канал занят, в очереди две заявки) и т. д.

Размеченный граф состояний одноканальной СМО с неограниченной очередью представлен на рис. 5.6.

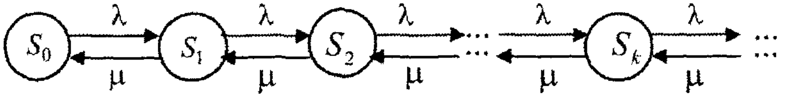


Рис. 5.6. Размеченный граф состояний одноканальной СМО с неограниченной очередью

Это процесс гибели и размножения, но с бесконечным числом состояний, в нем интенсивность потока заявок равна λ, а интенсивность потока обслуживания — μ.

При $\rho = \lambda/\mu < 1$, т. е. среднее число приходящих заявок меньше среднего числа обслуженных заявок (в единицу времени), предельные вероятности существуют. Если $\rho \geq 1$, очередь растет до бесконечности. *Найдем предельные вероятности.*

Для определения предельных вероятностей состояний воспользуемся формулами (5.1), (5.2) для процесса гибели и размножения. Получим

$$p_0 = \left[1 + \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu}\right)^2 + \ldots + \left(\frac{\lambda}{\mu}\right)^k + \ldots\right]^{-1} = \left(1 + \rho + \rho^2 + \ldots + \rho^k + \ldots\right)^{-1}.$$

Выражение в скобках $\left(1 + \rho + \rho^2 + \ldots + \rho^k + \ldots\right)$ есть сумма бесконечно убывающей геометрической прогрессии со знаменателем $\rho < 1$, которая сходится к числу $\frac{1}{1-\rho}$. Тогда $p_0 = 1 - \rho$. Это вероятность того, что канал свободен. Вероятность того, что канал занят, равна $P_{\text{зан}} = 1 - p_0 = \rho$.

С учетом формулы (5.2)

$$p_1 = \rho p_0, \; p_2 = \rho^2 p_0, \; \ldots, \; p_k = \rho^k p_0, \; \ldots$$

Тогда вероятность состояния S_k (канал занят, в очереди $k-1$ заявок) равна $p_k = \rho^k p_0 = \rho^k (1-\rho)$, $k = 1, 2, 3, \ldots$

Предельные вероятности p_0, p_1, p_2, ..., p_k, ... образуют убывающую геометрическую прогрессию со знаменателем $\rho < 1$, следовательно, вероятность p_0 — наибольшая. Это означает, что если СМО справляется с потоком заявок (при $\rho < 1$), то наиболее вероятным будет отсутствие заявок в системе.

Среднее число заявок в системе $L_{\text{СМО}}$ определим по формуле математического ожидания

$$L_{\text{СМО}} = \sum_{k=1}^{\infty} k p_k = (1-\rho) \sum_{k=1}^{\infty} k \rho^k$$

(суммирование от 1 до ∞, так как нулевой член $0 \cdot p_0 = 0$).

Это выражение преобразуется (при $\rho < 1$) к виду

$$L_{\text{СМО}} = \frac{\rho}{1-\rho}.$$

Найдем среднее число заявок в очереди $L_{\text{оч}}$. Очевидно, что

$$L_{\text{оч}} = L_{\text{СМО}} - L_{\text{об}},$$

где $L_{\text{об}}$ — среднее число заявок, находящихся на обслуживании.

Среднее число заявок на обслуживании определим по формуле математического ожидания числа заявок на обслуживании, принимающего значение 0 (если канал свободен) с вероятностью p_0, либо 1 (если канал занят) с вероятностью $(1 - p_0)$:

$$L_{\text{об}} = 0 \cdot p_0 + 1 \cdot (1 - p_0),$$

т. е. среднее число заявок на обслуживании равно вероятности того, что канал занят:

$$L_{\text{об}} = P_{\text{зан}} = 1 - p_0 = \rho.$$

Тогда среднее число заявок в очереди (длина очереди):

$$L_{\text{оч}} = L_{\text{СМО}} - L_{\text{об}} = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho}.$$

Среднее время пребывания заявки в системе $\bar{t}_{\text{СМО}}$ и в очереди $\bar{t}_{\text{оч}}$ соответственно равны

$$\bar{t}_{\text{СМО}} = \frac{L_{\text{СМО}}}{\lambda}, \; \bar{t}_{\text{оч}} = \frac{L_{\text{оч}}}{\lambda} \text{ — формулы Литтла.}$$

Формулы Литтла справедливы при любом потоке заявок, при любом распределении времени обслуживания и при любой дисциплине обслуживания.

На основании формул Литтла среднее время пребывания заявки в системе $\bar{t}_{\text{СМО}}$ и среднее время пребывания заявки в очереди $\bar{t}_{\text{оч}}$ определяются по формулам

$$\bar{t}_{\text{СМО}} = \frac{\rho}{\lambda(1-\rho)}, \; \bar{t}_{\text{оч}} = \frac{\rho^2}{\lambda(1-\rho)},$$

при этом выполняется равенство

$$\bar{t}_{\text{СМО}} = \bar{t}_{\text{оч}} + \bar{t}_{\text{об}}$$

Пример 5.9. Автоматическая мойка для автомобилей имеет только один моечный бокс. На нее прибывают (в соответствии с распределением Пуассона) в среднем 4 автомашины в час, которые могут ожидать обслуживания на улице рядом с мойкой (на очередь ограничений нет). Среднее время мойки автомобиля 10 минут. Найти функциональные характеристики системы.

▼По условию $\lambda = 4$ автомобиля в час, $\bar{t}_{\text{об}} = 10$ мин. Имеем $\mu = 1/\bar{t}_{\text{об}} = 60/10 = 6$ автомобилей в час, $\rho = \lambda/\mu = 4/6 = 2/3 < 1$.

Вероятность простоя (отсутствия клиентов) $p_0 = 1 - \rho = 1 - 2/3 = 1/3$.

Вероятность, что мойка занята $P_{\text{зан}} = 1 - p_0 = \rho = 2/3$.

Вероятность , что в системе находится k клиентов

$$p_k = \rho^k p_0 = \rho^k (1-\rho),\ k = 1, 2, 3, \ldots$$

Среднее число находящихся в системе клиентов:

$$L_{\text{СМО}} = \frac{\rho}{1-\rho} = \frac{2/3}{1-2/3} = 2.$$

Среднее число клиентов в очереди (длина очереди):

$$L_{\text{оч}} = \frac{\rho^2}{1-\rho} = \frac{(2/3)^2}{1-2/3} = 4/3 \approx 1{,}33.$$

Средняя продолжительность пребывания клиента в системе:

$$\bar{t}_{\text{СМО}} = \frac{L_{\text{СМО}}}{\lambda} = \frac{2}{4} = \frac{1}{2}\ \text{ч} = 30\ \text{мин}.$$

Средняя продолжительность пребывания клиента в очереди:

$$\bar{t}_{\text{оч}} = \frac{L_{\text{оч}}}{\lambda} = \frac{4/3}{4} = \frac{1}{3}\ \text{ч} = 20\ \text{мин}.$$

Выполняется равенство $\bar{t}_{\text{СМО}} = \bar{t}_{\text{оч}} + \bar{t}_{\text{об}}$. ▲

Упражнение 5.3. Магазин работает с одним продавцом. Предполагается, что простейший поток покупателей поступает с интенсивностью $\lambda = 20$ человек/ч. Время обслуживания заявки — случайная величина, которая подчиняется экспоненциальному закону распределения с параметром $\mu = 25$ человек/ч. Определить показатели эффективности работы магазина.

Ответ: $L_{\text{оч}} = 3{,}2$; $\bar{t}_{\text{оч}} = 9{,}6$ мин; $L_{\text{СМО}} = 4$; $\bar{t}_{\text{СМО}} = 12$ мин.

Многоканальная СМО с неограниченной очередью. СМО содержит n обслуживающих каналов. На вход поступает простейший поток заявок с интенсивностью λ. Если в момент поступления требования имеется свободный канал, то он немедленно приступает к обслуживанию поступившего требования. Каждый канал может одновременно обслужить только одно требование. Все каналы функционируют независимо. Если заявка застала все обслуживающие каналы занятыми, то она встает в очередь и ожидает начала обслуживания.

Время обслуживания заявки есть случайная величина, которая подчиняется экспоненциальному закону распределения с параметром μ. Среднее время обслуживания одной заявки $\bar{t}_{об} = 1/\mu$.

Возможные состояния СМО: S_0 (все каналы свободны), S_1 (один канал занят, остальные свободны), S_2 (два канала заняты, остальные свободны), …, S_n (все каналы заняты), S_{n+1} (все каналы заняты, в очереди одна заявка), S_{n+2} (все каналы заняты, в очереди две заявки) и т. д.

Приведенная интенсивность потока заявок $\rho = \lambda/\mu$. Размеченный граф состояний многоканальной СМО с неограниченной очередью приведен на рис. 5.7.

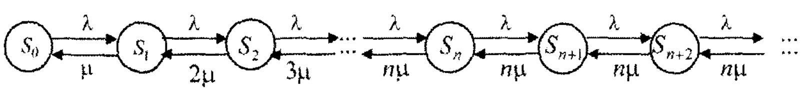


Рис. 5.7. Размеченный граф состояний многоканальной СМО с неограниченной очередью

Интенсивность потока обслуживаний (переводящего систему из одного состояния в другое справа налево) не остается постоянной, а по мере увеличения числа заявок в СМО от 0 до n увеличивается от величины μ до $n\mu$, так как соответственно увеличивается число каналов обслуживания. При числе заявок в СМО большем, чем n, интенсивность потока обслуживания сохраняется равной $n\mu$.

При $\rho/n \geq 1$ очередь растет до бесконечности. При $\rho/n < 1$ существуют предельные вероятности.

Используя формулы (5.1) и (5.2) для процесса гибели и размножения, получим формулы для предельных вероятностей состояний n-канальной СМО с неограниченной очередью:

$$p_0 = \left(\sum_{k=0}^{n} \frac{\rho^k}{k!} + \frac{\rho^{n+1}}{n!(n-\rho)} \right)^{-1},$$

$$p_1 = \frac{\rho}{1!} p_0, \ \ldots, \ p_k = \frac{\rho^k}{k!} p_0, \ldots, \ p_n = \frac{\rho^n}{n!} p_0,$$

$$p_{n+1} = \frac{\rho^{n+1}}{n \cdot n!} p_0, \ldots, \ p_{n+r} = \frac{\rho^{n+r}}{n^r n!} p_0, \ \ldots$$

Вероятность простоя каналов (все каналы свободны):

$$p_0 = \left(\sum_{k=0}^{n} \frac{\rho^k}{k!} + \frac{\rho^{n+1}}{(n-\rho)n!} \right)^{-1}.$$

Вероятность занятости обслуживанием k каналов:

$$p_k = \frac{\rho^k}{k!} p_0, \; (1 \le k \le n).$$

Вероятность занятости обслуживанием всех каналов при отсутствии очереди:

$$p_n = \frac{\rho^n}{n!} p_0.$$

Вероятность наличия очереди есть вероятность того, что число заявок в системе больше числа каналов:

$$P_{\text{оч}} = \frac{\rho^{n+1}}{(n-\rho)n!} p_0.$$

Вероятность для заявки попасть в очередь есть вероятность занятости всех каналов, которая равна сумме вероятностей наличия очереди и занятости всех n каналов при отсутствии очереди

$$P_{\text{зан}} = P_{\text{оч}} + p_n = \frac{\rho^n}{(n-\rho)(n-1)!} p_0.$$

Среднее число заявок в очереди (длина очереди):

$$L_{\text{оч}} = \frac{\rho^{n+1}}{(n-\rho)^2 (n-1)!} p_0.$$

Среднее число занятых обслуживанием каналов:

$$\bar{k} = \frac{\lambda}{\mu} = \rho.$$

Доля каналов, занятых обслуживанием:

$$q = \frac{\bar{k}}{n}.$$

Среднее число заявок в системе:

$$L_{\text{СМО}} = L_{\text{оч}} + \bar{k} = L_{\text{оч}} + \rho.$$

Среднее число свободных каналов:

$$\bar{n}_{\text{св}} = n - \bar{k} = n - \rho.$$

Среднее время пребывания заявки в очереди $\bar{t}_{\text{оч}}$ и среднее время пребывания заявки в системе $\bar{t}_{\text{СМО}}$ находятся по формулам Литтла:

$$\bar{t}_{\text{оч}} = \frac{L_{\text{оч}}}{\lambda},\ \bar{t}_{\text{СМО}} = \frac{L_{\text{СМО}}}{\lambda}.$$

Имеет место соотношение

$$\bar{t}_{\text{СМО}} = \bar{t}_{\text{оч}} + \bar{t}_{\text{об}}.$$

В силу неограниченности очереди при $\rho < 1$ любая заявка, пришедшая в систему, будет обслужена, т. е. вероятность отказа $P_{\text{отк}} = 0$, относительная пропускная способность $Q = 1$, а абсолютная пропускная способность равна интенсивности входящего потока заявок, т. е. $A = \lambda$.

Пример 5.10. К двум продавцам поступает на обслуживание поток покупателей с интенсивностью 220 чел./ч. Каждый из продавцов затрачивает на обслуживание одного покупателя в среднем 30 с. Определим среднее число покупателей в очереди и показатели занятости продавцов.

▼По условию $n = 2$, $\lambda = 220$ чел./ч, $\bar{t}_{\text{об}} = 30$ с.

Интенсивность обслуживания $\mu = 1/\bar{t}_{\text{об}} = 3600/30 = 120$ чел./ч, параметр нагрузки $\rho = \lambda/\mu = 220/120 = 11/6$.

Поскольку $\rho < n$ ($11/6 < 2$), очередь не будет расти до ∞.

Вероятность простоя обоих продавцов:

$$p_0 = \left(1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \frac{\rho^3}{(2-\rho)\cdot 2!}\right)^{-1} =$$

$$= \left(1 + \frac{11}{6} + \frac{11^2}{6^2 \cdot 2} + \frac{11^3}{6^3(2 - 11/6)}\right)^{-1} = \frac{1}{23} = 0{,}04.$$

Вероятность занятости одного из продавцов:

$$p_1 = \frac{\rho^1}{1!} p_0 = \frac{11}{6} \cdot \frac{1}{23} = \frac{11}{138} = 0{,}08\,.$$

Поскольку $p_0 + p_1 + p_2 = 1$, то вероятность занятости обоих продавцов:

$$P_{\text{зан}} = p_2 = 1 - p_0 - p_1 = 1 - 0{,}04 - 0{,}08 = 0{,}88\,.$$

Таким образом, 4 % рабочего времени оба продавца будут свободны, 8 % времени будет занят обслуживанием один из них, а 88 % времени будут заняты оба.

Средняя длина очереди:

$$L_{\text{оч}} = \frac{\rho^3}{(2-\rho)^2(2-1)!} p_0 = \frac{11^3}{6^3 \cdot (2 - 11/6)^2} \cdot \frac{1}{23} = 10 \text{ чел.} \ \blacktriangle$$

Пример 5.11. Сберкасса имеет трех контролеров-кассиров ($n = 3$) для обслуживания вкладчиков. Поток вкладчиков поступает в сберкассу с интенсивностью $\lambda = 30$ чел./ч. Средняя продолжительность обслуживания контролером-кассиром одного вкладчика $\overline{t}_{\text{об}} =$ $= 3$ мин. Определим характеристики сберкассы как объекта СМО.

▼Имеем $\lambda = 30$ чел./ч, $\overline{t}_{\text{об}} = 3$ мин.

Интенсивность обслуживания $\mu = 1 / \overline{t}_{\text{об}} = 60 / 3 = 20$ чел./ч.

Параметр нагрузки $\rho = \lambda / \mu = 30/20 = 1{,}5$.

Вероятность простоя контролеров-кассиров в течение рабочего дня:

$$p_0 = \left(1 + 1{,}5 + \frac{1{,}5^2}{2!} + \frac{1{,}5^3}{3!} + \frac{1{,}5^4}{3!(3 - 1{,}5)}\right)^{-1} =$$

$$= (1 + 1{,}5 + 1{,}125 + 0{,}5625 + 0{,}5625)^{-1} = (4{,}75)^{-1} = 0{,}210.$$

Вероятность застать всех контролеров-кассиров занятыми:

$$p_3 = \frac{\rho^3}{3!} p_0 = 0,5625 \cdot 0,210 = 0,118.$$

Вероятность наличия очереди:

$$P_{оч} = \frac{1,5^4}{(3-1,5)\cdot 3!} p_0 = 0,5625 \cdot 0,210 = 0,118.$$

Среднее число заявок в очереди:

$$L_{оч} = \frac{\rho^4}{(n-\rho)^2 (n-1)!} p_0 = \frac{1,5^4}{(3-1,5)^2 \cdot (3-1)!} \cdot 0,210 = 0,236.$$

Среднее время ожидания заявки в очереди:

$$\overline{t}_{оч} = L_{оч} / \lambda = 0,236 / 30 = 0,00786 \text{ ч} = 0,472 \text{ мин}.$$

Среднее время пребывания заявки в СМО:

$$\overline{t}_{СМО} = \overline{t}_{оч} + \overline{t}_{об} = 0,472 + 3 = 3,472 \text{ мин}.$$

Среднее число занятых обслуживанием каналов:

$$\overline{k} = \frac{\lambda}{\mu} = \rho = 1,5.$$

Среднее число свободных каналов:

$$\overline{n}_{св} = n - \overline{k} = n - \rho = 3 - 1,5 = 1,5.$$

Коэффициент занятости каналов обслуживания:

$$q = \overline{k} / n = \rho / n = 1,5 / 3 = 0,5.$$

Среднее число посетителей в сберкассе:

$$L_{СМО} = L_{оч} + \overline{k} = 0,236 + 1,5 = 1,736 \text{ чел}.$$

Таким образом, вероятность простоя контролеров-кассиров равна 21 % рабочего времени, вероятность для посетителей оказаться в очереди составляет 11,8 %, среднее число посетителей в очереди 0,236 чел., среднее время ожидания посетителями обслуживания 0,472 мин. ▲

Упражнение 5.4. По данным упражнения 5.3 определить показатели эффективности работы магазина, считать число продавцов $n = 2$.

Ответ: $L_{оч} = 0{,}152$; $\bar{t}_{оч} = 0{,}46$ мин.; $L_{СМО} = 0{,}952$; $\bar{t}_{СМО} = 2{,}86$ мин.

5.4. Средства массового обслуживания с ограниченной очередью

Одноканальная СМО с ограниченной очередью. СМО содержит один обслуживающий канал. На вход поступает простейший поток заявок с интенсивностью λ. Если заявка застала обслуживающий канал занятым, то она встает в очередь и ожидает начала обслуживания. Число мест в очереди ограничено и равно m. Если заявка застала обслуживающий канал занятым и в очереди нет свободных мест, то она покидает систему необслуженной.

Время обслуживания заявки есть случайная величина, которая подчиняется экспоненциальному закону распределения с параметром μ. Среднее время обслуживания одной заявки $\bar{t}_{об} = 1/\mu$, если $Q = 1$.

Возможные состояния СМО: S_0 (канал свободен), S_1 (канал занят, очереди нет), S_{1+1} (канал занят, в очереди одна заявка), S_{1+2} (канал занят, в очереди две заявки), ..., S_{1+m} (канал занят, в очереди m заявок).

Размеченный граф состояний одноканальной СМО с ограниченной очередью показан на рис. 5.8.

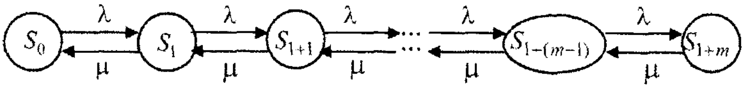


Рис. 5.8. Размеченный граф состояний одноканальной СМО с ограниченной очередью

Положим $\rho = \lambda/\mu$. Так как число состояний конечно, то существуют предельные вероятности.

Для вычисления предельных вероятностей состояний и показателей эффективности таких СМО может быть использован тот же подход, что и выше, с той разницей, что суммировать надо не бесконечную прогрессию, а конечную.

Предельные вероятности:

$$p_0 = \frac{1-\rho}{1-\rho^{m+2}}, \quad p_1 = \rho p_0,$$
$$p_2 = \rho^2 p_0, \ldots, \quad p_k = \rho^k p_0.$$

Вероятность отказа:

$$P_{\text{отк}} = p_{1+m} = \rho^{1+m} p_0.$$

Относительная пропускная способность:

$$Q = 1 - P_{\text{отк}} = 1 - \rho^{1+m} p_0.$$

Абсолютная пропускная способность:

$$A = \lambda Q = \lambda(1 - \rho^{1+m} p_0).$$

Среднее число заявок в очереди:

$$L_{\text{оч}} = \rho^2 \cdot \frac{1-\rho^m(m+1-m\rho)}{(1-\rho^{m+2})(1-\rho)}.$$

Среднее число заявок на обслуживании (среднее число занятых каналов):

$$L_{\text{об}} = 1 - p_0.$$

Среднее время пребывания заявки в очереди:

$$\bar{t}_{\text{оч}} = \frac{L_{\text{оч}}}{\lambda}.$$

Среднее время пребывания заявки в системе:

$$\overline{t}_{\text{СМО}} = \frac{L_{\text{СМО}}}{\lambda}.$$

Считаем $\overline{t}_{\text{об}} = \frac{1}{\mu}$, если $Q = 1$; иначе $\overline{t}_{\text{об}} = \frac{Q}{\mu}$. Тогда $\overline{t}_{\text{СМО}} = \overline{t}_{\text{оч}} + \overline{t}_{\text{об}}$.

Пример 5.12. Автозаправочная станция имеет $n = 1$ бензоколонку с площадкой, допускающей пребывание в очереди на заправку не более двух автомашин одновременно. Если в очереди находится $m = 2$ автомашины, то очередная прибывшая автомашина проезжает мимо автозаправочной станции. Предполагается, что простейший поток автомашин поступает на станцию с интенсивностью $\lambda = 10$ автомашин/ч. Время обслуживания заявки есть случайная величина, которая подчиняется экспоненциальному закону распределения с параметром $\mu = 12$ автомашин/ч. Определить параметры системы.

▼По условию $\lambda = 10$ автомашин/ч, $\mu = 12$ автомашин/ч, $m = 2$.

Имеем $\rho = \lambda/\mu = 10/12 \approx 0{,}833$.

Вероятность того, что на станции нет автомашин:

$$p_0 = \frac{1-\rho}{1-\rho^{m+2}} = \frac{1-0{,}833}{1-0{,}833^4} \approx 0{,}322.$$

Вероятность отказа:

$$P_{\text{отк}} = \rho^{1+m} p_0 = 0{,}833^3 \cdot 0{,}322 \approx 0{,}186.$$

Относительная пропускная способность:

$$Q = 1 - P_{\text{отк}} = 1 - 0{,}186 = 0{,}814.$$

Это вероятность того, что заявка будет обслужена.

Абсолютная пропускная способность:

$$A = \lambda Q = 10 \cdot 0{,}814 = 8{,}14 \text{ автомашин/ч}.$$

Среднее число заявок в очереди:

$$L_{\text{оч}} = \rho^2 \cdot \frac{1-\rho^m(m+1-m\rho)}{(1-\rho^{m+2})(1-\rho)} =$$

$$= 0{,}833^2 \cdot \frac{1-0{,}833^2(2+1-2\cdot 0{,}833)}{(1-0{,}833^4)(1-0{,}833)} = 0{,}596 \text{ автомашин,}$$

т. е. среднее число автомашин, ожидающих в очереди на заправку, равно 0,596.

Среднее число автомашин на обслуживании (среднее число занятых каналов):

$$L_{\text{об}} = 1 - p_0 = 1 - 0{,}322 = 0{,}678.$$

Среднее число автомашин на станции:

$$L_{\text{СМО}} = L_{\text{оч}} + L_{\text{об}} = 0{,}596 + 0{,}678 = 1{,}274.$$

Среднее время пребывания заявки в очереди

$$\overline{t}_{\text{оч}} = \frac{L_{\text{оч}}}{\lambda} = \frac{0{,}596}{10} = 0{,}0596 \text{ ч} = 0{,}0596 \cdot 60 = 3{,}576 \text{ мин.}$$

Среднее время обслуживания:

$$\overline{t}_{\text{об}} = \frac{Q}{\mu} = \frac{0{,}814}{12} = 0{,}0678 \text{ ч} = 0{,}0678 \cdot 60 = 4{,}07 \text{ мин.}$$

Среднее время пребывания заявки в системе:

$$\overline{t}_{\text{СМО}} = \frac{L_{\text{СМО}}}{\lambda} = \frac{1{,}274}{10} = 0{,}1274 \text{ ч} = 0{,}1274 \cdot 60 = 7{,}644 \text{ мин,}$$

причем $\overline{t}_{\text{СМО}} = \overline{t}_{\text{оч}} + \overline{t}_{\text{об}} = 3{,}576 + 4{,}07 = 7{,}644$. ▲

Упражнение 5.5. Автозаправочная станция имеет $n = 1$ бензоколонку с площадкой, допускающей пребывание в очереди на заправку не более трех автомашин одновременно. Если в очереди находится $m = 3$ автомашины, то очередная прибывшая автомашина проезжает мимо автозаправочной станции. Предполагается, что простейший поток автомашин поступает на станцию с интенсивностью $\lambda = 8$ автомашин/ч. Время обслуживания заявки есть случайная величина, которая подчиняется экспоненциальному закону распределения с параметром $\mu = 10$ автомашин/ч. Определить параметры системы.

Ответ: $Q = 0{,}878$; $A = 7{,}02$ автомашин/ч; $L_{оч} = 0{,}860$ автомашин; $L_{СМО} = 1{,}562$ автомашин; $\bar{t}_{оч} = 6{,}45$ мин; $\bar{t}_{СМО} = 11{,}72$ мин.

Многоканальная СМО с ограниченной очередью. СМО содержит n обслуживающих каналов. На вход поступает простейший поток заявок с интенсивностью λ. Если в момент поступления требования имеется свободный канал, то он немедленно приступает к обслуживанию требования. Каждый канал может одновременно обслуживать только одно требование. Все каналы функционируют независимо. Если заявка застала все обслуживающие каналы занятыми, то она встает в очередь и ожидает начала обслуживания. Число мест в очереди ограничено и равно m. Если заявка застала все обслуживающие каналы занятыми и в очереди нет свободных мест, то она покидает систему необслуженной.

Время обслуживания заявки есть случайная величина, которая подчиняется экспоненциальному закону распределения с параметром μ. Среднее время обслуживания одной заявки $\bar{t}_{об} = 1/\mu$, если $Q = 1$.

Возможные состояния СМО: S_0 (все каналы свободны), S_1 (один канал занят, остальные свободны), S_2 (два канала заняты, остальные свободны), ..., S_n (все каналы заняты), S_{n+1} (все каналы заняты, в очереди одна заявка), S_{n+2} (все каналы заняты, в очереди две заявки), ..., S_{n+m} (все каналы заняты, в очереди m заявок).

Размеченный граф состояний многоканальной СМО с ограниченной очередью представлен на рис. 5.9.

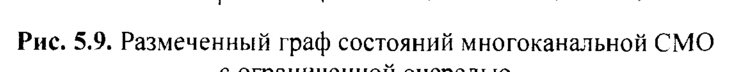


Рис. 5.9. Размеченный граф состояний многоканальной СМО с ограниченной очередью

Приведенная интенсивность потока заявок $\rho = \lambda/\mu$. Так как число состояний конечно, то существуют предельные вероятности.

Предельные вероятности:

$$p_0 = \left(\sum_{k=0}^{n} \frac{\rho^k}{k!} + \frac{\rho^{n+1}\left(1-(\rho/n)^m\right)}{n \cdot n!\left(1-\rho/n\right)} \right)^{-1}, \; p_1 = \frac{\rho}{1!} p_0, \ldots, \; p_n = \frac{\rho^n}{n!} p_0,$$

$$p_{n+1} = \frac{\rho^{n+1}}{n \cdot n!} p_0, \ldots, \; p_{n+k} = \frac{\rho^{n+k}}{n^k \cdot n!} p_0.$$

$$k = 1, \ldots, m.$$

Вероятность отказа:

$$P_{\text{отк}} = p_{n+m} = \frac{\rho^{n+m}}{n^m \cdot n!} p_0.$$

Относительная пропускная способность:

$$Q = 1 - P_{\text{отк}} = 1 - \frac{\rho^{n+m}}{n^m \cdot n!} p_0.$$

Абсолютная пропускная способность:

$$A = \lambda Q = \lambda\left(1 - \frac{\rho^{n+m}}{n^m \cdot n!} p_0\right).$$

Среднее число заявок в очереди:

$$L_{\text{оч}} = \frac{\rho^{n+1} p_0 \left[1 - \left(m + 1 - m \cdot \frac{\rho}{n}\right)\left(\frac{\rho}{n}\right)^m\right]}{n \cdot n!\left(1 - \frac{\rho}{n}\right)^2}.$$

Среднее число заявок на обслуживании (среднее число занятых каналов):

$$L_{\text{об}} = \rho Q.$$

Среднее число заявок в системе:

$$L_{\text{СМО}} = L_{\text{оч}} + L_{\text{об}}.$$

Среднее время пребывания заявки в очереди:

$$\overline{t}_{\text{оч}} = \frac{L_{\text{оч}}}{\lambda}.$$

Среднее время пребывания заявки в системе:

$$\overline{t}_{\text{СМО}} = \frac{L_{\text{СМО}}}{\lambda}.$$

Считаем $\overline{t}_{\text{об}} = \dfrac{1}{\mu}$, если $Q = 1$; иначе $\overline{t}_{\text{об}} = \dfrac{Q}{\mu}$. Тогда $\overline{t}_{\text{СМО}} = \overline{t}_{\text{оч}} + \overline{t}_{\text{об}}$.

Пример 5.13. Автозаправочная станция имеет $n = 2$ бензоколонки с площадкой, допускающей пребывание в очереди на заправку не более трех автомашин одновременно. Если в очереди находится $m = 3$ автомашины, то очередная прибывшая автомашина проезжает мимо автозаправочной станции. Предполагается, что простейший поток автомашин поступает на станцию с интенсивностью $\lambda = 10$ автомашин/ч. Время обслуживания заявки есть случайная величина, которая подчиняется экспоненциальному закону распределения с параметром $\mu = 12$ автомашин/ч. Определить параметры системы.

▼По условию $\lambda = 10$ автомашин/ч, $\mu = 12$ автомашин/ч, $n = 2$, $m = 3$.

Имеем $\rho = \lambda/\mu = 10/12 \approx 0{,}833$, $\rho/n = 0{,}833/2 = 0{,}417$.

Вероятность того, что на станции нет автомашин:

$$p_0 = \left(1 + 0{,}833 + \frac{0{,}833^2}{2!} + \frac{0{,}833^3(1-0{,}417^3)}{2\cdot 2!\cdot(1-0{,}417)}\right)^{-1} = (2{,}41)^{-1} = 0{,}415.$$

Вероятность отказа (обе бензоколонки заняты, в очереди нет свободных мест):

$$P_{\text{отк}} = p_{n+m} = \frac{\rho^{n+m}}{n^m \cdot n!} p_0 = \frac{0{,}833^5}{2^3 \cdot 2!} \cdot 0{,}415 = 0{,}010.$$

Относительная пропускная способность:

$$Q = 1 - P_{\text{отк}} = 1 - 0{,}01 = 0{,}99.$$

Это вероятность того, что заявка будет обслужена.
Абсолютная пропускная способность:

$$A = \lambda Q = 10 \cdot 0{,}99 = 9{,}9 \text{ автомашин/ч.}$$

Среднее число заявок в очереди:

$$L_{\text{оч}} = \frac{\rho^{n+1} p_0 \left[1 - \left(m + 1 - m \cdot \frac{\rho}{n}\right)\left(\frac{\rho}{n}\right)^m\right]}{n \cdot n! \left(1 - \frac{\rho}{n}\right)^2} =$$

$$= \frac{0{,}833^3 \cdot 0{,}415\left[1 - (3 + 1 - 3 \cdot 0{,}417)0{,}417^3\right]}{2 \cdot 2! \cdot (1 - 0{,}417)^2} = 0{,}141.$$

Среднее число автомашин на обслуживании:

$$L_{\text{об}} = \rho Q = 0{,}833 \cdot 0{,}99 = 0{,}825.$$

Среднее число автомашин на станции:

$$L_{\text{СМО}} = L_{\text{оч}} + L_{\text{об}} = 0{,}141 + 0{,}825 = 0{,}966.$$

Среднее время пребывания заявки в очереди:

$$\bar{t}_{\text{оч}} = \frac{L_{\text{оч}}}{\lambda} = \frac{0{,}141}{10} = 0{,}0141 \text{ ч} = 0{,}0141 \cdot 60 = 0{,}85 \text{ мин.}$$

Среднее время обслуживания:

$$\bar{t}_{\text{об}} = \frac{Q}{\mu} = \frac{0{,}99}{12} = 0{,}0825 \text{ ч} = 0{,}0825 \cdot 60 = 4{,}95 \text{ мин.}$$

Среднее время пребывания заявки в системе:

$$\overline{t}_{\text{СМО}} = \frac{L_{\text{СМО}}}{\lambda} = \frac{0,966}{10} = 0,0966 \text{ ч} = 0,0966 \cdot 60 = 5,8 \text{ мин},$$

причем $\overline{t}_{\text{СМО}} = \overline{t}_{\text{оч}} + \overline{t}_{\text{об}} = 0,85 + 4,95 = 5,8$ мин. ▲

Пример 5.14. В парикмахерской работают 3 мастера, а в зале ожидания расположены 3 стула. Поток клиентов имеет интенсивность $\lambda = 12$ клиентов в час. Время обслуживания клиента есть случайная величина, которая подчиняется экспоненциальному закону распределения с параметром $\mu = 3$ клиента в час. Определим относительную и абсолютную пропускную способность системы, среднее число занятых кресел, среднюю длину очереди, среднее время, которое клиент проводит в парикмахерской.

▼По условию $n = 3$, $m = 3$, $\lambda = 12$ клиентов в час, $\mu = 3$ клиента в час.

Имеем $\rho = \lambda/\mu = 12/3 = 4$, $\rho/n = 4/3$.

Вероятность простоя трех мастеров:

$$p_0 = \left(1 + 4 + \frac{4^2}{2!} + \frac{4^3}{3!} + \frac{4^{3+1}}{3!(3-4)}\left(1 - \left(\frac{4}{3}\right)^3\right)\right)^{-1} = 0,012 .$$

Вероятность отказа в обслуживании:

$$P_{\text{отк}} = pn + m = \frac{\rho^{n+m}}{n^m n!} p_0 = \frac{4^{3+3}}{3^3 \cdot 3!} \cdot 0,012 = 0,307 .$$

Вероятность обслуживания (относительная пропускная способность):

$$Q = P_{\text{об}} = 1 - P_{\text{отк}} = 1 - 0,307 = 0,693.$$

Абсолютная пропускная способность:

$$A = \lambda Q = 12 \cdot 0,693 = 8,32.$$

Среднее число занятых каналов:

$$L_{\text{об}} = \rho Q = 4 \cdot 0,693 = 2,77.$$

Среднее время обслуживания:

$$\bar{t}_{\text{об}} = \frac{Q}{\mu} = \frac{0,693}{3} = 0,23 \text{ ч.}$$

Средняя длина очереди (среднее число заявок в очереди):

$$L_{\text{оч}} = \frac{\rho^{n+1} \cdot p_0 \left[1 - \left(m + 1 - m \cdot \rho / n\right)\left(\rho / n\right)^m\right]}{n \cdot n!\left(1 - \rho / n\right)^2} =$$

$$= \frac{4^4 \cdot 0,012 \cdot \left[1 - (3 + 1 - 3 \cdot 4 / 3)(4 / 3)^3\right]}{3 \cdot 3!(1 - 4 / 3)^2} = 1,56.$$

Среднее время ожидания обслуживания в очереди:

$$\bar{t}_{\text{оч}} = L_{\text{оч}} / \lambda = 1,56 / 12 = 0,13 \text{ ч.}$$

Среднее число заявок в СМО:

$$L_{\text{СМО}} = L_{\text{оч}} + L_{\text{об}} = 1,56 + 2,77 = 4,33.$$

Среднее время пребывания заявки в СМО:

$$\bar{t}_{\text{СМО}} = L_{\text{СМО}} / \lambda = 4,33 / 12 = 0,36 \text{ ч,}$$

причем $\bar{t}_{\text{СМО}} = \bar{t}_{\text{оч}} + \bar{t}_{\text{об}} = 0,13 + 0,23 = 0,36$ ч. ▲

5.5. Замкнутые средства массового обслуживания

До сих пор рассматривались системы, в которых входящий поток никак не связан с выходящим. Такие системы называются *разомкнутыми*. В некоторых же случаях обслуженные требования после задержки опять поступают на вход. Такие СМО называются *замкнутыми*.

В замкнутых СМО интенсивность потока заявок зависит от состояния СМО: чем больше заявок находится в СМО, тем меньше интенсивность входного потока.

Например, если рассматривать как СМО бригаду рабочих, обслуживающих станки в некотором цехе, то такая СМО является замкнутой. Действительно, если станок испортился и в данный момент либо ремонтируется, либо ожидает в очереди на ремонт, то он перестает быть источником заявок. Поскольку количество станков в цехе ограничено, то выход из строя одного или нескольких станков существенно сказывается на интенсивности потока заявок на обслуживание.

Зависимость потока заявок от состояний СМО проявляется существенно при *конечном небольшом* числе источников заявок. Но если число источников достаточно велико, то практически можно считать, что интенсивность потока заявок не зависит от состояний СМО.

Одноканальная замкнутая СМО. Пусть одноканальная СМО содержит m источников заявок, каждый из которых порождает простейший поток заявок с интенсивностью λ. Заявка, пришедшая от источника в момент, когда канал занят, становится в очередь и ждет обслуживания. При этом источник может подать следующую заявку только в том случае, если поданная им предыдущая заявка уже обслужена. Среднее время обслуживания каналом одной заявки (безразлично из каких источников) $\bar{t}_{об} = 1/\mu$, где μ — интенсивность простейшего потока обслуживания.

Таким образом, имеем замкнутую СМО, содержащую конечное число источников заявок, каждый из которых может находиться в одном из двух состояний: активном или пассивном. *Активное состояние* источника — это такое состояние, при котором уже обслужена поданная им последняя заявка. *Пассивное состояние* характеризуется тем, что поданная источником последняя заявка еще не обслужена, т. е. либо стоит в очереди, либо находится на обслуживании.

В активном состоянии источник может подавать заявки, а в пассивном — нет. Следовательно, интенсивность общего потока заявок зависит от того, сколько источников находится в пассивном состоянии, т. е. сколько заявок связано с процессом обслуживания (стоит в очереди или непосредственно обслуживается).

Пример 5.15. Один рабочий (n = 1) обслуживает 4 станка (m = 4). Предполагается, что поломки станков образуют простейший поток заявок с интенсивностью λ = 0,5 раз/ч. Время ремонта каждого станка есть случайная величина, которая подчиняется экспоненциальному закону распределения с параметром $\mu = 1,25$ станка/ч. Определить параметры системы.

▼Это замкнутая СМО. Занумеруем состояния СМО по числу источников (станков), находящихся в пассивном состоянии, т. е. по числу заявок, находящихся в очереди и на обслуживании: S_0 (все четыре источника находятся в активном состоянии, канал (рабочий) свободен, очереди нет), S_1 (один источник находится в пассивном состоянии, канал обслуживает поданную этим источником заявку, очереди нет), S_2 (два источника находятся в пассивном состоянии, заявка, поданная одним из них, обслуживается, а заявка, поданная другим источником, стоит в очереди), S_3 (три источника находятся в пассивном состоянии, заявка, поданная одним из них, обслуживается, а две заявки, поданные другими источниками, стоят в очереди), S_4 (все четыре источника находятся в пассивном состоянии, заявка, поданная одним из них, обслуживается, а три заявки, поданные остальными источниками, стоят в очереди).

Размеченный граф состояний приведен на рис. 5.10.

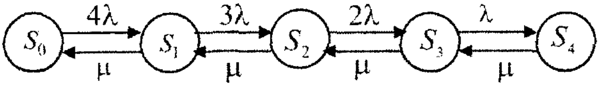


Рис. 5.10. Размеченный граф состояний для замкнутой одноканальной СМО ($m = 4$)

По условию $\lambda = 0,5$ раз/ч, $\mu = 1,25$ станка/ч, $m = 4$. Приведенная интенсивность потока заявок $\rho = \frac{\lambda}{\mu} = \frac{0,5}{1,25} = 0,4$.

Так как число возможных состояний конечно, существуют предельные вероятности.

Вероятность, что все станки исправны (предельная вероятность состояния S_0), равна p_0.

Состояние S_0 связано с состоянием S_1 двумя стрелками с интенсивностями 4λ и μ. Предельная вероятность p_1 состояния S_1 равна $p_1 = \frac{4\lambda}{\mu} p_0$ (интенсивность верхней стрелки пишем в числителе, интенсивность нижней стрелки пишем в знаменателе) = $= 4\rho p_0 = 4 \cdot 0,4 p_0 = 1,6 p_0$.

Аналогично $p_2 = \frac{3\lambda}{\mu} p_1 = 3\rho p_1 = 3 \cdot 0{,}4 \cdot 1{,}6p_0 = 1{,}92p_0$; $p_3 = \frac{2\lambda}{\mu} p_2 = 2\rho p_2 = 2 \cdot 0{,}4 \cdot 1{,}92p_0 = 1{,}536p_0$; $p_4 = \frac{\lambda}{\mu} p_3 = \rho p_3 = 0{,}4 \cdot 1{,}536p_0 = 0{,}614p_0$. Поскольку $1 = p_0 + p_1 + p_2 + p_3 + p_4$, то $1 = p_0 + 1{,}6p_0 + 1{,}92p_0 + 1{,}536p_0 + 0{,}614p_0 = 6{,}67p_0$.

Следовательно, $p_0 = 1/6{,}67 \approx 0{,}15$; $p_1 = 1{,}6p_0 = 1{,}6 \cdot 0{,}15 = 0{,}24$; $p_2 = 1{,}92p_0 = 1{,}92 \cdot 0{,}15 = 0{,}288$; $p_3 = 1{,}536p_0 = 1{,}536 \cdot 0{,}15 = 0{,}230$; $p_4 = 0{,}614p_0 = 0{,}614 \cdot 0{,}15 = 0{,}092$.

Вероятность занятости рабочего $P_{зан} = 1 - p_0 = 0{,}85$.

Абсолютная пропускная способность (среднее число неисправностей, которое рабочий ликвидирует в час) $A = P_{зан}\mu = 0{,}85 \cdot 1{,}25 = 1{,}062$ станка/ч.

Относительная пропускная способность $Q = 1$.

Средняя длина очереди $L_{оч} = 1 \cdot p_2 + 2 \cdot p_3 + 3 \cdot p_4 = 1 \cdot 0{,}288 + 2 \cdot 0{,}230 + 3 \cdot 0{,}092 = 1{,}024$.

Среднее число заявок, находящихся на обслуживании $L_{об} = 0 \cdot p_0 + 1 \cdot (1 - p_0) = 1 - p_0 = 0{,}85$.

Среднее число заявок в системе (т. е. среднее число источников, находящихся в пассивном состоянии) $L_{СМО} = 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 + 4 \cdot p_4 = 1 \cdot 0{,}24 + 2 \cdot 0{,}288 + 3 \cdot 0{,}230 + 4 \cdot 0{,}092 = 1{,}874$, причем $L_{СМО} = L_{оч} + L_{об}$. ▲

Многоканальная замкнутая СМО. Пусть СМО содержит n обслуживающих каналов. На вход поступает простейший поток заявок с интенсивностью λ. Если в момент поступления требования имеется свободный канал, то он немедленно приступает к обслуживанию поступившего требования. Каждый канал может одновременно обслуживать только одно требование. Все каналы функционируют независимо. Если заявка застала все обслуживающие каналы занятыми, то она встает в очередь и ожидает начала обслуживания. Требования на обслуживания поступают от m обслуживаемых объектов, т. е. поток поступающих требований ограничен.

Время обслуживания заявки есть случайная величина, которая подчиняется экспоненциальному закону распределения с параметром μ. Среднее время обслуживания одной заявки $\overline{t}_{об} = 1/\mu$.

Пример 5.16. Бригада рабочих из 2 человек ($n = 2$) обслуживает 4 станка ($m = 4$). Предполагается, что поломки станков образуют простейший поток заявок с интенсивностью $\lambda = 0{,}1$ раз/ч. Время ремонта каждого станка есть случайная величина, которая подчиняется экспоненциальному закону распределения с параметром $\mu = 0{,}5$ станка/ч. *Определим параметры системы.*

▼Это замкнутая СМО. Занумеруем состояния СМО по числу станков, находящихся в пассивном состоянии, т. е. по числу заявок, находящихся в очереди и на обслуживаниим: S_0 (все станки работают, рабочие не заняты, очереди нет), S_1 (один станок остановился, один рабочий занят, очереди нет), S_2 (два станка остановились, два рабочих заняты, очереди нет), S_3 (три станка остановились, два рабочих заняты, одна заявка стоит в очереди), S_4 (все четыре станка остановились, два рабочих заняты, две заявки стоят в очереди).

Размеченный граф состояний приведен на рис. 5.11

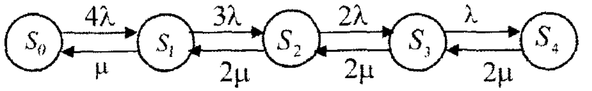


Рис. 5.11. Размеченный граф состояний для замкнутой многоканальной СМО ($n = 2$, $m = 4$)

По условию $\lambda = 0{,}1$ раз/ч, $\mu = 0{,}25$ станка/ч, $m = 4$. Приведенная интенсивность потока заявок $\rho = \dfrac{\lambda}{\mu} = \dfrac{0{,}1}{0{,}5} = 0{,}2$.

Так как число возможных состояний конечно, существуют предельные вероятности. Вероятность, что все станки исправны (предельная вероятность состояния S_0), равна p_0.

Состояние S_0 связано с состоянием S_1 двумя стрелками с интенсивностями 4λ и μ. Предельная вероятность p_1 состояния S_1 равна $p_1 = \dfrac{4\lambda}{\mu} p_0$ (интенсивность верхней стрелки пишем в числителе, интенсивность нижней стрелки — в знаменателе) $= 4\rho p_0 = 4 \cdot 0{,}2 p_0 = 0{,}8 p_0$.

Аналогично $p_2 = \dfrac{3\lambda}{2\mu} p_1 = 1{,}5\rho p_1 = 1{,}5 \cdot 0{,}2 \cdot 0{,}8 p_0 = 0{,}24 p_0$;
$p_3 = \dfrac{2\lambda}{2\mu} p_2 = \rho p_2 = 0{,}2 \cdot 0{,}24 p_0 = 0{,}048 p_0$; $p_4 = \dfrac{\lambda}{2\mu} p_3 = 0{,}5\rho p_3 =$

$= 0{,}5 \cdot 0{,}2 \cdot 0{,}048p_0 = 0{,}0048p_0$. Поскольку $1 = p_0 + p_1 + p_2 + p_3 + p_4$, то $1 = p_0 + 0{,}8p_0 + 0{,}24p_0 + 0{,}048p_0 + 0{,}0048p_0 = 2{,}0928p_0$.

Следовательно, $p_0 = 1/2{,}0928 \approx 0{,}478$; $p_1 = 0{,}8p_0 = 0{,}8 \cdot 0{,}478 = 0{,}382$; $p_2 = 0{,}24p_0 = 0{,}24 \cdot 0{,}478 = 0{,}115$; $p_3 = 0{,}048p_0 = 0{,}048 \cdot 0{,}478 = 0{,}023$; $p_4 = 0{,}0048p_0 = 0{,}0048 \cdot 0{,}478 = 0{,}002$.

Средняя длина очереди $L_{оч} = 1 \cdot p_3 + 2 \cdot p_4 = 1 \cdot 0{,}023 + 2 \cdot 0{,}002 = 0{,}027$.

Среднее число заявок в системе $L_{СМО} = 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 + 4 \cdot p_4 = 1 \cdot 0{,}382 + 2 \cdot 0{,}115 + 3 \cdot 0{,}023 + 4 \cdot 0{,}002 = 0{,}689$.

Среднее число свободных от обслуживания рабочих $\overline{n}_{св} = 2 \cdot p_0 + 1 \cdot p_1 = 2 \cdot 0{,}478 + 1 \cdot 0{,}382 = 1{,}338$.

Среднее число заявок, находящихся на обслуживании $L_{об} = n - \overline{n}_{св} = 2 - 1{,}338 = 0{,}662$. ▲

5.6. Средства массового обслуживания с ограниченным временем ожидания

До сих пор мы рассматривали СМО с ожиданием, ограниченным только длиной очереди (числом m заявок, одновременно находящихся в очереди). В такой СМО заявка, раз ставшая в очередь, уже не покидает ее и «терпеливо» дожидается обслуживания. На практике нередко встречаются СМО другого типа, в которых заявка, подождав некоторое время, может уйти из очереди (так называемые нетерпеливые заявки).

Предполагается, что заявка может находиться в очереди случайное время, которое подчиняется экспоненциальному закону распределения с параметром ν, то есть ν — среднее количество требований, покидающих очередь в единицу, вызванное появлением в очереди одного требования.

СМО содержит n обслуживающих каналов. На вход поступает простейший поток заявок с интенсивностью λ. Если в момент поступления требования имеется свободный канал, то он немедленно приступает к обслуживанию. Каждый канал может одновременно обслуживать только одно требование. Все каналы функционируют независимо. Если заявка застала все обслуживающие каналы занятыми, то она встает в очередь и ожидает начала обслуживания.

Время обслуживания заявки есть случайная величина, подчиняющая экспоненциальному закону распределения с параметром μ.

Рассмотрим СМО подобного типа, оставаясь в рамках марковской схемы.

Пример 5.17. В пункте химчистки имеется $n = 2$ аппарата для чистки. Поток посетителей предполагается простейшим с интенсивностью $\lambda = 5$ человек/ч. Время обслуживания каждого посетителя есть случайная величина, подчиненная экспоненциальному закону распределения с параметром $\mu = 4$ человека/ч. Среднее число посетителей, покидающих очередь, не дождавшись обслуживания, равно $\nu = 3$ человека/ч. Определим параметры системы.

▼Будем снова нумеровать состояния системы по числу заявок, связанных с системой — как обслуживаемых, так и стоящих в очереди: S_0 (все каналы свободны, очереди нет), S_1 (один канал занят, очереди нет), S_2 (два канала заняты, очереди нет), S_3 (два канала заняты, в очереди одна заявка), S_4 (два канала заняты, в очереди две заявки), S_5 (два канала заняты, в очереди три заявки) и т. д. Размеченный граф состояний этой многоканальной СМО приведен на рис. 5.12.

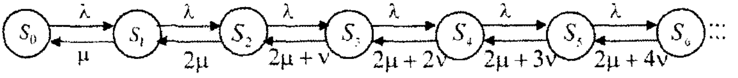


Рис. 5.12. Размеченный граф состояний многоканальной СМО с ограниченным временем ожидания ($n = 2$)

По условию $n = 2$, $\lambda = 5$ человек/ч, $\mu = 4$ человека/ч, $\nu = 3$ человека/ч.

Вероятность, что оба аппарата свободны (предельная вероятность состояния S_0) равна p_0.

Состояние S_1 связано с состоянием S_0 двумя стрелками с интенсивностями λ и μ. Предельная вероятность состояния S_1 есть $p_1 = \dfrac{\lambda}{\mu} p_0$ (интенсивность верхней стрелки пишем в числитель, интенсивность нижней стрелки — в знаменатель) $= \dfrac{5}{4} p_0 = 1{,}25 p_0$.

Аналогично $p_2 = \frac{\lambda}{2\mu} p_1 = \frac{5}{2 \cdot 4} \cdot 1{,}25 p_0 = 0{,}781 p_0$;

$$p_3 = \frac{\lambda}{2\mu + \nu} p_2 = \frac{5}{2 \cdot 4 + 3} \cdot 0{,}781 p_0 = 0{,}355 p_0;$$

$$p_4 = \frac{\lambda}{2\mu + 2\nu} p_3 = \frac{5}{2 \cdot 4 + 2 \cdot 3} \cdot 0{,}355 p_0 = 0{,}127 p_0;$$

$$p_5 = \frac{\lambda}{2\mu + 3\nu} p_4 = \frac{5}{2 \cdot 4 + 3 \cdot 3} \cdot 0{,}127 p_0 = 0{,}037 p_0;$$

$$p_6 = \frac{\lambda}{2\mu + 4\nu} p_5 = \frac{5}{2 \cdot 4 + 4 \cdot 3} \cdot 0{,}037 p_0 = 0{,}009 p_0;$$

$$p_7 = \frac{\lambda}{2\mu + 5\nu} p_6 = \frac{5}{2 \cdot 4 + 5 \cdot 3} \cdot 0{,}009 p_0 = 0{,}002 p_0;$$

$$p_8 = \frac{\lambda}{2\mu + 6\nu} p_7 = \frac{5}{2 \cdot 4 + 6 \cdot 3} \cdot 0{,}002 p_0 \approx 0{,}000 p_0.$$

Поэтому $1 = p_0 + p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 + \ldots \approx p_0 + 1{,}25p_0 + 0{,}781p_0 + 0{,}355p_0 + 0{,}127p_0 + 0{,}037p_0 + 0{,}009p_0 + 0{,}002p_0 = 3{,}561p_0$.

Следовательно, $p_0 = 1/3{,}561 = 0{,}281$;

$p_1 = 1{,}25p_0 = 1{,}25 \cdot 0{,}281 = 0{,}351$;

$p_2 = 0{,}781p_0 = 0{,}781 \cdot 0{,}281 = 0{,}219$;

$p_3 = 0{,}355p_0 = 0{,}355 \cdot 0{,}281 = 0{,}100$;

$p_4 = 0{,}127p_0 = 0{,}127 \cdot 0{,}281 = 0{,}036$;

$p_5 = 0{,}037p_0 = 0{,}037 \cdot 0{,}281 = 0{,}010$;

$p_6 = 0{,}009p_0 = 0{,}009 \cdot 0{,}281 = 0{,}003$;

$p_7 = 0{,}002p_0 = 0{,}002 \cdot 0{,}281 = 0{,}000$.

Остальные предельные вероятности также полагаются равными нулю.

Средняя длина очереди — это математическое ожидание числа посетителей, находящихся в очереди: $L_{оч} \approx 1 \cdot p_3 + 2 \cdot p_4 + 3 \cdot p_5 + 4 \cdot p_6 = 1 \cdot 0{,}100 + 2 \cdot 0{,}036 + 3 \cdot 0{,}010 + 4 \cdot 0{,}003 = 0{,}214$.

Среднее число заявок в системе — это математическое ожидание числа заявок в системе: $L_{СМО} \approx 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 + 4 \cdot p_4 + 5 \cdot p_5 + 6 \cdot p_6 = 1 \cdot 0{,}351 + 2 \cdot 0{,}219 + 3 \cdot 0{,}100 + 4 \cdot 0{,}036 + 5 \cdot 0{,}010 + 6 \cdot 0{,}003 = 1{,}301$.

Некоторые посетители, не дождавшись обслуживания, уходят из очереди с интенсивностью ν. Поэтому без обслуживания систему покидают в среднем $\nu L_{\text{СМО}}$ человек/ч, то есть из λ посетителей будет обслужено лишь $A = \lambda - \nu L_{\text{СМО}} = 5 - 3 \cdot 1{,}301 = 1{,}097$ человек/ч. Это абсолютная пропускная способность. Относительная пропускная способность $Q = A/\lambda = 1{,}097/5 = 0{,}2194$.

Среднее число занятых аппаратов $\bar{n}_{\text{зан}} = A/\mu = 1{,}097/4 = 0{,}274$. ▲

Глава 6. ДИНАМИЧЕСКОЕ ПРОГРАММИРОВАНИЕ

6.1. Постановка задачи динамического программирования

Динамическое программирование (ДП) — метод оптимизации, приспособленный к операциям, в которых процесс принятия решения может быть разбит на этапы (шаги). Такие операции называются многошаговыми.

Рассматривается управляемый процесс, например экономический процесс распределения средств между предприятиями, замены оборудования и т. п. В результате управления система (объект управления) S переводится из начального состояния S_0 в конечное S_n. Предположим, что управление можно разбить на n шагов, т. е. решение принимается последовательно на каждом шаге, а управление, переводящее систему S из начального состояния в конечное, представляет собой совокупность n пошаговых управлений.

На каждом шаге необходимо определить два типа переменных: переменную состояния системы S_k и переменную управления x_k. Переменная S_k определяет, в каких состояниях может оказаться система на рассматриваемом k-м шаге. В зависимости от состояния S на этом шаге можно применить некоторые управления, которые характеризуются переменной x_k, удовлетворяют определенным ограничениям и называются *допустимыми*.

Пусть $X = (x_1, x_2, \ldots, x_k, \ldots, x_n)$ — управление, переводящее систему из состояния S_0 в состояние S_n, а S_k — состояние системы после k-го шага управления. Тогда последовательность состояний системы можно представить в виде графа (рис. 6.1).

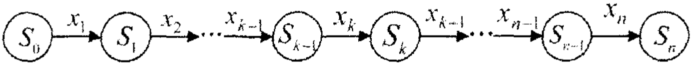


Рис. 6.1. Граф состояний системы управления

Показатель эффективности рассматриваемой управляемой операции (целевая функция) зависит от начального состояния и управления:

$$Z = F(S_0, X).$$

Сделаем ряд предположений.

1. Состояние S_k системы в конце k-го шага зависит только от предшествующего состояния S_{k-1} и управления x_k на k-м шаге (и не зависит от предшествующих состояний и управлений). Это требование называется «отсутствием последействия». Сформулированное положение записывается в виде уравнений

$$S_k = S_k(S_{k-1}, x_k),\ k = 1, 2, \ldots, n,$$

которые называются *уравнениями состояний.*

2. Целевая функция является аддитивной от показателя эффективности каждого шага:

$$Z = \sum_{k=1}^{n} F_k(S_{k-1}, x_k),$$

где $Z_k = F_k(S_{k-1}, x_k)$ — показатель эффективности на k-м шаге.

Задача пошаговой оптимизации (задача ДП) формулируется так: определить такое допустимое управление X, переводящее систему за n шагов из начального состояния S_0 в конечное состояние S_n, чтобы достичь экстремума (максимума или минимума в зависимости от вида решаемой задачи) целевой функции Z.

Особенности модели ДП следующие:

• задача оптимизации формулируется как конечный многошаговый процесс управления;

• целевая функция равна сумме целевых функций каждого шага;

• выбор управления x_k на каждом шаге зависит только от состояния системы к этому шагу S_{k-1} и не влияет на предшествующие шаги (нет обратной связи);

• состояние системы S_k после каждого шага управления зависит только от предшествующего состояния S_{k-1} и этого управляющего воздействия x_k (отсутствие последействия) и может быть записано в виде уравнения состояния;

• на каждом шаге управление x_k зависит от конечного числа управляющих переменных, а состояние S_k зависит от конечного числа параметров.

Принцип оптимальности Беллмана. Каково бы ни было состояние S системы в результате какого-либо числа шагов, на ближайшем шаге нужно выбирать управление так, чтобы оно в совокупности с оптимальным управлением на всех последующих шагах приводило к оптимальному выигрышу на всех оставшихся шагах, включая данный.

Функциональные уравнения Беллмана. Функциональные уравнения Беллмана — это математическая формулировка принципа оптимальности Беллмана. Из принципа оптимальности следует, что оптимальную стратегию управления можно получить, если сначала найти оптимальную стратегию управления на n-м шаге, затем на двух последних шагах, затем на трех последних шагах и т. д., вплоть до первого шага. Таким образом, решение рассматриваемой задачи ДП целесообразно начинать с определения оптимального решения на последнем, n-м шаге. Будем считать, что ищем максимум целевой функции.

На n-м шаге $Z_n(S_{n-1}) = \max\limits_{x_n} F_n\left(S_{n-1}, x_n\right)$, т. е. выигрыш на n-м шаге зависит только от выбора управления x_n на этом шаге.

На (n − 1)-м шаге $Z_{n-1}(S_{n-2}) = \max\limits_{x_{n-1}}\left\{F_{n-1}(S_{n-2}, x_{n-1}) + Z_n(S_{n-1})\right\} =$

$= \max\limits_{x_{n-1}}\left\{F_{n-1}(S_{n-2}, x_{n-1}) + Z_n(S_{n-1}(S_{n-2}, x_{n-1}))\right\}$, т. е. на $(n-1)$-м шаге надо так подобрать управление x_{n-1}, чтобы сумма выигрышей на $(n-1)$-м шаге $(F_{n-1}(S_{n-2}, x_{n-1}))$ и на n-м шаге $(Z_n(S_{n-1}(S_{n-2}, x_{n-1})))$ была максимальна, и т.д.

На k-м шаге $Z_k(S_{k-1}) = \max\limits_{x_k}\{F_k(S_{k-1}, x_k) + Z_{k+1}(S_k)\} =$

$= \max\limits_{x_k}\{F_k(S_{k-1}, x_k) + Z_{k+1}(S_k(S_{k-1}, x_k))\}$ $(k = n-1, n-2, \dots, 2, 1)$, т. е.

на k-м шаге надо так подобрать управление x_k, чтобы сумма выигрышей на k-м шаге $(F_k(S_{k-1}, x_k))$ и на $n-k$ последующих шагах $(Z_{k+1}(S_k(S_{k-1}, x_k)))$ была максимальна.

Последовательно осуществляя описанный выше итерационный процесс, дойдем, наконец, до первого шага. На этом шаге известно, в каком состоянии может находиться система.

Чтобы найти оптимальную стратегию управления, т. е. определить искомое решение задачи, нужно теперь пройти всю последовательность шагов, только на этот раз от начала к концу.

Из уравнения Беллмана для $Z_1(S_0)$ по S_0 находим оптимальное управление на 1-м шаге x_1^*. По S_0 и x_1^* определяем состояние системы после 1-го шага: $S_1 = S_1(S_0, x_1^*)$.

Из уравнения Беллмана для $Z_2(S_1)$ по S_1 находим оптимальное управление на 2-м шаге x_2^*. По S_1 и x_2^* определяем состояние системы после 2-го шага: $S_2 = S_2(S_1, x_2^*)$. И так далее.

Условно этот процесс можно представить так:

$$S_0 \to x_1^* \to S_1 \to x_2^* \to \dots \to S_{n-1} \to x_n^*.$$

Получаем оптимальное решение задачи ДП: $X^* = (x_1^*, x_2^*, \dots, x_n^*)$.

6.2. Задача распределения ресурсов

Пример 6.1. Между тремя предприятиями распределяются 60 млн руб. Прирост выпуска продукции на каждом предприятии зависит от выделенной суммы средств x. Значения прироста задаются в виде таблицы $g_i(x)$, $i = 1, 2, 3$ (табл. 6.1).

Найти такой план распределения выделенных средств между предприятиями, при котором общий прирост выпуска продукции будет максимальным.

▼Начальное значение $S_0 = 60$ млн руб. Разобьем весь процесс выделения средств предприятиям на 3 шага.

Таблица 6.1

Исходные данные для примера 6.1, млн руб.

Средства (x)	$g_1(x)$	$g_2(x)$	$g_3(x)$
0	0	0	0
20	6	14	14
40	23	21	20
60	30	34	35

Уравнения состояния в данной задаче имеют вид:

$$S_k = S_{k-1} - x_k,\ k = 1, 2, 3.$$

На 1-м шаге выделим x_1 млн руб. 1-му предприятию. После этого останется $S_1 = S_0 - x_1$ млн руб.

На 2-м шаге выделим x_2 млн руб. 2-му предприятию. После этого останется $S_2 = S_1 - x_2$ млн руб.

На 3-м шаге выделим x_3 млн руб. 3-му предприятию.

Уравнения Беллмана:

$$Z_k(S_{k-1}) = \max_{x_k}\left\{F_k(S_{k-1}, x_k) + Z_{k+1}(S_k)\right\} =$$
$$= \max_{0 \le x_k \le S_{k-1}}\left\{g_k(x_k) + Z_{k+1}(S_{k-1} - x_k)\right\},$$

т. е. на k-м шаге из оставшихся S_{k-1} средств надо выделить x_k средств k-му предприятию, чтобы прирост выпуска продукции на k-м и оставшихся предприятиях был максимальным.

Пусть $k = 3 \Rightarrow$ $Z_3(S_2) = \max\limits_{0 \le x_3 \le S_2} g_3(x_3)$.

Заполним табл. 6.2. В столбце S_2 и заголовках столбцов 2–5 x_3 указаны все возможные значения. Все оставшиеся перед 3-м шагом средства нужно выделить 3-му предприятию. Поэтому числа из столбца $g_3(x)$ исходной табл. 6.1 запишем в табл. 6.2 в столбцы со 2-го по 5-й. Определим максимум в каждой строке и результат записываем в 6-й столбец. Те x_3, которым соответствуют числа из 6-го столбца, запишем в 7-й столбец.

Определим оптимальную стратегию при распределении средств между 2-м и 3-м предприятиями. По табл. 6.1 и 6.2 заполним табл. 6.3. В 1-м столбце S_1 указано, сколько средств оста-

Таблица 6.2

Определение оптимальной стратегии при выделении средств 3-му предприятию, млн руб.

S_2	x_3				$Z_3(S_2)$	x_3^*
	0	**20**	**40**	**60**		
1	**2**	**3**	**4**	**5**	**6**	**7**
0	0				0	0
20		14			14	20
40			20		20	40
60				35	35	60

лось для 2-го и 3-го предприятий. В заголовках столбцов 2–5 x_2 дана информация о том, сколько из этих оставшихся средств досталось 2-му предприятию.

Пусть $k = 2$. Тогда $Z_2(S_1) = \max\limits_{0 \le x_2 \le S_1} \{g_2(x_2) + Z_3(S_1 - x_2)\}$.

Таблица 6.3

Определение оптимальной стратегии при распределении средств между 2-м и 3-м предприятиями, млн руб.

S_1	x_2				$Z_2(S_1)$	x_2^*
	0	**20**	**40**	**60**		
1	**2**	**3**	**4**	**5**	**6**	**7**
0	0 + 0 = 0				0	0
20	0 + 14 = 14	14 + 0 = 14			14	0 или 20
40	0 + 20 = 20	14 + 14 = 28	21 + 0 = 21		28	20
60	0 + 35 = 35	14 + 20 = 34	21 + 14 = 35	34 + 0 = 34	35	0 или 40

Поясним заполнение столбцов.

В клетке $S_1 = 0$, $x_2 = 0$:

$Z_2(0) = g_2(0) + Z_3(0) = 0 + 0 = 0.$

В клетке $S_1 = 20$, $x_2 = 0$:

$Z_2(20) = g_2(0) + Z_3(20) = 0 + 14 = 14.$

В клетке $S_1 = 20$, $x_2 = 20$:

$Z_2(20) = g_2(20) + Z_3(0) = 14 + 0 = 14.$

В клетке $S_1 = 40$, $x_2 = 0$:

$Z_2(40) = g_2(0) + Z_3(40) = 0 + 20 = 20.$

В клетке $S_1 = 40$, $x_2 = 20$:

$Z_2(40) = g_2(20) + Z_3(20) = 14 + 14 = 28$.

В клетке $S_1 = 40$, $x_2 = 40$:

$Z_2(40) = g_2(40) + Z_3(0) = 21 + 0 = 21$.

В клетке $S_1 = 60$, $x_2 = 0$:

$Z_2(60) = g_2(0) + Z_3(60) = 0 + 35 = 35$.

В клетке $S_1 = 60$, $x_2 = 20$:

$Z_2(60) = g_2(20) + Z_3(40) = 14 + 20 = 34$.

В клетке $S_1 = 60$, $x_2 = 40$:

$Z_2(60) = g_2(40) + Z_3(20) = 21 + 14 = 35$.

В клетке $S_1 = 60$, $x_2 = 60$:

$Z_2(60) = g_2(60) + Z_3(0) = 34 + 0 = 34$.

В столбцах со 2-го по 5-й определяем максимум в каждой строке и результат пишем в 6-й столбец. Те x_2, которым соответствуют числа из 6-го столбца, запишем в 7-й столбец.

Пусть k = 1. Тогда $Z_1(S_0) = \max\limits_{0 \le x_1 \le S_0} \{g_1(x_1) + Z_2(S_0 - x_1)\}$.

Определим оптимальную стратегию при распределении средств между предприятиями. По табл. 6.1 и 6.3 заполним табл. 6.4. В 1-м столбце S_0 указано общее количество средств, приходящихся на долю предприятий. В заголовках столбцов 2–5 x_1 дана информация о том, сколько из этих средств досталось 1-му предприятию.

Таблица 6.4

Определение оптимальной стратегии при выделении средств 1-му предприятию, млн руб.

S_0	x_1				$Z_1(S_0)$	x_1^*
	0	20	40	60		
1	2	3	4	5	6	7
20	0 + 14 = 14	6 + 0 = 6			14	0
40	0 + 28 = 28	6 + 14 = 20	23 + 0 = 23		28	0
60	0 + 35 = 35	6 + 28 = 34	23 + 14 = 37	30 + 0 = 30	37	40

Поясним заполнение столбцов.

В клетке $S_0 = 0$, $x_1 = 0$:

$Z_1(0) = g_1(0) + Z_2(0) = 0 + 0 = 0$.

В клетке $S_0 = 20$, $x_1 = 0$:
$Z_1(20) = g_1(0) + Z_2(20) = 0 + 14 = 14.$
В клетке $S_0 = 20$, $x_1 = 20$:
$Z_1(20) = g_1(20) + Z_2(0) = 6 + 0 = 6.$
В клетке $S_0 = 40$, $x_1 = 0$:
$Z_1(40) = g_1(0) + Z_2(40) = 0 + 28 = 28.$
В клетке $S_0 = 40$, $x_1 = 20$:
$Z_1(40) = g_1(20) + Z_2(20) = 6 + 14 = 20.$
В клетке $S_0 = 40$, $x_1 = 40$:
$Z_1(40) = g_1(40) + Z_2(0) = 23 + 0 = 23.$
В клетке $S_0 = 60$, $x_1 = 0$:
$Z_1(60) = g_1(0) + Z_2(60) = 0 + 35 = 35.$
В клетке $S_0 = 60$, $x_1 = 20$:
$Z_1(60) = g_1(20) + Z_2(40) = 6 + 28 = 34.$
В клетке $S_0 = 60$, $x_1 = 40$:
$Z_1(60) = g_1(40) + Z_2(20) = 23 + 14 = 37.$
В клетке $S_0 = 60$, $x_1 = 60$:
$Z_1(60) = g_1(60) + Z_2(0) = 30 + 0 = 30.$

В столбцах со 2-го по 5-й определяем максимум в каждой строке и результат пишем в 6-й столбец. Те x_1, которым соответствует число из 6-го столбца, запишем в 7-й столбец.

Просматриваем решение в обратном порядке.

Максимальное значение $Z_1(S_0) = 37$ при $S_0 = 60$ и $x_1^* = 40$. Если $x_1^* = 40$, то $S_1 = S_0 - x_1^* = 60 - 40 = 20$. Из табл. 6.3 и $S_1 = 20$ находим в последнем столбце $x_2^* = 0$ или 20.

Если $x_2^* = 0$, то $S_2 = S_1 - x_2^* = 20 - 0 = 20$. Из табл. 6.2 и $S_2 = 20$ находим в последнем столбце $x_3^* = 20$. Получили один оптимальный вариант распределения средств $x^* = (40; 0; 20)$.

Если $x_2^* = 20$, то $S_2 = S_1 - x_2^* = 20 - 20 = 0$. Из табл. 6.2 и $S_2 = 0$ находим в последнем столбце $x_3^* = 0$. Получили другой оптимальный вариант распределения средств $x^* = (40; 20; 0)$.

Общий прирост выпуска продукции в каждом из вариантов равен 37 млн руб.

Отметим, что эти же результаты были бы получены и при любом другом порядке распределения денег между предприятиями (например, не $3 \rightarrow 2 \rightarrow 1$, а $1 \rightarrow 2 \rightarrow 3$). ▲

Упражнение 6.1. Между тремя предприятиями распределяются 5 млн руб. Прирост выпуска продукции на каждом предприятии зависит от выделенной суммы средств x. Значения прироста задаются в виде таблицы $g_i(x)$, $i = 1, 2, 3$ (табл. 6.5).

Таблица 6.5

Исходные данные для упражнения 6.1, млн руб.

Средства (x)	$g_1(x)$	$g_2(x)$	$g_3(x)$
0	0	0	0
1	3	1	2
2	4	3	4
3	5	6	5
4	7	7	6
5	8	8	8

Найти такой план распределения выделенных средств между предприятиями, при котором общий прирост выпуска продукции будет максимальным.

Ответ: Оптимальный план $x^* = (1; 3; 1)$, общий прирост выпуска продукции равен 11 млн руб.

6.3. Задача замены оборудования

Чем дольше механизм эксплуатируется, тем выше затраты на его обслуживание и ниже его производительность. Когда срок эксплуатации механизма достигает определенного уровня, может оказаться более выгодной его замена. Задача замены оборудования, таким образом, сводится к определению оптимального срока эксплуатации механизма.

Предположим, что занимаемся заменой механизмов на протяжении n лет. В начале каждого года принимается решение либо об эксплуатации механизма еще один год, либо о замене его новым.

Обозначим через $r(t)$ и $c(t)$ прибыль от эксплуатации t-летнего механизма на протяжении года и затраты на его обслуживание за этот же период. Далее, пусть $s(t)$ — стоимость продажи механизма, который эксплуатировался t лет. Стоимость приобретения нового механизма остается неизменной на протяжении всех лет и равна I.

Элементы модели динамического программирования таковы:

• этап i представляется порядковым номером года i, $i = 1, 2, ..., n$;

• вариантами решения на i-м этапе (т. е. для i-го года) являются альтернативы: продолжить эксплуатацию или заменить механизм в начале i-го года.

• состоянием на i-м этапе является срок эксплуатации t (возраст) механизма к началу i-го года.

Пусть $f_i(t)$ — максимальная прибыль, получаемая за годы от i до n при условии, что в начале i-го года имеется механизм t-летнего возраста.

Рекуррентное уравнение имеет вид

$$f_i(t) = \max \begin{cases} r(t) - c(t) + f_{i+1}(t+1), & \text{если эксплуатировать механизм,} \\ r(0) - c(0) + s(t) - I + f_{i+1}(1), & \text{если заменить механизм,} \end{cases}$$

где $f_{n+1}(t) \equiv 0$.

Пример 6.2. К началу текущей пятилетки на предприятии установлено новое оборудование. Зависимость производительности этого оборудования в стоимостном выражении от времени его использования предприятием, а также зависимость затрат на содержание и ремонт оборудования при различном времени его использования приведены в табл. 6.6

Таблица 6.6

Исходные данные для примера 6.2

Показатель	**Время *t*, в течение которого используется оборудование, лет**					
	0	**1**	**2**	**3**	**4**	**5**
Годовой выпуск продукции ($r(t)$), тыс. руб.	80	75	65	60	60	55
Ежегодные затраты ($c(t)$), связанные с содержанием и ремонтом оборудования, тыс. руб.	20	25	30	35	45	55

Зная, что затраты, связанные с приобретением и установкой нового оборудования, идентичного с установленным, составляют 40 тыс. руб., а заменяемое оборудование списывается, составить такой план замены оборудования в течение пятилетки, при котором общая прибыль за данный период времени максимальна.

▼Обозначим через С решение о сохранении оборудования, а через З — решение о замене оборудования. Тогда задача состоит в нахождении такой стратегии управления, определяемой решениями, применяемыми к началу каждого года, при которой общая прибыль предприятия за пятилетку является максимальной.

При решении данной задачи используем табличную форму записи (табл. 6.7–6.10).

Таблица 6.7

Этап 5 (к началу 5-го года пятилетки возраст оборудования t = 1, 2, 3, 4)

t	С	З	Оптимум	
	$r(t) - c(t)$	$r(0) - c(0) - 40$	$f_5(t)$	Решение
1	75 – 25 = 50	80 – 20 – 40 = 20	50	С
2	65 – 30 = 35	80 – 20 – 40 = 20	35	С
3	60 – 35 = 25	80 – 20 – 40 = 20	25	С
4	60 – 45 = 15	80 – 20 – 40 = 20	20	З

Таблица 6.8

Этап 4 (к началу 4-го года пятилетки возраст оборудования t = 1, 2, 3)

t	С	З	Оптимум	
	$r(t) - c(t) + f_5(t + 1)$	$r(0) - c(0) - 40 + f_5(1)$	$f_4(t)$	Решение
1	75 – 25 + 35 = 85	80 – 20 – 40 + 50 = 70	85	С
2	65 – 30 + 25 = 60	80 – 20 – 40 + 50 = 70	70	З
3	60 – 35 + 20 = 45	80 – 20 – 40 + 50 = 70	70	З

Таблица 6.9

Этап 3 (к началу 3-го года пятилетки возраст оборудования t = 1, 2)

t	С	З	Оптимум	
	$r(t) - c(t) + f_4(t + 1)$	$r(0) - c(0) - 40 + f_4(1)$	$f_3(t)$	Решение
1	75 – 25 + 70 = 120	80 – 20 – 40 + 85 = 105	120	С
2	65 – 30 + 70 = 105	80 – 20 – 40 + 85 = 105	105	С или З

Таблица 6.10

Этап 2 (к началу 2-го года пятилетки возраст оборудования $t = 1$)

t	С	З	Оптимум	
	$r(t) - c(t) + f_3(t+1)$	$r(0) - c(0) - 40 + f_3(1)$	$f_2(t)$	Решение
1	75 – 25 + 105 = 155	80 – 20 – 40 + 120 = 140	155	С

Этап 1. Поскольку к началу пятилетки установлено новое оборудование ($t = 0$), то $f_1(0) = r(0) - c(0) + f_2(1) = 80 - 20 + 155 = 215$.

Просматривая полученные результаты в обратном порядке, получим: для 1-го года пятилетки решение единственное (этап 1) — следует сохранить оборудование. Значит, возраст оборудования к началу 2-го года пятилетки равен одному году. Тогда оптимальным решением для 2-го года пятилетки (этап 2) является решение о сохранении оборудования. Реализация такого решения приводит к тому, что возраст оборудования к началу 3-го года пятилетки становится равным двум годам. При таком возрасте (этап 3) оборудование было заменено по решению руководства. После замены оборудования его возраст к началу 4-го года пятилетки составит один год. При таком возрасте (этап 4) его менять не следует. Поэтому возраст оборудования к началу 5-го года пятилетки составит два года. При таком возрасте (этап 5) оборудование менять не следует.

Оптимальным планом замены оборудования будет u = (С; С; З; С; С). Максимальная прибыль предприятия, соответствующая оптимальному плану, равна 215 тыс. руб. ▲

Упражнение 6.2. Найти оптимальную стратегию эксплуатации оборудования на период продолжительностью 6 лет, если годовой доход $r(t)$ и остаточная стоимость $s(t)$ в зависимости от возраста заданы в табл. 6.11.

Таблица 6.11

Исходные данные для упражнения 6.2

Показатель	t, лет						
	0	1	2	3	4	5	6
$r(t)$, усл. ед.	8	8	7	7	6	6	5
$s(t)$, усл. ед.	10	7	6	5	4	3	2

Стоимость нового оборудования равна $I = 10$ усл. ед., а возраст оборудования к началу эксплуатационного периода составлял 1 год.

Ответ: за 6 лет эксплуатации оборудования замену надо произвести в начале 2-го года эксплуатации. Максимальный доход составит 40 усл. ед.

6.4. Задача о загрузке

Задача о загрузке — это задача о рациональной загрузке судна (самолета, автомашины и т. п.), которое имеет ограничения по объему или грузоподъемности. Каждый помещенный на судно груз приносит определенную прибыль. Задача состоит в определении загрузки судна такими грузами, которые приносят наибольшую суммарную прибыль.

Рекуррентное уравнение процедуры обратной прогонки выводится для общей задачи загрузки судна грузоподъемностью W предметов (грузов) n наименований.

Пусть m_i — количество предметов i-го наименования, подлежащих загрузке, r_i — прибыль, которую приносит один загруженный предмет i-го наименования, w_i — вес одного предмета i-го наименования.

Общая задача имеет вид следующей целочисленной задачи линейного программирования:

$$z = r_1 m_1 + r_2 m_2 + \ldots + r_n m_n \to \max$$

при условии

$$w_1 m_1 + w_2 m_2 + \ldots + w_n m_n \le W,$$

$$m_1, m_2, \ldots, m_n \ge 0 \text{ и целые.}$$

Три элемента модели динамического программирования определяется следующим образом.

1. Этап i ставится в соответствие предмету i-го наименования, $i = \overline{1, n}$.

2. Варианты решения на этапе i описываются количеством m_i предметов i-го наименования, подлежащих загрузке. Соответ-

ствующая прибыль равна $r_i m_i$. Значение m_i заключено в пределах от 0 до $[W/w_i]$, где $[W/m_i]$ — целая часть числа.

3. Состояние x_i, на этапе i выражает суммарный вес предметов, решения о погрузке которых приняты на этапах i, $i + 1$, ..., n. Это определение отражает тот факт, что ограничение по весу является единственным, которое связывает n этапов вместе.

Пусть $f_i(x_i)$ — максимальная суммарная прибыль от этапов i, $i + 1$, ..., n при заданном состоянии x_i. Проще всего рекуррентное уравнение определяется с помощью следующей двухшаговой процедуры.

Шаг 1. Выразим $f_i(x_i)$ как функцию $f_{i-1}(x_{i-1})$ в виде

$$f_i(x_i) = \max_{\substack{m_i = 0,1,\ldots,\left[\frac{W}{w_i}\right] \\ x_i = 0,1,\ldots,W}} \{r_i m_i + f_{i+1}(x_{i+1})\},\ i = 1, 2, \ldots, n,$$

где $f_{n-1}(x_{n-1}) \equiv 0$.

Шаг 2. Выразим x_{i-1} как функцию x_i гарантии того, что левая часть последнего уравнения является функцией лишь x_i. По определению $x_i - x_{i-1}$ представляет собой вес, загруженный на этапе i, т. е. $x_i - x_{i+1} = w_i m_i$. Следовательно, рекуррентное уравнение приобретает вид

$$f_i(x_i) = \max_{\substack{m_i = 0,1,\ldots,\left[\frac{W}{w_i}\right] \\ x_i = 0,1,\ldots,W}} \{r_i m_i + f_{i+1}(x_i - w_i m_i)\},\ i = 1, 2, \ldots, n.$$

Пример 6.3. В самолет грузоподъемностью 4 т загружаются предметы трех наименований. Данные о весе одного предмета w_i (в тоннах) и прибыли r_i (в тыс. руб), получаемой от одного загруженного предмета, приведены в табл. 6.12.

Таблица 6.12

Исходные данные для примера 6.3

Предмет (i)	**Вес одного предмета (w_i), т**	**Прибыль, получаемая от одного загруженного предмета, (r_i), тыс. руб.**
1	2	31
2	3	47
3	1	14

Необходимо загрузить самолет так, чтобы получить максимальную прибыль.

▼Так как вес одного предмета w_i для всех наименований и максимальный вес W принимают целочисленные значения, состояние x_i может принимать лишь целочисленные значения.

Этап 3. Точный вес, который может быть загружен на этапе 3 (предмет 3), заранее неизвестен, но он должен принимать одно из значений 0, 1, ..., 4 (так как $W = 4$ т). Состояния $x_3 = 0$ и $x_3 = 4$ представляют собой крайние случаи, когда предметы третьего наименования совсем не загружаются или загружают самолет полностью. Остальные значения x_3 (= 1, 2 или 3) предполагают частичную загрузку самолета предметами третьего наименования. Действительно, при этих значениях x_3 весь оставшийся объем самолета может быть заполнен предметами третьего наименования.

Так как вес w_3 одного предмета третьего типа равен 1 т, максимальное количество единиц этого типа, которое может быть загружено, равно [4/1] = 4. Это означает, что возможными значениями x_3 будут 0, 1, 2, 3 и 4.

Решение m_i является допустимым лишь при условии, что $w_3 m_3 \leq x_3$. Следовательно, все недопустимые альтернативы (те, для которых $w_3 m_3 > x_3$) исключены. Следующее уравнение является основой для сравнения альтернатив на этапе 3:

$$f_3(x_3) = \max_{m_3}\{14m_3\}, \ \max\{m_3\} = \left[\frac{4}{1}\right] = 4$$

В табл. 6.13 сравниваются допустимые решения для каждого значения x_3.

Таблица 6.13

Расчеты этапа 3 для примера 6.3

x_3	$14m_3$					Оптимальное решение	
	$m_3 = 0$	$m_3 = 1$	$m_3 = 2$	$m_3 = 3$	$m_3 = 4$	$f_3(x_3)$	m_3^*
0	0	—	—	—	—	0	0
1	0	14	—	—	—	14	1
2	0	14	28	—	—	28	2
3	0	14	28	42	—	42	3
4	0	14	28	42	56	56	4

Этап 2 (табл. 6.14). $f_2(x_2) = \max_{m_2}\{47m_2 + f_3(x_2 - 3m_2)\}$, $\max\{m_2\} = \left[\frac{4}{3}\right] = 1$.

Таблица 6.14

Расчеты этапа 2 для примера 6.3

x_2	$47m_2 + f_3(x_2 - 3m_2)$		Оптимальное решение	
	$m_2 = 0$	$m_2 = 1$	$f_2(x_2)$	m_2*
0	0 + 0 = 0	—	0	0
1	0 + 14 = 14	—	14	0
2	0 + 28 = 28	—	28	0
3	0 + 42 = 42	47 + 0 = 47	47	1
4	0 + 56 = 56	47 + 14 =61	61	1

Этап 1 (табл. 6.15). $f_1(x_1) = \max_{m_1}\{31m_1 + f_2(x_1 - 2m_1)\}$,

$\max\{m_1\} = \left[\frac{4}{2}\right] = 2$

Таблица 6.15

Расчеты этапа 1 для примера 6.3

x_1	$31m_1 + f_2(x_1 - 2m_1)$			Оптимальное решение	
	$m_1 = 0$	$m_1 = 1$	$m_1 = 2$	$f_1(x_1)$	m_1*
0	0 + 0 = 0	—	—	0	0
1	0 + 14 = 14	—	—	14	0
2	0 + 28 = 28	31 + 0 = 31	—	31	1
3	0 + 47 = 47	31 + 14 = 45	—	47	0
4	0 + 61 = 61	31 + 28 = 59	62 + 0 = 62	<u>62</u>	<u>2</u>

Оптимальное решение определяется теперь следующим образом. Из условия $W = 4$ следует, что первый этап решения задачи при $x_1 = 4$ дает оптимальное решение $m_1^* = 2$, которое означает, что два предмета первого наименования будут загружены в самолет. Эта загрузка оставляет $x_2 = x_1 - 2m_1^* = 4 - 2 \cdot 2 = 0$. Решение на втором этапе при $x_2 = 0$ приводит к оптимальному решению $m_2^* = 0$, которое, в свою очередь, дает $x_3 = x_2 - 3m_3^* = 0 - 3 \cdot 0 = 0$. Далее этап 3 при $x_3 = 0$ приводит к $m_3^* = 0$. Следовательно, оптимальным решением задачи является $m_1^* = 2$, $m_2^* = 0$ и $m_3^* = 0$. Соответствующая прибыль равна 62 тыс. руб. ▲

Задача о загрузке является типичным представителем задачи распределения ресурсов, в которой ограниченный ресурс распределяется между конечным числом видов (экономической) деятельности. При этом целью является максимизация соответствующей функции прибыли.

Решим пример 6.3 как задачу ЛП, используя *Excel* (*Поиск решения*).

$$F(m) = r_1 m_1 + r_2 m_2 + r_3 m_3 \to \max,$$

$$w_1 m_1 + w_2 m_2 + w_3 m_3 \le W,$$

$$m_1, m_2, m_3 \ge 0,$$

$$m_1, m_2, m_3 \text{ — целые.}$$

▼Для рассматриваемого примера рабочий лист *Excel*, заполненные окна *Поиск решения* и *Параметры поиска решения* имеют вид, представленный на рис. 6.2.

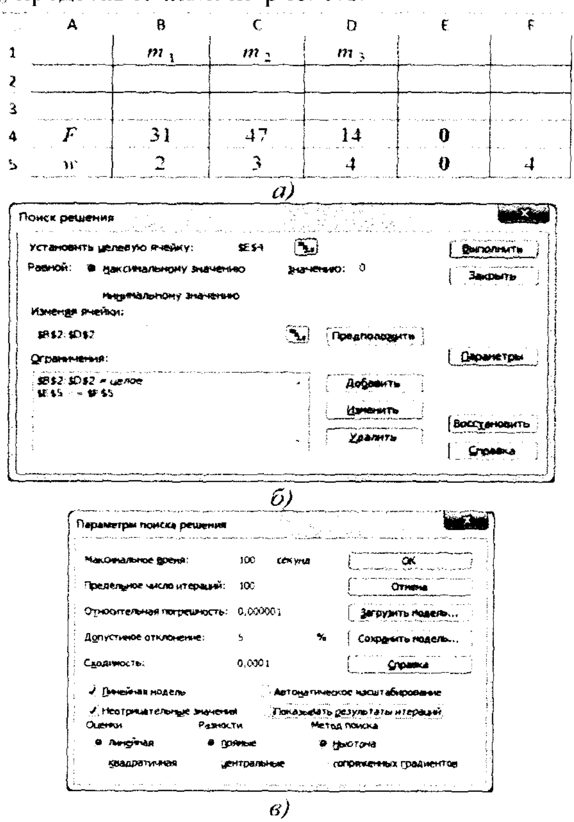


Рис. 6.2. Исходные данные в *Excel* (*а*), заполненные диалоговые окна *Поиск решения* (*б*) и *Параметры поиска решения* (*в*) для примера 6.3

Вид рабочего листа *Excel* после выполнения команды *Поиск решения* представлен на рис. 6.3.

	A	B	C	D	E	F
1		m_1	m_2	m_3		
2		2	0	0		
3						
4	F	31	47	14	**62**	
5	w	2	3	4	**4**	4

Рис. 6.3. Пример 6.3 в *Excel* после выполнения команды *Поиск решения*

Оптимальное решение: $m = (2;\ 0;\ 0)$, $F_{\max}(m) = 62$. ▲

Упражнение 6.3. В примере 6.3 определите оптимальное решение, предполагая, что максимальная грузоподъемность самолета составляет 7 т.

Ответ: $m^* = (2;\ 1;\ 0)$, доход равен 109 тыс. руб.

6.5. Задача планирования рабочей силы

При выполнении некоторых проектов необходимое число рабочих регулируется путем их найма и увольнения. Поскольку как наем, так и увольнение рабочих связаны с дополнительными затратами, необходимо определить, каким образом должна регулироваться численность рабочих в период реализации проекта.

Предположим, что проект будет выполняться в течение n недель и минимальная потребность в рабочей силе на протяжении i-й недели составит b_i, рабочих. При идеальных условиях хотелось бы на протяжении i-й недели иметь ровно b_i рабочих. Однако в зависимости от стоимостных показателей может быть более выгодным отклонение численности рабочей силы как в одну, так и в другую сторону от минимальных потребностей. Если x_i — количество работающих на протяжении i-й недели, то возможны затраты двух видов: 1) $C_1(x_i - b_i)$ — затраты, связанные с необходимостью содержать избыток $x_i - b_i$ рабочей силы, и 2) $C_2(x_i - x_{i-1})$ — затраты, связанные с необходимостью дополнительного найма $x_i - x_{i-1}$ рабочих.

Элементы модели динамического программирования определяются следующим образом.

1. *Этап* i представляется порядковым номером недели i, $i = 1, 2, \ldots, n$.

2. *Вариантами решения* на i-м этапе являются значения x_i — количество работающих на протяжении i-й недели.

3. *Состоянием* на i-м этапе является x_{i-1} — количество работающих на протяжении $(i-1)$-й недели (этапа).

Рекуррентное уравнение динамического программирования представляется в виде

$$f_i(x_{i-1}) = \min\{C_1(x_i - b_i) + C_2(x_i - x_{i-1}) + f_{i+1}(x_i)\}, \ i = \overline{1, n},$$

где $f_{n+1}(x_n) \equiv 0$. Вычисления начинаются с этапа n при $x_n = b_n$ и заканчиваются на этапе 1.

Пример 6.4. Строительный подрядчик оценивает минимальные потребности в рабочей силе на каждую из последующих пяти недель следующим образом: 5, 7, 8, 4 и 6 рабочих соответственно. Содержание избытка рабочей силы обходится подрядчику в 300 долл. за одного рабочего в неделю, а наем рабочей силы на протяжении одной недели обходится в 400 долл. плюс 200 долл. за одного рабочего в неделю.

▼Выражая C_1 и C_2 в сотнях долларов, имеем следующее:

$$b_1 = 5, b_2 = 7, b_3 = 8, b_4 = 4, b_5 = 6,$$

$$C_1(x_i - b_i) = 3(x_i - b_i), x_i > b_i, i = 1, 2, \ldots, 5,$$

$$C_2(x_i - x_{i-1}) = 4 + 2(x_i - x_{i-1}), x_i > x_{i-1}, i = 1, 2, \ldots, 5.$$

Этап 5 (табл. 6.16). $b_5 = 6$.

Таблица 6.16

Расчеты этапа 5 для примера 6.4

x_4	$C_1(x_5 - 6) + C_2(x_5 - x_4)$	Оптимальное решение	
	$x_5 = 6$	$f_5(x_4)$	x_5^*
4	3(0) + 4 + 2(2) = 8	8	6
5	3(0) + 4 + 2(1) = 6	6	6
6	3(0) + 0 = 0	0	6

Этап 4 (табл. 6.17). $b_4 = 4$.

Таблица 6.17

Расчеты этапа 4 для примера 6.4

x_3	$C_1(x_4 - 4) + C_2(x_4 - x_3) + f_5(x_4)$			Оптимальное решение	
	$x_4 = 4$	$x_4 = 5$	$x_4 = 6$	$f_4(x_3)$	x_4*
8	3(0) + 0 + 8 = 8	3(1) + 0 + 6 = 9	3(2) + 0 + 0 = 6	6	6

Этап 3 (табл. 6.18). $b_3 = 8$.

Таблица 6.18

Расчеты этапа 3 для примера 6.4

x_2	$C_1(x_3 - 8) + C_2(x_3 - x_2) + f_4(x_3)$	Оптимальное решение	
	$x_3 = 8$	$f_3(x_2)$	x_3*
7	3(0) + 4 + 2(1) + 6 = 12	12	8
8	3(0) + 0 + 6 = 6	6	8

Этап 2 (табл. 6.19). $b_2 = 7$.

Таблица 6.19

Расчеты этапа 2 для примера 6.4

x_1	$C_1(x_2 - 7) + C_2(x_2 - x_1) + f_3(x_2)$		Оптимальное решение	
	$x_2 = 7$	$x_2 = 8$	$f_2(x_1)$	x_2*
5	3(0) + 4 + 2(2) + 12 = 20	3(1) + 4 + 2(3) + 6 = 19	19	8
6	3(0) + 4 + 2(1) + 12 = 18	3(1) + 4 + 2(2) + 6 = 17	17	8
7	3(0) + 0 + 12 = 12	3(1) + 4 + 2(1) + 6 = 15	12	7
8	3(0) + 0 + 12 = 12	3(1) + 0 + 6 = 9	9	8

Этап 1 (табл. 6.20). $b_1 = 5$.

Таблица 6.20

Расчеты этапа 1 для примера 6.4

x_0	$C_1(x_1 - 5) + C_2(x_1 - x_0) + f_2(x_1)$				Оптимальное решение	
	$x_1 = 5$	$x_1 = 6$	$x_1 = 7$	$x_1 = 8$	$f_1(x_0)$	x_1*
0	3(0) + 4 + + 2(5) + 19 = = 33	3(1) + 4 + + 2(6) + 17 = = 36	3(2) + 4 + + 2(7) + 12 = = 36	3(3) + 4 + + 2(8) + 9 = = 38	33	5

Оптимальное решение определяется последовательно таким образом:

$$x_0 = 0 \to x_1^* = 5 \to x_2^* = 8 \to x_3^* = 8 \to x_4^* = 6 \to x_5^* = 6.$$

Полученному решению соответствует план, представленный в табл. 6.21.

Таблица 6.21

Планирование рабочей силы для примера 6.4

Номер недели (i)	Минимум рабочей силы (b_i)	Количество фактически работающих (x_i)	Решение
1	5	5	Нанять 5 рабочих
2	7	8	Нанять 3 рабочих
3	8	8	Ничего не менять
4	4	6	Уволить 2 рабочих
5	6	6	Ничего не менять

▲

Упражнение 6.4. Решите задачу из примера 6.5 при следующих минимальных потребностях в рабочей силе: $b_1 = 6$, $b_2 = 5$, $b_3 = 3$, $b_4 = 6$, $b_5 = 8$.

Ответ: Количество рабочих, занятых в каждую неделю, равно соответственно 6, 5, 3, 6 и 8 человек.

6.6. Задача о кратчайшем пути

Пример 6.5. Сеть дорог, показанная на рис. 6.4, представляет возможные маршруты между исходным городом, находящимся в узле 1, и конечным пунктом, который находится в узле 7. Маршруты проходят через промежуточные города, обозначенные на сети узлами с номерами 2–6. Между различными узлами указаны расстояния, км.

Требуется определить маршрут доставки груза из пункта 1 в пункт 7, обеспечивающий наименьшие транспортные расходы.

▼Разделим решение задачи на *этапы*. На рис. 6.5 вертикальные пунктирные линии очерчивают три этапа задачи.

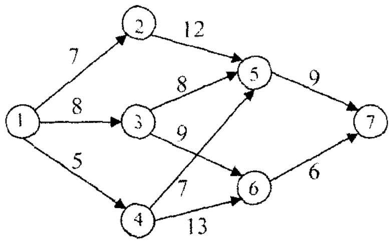


Рис. 6.4. Сеть дорог и расстояния между городами для примера 6.5

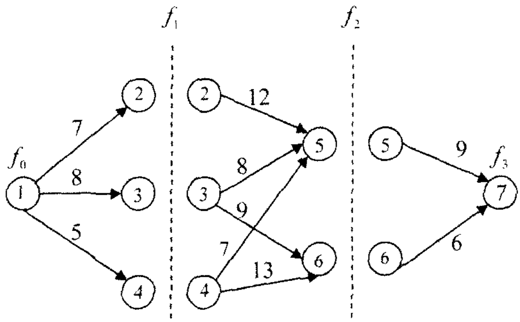


Рис. 6.5. Определение трех этапов задачи для примера 6.5

Общая задача состоит в вычислении кратчайших (постепенно накапливаемых) расстояний ко всем вершинам этапа с последующим использованием этих расстояний в качестве исходных данных для следующего этапа.

Последовательность вычислений можно проводить, используя прямую или обратную прогонки.

Прямая прогонка. Вычисления проводятся от первого этапа до третьего.

Пусть $f_i(x_i)$ — кратчайшее расстояние до вершины x_i на этапе i, $d(x_{i-1}, x_i)$ — расстояние от узла x_{i-1} до узла x_i. Тогда f_i вычисляется из f_{i-1} с помощью следующего рекуррентного уравнения:

$$f_i(x_i) = \min_{\substack{\text{допустимые} \\ (x_{i-1},x_i)}} \left\{ d(x_{i-1}, x_i) + f_{i-1}(x_{i-1}) \right\},\ i = 1, 2, 3,$$

где $f_0(x_0) \equiv 0$ для $x_0 = 1$. Соответствующей последовательностью вычислений будет $f_1 \to f_2 \to f_3$. При решении данного примера используем табличную форму записи.

Этап 1. Рассматривая узлы, относящиеся к первому этапу, замечаем, что каждый из узлов 2, 3 и 4 связан с начальным узлом 1 единственной дугой (см. рис. 6.4). Составим табл. 6.22.

Таблица 6.22

Расчеты этапа 1 (прямая прогонка) для примера 6.5

x_1	$d(x_0, x_1)$	Оптимальное решение	
	$x_0 = 1$	$f_1(x_1)$	x_0*
2	7	7	1
3	8	8	1
4	5	5	1

Этап 2. Вычисления кратчайших (накопленных) расстояний к узлам 5 и 6 представлены в табл. 6.23.

Таблица 6.23

Расчеты этапа 2 (прямая прогонка) для примера 6.5

x_2	$d(x_1, x_2) + f_1(x_1)$			Оптимальное решение	
	$x_1 = 2$	$x_1 = 3$	$x_1 = 4$	$f_2(x_2)$	x_1*
5	12 + 7 = 19	8 + 8 = 16	7 + 5 = 12	12	4
6	—	9 + 8 = 17	13 + 5 = 18	17	3

Этап 3. Конечный узел 7 можно достигнуть как из узла 5, так и из узла 6. Используя результаты этапа 2 и расстояния от узлов 5 и 6 к узлу 7, составляем табл. 6.24.

Таблица 6.24

Расчеты этапа 3 (прямая прогонка) для примера 6.5

x_3	$d(x_2, x_3) + f_2(x_2)$		Оптимальное решение	
	$x_2 = 5$	$x_2 = 6$	$f_3(x_3)$	x_2*
7	9 + 12 = 21	6 + 17 = 23	21	5

Приведенные вычисления показывают, что кратчайшее расстояние между узлами 1 и 7 равно 21 км. Из результатов этапа 3 следует, что узел $x_3 = 7$ связывается с узлом $x_2 = 5$. Из результатов этапа 2 следует, что узел $x_2 = 5$ связывается с узлом $x_1 = 4$. Наконец, из результатов этапа 1 следует, что узел $x_1 = 4$ связывается с узлом $x_0 = 1$. Следовательно, оптимальным маршрутом является последовательность $1 \to 4 \to 5 \to 7$.

Этот же пример может быть решен с помощью обратной прогонки.

Обратная прогонка. Вычисления проводятся от третьего этапа до первого. Рекуррентное уравнение для алгоритма обратной прогонки примера имеет вид

$$f_i(x_i) = \min_{\substack{\text{допустимые} \\ (x_{i-1}, x_i)}} \{d(x_i, x_{i+1}) + f_{i+1}(x_{i+1})\}, \; i = 1, 2, 3,$$

где $f_4(x_4) \equiv 0$ для $x_4 = 7$. Соответствующей последовательностью вычислений будет $f_3 \to f_2 \to f_1$. При решении данного примера используем также табличную форму записи.

Этап 3. Поскольку узел 7 ($x_4 = 7$) связан с узлами 5 и 6 ($x_3 = 5$ и 6) только одним маршрутом, результаты третьего этапа можно подытожить следующим образом в табл. 6.25.

Таблица 6.25

Расчеты этапа 3 (обратная прогонка) для примера 6.5

x_3	$d(x_3, x_4)$	Оптимальное решение	
	$x_4 = 7$	$f_3(x_3)$	x_4*
5	9	9	7
6	6	6	7

Этап 2. Используя значения $f_3(x_3)$, полученные на третьем этапе, можем сравнить допустимые альтернативные решения, как показано в табл. 6.26.

Таблица 6.26

Расчеты этапа 2 (обратная прогонка) для примера 6.5

x_2	$d(x_2, x_3) + f_3(x_3)$		*Оптимальное решение*	
	$x_3 = 5$	$x_3 = 6$	$f_2(x_2)$	x_3*
2	12 + 9 = 21	—	21	5
3	8 + 9 = 17	9 + 6 = 15	15	6
4	7 + 9 = 16	13 + 6 = 19	16	5

Этап 1. Из узла 1 начинаются три альтернативных маршрута: (1, 2), (1, 3) и (1, 4). Используя значения $f_2(x_2)$, полученные на втором этапе, вычисляем данные табл. 6.27.

Таблица 6.27

Расчеты этапа 1 (обратная прогонка) для примера 6.5

x_1	$d(x_1, x_2) + f_2(x_2)$			Оптимальное решение	
	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$	$f_1(x_1)$	x_2^*
1	7 + 21 = 28	8 + 15 = 23	5 + 16 = 21	21	4

Просматриваем решение в обратном порядке. Оптимальное решение на этапе 1 показывает, что кратчайший путь проходит через узел 4 ($x_2 = 4$). Далее из оптимального решения на этапе 2 следует, что узел $x_2 = 4$ связывается с узлом $x_3 = 5$. Наконец, из результатов этапа 3 следует, что узел $x_3 = 5$ связывается с узлом $x_4 = 7$. Следовательно, оптимальным маршрутом является последовательность $1 \to 4 \to 5 \to 7$, и его длина равна 21 км. ▲

Алгоритмы прямой и обратной прогонки приводят к одному и тому же решению. Несмотря на то что алгоритм прямой прогонки представляется более логичным, в специальной литературе, посвященной динамическому программированию, неизменно используется алгоритм обратной прогонки. Причина этого в том, что в общем случае алгоритм обратной прогонки может быть более эффективным с вычислительной точки зрения.

6.7. Задача выбора оптимального маршрута перевозки грузов

Пример 6.6. Транспортная сеть состоит из 10 узлов, часть из которых соединена магистралями. На рис. 6.6 показаны сеть дорог и стоимости перевозки единицы груза между отдельными пунктами сети, которые проставлены у соответствующих ребер.

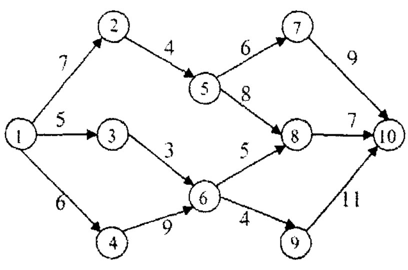


Рис. 6.6. Сеть дорог и стоимости перевозки единицы груза между отдельными пунктами для примера 6.6

▼Разделим решение задачи на *этапы*. На рис. 6.7 вертикальные пунктирные линии отделяют четыре этапа задачи.

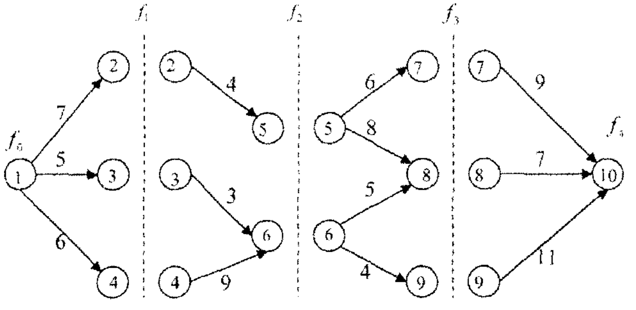


Рис. 6.7. Определение четырех этапов задачи для примера 6.6

Вычисления проводятся от четвертого этапа до первого (обратная прогонка). Пусть $f_i(x_i)$ — минимальные затраты на перевозку груза до вершины x_i на этапе i, $c(x_i, x_{i+1})$ — стоимость перевозки груза от узла x_i до узла x_{i+1}. Тогда f_i вычисляется из f_{i+1} с помощью следующего рекуррентного уравнения:

$$f_i(x_i) = \min_{\substack{\text{допустимые} \\ (x_i, x_{i+1})}} \left\{ c(x_i, x_{i+1}) + f_{i+1}(x_{i+1}) \right\},\ i = 1, 2, 3, 4,$$

где $f_5(x_5) \equiv 0$ для $x_5 = 10$. Соответствующей последовательностью вычислений будет $f_4 \to f_3 \to f_2 \to f_1$. При решении данного примера используем табличную форму записи.

Этап 4. Поскольку узел 10 ($x_5 = 10$) связан с узлами 7, 8 и 9 (x_4 = 7, 8 и 9) только одним маршрутом, результаты четвертого этапа можно подытожить следующим образом в табл. 6.28.

Таблица 6.28

Расчеты этапа 4 для примера 6.6

x_4	$c(x_4, x_5)$	**Оптимальное решение**	
	$x_5 = 10$	$f_4(x_4)$	x_5^*
7	9	9	10
8	7	7	10
9	11	11	10

Этап 3. Используя значения $f_4(x_4)$, полученные на четвертом этапе, можно сравнить допустимые альтернативные решения, как показано в табл. 6.29.

Таблица 6.29

Расчеты этапа 3 для примера 6.6

x_3	$c(x_3, x_4) + f_4(x_4)$			Оптимальное решение	
	$x_4 = 7$	$x_4 = 8$	$x_4 = 9$	$f_3(x_3)$	x_4*
5	6 + 9 = 15	8 + 7 = 15	—	15	7 или 8
6	—	5 + 7 = 12	4 + 11 = 15	12	8

Этап 2. Используя значения $f_3(x_3)$, полученные на третьем этапе, можно сравнить допустимые альтернативные решения, как показано в табл. 6.30.

Таблица 6.30

Расчеты этапа 2 для примера 6.6

x_2	$c(x_2, x_3) + f_3(x_3)$		Оптимальное решение	
	$x_3 = 5$	$x_3 = 6$	$f_2(x_2)$	x_3*
2	4 + 15 = 19	—	19	5
3	—	3 + 12 = 15	15	6
4	—	9 + 12 = 21	21	6

Этап 1. Из узла 1 начинаются три альтернативных маршрута: (1, 2), (1, 3) и (1, 4). Используя значения $f_2(x_2)$, полученные на втором этапе, вычисляем данные табл. 6.31.

Таблица 6.31

Расчеты этапа 1 для примера 6.6

x_1	$c(x_1, x_2) + f_2(x_2)$			Оптимальное решение	
	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$	$f_1(x_1)$	x_2*
1	7 + 19 = 26	5 + 15 = 20	6 + 21 = 27	20	3

Просматриваем решение в обратном порядке. Из этапа 1 следует, что минимальные затраты на перевозку груза между узлами 1 и 10 составляют $f_1(1) = 20$. Это достигается при движении груза из узла $x_1 = 1$ в узел $x_2 = 3$. Из результатов этапа 2 следует, что узел $x_2 = 3$ связывается с узлом $x_3 = 6$. Далее из результатов этапа 3 следует, что узел $x_3 = 6$ связывается с узлом $x_4 = 8$. Наконец, из результатов этапа 4 следует, что узел $x_4 = 8$ связывается с узлом $x_5 = 10$. Следовательно, оптимальным маршрутом является последовательность $1 \to 3 \to 6 \to 8 \to 10$. ▲

Глава 7.
МОДЕЛИ УПРАВЛЕНИЯ ЗАПАСАМИ

7.1. Постановка задачи

Как в торговле, так и в производстве обычно принято поддерживать разумный запас материальных ресурсов или комплектующих для обеспечения непрерывности производственного процесса. Традиционно запас рассматривается как неизбежные издержки, когда слишком низкий его уровень приводит к дорогостоящим остановкам производства, а слишком высокий — к «омертвлению» капитала. Задача управления запасами состоит в определении уровня запаса, который уравновешивает два упомянутых крайних случая.

Природа задачи управления запасами определяется неоднократным размещением и получением заказов заданных объемов продукции (в дальнейшем — хранимых запасов) в определенные моменты времени. С этой точки зрения стратегия управления запасами должна отвечать на следующие два вопроса.

1. Какое количество хранимого запаса следует заказать?

2. Когда заказывать?

Ответ на первый вопрос определяет экономичный размер заказа путем минимизации следующей функции затрат:

$$\begin{pmatrix}\text{Суммарные}\\ \text{затраты системы}\\ \text{управления}\\ \text{запасами}\end{pmatrix}=\begin{pmatrix}\text{Затраты на}\\ \text{приобретение}\end{pmatrix}+\begin{pmatrix}\text{Затраты на}\\ \text{оформление}\\ \text{заказа}\end{pmatrix}+$$

$$+\begin{pmatrix}\text{Затраты на}\\ \text{хранение}\\ \text{заказа}\end{pmatrix}+\begin{pmatrix}\text{Потери от}\\ \text{дефицита}\\ \text{запаса}\end{pmatrix}.$$

Все эти стоимости должны быть выражены как функции искомого объема заказа и интервала времени между заказами:

1) затраты на приобретение определяются стоимостью единицы приобретаемой продукции (хранимого запаса). Эта стоимость может быть постоянной или со скидкой, которая зависит от объема заказа;

2) затраты на оформление заказа представляют собой постоянные расходы, связанные с его размещением (для изготовления продукции) на других производствах. Эти затраты не зависят от объема заказа;

3) затраты на хранение запаса представляют собой затраты на содержание запаса на складе. Этот вид затрат включает как процент на инвестированный капитал, так и стоимость хранения, содержания и ухода;

4) потери от дефицита запаса — это расходы, обусловленные отсутствием запаса необходимой продукции. Они включают как потенциальные потери прибыли, так и более субъективную стоимость, связанную с потерей доверия клиентов.

Ответ на второй вопрос (когда заказывать?) зависит от типа системы управления запасами, с которой мы имеем дело. Если система предусматривает периодический контроль состояния запаса (например, каждую неделю или месяц), момент поступления нового заказа совпадает с началом периода. Если же в системе предусмотрен непрерывный контроль состояния запаса, новые заказы размещаются тогда, когда уровень запаса опускается до заранее определенного значения, называемого точкой возобновления заказа.

7.2. Классическая модель экономичного размера заказа

Простейшие модели управления запасами характеризуются постоянным во времени спросом, мгновенным пополнением запаса и отсутствием дефицита. Введем обозначения:

- y — объем заказа (количество единиц продукции);
- D — интенсивность спроса (измеряется в единицах продукции на единицу времени);
- t_0 — продолжительность цикла заказа (измеряется во временных единицах).

Уровень запаса изменяется в соответствии с функцией, показанной на рис. 7.1, где использованы приведенные выше обозначения. Заказ объема y единиц размещается и пополняется мгновенно, когда уровень запаса равен нулю. Затем запас равномерно расходуется с постоянной интенсивностью спроса D. Продолжительность цикла заказа для этого примера равна

$$t_0 = \frac{y}{D} \text{ единиц времени.}$$

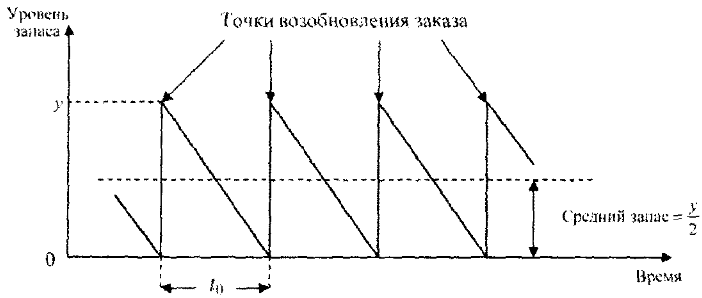


Рис. 7.1. *Изменение запаса в классической модели*

Средний уровень запаса определяется соотношением $\frac{y}{2}$.

Для построения функции затрат требуются два стоимостных параметра:

- K — затраты на оформление, связанные с размещением заказа;
- h — затраты на хранение (затраты на единицу складируемой продукции в единицу времени).

Суммарные затраты *в единицу времени* (TCU) можно представить как функцию от y:

$$TCU(y) = \frac{K + h\left(\frac{y}{2}\right)t_0}{t_0} = \frac{KD}{y} + h\left(\frac{y}{2}\right).$$

Оптимальное значение объема заказа y определяется путем минимизации по y функции $TCU(y)$. Предполагая, что y является непрерывной переменной, получаем необходимое условие минимума (в виде уравнения), из которого можно найти оптимальное значение y:

$$\frac{\partial TCU(y)}{\partial y} = -\frac{KD}{y^2} + \frac{h}{2} = 0.$$

Это условие является также и достаточным, так как функция $TCU(y)$ выпуклая. Решение данного уравнения определяет экономичный объем заказа

$$y^* = \sqrt{\frac{2KD}{h}}.$$

Оптимальная стратегия управления запасами для рассмотренной модели формулируется следующим образом: *заказывать y^* единиц продукции через каждые $t_0^* = y^* / D$ единиц времени.*

В действительности пополнение запаса не может произойти мгновенно в момент размещения заказа, как предполагалось ранее. Для большинства реальных ситуаций существует положительный срок выполнения заказа L (временное запаздывание) от момента его размещения до реальной поставки, как показано на рис. 7.2. В этом случае точка возобновления заказа имеет место, когда уровень запаса опускается до LD единиц.

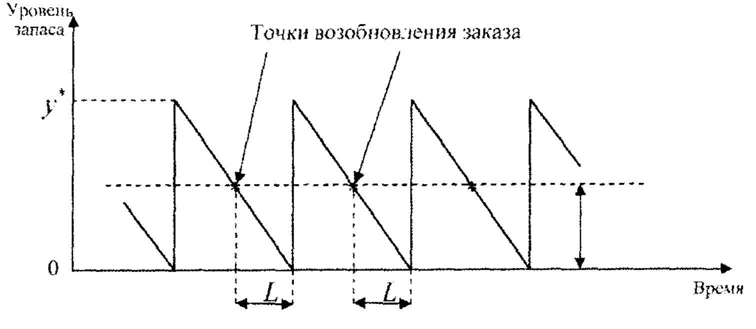


Рис. 7.2. Точки возобновления заказа в классической модели

На рис. 7.2 представлено изменение уровня запаса во времени при условии, что срок выполнения заказа L меньше продолжительности цикла заказа t_0^*, что в общем случае выполняется не всегда. В противном случае определяется эффективный срок L_e выполнения заказа в виде

$$L_e = L - n t_0^*,$$

где n — наибольшее целое, не превышающее L/t_0^*.

Такое решение оправдывается тем, что после n циклов (длиной t_0^* каждый) ситуация управления запасами становится такой же, как если бы интервал между размещением одного заказа и получением другого был равен L_e.

Следовательно, точка возобновления заказа имеет место при уровне запаса L_eD единиц продукции, и стратегия управления запасами может быть переформулирована следующим образом: *Заказывать* y^* *единиц продукции, как только уровень запаса опускается до* L_eD *единиц.*

Пример 7.1. Неоновые лампы в университетском городке заменяются с интенсивностью 100 шт. в день. Подразделение материального обеспечения городка заказывает эти лампы с определенной периодичностью. Стоимость размещения заказа на покупку ламп составляет 100 долл. Стоимость хранения лампы на складе оценивается в 0,02 долл. в день. Срок выполнения заказа от момента его размещения до реальной поставки равен 12 дней.

Требуется определить оптимальную стратегию заказа неоновых ламп.

▼По условию D = 100 шт. в день, K = 100 долл. за заказ, h = 0,02 долл. за хранение одной лампы в день, L = 12 дн.

Оптимальный размер заказа

$$y^* = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2\cdot 100\cdot 100}{0{,}02}} = 1000 \text{ ламп.}$$

Соответствующая длина цикла составляет

$$t_0^* = \frac{y^*}{D} = \frac{1000}{100} = 10 \text{ дн.}$$

Так как срок выполнения заказа L = 12 дней превышает продолжительность цикла t_0^* (10 дней), необходимо вычислить L_e.

Число целых циклов, заключенных в L, равно n = (наибольшее целое $\leq L/t_0^*$) = (наибольшее целое $\leq 12/10$) = 1.

Следовательно,

$$L_e = L - n t_0^* = 12 - 1 \cdot 10 = 2 \text{ дн.}$$

Поэтому точка возобновления заказа имеет место при уровне запаса

$$L_e D = 2 \cdot 100 = 200 \text{ неоновых ламп.}$$

Оптимальная стратегия заказа неоновых ламп может быть сформулирована следующим образом: *заказать 1000 ламп, как только уровень их запаса уменьшается до 200 единиц.*

Дневные расходы, связанные с содержанием запаса в соответствии с оптимальной стратегией, равны

$$TCU(y) = \frac{KD}{y} + h\left(\frac{y}{2}\right) = \frac{100 \cdot 100}{1000} + 0{,}02\left(\frac{1000}{2}\right) = 20 \text{ долл. в день. } \blacktriangle$$

Пример 7.2. Владелец магазина приобретает в течение года 1500 телевизоров для розничной продажи. Издержки хранения каждого телевизора равны 45 руб. в год. Издержки заказа — 150 руб. Количество рабочих дней в году равно 300, время выполнения заказа — 6 дней.

Определить: оптимальный размер заказа, точку возобновления запаса, годовые издержки.

$\blacktriangledown$ D = 1500 шт./год, h = 45 руб./год, K = 150 руб., T = 300 дн., L = 6 дн.

Оптимальный размер заказа

$$y^* = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2 \cdot 150 \cdot 1500}{45}} = 100 \text{ шт.}$$

Соответствующая длина цикла составляет

$$t_0^* = \frac{y^* T}{D} = \frac{100 \cdot 300}{1500} = 20 \text{ дн.}$$

Так как срок выполнения заказа $L = 6$ дн. не превышает продолжительность цикла t_0^* (20 дн.), то точка возобновления заказа имеет место при уровне запаса $LD/T = 6 \cdot 1500 : 300 = 30$ шт. Каждый раз, когда на складе останется 30 шт., подается заказ на 100 шт. Число циклов (количество заказов) за год равно $D/y^* = 1500/100 = 15$, а расстояние между циклами (заказами) — 20 рабочих дней.

Издержки, связанные с содержанием запаса в соответствии с оптимальной стратегией, равны

$$TCU(y) = \frac{KD}{y} + h\left(\frac{y}{2}\right) = \frac{150 \cdot 1500}{100} + 45 \cdot \left(\frac{100}{2}\right) = 4500 \text{ руб./год. } \blacktriangle$$

Пример 7.3. Объем продажи некоторого магазина составляет в год 500 упаковок супа в пакетах. Величина спроса равномерно распределяется в течение года. Цена покупки одного пакета равна 2 усл. ед. За доставку заказа владелец магазина должен заплатить 10 усл. ед. Время доставки заказа от поставщика составляет 12 рабочих дней (при 6-дневной рабочей неделе). По оценкам специалистов, издержки хранения в год составляют 0,4 усл. ед. за один пакет.

Определить: оптимальный размер заказа, точку возобновления запаса, годовые издержки. Известно, что магазин работает 300 дней в году.

▼$D = 500$ шт./год, $h = 0{,}4$ усл. ед./год, $K = 10$ усл. ед., $T = 300$ дн., $L = 12$ дн.

Оптимальный размер заказа

$$y^* = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2 \cdot 10 \cdot 500}{0{,}4}} = 158{,}11 \approx 158 \text{ шт.}$$

Соответствующая длина цикла составляет

$$t_0^* = \frac{y^* T}{D} = \frac{158 \cdot 300}{500} = 94{,}8 \approx 95 \text{ дн.}$$

Так как срок выполнения заказа $L = 12$ дней не превышает продолжительность цикла t_0^* (95 дней), то точка возобновления заказа имеет место при уровне запаса $LD/T = 12 \cdot 500 : 300 = 20$ шт. Каждый раз, когда на складе останется 20 пакетов, подается заказ на 158 пакетов. Подача нового заказа должна производиться через 95 дней.

Годовые издержки, связанные с содержанием запаса в соответствии с оптимальной стратегией, равны

$$TCU(y) = \frac{KD}{y} + h\left(\frac{y}{2}\right) = \frac{10 \cdot 500}{158} + 0,4 \cdot \left(\frac{158}{2}\right) = 63,25 \text{ усл. ед./год.} \ \blacktriangle$$

Упражнение 7.1. Потребность станкосборочного цеха в заготовках некоторого типа составляет 35 тыс. шт. в год. (1 год = = 365 дней). Дефицит не допускается. Издержки размещения заказа составляют 50 руб., издержки содержания одной заготовки в год равны 5 руб., среднее время реализации заказа равно 10 дней. Определить оптимальную партию поставки, периодичность возобновления поставок, точку размещения заказа, суммарные годовые затраты.

Ответ: заказать 837 шт. заготовок, как только уровень их запаса уменьшается до 125 шт. Подача нового заказа должна производиться через 9 дней. Годовые расходы составляют 4183,3 руб.

7.3. Модель экономичного размера заказа с разрывами цен

Пусть продукция может быть приобретена со скидкой, если объем заказа y превышает некоторый фиксированный уровень q; таким образом, стоимость единицы продукции c определяется как

$$c = \begin{cases} c_1, \text{если } y < q, \\ c_2, \text{если } y \geq q. \end{cases}$$

Затраты на приобретение продукции в единицу времени равны

$$\frac{cy}{t_0} = \frac{cy}{y / D} = Dc .$$

Запишем общие затраты в единицу времени

$$TCU(y) = Dc + \frac{KD}{y} + \frac{h}{2}y.$$

Необходимо выяснить, стоит ли воспользоваться скидкой.

Задача покупателя — выбор объема заказа, при котором общие расходы (подача заказа, закупка и хранение) будут минимальны.

Пример 7.4. Мебельный салон продает в год около 1000 спальных гарнитуров по цене 50 тыс. руб. Размещение одного заказа на поставку гарнитуров обходится в 40 тыс. руб. Годовая стоимость хранения гарнитура составляет 25 % его цены. Салон может получить у поставщика скидку в 3 %, если размер заказа составит не менее 200 гарнитуров. Следует ли салону воспользоваться этой скидкой?

▼D = 1000 ед., K = 40 тыс. руб./заказ.

Без скидки c = 50 тыс. руб./ед., $h = 0{,}25c = 0{,}25 \cdot 50 =$ $= 12{,}5$ тыс. руб./ед.

Оптимальный размер заказа

$$y = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2 \cdot 40 \cdot 1000}{12{,}5}} = 80 \text{ ед.}$$

Общие издержки равны

$$TCU_1(80) = 1000 \cdot 50 + \frac{40 \cdot 1000}{80} + \frac{12{,}5}{2} \cdot 80 = 51\ 000 \text{ тыс. руб./год.}$$

Со скидкой $c = 0{,}97 \cdot 50 = 48{,}5$ тыс. руб./ед., $h = 0{,}25c = 0{,}25 \times$ $\times\ 48{,}5 = 12{,}125$ тыс. руб./ед.

В этом случае оптимальный размер заказа равен

$$y = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2 \cdot 40 \cdot 1000}{12{,}125}} = 81 \text{ ед.}$$

Но скидка предоставляется, если объем заказа $q \geq 200$ ед., поэтому положим $y = 200$ ед.

Тогда общие издержки равны

$$TCU_2(200) = 1000 \cdot 48,5 + \frac{40 \cdot 1000}{200} + \frac{12,125}{2} \cdot 200 =$$
$$= 49\,912,5 \text{ тыс. руб./год.}$$

Поскольку общие издержки уменьшились, то следует воспользоваться скидкой, заказывая каждый раз 200 ед.

На рис. 7.3 показан выбор объема заказа при скидке на количество.

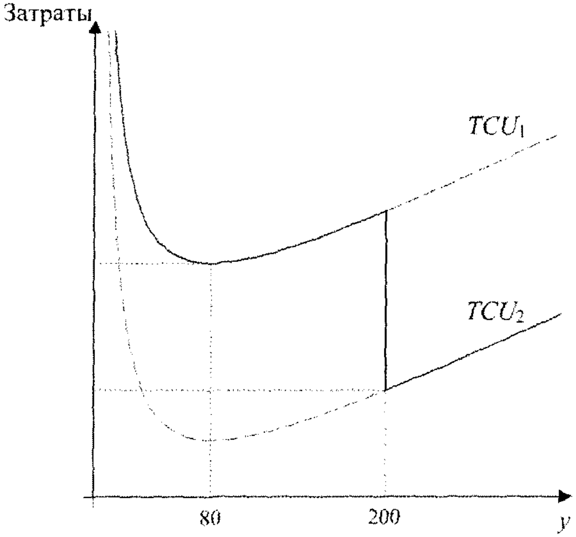


Рис. 7.3. Графики функций затрат для примера 7.4

Число циклов за год равно D/y = 1000/200 = 5, а интервал между заказами y/D = 200/1000 = 0,2 года = 73 дн. ▲

Пример 7.5. Магазин продает игрушечные гоночные машинки. В зависимости от размера заказа фирма предлагает скидки (табл. 7.1).

Таблица 7.1

Исходные данные для примера 7.5

Размер заказа, шт.	0–999	1000–1999	2000 и более
Цена со скидкой, руб.	5,00	4,80	4,75

Издержки заказа составляют 49 руб. Годовой спрос на машинки равен 5000 ед. Годовые издержки хранения в процентном отношении к цене составляют 20 %. Найдите размер заказа, минимизирующий общие издержки.

▼D = 5000 ед., K = 49 руб./заказ, c = 5,00; 4,80 и 4,75 руб./ед., h = 1,00; 0,96 и 0,95 руб./ед.

Оптимальный размер заказа для каждого вида цены

$$y_1 = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2 \cdot 49 \cdot 5000}{1,0}} = 700 \text{ ед.},$$

$$y_2 = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2 \cdot 49 \cdot 5000}{0,96}} = 714 \text{ ед.},$$

$$y_3 = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2 \cdot 49 \cdot 5000}{0,95}} = 718 \text{ ед.}$$

Так как y_1 находится в интервале (0; 1000), то его необходимо взять равным 700 ед. Оптимальный объем со скидкой y_2 меньше количества, необходимого для получения скидки, следовательно, его необходимо взять равным 1000 ед. Аналогично y_3 берем равным 2000 ед.

Рассчитаем общие издержки для каждого размера заказа и вида скидок, а затем выберем среди них наименьшее значение.

$$TCU_1(700) = 5000 \cdot 5 + \frac{49 \cdot 5000}{700} + \frac{1}{2} \cdot 700 = 25\ 700,0 \text{ руб.}$$

$$TCU_2(1000) = 5000 \cdot 4,8 + \frac{49 \cdot 5000}{1000} + \frac{0,96}{2} \cdot 1000 = 24\ 725,0 \text{ руб.}$$

$$TCU_3(2000) = 5000 \cdot 4,75 + \frac{49 \cdot 5000}{2000} + \frac{0,95}{2} \cdot 2000 = 24\ 822,5 \text{ руб.}$$

Получили, что заказ в размере 1000 игрушечных машинок будет минимизировать общие издержки.

На рис. 7.4 показан выбор объема заказа при скидке на количество.

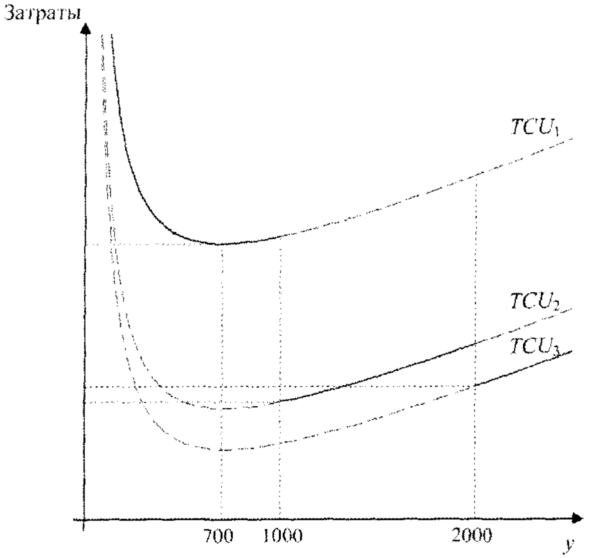


Рис. 7.4. Графики функций затрат для примера 7.5

Число циклов равно D/y = 5000/1000 = 5, а интервал между заказами y/D = 1000/5000 = 0,2 года = 73 дн. ▲

Упражнение 7.2. Магазин «Все для дома» закупает линолеум размером 2×3 м^2 в компании «Химические товары». В зависимости от размера заказа компания предлагает скидки (табл. 7.2).

Таблица 7.2

Исходные данные для упражнения 7.2

Размер заказа, кусков	9 или менее	10–49	50 и более
Цена 1 куска, тыс. руб.	18	17,5	17,25

Издержки заказа равны 45 тыс. руб. Годовые издержки хранения составляют 50 % от закупочной цены, годовой спрос на линолеум равен 100 кускам. Определите оптимальный размер заказа.

Ответ: 32 или 50 шт.

7.4. Модель с ограниченной вместимостью склада

Эта модель рассматривает задачу управления запасами n различных товаров, которые хранятся на одном складе ограниченной вместимости. Характер изменения запаса каждого товара в отдельности определяется функцией, показанной на рис. 7.1; предполагается, что дефицит отсутствует. Отличие от ранее рассмотренных моделей состоит в том, что товары конкурируют между собой за ограниченное складское пространство.

Определим для товара i, $i = 1, 2, n$, следующие параметры: D_i — интенсивность спроса, K_i — стоимость размещения заказа, h_i — стоимость хранения единицы товара в единицу времени, y_i — объем заказа, a_i — необходимое пространство для хранения единицы товара, A — максимальное складское пространство для хранения товаров n видов.

При отсутствии дефицита математическая модель сформулированной задачи имеет следующий вид:

$$\text{Минимизировать } TCU(y_1, y_2, \dots, y_n) = \sum_{i=1}^{n}\left(\frac{K_i D_i}{y_i} + \frac{h_i y_i}{2}\right)$$

при ограничениях

$$\sum_{i=1}^{n} a_i y_i \le A,$$

$$y_i > 0, i = \overline{1,n}.$$

Ограничение по вместимости склада должно удовлетворяться в форме равенства. Воспользуемся методом Лагранжа для определения оптимальных объемов заказа для задачи с ограничением.

Метод Лагранжа состоит в построении функции Лагранжа

$$L(\lambda, y_1, y_2, \dots, y_n) = TCU(y_1, y_2, \dots, y_n) - \lambda\left(\sum_{i=1}^{n} a_i y_i - A\right) =$$

$$= \sum_{i=1}^{n}\left(\frac{K_i D_i}{y_i} + \frac{h_i y_i}{2}\right) - \lambda\left(\sum_{i=1}^{n} a_i y_i - A\right),$$

где λ (< 0) — множитель Лагранжа.

Поскольку функция Лагранжа является выпуклой, оптимальные значения y_i и λ находятся из следующей системы уравнений, которые представляют собой необходимые условия экстремума функции Лагранжа:

$$\begin{cases} \dfrac{\partial L}{\partial y_i} = -\dfrac{K_i D_i}{y_i^2} + \dfrac{h_i}{2} - \lambda a_i = 0, \\ \dfrac{\partial L}{\partial \lambda} = \sum\limits_{i=1}^{n} a_i y_i + A = 0. \end{cases}$$

Второе уравнение показывает, что ограничение по вместимости склада в оптимальной точке должно удовлетворяться в форме равенства.

Из первого уравнения следует, что

$$y_i^* = \sqrt{\frac{2K_i D_i}{h_i - 2\lambda^* a_i}}\ .$$

Полученная формула показывает, что y_i^* зависит от оптимального значения λ^* множителя Лагранжа. Кроме того, при $\lambda^* = 0$ получаем задачу без ограничения. Если при этом выполняется ограничение

$$\sum_{i=1}^{n} a_i y_i \le A\ ,$$

то задача решена.

В случае невыполнения ограничения систему уравнений решаем численными методами. Для этого на каждом шаге, уменьшая λ, рассчитываем y_i^* и проверяем ограничение. Итерационную процедуру заканчиваем, когда будет выполнено ограничение.

Пример 7.6. Количество продаваемого товара в магазине составляет 1880 т, расходы по хранению единицы товара — 9,8 усл. ед./т, расходы по заказу и завозу одной партии товара — 29,7 усл. ед. Складская площадь магазина составляет 86 м^2, а для хранения 1 т товара требуется 1,2 м^2 складского помещения.

Требуется найти оптимальный объем одной поставки и общие затраты по управлению товарными запасами.

▼D = 1880 т, h = 9,8 усл. ед./т, K = 29,7 усл. ед., A = 86 м^2, a = 1,2 м^2/т.

Оптимальный объем заказа при $\lambda^* = 0$ (*без учета ограничения*) составляет

$$y^* = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2 \cdot 29,7 \cdot 1880}{9,8}} = 106,75 \text{ т.}$$

Требуемая площадь 1,2 · 106,75 = 128,1 м^2 для размещения данного объема заказа больше складской площади 86 м^2 магазина на 42,1 м^2, т. е. ограничение не выполняется.

Решаем задачу численным методом. Для этого в *Excel* записываем расчетную таблицу (рис. 7.5).

Задаем	Вычисляем	
λ	y	$ay - A$
0	106.75	42.10
-1	95,67	28,81
-2	87,46	18,95
-3	81,05	11,26
-4	75,87	5,04
-5	71,57	-0,11

Рис. 7.5. Расчет объема поставки в *Excel* для примера 7.6

Последний столбец таблицы показывает, что ограничение по вместимости склада в форме равенства выполняется при $\lambda = -5$. Таким образом, оптимальный объем одной поставки с учетом ограничений равен 71,57 т, при этом общие затраты составят

$$TCU(71,57) = \frac{29,7 \cdot 1880}{71,57} + \frac{9,8}{2} \cdot 71,57 = 1131 \text{ усл. ед.} \ \blacktriangle$$

Пример 7.7. Решим задачу управления запасами, исходные данные для которой приведены в табл. 7.3.

Таблица 7.3

Исходные данные для примера 7.7

Товар	*K*, долл.	*D*, ед. в день	*h*, долл.	*a*, кв. футов
1	10	2	0,3	1
2	5	4	0,1	1
3	15	4	0,2	1
A = 25 кв. футов				

▼Решаем задачу численным методом. Для этого в *Excel* записываем расчетную таблицу (рис. 7.6).

Задаем	Вычисляем			
λ	y_1	y_2	y_3	$\Sigma a_i y_i$
0	11.55	20,00	24,49	31,04
-0,1	8.94	11,55	17,32	12,81
-0,2	7.56	8,94	14,14	5,65
-0,3	6,67	7,56	12,25	1,47
-0,4	6,03	6,67	10,95	-1,35

Рис. 7.6. Расчет объема поставок в *Excel* для примера 7.7

Последний столбец таблицы показывает, что ограничение по вместимости склада в форме равенства выполняется при $\lambda = -0,4$. Таким образом, оптимальные объемы одной поставки с учетом ограничений равны 6,03; 6,67 и 10,95 ед. товаров 1, 2 и 3 соответственно, при этом общие затраты составят

$$TCU(6,03; 6,67; 10,95) = \left(\frac{10\cdot 2}{6,03} + \frac{0,3\cdot 6,03}{2}\right) +$$

$$+\left(\frac{5\cdot 4}{6,67} + \frac{0,1\cdot 6,67}{2}\right) + \left(\frac{15\cdot 4}{10,95} + \frac{0,2\cdot 10,95}{2}\right) = 14,1 \text{ долл.}$$ ▲

7.5. Модель производственных поставок

В основной модели (классическая модель экономичного размера заказа) предполагалось, что поступление товара на склад происходит мгновенно. Если товары поставляются с работающей производственной линии, необходимо модифицировать эту модель. В этом случае к параметрам основной модели (D — интенсивность спроса, K — затраты на оформление заказа, h — затраты на хранение) добавляется еще один (P — производительность производственной линии, причем $P > D$). Будем считать величины P, D заданными, постоянными и выражать в количестве единиц продукции в год. Величина y по-прежнему обозначает размер партии заказа.

На рис. 7.7 показано изменение уровня запаса.

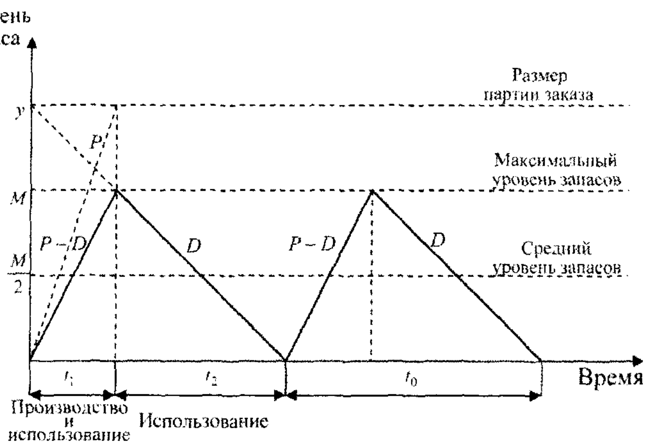


Рис. 7.7. График циклов изменения уровня запаса в случае производственных поставок

В течение времени t_1 продукция производится и используется одновременно, вследствие чего запас накапливается с интенсивностью ($P - D$). При достижении некоторой максимальной величины уровня запасов M производство останавливается. В течение времени t_2 продукция только используется.

Длина цикла $t_0 = t_1 + t_2$.

Максимальный уровень запаса M вычисляется по формуле $M = (P - D)t_1$. За время t_1 будет произведено $y = Pt_1$ продукции. Из последних двух равенств следует, что максимальный уровень запаса

$$M = \frac{(P-D)y}{P}.$$

В отличие от основной модели здесь максимальный уровень запаса M не совпадает с размером заказа y. Средний уровень запаса, как и в основной модели, равен половине максимального, т. е. $M/2$.

Суммарные издержки за рассматриваемый период можно представить в виде

$$TCU(y) = \frac{KD}{y} + \frac{(P-D)yh}{2P}.$$

Из условия минимизации суммарных издержек получается следующий оптимальный размер партии:

$$y^* = \sqrt{\frac{2KPD}{(P-D)h}}.$$

Соответствующие длина цикла $t_0 = y^*/D$, продолжительность поставки $t_1 = y^*/P$, число циклов за рассматриваемый период $n = D/y^*$.

Пример 7.8. Интенсивность равномерного спроса составляет 1000 ед. товара в год. Товар поставляется с конвейера, производительность которого 5000 ед. в год. Организационные издержки равны 10 усл. ед., издержки на хранение товара — 2 усл. ед./ед. Найти оптимальные параметры товарооборота.

▼$D = 1000$ ед., $P = 5000$ ед., $K = 10$ усл. ед., $h = 2$ усл. ед.

Оптимальный размер поставки:

$$y^* = \sqrt{\frac{2KPD}{(P-D)h}} = \sqrt{\frac{2 \cdot 10 \cdot 5000 \cdot 1000}{(5000-1000) \cdot 2}} = 112 \text{ ед.}$$

Суммарные издержки за год:

$$TCU(y) = \frac{KD}{y} + \frac{(P-D)yh}{2P} =$$

$$= \frac{10 \cdot 1000}{112} + \frac{(5000-1000) \cdot 112 \cdot 2}{2 \cdot 5000} \approx 179 \text{ усл. ед.}$$

Число циклов (поставок) за год равно $D/y^* = 1000/112 = 9$, длина цикла $y^*/D = 112/1000 = 0{,}112$ года $= 0{,}112 \cdot 365 = 41$ дн., продолжительность поставки $y^*/P = 112/5000 = 0{,}0224$ года $= 8$ дн. ▲

Упражнение 7.3. Компания выпускает электрические ножи. В среднем она может производить 150 ножей в день. Дневной спрос на ножи равен 40 шт. Фиксированные издержки производства составляют 100 руб., издержки хранения — 8 руб. за нож в год. В году 250 рабочих дней. Определить оптимальный размер производственного заказа и суммарные издержки за год.

Ответ: 583,9 шт., 3425,4 руб.

7.6. Модель оптимального размера с дефицитом

Во многих случаях при закупке товаров у поставщиков экономически выгоднее бывает допустить отсутствие товаров в течение какого-либо промежутка времени, чем поддерживать их постоянное наличие. Для управления запасами в таких системах используется модель, в которой в течение определенного времени запас отсутствует, т. е. допускаются дефицит запасов и убытки из-за дефицита запасов. При этом возможны два подхода:

1) полученная новая продукция не идет на выполнение заявок на товар во время его отсутствия;

2) часть полученной новой продукции идет на погашение всех заявок, поступивших во время отсутствия товара.

Случай невыполнения заявок. В рамках этого подхода спрос, возникающий на товары в течение времени дефицита, не удовлетворяется. Поэтому максимальный уровень запасов совпадает с объемом заказа (рис. 7.8).

На графике периоды дефицита условно изображаются ниже оси времени.

Пусть c_b — годовая стоимость отсутствия запаса единицы продукции.

Цикл заказа t_0 состоит из времени потребления запасов t_1 и времени отсутствия запасов t_2. Таким образом, в течение цикла запаса t_0 на складе хранится следующее количество запасов:

$$\bar{y} = \frac{yt_1}{2t_0}.$$

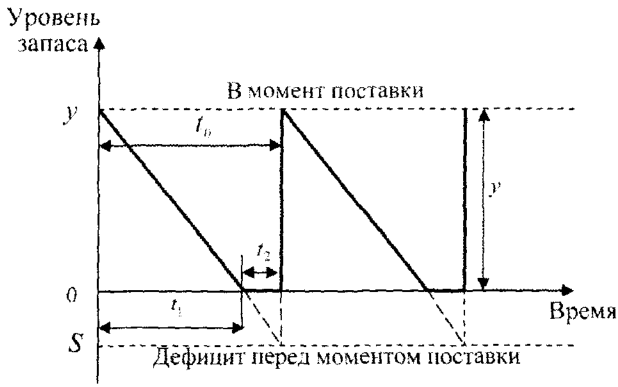


Рис. 7.8. График циклов изменения уровня запаса в случае невыполнения заявок, поступивших во время отсутствия товара

Аналогично определяется средний уровень дефицита в течение времени t_2 по формуле

$$\bar{S} = \frac{St_2}{2t_0}.$$

В условиях известного и линейного спроса D за период (год) количество заказанных партий товара будет составлять $n = D/(y+S)$, а интервал заказа $t_0 = (y+S)/D$. Таким образом, можно определить t_1 и t_2:

$$t_1 = \frac{y}{D},\ t_2 = \frac{S}{D}.$$

Суммарные затраты по управлению запасами включают три слагаемых:

- годовую сумму затрат на размещение заказов

$$nK = \frac{KD}{y+S};$$

- годовую сумму затрат на хранение запасов

$$h\bar{y} = h\frac{y}{2}\cdot\frac{y}{D}:\frac{y+S}{D} = \frac{hy^2}{2(y+S)};$$

• годовую сумму издержек из-за отсутствия запасов

$$c_b\overline{S} = c_b\frac{St_2}{2t_0} = c_b\frac{S}{2}\cdot\frac{S}{D}:\frac{y+S}{D} = \frac{c_bS^2}{2(y+S)}.$$

Годовые суммарные затраты по управлению запасами составляют

$$TCU(y,S) = \frac{KD}{y+S} + \frac{hy^2}{2(y+S)} + \frac{c_bS^2}{2(y+S)} \to \min.$$

Минимальное значение $TCU(y,S)$ можно найти, используя необходимое условие минимума (частные производные функции $TCU(y,S)$ по y и S должны обращаться в ноль в точке минимума). В результате из системы уравнений получим:

• оптимальный размер заказа $y^* = \sqrt{\frac{2KD}{h}}\sqrt{\frac{c_b}{c_b+h}}$;

• максимальный размер дефицита $S = \sqrt{\frac{2KD}{c_b}}\sqrt{\frac{h}{c_b+h}}$.

Пример 7.9. Годовой спрос 500 ед., стоимость подачи заказов 40 усл. ед./зак., издержки хранения одной единицы 5 усл. ед./год, годовая стоимость отсутствия запасов 100 усл. ед./ед. Модель с дефицитом (заявки не выполняются). Найти издержки.

▼ D = 500 ед., K = 40 усл. ед./зак., h = 5 усл. ед./год, c_b = 100 усл. ед./ед.

Сравним две модели: основную и с дефицитом (заявки не выполняются).

Основная модель:

Оптимальный размер заказа:

$$y^* = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2\cdot 40\cdot 500}{5}} \approx 89 \text{ ед.}$$

Суммарные издержки:

$$TCU(y) = \frac{KD}{y} + h\left(\frac{y}{2}\right) = \frac{40\cdot 500}{89} + 5\cdot\left(\frac{89}{2}\right) \approx 447 \text{ усл. ед./год.}$$

Модель с дефицитом:

Оптимальный размер заказа:

$$y^* = \sqrt{\frac{2KD}{h}}\sqrt{\frac{c_h}{c_h+h}} = \sqrt{\frac{2\cdot 40\cdot 500}{5}}\sqrt{\frac{100}{100+5}} \approx 87 \text{ ед.}$$

Максимальный размер дефицита:

$$S = \sqrt{\frac{2KD}{c_h}}\sqrt{\frac{h}{c_h+h}} = \sqrt{\frac{2\cdot 40\cdot 500}{100}}\sqrt{\frac{5}{100+5}} \approx 4 \text{ ед.}$$

Суммарные издержки:

$$TCU = \frac{KD}{y+S} + \frac{hy^2}{2(y+S)} + \frac{c_h S^2}{2(y+S)} = \frac{40\cdot 500}{87+4} +$$

$$+ \frac{5\cdot 87^2}{2\cdot(87+4)} + \frac{100\cdot 4^2}{2\cdot(87+4)} \approx 437 \text{ усл. ед./год.}$$

Таким образом, в модели с дефицитом годовые издержки меньше. ▲

Упражнение 7.4. Годовой спрос 600 ед., стоимость подачи заказов 50 усл. ед./зак., издержки хранения одной единицы 6 усл. ед./год, годовая стоимость отсутствия запасов 110 усл. ед./ед. Сравнить издержки при использовании основной модели и в модели с дефицитом (заявки не выполняются).

Ответ: лучше модель с дефицитом.

Случай выполнения заявок. В рамках этого подхода дефицит товаров по заказам покупателей восполняется из следующей поставки. В этом случае максимальная величина запасов равна разнице объема заказа y и максимального неудовлетворенного спроса S, возникающего в течение времени дефицита $(y - S)$ (рис. 7.9).

На графике периоды дефицита условно изображаются ниже оси времени.

Пусть c_h — годовая стоимость отсутствия запаса единицы продукции.

Цикл заказа t_0 состоит из времени потребления запасов t_1 и времени отсутствия запасов t_2. Таким образом, в течение цикла t_0 на складе хранится следующее количество запасов:

$$\bar{y} = \frac{(y-S)t_1}{2t_0}.$$

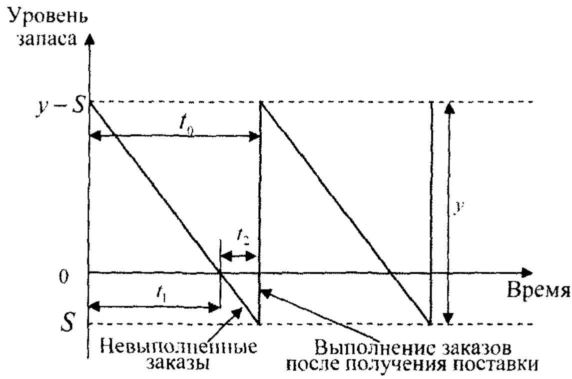


Рис. 7.9. График циклов изменения уровня запаса в случае выполнения заявок, поступивших во время отсутствия товара

Аналогично определяется средний уровень дефицита в течение времени t_2 по формуле:

$$\bar{S} = \frac{St_2}{2t_0}.$$

В условиях известного и линейного спроса D за период (год) количество заказанных партий товара будет составлять $n = D/y$, а интервал заказа $t_0 = y/D$. Таким образом, можно определить t_1 и t_2:

$$t_1 = \frac{y-S}{D}, \; t_2 = \frac{S}{D}.$$

Суммарные затраты по управлению запасами включают три слагаемых:

• годовую сумму затрат на размещение заказов

$$nK = \frac{KD}{y};$$

• годовую сумму затрат на хранение запасов

$$h\bar{y} = h\frac{(y-S)}{2} \cdot \frac{y-S}{D} : \frac{y}{D} = \frac{h(y-S)^2}{2y};$$

• годовую сумму издержек из-за отсутствия запасов

$$c_b\overline{S} = c_b\frac{St_2}{2t_0} = c_b\frac{S}{2}\cdot\frac{S}{D}:\frac{y}{D} = \frac{c_bS^2}{2y}.$$

Годовые суммарные затраты по управлению запасами составляют

$$TCU(y,S) = \frac{KD}{y} + \frac{h(y-S)^2}{2y} + \frac{c_bS^2}{2y} \to \min.$$

Минимальное значение $TCU(y,S)$ можно найти, используя необходимое условие минимума (частные производные функции $TCU(y,S)$ по y и S должны обращаться в ноль в точке минимума). В результате из системы уравнений получим:

• оптимальный размер заказа $y^* = \sqrt{\frac{2KD}{h}}\sqrt{\frac{c_b+h}{c_b}}$;

• максимальный размер дефицита $S = \sqrt{\frac{2KD}{c_b}}\sqrt{\frac{h}{c_b+h}}$.

Пример 7.10. Годовой спрос 3000 ед., стоимость подачи заказов 25 усл. ед./зак., издержки хранения одной единицы 120 усл. ед./год, годовая стоимость отсутствия запасов 225 усл. ед./ед. Модель с дефицитом (заявки выполняются). Найти издержки.

▼D = 3000 ед., K = 25 усл. ед./зак., h = 120 усл. ед./год, c_b = 225 усл. ед./ед.

Сравним две модели: основную и с дефицитом (заявки выполняются).

Основная модель:

Оптимальный размер заказа:

$$y^* = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2\cdot 25\cdot 3000}{120}} \approx 35 \text{ ед.}$$

Суммарные издержки:

$$TCU(y) = \frac{KD}{y} + h\left(\frac{y}{2}\right) = \frac{25\cdot 3000}{35} + 120\cdot\left(\frac{35}{2}\right) \approx 4243 \text{ усл. ед./год.}$$

Модель с дефицитом:

Оптимальный размер заказа:

$$y^* = \sqrt{\frac{2KD}{h}}\sqrt{\frac{c_b + h}{c_b}} = \sqrt{\frac{2 \cdot 25 \cdot 3000}{120}}\sqrt{\frac{225+120}{225}} \approx 44 \text{ ед.}$$

Максимальный размер дефицита:

$$S = \sqrt{\frac{2KD}{c_b}}\sqrt{\frac{h}{c_b + h}} = \sqrt{\frac{2 \cdot 25 \cdot 3000}{225}}\sqrt{\frac{120}{225+120}} \approx 15 \text{ ед.}$$

Суммарные издержки:

$$TCU = \frac{KD}{y} + \frac{h(y-S)^2}{2y} + \frac{c_b S^2}{2y} = \frac{25 \cdot 3000}{44} +$$

$$+ \frac{120 \cdot (44-15)^2}{2 \cdot 44} + \frac{225 \cdot 15^2}{2 \cdot 44} = 3427 \text{ усл. ед./год.}$$

Таким образом, в модели с дефицитом годовые издержки меньше. ▲

Упражнение 7.5. Годовой спрос 2000 ед., стоимость подачи заказов 20 усл. ед./зак., издержки хранения одной единицы 100 усл. ед./год, годовая стоимость отсутствия запасов 220 усл. ед./ед. Модель с дефицитом (заявки выполняются). Сравнить издержки при использовании основной модели и в модели с дефицитом (заявки выполняются).

Ответ: лучше модель с дефицитом.

Глава 8.
МОДЕЛИ СЕТЕВОГО ПЛАНИРОВАНИЯ И УПРАВЛЕНИЯ

8.1. Основные понятия сетевой модели

Сетевая модель — это план выполнения некоторого комплекса взаимосвязанных работ, заданного в форме сети, графическое изображение которой называется *сетевым графиком.* Отличительной особенностью сетевой модели является четкое определение всех временных взаимосвязей предстоящих работ. Главными элементами сетевого графика являются работа и событие.

Работа — это любые процессы, сопровождающиеся затратами ресурсов или времени и приводящие к определенным результатам. *Фиктивная работа* — это связь между результатами работ, не требующая затрат времени и ресурсов. На сетевом графике работы обозначаются стрелками, а фиктивные работы — пунктирными стрелками.

Событие — это факт окончания всех входящих в него работ. Событие не имеет протяженности во времени. События на сетевом графике (или на графе) изображаются кружками (вершинами графа). Ни одна выходящая из данного события работа не может начаться до окончания всех работ, входящих в это событие.

Среди событий сетевой модели выделяют исходное и завершающее события. *Исходное событие* не имеет предшествующих работ (с него начинается выполнение проекта). *Завершающее событие* не имеет последующих работ (на нем заканчивается выполнение проекта).

После построения сетевого графика необходимо оценить продолжительность выполнения каждой работы и выделить работы, которые определяют завершение проекта в целом. Нужно оценить потребность каждой работы в ресурсах и пересмотреть план с учетом обеспечения ресурсами.

При построении сетевого графика необходимо соблюдать ряд правил:

• не должно быть тупиковых событий (т. е. не имеющих последующих событий), кроме завершающего события;

• не должно быть висячих событий (т. е. не имеющих предшествующих событий), кроме исходного события;

• не должно быть замкнутых циклов;

• любые два события должны быть непосредственно связаны не более чем одной работой-стрелкой;

• если два события связаны более чем одной работой, рекомендуется ввести дополнительное событие и фиктивную работу (рис. 8.1).

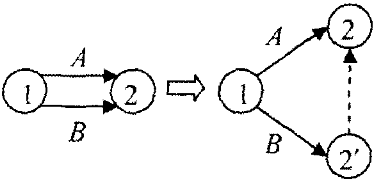


Рис. 8.1. Модификация сетевого графика, содержащего события, связанные двумя работами

8.2. Метод критического пути

Метод критического пути (*Critical Path Method — CPM*) используется для контроля сроков выполнения проекта. При реализации проекта составляется график выполнения проекта. Для этого проводятся специальные вычисления, в результате чего:

• получают общую длительность выполнения проекта;

• производят разделение множества процессов, составляющих проект, на критические и некритические. Процесс является критическим, если он не имеет «зазора» по времени своего начала и завершения. Таким образом, чтобы весь проект завершился без задержек, необходимо, чтобы все критические процессы начинались и заканчивались в строго определенное время. Для некритического процесса возможен некоторый «дрейф» времени его начала, но в определенных границах, когда время его начала не влияет на длительность выполнения всего проекта.

Для проведения необходимых вычислений определим *событие* как точку на временной оси, где завершается один процесс и начинается другой. В терминах сети событие — это сетевой узел.

Одно их важнейших понятий сетевого графика — понятие пути.

Путь — последовательность взаимосвязанных работ, ведущая из одной вершины проекта в другую вершину. *Длина пути* — суммарная продолжительность выполнения всех работ пути.

Самый продолжительный путь сетевого графика от исходного события к завершающему называется *критическим*. Все события и работы критического пути называются *критическими*. Продолжительность критического пути и определяет срок выполнения проекта. Критических путей на сетевом графике может быть несколько.

Рассмотрим основные временные параметры сетевых графиков.

Обозначим $t(i, j)$ — продолжительность работы с начальным событием i и конечным событием j.

Ранний срок $t_р(j)$ совершения события j — это самый ранний момент, к которому завершаются все работы, предшествующие этому событию. Правило вычисления:

$$t_р(j) = \max_i \left\{ t_р(i) + t(i, j) \right\},$$

где максимум берется по всем событиям i, непосредственно предшествующим событию j (соединены стрелками).

Поздний срок $t_п(i)$ совершения события i — это такой предельный момент, после которого остается ровно столько времени, сколько необходимо для выполнения всех работ, следующих за этим событием. Правило вычисления:

$$t_п(i) = \min_j \left\{ t_п(j) - t(i, j) \right\},$$

где минимум берется по всем событиям j, непосредственно следующим за событием i.

Резерв $R(i)$ события i показывает, на какой предельно допустимый срок может задержаться свершение события i без нарушения срока наступления завершающего события:

$$R(i) = t_{\text{п}}(i) - t_{\text{р}}(i).$$

Критические события резервов не имеют.

При расчете сетевого графика каждый круг, изображающий событие, делят диаметрами на четыре сектора (рис. 8.2).

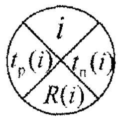


Рис. 8.2. Элемент сетевого графика

Вычисление критического пути включает два этапа (прохода). При **проходе вперед** вычисляются *самые ранние времена* наступления событий, а при **проходе назад** — *самые поздние времена* наступления тех же событий.

Процесс (i, j) будет *критическим*, если выполняются три условия:

1. $t_{\text{п}}(i) = t_{\text{р}}(i)$.
2. $t_{\text{п}}(j) = t_{\text{р}}(j)$
3. $t_{\text{п}}(j) - t_{\text{п}}(i) = t_{\text{р}}(j) - t_{\text{р}}(i) = t(i, j)$.

Если эти условия не выполняются, то процесс *некритический.*

Критические процессы должны образовывать непрерывный путь через всю сеть от начального события до конечного.

Пример 8.1. Рассмотрим сеть проекта, представленную данными табл. 8.1.

Таблица 8.1

Исходные данные для примера 8.1

Работа	Непосредственный предшественник	Продолжительность работы, дн.
A	—	5
B	—	6
C	*A*	3

Окончание табл. 8.1

Работа	Непосредственный предшественник	Продолжительность работы, дн.
D	*A*	8
E	*B, C*	2
F	*B, C*	11
G	*E*	12
H	*D*	1

Найдем критический путь для сети проекта.

▼Рисуем сетевой график (рис. 8.3).

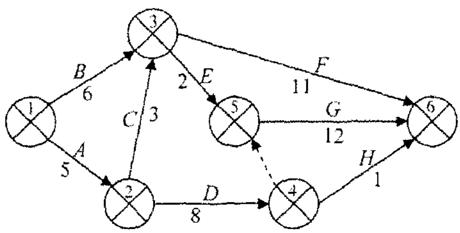


Рис. 8.3. Сетевой график для примера 8.1

Проход вперед. При вычислении $t_р(i)$ перемещаемся по сетевому графику от исходного события 1 к завершающему событию 6.

Событие 1. Полагаем $t_р(1) = 0$.

Событие 2. $t_р(2) = t_р(1) + t(1, 2) = 0 + 5 = 5$.

Событие 3. $t_р(3) = \max\begin{Bmatrix} t_р(1) + t(1,3) \\ t_р(2) + t(2,3) \end{Bmatrix} = \max\begin{Bmatrix} 0+6 \\ 5+3 \end{Bmatrix} = \mathbf{8}.$

Событие 4. $t_р(4) = t_р(2) + t(2, 4) = 5 + 8 = 13$.

Событие 5. $t_р(5) = \max\begin{Bmatrix} t_р(3) + t(3,5) \\ t_р(4) + t(4,5) \end{Bmatrix} = \max\begin{Bmatrix} 8+2 \\ 13+0 \end{Bmatrix} = 13.$

Событие 6. $t_р(6) = \max\begin{Bmatrix} t_р(3) + t(3,6) \\ t_р(4) + t(4,6) \\ t_р(5) + t(5,6) \end{Bmatrix} = \max\begin{Bmatrix} 8+11 \\ 13+1 \\ 13+12 \end{Bmatrix} = 25.$

Таким образом, расчеты показывают, что проект можно выполнить за 25 дней.

Проход назад. При вычислении $t_п(i)$ перемещаемся от завершающего события 6 к исходному событию 1 по сетевому графику против стрелок.

Событие 6. Полагаем $t_п(6) = t_р(6) = 25$.

Событие 5. $t_п(5) = t_п(6) - t(5, 6) = 25 - 12 = 13$.

Событие 4. $t_п(4) = \min\begin{Bmatrix} t_п(6) - t(4,6) \\ t_п(5) - t(4,5) \end{Bmatrix} = \min\begin{Bmatrix} 25-1 \\ 13-0 \end{Bmatrix} = 13.$

Событие 3. $t_п(3) = \min\begin{Bmatrix} t_п(6) - t(3,6) \\ t_п(5) - t(3,5) \end{Bmatrix} = \min\begin{Bmatrix} 25-11 \\ 13-2 \end{Bmatrix} = 11.$

Событие 2. $t_п(2) = \min\begin{Bmatrix} t_п(4) - t(2,4) \\ t_п(3) - t(2,3) \end{Bmatrix} = \min\begin{Bmatrix} 13-8 \\ 11-3 \end{Bmatrix} = 5.$

Событие 1. $t_п(1) = \min\begin{Bmatrix} t_п(3) - t(1,3) \\ t_п(2) - t(1,2) \end{Bmatrix} = \min\begin{Bmatrix} 11-6 \\ 5-5 \end{Bmatrix} = 0.$

Вычисления без ошибок всегда приводят к результату $t_п(1) = 0$.

Результаты вычислений, выполняемых при проходах вперед и назад, показаны на рис. 8.4.

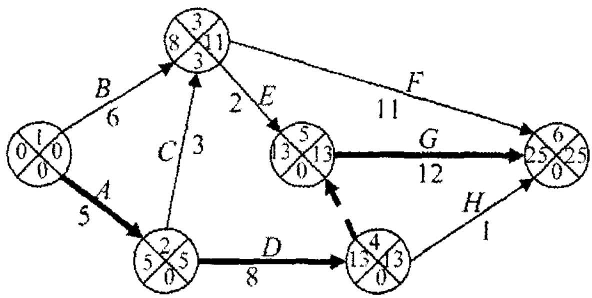


Рис. 8.4. Рассчитанный сетевой график для примера 8.1

Правила определения критических процессов показывают, что критический путь составляют процессы 1 → 2 → 4 → 5 → 6, т. е. этот путь проходит от начального узла 1 до конечного узла 6. Сумма длительности критических процессов (1, 2), (2, 4), (4, 5) и (5, 6) равна длительности всего проекта (т. е. 25 дней).

Отметим, что процесс (4, 6) удовлетворяет первым двум условиям критического пути ($t_п(4) = t_р(4) = 13$ и $t_п(6) = t_р(6) = 25$), но не удовлетворяет третьему условию ($t_р(6) - t_р(4) \neq t(4, 6)$. Поэтому данный процесс не является критическим. ▲

Упражнение 8.1. Рассмотрим сеть проекта, представленную данными табл. 8.2.

Таблица 8.2

Исходные данные для упражнения 8.1

Работа	**Непосредственный предшественник**	**Продолжительность работы, нед.**
A	—	5
B	—	3
C	*A*	7
D	*A*	6
E	*B*	7
F	*D*, *E*	3
G	*D*, *E*	10
H	*C*, *F*	8

Найти критический путь. Сколько времени потребуется для завершения проекта? Можно ли отложить выполнение работы *D* без отсрочки проекта в целом? На сколько недель можно отложить выполнение работы *C* без отсрочки завершения проекта в целом?

▼ *Указание.* Сетевой график, построенный по данным табл. 8.2, представлен на рис. 8.5.

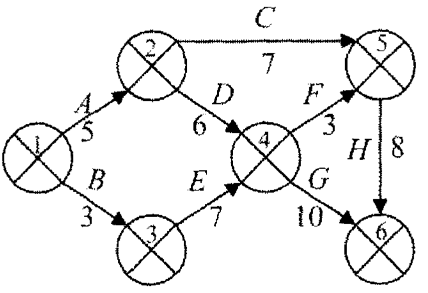


Рис. 8.5. Сетевой график для упражнения 8.1

Результаты вычислений, выполненные при проходах вперед и назад, показаны на рис. 8.6.

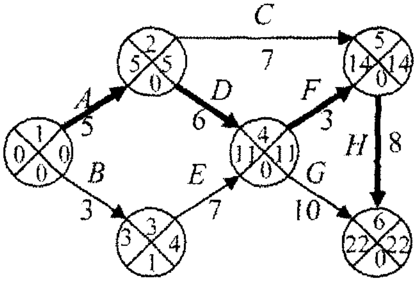


Рис. 8.6. Рассчитанный сетевой график для упражнения 8.1

Критический путь $1 \to 2 \to 4 \to 5 \to 6$ отмечен на рис. 8.6 жирными стрелками. Для завершения проекта потребуется 22 недели. Работа $D = (2, 4)$ расположена на критическом пути. Поэтому ее нельзя отложить без отсрочки завершения проекта в целом. Работа $C = (2, 5)$ не расположена на критическом пути, ее можно задержать на $t_п(5) - t_р(2) - t(2, 5) = 14 - 5 - 7 = 2$ нед. ▲

Построение предварительного графика

Построение предварительного временного графика выполнения проекта покажем на результатах примера 8.1.

▼Предварительный временной график проекта можно начертить, используя максимальные интервалы выполнения каждого процесса. Пара величин $(t_р(i), t_п(j))$ ограничивает максимальный интервал времени, в течение которого может выполняться процесс (i, j). В результате получим график, представленный на рис. 8.7.

Критические процессы (показаны на рис. 8.7 сплошными линиями) располагаются последовательно друг за другом без временных зазоров и перекрытий. Таким образом, их суммарная длительность равна длительности выполнения всего проекта (в данном случае 25 дней).

Некритические процессы (показаны на рис. 8.7 пунктирными линиями) представлены максимальными интервалами выпол-

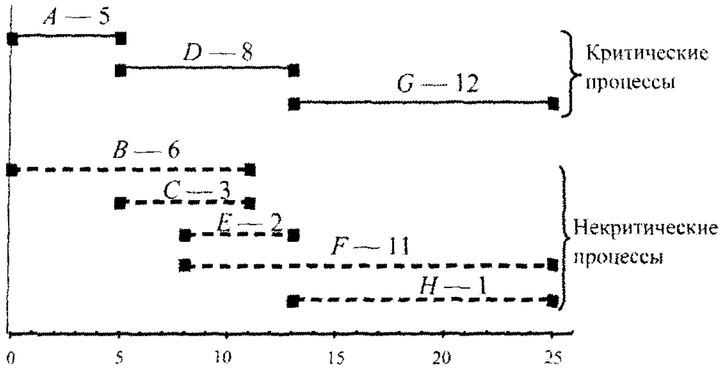


Рис. 8.7. Время выполнения процессов примера 5.1

нения, которые превышают реальную длительность выполнения этих процессов. Поэтому необходимо каким-то образом определиться с началом выполнения этих процессов. ▲

Определение запасов времени

Если выполнение какой-либо критической работы будет задержано на некоторый срок, то это вызовет *запаздывание выполнения* всего комплекса работ на тот же срок. Чтобы ускорить выполнение комплекса, необходимо сократить сроки выполнения критических работ. Некритические работы допускают некоторое запаздывание с их выполнением, и это не вызывает задержки срока реализации всего комплекса работ. Чтобы определить время, на которое можно задержать выполнение некритических работ, вводятся понятия резервов времени событий и работ, которые, в свою очередь, выражаются через ранние и поздние сроки свершения событий. Критические работы, как и критические события, резервов времени не имеют.

Полный резерв времени $R_п(i, j)$ работы (i, j) — это максимальное время, на которое можно задержать начало работы или увеличить продолжительность ее выполнения, не изменяя срока завершения всего комплекса работ. Вычисляется он по формуле

$$R_п(i, j) = t_п(j) - t_р(i) - t(i, j).$$

Если на некоторой работе использовать ее полный резерв, то путь, проходящий через эту работу, станет критическим. Полный резерв времени любой работы на этом пути станет равным нулю.

Свободный резерв времени $R_с(i, j)$ работы (i, j) — это максимальный запас времени, на который можно отложить начало работы или (если она началась в свой ранний срок) увеличить продолжительность ее выполнения при условии, что ранние сроки начала всех последующих работ не изменятся. Вычисляется он по формуле

$$R_с(i, j) = t_р(j) - t_р(i) - t(i, j).$$

Свободный резерв — это время, которое можно дополнительно выделить для выполнения работы без введения дополнительных ограничений на время выполнения последующих работ.

По определению $R_с(i, j) \le R_п(i, j)$.

Заметим, что только критические процессы должны иметь нулевой полный резерв времени. Когда полный резерв равен нулю, свободный резерв также должен быть равен нулю. Однако обратное неверно, поскольку свободный резерв некритического процесса также может быть нулевым.

Правила. Для некритического процесса (i, j):

1) если $R_с(i, j) = R_п(i, j)$, то данный процесс может выполняться в любое время между его ранним началом $t_р(i)$ и поздним окончанием $t_п(j)$ без нарушения отношений следования;

2) если $R_с(i, j) < R_п(i, j)$, то без нарушения отношений следования данный процесс может начаться со сдвигом, не превышающим $R_с(i, j)$, относительно самого раннего момента начала процесса $t_р(i)$. Сдвиг начала процесса на величину времени, превышающую $R_с(i, j)$ (но не более $R_п(i, j)$), должен сопровождаться равным сдвигом относительно $t_р(j)$ всех процессов, начинающихся с события j.

Пример 8.2. Определим запасы времени для некритических процессов в сети проекта примера 8.1.

▼Полные и свободные резервы времени представлены в табл. 8.3.

Правило 1 применимо только к процессам E, F и H, поскольку для них $R_с(i, j) = R_п(i, j)$, поэтому они могут выполняться в любое время внутри своих максимальных интервалов времени выполнения.

Таблица 8.3

Полные и свободные резервы времени для некритических процессов в сети проекта примера 8.1, дн.

Работа (i, j)	$t(i, j)$	$t_р(i)$	$t_р(j)$	$t_п(j)$	Резервы времени	
					$R_п(i, j) = t_п(j) - t_р(i) - t(i, j)$	$R_с(i, j) = t_р(j) - t_р(i) - t(i, j)$
$B(1, 3)$	6	0	8	11	$11 - 0 - 6 = 5$	$8 - 0 - 6 = 2$
$C(2, 3)$	3	5	8	11	$11 - 5 - 3 = 3$	$8 - 5 - 3 = 0$
$E(3, 5)$	2	8	13	13	$13 - 8 - 2 = 3$	$13 - 8 - 2 = 3$
$F(3, 6)$	11	8	25	25	$25 - 8 - 11 = 6$	$25 - 8 - 11 = 6$
$H(4, 6)$	1	13	25	25	$25 - 13 - 1 = 11$	$25 - 13 - 1 = 11$

Правило 2 следует применять только к процессам B и C, поскольку для них $R_с(i, j) < R_п(i, j)$.

Рассмотрим процесс B. Поскольку для этого процесса $R_п = 5$ дн., он может начаться в любой день из интервала 0–5 дн. от начала выполнения всего проекта (см. рис. 8.7). Но если $R_с = 2$ дн., то, если процесс B начнется в 0-й, 1-й или 2-й день от начала выполнения проекта, это не окажет никакого влияния на последующие процессы E и F. Однако, если процесс B начнется в $(2 + d)$-й день $(2 + d \leq 5)$, начало выполнения процессов E и F необходимо сдвинуть от самого раннего срока их начала (8-й день от начала выполнения проекта) на величину, не меньше d; только при таком условии не нарушатся отношения следования между процессами B, E и F.

Для процесса C имеем $R_с = 0$. Это означает, что любой сдвиг начала выполнения этого процесса должен сопровождаться таким же (не меньшим) сдвигом начала выполнения процессов E и F. ▲

Упражнение 8.2. Определить резервы работ для сетевого графика, представленного на рис. 8.8.

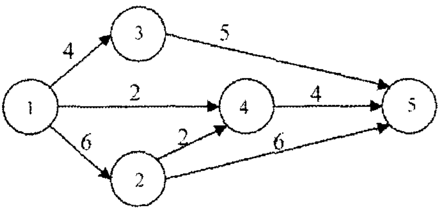


Рис. 8.8. Сетевой график для упражнения 8.2

▼*Указание.* Результаты вычислений, выполненных при проходах вперед и назад, показаны на рис. 8.9.

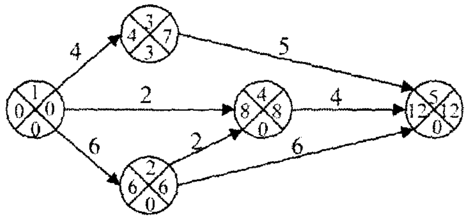


Рис. 8.9. Рассчитанный сетевой график для упражнения 8.2

Составляем табл. 8.4, в которой значения первых пяти колонок берем из сетевого графика, а остальные колонки просчитаем по этим данным.

Таблица 8.4

Полные и свободные резервы времени для сети из упражнения 8.2, дн.

Работа (i, j)	$t(i, j)$	$t_р(i)$	$t_р(j)$	$t_п(j)$	Резервы времени	
					$R_п(i, j) = t_п(j) - t_р(i) - t(i, j)$	$R_с(i, j) = t_р(j) - t_р(i) - t(i, j)$
(1, 2)	6	0	6	6	6 – 0 – 6 = 0	6 – 0 – 6 = 0
(1, 3)	4	0	4	7	7 – 0 – 4 = 3	4 – 0 – 4 = 0
(1, 4)	2	0	8	8	8 – 0 – 2 = 6	8 – 0 – 2 = 6
(2, 4)	2	6	8	8	8 – 6 – 2 = 0	8 – 6 – 2 = 0
(2, 5)	6	6	12	12	12 – 6 – 6 = 0	12 – 6 – 6 = 0
(3, 5)	5	4	12	12	12 – 4 – 5 =3	12 – 4 – 5 =3
(4, 5)	4	8	12	12	12 – 8 – 4 = 0	12 – 8 – 4 = 0

Критические работы (работы с нулевыми резервами): (1, 2), (2,4), (2, 5), (4, 5). У нас два критических пути: 1 → 2 → 5 и 1 → → 2 → 4 → 5. ▲

8.3. Стоимость проекта. Оптимизация сетевого графика

Стоимость выполнения каждой работы плюс дополнительные расходы определяют стоимость проекта. С помощью дополнительных ресурсов можно добиться сокращения времени выполнения критических работ. Тогда стоимость этих работ возрастет, но общее время выполнения проекта уменьшится, что может привести к снижению общей стоимости проекта. Предполагается, что работы можно выполнить либо в стандартные, либо в минимальные сроки, но не в промежутке между ними.

Пример 8.3. Рассмотрим сеть проекта, представленную данными табл. 8.5.

Таблица 8.5

Исходные данные для примера 8.3

Работа	Непосредственный предшественник	Стандартное время, дн.	Минимальное время, дн.	Затраты на работы, тыс. руб.	
				при стандартном времени	при минимальном времени
A	—	3	2	800	1400
B	—	2	1	1200	1900
C	*A*	5	3	2000	2800
D	*B*	5	3	1500	2300
E	*C*, *D*	6	4	1800	2800
F	*C*, *D*	2	1	600	1000
G	*F*	2	1	500	1000

Минимизируем общее время выполнения проекта с наименьшими дополнительными затратами.

▼Построим сетевой график и найдем критический путь при условии, что все работы совершаются в минимальное время (рис. 8.10).

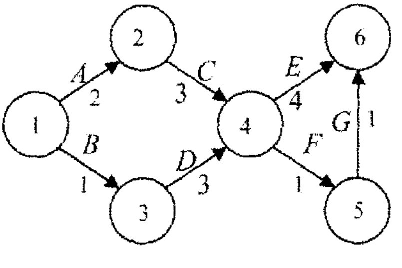


Рис. 8.10. Сетевой график при минимальных сроках (пример 8.3)

Результаты вычислений, выполненные при проходах вперед и назад, показаны на рис. 8.11.

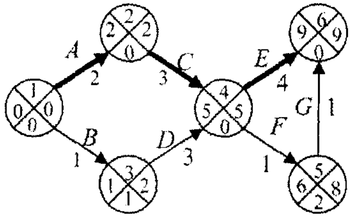


Рис. 8.11. Рассчитанный сетевой график при минимальных сроках (пример 8.3)

Минимальное время, за которое может быть завершен проект, — 9 дн. Критический путь включает работы *A*, *C*, *E*. Работы *B*, *D*, *F*, *G* не лежат на критическом пути.

Посмотрим, нельзя ли их выполнить в стандартные сроки без увеличения общего времени выполнения проекта (9 дней). Выполнение этих работ в стандартное время дает следующую экономию: 700 (*B*), 800 (*D*), 400 (*F*), 500 (*G*).

D: мы не можем увеличить продолжительность работы $D = (3, 4)$ с 3 до 5 дней, так как тогда изменятся оценка $t_р(4)$ и критический путь, т. е. общее время выполнения проекта увеличится.

B: увеличение продолжительности работы $B = (1, 3)$ с 1 до 2 дн. возможно (рис. 8.12).

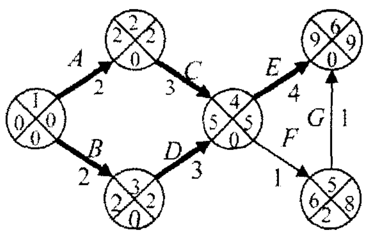


Рис. 8.12. Рассчитанный сетевой график при стандартном времени для *B* и минимальных сроках для остальных работ (пример 8.3)

Появятся 2 критических пути, включающих работы *A*, *C*, *E* и *B*, *D*, *E* соответственно. Работы *A* и *C* мы должны по-прежнему выполнять в минимальное время, иначе изменится критический путь.

G: увеличение продолжительности с 1 до 2 дн. возможно (рис. 8.13).

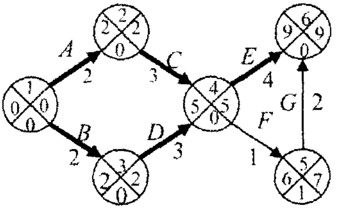


Рис. 8.13. Рассчитанный сетевой график при стандартном времени для *B*, *G* и минимальных сроках для остальных работ (пример 8.3)

F: увеличение продолжительности с 1 до 2 дн. возможно (рис. 8.14).

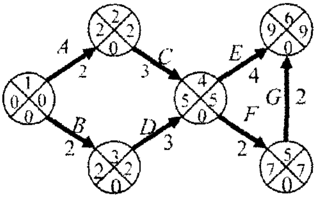


Рис. 8.14. Рассчитанный сетевой график при стандартном времени для *B*, *G*, *F* и минимальных сроках для остальных работ (пример 8.3)

Мы видим, что работы *A*, *C*, *D* и *E* выполняются в минимальное время, а работы *B*, *F*, *G* — в стандартное. Общая стоимость проекта составит: 1400 (*A*) + 1200 (*B*) + 2800 (*C*) + 2300 (*D*) + + 2800 (*E*) + 600 (*F*) + 500 (*G*) = 11 600 тыс. руб.

Таким образом, мы минимизировали общее время выполнения проекта с наименьшими дополнительными затратами. ▲

8.4. Сетевые модели в условиях неопределенности

Сетевые модели в условиях неопределенности характеризуются тем, что время выполнения отдельных операций и всего проекта в целом представляют собой случайные величины. Другими словами, продолжительность отдельной работы $t(i, j)$ является случайной величиной со своим законом распределения и своими числовыми характеристиками: математическим ожиданием (средним значением) $\overline{t}(i, j)$ и дисперсией $\sigma^2(i, j)$. Обычно предполагается, что продолжительность работ подчинена β-распределению.

Для каждой работы вводят три оценки:

• оптимистическое время a — наименьшее возможное время выполнения работы;

• пессимистическое время b — наибольшее возможное время выполнения работы;

• наиболее вероятное время m — время выполнения работы в нормальных условиях.

Ожидаемое время выполнения работы и его дисперсию можно оценить по формулам

$$\overline{t} = \frac{a + 4m + b}{6}, \quad \sigma^2 = \left(\frac{b - a}{6}\right)^2.$$

Используя значения $\overline{t}(i, j)$, можно определить критический путь сетевого графика.

Распределение времени T завершения проекта является нормальным со средним $\overline{T} = M(T)$, равным сумме ожидаемых значений времени работ на критическом пути, и дисперсией $\sigma^2(T)$, равной сумме дисперсий работ критического пути, если время выполнения каждой из работ можно считать независимым от времени других работ.

Вероятность завершения проекта в установленный срок T_0 есть

$$P(T \le T_0) = \frac{1}{2} + \frac{1}{2}\Phi\left(\frac{T_0 - \overline{T}}{\sigma(T)}\right),$$

где $\Phi(x)$ — функция Лапласа (табулирована).

Пример 8.4. Проект строительства плавательного бассейна состоит из девяти основных работ. Работы, их непосредственные предшественники и оценки времени выполнения работ приведены в табл. 8.6.

Таблица 8.6

Исходные данные для примера 8.4

Работа	Непосредственный предшественник	Время выполнения работы, дн.		
		Оптимистическое (a)	Наиболее вероятное (m)	Пессимистическое (b)
A	—	3	5	6
B	—	2	4	6
C	*A*, *B*	5	6	7
D	*A*, *B*	7	9	10
E	*B*	2	4	6
F	*C*	1	2	3
G	*D*	5	8	10
H	*D*, *F*	6	8	10
I	*E*, *G*, *H*	3	4	5

Определим ожидаемый срок завершения проекта и вероятность того, что выполнение проекта займет не более 25 дн.

▼Исходные данные и расчетные показатели представлены в табл. 8.7.

Таблица 8.7

Расчет ожидаемого времени выполнения работ и его дисперсии для примера 8.4

Работа	a	m	b	$\bar{t}=\dfrac{a+4m+b}{6}$	$\sigma^2=\left(\dfrac{b-a}{6}\right)^2$
A	3	5	6	4,83	0,25
B	2	4	6	4,00	0,44
C	5	6	7	6,00	0,11
D	7	9	10	8,83	0,25
E	2	4	6	4,00	0,44
F	1	2	3	2,00	0,11
G	5	8	10	7,83	0,69
H	6	8	10	8,00	0,44
I	3	4	5	4,00	0,11

Построим сетевой график с указанием ожидаемой продолжительности каждой работы и найдем длину критического пути (рис. 8.15).

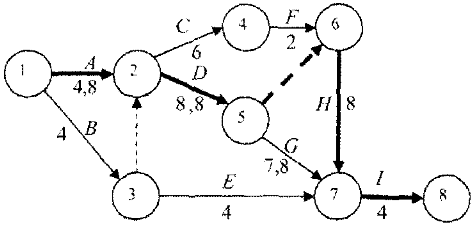


Рис. 8.15. Сетевой график при ожидаемой продолжительности работ для примера 8.4

Критический путь можно найти, например, перебором всех возможных путей:

- A, C, F, H, I: $4{,}8 + 6 + 2 + 8 + 4 = 24{,}8$;
- A, D, H, I: $4{,}8 + 8{,}8 + 0 + 8 + 4 = \underline{25{,}6}$;
- A, D, G, I: $4{,}8 + 8{,}8 + 7{,}8 + 4 = 25{,}4$;
- B, E, I: $4 + 4 + 4 = 12$;
- B, C, F, H, I: $4 + 6 + 2 + 8 + 4 = 24$;
- B, D, H, I: $4 + 8{,}8 + 8 + 4 = 24{,}8$;
- B, D, G, I: $4 + 8{,}8 + 7{,}8 + 4 = 24{,}6$.

Длина критического пути A, D, H, I в среднем равна $\overline{T} = 25{,}6$ дн. Дисперсия ожидаемого времени выполнения проекта равна сумме дисперсий критических работ: $\sigma^2(T) = 0{,}25 + 0{,}25 + 0{,}44 + 0{,}11 = 1{,}05$. Тогда стандартное отклонение времени выполнения проекта составит: $\sigma(T) = \sqrt{1{,}05} = 1{,}02$ дн.

Вероятность того, что выполнение проекта займет не более $T_0 = 25$ дн.:

$$P(T \le T_0) = \frac{1}{2} + \frac{1}{2}\Phi\left(\frac{25 - 25{,}6}{1{,}02}\right) = \frac{1}{2} + \frac{1}{2}\Phi(-0{,}59) =$$

$$= \frac{1}{2} - \frac{1}{2}\Phi(0{,}59) \approx 0{,}278. \ \blacktriangle$$

Упражнение 8.3. Рассмотрим сеть проекта с оценками продолжительности работ, указанными в табл. 8.8.

Таблица 8.8

Исходные данные для упражнения 8.3

Работа	**Непосредственный предшественник**	**Время выполнения работы, нед.**		
		Оптимистическое (a)	**Наиболее вероятное (m)**	**Пессимистическое (b)**
A	—	4	5	12
B	—	1	1,5	5
C	*A*	2	3	4
D	*A*	3	4	11
E	*A*	2	3	4
F	*C*	1,5	2	2,5
G	*D*	1,5	3	4,5
H	*B, E*	2,5	3,5	7,5
I	*H*	1,5	2	2,5
J	*F, G, I*	1	2	3

Требуется определить ожидаемое время выполнения проекта и вероятность того, что проект может быть выполнен за 20 нед.

Ответ: 17 нед.; 0,9656.

Глава 9.
ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ

9.1. Понятия

Имитационное моделирование является мощным инструментом исследования поведения реальных систем. Методы имитационного моделирования позволяют собрать необходимую информацию о поведении системы путем создания ее компьютеризованной модели. Эта информация используется затем для проектирования системы. Имитационное моделирование не решает оптимизационных задач, а скорее представляет собой технику оценки значений функциональных характеристик моделируемой системы.

Современное имитационное моделирование применяется в основном для исследования ситуаций и систем, которые можно описать как системы массового обслуживания. Это не ограничивает применение имитационного моделирования, поскольку на практике любую ситуацию исследования операций или принятия решений можно в той или иной мере рассматривать как систему массового обслуживания. По этой причине методы имитационного моделирования находят широкое применение в задачах, возникающих в процессе создания систем массового обслуживания, систем связи; в экономических и коммерческих задачах, включая оценки поведения потребителя, определение цен, экономическое прогнозирование деятельности фирм; в социальных и социально-психометрических задачах; в задачах анализа военных стратегий и тактик.

Предшественником современного имитационного моделирования считается метод Монте-Карло, основная идея которого состоит в использовании выборки случайных чисел для получения вероятностных или детерминированных оценок каких-либо величин. Основное различие между современными методами имитации и методом Монте-Карло заключается в том, что в последнем время не является обязательным фактором, а получаемые оценки «статичны». Метод Монте-Карло применяется для вычисления площадей фигур, ограниченных кривыми, или в более общем случае для вычисления кратных интегралов, констант, например числа $\pi \approx 3{,}14159$; для обращения матриц и т. п.

Имитация является случайным экспериментом, поэтому любой результат, полученный путем имитационного моделирования, подвержен экспериментальным ошибкам и, следовательно, как в любом статистическом эксперименте, должен основываться на результатах соответствующих статистических проверок.

9.2. Метод Монте-Карло

Метод Монте-Карло — это численный метод решения математических задач при помощи моделирования случайных величин.

Рассмотрим метод Монте-Карло с использованием инструмента *Генерация случайных чисел*. Этот инструмент предназначен для автоматической генерации множества данных (генеральной совокупности) заданного объекта, элементы которого характеризуются определенным распределением вероятностей.

Пример 9.1. Вычислим число π методом Монте-Карло.

▼Существует много способов вычисления числа π. Самым простым и понятным является численный метод Монте-Карло, суть которого сводится к простейшему перебору точек на площади.

Суть расчета заключается в том, что берем квадрат со стороной $a = 2r$, вписываем в него круг радиусом r. И начинаем наугад ставить точки внутри квадрата. Геометрически вероятность P_1 того, что точка попадет в круг, равна отношению площадей круга и квадрата:

$$P_1 = S_{\text{круг}}/S_{\text{квадрат}} = \pi r^2/a^2 = \pi r^2/(2r)^2 = \pi r^2/(2r)^2 = \pi/4.$$

Вероятность попадания точки в круг можно также посчитать после численного эксперимента: посчитать количество точек m, попавших в круг, и поделить их на общее количество n поставленных точек:

$$P_2 = m/n.$$

При большом количестве точек в численном эксперименте

$$P_1 \approx P_2 \Rightarrow \pi \approx 4 \cdot \frac{m}{n}.$$

Пусть круг радиуса $r = 1$ с центром в точке (0, 0) заключен в описанный около него квадрат, сторона которого равна диаметру круга (рис. 9.1).

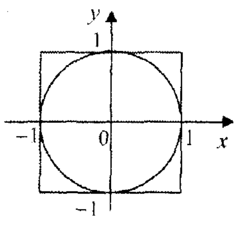


Рис. 9.1. График к примеру 9.1

Уравнение окружности имеет вид $x^2 + y^2 = 1$. Предположим, что выборка состоит из наблюдений n точек квадрата и m из них попали внутрь круга или на окружность. Процедура вычисления выборочных значений (x, y) начинается с генерирования независимых случайных чисел, равномерно распределенных на интервале [0, 1].

Пусть R_1 и R_2 — различные случайные числа из интервала [0, 1]. Тогда координаты (x, y) точек квадрата можно выразить через эти случайные числа:

$$x = -1 + 2R_1,$$
$$y = -1 + 2R_2.$$

Используя приведенные формулы, можно сгенерировать равномерно распределенную случайную точку (x, y) квадрата для каждой пары случайных чисел (R_1, R_2).

Сгенерированная точка (x, y) попадает внутрь круга, если

$$x^2 + y^2 \le 1 .$$

На рис. 9.2 приведены результаты испытаний (n = 20) и заполненное диалоговое окно *Генерация случайных чисел* из пакета *Анализа данных Excel* (в поле *Случайное рассеивание* заносится порядковый номер прогонки)

	A	B	C	D	E	F
1	n	R_1	R_2	x	y	m
2	1	0,001251	0,563585	-0,9975	0,127171	0
3	2	0,193304	0,808741	-0,61339	0,617481	1
4	3	0,585009	0,479873	0.170019	-0,04025	1
5	4	0,350291	0,895962	-0,29942	0,791925	1
6	5	0,82284	0,746605	0,64568	0.49321	1
7	6	0,174108	0,858943	-0,65178	0,717887	1
8	7	0,710501	0,513535	0,421003	0.02707	1
9	8	0,303995	0,014985	0,39201	0.97003	0
10	9	0,091403	0,364452	-0,81719	-0,2711	1
11	10	0,147313	0,165899	0,70537	0,6682	1
12	11	0,988525	0,445692	0,97705	-0,10862	1
13	12	0,119083	0,004669	0,76183	0,99066	0
14	13	0,008911	0,37788	-0,98218	-0,24424	0
15	14	0,531663	0,571184	0,063326	0,142369	1
16	15	0,601764	0,607166	0.203528	0,214331	1
17	16	0,166234	0,663045	-0,66753	0,32609	1
18	17	0,450789	0,352123	-0,09842	-0.29575	1
19	18	0,057039	0,607685	0,88592	0,215369	1
20	19	0,783319	0.802606	0.566637	0,605213	1
21	20	0.519883	0,30195	0.039766	-0.3961	1
22	Итого					16

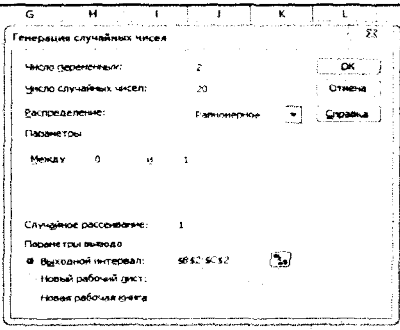


Рис. 9.2. Результаты испытаний и заполненное диалоговое окно *Генерация случайных чисел* для примера 9.1 в *Excel*

Окончательно имеем $\pi \approx 4m/n = 4 \cdot 16/20 = 3{,}2$.

Точность оценки можно повысить, увеличив объем одной выборки и (или) повторив эксперименты (прогоны) на разных выборках (но одинакового размера).

Ввиду того что оценки площади имеют разброс, важно, чтобы результаты эксперимента, связанного с моделированием, были выражены в виде доверительных интервалов, показывающих величину отклонения от точного значения. В рассматриваемом примере если π представляет собой точное значение площади, а $\overline{\pi}$ и s^2 — среднее и дисперсию при числе экспериментов N, то $100(1 - \alpha)$-процентный доверительный интервал для π задается в виде

$$\overline{\pi} - t_{\alpha/2, N-1} \frac{s}{\sqrt{N}} < \pi < \overline{\pi} + t_{\alpha/2, N-1} \frac{s}{\sqrt{N}} ,$$

где $t_{\alpha/2, N-1}$ — *точка t-распределения* (*распределения Стьюдента*) с $N - 1$ степенями свободы. Заметим, что N обозначает число экспериментов (прогонов), и его следует отличать от n, которое обозначает объем выборки («продолжительность» прогона модели).

В табл. 9.1 приведены результаты расчета числа π для $N = 5$ прогонов.

Таблица 9.1

Результаты расчета числа π

Показатель	**Номер прогона**				
	1	**2**	**3**	**4**	**5**
m	16	16	18	14	13
π	3,2	3,2	3,6	2,8	2,6

Тогда среднее значение $\overline{\pi} = 3{,}08$, стандартное отклонение $s = 0{,}39$. Примем $\alpha = 5$ %. Табличное значение распределения Стьюдента $t_{0,025,4} = 2{,}776$, поэтому нижняя и верхняя границы доверительного интервала для числа π соответственно равны $\pi_L = 2{,}596$, $\pi_R = 3{,}564$.

Таким образом, при объеме выборки $n = 20$ с вероятностью 95 % доверительный интервал числа π есть (2,596; 3,564).

Можно показать, что при объеме выборки $n = 3000$ с вероятностью 95 % доверительный интервал числа π составляет (3,137; 3,151). ▲

Рассмотренный пример ставит два вопроса, характерных для любого эксперимента, связанного с моделированием:

- каким должен быть объем выборки n для достижения необходимого значения доверительных интервалов;
- сколько для этого требуется прогонов N?

Как и в любом статистическом эксперименте, большие значения n и N обеспечивают более надежные результаты. Препятствием может быть стоимость (время) проведения эксперимента, которая возрастает пропорционально увеличению n и N.

Интегрирование с помощью метода Монте-Карло

Рассмотрим интеграл вида

$$I = \int_a^b \varphi(x)dx\,.$$

Введем в рассмотрение случайную величину X, распределенную равномерно в интервале интегрирования (a, b) с плотностью $f(x) = \dfrac{1}{b-a}$. Тогда математическое ожидание

$$M\left[\varphi(x)\right] = \int_a^b \varphi(x) f(x) dx = \frac{1}{b-a} \int_a^b \varphi(x) dx .$$

Отсюда

$$\int_a^b \varphi(x) dx = (b-a) M\left[\varphi(x)\right].$$

Заменив математическое ожидание $M\left[\varphi(x)\right]$ его оценкой — выборочной средней, получим оценку интеграла

$$I^* = (b-a) \frac{\sum_{i=1}^{n} \varphi(x_i)}{n},$$

где x_i — возможные значения случайной величины X;

n — число испытаний.

Дисперсия усредняемой функции $\varphi(x)$ равна

$$\sigma^2 = \int_a^b \varphi^2(x) f(x) dx - \left(M\left[\varphi(X)\right]\right)^2 .$$

Оценкой дисперсии σ^2 является выборочная дисперсия (при $n > 30$)

$$\sigma_e^2 = \frac{\sum_{i=1}^{n} \varphi^2(x_i)}{n} - \left[\frac{\sum_{i=1}^{n} \varphi(x_i)}{n}\right]^2$$

или исправленная дисперсия (при $n < 30$)

$$s^2 = \frac{n}{n-1} \sigma_e^2 .$$

Так как случайная величина распределена равномерно в интервале (a, b) с плотностью $f(x)$, то x_i разыгрывается по формуле

$$\frac{1}{b-a}\int_a^{x_i}\varphi(x)dx = r_i .$$

Отсюда $x_i = a + (b-a)r_i$, где r_i — случайное число.

Пример 9.2. Найти оценку интеграла $I = \int_0^2 (4-x^2)dx$, абсолютную погрешность $|I - I^*|$, минимальное число испытаний, которое с надежностью 0,95 обеспечит верхнюю границу ошибки $\delta = 0{,}1$.

▼По условию $a = 0$, $b = 2$, $\varphi(x) = 4 - x^2$. Примем число испытаний $n = 10$, тогда оценка интеграла равна

$$I^* = \frac{2\sum_{i=1}^{10}\varphi(x_i)}{10},$$

где $x_i = 2r_i$.

Результаты испытаний (рассчитано в Excel) представим в табл. 9.2.

Таблица 9.2

Результаты испытаний и расчеты для примера 9.2

i	r_i	x_i	$\varphi(x_i)$	$\varphi^2(x_i)$
1	0,001	0,003	4,000	16,000
2	0,564	1,127	2,729	7,450
3	0,193	0,387	3,851	14,827
4	0,809	1,617	1,384	1,915
5	0,585	1,170	2,631	6,922
6	0,480	0,960	3,079	9,480
7	0,350	0,701	3,509	12,314
8	0,896	1,792	0,789	0,623
9	0,823	1,646	1,292	1,669
10	0,747	1,493	1,770	3,134
Итого			25,034	74,333

Из табл. 9.2 $\sum \varphi(x_i)$ = 25,03. Оценка интеграла $I^* = 2 \cdot 25{,}03/10 = 5{,}01$.

Истинное значение интеграла $I = 16/3 \approx 5{,}33$. Тогда абсолютная погрешность $\delta = \left| I - I^* \right| = 0{,}32$.

Из табл. 9.2 $\sum \varphi^2(x_i) = 74{,}33$. Выборочная и исправленная дисперсии:

$$\sigma_e^2 = \frac{74{,}33}{10} - \left[\frac{25{,}03}{10}\right]^2 = 1{,}17;\ s^2 = \frac{10}{9} \cdot 1{,}17 = 1{,}3.$$

Вероятность того, что отклонение выборочной средней от генеральной средней не превзойдет число δ, равна

$$P(\left| I - I^* \right| \le \delta) = \Phi(t),\ t = \frac{\delta\sqrt{n}}{s},$$

где $\Phi(t)$ — функция Лапласа.

Из соотношения $\Phi(t) = 0{,}95$ по таблице находим $t = 1{,}96$.

Минимальное число испытаний, которое с надежностью 0,95 обеспечит верхнюю границу ошибки $\delta = 0{,}1$, есть

$$n = \frac{t^2 s^2}{\delta^2} = \frac{1{,}96^2 \cdot 1{,}3}{0{,}1^2} \approx 499.\ \blacktriangle$$

Упражнение 9.1. Оцените значение интеграла $I = \int_2^5 x^2 dx$ с помощью метода Монте-Карло.

9.3. Элементы дискретного моделирования

Использование современных имитационных моделей базируется в основном на идее метода Монте-Карло. Отличие состоит в том, что имитационная модель обычно связана с изучением реально существующей системы, поведение которой является функцией времени. Существуют два типа имитационных моделей: непрерывные и дискретные.

Непрерывные модели используются для систем, поведение которых изменяется непрерывно во времени. Непрерывные имитационные модели обычно представляются в виде разностно-дифференциальных уравнений, которые описывают взаимодействие между различными элементами системы.

Дискретные модели имеют дело с системами, поведение которых изменяется лишь в заданные моменты времени. Типичным примером такой модели является очередь. Все имитационные модели с дискретными событиями описывают прямо или косвенно ситуации с очередью, в которую клиенты прибывают, при необходимости ожидают в ней, потом обслуживаются перед тем, как оставить систему. В общем случае любая модель с дискретными событиями состоит из сети взаимосвязанных очередей.

Имитационная модель с дискретными событиями в действительности является композицией очередей. В целях сбора статистических данных (показателей функционирования системы) заметим, что изменения в системе (например, изменение длины очереди или состояния средств обслуживания) возникают лишь тогда, когда клиент поступает в очередь или покидает систему после обслуживания. Это означает, что двумя главными событиями в любой дискретной имитационной модели являются прибытие и уход клиентов. Это единственные показатели, по которым необходимо исследовать систему. В другие моменты времени никаких изменений, влияющих на статистические данные системы, не происходит.

По сути, имитация возникает в те моменты, когда происходят события.

Рассмотрим, как имитационное моделирование определяет время наступления событий. В системе события, связанные с прибытием, определяются временем между поступлениями клиентов, а события, связанные с их уходом, — временем обслуживания.

Время наступления этих событий может быть детерминированным (например, прибытие поездов метро на станцию каждые пять минут) или случайным (например, прибытие клиентов в банк). Если время между наступлениями событий является детерминированным, то процедура определения времени их наступления проста. Если же указанное время является случайным, то используется специальная процедура для получения выборочных значений времени между событиями в системе, соответствующей заданному вероятностному распределению.

Генерирование выборочных значений. Предположим, что время t между прибытиями клиентов в парикмахерскую распределено по экспоненциальному закону с математическим ожиданием $M(t) = 1/\lambda$ единиц времени, т. е. плотность вероятности задается формулой

$$f(t) = \lambda e^{-\lambda t}, \; t > 0.$$

Найдем случайное значение времени t.

Функция распределения вычисляется стандартным образом:

$$F(t) = \int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda t}, \; t > 0.$$

Если R — случайное число из интервала [0, 1], то, если $R = F(t)$, получим

$$t = -\left(\frac{1}{\lambda}\right) \ln(1 - R).$$

Так как R — случайное число из интервала [0,1], то и $(1 - R)$ представляет собой случайное число из того же интервала, поэтому можно заменить $(1 - R)$ на R.

Пусть в имитационной модели события происходят через t единиц времени. Тогда, например, при $\lambda = 4$ посетителя в час и $R_1 = 0{,}9$ интервал времени между прибытиями посетителей вычисляется как

$$t_1 = -\left(\frac{1}{4}\right) \ln(1 - 0{,}9) = 0{,}576 \text{ ч} = 34{,}5 \text{ мин}.$$

Значения R, используемые для получения последовательных случайных чисел, должны выбираться случайным образом из интервала [0,1], подчиняясь равномерному закону распределения.

Имитация модели очереди с одним сервисом

Пример 9.3. В парикмахерской работает один мастер, который выполняет мужскую стрижку от 10 до 15 мин, время стрижки имеет равномерное распределение на этом интервале. Клиенты обслуживаются в порядке очереди. Время между приходом клиентов в парикмахерскую является случайной величиной, изменяющейся по экспоненциальному закону с математическим ожиданием 15 мин.

Требуется определить следующие параметры работы парикмахерской: среднюю занятость парикмахера, среднее количество ожидающих клиентов, среднее время ожидания клиента в очереди.

▼В рассматриваемом примере время между приходом клиентов изменяется по экспоненциальному закону с математическим ожиданием 15 мин, а время обслуживания распределено равномерно в интервале от 10 до 15 мин. Обозначим через p и q случайные значения времени между прибытием клиентов и их обслуживанием соответственно, тогда

$$p = -15 \cdot \ln(R) \text{ мин}, \ 0 < R \leq 1,$$
$$q = 10 + 5R \text{ мин}, \ 0 < R \leq 1.$$

В этом примере используем следующие случайные числа R, начиная с первой строки:

0,0589; 0,6733; 0,4799; 0,9486;
0,6139; 0,5933; 0,9341; 0,1782;
0,3473; 0,5644; 0,3529; 0,3646;
0,7676; 0,8931; 0,3919; 0,7876;
0,5199; 0,6358; 0,7472; 0,8954.

Обозначим через T время моделирования (время наблюдения над моделью). Предположим также, что первый клиент приходит в момент времени $T = 0$ и средство обслуживания свободно.

Поскольку в процессе имитации выполняется много вычислений, ограничимся имитацией прихода только пяти клиентов. При этом рассмотрим все возможные ситуации, возникающие в процессе имитации.

Прибытие клиента 1 в момент $T = 0$. Второй клиент прибудет в момент времени $T = 0 + p = -15 \cdot \ln(0{,}0589) = 42{,}48$ мин.

Так как при $T = 0$ система свободна, немедленно начинается обслуживание первого клиента. Время его обслуживания $q_1 = 10 + 5 \cdot 0{,}6733$ вычисляется с использованием равномерно распределенного случайного числа $R = 0{,}6733$.

Первый клиент уйдет из парикмахерской в момент времени

$$T = 0 + q_1 = 0 + 10 + 5 \cdot 0{,}6733 = 13{,}37 \text{ мин.}$$

Получаем следующий *хронологический* перечень будущих событий:

- 13,37 мин — уход клиента 1;
- 42,48 мин — приход клиента 2.

Уход клиента 1 в момент $T = 13{,}37$. Так как очередь пуста, средство обслуживания становится свободным. Система была занята от момента времени $T = 0$ до $T = 13{,}37$. Имеем следующий скорректированный перечень будущих событий (в данном случае имеется только одно событие):

- 42,48 мин — приход клиента 2.

Прибытие клиента 2 в момент $T = 42{,}48$. Третий клиент прибудет в момент времени

$$T = 42{,}48 + [-15 \cdot \ln(0{,}4799)] = 53{,}49 \text{ мин.}$$

Так как средство обслуживания свободно, начинается обслуживание второго клиента, и система объявляется занятой. Время обслуживания второго клиента $q_2 = 10 + 5 \cdot 0{,}9486 = 14{,}74$ мин. Второй клиент оставит систему в момент времени

$$T = 42{,}48 + 14{,}74 = 57{,}22 \text{ мин.}$$

Список будущих событий изменяется следующим образом:

- 53,49 — приход клиента 3;
- 57,22 — уход клиента 2.

Прибытие клиента 3 в момент $T = 53{,}49$. Клиент 4 прибудет в момент времени

$$T = 53{,}49 + [-15 \cdot \ln(0{,}6139)] = 60{,}81 \text{ мин.}$$

Так как система в это время занята (она занята до $T = 57{,}22$), третий клиент помещается в очередь в момент $T = 53{,}49$. Откорректированный список будущих событий принимает следующий вид:

- 57,22 — уход клиента 2;
- 60,81 — приход клиента 4.

Уход клиента 2 в момент $T = 57{,}22$. Клиент 3 покидает очередь для начала обслуживания. Время его ожидания в очереди равно

$$W_3 = 57{,}22 - 53{,}49 = 3{,}73 \text{ мин.}$$

Время обслуживания третьего клиента $q_3 = 10 + 5 \cdot 0{,}5933 = 12{,}97$ мин.

Время ухода (время окончания обслуживания) третьего клиента:

$$T = 57{,}22 + 12{,}97 = 70{,}19 \text{ мин.}$$

Список будущих событий принимает следующий вид:

- 60,81 — приход клиента 4;
- 70,19 — уход клиента 3.

Прибытие клиента 4 в момент $T = 60{,}81$. Клиент 5 прибудет в момент времени

$$T = 60{,}81 + [-15 \cdot \ln(0{,}9341)] = 61{,}83 \text{ мин.}$$

Так как система занята до $T = 70{,}19$, четвертый клиент помещается в очередь. Откорректированный список будущих событий принимает следующий вид:

- 61,83 — приход клиента 5;
- 70,19 — уход клиента 3.

Приход клиента 5 в момент $T = 61{,}83$. Поскольку наша имитация ограничивается пятью клиентами, время прибытия шестого клиента генерировать не будем. Так как средство обслуживания в момент прихода клиента 5 занято, он в момент времени $T = 61{,}83$ помещается в очередь. Имеем список будущих событий:

- 70,19 — уход клиента 3.

Уход клиента 3 в момент $T = 70{,}19$. Первый клиент в очереди (клиент 4) начинает обслуживаться. Его время ожидания в очереди составляет

$$W_4 = 70{,}19 - 60{,}81 = 9{,}38 \text{ мин.}$$

Время обслуживания четвертого клиента $q_4 = 10 + 5 \cdot 0{,}1782 = 10{,}89$ мин.

Время ухода клиента 4:

$$T = 70{,}19 + 10{,}89 = 81{,}08 \text{ мин.}$$

Откорректированный список будущих событий принимает такой вид:

• 81,08 — уход клиента 4.

Уход клиента 4 в момент $T = 81{,}08$. Клиент 5 начинает обслуживаться. Его время ожидания в очереди составляет

$$W_5 = 81{,}08 - 61{,}83 = 19{,}25 \text{ мин.}$$

Время обслуживания пятого клиента $q_5 = 10 + 5 \cdot 0{,}3473 = 11{,}74$ мин.

Время ухода клиента 5:

$$T = 81{,}08 + 11{,}74 = 92{,}82 \text{ мин.}$$

Откорректированный список будущих событий принимает такой вид:

• 92,82 — уход клиента 5.

Уход клиента 5 в момент $T = 92{,}82$. Клиентов в системе (в очереди и на обслуживании) больше нет. Имитация заканчивается.

На рис. 9.3 показаны изменения длины очереди и использование сервиса (занятость системы) как функции времени имитации.

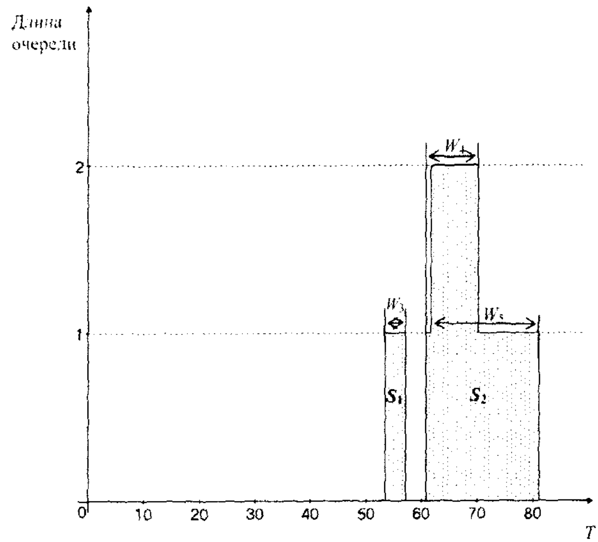
Длина очереди и занятость системы (т. е. парикмахера) являются переменными, зависящими от времени. Поэтому их средние значения вычисляются следующим образом.

$$\begin{pmatrix}\text{Среднее значение переменной,}\\ \text{зависящей от времени}\end{pmatrix} = \left(\frac{\text{Площадь под кривой}}{\text{Период имитации}}\right).$$

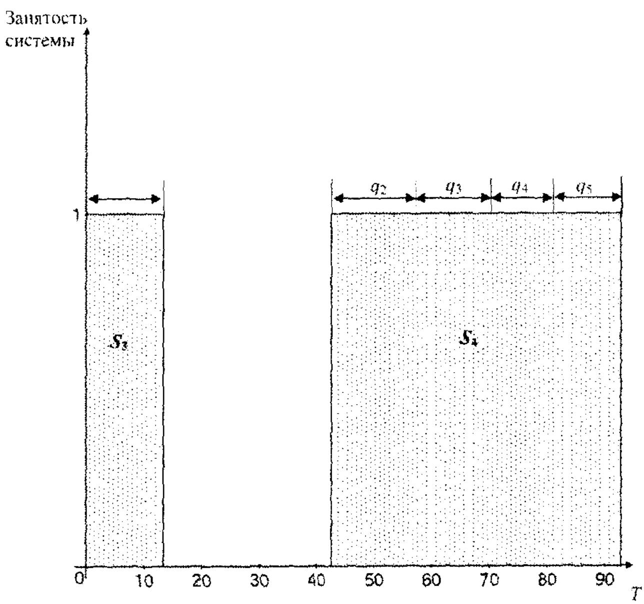
Применяя эту формулу для данных, показанных на рис. 9.3, получаем

$$S_1 + S_2 = W_3 \cdot 1 + (W_4 \cdot 1 + W_5 \cdot 1) = 3{,}73 + (9{,}38 + 19{,}25) = 32{,}36.$$

$$S_3 + S_4 = q_1 \cdot 1 + (q_2 + q_3 + q_4 + q_5) \cdot 1 =$$
$$= 13{,}37 + (14{,}74 + 12{,}97 + 10{,}89 + 11{,}74) = 63{,}71.$$



а)



б)

Рис. 9.3. Изменение длины очереди (*а*) и использование сервиса (*б*) во время имитации

Средняя длина очереди = $(S_1 + S_2)/92{,}82 = 32{,}36/92{,}82 = 0{,}349$ (клиента).

Средняя занятость системы = $(S_3 + S_4)/92{,}82 = 63{,}71/92{,}82 =$ $= 0{,}686$ (парикмахера).

Среднее время ожидания в очереди является переменной, зависящей от количества происшедших событий, и ее значение вычисляется по формуле

$$\left(\begin{array}{c}\text{Среднее значение переменной,}\\ \text{зависящей от количества событий}\end{array}\right) = \left(\frac{\text{Сумма наблюдений}}{\text{Количество событий}}\right).$$

Из рис. 9.3, *а* видно, что площадь под кривой длины очереди в действительности равна сумме времен ожидания трех клиентов, которые формировали очередь. Поэтому

$$W_1 + W_2 + W_3 + W_4 + W_5 = 0 + 0 + 3{,}73 + 9{,}38 + 19{,}25 = 32{,}36 \text{ мин.}$$

Среднее время ожидания в очереди равно $W_q = 32{,}36/5 =$ $= 6{,}47$ мин. ▲

Конечно, модель с одним сервисом очень проста и ее несложно смоделировать в электронной таблице. Для моделирования других моделей придется затратить больше усилий, но использование специального программного обеспечения значительно облегчает эту работу.

Имитационное моделирование управления запасами

Рассмотрим одно из направлений имитационного моделирования. Пусть моделируется некоторая случайная величина. Вначале из опытных данных определяются частоты появления возможных значений этой величины. По частотам вычисляются вероятности, а по вероятностям — кумулятивные вероятности. По кумулятивным вероятностям устанавливается соответствие между случайными числами и значениями случайной величины. Затем выбираются случайные числа с использованием или функции *Excel* СЛУЧМЕЖДУ(1;100), или специальных таблиц случайных чисел, восстанавливаются по ним значения случайной величины и определяются нужные характеристики.

Замечание. При использовании функции СЛУЧМЕЖДУ (1;100) результаты фиксируется, так как эти числа могут изменяться со временем.

Пример 9.4. Фирма занимается продажей некоторых изделий. Начальный запас — 10 ед., стоимость подачи заказов на изделия K = 10 руб./зак., стоимость хранения h = 5 руб./ед. в день, стоимость одной упущенной продажи c_b = 80 руб. При наличии на складе не более 5 ед. подается заказ на 10 ед. Считается, что все заказы подаются и выполняются в начале рабочего дня. На основе прошлого опыта известны частота спроса (табл. 9.3) и частота времени выполнения заказа (табл. 9.4).

Таблица 9.3

Частота спроса для примера 9.4

Спрос в день, ед.	0	1	2	3	4	5
Частота	15	30	60	120	45	30

Таблица 9.4

Частота времени выполнения заказа для примера 9.4

Время выполнения заказа, дни	1	2	3
Частота	10	25	15

Оценим общие издержки в день, смоделировав работу склада за 10 дней.

▼Построим функцию распределения (кумулятивную вероятность) спроса и интервалы (*от* и *до*) случайных чисел для значений случайной переменной (табл. 9.5).

Аналогично построим кумулятивную вероятность и интервалы случайных чисел для времени выполнения заказа (табл. 9.6).

Смоделируем работу склада за 10 дней в табл. 9.7.

Вначале заполняются третий и четвертый столбец табл. 9.7 с использованием функции *Excel* СЛУЧМЕЖДУ(0;100) и табл. 9.5 соответственно.

Затем последовательно рассчитывается запас на конец дня, равный запасу в начале дня за вычетом спроса (величины продаж). Полученный запас переносится на запас в начале следующего дня.

Таблица 9.5

Расчет кумулятивной вероятности спроса и соответствующих интервалов случайных чисел для примера 9.4

Спрос в день	Частота	Вероятность	Кумулятивная вероятность	Интервал случайных чисел			
0	15	0,05	0,05	*от*	0	*до*	5
1	30	0,10	0,15	*от*	6	*до*	15
2	60	0,20	0,35	*от*	16	*до*	35
3	120	0,40	0,75	*от*	36	*до*	75
4	45	0,15	0,90	*от*	76	*до*	90
5	30	0,10	1,00	*от*	91	*до*	100
Итого	300	1,00					

Таблица 9.6

Расчет кумулятивной вероятности времени выполнения заказа и соответствующих интервалов случайных чисел для примера 9.4

Время выполнения заказа, дни	Частота	Вероятность	Кумулятивная вероятность	Интервал случайных чисел			
1	10	0,20	0,20	*от*	0	*до*	20
2	25	0,50	0,70	*от*	21	*до*	70
3	15	0,30	1,00	*от*	71	*до*	100
Итого	50	1,00					

Таблица 9.7

Моделирование работы склада для примера 9.4

День	Запас в начале дня	Случайное число	Спрос	Запас на конец дня	Делать заказ?	Случайное число	Время выполнения	Дефицит
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	10	22	2	8				

Окончание табл. 9.7

День	Запас в начале дня	Случайное число	Спрос	Запас на конец дня	Делать заказ?	Случайное число	Время выполнения	Дефицит
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
2	8	19	2	6				
3	6	51	3	3				
4	3	78	4	0	да	15	1	1
5	10	44	3	7				
6	7	77	4	3				
7	3	3	0	3	да	35	2	
8	3	79	4	0				1
9	10	74	3	7				
10	7	89	4	3				
Итого				40				2

Определяется, снизился ли запас до точки восстановления (в примере — 5 ед.). Если да, то делается заказ и определяется время его выполнения с использованием табл. 9.6 и функции *Excel* СЛУЧМЕЖДУ(0;100). Рассчитывается дефицит.

Результаты имитационного эксперимента:

• средний запас = суммарный конечный запас/общее число дней = 40/10 = 4,0 ед./день;

• среднее число упущенных продаж = общее число упущенных продаж/общее число дней = 2/10 = 0,2 продажи/день;

• среднее число заказов = общее число заказов/общее число дней = 2/10 = 0,2 заказа/день;

• общие издержки = подача заказов + хранение + штраф за дефицит = K · (среднее число заказов) + h · (средний запас) + c_b × ×·(среднее число упущенных заказов) = 10 · 0,2 + 5 · 4,0 + 80 · 0,2 = = 38,0 руб./день. ▲

Упражнение 9.2. Фирма занимается продажей некоторых изделий. Начальный запас — 11 ед., стоимость подачи заказов на изделия K = 15 руб./заказ, стоимость хранения h = 6 руб./ед. в день, стоимость одной упущенной продажи c_b = 70 руб. При наличии на складе не более 5 ед. подается заказ на 11 ед. Считается, что все заказы подаются и выполняются в начале рабочего дня. На основе прошлого опыта известны частота спроса (табл. 9.8) и частота времени выполнения заказа (табл. 9.9).

Таблица 9.8

Частота спроса для упражнения 9.2

Спрос в день, ед.	0	1	2	3	4	5
Частота	10	15	25	20	20	10

Таблица 9.9

Частота времени выполнения заказа для упражнения 9.2

Время выполнения заказа, дни	1	2	3
Частота	5	30	15

Требуется оценить общие издержки в день, смоделировав работу склада за 10 дней.

Для упрощения случайные числа выбирайте из следующего списка:

• для спроса:

35; 91; 92; 94; 26; 57; 34; 30; 91; 1;

• времени выполнения заказа:

71; 1; 72; 92; 55; 45; 15; 51; 37; 49.

Ответ: 91 руб./день.

Глава 10. ЭКОНОМЕТРИЧЕСКИЕ МОДЕЛИ

10.1. Общие понятия эконометрических моделей

Закономерности в экономике выражаются в виде зависимостей экономических показателей и математических моделей их поведения. Такие зависимости и модели могут быть получены только путем обработки реальных статистических данных с учетом внутренних связей и случайных факторов.

Эконометрика — наука, изучающая количественные закономерности и взаимозависимости в экономике методами математической статистики.

Цель эконометрики — эмпирический вывод экономических законов.

Задачи эконометрики — построение экономических моделей и оценивание их параметров, проверка гипотез о свойствах экономических показателей и формах их связи.

Эконометрический анализ служит основой для экономического анализа и прогнозирования, создавая возможность для принятия обоснованных экономических решений.

Типы данных. При моделировании экономических процессов оперируют с двумя типами данных: пространственными и временными.

Пространственные данные — это данные по какому-либо экономическому показателю, полученные от разных однотипных объектов (фирм, регионов и т. п.), но относящиеся к одному и тому же моменту времени (пространственный срез). Например, данные об объеме производства, количестве работников, доходе разных фирм в один и тот же момент времени.

Временные ряды — это данные, характеризующие один и тот же объект в различные моменты времени (временной срез). Например, ежеквартальные данные об инфляции, средней заработной плате, данные о национальном доходе за последние годы. Отличительная черта временных данных — упорядоченность во времени. Кроме того, наблюдения в близкие моменты времени часто бывают зависимы.

Любые экономические данные представляют собой характеристики какого-либо экономического объекта. Они формируются под воздействием множества факторов, не все из которых доступны внешнему контролю. Неконтролируемые (неучтенные) факторы обусловливают случайность данных, которые они определяют.

Поскольку экономические данные имеют статистическую природу, для их анализа и обработки необходимо применять статистические методы.

Классы моделей. Можно выделить три основных класса моделей: модели временных рядов, регрессионные модели с одним уравнением и системы одновременных уравнений.

К *моделям временных рядов* относятся модели тренда и сезонности. Тренд представляет собой устойчивое изменение уровня показателя в течение длительного времени. Сезонность характеризует устойчивые внутригодовые колебания уровня показателя. Кроме того, к этому классу относится множество более сложных моделей, таких, например, как модель с распределенным лагом, модель авторегрессии и модель адаптивных ожиданий. Их общей чертой является то, что они объясняют поведение временного ряда исходя только из его предыдущих и будущих значений.

Модели временных данных подразделяются также на модели, построенные по стационарным и нестационарным временным рядам.

Стационарные временные ряды — это ряды, которые имеют постоянное среднее значение и колеблются вокруг него с постоянной дисперсией. Распределение показателей уровня ряда в них не зависит от времени, т. е. такой временной ряд не содержит трендовой или сезонной компоненты.

В нестационарных временных рядах распределение уровня ряда зависит от времени, т. е. ряд имеет трендовую или сезонную компоненту.

В *регрессионных моделях с одним уравнением* объясняемая переменная представляется в виде функции от объясняющих переменных. Например, модель спроса на некоторый товар в зависимости от его цены и дохода потребителей.

По виду функции регрессионные модели подразделяются на линейные и нелинейные. Существуют эффективные методы оценки и анализа линейных регрессионных моделей. Их анализ является базовым в прикладной эконометрике.

Область применения регрессионных моделей, даже линейных, значительно шире, чем моделей временных рядов.

Системы одновременных уравнений описываются системами уравнений, состоящими из тождеств и регрессионных уравнений, в каждом из которых аргументы помимо объясняющих переменных содержат объясняемые переменные из других уравнений системы. Например, модель формирования доходов.

Все три класса моделей могут использоваться при моделировании экономических процессов.

Обычно предполагают, что все факторы, не учтенные явно в экономической модели, оказывают на объект некое результирующее воздействие, величина которого задается случайной компонентой.

Введение случайной компоненты в экономическую модель делает ее доступной для эмпирической проверки на основе статистических данных.

Типы зависимостей. В экономических исследованиях одной из основных задач является анализ зависимостей между переменными. Зависимость может быть функциональной либо статистической.

Функциональная зависимость задается в виде точной формулы, в которой каждому значению одной переменной соответствует строго определенное значение другой, воздействием случайных факторов при этом пренебрегают.

В экономике функциональная зависимость между переменными проявляется редко.

Статистической зависимостью называется связь переменных, на которую накладывается воздействие случайных факторов. При этом изменение одной переменной приводит к изменению математического ожидания другой переменной.

По направлению выделяют прямые и обратные связи. Направление изменения результативного признака при прямой связи совпадает с направлением изменения факторного признака, а при обратной связи противоположно ему.

Уравнение регрессии — это формула статистической связи между переменными. Если эта формула линейна, то имеем *линейную регрессию*.

Формула статистической связи двух переменных называется *парной регрессией*, зависимость от нескольких переменных — *множественной регрессией*.

10.2. Элементы математической статистики

Случайной величиной (переменной) называется величина, которая под воздействием случайных факторов может с определенными вероятностями принимать те или иные значения из некоторого множества чисел.

Случайные величины обозначаются заглавными буквами латинского алфавита, например (X, Y и т. п.), а их возможные значения — строчными (x, y и т. п.).

Универсальным способом задания случайной величины X является задание ее функции распределения.

Функцией распределения $F(x)$ случайной величины X называется вероятность того, что величина X принимает значение, меньшее x, т. е.

$$F(x) = P(X < x), \ \ x \in R .$$

Различают дискретные и непрерывные случайные величины.

Дискретной называется случайная величина, которая принимает отдельные, изолированные друг от друга значения.

Непрерывной называется случайная величина, множество значений которой непрерывно заполняет некоторый числовой промежуток.

В основе математической статистики лежат понятия генеральной и выборочной совокупностей.

Генеральная совокупность — множество всех значений случайной величины, которые она может принять в процессе наблюдения. Например, данные о доходах всех жителей страны.

Выборочная совокупность (выборка) — множество наблюдений, составляющих лишь часть генеральной совокупности.

Для любой случайной величины важную роль помимо функции распределения играют **числовые характеристики ее распределения**.

Пусть из генеральной совокупности извлекается выборка объема n.

Выборочной средней называется среднее арифметическое наблюдаемых значений случайной величины в выборке, т. е.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

Выборочной дисперсией (вариацией) называется среднее арифметическое квадратов отклонения наблюдаемых значений случайной величины в выборке от их среднего значения, т. е.

$$\mathrm{var}(x) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \text{, или } \mathrm{var}(x) = \overline{x^2} - (\overline{x})^2 ,$$

где $\overline{x^2} = \dfrac{\sum_{i=1}^{n} x_i^2}{n}$.

Значения $\overline{x}$, $\mathrm{var}(x)$ являются числовыми характеристиками выборочной совокупности.

Для разных выборок, взятых из одной и той же генеральной совокупности, выборочные средние и дисперсии будут различны, т. е. выборочные характеристики являются случайными величинами.

Генеральной средней называется математическое ожидание случайной величины X, т. е.

$$\mu X = M(X),$$

которое характеризует среднее значение случайной величины в генеральной совокупности.

Генеральной дисперсией называется математическое ожидание квадрата отклонения случайной величины X относительно ее средней, т. е.

$$\sigma_X^2 = D(X) = M(X - \mu_X)^2,$$

которое характеризует меру рассеяния случайной величины в генеральной совокупности относительно ее средней.

Стандартным отклонением случайной величины X называется квадратный корень из ее дисперсии, т. е.

$$\sigma_X = \sqrt{D(X)},$$

который показывает, насколько в среднем отклоняется случайная величина в генеральной совокупности относительно ее средней.

Значения μ_X, σ_X^2 — это числовые характеристики генеральной совокупности (числа), которые в отличие от выборочных характеристик $\overline{x}$, $\mathrm{var}(x)$ являются неслучайными величинами (истинными).

Две случайные величины называются **независимыми**, если закон распределения одной из них не меняется от того, какие возможные значения приняла другая величина.

Если случайные величины X, Y независимы, то

$$M(XY) = M(X)M(Y);$$
$$D(X+Y) = D(X) + D(Y).$$

Из условия, что выборочные наблюдения $X_1, X_2, \ldots, X_n$ независимы и имеют одинаковые распределения, вытекают следующие соотношения:

$$\mu_{\overline{x}} = M(\overline{x}) = \mu_X,\ \sigma_{\overline{x}}^2 = D(\overline{x}) = \sigma_X^2 / n, \sigma_{\overline{x}} = \sigma_X / \sqrt{n}.$$

Для задания непрерывной случайной величины обычно используется плотность распределения вероятностей $f(x) = F'(x)$.

Площадь S под кривой распределения $f(x)$ в интервале $(a; b)$ равна вероятности попадания случайной величины X в этот интервал:

$$S = \int_a^b f(x)dx = P(a < X < b), \text{ причем } \int_{-\infty}^{\infty} f(x)dx = 1.$$

Наблюдаемое распределение $f(x)$ сводят к одному из хорошо исследованных видов распределения (нормальное распределение или специальные распределения, получаемые из нормального).

В эконометрических исследованиях чаще всего используются нормальное распределение, распределение Стьюдента и распределение Фишера. На основе этих распределений построены статистические критерии, служащие для проверки гипотез.

Нормальное распределение случайной величины X характеризуется лишь двумя параметрами: средним значением μ и дисперсией σ^2.

Это обозначается как $X \sim N(\mu; \sigma^2)$.

График плотности нормального распределения $f(x)$ имеет колоколообразный вид, симметричный относительно центра распределения (рис. 10.1).

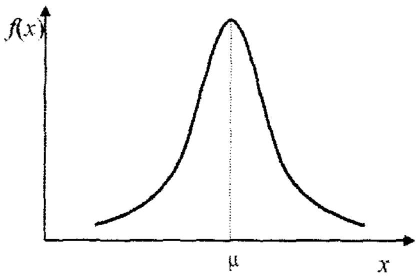


Рис. 10.1. График плотности нормального распределения

Максимум этой функции находится в точке $x = \mu$, а разброс относительно этой точки определяется параметром σ^2. Чем меньше значение σ^2, тем более острый и высокий максимум $f(x)$.

Если от точки среднего значения (или точки максимума плотности нормального распределения) отложить вправо и влево три величины стандартного отклонения, то площадь под графиком плотности распределения, подсчитанная по этому промежутку, будет равна 99,73 % всей площади под графиком (правило трех сигм).

Другими словами, 99,73 % всех независимых наблюдений из нормального распределения лежат в радиусе трех стандартных отклонений от среднего значения.

Статистические оценки и их свойства

Характеристики генеральной совокупности обычно неизвестны. Задача заключается в их оценке по характеристикам выборочной совокупности.

Характеристики генеральной совокупности принято называть *параметрами*, а выборочной совокупности — *оценками*.

Пусть искомый параметр генеральной совокупности есть θ_0, а на основе выборки объема n определяется оценка θ.

Для того чтобы выборочная оценка давала хорошее приближение оцениваемого параметра, она должна удовлетворять определенным требованиям (несмещенности, эффективности и состоятельности).

Несмещенность оценок. Оценка θ является несмещенной, если ее математическое ожидание равно оцениваемому параметру θ_0 при любом объеме выборки, т. е. $M(\theta) = \theta_0$. Если это не так, то оценка называется смещенной.

Требование несмещенности гарантирует отсутствие систематических ошибок при оценивании.

Выборочная средняя $\overline{x}$ является несмещенной оценкой генеральной средней μ, т. е. $M(\overline{x}) = \mu$.

Выборочная дисперсия $\text{var}(x)$ является смещенной оценкой генеральной дисперсии σ^2, т. е. $M[\text{var}(x)] \neq \sigma^2$. В качестве несмещенной оценки генеральной дисперсии используется *исправленная дисперсия*

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}, \text{ для которой } M(s^2) = \sigma^2 .$$

Величина s называется *стандартным отклонением* случайной величины X в выборке.

Эффективность оценок. Несмещенная оценка θ называется эффективной, если она имеет минимальную дисперсию по сравнению с другими выборочными оценками, т. е. $\min\sigma^2(\theta)$.

Выборочная средняя $\overline{x}$ является эффективной оценкой генеральной средней μ, т. е. имеет наименьшую дисперсию в классе несмещенных оценок.

Состоятельность оценок. Оценка θ называется состоятельной, если при $n \to \infty$ она стремится по вероятности к оцениваемому параметру θ_0, т.е.

$$\lim_{n\to\infty} P(|\theta_n - \theta_0| < \varepsilon) = 1 .$$

Иначе говоря, состоятельной является такая оценка, которая дает точное значение для большой выборки.

Теорема Чебышева закона больших чисел утверждает, что

$$\lim_{n\to\infty} P(|\bar{x} - \mu| < \varepsilon) = 1 ,$$

т. е. выборочная средняя $\bar{x}$ является состоятельной оценкой генеральной средней μ.

Ковариация и корреляция

Различают выборочную и теоретическую ковариацию.

Выборочной ковариацией двух переменных x, y называется средняя величина произведения отклонений этих переменных от своих средних, т. е.

$$\operatorname{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}), \text{ или } \operatorname{cov}(x,y) = \overline{xy} - \bar{x}\,\bar{y},$$

где $\bar{x}, \bar{y}$ — выборочные средние переменных x, y, а $\overline{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n}$.

Выборочная ковариация является статистической мерой взаимосвязи между переменными.

Заметим, что $\operatorname{cov}(x,x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \operatorname{var}(x)$.

Теоретической ковариацией случайных величин X, Y называется математическое ожидание произведения отклонений этих величин от своих средних значений, т. е.

$$\operatorname{cov}(X, Y) = M[(X - \mu X)(Y - \mu Y)].$$

Заметим, что $\mathrm{cov}(X, X) = M(X - \mu X)^2 = \sigma_X^2$.

Ковариация $\mathrm{cov}(X, Y)$ есть мера связи двух случайных величин.

Если случайные величины X, Y *независимы*, то $\mathrm{cov}(X, Y) = 0$, а если $\mathrm{cov}(X, Y) \neq 0$, то X, Y *зависимы*.

Ковариация зависит от единиц, в которых измеряются переменные X и Y, она является ненормированной величиной. Поэтому для измерения силы связи между двумя переменными используется другая статистическая характеристика — коэффициент корреляции.

Различают теоретический и выборочный коэффициенты корреляции.

Теоретический коэффициент корреляции определяется выражением

$$\rho = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y},$$

где σ_X, σ_Y — стандартные отклонения случайных величин X, Y.

Коэффициент корреляции является безразмерной величиной, изменяющейся в пределах $-1 \leq \rho \leq 1$, и показывает степень линейной связи двух случайных величин. Коэффициент корреляции ρ является неслучайной величиной (истинной).

Случайные величины X, Y называются *некоррелированными*, если $\rho = 0$, и *коррелированными*, если $\rho \neq 0$.

Независимость случайных величин X, Y означает отсутствие любой связи (линейной и нелинейной), а некоррелированность — отсутствие только линейной связи.

Если случайные величины X, Y независимы, то они некоррелированы ($\rho = 0$), но из некоррелированности не следует их независимость, т. е. равенство $\rho = 0$ указывает лишь на отсутствие линейной связи между переменными, но не на отсутствие связи между ними вообще.

Выборочный коэффициент корреляции определяется выражением

$$r = \frac{\mathrm{cov}(x, y)}{\sqrt{\mathrm{var}(x)\,\mathrm{var}(y)}}.$$

При каждом конкретном значении ρ выборочный коэффициент корреляции является случайной величиной, изменяющейся в пределах $-1 \le r \le 1$.

Если $\rho = 0$ для генеральной совокупности, то это необязательно означает, что $r = 0$ для выборочной совокупности.

Пример 10.1. По исходным данным вычислим ковариацию и коэффициент корреляции между переменными x, y.

▼Представим исходные данные (x, y) и расчетные показатели в виде расчетной таблицы (табл. 10.1).

Таблица 10.1

Расчетная таблица для примера 10.1

i	x	y	x^2	xy	y^2
1	2	1	4	2	1
2	6	2	36	12	4
3	10	4	100	40	16
4	14	11	196	154	121
5	18	12	324	216	144
Итого	50	30	660	424	286
Среднее	10	6	132	84,8	57,2

Окончательно имеем:

$$\mathrm{var}(x) = \overline{x^2} - (\bar{x})^2 = 132 - 100 = 32;$$
$$\mathrm{var}(y) = \overline{y^2} - (\bar{y})^2 = 57,2 - 36 = 21,2;$$
$$\mathrm{cov}(x, y) = \overline{xy} - \bar{x}\,\bar{y} = 84,8 - 60 = 24,8;$$
$$r = \frac{\mathrm{cov}(x, y)}{\sqrt{\mathrm{var}(x)\,\mathrm{var}(y)}} = \frac{24,8}{\sqrt{32 \cdot 21,2}} = 0,952. \blacktriangle$$

Проверка статистических гипотез

Статистической гипотезой H называется предположение относительно параметра или вида распределения случайной величины.

Нулевой гипотезой H_0 называют выдвигаемую гипотезу. Обычно считают, что H_0 — гипотеза об отсутствии различий.

Конкурирующей гипотезой H_1 называют гипотезу, которая противоречит нулевой. Таким образом, H_1 — гипотеза о значимости различий.

Статистическим критерием называется случайная величина, которая служит для проверки нулевой гипотезы. В качестве статистического критерия выбирается такая случайная величина, точное или приближенное распределение которой известно.

Если в качестве критерия проверки нулевой гипотезы используется случайная величина, подчиненная распределению Стьюдента, то ее обозначают через t (t-статистика), а подчиненная распределению Фишера — через F (F-статистика).

Проверку статистической гипотезы выполняют на основании результатов выборки. Поскольку выборка производится случайным образом, то появляется возможность принятия ошибочного решения.

Уровнем значимости α называется вероятность того, что будет отвергнута правильная нулевая гипотеза, т. е. $P_{H_0}(\overline{H_0}) = \alpha$.

Уровень значимости α устанавливается заранее.

В экономических исследованиях проверку гипотез осуществляют при 5%-м и 1%-м уровнях значимости (стандартные уровни).

Выбор, например, 5%-го уровня значимости означает, что в пяти случаях применения критерия из ста верная гипотеза будет отвергнута.

Расчетным значением критерия называют значение критерия, вычисленное по данным выборки.

Для проверки гипотезы используют *критические точки*, вычисляемые по известному распределению критерия.

Из сравнения расчетного значения критерия с критическими точками делается вывод о принятии или отклонении нулевой гипотезы.

Правила проверки нулевой гипотезы

***t*-статистика**. Определяется расчетное значение критерия $t_р$. Критическая точка $t_{кр}$ при заданном уровне значимости α и числе степеней свободы ν определяется по таблице t-распределения

Стьюдента для двусторонней критической области или с помощью функции *Excel* СТЬЮДРАСПОБР(α; ν).

Из сравнения $t_р$ и $t_{кр}$, получаем:

- если $t_р < t_{кр}$, то H_0 принимается;
- если $t_р > t_{кр}$, то H_0 отклоняется.

***F*-статистика.** Определяется расчетное значение критерия $F_р$. Критическая точка $F_{кр}$ при заданном уровне значимости α и числе степеней свободы ν_1, ν_2 определяется по таблице F-распределения Фишера для правосторонней критической области или с помощью функции *Excel* FРАСПОБР(α; ν_1; ν_2).

Из сравнения расчетного значения $F_р$ с критическим $F_{кр}$, получаем:

- если $F_р < F_{кр}$, то H_0 принимается;
- если $F_р > F_{кр}$, то H_0 отклоняется.

В статистических пакетах вместо критических точек используется p-уровень значимости, соответствующий вычисленному значению критерия.

***p*-уровень** — это вероятность того, что наблюдаемые различия случайны. Чем меньше эта величина, тем выше статистическая значимость результата (различия, связи).

В *Excel* принято обозначение:

$$p\text{-уровень} = \begin{cases} P\text{-значение}, & (t\text{-статистика}) \\ \text{Значимость } F, & (F\text{-статистика}) \end{cases},$$

где

$$P\text{-значение} = \text{СТЬЮДРАСП}(t_р;\ \nu;\ 2);$$
$$\text{Значимость } F = \text{FРАСП}(F_р;\ \nu_1;\ \nu_2).$$

В общем случае при проверке статистических гипотез с помощью статистических пакетов выводится значение p-уровня значимости.

Если вычисленный p-уровень меньше заданного стандартного уровня ($p < \alpha$), то нулевая гипотеза отклоняется; в противном случае ($p > \alpha$) нулевая гипотеза принимается.

10.3. Модель парной линейной регрессии

В модели парной линейной регрессии зависимость между переменными в генеральной совокупности представляется в виде

$$Y = \alpha + \beta X + \varepsilon,$$

где X — неслучайная величина;

Y, ε — случайные величины.

Величина Y называется объясняемой (зависимой) переменной или результативным признаком, а X — объясняющей (независимой) переменной или факторным признаком. Постоянные α, β — параметры уравнения.

Наличие случайного члена ε (*ошибки регрессии*) связано с воздействием на зависимую переменную других, не учтенных в уравнении факторов.

На основе выборочного наблюдения объема n оценивается выборочное уравнение регрессии (*линия регрессии*):

$$\hat{y} = a + bx,$$

где a, b — оценки параметров α, β.

Метод наименьших квадратов. Рассмотрим задачу наилучшей аппроксимации набора наблюдений (x_1, y_1), (x_2, y_2), ... линейным уравнением (линией регрессии).

На рис. 10.2 приведены диаграмма рассеяния наблюдений и линия регрессии.

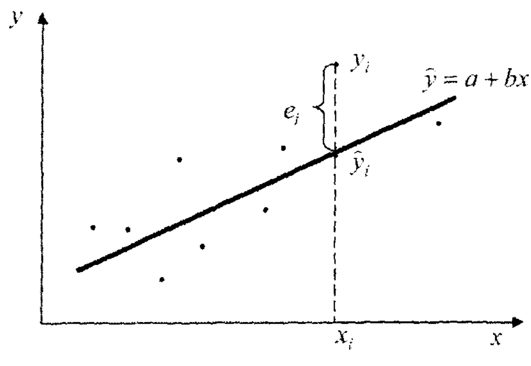


Рис. 10.2. Диаграмма рассеяния наблюдений и линия регрессии

Расчетные значения $\hat{y}_i$ зависимой переменной в каждом наблюдении лежат на линии регрессии (прямой) и не совпадают в целом с соответствующими наблюдаемыми значениями y_i.

Определим *остаток* e_i в i-м наблюдении как разность между наблюдаемым и расчетным значениями зависимой переменной, т. е.

$$e_i = y_i - \hat{y}_i .$$

Неизвестные значения a, b определяются методом наименьших квадратов (МНК), сущность которого заключается в *минимизации суммы квадратов остатков*:

$$Q = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - a - bx_i)^2 \to \min ,$$

где (x_i, y_i) — известные значения (числа);

a, b — неизвестные.

Запишем необходимые условия экстремума:

$$\begin{cases} Q'_a = -2\sum (y_i - a - bx_i) = 0, \\ Q'_b = -2\sum (y_i - a - bx_i)x_i = 0. \end{cases}$$

После преобразования получим систему нормальных уравнений

$$\begin{cases} a + b\overline{x} = \overline{y}, \\ a\overline{x} + b\overline{x^2} = \overline{xy}, \end{cases}$$

решение которой есть

$$\begin{cases} b = \dfrac{\operatorname{cov}(x, y)}{\operatorname{var}(x)} = \dfrac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - (\overline{x})^2}, \\ a = \overline{y} - b\overline{x}. \end{cases}$$

Таким образом, линия регрессии (расчетное значение зависимой переменной) есть

$$\hat{y} = a + bx \text{, или } \hat{y} - \overline{y} = b(x - \overline{x}) .$$

Линия регрессии проходит через точку $(\overline{x}, \overline{y})$, и выполняются равенства

$$\overline{y} = \overline{\hat{y}}, \overline{e} = 0 .$$

Коэффициент b есть *угловой коэффициент линии регрессии*, который показывает, на сколько единиц в среднем изменяется переменная y при увеличении независимой переменной x на единицу.

Постоянная a дает *прогнозируемое значение* зависимой переменной при $x = 0$.

После построения уравнения регрессии наблюдаемые значения y можно представить в виде $y_i = \hat{y}_i + e_i$. Остатки e_i, как и ошибки ε_i, являются случайными величинами, однако они, в отличие от ошибок ε_i, являются наблюдаемыми.

Определим выборочные дисперсии величин $y, \hat{y}, e$:

$$\operatorname{var}(y) = \frac{1}{n}\sum (y_i - \bar{y})^2$$ — дисперсия *наблюдаемых значений y*;

$$\operatorname{var}(\hat{y}) = \frac{1}{n}\sum (\hat{y}_i - \bar{y})^2$$ — дисперсия *расчетных значений y*;

$$\operatorname{var}(e) = \frac{1}{n}\sum (y_i - \hat{y}_i)^2$$ — дисперсия *остатков e*.

Разброс наблюдаемых значений зависимой переменной характеризуется выборочной дисперсией var(y).

Разложим дисперсию var(y) на две части:

$$\operatorname{var}(y) = \operatorname{var}(\hat{y} + e) = \operatorname{var}(\hat{y}) + \operatorname{var}(e),$$

где $\operatorname{var}(\hat{y})$ — часть, *объясненная* регрессионным уравнением;

$\operatorname{var}(e)$ — *необъясненная* часть.

В общем случае при наличии m независимых переменных и свободного члена имеет место следующее уравнение:

$$\underset{(n-1)}{SS_Y} = \underset{(m)}{SS_\text{p}} + \underset{(n-m-1)}{SS_e},$$

где в скобках указано число степеней свободы, соответствующее каждому члену уравнения;

$SS_Y = \sum (y_i - \bar{y})^2$ — общий разброс зависимой переменной;

$SS_р = \sum(\hat{y}_i - \bar{y})^2$ — разброс, объясненный регрессией;

$SS_e = \sum e_i^2 = \sum(y_i - \hat{y}_i)^2$ — разброс, не объясненный регрессией.

Любая сумма квадратов связана с числом степеней свободы, т. е. с числом независимого варьирования переменной. Существует равенство между числом степеней свободы для этого уравнения. Отнесение каждой суммы квадратов этого уравнения на одну степень свободы приводит их к сравнимому виду.

Рассчитываются средние квадраты MS:

$$MS_р = \frac{SS_р}{m}, \; MS_e = \frac{SS_e}{n-m-1}.$$

Коэффициентом детерминации R^2 называется отношение

$$R^2 = \frac{\mathrm{var}(\hat{y})}{\mathrm{var}(y)} = \frac{SS_р}{SS_y}, \; 0 \le R^2 \le 1,$$

характеризующее долю вариации (разброса) зависимой переменной, объясненную уравнением регрессии:

- если $R^2 = 1$, то это означает точную подгонку, т. е. все точки наблюдения лежат на регрессионной прямой;
- если $R^2 = 0$, то регрессия ничего не дает, т. е. переменные x не улучшают качество предсказания y по сравнению с горизонтальной прямой $\hat{y} = \bar{y}$.

Чем ближе к единице R^2, тем лучше качество подгонки, т. е. $\hat{y}$ более точно аппроксимирует y.

***F*-тест на качество оценивания.** Для определения статистической значимости коэффициента детерминации R^2 проверяется гипотеза H_0: $F = 0$ для F-статистики, рассчитываемой по формуле

$$F = \frac{MS_р}{MS_e} = \frac{SS_р}{SS_e} \cdot \frac{n-m-1}{m} = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}.$$

Величина F имеет распределение Фишера с $\nu_1 = m$, $\nu_2 = n - m - 1$ степенями свободы. По установленным правилам F-статистики производится процедура проверки нулевой гипотезы.

Значения *SS* (сумма квадратов), *MS* (средний квадрат), *F*-критерий и его значимость обычно приводятся в таблице «Дисперсионный анализ».

Пример 10.2. Построим регрессионную зависимость расходов на питание от личного дохода по данным табл. 10.2 и оценим качество подгонки.

Таблица 10.2

Исходные данные для примера 10.2, усл. ед.

Показатель	Год				
	2011	2012	2013	2014	2015
Расходы на питание (X)	2	6	10	14	18
Личный доход (Y)	1	2	4	11	12

▼Пусть истинная модель описывается выражением $Y = \alpha + \beta X + \varepsilon$, по выборочным наблюдениям определим оценки a, b.

Исходные данные и расчетные показатели представим в *Excel* в виде табл. 10.3.

Таблица 10.3

Расчетная таблица для примера 10.2

Год	x	y	x^2	xy	$\hat{y}$	$(y-\bar{y})^2$	$(\hat{y}-\bar{y})^2$	$(y-\hat{y})^2$
2011	2	1	4	2	–0,2	25	38,44	1,44
2012	6	2	36	12	2,9	16	9,61	0,81
2013	10	4	100	40	6	4	0	4
2014	14	11	196	154	9,1	25	9,61	3,61
2015	18	12	324	216	12,2	36	38,44	0,04
Итого	50	30	660	424	30	106	96,1	9,9
Среднее	10	6	132	84,8	6	21,2 (var(y))	19,22 (var($\hat{y}$))	1,98 (var(e))

Окончательно имеем:

$$\text{var}(x) = \overline{x^2} - (\overline{x})^2 = 132 - 100 = 32;$$

$$\text{cov}(x, y) = \overline{xy} - \overline{x} \cdot \overline{y} = 84{,}8 - 10 \cdot 6 = 24{,}8;$$

$$b = \frac{\text{cov}(x, y)}{\text{var}(x)} = 24{,}8/32 = 0{,}775;$$

$$a = \overline{y} - b\overline{x} = 6 - 0{,}775 \cdot 10 = -1{,}75,$$

следовательно, $\hat{y} = -1{,}75 + 0{,}775x$ — линия регрессии.

Коэффициент $b = 0{,}775$ показывает, что при увеличении дохода на 1 усл. ед. расходы на питание увеличиваются в среднем на 0,775 усл. ед.

Качество подгонки оцениваем коэффициентом детерминации:

$$R^2 = \frac{\text{var}(\hat{y})}{\text{var}(y)} = \frac{19{,}22}{21{,}2} = 0{,}907,$$

т. е. 90,7 % разброса зависимой переменной (расходы на питание) объясняется регрессией.

Значимость коэффициента R^2 проверяем по F-тесту.

Зададим $\alpha = 0{,}05$. Число степени свободы $\nu_1 = m = 1$, $\nu_2 = n - m - 1 = 3$, где n — число наблюдений, m — число объясняющих переменных.

Расчетное значение критерия $F_\text{р}$ есть

$$F_\text{р} = \frac{R^2(n-2)}{1-R^2} = \frac{0{,}907 \cdot 3}{0{,}093} = 29{,}12.$$

Критическое значение критерия определяем с помощью функции

$$F_\text{кр} = \text{FРАСПОБР}(\alpha; \nu_1; \nu_2) = \text{FРАСПОБР}(0{,}05; 1; 3) = 10{,}13.$$

Поскольку $F_\text{р} = 29{,}12 > F_\text{кр} = 10{,}13$, то H_0 отклоняется, т. е. $R^2 = 0{,}907$ статистически значим.

Этот же вывод следует, если использовать значимость F:

$$\text{Значимость } F = \text{FРАСП}(29{,}12;\ 1;\ 3) = 0{,}0125.$$

Поскольку значимость $F = 0{,}0125 < \alpha = 0{,}05$, то R^2 значим при уровне 5 %.

Для данного примера приведем таблицу дисперсионного анализа (табл. 10.4).

Таблица 10.4

Дисперсионный анализ для примера 10.2

Компонента	**Число степеней свободы (*df*)**	**Сумма квадратов отклонений (*SS*)**	**Средний квадрат (*MS*)**	**Расчетное значение критерия ($F_р$)**	**Значимость *F***
Регрессия	1	96,1	96,1	29,12	0,01247
Остаток	3	9,9	3,3		
Итого	4	106			

▲

Предпосылки регрессионного анализа

Линейная регрессионная модель с двумя переменными имеет вид

$$Y_i = \alpha + \beta X_i + \varepsilon_i,\ (i = \overline{1,n}),$$

где Y — объясняемая переменная;

X — объясняющая переменная;

ε — случайный член.

Для того чтобы регрессионный анализ, основанный на МНК, давал наилучшие из всех возможных результаты, должны выполняться определенные условия (*условия Гаусса — Маркова*).

1. Математическое ожидание случайного члена в любом наблюдении должно быть равно нулю, т. е.

$$M(\varepsilon_i) = 0,\ (i = \overline{1,n}).$$

2. Дисперсия случайного члена должна быть постоянной для всех наблюдений, т. е.

$$D(\varepsilon_i) = \sigma^2,\ (i = \overline{1,n}).$$

3. Случайные члены должны быть статистически независимы (некоррелированы) между собой, т. е.

$$M(\varepsilon_i\varepsilon_j) = 0, (i \neq j).$$

4. Объясняющая переменная X есть величина неслучайная.

При выполнении условий Гаусса — Маркова модель называется *классической нормальной линейной регрессионной моделью.*

Наряду с условиями Гаусса — Маркова обычно предполагается, что случайный член распределен нормально, т. е. $\varepsilon_i \sim N(0; \sigma^2)$.

Рассмотрим подробнее условия и предположения, лежащие в основе регрессионного анализа.

Первое условие означает, что случайный член не должен иметь систематического смещения. Если постоянный член α включен в уравнение регрессии, то это условие автоматически выполняется.

Второе условие означает, что дисперсия случайного члена в каждом наблюдении имеет только одно значение σ^2.

Условие независимости дисперсии случайного члена от номера наблюдения называется *гомоскедастичностью* (одинаковый разброс).

Зависимость дисперсии случайного члена от номера наблюдения называется *гетероскедастичностью.*

Таким образом:

- $D(\varepsilon_i) = \sigma^2, (i = \overline{1,n})$ — *гомоскедастичность*;
- $D(\varepsilon_i) = \sigma_i^2, (i = \overline{1,n})$ — *гетероскедастичность.*

Если условие гомоскедастичности не выполняется, то оценки коэффициентов регрессии будут неэффективными, хотя и несмещенными.

Существуют специальные методы диагностирования и устранения гетероскедастичности.

Третье условие указывает на некоррелированность случайных членов для каждых двух соседних наблюдений. Это условие часто нарушается, когда данные являются временными рядами. В случае когда третье условие не выполняется, говорят об *автокорреляции остатков.*

Если условие независимости случайных членов не выполняется, то оценки коэффициентов регрессии, полученные по МНК, оказываются неэффективными, хотя и несмещенными.

Существуют методы диагностирования и устранения автокорреляции.

Четвертое условие о неслучайности объясняющей переменной является особенно важным. Если это условие нарушается, то оценки коэффициентов регрессии оказываются смещенными и несостоятельными.

Нарушение этого условия может быть связано с ошибками измерения объясняющих переменных или с использованием лаговых переменных.

Предположение о нормальности распределения случайного члена необходимо для проверки значимости параметров регрессии и для их интервального оценивания.

Теорема Гаусса — Маркова. Если предпосылки регрессионного анализа выполняются, то оценки (a, b), сделанные с помощью МНК, являются:

• *несмещенными*: $M(a) = \alpha$, $M(b) = \beta$, что означает отсутствие систематической ошибки в положении линии регрессии;

• *эффективными*: имеют наименьшую дисперсию в классе всех линейных несмещенных оценок, равную

$$D(a) = \frac{\sigma^2 \overline{x^2}}{n \operatorname{var}(x)}, \quad D(b) = \frac{\sigma^2}{n \operatorname{var}(x)};$$

• *состоятельными:* $\lim\limits_{n \to \infty} D(a) = 0, \lim\limits_{n \to \infty} D(b) = 0$, т. е. при достаточно большом n оценки (a, b) близки к (α, β).

Расчет стандартной ошибки коэффициентов регрессии. Полученные теоретические дисперсии $D(a)$, $D(b)$ зависят от дисперсии σ^2 случайного члена. По данным выборки отклонения ε_i, а следовательно, и их дисперсии σ^2 неизвестны, поэтому они заменяются наблюдаемыми остатками e_i и их выборочной дисперсией $\operatorname{var}(e)$. Однако оценка $\operatorname{var}(e)$ является смещенной, т. е. $M[\operatorname{var}(e)] \neq \sigma^2$.

Несмещенной оценкой дисперсии σ^2 является *остаточная дисперсия*

$$s^2 = \sum_{i=1}^{n} e_i^2 / (n-2),$$

которая служит мерой разброса зависимой переменной вокруг линии регрессии. Отметим, что в знаменателе остаточной дисперсии стоит число степеней свободы ($n-2$), а не n, так как две степени свободы теряются при определении двух параметров (a; b).

Величина s называется *стандартной ошибкой регрессии.*

Заменив в теоретических дисперсиях неизвестную σ^2 на оценку s^2, получим оценки дисперсий:

$$s_a^2 = \frac{\overline{x^2}\, s^2}{n \operatorname{var}(x)}, \quad s_b^2 = \frac{s^2}{n \operatorname{var}(x)}.$$

Величины s_a, s_b называются стандартными ошибками коэффициентов регрессии.

Проверка гипотезы H_0: $\beta = \beta_0$. Пусть в теоретической зависимости $Y = \alpha + \beta X + \varepsilon$ случайный член ε распределен нормально с неизвестной дисперсией σ^2. Хотя величина β неизвестна, имеется основание предполагать, что она равна заданной величине β_0. Выдвигаются гипотезы

$$\begin{cases} H_0:\ \beta = \beta_0, \\ H_1:\ \beta \neq \beta_0. \end{cases}$$

Задача заключается в проверке нулевой гипотезы на основании выборочных данных.

Пусть по выборочным данным получена оценка b.

В качестве *критерия проверки нулевой гипотезы* принимают случайную величину

$$t = \frac{b - \beta_0}{s_b},$$

которая имеет распределение Стьюдента с $\nu = n - 2$ степенями свободы.

По таблице критических точек распределения Стьюдента по заданному уровню значимости α и числу ν степеней свободы находят критическую точку $t_{кр}$. Сравнивая наблюдаемое значение критерия с критическим, можно принять или отвергнуть нулевую гипотезу.

Результаты оценивания регрессии совместимы не только с конкретной гипотезой H_0: $\beta = \beta_0$, но и с некоторым их множеством. Любое значение β, совместимое с оценкой b, удовлетворяет условию

$$\left|\frac{b-\beta}{s_b}\right| < t_{\text{кр}}, \text{ или } -t_{\text{кр}} < \frac{\beta - b}{s_b} < t_{\text{кр}}$$

Разрешив это неравенство относительно β, получим:

$$b - t_{\text{кр}}s_b < \beta < b + t_{\text{кр}}s_b,$$

т. е. доверительный интервал для величины β.

Посредине интервала лежит величина b. Границы интервала одинаково отстоят от b, зависят от выбора уровня значимости и являются случайными числами.

Доверительный интервал покрывает значение параметра β с заданной вероятностью $(1 - \alpha)$, т. е. $P(b - t_{\text{кр}}s_b < \beta < b + t_{\text{кр}}s_b) = 1 - \alpha$.

Проверка гипотезы H_0: $\beta = 0$. Пусть по выборке получена оценка коэффициента *регрессии* b. *Для определения* статистической значимости коэффициента регрессии b проверяется гипотеза H_0: $\beta = 0$ для t-статистики, рассчитываемой по формуле

$$t = b/s_b.$$

Величина t имеет распределение Стьюдента с $\nu = n - 2$ степенями свободы.

По установленным правилам производится процедура проверки нулевой гипотезы.

Пример 10.3. По данным и результатам примера 10.2 рассчитаем стандартную ошибку, оценим значимость и построим доверительный интервал коэффициентов модели.

▼К результатам примера 10.2 добавятся расчеты следующих показателей.

Стандартная ошибка регрессии:

$$s = \sqrt{\frac{\sum(y-\hat{y})^2}{n-2}} = \sqrt{\frac{9,9}{3}} = 1,817.$$

Стандартная ошибка коэффициента b:

$$s_b = \frac{s}{\sqrt{n\,\text{var}(x)}} = \frac{1,817}{\sqrt{5\cdot 32}} = 0,144.$$

Расчетное значение критерия $t_р = b/s_b = 0{,}775/0{,}144 = 5{,}4$.

Число степени свободы $\nu_2 = n - 2 = 3$.

Критическое значение $t_{кр}$ = СТЬЮДРАСПОБР(0,05; 3) = 3,18.

Поскольку $t_р = 5{,}4 > t_{кр} = 3{,}18$, то $b = 0{,}775$ значим на уровне 5 %.

Этот же вывод следует, если использовать P-значение:

P-значение = СТЬЮДРАСП(5,4; 3; 2) = 0,0124.

Поскольку P-значение $= 0{,}0124 < \alpha = 0{,}05$, то $b = 0{,}775$ значим.

Доверительный интервал для b есть $(b - t_{кр}s_b;\ b + t_{кр}s_b) =$ $= (0{,}775 - 3{,}18 \cdot 0{,}144;\ 0{,}775 + 3{,}18 \cdot 0{,}144) = (0{,}32;\ 1{,}23)$.

Проведем подобные расчеты для коэффициента $a = -1{,}75$:

$$s_a = s \cdot \sqrt{\frac{\overline{x^2}}{n\,\mathrm{var}(x)}} = 1{,}817 \cdot \sqrt{\frac{132}{5 \cdot 32}} = 1{,}65\,;$$

$$t_р = a/s_a = -1{,}75/1{,}65 = -1{,}06;$$

P-значение = СТЬЮДРАСП(1,06; 3; 2) = 0,367.

Поскольку P-значение $= 0{,}367 > \alpha = 0{,}05$, то $a = -1{,}75$ незначим.

Доверительный интервал: $(a - t_{кр}s_a;\ a + t_{кр}s_a) = (-7{,}00;\ 3{,}50)$. ▲

Взаимозависимость критериев. В парном регрессионном анализе эквивалентны критерии

$$t_b = \frac{b}{s_b},\quad t_r = r\sqrt{\frac{n-2}{1-r^2}},\quad F = \frac{R^2(n-2)}{1-R^2}\,.$$

Связь между критериями выражается равенством

$$t_b = t_r = \sqrt{F}\,,$$

причем для критических значений критериев при любом уровне значимости

$$(t_b)_{кр} = (t_r)_{кр} = \sqrt{F_{кр}}\,,$$

и эти критерии дают один и тот же результат.

Вывод. В парной линейной регрессии проверки значимости коэффициента b, коэффициента корреляции r и коэффициента детерминации R^2 эквивалентны.

Надстройка
Пакет анализа Excel (Регрессия)

Расчет параметров уравнения линейной регрессии, проверку их статистической значимости и построение интервальных оценок можно выполнить значительно быстрее автоматически при использовании *Пакета анализа Excel* (инструмент *Регрессия*).

Пусть исходные данные примера 10.2 (расходы на питание ↔ ↔ личный доход) представлены в *Excel.*

Выбираем команду *Анализ данных* ⇒ *Регрессия*.

В диалоговом окне инструмента *Регрессия* задаются следующие параметры:

- *Входной интервал Y* — вводится ссылка на ячейки, содержащие данные по результативному признаку;
- *Входной интервал X* — вводится ссылка на ячейки, содержащие факторные признаки;
- *Метки* — поставьте галочку, если выделены и заголовки столбцов;
- *Константа — ноль* — поставьте галочку, если оценивайте регрессионное уравнение без свободного члена.

При необходимости задаются и другие параметры.

Результаты расчетов с использованием инструмента *Регрессия* выводятся в виде таблиц под общим названием «Вывод итогов» (рис. 10.3).

Результаты работы инструмента *Регрессия* полностью совпадают с ранее рассчитанными.

При необходимости выводятся предсказанные значения $\hat{y}_i$ результативного признака и значения остатков (рис. 10.4).

Коэффициенты регрессии, их стандартные ошибки и коэффициент детерминации составляют:

$$a = -1{,}75;\ b = 0{,}775;\ s_a = 1{,}65;\ s_b = 0{,}144;\ R^2 = 0{,}907.$$

Регрессионная статистика	
Множественный *R*	0,952
R-квадрат	0,907
Нормированный R-квадрат	0,875
Стандартная ошибка	1,817
Наблюдения	5

Дисперсионный анализ					
	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	96,1	96,1	29,12	0,01247
Остаток	3	9,9	3,3		
Итого	4	106			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	-1,75	1,65	-1,06	0,36669	-7,001	3,501
Переменная X1	0,775	0,14361	5,40	0,01247	0,318	1,232

Рис. 10.3. Результат работы инструмента *Регрессия* в *Excel* для примера 10.2

ВЫВОД ОСТАТКА		
Наблюдение	*Предсказанное Y*	*Остатки*
1	-0,2	1,2
2	2,9	-0,9
3	6	-2
4	9,1	1,9
5	12,2	-0,2

Рис. 10.4. Вывод остатка с помощью инструмента *Регрессия* в *Excel*

Результаты регрессионного анализа принято записывать в виде

$$\hat{y} = -1,75 + 0,775x; \; R^2 = 0,907,$$
$$(1,65)\;(0,144)$$

где в скобках указаны стандартные ошибки коэффициентов регрессии.

Статистическая значимость коэффициента $R^2 = 0,907$ устанавливается по F-тесту. Поскольку значимость $F = 0,0125 < \alpha = 0,05$, то $R^2 = 0,907$ значим при уровне 5 %. Модель в целом значима.

Обычно проверка значимости коэффициента a не производится. Оценим статистическую значимость коэффициента b.

Поскольку P-значение = 0,0125 < 0,05, то коэффициент $b = 0,775$ значим на уровне 5 %.

Результаты оценивания регрессии совместимы не только с полученным значением коэффициента регрессии $b = 0,775$, но и с некоторым его множеством (доверительным интервалом). С вероятностью 95 % доверительный интервал коэффициента b есть (0,318; 1,232).

Упражнение 10.1. Используя *Пакет анализа Excel* (инструмент *Регрессия*), по данным примера 10.2 построить модель регрессии без свободного члена и проверить ее адекватность.

Ответ: $\hat{y} = 0,064x,\ R^2 = 0,952.$

Пример 10.4. Имеются данные о расходах на питание и душевом доходе для девяти групп семей (табл. 10.5).

Таблица 10.5

Исходные данные для примера 10.4, усл. ед.

Душевой доход (x)	63	158	260	370	480	593	728	935	1880
Расходы на питание (y)	43	62	90	111	130	149	165	191	241

Используя результаты работы инструмента *Регрессия*, проанализируем зависимость расходов на питание от величины душевого дохода.

▼ Результаты регрессионного анализа записываем в виде:

$$\hat{y} = 66,04 + 0,107x,\ R^2 = 0,885,$$
$$(11,72)\ (0,015)$$

где в скобках указаны стандартные ошибки коэффициентов регрессии.

Качество модели оценивается коэффициентом детерминации R^2.

Величина $R^2 = 0,885$ означает, что фактором душевого дохода можно объяснить 88,5 % вариации (разброса) расходов на питание.

Установим статистическую значимость коэффициента R^2.

Поскольку значимость $F = 0{,}000158 < \alpha = 0{,}05$, то $R^2 = 0{,}885$ значим при уровне 5 %.

Направление связи между y и x определяет знак коэффициента $b = 0{,}107 > 0$, т. е. связь является прямой (положительной).

Коэффициент $b = 0{,}107$ показывает, что при увеличении душевого дохода на 1 усл. ед. расходы на питание в среднем увеличиваются на 0,107 усл. ед.

Оценим статистическую значимость коэффициента b. Поскольку P-значение $= 0{,}000158 < \alpha = 0{,}05$, то коэффициент $b = 0{,}107$ значим на 5%-м уровне. С вероятностью 95 % доверительный интервал коэффициента b есть (0,073; 0,142). ▲

Упражнение 10.2. По десяти туристическим фирмам были установлены затраты на рекламу и количество туристов, воспользовавшихся услугами фирм (табл. 10.6).

Таблица 10.6

Исходные данные для упражнения 10.2

Туристическая фирма	Затраты на рекламу, усл. ед. (x)	Количество туристов (y)
1	1	8
2	1	16
3	3	15
4	3	25
5	6	20
6	6	25
7	6	27
8	8	25
9	8	30
10	8	29

Используя результаты работы инструмента *Регрессия*, проанализируйте зависимость числа туристов, воспользовавшихся услугами туристической фирмы, от затрат на рекламу.

10.4. Нелинейные регрессии

Одним из недостатков линейного регрессионного анализа, как это следует из самого названия, является то, что он может быть применим только к линейным уравнениям. В случае простого регрессионного анализа речь идет об уравнении вида $y = \alpha + \beta x$, состоящем из постоянной величины (которая может и отсутствовать), независимой переменной, умноженной на некоторый коэффициент, и из случайного остаточного члена возмущения, которым можно временно пренебречь.

В общем случае линейное уравнение выглядит так, что каждый объясняющий элемент, за исключением постоянной величины, записан в виде произведения постоянной и коэффициента: $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots$

Уравнения вида $y = \alpha + \beta\sqrt{x}$, $y = \alpha + \beta / x$ и $y = \alpha x^{\beta}$ являются нелинейными, и задача состоит в определении параметров α и β в каждом уравнении по исходным данным (y, x).

Нелинейность регрессии проявляется как по переменным, так и по параметрам.

Нелинейность по переменным устраняется путем замены переменной. Например, нелинейное уравнение $y = \alpha + \beta\sqrt{x} + \varepsilon$ после замены переменной $z = \sqrt{x}$ становится линейным: $y = \alpha + \beta z + \varepsilon$, и для оценки его параметров используется МНК.

Нелинейность по параметру часто устраняется путем логарифмического преобразования уравнения. Например, следующие нелинейные уравнения после логарифмирования сводятся к линейным:

- степенная функция $y = \alpha x^{\beta}\varepsilon \sim \ln y = \ln\alpha + \beta\ln x + \ln\varepsilon$;
- экспоненциальная функция $y = \alpha e^{\beta x}\varepsilon \sim \ln y = \ln\alpha + \beta x + \ln\varepsilon$.

Пример 10.5. Имеются данные о ежегодном потреблении бананов y и годовом доходе x десяти американских семей (табл. 10.7).

Таблица 10.7

Исходные данные для примера 10.5

Годовой доход (x)	1	2	3	4	5	6	7	8	9	10
Ежегодное потребление бананов (y)	2	8	7	12	9	11	10	8	13	9

Рассмотрим следующие варианты уравнения регрессии:

a)
1) $y = \alpha + \beta x + \varepsilon$;
2) $y = \alpha + \beta\sqrt{x} + \varepsilon$;
3) $y = \alpha + \beta / x + \varepsilon$;

b)
1) $y = \alpha x^{\beta}\varepsilon$;
2) $y = \alpha e^{\beta x}\varepsilon$.

▼*a*) Исходные и преобразованные данные для построения искомых уравнений регрессии представлены в табл. 10.8.

Таблица 10.8

Исходные и преобразованные данные для примера 10.5 (вариант *a*)

y	x	$\sqrt{x}$	$1/x$
2	1	1,0000	1,0000
8	2	1,4142	0,5000
7	3	1,7321	0,3333
12	4	2,0000	0,2500
9	5	2,2361	0,2000
11	6	2,4495	0,1667
10	7	2,6458	0,1429
8	8	2,8284	0,1250
13	9	3,0000	0,1111
9	10	3,1623	0,1000

1) Оценка первого уравнения по исходным данным (y, x):

$$\hat{y} = 5,60 + 0,60x,\ R^2 = 0,350,\ s = 2,63.$$

2) Оценка второго уравнения по преобразованным данным $(y, \sqrt{x})$:

$$\hat{y} = 2,39 + 2,90\sqrt{x},\ R^2 = 0,446,\ s = 2,42.$$

3) Оценка третьего уравнения по преобразованным данным (y, $1/x$):

$$\hat{y} = 11,58 - 9,15 / x,\ R^2 = 0,682,\ s = 1,84.$$

Качество оценивания последнего варианта уравнения выше других.

b) Исходные и преобразованные данные для построения искомых уравнений регрессии представлены в табл. 10.9.

1) Оценив регрессию между ($\ln y$, $\ln x$), получим преобразованное уравнение $\ln\hat{y} = 1,24 + 0,56\ln x$. Выполнив обратные преобразования, получим

$$\hat{y} = 3,47x^{0,56},\ R^2 = 0,610,\ s = 2,7.$$

Таблица 10.9

Исходные и преобразованные данные для примера 10.5 (вариант *b*)

y	x	$\ln y$	$\ln x$
2	1	0,693	0,000
8	2	2,079	0,693
7	3	1,946	1,099
12	4	2,485	1,386
9	5	2,197	1,609
11	6	2,398	1,792
10	7	2,303	1,946
8	8	2,079	2,079
13	9	2,565	2,197
9	10	2,197	2,303

2) Оценив регрессию между ($\ln y$, x), получим преобразованное уравнение $\ln\hat{y} = 1,52 + 0,105x$. Выполнив обратные преобразования, получим

$$\hat{y} = 4,57e^{0,105x}, R^2 = 0,359, s = 3,03. \blacktriangle$$

В экономике функции вида $y = \alpha\, x^{\beta}\varepsilon$ применяются при моделировании кривых спроса, а вида $y = \alpha e^{\beta x}\varepsilon$ — при моделировании временных трендов, при этом вместо x используется время t, а вместо β — постоянный темп прироста r, т. е. $y = \alpha e^{rt}\varepsilon$.

В экономическом анализе часто используется эластичность функции $y = f(x)$, которая рассчитывается как отношение относительных изменений y и x:

$$Э = \left(\frac{dy}{y}\right) / \left(\frac{dx}{x}\right) = \frac{x}{y} f'(x).$$

Эластичность функции показывает, на сколько процентов изменяется функция $y = f(x)$ при изменении независимой переменной на 1 %.

Для степенной функции $y = ax^b$ эластичность представляет собой постоянную величину, равную b.

В силу того что эластичность линейной функции $y = a + bx$ не является постоянной величиной, а зависит от x, т. е.

$$Э = b\frac{x}{y},$$

обычно вычисляется средний показатель эластичности по формуле

$$\bar{Э} = b\frac{\bar{x}}{\bar{y}},$$

где $\bar{x}, \bar{y}$ — средние значения переменных x, y в выборке.

Например, для зависимости расходов на питание от доходов $\hat{y} = -1,75 + 0,775x$ $(\bar{x} = 10, \bar{y} = 6)$ средний показатель эластичности есть 1,29. Он показывает, что с увеличением дохода на 1 % расходы на питание возрастут в среднем на 1,29 %.

Пример 10.6. По данным примера 10.2 построим зависимость расходов на питание от доходов в виде: *а*) степенной функции; *b*) экспоненциального временного тренда.

▼Исходные и преобразованные данные для построения указанных уравнений представлены в табл. 10.10.

Таблица 10.10

Исходные и преобразованные данные для примера 10.6

t	x	y	$\ln x$	$\ln y$
1	2	1	0,69315	0
2	6	2	1,79176	0,69315
3	10	4	2,30259	1,38629
4	14	11	2,63906	2,39790
5	18	12	2,89037	2,48490

а) Оценив регрессию между ($\ln y$, $\ln x$), получим преобразованное уравнение $\ln\hat{y} = -1,0496 + 1,183\ln x$. Выполнив обратные преобразования, получим $\hat{y} = 0,350x^{1,183}$.

Эластичность спроса на продукты питания по доходу составляет 1,183.

Это означает, что увеличение дохода на 1 % приведет к увеличению расходов на питание на 1,183 %.

b) Оценив регрессию между ($\ln y$, t), получим преобразованное уравнение $\ln \hat{y} = -0,61 + 0,667t$. Выполнив обратные преобразования, получим $\hat{y} = 0,543e^{0.667t}$.

Постоянный темп роста r = 0,667. Это означает, что расходы на продукты питания в течение выборочного периода росли с темпом 66,7 % в год. ▲

Упражнение 10.3. Директор по маркетингу определил, что общий объем продаж товара зависит от его цены (табл. 10.11).

Таблица 10.11

Исходные данные для упражнения 10.3, усл. ед.

Цена товара (x)	10	20	30	40	50	60
Спрос на товар (y)	700	350	300	200	110	50

Требуется установить степенную зависимость между ценой и спросом, предсказать величину спроса для цены 45 усл. ед. и определить, на сколько процентов уменьшится спрос при росте цены на 1 %.

Ответ: функция спроса $y = 17662x^{-1,307}$. Прогнозное значение спроса на товар при цене 45 усл. ед. равно 122,08 усл. ед. При увеличении цены товара на 1 % спрос уменьшится на 1,307 %.

Упражнение 10.4. По данным табл. 10.12 об объеме продаж компании за 10 лет постройте модель экспоненциального роста и сделайте прогноз объема продаж на два года вперед.

Таблица 10.12

Исходные данные для упражнения 10.4, усл. ед.

Год (t)	**Объем продаж (y)**	**Год (t)**	**Объем продаж (y)**
1	70	6	1979
2	183	7	4096
3	340	8	6440
4	649	9	8459
5	1243	10	12 154

Ответ: модель имеет вид $y = 58{,}55e^{0{,}569t}$. Прогноз $y(11) =$ $= 30\ 729$; $y(12) = 54\ 303$.

Прогнозирование в регрессионных моделях

Под **прогнозированием** в эконометрике понимается построение оценки зависимой переменной для некоторого набора независимых переменных, которых нет в исходных наблюдениях.

Различают *точечное* и *интервальное* прогнозирование. В первом случае оценка — некоторое число, во втором — интервал, в котором находится истинное значение зависимой переменной с заданным уровнем значимости.

Рассмотрим регрессионную модель

$$y = \alpha + \beta x + \varepsilon.$$

Действительное значение зависимой переменной при $x = x_p$

$$y_p = \alpha + \beta x_p + \varepsilon_p,$$

где значения α, β, ε_p неизвестны и $M(\varepsilon_p) = 0$, $D(\varepsilon_p) = \sigma^2$.

Предсказанным значением является оценка $\hat{y}_p$ (точечный прогноз):

$$\hat{y}_p = a + bx_p.$$

Ошибка предсказания равна разности между предсказанным и действительным значениями:

$$\Delta_p = \hat{y}_p - y_p.$$

Ошибка предсказания имеет нулевое математическое ожидание

$$M(\Delta_p) = 0.$$

Действительно,

$$M(\Delta_p) = M(\hat{y}_p) - M(y_p) = M(a + bx_p) - M(\alpha + \beta x_p + \varepsilon_p) = 0.$$

Вычислим дисперсию прогноза. Учитывая, что в случае парной регрессии

$$\hat{y}_p = \bar{y} + b(x_p - \bar{x}), D(\bar{y}) = \frac{\sigma^2}{n}, D(y_p) = D(\varepsilon_p) = \sigma^2,$$

для дисперсии прогноза получим

$$D(\Delta_p) = \left[1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n\,\mathrm{var}(x)}\right]\sigma^2.$$

Из формулы следует, что чем больше x_p отклоняется от выборочного среднего $\bar{x}$, тем больше дисперсия ошибки предсказания, и чем больше объем выборки n, тем меньше дисперсия.

Заменяя в дисперсии прогноза σ^2 на ее оценку s^2 (остаточная дисперсия) и извлекая квадратный корень, получим *стандартную ошибку предсказания*

$$s_p = s\sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n\,\mathrm{var}(x)}}\ .$$

Доверительный интервал для действительного значения y_p определяется выражением

$$\hat{y}_p - t_{\text{кр}} s_p < y_p < \hat{y}_p + t_{\text{кр}} s_p\,,$$

где $t_{\text{кр}}$ — критическое значение t-статистики при заданном уровне значимости и числе $\nu = n - 2$ степени свободы.

Пример 10.7. По данным о зависимости объема продаж (y) фирмы от затрат (x) на рекламу оценим объем продаж при затратах на рекламу, равных 5,5 усл. ед. Найдем стандартную ошибку предсказания и 99%-й доверительный интервал для полученной оценки. Построим график зависимости предсказания объема продаж с доверительным интервалом от затрат на рекламу.

▼Исходные данные и расчетные показатели представим в *Excel* в виде расчетной таблицы (рис. 10.5).

	A	B	C	D	E	F
1	x	y	$\hat{y}_p$	s_p	y_{min}	y_{max}
2	1	68	64.84	4.79	49.97	79.71
3	5	65	68.53	4.56	54.36	82.70
4	9	77	72.22	4.42	58.51	85.94
5	10	70	73.14	4.39	59.51	86.78
6	11	80	74.07	4.37	60.49	87.65
7	12	70	74.99	4.36	61.45	88.53
8	13	80	75.91	4.35	62.39	89.43
9	14	75	76.83	4.35	63.32	90.35
10	15	74	77.76	4.35	64.23	91.28
11	16	80	78.68	4.36	65.13	92.23
12	18	75	80.52	4.40	66.86	94.18
13	26	88	87.90	4.74	73.17	102.64
14	30	95	91.59	5.02	75.99	107.20
15	5.5		68.99	4.54	54.89	83.09
16	$\bar{x}$	13.85				
17	var(x)	55.82				
18	s	4,19				
19	$t_{кр}$	3,11				

Рис. 10.5. Расчетная таблица для примера 10.7 в *Excel*

На рис. 10.5 (x, y) — исходные данные, расположенные в порядке возрастания x; $\hat{y}_p, s_p$ — предсказание и его стандартная ошибка; $y_{\min} = \hat{y}_p - t_{кр}s_p, y_{\max} = \hat{y}_p + t_{кр}s_p$ — нижняя и верхняя границы доверительного интервала предсказания.

При вычислении расчетных показателей использовались функции *Excel* и формула ошибки предсказания s_p. Поясним ввод формул в соответствующие ячейки таблицы (табл. 10.13).

При затратах на рекламу x_p = 5,5 объем продаж фирмы составляет $\hat{y}_p$ = 68,99 с доверительным интервалом (54,89; 83,09).

По данным расчетной таблицы построим график зависимости предсказания продаж с доверительным интервалом от затрат на рекламу (рис. 10.6).

Таблица 10.13

Использование формул *Excel* для решения примера 10.7

Ячейка	Формула
B16	= СРЗНАЧ(A2:A14)
B17	= ДИСПР(A2:A14)
B18	= СТОШYX(B2:B14;A2:A14)
B19	= СТЬЮДРАСПОБР(0,01;11)
C2	= ПРЕДСКАЗ(A2;$B$2:$B$14;$A$2:$A$14)
D2	= $B$18*КОРЕНЬ(1+1/13+(A2-$B$16)^2/13/$B$17)
E2	= C2-$B$19*D2
F2	= C2+$B$19*D2
C3:F15	«Протягиваем» формулы из ячеек C2:F2

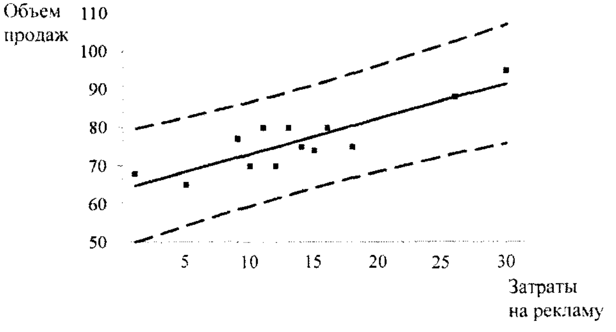


Рис. 10.6. График зависимости предсказания продаж с доверительным интервалом от затрат на рекламу для примера 10.7▲

10.5. Модель множественной регрессии

Обобщением линейной регрессионной модели с одной объясняющей переменной является линейная регрессионная модель с k объясняющими переменными (**модель множественной регрессии**):

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon,$$

где $\beta_0, \beta_1, \ldots, \beta_k$ — параметры модели;

$x_1, x_2, \ldots, x_k$ — объясняющие переменные;

ε — случайный член.

Случайный член ε удовлетворяет тем же предпосылкам, что и в модели с парной регрессией (условия Гаусса — Маркова), но на объясняющие переменные наложено условие: случайные члены в любом наблюдении должны быть статистически независимы от объясняющих переменных.

При выполнении условий Гаусса — Маркова модель называется *классической нормальной линейной регрессионной моделью*.

Кроме того, предполагается, что объясняющие переменные некоррелированы друг с другом.

На основе n наблюдений оценивается выборочное уравнение регрессии

$$\hat{y} = b_0 + b_1 x_1 + \ldots + b_k x_k,$$

где $b_0, b_1, \ldots, b_k$ — оценки параметров $\beta_0, \beta_1, \ldots, \beta_k$.

Для оценки параметров регрессии используется метод наименьших квадратов, в соответствии с которым минимизируется сумма квадратов остатков:

$$Q = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 \to \min .$$

Необходимым условием ее минимума является равенство нулю всех ее частных производных по $b_0, b_1, \ldots, b_k$.

В результате приходим к системе из $(k + 1)$ линейных уравнений с $(k + 1)$ неизвестными, называемой *системой нормальных уравнений.*

Ее решение в явном виде обычно записывается в матричной форме, иначе оно становится слишком громоздким.

Оценки параметров модели и их теоретические дисперсии в матричной форме определяются выражениями

$$b = (X^T X)^{-1} X^T Y, \; D(b_j) = (X^T X)_{jj}^{-1} \sigma^2,$$

где b — вектор с компонентами $b_0, b_1, \ldots, b_k$;

X — матрица значений объясняющих переменных;

Y — вектор значений зависимой переменной;

σ^2 — дисперсия случайного члена.

Несмещенной оценкой σ^2 является величина s^2 (остаточная дисперсия):

$$s^2 = \frac{\sum e_i^2}{n-k-1}.$$

Величина s называется стандартной ошибкой регрессии.

Заменяя в теоретических дисперсиях неизвестную дисперсию σ^2 на ее оценку s^2 и извлекая квадратный корень, получим *стандартные ошибки коэффициентов регрессии*:

$$s_{b_j} = s\sqrt{(X^T X)_{jj}^{-1}} .$$

Если предпосылки относительно случайного члена ε выполняются, оценки параметров множественной регрессии являются несмещенными, состоятельными и эффективными.

t-тесты для коэффициентов множественной регрессии и F-тест для проверки значимости R^2 выполняются так же, как и в парном регрессионном анализе с учетом числа степеней свободы.

Чаще всего F-тест используется для оценки того, значимо ли объяснение, даваемое уравнением *в целом*.

Проверка гипотезы H_0: $F = 0$ равнозначна проверке гипотезы

$$H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0$$

об одновременном равенстве нулю всех коэффициентов линейной регрессии, за исключением свободного члена.

Если объясняющие способности независимых переменных перекрываются (сильная корреляция между ними), то t-тест для каждой переменной окажется незначим, в то время как F-тест для уравнения в целом вполне может быть значим.

При использовании *Пакета анализа Excel* (инструмент *Регрессия*) коэффициенты множественной регрессии, их стандартные отклонения и другие показатели вычисляются автоматически и одновременно. При этом интерпретация получаемых показателей такая же, как в парной регрессии с учетом числа степеней свободы.

Пример 10.8. По данным бюджетного обследования семи случайно выбранных семей изучалась зависимость накопления (y) от дохода (x_1) и стоимости имущества (x_2) (табл. 10.14).

Таблица 10.14

Исходные данные для примера 10.8, усл. ед.

Накопление (y)	3	7	5	4	2	7	6
Доход (x_1)	40	55	45	30	30	60	50
Стоимость имущества (x_2)	60	40	40	15	90	30	30

Используя *Пакет анализа Excel* (инструмент *Регрессия*), проанализируем зависимость накопления от дохода и стоимости имущества.

▼Результаты расчетов с использованием инструмента *Регрессия* выводятся под общим названием *Вывод итогов* в виде таблиц (рис. 10.7).

Регрессионная статистика	
Множественный R	0,970
R-квадрат	0,942
Нормированный R-квадрат	0,913
Стандартная ошибка	0,577
Наблюдения	7

Дисперсионный анализ					
	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	2	21,527	10,763	32,355	0,0034
Остаток	4	1,331	0,333		
Итого	6	22,857			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-значение*	*Нижние 95 %*	*Верхние 95 %*
Y-пересечение	0,493	1,229	0,402	0,708	-2,917	3,904
Переменная X_1	0,128	0,022	5,866	0,004	0,067	0,188
Переменная X_2	-0,030	0,010	-2,860	0,046	-0,058	-0,001

Рис. 10.7. Результат работы инструмента *Регрессия Excel* для примера 10.8

Коэффициенты регрессии $a = 0{,}493$; $b_1 = 0{,}128$; $b_2 = -0{,}030$.

Запишем оцененное уравнение регрессии в виде

$$\hat{y} = \underset{}{0{,}493} + \underset{(0{,}022)}{0{,}128x_1} - \underset{(0{,}010)}{0{,}030x_2}, R^2 = 0{,}942,$$

где в скобках указаны стандартные ошибки.

Качество модели оценивается коэффициентом детерминации R^2.

Величина $R^2 = 0{,}942$ означает, что факторами дохода и стоимости имущества можно объяснить 94,2 % вариации (разброса) накопления.

Установим статистическую значимость коэффициента R^2.

Поскольку значимость $F = 0{,}0034 < \alpha = 0{,}05$, то $R^2 = 0{,}942$ значим при 5%-м уровне, т. е. модель в целом адекватна исходным данным.

Коэффициент $b_1 = 0{,}128 > 0$ показывает, что при увеличении дохода на 1 усл. ед. накопление увеличивается на 0,128 усл. ед.

Коэффициент $b_2 = -0{,}030 < 0$ показывает, что при увеличении стоимости имущества на 1 усл. ед. накопление уменьшается на 0,030 усл. ед.

Оценим статистическую значимость коэффициентов b_1, b_2.

Поскольку P-значение $(b_1) = 0{,}004 < \alpha = 0{,}05$, то $b_1 = 0{,}128$ значим на уровне 5 %. Поскольку P-значение $(b_2) = 0{,}046 < \alpha = 0{,}05$, то $b_2 = -0{,}030$ также значим.

Результаты оценивания регрессии совместимы не только с полученными значениями коэффициентов регрессии, но и с некоторым их множеством (доверительным интервалом). С вероятностью 95 % доверительный интервал коэффициента b_1 есть (0,067; 0,188), а для b_2 есть (–0,058; –0,001).

Из полученного уравнения регрессии можно сделать следующие **выводы**:

• прогнозируемые накопления семьи с доходом 40 усл. ед. и имуществом стоимостью 25 усл. ед. составляют

$$\hat{y} = 0{,}493 + 0{,}128 \cdot 40 - 0{,}030 \cdot 25 = 4{,}863 \text{ усл. ед.};$$

• если доход семьи возрастет на 10 усл. ед., а стоимость имущества не изменится, то накопления возрастут на величину

$$\Delta y = 0{,}128\Delta x_1 = 0{,}128 \cdot 10 = 1{,}28 \text{ усл. ед.};$$

• если доход семьи увеличится на 5 усл. ед., а стоимость имущества — на 15 усл. ед., то накопления возрастут на величину

$$\Delta y = 0{,}128\Delta x_1 - 0{,}030\Delta x_2 = 0{,}128 \cdot 5 - 0{,}030 \cdot 15 = 0{,}19 \text{ усл. ед.} \blacktriangle$$

Мультиколлинеарность

Мультиколлинеарность — это коррелированность двух или нескольких объясняющих переменных в уравнении регрессии. При наличии мультиколлинеарности МНК-оценки формально существуют, но обладают рядом недостатков:

• небольшое изменение исходных данных приводит к существенному изменению оценок коэффициентов регрессии;

• оценки коэффициентов регрессии имеют большие стандартные ошибки, малую значимость, в то время как модель в целом является значимой (высокое значение R^2).

Если при оценке уравнения регрессии несколько факторов оказались незначимы, то нужно выяснить, нет ли среди них сильно коррелированных между собой.

Для отбора факторов в модель регрессии и оценки их мультиколлинеарности анализируют *корреляционную матрицу*. Общий вид корреляционной матрицы, составленной из переменных y, x_1, x_2, x_3 приведен в табл. 10.15.

Таблица 10.15

Общий вид корреляционной матрицы

Переменная	y	x_1	x_2	x_3
y	1			
x_1	$r(y,x_1)$	1		
x_2	$r(y,x_2)$	$r(x_1,x_2)$	1	
x_3	$r(y,x_3)$	$r(x_1,x_3)$	$r(x_2,x_3)$	1

При наличии корреляции один из пары связанных между собой факторов исключается. Если статистически незначим лишь один фактор, то он должен быть исключен.

В модель регрессии включаются те факторы, которые более сильно связаны с зависимой переменной, но слабо связаны с другими факторами.

Корреляционную матрицу можно получить, используя *Пакет анализа Excel* (инструмент *Корреляция*).

Пример 10.9. По данным бюджетного обследования семи случайно выбранных семей изучалась зависимость накопления (y) от дохода (x_1), стоимости имущества (x_2) и расходов на питание (x_3) (табл. 10.16).

Таблица 10.16

Исходные данные для примера 10.9, усл. ед.

Накопление (y)	3	7	5	4	2	7	6
Доход (x_1)	40	55	45	30	30	60	50
Стоимость имущества (x_2)	60	40	40	15	90	30	30
Расходы на питание (x_3)	10	15	12	8	10	20	15

Используя *Пакет анализа Excel* (инструмент *Корреляция*), проанализируем целесообразность включения в модель каждого фактора.

▼Выборочное уравнение регрессии есть

$$\hat{y} = \underset{}{0{,}590} + \underset{(0{,}070)}{0{,}116x_1} - \underset{(0{,}012)}{0{,}030x_2} + \underset{(0{,}192)}{0{,}035x_3},\ R^2 = 0{,}942,\ \overline{R^2} = 0{,}885,$$

где коэффициенты при факторных переменных незначимы.

Корреляционная матрица из этих переменных представлена в табл. 10.17.

Таблица 10.17

Корреляционная матрица для примера 10.9

Переменная	y	x_1	x_2	x_3
y	1			
x_1	0,907	1		
x_2	–0,664	–0,380	1	
x_3	0,830	0,936	–0,283	1

Из рассмотрения корреляционной матрицы заключаем, что факторы x_1, x_3 дублируют друг друга, $r(x_1, x_3) = 0{,}936$. В анализ целесообразно включить фактор x_1, а не x_3, так как его связь с результативным признаком y сильнее, чем у фактора x_3, $r(y, x_1) = 0{,}907 > r(y, x_3) = 0{,}830$.

Оценив регрессию y на x_1, x_2, получим

$$\hat{y} = \underset{}{0{,}493} + \underset{(0{,}022)}{0{,}128x_1} - \underset{(0{,}010)}{0{,}030x_2},\ R^2 = 0{,}942,\ \overline{R^2} = 0{,}913,$$

где коэффициенты при факторных переменных значимы.

Заметим, что первоначальное скорректированное значение $\overline{R^2} = 0{,}885$ при отбраковке дублирующего фактора x_3 увеличилось до $\overline{R^2} = 0{,}913$. ▲

Спецификация модели

Построение экономической модели включает **выбор объясняющих переменных**. Свойства оценок коэффициентов регрессии в значительной мере зависят от правильной спецификации модели.

Рассмотрим два случая:

• отсутствие в модели переменной, которая должна быть включена;

• наличие в модели переменной, которая не должна быть включена.

1. Влияние отсутствия в модели переменной, которая должна быть включена. Предположим, что переменная y имеет следующую зависимость от двух переменных x_1 и x_2: $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$. Однако считается, что модель выглядит как $y = \alpha + \beta_1 x_1 + \varepsilon$, и оценивается регрессия $\hat{y} = a + b_1 x_1$. В этом случае оценка b_1 является смещенной.

Смещенность оценки b_1 связана с тем, что если не учесть x_2 в регрессии, то переменная x_1 будет играть двойную роль: отражать свое прямое влияние и заменять переменную x_2 в описании ее влияния.

Коэффициент R^2 для данной регрессии отражает общую объясняющую способность переменной x_1 в обеих ролях и является завышенной оценкой.

2. *Влияние включения в модель переменной, которая не должна быть включена.* Допустим, что истинная модель есть $y = \alpha + \beta_1 x_1 + \varepsilon$. Однако считается, что ею является $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ и оценивается регрессия $\hat{y} = a + b_1 x_1 + b_2 x_2$.

Оценки коэффициентов регрессии в этом случае являются несмещенными, но неэффективными. Практически обнаруживается, что коэффициент b_2 статистически незначим, и переменная x_2 исключается из модели.

Замещающие переменные. Предположим, что истинной *моделью является $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$, и допустим,* что не имеется данных по существенной переменной x_1.

Если не включить в модель эту переменную, то регрессия может пострадать от смещения оценок и статистическая проверка будет некорректной.

Если вместо отсутствующей переменной x_1 использовать ее заменитель z, линейно связанный с x_1, и построить регрессию

$$\hat{y} = a + b_2 x_2 + \ldots + b_k x_k + cz,$$

то оценки $b_2, \ldots, b_k$, их стандартные ошибки и коэффициент R^2 будут такими же, как с использованием x_1. Единственным недостатком является то, что отсутствует оценка коэффициента при самой величине x_1, а величина a не является оценкой α.

В качестве замещающей переменной, например, для показателя технического прогресса может использоваться время.

Лаговые переменные. При использовании данных временного ряда на текущее значение зависимой переменной могут влиять не только текущие значения объясняющих переменных, но также их значения с некоторым запаздыванием.

В общем случае, если какая-то переменная появляется в модели с запаздыванием на τ периодов, она записывается с нижним индексом $(t - \tau)$, например:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-\tau} + \varepsilon_t.$$

Сдвиг τ, характеризующий запаздывание в воздействии фактора на результат, называется *лагом*. Переменная, влияние которой характеризуется некоторым запаздыванием, — *лаговой*.

Инструментальные переменные. При наличии корреляции между объясняющими переменными и случайным членом МНК-оценки являются смещенными и несостоятельными.

Для получения состоятельных оценок используется *метод инструментальных переменных* (ИП). Сущность метода ИП заключается в замене непригодной объясняющей переменной такой переменной, которая некоррелирована со случайным членом и коррелирована с исходной переменной.

Пусть в модели $y = \alpha + \beta x + \varepsilon$ переменная x коррелирована со случайным членом ε. Предположим, что можно найти другую переменную z, которая коррелирована с x, но не коррелирована с ε.

Основанная на использовании ИП оценка параметра β, определяемая как

$$b_{\text{ИП}} = \frac{\operatorname{cov}(z, y)}{\operatorname{cov}(z, x)},$$

является состоятельной.

В больших выборках ошибка исчезает при условии, что переменная z действительно распределена независимо от ε и $\operatorname{cov}(z,\ x) \neq 0$. Следовательно, на больших выборках $b_{\text{ИП}}$ будет стремиться к истинному значению β.

Таким образом, оценка $b_{\text{ИП}}$ является состоятельной, но в общем случае смещенной и неэффективной.

Фиктивные переменные. При исследовании влияния *качественных признаков* в модель можно вводить фиктивные переменные, принимающие, как правило, два значения: единица, если данный признак присутствует в наблюдении, и нуль при его отсутствии.

Если включаемый в рассмотрение качественный признак имеет не два, а несколько значений, то используют несколько фиктивных переменных, число которых должно быть на единицу меньше числа значений признака.

При назначении фиктивных переменных исследуемая совокупность по числу значений качественного признака разбивается на группы. Одну из групп выбирают как эталонную (группа 0) и определяют фиктивные переменные для остальных.

Например, если качественный признак имеет три значения, то две фиктивные переменные определяются следующим образом:

• группа 0: $z_1 = z_2 = 0$;
• группа 1: $z_1 = 1, z_2 = 0$;
• группа 2: $z_1 = 0, z_2 = 1$,

или

$$z_1 = \begin{cases} 1, & \text{(группа 1)} \\ 0; & \text{(остальные)} \end{cases} \qquad z_2 = \begin{cases} 1, & \text{(группа 2)} \\ 0. & \text{(остальные)} \end{cases}$$

Введение в регрессию фиктивных переменных существенно улучшает качество ее оценивания.

Пример 10.10. Построим линейную регрессионную модель зависимости заработной платы (y) 16 работников фирмы от возраста (x) с использованием фиктивной переменной (z) по фактору «пол», взяв женский пол в качестве эталонной группы.

▼Исходные данные представим в виде табл. 10.18.

Таблица 10.18

Исходные данные для примера 10.10

№	Пол	Заработная плата, усл. ед. (y)	Возраст, лет (x)	z
1	Мужской	338	35	1
2	Женский	329	55	0
3	Женский	299	32	0
4	Мужской	334	44	1
5	Женский	305	33	0
6	Женский	304	58	0
7	Мужской	336	38	1
8	Женский	310	50	0
9	Мужской	340	45	1
10	Женский	300	47	0
11	Мужской	323	38	1
12	Мужской	321	35	1
13	Женский	293	35	0
14	Женский	313	48	0
15	Мужской	340	58	1
16	Мужской	340	54	1

Оценив регрессию между y и x, получим

$$\hat{y} = 299{,}19 + 0{,}48x, \; R^2 = 0{,}067.$$

Уравнение регрессии статистически незначимо.

Для учета качественного фактора «пол» введем в модель фиктивную переменную:

$$z = \begin{cases} 0, & \text{(Женский)} \\ 1. & \text{(Мужской)} \end{cases}$$

Оценив регрессию между y и x, z, получим

$$\hat{y} = 279{,}64 + 0{,}60x + 28{,}2z, \quad R^2 = 0{,}809.$$

Уравнение регрессии статистически значимо.

Из полученного уравнения регрессии следует, что при одном и том же возрасте заработная плата у мужчин на 28,2 усл. ед. выше, чем у женщин.

Коэффициент 28,2 при фиктивной переменной z статистически значим, следовательно, имеем две различные модели:

$$\hat{y} = 279{,}64 + 0{,}60x \text{ — для женщин, } (z = 0);$$

$$\hat{y} = 307{,}84 + 0{,}60x \text{ — для мужчин, } (z = 1). \;\blacktriangle$$

10.6. Гетероскедастичность и автокоррелированность случайного члена

Обнаружение гетероскедастичности

Одной из предпосылок регрессионного анализа является предположение о постоянстве дисперсии случайного члена для всех наблюдений (*гомоскедастичность*). Это значит, что для каждого значения объясняющей переменной случайные члены имеют одинаковые дисперсии. Если это условие не соблюдается, то имеет место *гетероскедастичность*.

При отсутствии гетероскедастичности коэффициенты регрессии имеют наименьшую дисперсию среди всех несмещенных оценок, являющихся линейными функциями от наблюдений y.

Если наблюдается гетероскедастичность, то МНК-оценки будут неэффективными (они не будут иметь наименьшую дисперсию по сравнению с другими оценками данного параметра).

Оценки стандартных ошибок коэффициентов регрессии вычисляются в предположении, что распределение случайного члена гомоскедастично; если это не так, то они неверны (занижены), а, следовательно, t-статистика завышена. Это может привести к статистически значимым коэффициентам регрессии, тогда как в действительности это неверно.

Проблема гетероскедастичности характерна для пространственных данных, полученных от неоднородных объектов. Например, если исследуется зависимость прибыли предприятий от размера основного фонда, то можно ожидать, что для больших предприятий колебание прибыли будет выше, чем для малых.

Рассмотрим ряд тестов для обнаружения гетероскедастичности.

I. Тест ранговой корреляции Спирмена. Выдвигается нулевая гипотеза об отсутствии гетероскедастичности случайного члена.

Предполагается, что дисперсия случайного члена будет либо увеличиваться, либо уменьшаться по мере увеличения x, и поэтому в регрессии, оцениваемой с помощью МНК, абсолютные величины остатков $|e|$ и значения x будут коррелированы.

Данные по x и остатки $|e|$ ранжируются по переменной x, и определяются их *ранги*. Ранг — это порядковый номер значения переменной в ранжированном ряду.

Определяется *коэффициент ранговой корреляции Спирмена* по формуле

$$r = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2 - 1)},$$

где D_i — разность между рангами x и $|e|$.

Коэффициент Спирмена может быть использован для определения тесноты связи как между количественными, так и между качественными признаками при условии, что значения этих признаков могут быть упорядочены по убыванию или возрастанию.

Значимость r проверяется по t-тесту. В случае установления значимости коэффициента ранговой корреляции Спирмена делается вывод о наличии гетероскедастичности.

Пример 10.11. Имеются данные о зависимости выпуска продукции обрабатывающей промышленности на душу населения (y) от валового внутреннего продукта на душу населения (x) в том же году для 17 стран (табл. 10.19).

Таблица 10.19

Исходные данные для примера 10.11, усл. ед. на душу населения

№	Выпуск продукции обрабатывающей промышленности (y)	ВВП (x)	№	Выпуск продукции обрабатывающей промышленности (y)	ВВП (x)
1	18	3	10	100	24
2	27	6	11	63	25
3	18	7	12	130	26
4	45	9	13	135	27
5	55	13	14	60	28
6	68	15	15	70	35
7	51	18	16	80	37
8	84	21	17	180	44
9	85	22			

Построим регрессионную модель зависимости выпуска продукции обрабатывающей промышленности от валового внутреннего продукта. Проверим наличие гетероскедастичности, используя тест Спирмена.

▼Исходные данные и расчетные показатели представим в виде табл. 10.20.

Пусть модель описывается выражением $y = \alpha + \beta x + \varepsilon$.

Оцененное уравнение регрессии обычным МНК:

$$\hat{y} = 12{,}84 + 2{,}92\,x,\ \alpha = 12{,}84;\ \beta = 2{,}92.$$

Остатки e вносим в соответствующий столбец табл. 10.20. Далее последовательно заполняются столбцы $|e|$, ранг x, ранг $|e|$, D, D^2.

Таблица 10.20

Расчетная таблица для примера 10.11

y	x	e	$\|e\|$	Ранг x	Ранг $\|e\|$	D	D^2
18	3	–3,60	3,60	1	2	–1	1
27	6	–3,35	3,35	2	1	1	1
18	7	–15,27	15,27	3	9	–6	36
45	9	5,89	5,89	4	4	0	0
55	13	4,22	4,22	5	3	2	4
68	15	11,38	11,38	6	7	–1	1
51	18	–14,38	14,38	7	8	–1	1
84	21	9,87	9,87	8	6	2	4
85	22	7,95	7,95	9	5	4	16
100	24	17,11	17,11	10	10	0	0
63	25	–22,81	22,81	11	11	0	0
130	26	41,28	41,28	12	15	–3	9
135	27	43,36	43,36	13	16	–3	9
60	28	–34,56	34,56	14	12	2	4
70	35	–44,99	44,99	15	17	–2	4
80	37	–40,83	40,83	16	14	2	4
180	44	38,74	38,74	17	13	4	16
						Итого	110

Коэффициент ранговой корреляции:

$$r = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)} = 1 - \frac{6 \cdot 110}{17 \cdot 288} = 0,865.$$

Для проверки нулевой гипотезы используем t-тест. Расчетное значение критерия

$$t_p = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,865\sqrt{17-2}}{\sqrt{1-0,865^2}} = 6,68.$$

Установим $\alpha = 0{,}05$. Число степени свободы $\nu = n - 2 = 15$. Критическую точку находим с помощью функции *Excel*:

$$t_{кр} = \text{СТЬЮДРАСПОБР}(0{,}05;\ 15) = 2{,}13.$$

Поскольку $|t_р| = 6{,}68 > t_{кр} = 2{,}13$, гипотеза H_0 отклоняется, следовательно, имеет место гетероскедастичность. ▲

II. Тест Глейзера. Тест Глейзера основывается на более общих представлениях о зависимости стандартной ошибки случайного члена от значений объясняющей переменной. Например, зависимость может быть представлена в виде $\sigma_i = \alpha + \beta x_i + \varepsilon_i$.

Используя абсолютные значения остатков в качестве оценки σ_i, осуществляют оценивание данной регрессионной зависимости.

Таким образом, гетероскедастичность аппроксимируется уравнением:

$$s_i = a + bx_i,$$

где $s_i = |e_i|$ — оценка σ_i.

Нулевая гипотеза об отсутствии гетероскедастичности отклоняется, если оценка b значимо отличается от нуля.

Пример 10.12. По данным примера 10.11 установим наличие гетероскедастичности, используя тест Глейзера.

▼ Оценив регрессию между $|e|$ и x, получим оцененное уравнение

$$s = -3{,}13 + \underset{(0{,}19)}{1{,}15}x,$$

где коэффициент $b = 1{,}15$ значим, следовательно, имеет место гетероскедастичность. ▲

III. Тест Голдфелда — Квандта. При проведении проверки по этому тесту предполагается, что стандартное отклонение σ случайного члена пропорционально значению независимой переменной x.

Тест включает следующие шаги:

- все n наблюдений в выборке упорядочиваются по возрастанию переменной x;
- оцениваются отдельные регрессии для первых n_0 и для последних n_0 наблюдений, а средние $(n - 2n_0)$ наблюдений отбрасываются;

• составляется статистика: $F = RSS_2/RSS_1$, где RSS_2, RSS_1 — суммы квадратов остатков для первых и последних n_0 наблюдений соответственно.

Если верна гипотеза H_0 об отсутствии гетероскедастичности, то F имеет распределение Фишера с $\nu_1 = n_0 - k - 1$, $\nu_2 = n_0 - k - 1$ степенями свободы, где k — число объясняющих переменных.

Определяется критическое значение критерия $F_{кр}$.

Если $F > F_{кр}$, то нулевая гипотеза об отсутствии гетероскедастичности отклоняется.

Замечание. Тест Голдфелда — Квандта можно также использовать для проверки на гетероскедастичность при предположении, что σ_i обратно пропорционально x_i. В этом случае тестовой статистикой является величина $F = RSS_1/RSS_2$.

Пример 10.13. По данным примера 10.11 установим наличие гетероскедастичности, используя тест Голдфелда — Квандта.

▼ С помощью обычного МНК оценим регрессии для шести стран с наименьшими значениями показателя x и для шести стран с наибольшими значениями этого показателя.

Получим соответственно уравнения

$$\hat{y}_1 = -0{,}18 + 4{,}38x, RSS_1 = 224{,}4;$$

$$\hat{y}_2 = 39{,}9 + 2{,}11x, RSS_2 = 9804{,}5.$$

Суммы квадратов отклонений составляют $RSS_1 = 224{,}4$; $RSS_2 = 9804{,}5$, при этом $F = 9804{,}5/224{,}4 = 43{,}7$. Число степеней свободы $\nu_1 = \nu_2 = 4$. Критическое значение $F_{кр} = 6{,}39$ при 5%-м уровне значимости.

Поскольку $F = 43{,}7 > F_{кр} = 6{,}39$, то нулевая гипотеза об отсутствии гетероскедастичности отклоняется. ▲

Обобщенный метод наименьших квадратов

При наличии гетероскедастичности используют **обобщенный (взвешенный) МНК** (ОМНК). Суть метода заключается в уменьшении вклада данных наблюдений, имеющих большую дисперсию, в результат расчета.

Пусть в исходной модели регрессии

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

случайные члены гетероскедастичны, т. е.

$$M(\varepsilon_i) = 0,\ D(\varepsilon_i) = \sigma_i^2,\ (i = \overline{1,n}).$$

Допустим, что дисперсии σ_i^2 в каждом наблюдении известны. Разделив каждое наблюдение на соответствующее ему значение σ_i, получим преобразованную модель

$$\frac{y_i}{\sigma_i} = \alpha \cdot \frac{1}{\sigma_i} + \beta \cdot \frac{x_i}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i},$$

для которой выполнены условия гомоскедастичности, так как

$$M\left(\frac{\varepsilon_i}{\sigma_i}\right) = \frac{M(\varepsilon_i)}{\sigma_i} = 0,\ D\left(\frac{\varepsilon_i}{\sigma_i}\right) = \frac{D(\varepsilon_i)}{\sigma_i^2} = \frac{\sigma_i^2}{\sigma_i^2} = 1, (i = \overline{1,n}).$$

Оценки параметров преобразованной модели обычным МНК дают непосредственно оценки параметров исходной модели.

На практике дисперсии σ_i^2 неизвестны, поэтому их заменяют оценками s_i^2, т. е. оценивается обычным МНК преобразованная модель

$$\frac{y}{s} = \alpha \cdot \frac{1}{s} + \beta \cdot \frac{x}{s} + \frac{\varepsilon}{s},$$

в которой отсутствует постоянный член.

Для экономических данных σ_i часто пропорциональны значениям объясняющей переменной x. При этом обычным МНК оценивается преобразованная модель

$$\frac{y}{x} = \alpha \cdot \frac{1}{x} + \beta + \frac{\varepsilon}{x}.$$

Коэффициент при $1/x$ будет *эффективной оценкой* α, а постоянный член — *эффективной оценкой* β.

При применении взвешенного МНК оценки параметров будут несмещенными, кроме того, они имеют меньшую дисперсию, чем невзвешенные оценки.

Пример 10.14. В примере 10.11 обычным МНК была получена регрессионная зависимость

$$\tilde{y} = 12{,}84 + 2{,}92\,x\,,\ \alpha = 12{,}84;\ \beta = 2{,}92.$$

По этим же данным оценим регрессионную зависимость взвешенным МНК в двух вариантах:

a) с делением на s, где $s = -3{,}13 + 1{,}15x$;

b) делением на x.

▼Исходные и преобразованные данные представлены в табл. 10.21.

Таблица 10.21

Расчетная таблица для примера 10.14

y	x	s	y/s	y/x	$1/s$	x/s	$1/x$
18	3	0,31	58,75	6,00	3,264	9,791	0,333
27	6	3,75	7,21	4,50	0,267	1,601	0,167
18	7	4,89	3,68	2,57	0,204	1,430	0,143
45	9	7,19	6,26	5,00	0,139	1,252	0,111
55	13	11,78	4,67	4,23	0,085	1,104	0,077
68	15	14,07	4,83	4,53	0,071	1,066	0,067
51	18	17,51	2,91	2,83	0,057	1,028	0,056
84	21	20,95	4,01	4,00	0,048	1,002	0,048
85	22	22,10	3,85	3,86	0,045	0,996	0,045
100	24	24,39	4,10	4,17	0,041	0,984	0,042
63	25	25,54	2,47	2,52	0,039	0,979	0,040
130	26	26,68	4,87	5,00	0,037	0,974	0,038
135	27	27,83	4,85	5,00	0,036	0,970	0,037
60	28	28,98	2,07	2,14	0,035	0,966	0,036
70	35	37,01	1,89	2,00	0,027	0,946	0,029
80	37	39,30	2,04	2,16	0,025	0,941	0,027
180	44	47,33	3,80	4,09	0,021	0,930	0,023

Регрессионные зависимости, полученные взвешенным МНК в двух вариантах, следующие:

$$a)\ \frac{\tilde{y}}{s} = 8,64 \cdot \frac{1}{s} + 3,11 \cdot \frac{x}{s};\ \alpha = 8,64;\ \beta = 3,11;$$

$$b)\ \frac{\tilde{y}}{x} = 7,78 \cdot \frac{1}{x} + 3,20;\ \alpha = 7,78;\ \beta = 3,20,$$

где вариант a — с делением на s, а вариант b — с делением на x.

Полученные регрессионные зависимости не обнаруживают гетероскедастичности и дают более эффективные оценки коэффициентов регрессии, чем с использованием обычного МНК. ▲

Обнаружение автокорреляции

Одной из предпосылок регрессионного анализа является независимость случайного члена в любом наблюдении от его значений во всех других наблюдениях, т. е. $M(\varepsilon_i\varepsilon_j) = 0$, $(i \neq j)$.

Если данное условие не выполняется, то говорят, что случайный член подвержен автокорреляции. В этом случае коэффициенты регрессии, получаемые по МНК, оказываются неэффективными, хотя и несмещенными, а их стандартные ошибки рассчитываются некорректно (занижаются).

Причиной автокорреляции может быть либо неверная спецификация модели, либо наличие неучтенных факторов. Устранение этих причин не всегда дает нужные результаты. Автокорреляция имеет собственные внутренние причины, связанные с автокорреляционной зависимостью.

Автокорреляция обычно встречается только в регрессионном анализе при использовании данных временных рядов. В силу этого в дальнейшем вместо символа i (порядковый номер наблюдения) будем использовать символ t (момента наблюдения). Таким образом, базовая модель будет записана в виде

$$y_t = \alpha + \beta x_t + \varepsilon_t, \qquad (10.1)$$

где ε_t — случайный член.

Пусть автокорреляция подчиняется простой авторегрессионной схеме первого порядка:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t,$$

где ρ — *коэффициент авторегрессии*, а u_t — новый случайный член, удовлетворяющий предпосылкам МНК.

Это означает, что величина случайного члена ε_t в любом наблюдении равна его значению ε_{t-1} в предшествующем наблюдении, умноженному на ρ, плюс новый u_t.

Данная схема оказывается авторегрессионной, поскольку ε определяется значением этой же самой величины с запаздыванием, и схемой первого порядка, потому что в этом простом случае максимальное запаздывание равно единице.

Если $\rho > 0$, то автокорреляция *положительная*; если $\rho < 0$, то автокорреляция *отрицательная*; если $\rho = 0$, то автокорреляции нет и третье условие Гаусса — Маркова удовлетворяется.

Коэффициент авторегрессии ρ есть в точности коэффициент корреляции между двумя соседними ошибками ε_t, ε_{t-1}.

Положительная автокорреляция остатков имеет место при анализе экономических данных.

Поскольку значения случайных членов ε_t, ε_{t-1} неизвестны, то проверяется статистическая некоррелированность остатков e_t, e_{t-1} с использованием обычного МНК.

Соответствующей оценкой коэффициента авторегрессии ρ является *коэффициент автокорреляции остатков первого порядка*, который при достаточно большом числе наблюдений имеет вид

$$r = \frac{\operatorname{cov}(e_t, e_{t-1})}{\sqrt{\operatorname{var}(e_t)\operatorname{var}(e_{t-1})}} \approx \frac{\sum_{t=2}^{T} e_t\, e_{t-1}}{\sum_{t=1}^{T} e_t^2}.$$

Выдвигается нулевая гипотеза об отсутствии корреляции первого порядка H_0: $\rho = 0$. В качестве альтернативной гипотезы может выступать либо H_1: $\rho > 0$, либо H_2: $\rho < 0$.

Для проверки нулевой гипотезы используют *статистику Дарбина — Уотсона*, рассчитываемую по формуле

$$DW = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \simeq 2(1-r), \quad 0 \le DW \le 4.$$

По таблице определяются критические значения критерия Дарбина — Уотсона d_1 и d_2 для заданного числа наблюдений, числа объясняющих переменных и уровня значимости. По этим значениям отрезок [0; 4] разбивается на пять зон (рис. 10.8).

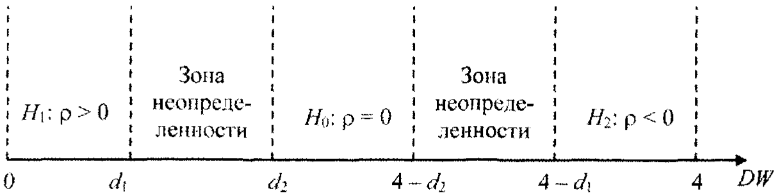


Рис. 10.8. Иллюстрация использования теста Дарбина — Уотсона

В зависимости от того, в какую зону попадает расчетное значение критерия DW, принимают или отвергают соответствующую гипотезу.

Замечание. Тест Дарбина — Уотсона построен в предположении, что объясняющие переменные некоррелированы со случайным членом. Поэтому этот тест неприменим к моделям, включающим в качестве объясняющих переменных лаговые значения зависимой переменной y.

Пример 10.15. Имеются данные об объеме предложения (y) товара, его цене (x_1) и зарплатах сотрудников (x_2) за 10 месяцев. Выявим на уровне значимости 0,05 наличие автокорреляции остатков в модели регрессии

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon.$$

▼Исходные данные и результаты промежуточных расчетов представлены в табл. 10.22.

Выборочная регрессия для этой модели

$$\hat{y} = 90,72 + 0,88x_1 - 7,32x_2,$$

где все коэффициенты уравнения статистически значимы.

Оценивая корреляцию между остатками e_t, e_{t-1} (функция =КОРРЕЛ), получим коэффициент автокорреляции остатков первого порядка $r = -0,0251$. Значение критерия $DW = 2(1 - r) = 2,05$.

По таблице распределения Дарбина — Уотсона (приложение 1) находим $d_1 = 0,70$ и $d_2 = 1,64$. Поскольку $d_2 < DW < 4 - d_2$, то нет оснований отклонять гипотезу H_0 об отсутствии автокорреляции в остатках.

Таблица 10.22

Расчетная таблица для примера 10.15

t	y	x_1	x_2	e_t	t_{t-1}
1	20	10	12	8,30	
2	35	15	10	4,26	8,30
3	30	20	9	–12,46	4,26
4	45	25	9	–1,86	–12,46
5	60	40	8	–7,38	–1,86
6	70	37	8	5,26	–7,38
7	75	43	6	–9,66	5,26
8	90	35	4	–2,26	–9,66
9	105	40	4	8,34	–2,26
10	110	55	5	7,46	8,34

▲

Авторегрессионное преобразование

Предположим, что истинная модель задается выражением (10.1), так что наблюдения t и $t-1$ формируются как

$$y_t = \alpha + \beta x_t + \varepsilon_t\,;$$
$$y_{t-1} = \alpha + \beta x_{t-1} + \varepsilon_{t-1}\,.$$

Пусть ρ *известен*. Вычтем из обеих частей первого уравнения второе уравнение, умноженное на ρ:

$$y_t - \rho\, y_{t-1} = \alpha(1-\rho) + \beta(x_t - \rho\, x_{t-1}) + (\varepsilon_t - \rho\,\varepsilon_{t-1})\,.$$

Обозначим:

$$\begin{cases} y'_t = y_t - \rho y_{t-1}, \\ x'_t = x_t - \rho x_{t-1}, \\ \alpha' = \alpha(1-\rho). \end{cases}$$

Это преобразование переменных называется *авторегрессионным преобразованием Бокса — Дженкинса.*

Тогда преобразованное уравнение

$$y'_t = \alpha' + \beta x'_t + u_t\,,\ (t \ge 2)$$

не содержит автокорреляцию, и для оценки его параметров α', β используется обычный МНК.

Оценка коэффициента β из этой зависимости непосредственно используется и для исходного уравнения, а коэффициент α рассчитывается по формуле $\alpha = \alpha' / (1-\rho)$.

Первое наблюдение в преобразованном уравнении можно сохранить, если положить:

$$y_1' = \sqrt{1-\rho^2}\, y_1;$$

$$x_1' = \sqrt{1-\rho^2}\, x_1.$$

Обычно первое наблюдение опускают, поэтому преобразование исходной модели становится единообразным.

На практике величина ρ *неизвестна*, ее оценка получается одновременно с оценками α, β в результате следующей итеративной процедуры:

• применяя МНК к исходному уравнению регрессии, получают первоначальные оценки параметров α, β;

• вычисляют остатки e и в качестве оценки ρ используют коэффициент автокорреляции остатков первого порядка, т. е. полагают $\rho = r$;

• применяя МНК к преобразованному уравнению, получают новые оценки параметров α, β.

Процесс обычно заканчивается, когда очередное приближение ρ мало отличается от предыдущего.

Указанная процедура является *обобщенным МНК при наличии автокорреляции*.

Пример 10.16. Оценим зависимость расходов на конечное потребление (y) от дохода (x) по данным некоторой страны за 15 лет (табл. 10.23).

▼Исходные данные и расчетные показатели приведены в табл. 10.24.

Оцененное уравнение регрессии обычным МНК есть

$$\hat{y}_t = 10{,}58 + 0{,}81 x_t, \; R^2 = 0{,}985$$

и статистически значимо.

Установим наличие автокорреляции с помощью статистики Дарбина — Уотсона. Коэффициент автокорреляции остатков первого порядка составляет $r = 0{,}495$, следовательно, $DW \approx 2(1 - r) = 1{,}01$.

Таблица 10.23

Исходные данные для примера 10.16, усл. ед.

t	Расходы на конечное потребление (y_t)	Доход (x_t)
1	70	73
2	73	76
3	78	83
4	83	89
5	86	95
6	89	100
7	96	107
8	96	108
9	103	113
10	109	119
11	112	121
12	114	122
13	115	131
14	118	135
15	122	139

Таблица 10.24

Расчетная таблица для примера 10.16

t	y_t	x_t	e_t	e_{t-1}	y_t'	x_t'
1	70	73	0,273			
2	73	76	0,842	0,273	38,327	39,841
3	78	83	0,170	0,842	41,841	45,355
4	83	89	0,309	0,170	44,364	47,888
5	86	95	–1,553	0,309	44,888	50,916
6	89	100	–2,604	–1,553	46,402	52,944
7	96	107	–1,276	–2,604	51,916	57,467
8	96	108	–2,086	–1,276	48,448	55,000
9	103	113	0,863	–2,086	55,448	59,505
10	109	119	2,001	0,863	57,981	63,028
11	112	121	3,381	2,001	58,009	62,056
12	114	122	4,570	3,381	58,523	62,065
13	115	131	–1,722	4,570	58,533	70,570
14	118	135	–1,963	–1,722	61,037	70,112
15	122	139	–1,204	–1,963	63,551	72,131

Пусть исходная модель имеет вид $y_t = \alpha + \beta x_t + \varepsilon_t$.

По таблице распределения Дарбина — Уотсона находим значение $d_1 = 1{,}08$ и $d_2 = 1{,}36$. Поскольку $0 < DW < d_1$, то имеется положительная автокорреляция остатков.

Для устранения автокорреляции воспользуемся обобщенным МНК.

Применяя обычный МНК к преобразованным данным

$$y'_t = y_t - 0{,}495 \cdot y_{t-1};$$
$$x'_t = x_t - 0{,}495 \cdot x_{t-1},\ t \geq 2,$$

получим оценку преобразованного уравнения:

$$\widehat{y'_t} = \underset{}{6{,}26} + \underset{(0{,}06)}{0{,}79x'_t}, R^2 = 0{,}944,$$

которое статистически значимо.

Коэффициент автокорреляции остатков первого порядка $r = 0{,}152$, следовательно, $DW = 1{,}7$ (автокорреляция отсутствует).

Пересчитывая оценку $\alpha = \alpha'/(1 - r) = 6{,}26/(1 - 0{,}152) = 7{,}38$, получим следующую оценку исходной модели:

$$\widehat{y}_t = 7{,}38 + 0{,}79x_t.$$

Это уравнение отличается от полученного ранее уравнения, оцененного обычным МНК, но в данном уравнении коэффициенты регрессии являются несмещенными и эффективными. ▲

Автокорреляция с лаговой зависимой переменной

Пусть имеется модель, в которой зависимая переменная, взятая с некоторым лагом, используется в качестве одной из объясняющих переменных.

Например, модель имеет вид

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t.$$

Допустим, что случайный член ε_t подвержен воздействию автокорреляции первого порядка. В этом случае влияние автокорреляции сделает оценки по обычному МНК несостоятельными.

Для обнаружения автокорреляции в случае, когда уравнение регрессии включает лаговую зависимую переменную, используется **h-статистика Дарбина**, которая также вычисляется на основе остатков:

$$h \approx r\sqrt{\frac{n}{1-n\operatorname{var}(b)}},$$

где r — коэффициент автокорреляции остатков первого порядка;

n — число наблюдений в выборке;

$\operatorname{var}(b)$ — оцененная дисперсия коэффициента β при лаговой зависимой переменной.

Значение h можно вычислить на основе обычных результатов оценивания регрессии. Этот тест предназначен только для *проверки на наличие автокорреляции первого порядка.*

При больших выборках h распределена как $N(0;\ 1)$. При применении двустороннего критерия и большой выборке гипотеза об отсутствии автокорреляции может быть отклонена:

- если $|h| > 1{,}96$ при уровне значимости 5%;
- если $|h| > 2{,}58$ при уровне значимости 1%.

Тест Дарбина *неприменим*, если $n\operatorname{var}(b) \geq 1$.

Пример 10.17. Исходные данные из 31 наблюдения для модели

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t$$

получены методом Монте-Карло со значениями $\alpha = 13$, $\beta = 0{,}8$. Оценим модель с использованием МНК и авторегрессионного преобразования.

▼Исходные данные и расчетные показатели представим в табл. 10.25.

Таблица 10.25

Расчетная таблица для примера 10.17

t	y_t	y_{t-1}	e_t	y'_t	y'_{t-1}
1	13,31			12,028	
2	20,53	13,31	–1,863	14,831	12,028
3	25,94	20,53	–2,587	17,150	14,831

Окончание табл. 10.25

t	y_t	y_{t-1}	e_t	y'_t	y'_{t-1}
4	41,45	25,94	8,326	30,344	17,150
5	48,49	41,45	2,189	30,743	30,344
6	53,17	48,49	0,888	32,409	30,743
7	58,38	53,17	2,122	35,615	32,409
8	61,09	58,38	0,406	36,095	35,615
9	62,11	61,09	–0,877	35,954	36,095
10	67,06	62,11	3,207	40,468	35,954
11	71,14	67,06	3,081	42,428	40,468
12	81,61	71,14	10,085	51,152	42,428
13	83,87	81,61	3,450	48,929	51,152
14	91,47	83,87	9,130	55,561	48,929
15	94,96	91,47	6,163	55,797	55,561
16	93,93	94,96	2,168	53,273	55,797
17	85,31	93,93	–5,577	45,094	53,273
18	78,75	85,31	–4,814	42,225	45,094
19	79,15	78,75	1,160	45,433	42,225
20	81,87	79,15	3,540	47,982	45,433
21	74,20	81,87	–6,441	39,148	47,982
22	72,89	74,20	–1,235	41,121	39,148
23	69,13	72,89	–3,882	37,922	41,121
24	66,71	69,13	–3,107	37,112	37,922
25	61,25	66,71	–6,511	32,688	37,112
26	51,14	61,25	–11,982	24,916	32,688
27	51,20	51,14	–3,333	29,305	24,916
28	60,72	51,20	6,136	38,799	29,305
29	60,31	60,72	–2,362	34,313	38,799
30	59,49	60,31	–2,834	33,668	34,313
31	56,98	59,49	–4,647	31,510	33,668

По исходным данным оцененное уравнение регрессии есть

$$\hat{y}_t = 11,085 + \underset{(0,047)}{0,850}\, y_{t-1},$$

где оценки α = 11,085; β = 0,850; var(b) = 0,047.

Коэффициент автокорреляции остатков первого порядка составляет $r = 0{,}428$. Значение h-критерия Дарбина:

$$h = 0,428\sqrt{\frac{30}{1 - 30 \cdot 0,047^2}} = 2,42 > 1,96,$$

что указывает на автокорреляцию.

Используя авторегрессионное преобразование (первое наблюдение в преобразованном уравнении сохранено), получим следующую оценку уравнения:

$$\widehat{y'_t} = \underset{}{7,925} + \underset{(0,079)}{0,806\, y_{t-1}},$$

где оценки $\alpha' = 7{,}925$; $\beta = 0{,}806$; $\mathrm{var}(b) = 0{,}079$.

Пересчитывая оценку $\alpha = \alpha'/(1 - r) = 7{,}925/(1 - 0{,}428) = 13{,}85$, получим следующую оценку исходной модели:

$$\hat{y}_t = 13{,}85 + 0{,}806 y_{t-1},\ \alpha = 13{,}85;\ \beta = 0{,}806.$$

Коэффициент автокорреляции остатков первого порядка составляет $r = 0{,}0357$. Значение h-критерия Дарбина:

$$h = 0,0357\sqrt{\frac{30}{1 - 30 \cdot 0,079^2}} = 0,22 < 1,96,$$

что указывает на отсутствие автокорреляции.

Ближе к истинным являются оценки $\alpha = 13{,}85$ и $\beta = 0{,}806$, полученные с учетом авторегрессии, чем оценки, полученные без ее учета. ▲

Глава 11. АНАЛИЗ ВРЕМЕННЫХ РЯДОВ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ

11.1. Понятие временных рядов

Временной ряд — это совокупность значений какого-либо показателя за несколько последовательных моментов времени.

Каждый уровень $y(t)$ временного ряда формируется под совместным влиянием длительных, кратковременных и случайных факторов:

• длительные, постоянно действующие факторы оказывают на изучаемое явление определяющее влияние и формируют основную тенденцию ряда — *тренд* $T(t)$;

• кратковременные, периодические факторы формируют *сезонные колебания ряда* $S(t)$;

• случайные факторы отражаются *случайными изменениями уровней ряда* $\varepsilon(t)$.

Модель, в которой временной ряд представлен как сумма перечисленных компонентов, т. е. $y(t) = T(t) + S(t) + \varepsilon(t)$, называется аддитивной.

Модель, в которой временной ряд представлен как произведение перечисленных компонентов, т. е. $y(t) = T(t)S(t)\varepsilon(t)$, называется *мультипликативной.*

Выбор одной из двух моделей осуществляется на основе анализа структуры сезонных колебаний:

• если амплитуда сезонных колебаний приближенно постоянная, то используют аддитивную модель;

• если амплитуда возрастает или уменьшается, то используют мультипликативную модель.

Основными задачами эконометрического исследования временного ряда является выявление каждой из перечисленных компонент ряда и прогнозирование поведения временного ряда на основе проведенных исследований.

Задача прогнозирования состоит в том, чтобы по значениям наблюдений, собранных к данному моменту, определить значения в следующие моменты.

11.2. Моделирование основной тенденции развития

Основной тенденцией развития (трендом) называется плавное и устойчивое изменение уровня явления во времени, свободное от случайных колебаний.

Методами выявления тренда являются:

- метод скользящей средней;
- аналитическое выравнивание.

Метод скользящей средней сводится к замене фактических уровней временного ряда *расчетными уровнями*, которые в меньшей степени подвержены колебаниям. Процедура данного метода заключается в следующем: рассматривается средний уровень из определенного числа первых по счету уровней ряда, затем из такого же числа уровней, но начиная со второго по счету и т. д. Средняя как бы «скользит» по ряду динамики, продвигаясь на один срок.

Пример 11.1. Рассчитаем скользящую среднюю (трех- и четырехлетнюю) по данным об урожайности зерновых культур за 10 лет.

▼Исходные данные и расчетные показатели представим в табл. 11.1.

Период скольжения может быть четным и нечетным.

Для *нечетного* (трехлетнего) периода первое значение скользящей средней есть $(15 + 13 + 15)/3 = 14{,}33$; второе — $(13 + 15 + 16)/3 = 14{,}67$ и т. д., причем полученные результаты скользящей средней отнесены к середине периода скольжения.

Для *четного* (четырехлетнего) периода первое значение скользящей средней есть $(15 + 13 + 15 + 16)/4 = 14{,}75$; второе — $(13 + 15 + 16 + 18)/4 = 15{,}50$ и т. д. Однако рассчитанные усредненные значения нельзя сопоставить каким-либо определенным

Таблица 11.1

Расчетная таблица для примера 11.1, ц/га

Год	Фактический уровень	Скользящая средняя		Центрированная скользящая средняя
		Трехлетняя	Четырехлетняя	
2006	15	—	—	—
2007	13	14,33	14,75	—
2008	15	14,67	15,50	15,125
2009	16	16,33	16,50	16,000
2010	18	17,00	16,75	16,625
2011	17	17,00	17,50	17,125
2012	16	17,33	17,25	17,375
2013	19	17,33	18,00	17,625
2014	17	18,67	—	—
2015	20	—	—	—

значениям t, поэтому применяют *процедуру центрирования* (вычисляют среднее значение из двух последовательных скользящих средних). Первое значение центрированной скользящей средней есть (14,75 + 15,50)/2 = 15,125, и результат будет отнесен к третьему году. Второе — (15,50 + 16,50)/2 = 16 и т. д. ▲

Метод аналитического выравнивания заключается в том, что фактические уровни y_t ряда заменяются плавно изменяющимися *теоретическими уровнями* $\hat{y}_t$.

На практике для описания тенденции широко используются *модели кривых роста*, представляющие собой различные функции времени $\hat{y}_t = f(t)$.

Параметры модели определяют обычным МНК, используя в качестве независимой переменной время $t = 1, 2, \ldots, n$, а в качестве зависимой переменной — фактические уровни временного ряда y_t.

При таком подходе изменение исследуемого показателя связывают лишь с течением времени; считается, что влияние других факторов несущественно или косвенно сказывается через фактор времени. Прогнозирование на основе модели кривой роста базируется на экстраполяции, т. е. на продлении в будущее тенденции, наблюдавшейся в прошлом.

В настоящее время компьютерные программы анализа временных рядов содержат широкий набор математических функций для построения уравнений тренда (линейных, логарифмических, полиномиальных, степенных, экспоненциальных, скользящих средних).

Модели тенденции можно сравнивать по величине коэффициента детерминации R^2. Чем больше R^2, тем в большей мере уравнение тренда подходит для описания тенденции временного ряда.

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

Измерить автокорреляцию в остатках можно с помощью коэффициента автокорреляции остатков. Чтобы судить об отсутствии (наличии) автокорреляции остатков, используют критерий Дарбина — Уотсона. Если установлено наличие автокорреляции остатков, то уравнение тренда не является наилучшим, так как нарушена предпосылка МНК об отсутствии автокорреляции остатков.

Пример 11.2. Имеются данные о розничном товарообороте региона за 10 лет (табл. 11.2).

Таблица 11.2

Исходные данные для примера 11.2

Номер года	1	2	3	4	5	6	7	8	9	10
Товарооборот (y), усл. ед.	11	13	22	18,5	20	19	25	23	24,5	35

Определим линейную трендовую модель товарооборота и оценим ее качество; построим точечный и интервальный прогнозы на три года вперед (при уровне значимости 0,05).

▼Исходные данные и расчетные показатели представим в *Excel* в виде табл. 11.3.

Таблица 11.3

Расчетная таблица для примера 11.2

Номер года	t	y	t^2	t_y	$\hat{y}$	e_t	e_{t-1}	$(e_t - e_{t-1})^2$	e_t^2
1	1	11	1	11	12,37	–1,37			1,88
2	2	13	4	26	14,31	–1,31	–1,37	0,0037	1,72

Окончание табл. 11.3

Номер года	t	y	t^2	t_y	$\hat{y}$	e_t	e_{t-1}	$(e_t - e_{t-1})^2$	e_t^2
3	3	22	9	66	16,25	5,75	–1,31	49,8522	33,05
4	4	18,5	16	74	18,19	0,31	5,75	29,5870	0,10
5	5	20	25	100	20,13	–0,13	0,31	0,1931	0,02
6	6	19	36	114	22,07	–3,07	–0,13	8,6400	9,42
7	7	25	49	175	24,01	0,99	–3,07	16,4885	0,98
8	8	23	64	184	25,95	–2,95	0,99	15,5188	8,69
9	9	24,5	81	220,5	27,89	–3,39	–2,95	0,1931	11,48
10	10	35	100	350	29,83	5,17	–3,39	73,2840	26,76
Итого	55	211	385	1320,5				193,7603	94,10
Среднее	5,5	21,1	38,5	132,05					

Окончательно имеем:

$$\mathrm{var}(t) = \overline{t^2} - \left(\overline{t}\right)^2 = 38{,}5 - 5{,}5^2 = 8{,}25;$$

$$\mathrm{cov}(t, y) = \overline{ty} - \overline{t} \cdot \overline{y} = 132{,}05 - 5{,}5 \cdot 21{,}1 = 16;$$

$$b = \frac{\mathrm{cov}(t, y)}{\mathrm{var}(t)} = 16/8{,}25 = 1{,}94;$$

$$a = \overline{y} - b\overline{t} = 21{,}1 - 1{,}94 \cdot 5{,}5 = 10{,}43,$$

следовательно, $\hat{y}_t = 10{,}43 + 1{,}94t$ — трендовая модель товарооборота.

Для проверки гипотезы об отсутствии автокорреляции остатков используется критерий Дарбина — Уотсона, рассчитываемый по формуле

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} = \frac{193{,}76}{94{,}10} = 2{,}06.$$

По таблице распределения Дарбина — Уотсона находим $d_1 = 0{,}88$ и $d_2 = 1{,}32$. Поскольку $d_2 < DW < 4 - d_2$, то нет оснований отклонять гипотезу H_0 об отсутствии автокорреляции в остатках.

Для вычисления точечного прогноза на три года вперед в построенную трендовую модель подставляем последовательно значения $t_p = 11;\ 12;\ 13$:

$$\hat{y}_{11} = 10{,}43 + 1{,}94 \cdot 11 = 31{,}77;$$
$$\hat{y}_{12} = 10{,}43 + 1{,}94 \cdot 12 = 33{,}71;$$
$$\hat{y}_{13} = 10{,}43 + 1{,}94 \cdot 13 = 35{,}65.$$

Для построения интервального прогноза рассчитаем доверительный интервал. Доверительный интервал для действительного значения y_p определяется выражением

$$\hat{y}_p - t_{\text{кр}} s_p < y_p < \hat{y}_p + t_{\text{кр}} s_p,$$

где s_p — стандартная ошибка предсказания;

$t_{\text{кр}}$ — критическое значение t-статистики при заданном уровне значимости и $\nu = n - 2$ степенях свободы.

Критическое значение $t_{\text{кр}} = 2{,}31$ можно получить в *Excel* с помощью функции =СТЬЮДРАСПОБР(α, ν).

Стандартная ошибка предсказания вычисляется по формуле

$$s_p = s\sqrt{1 + \frac{1}{n} + \frac{(t_p - \bar{t})^2}{n\,\mathrm{var}(t)}},$$

где $s = \sqrt{\dfrac{\sum\limits_{t=1}^{n} e_t^2}{n-2}} = \sqrt{\dfrac{94{,}10}{8}} = 3{,}4296.$

Для удобства расчета предсказания объема товарооборота с доверительным интервалом строим в *Excel* табл. 11.4.

В табл. 11.4 (t, y) — исходные данные; $\hat{y}_p, s_p$ — предсказание и его стандартная ошибка; $y_{\min} = \hat{y}_p - t_{\text{кр}} s_p$, $y_{\max} = \hat{y}_p + t_{\text{кр}} s_p$ — нижняя и верхняя границы доверительного интервала предсказания.

Таблица 11.4

Расчет предсказания объема товарооборота с доверительным интервалом для примера 11.2

t	y	$\hat{y}_p$	s_p	y_{min}	y_{max}
1	11	12,37	3,98	3,20	21,55
2	13	14,31	3,83	5,48	23,15
3	22	16,25	3,72	7,68	24,83
4	18,5	18,19	3,64	9,79	26,59
5	20	20,13	3,60	11,82	28,44
6	19	22,07	3,60	13,76	30,38
7	25	24,01	3,64	15,61	32,41
8	23	25,95	3,72	17,37	34,52
9	24,5	27,89	3,83	19,05	36,72
10	35	29,83	3,98	20,65	39,00
11		**31,77**	**4,15**	**22,19**	**41,34**
12		**33,71**	**4,35**	**23,66**	**43,75**
13		**35,65**	**4,58**	**25,09**	**46,20**

По данным табл. 11.4 строим график зависимости предсказания объема товарооборота с доверительным интервалом от времени (рис. 11.1).

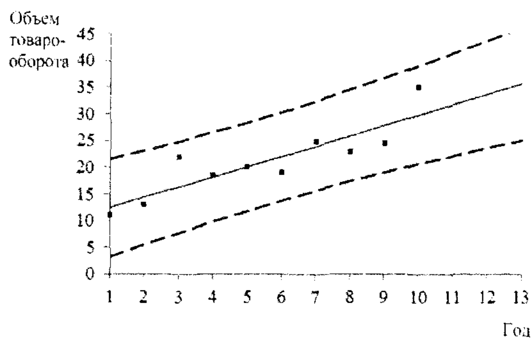


Рис. 11.1. График зависимости предсказания объема товарооборота с доверительным интервалом от времени для примера 11.2 ▲

11.3. Моделирование сезонных колебаний

Общий вид аддитивной модели: $Y = T + S + E$, мультипликативной модели: $Y = T \cdot S \cdot E$, при этом для аддитивной модели характеристики сезонности и случайной составляющей измеряются в абсолютных величинах, а в мультипликативной — в относительных.

Построение моделей включает в себя следующие шаги:

- выравнивание исходного ряда методом скользящей средней;
- расчет значений сезонной компоненты S;
- устранение сезонной компоненты из исходных уровней ряда $(Y - S)$ в аддитивной или (Y/S) в мультипликативной модели и получение выровненных данных $(T + E)$ в аддитивной или $(T \cdot E)$ в мультипликативной модели;
- аналитическое выравнивание уровней $(T + E)$ или $(T \cdot E)$ и расчет значений T с использованием полученного уравнения тренда;
- расчет полученных по модели значений $(T + E)$ или $(T \cdot E)$;
- расчет абсолютных или относительных ошибок, позволяющих оценить качество построенной модели.

Аддитивная модель временного ряда

Пример 11.3. Имеются поквартальные данные об объеме потребления электроэнергии (y) в некотором районе за четыре года (табл. 11.5).

Таблица 11.5

Исходные данные для примера 11.3, усл. ед.

Квартал	Год			
	1	2	3	4
I	60	72	80	90
II	44	48	56	66
III	50	60	64	70
IV	90	100	110	108

Построим аддитивную модель временного ряда. Определим прогнозное потребление электроэнергии на четыре квартала следующего года.

▼В качестве зависимой переменной при анализе временного ряда выступают фактические уровни ряда y_t, а в качестве независимой переменной — время (номер квартала) $t = 1, 2, \ldots, 16$.

По графику ряда можно установить наличие приблизительно линейного тренда и сезонных колебаний (период равен 4) одинаковой амплитуды, поэтому используется аддитивная модель. Произведем декомпозицию ряда (разложение на компоненты).

Для исключения влияния сезонной компоненты выполним выравнивание исходного ряда методом скользящей средней за четыре квартала и процедуру центрирования. Результаты расчетов представлены в табл. 11.6.

Таблица 11.6

Расчетная таблица для определения сезонной компоненты в примере 11.3

Номер квартала	Потребление электроэнергии (y_t)	Скользящая средняя за четыре квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
1	60	—	—	—
2	44	61	—	—
3	50	64	62,5	–12,5
4	90	65	64,5	25,5
5	72	67,5	66,25	5,75
6	48	70	68,75	–20,75
7	60	72	71	–11
8	100	74	73	27
9	80	75	74,5	5,5
10	56	77,5	76,25	–20,25
11	64	80	78,75	–14,75
12	110	82,5	81,25	28,75
13	90	84	83,25	6,75
14	66	83,5	83,75	–17,75
15	70	—	—	—
16	108	—	—	—

Оценки сезонной компоненты определяются как разность между фактическими уровнями ряда y_t и центрированными скользящими средними.

Расчет сезонной компоненты произведем в табл. 11.7, в которой оценки сезонной вариации записываются под соответствующим номером квартала в году.

Таблица 11.7

Расчет сезонной компоненты для примера 11.3

Год	Номер квартала в году				Сумма
	I	II	III	IV	
1	—	—	−12,50	25,50	—
2	5,75	−20,75	−11,00	27,00	—
3	5,5	−20,25	−14,75	28,75	—
4	6,75	−17,75	—	—	—
Итого	18,00	−58,75	−38,25	81,25	—
Среднее	6,00	−19,58	−12,75	27,08	0,75
Скорректированное S_i	5,81	−19,77	−12,94	26,90	0

В строке «Среднее» рассчитаны средние сезонной вариации по годам за каждый квартал и их сумма, равная 0,75.

В аддитивной модели предполагается, что сумма всех сезонных компонент по всем кварталам должна быть равна нулю (*условие взаимопогашаемости сезонных воздействий*).

В строке «Скорректированное S_i» рассчитаны значения сезонных компонент S_i как разность между средней сезонной вариацией и корректирующим коэффициентом 0,75/4, при этом $\sum_{i=1}^{4} S_i = 0$.

Таким образом, получены следующие значения сезонной компоненты:

$S_1 = 5{,}81$ (I квартал); $S_2 = -19{,}77$ (II квартал);

$S_3 = -12{,}94$ (III квартал); $S_4 = 26{,}90$ (IV квартал).

Расчет трендовой компоненты и ошибок произведем в табл. 11.8.

Каждый столбец таблицы имеет обозначение и формулу расчета.

В столбце «$T + E = Y - S$» исключается влияние сезонной компоненты: вычитая ее значение из каждого уровня исходного ряда, получим только тенденцию и случайную компоненту.

Таблица 11.8

Расчет трендовой компоненты и ошибок для примера 11.3

t	Y	S	$T + E = Y - S$	T	$Y_p = T + S$	$E = Y - Y_p$	$e = 100\ \% \cdot \lvert E \rvert / Y$
1	60	5,81	54,19	59,018	64,830	–4,830	8,05
2	44	–19,77	63,77	60,882	41,115	2,885	6,56
3	50	–12,94	62,94	62,746	49,809	0,191	0,38
4	90	26,90	63,11	64,611	91,503	–1,503	1,67
5	72	5,81	66,19	66,475	72,287	–0,287	0,40
6	48	–19,77	67,77	68,339	48,572	–0,572	1,19
7	60	–12,94	72,94	70,204	57,266	2,734	4,56
8	100	26,90	73,11	72,068	98,960	1,040	1,04
9	80	5,81	74,19	73,932	79,745	0,255	0,32
10	56	–19,77	75,77	75,796	56,029	–0,029	0,05
11	64	–12,94	76,94	77,661	64,723	–0,723	1,13
12	110	26,90	83,11	79,525	106,418	3,582	3,26
13	90	5,81	84,19	81,389	87,202	2,798	3,11
14	66	–19,77	85,77	83,254	63,486	2,514	3,81
15	70	–12,94	82,94	85,118	72,180	–2,180	3,11
16	108	26,90	81,11	86,982	113,875	–5,875	5,44
17		5,81		88,847	**94,659**		
18		–19,77		90,711	**70,943**		$\bar{e}$ **= 2,75 %**
19		–12,94		92,575	**79,638**		
20		26,90		94,439	**121,332**		

Проведя аналитическое выравнивание ряда ($T + E$) с помощью линейного тренда, получим следующее уравнение линии тренда:

$$T = 57{,}154 + 1{,}864t.$$

Уровни ряда T для каждого t = 1, 2, ..., 16 приведены в табл. 11.8.

Расчетное значение (подгонка) уровня $Y_р = T + S$ временного ряда в аддитивной модели есть сумма трендовой и сезонной компонент.

Рассчитываются ошибка $E = Y - Y_р$, абсолютная относительная ошибка $e = 100 \cdot |E| / Y$, % и средняя абсолютная относительная ошибка $\bar{e}$ = 2,75 %.

Прогнозные потребления электроэнергии в течении четырех кварталов следующего года (17–20-й кварталы) соответственно есть: 94,66; 70,94; 79,64; 121,33.

На рис. 11.2 показаны наблюдаемые и расчетные значения временного ряда.

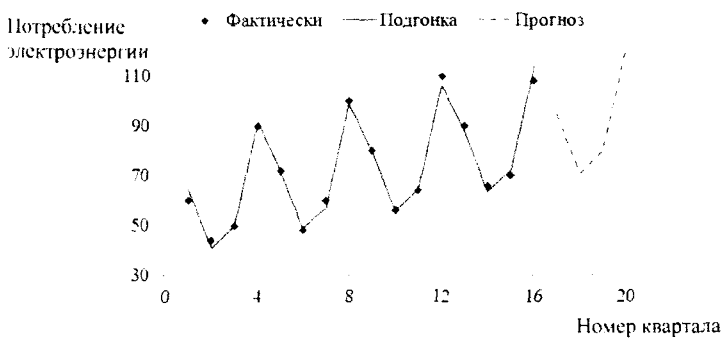


Рис. 11.2. График зависимости поквартального потребления электроэнергии от времени для примера 11.3 ▲

Мультипликативная модель временного ряда

Пример 11.4. Имеются поквартальные данные о выплате доходов компании акционерам в форме дивидендов (y) за четыре года (табл. 11.9).

Построим мультипликативную модель временного ряда и сделаем прогноз прибыли на четыре квартала следующего года (17–20-й кварталы).

Таблица 11.9

Исходные данные для примера 11.4, усл. ед.

Квартал	Год			
	1	2	3	4
I	40	50	60	70
II	60	80	100	110
III	50	70	80	130
IV	30	50	60	70

▼По графику ряда можно установить наличие приблизительно линейного тренда и сезонных колебаний (период равен 4) возрастающей амплитуды, поэтому используется мультипликативная модель. Произведем декомпозицию ряда.

Для исключения влияния сезонной компоненты проведем выравнивание исходного ряда методом скользящей средней за четыре квартала и процедуру центрирования. Результаты расчетов представлены в табл. 11.10.

Таблица 11.10

Расчетная таблица для определения сезонной компоненты в примере 11.4

Номер квартала	Размер дивидендов (y_t)	Скользящая средняя за четыре квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
1	40	—	—	—
2	60	45	—	—
3	50	47,5	46,25	1,081
4	30	52,5	50	0,600
5	50	57,5	55	0,909
6	80	62,5	60	1,333
7	70	65	63,75	1,098
8	50	70	67,5	0,741
9	60	72,5	71,25	0,842
10	100	75	73,75	1,356
11	80	77,5	76,25	1,049
12	60	80	78,75	0,762
13	70	92,5	86,25	0,811
14	110	95	93,75	1,173
15	130	—	—	—
16	70	—	—	—

Оценки сезонной компоненты для мультипликативной модели определяются как частное от деления фактических уровней ряда y_t на центрированные скользящие средние.

Расчет сезонной компоненты произведем в табл. 11.11, в которой оценки сезонной вариации записываются под соответствующим номером квартала в году.

Таблица 11.11

Расчет сезонной компоненты для примера 11.4

Год	**Номер квартала в году**				**Сумма**
	I	**II**	**III**	**IV**	
1	—	—	1,081	0,600	—
2	0,909	1,333	1,098	0,741	—
3	0,842	1,356	1,049	0,762	—
4	0,811	1,173	—	—	—
Итого	2,562	3,862	3,228	2,103	—
Среднее	0,854	1,287	1,076	0,701	3,918
Скорректированное S_i	0,872	1,314	1,098	0,715	4

В строке *Среднее* рассчитаны средние сезонной вариации по годам за каждый квартал и их сумма, равная 3,918.

В мультипликативной модели предполагается, что сумма всех сезонных компонент по всем кварталам должна быть равна четырем — числу сезонов в году (*условие взаимопогашаемости сезонных воздействий*).

В строке *Скорректированное* S_i рассчитаны значения сезонных компонент S_i как произведение соответствующей средней сезонной вариации на корректирующий коэффициент 4 : 3,918 = = 1,021, при этом $\sum_{i=1}^{4} S_i = 4$.

Получены следующие значения сезонной компоненты:

S_1 = 0,872 (I квартал); S_2 = 1,314 (II квартал);

S_3 = 1,098 (III квартал); S_4 = 0,715 (IV квартал).

Расчет трендовой компоненты и ошибок произведем в табл. 11.12.

Разделив каждый уровень исходного ряда на соответствующие значения сезонной компоненты (столбец $TE = Y/S$), исключим влияние сезонной компоненты и в результате получим только тенденцию и случайную компоненту.

Таблица 11.12

Расчет трендовой компоненты и ошибок для примера 11.4

t	Y	S	$TE = Y/S$	T	$Y_p = TS$	$E = Y - Y_p$	$e = 100\,\% \cdot \lvert E\rvert/Y$
1	40	0,87	45,87	38,972	33,982	6,018	15,04
2	60	1,31	45,65	43,047	56,574	3,426	5,71
3	50	1,10	45,52	47,123	51,760	–1,760	3,52
4	30	0,72	41,93	51,198	36,628	–6,628	22,09
5	50	0,87	57,34	55,273	48,197	1,803	3,61
6	80	1,31	60,87	59,349	77,997	2,003	2,50
7	70	1,10	63,73	63,424	69,665	0,335	0,48
8	50	0,72	69,89	67,500	48,290	1,710	3,42
9	60	0,87	68,81	71,575	62,411	–2,411	4,02
10	100	1,31	76,09	75,650	99,421	0,579	0,58
11	80	1,10	72,83	79,726	87,571	–7,571	9,46
12	60	0,72	83,87	83,801	59,952	0,048	0,08
13	70	0,87	80,28	87,876	76,625	–6,625	9,46
14	110	1,31	83,70	91,952	120,845	–10,845	9,86
15	130	1,10	118,35	96,027	105,476	24,524	18,86
16	70	0,72	97,85	100,102	71,614	–1,614	2,31
17		0,87		104,178	**90,840**		
18		1,31		108,253	**142,268**		$\bar{e}$ **= 6,94 %**
19		1,10		112,328	**123,382**		
20		0,72		116,404	**83,276**		

Производя аналитическое выравнивание ряда (TE) с помощью линейного тренда, получим следующее уравнение линии тренда:

$$T = 34{,}9 + 4{,}08t.$$

Уровни ряда T для каждого t = 1, 2, …, 16 приведены в табл. 11.12.

Столбец $Y_р = TS$ – расчетные значения временного ряда.

Ошибка в мультипликативной модели определяется по формуле

$$E = Y/Y_р.$$

Однако чтобы сравнивать мультипликативную модель с другими моделями временного ряда, расчет ошибки E произведем как $E = Y - Y_р$.

Средняя абсолютная относительная ошибка $\overline{e}$ = 6,94 %.

Прогнозное значение уровня временного ряда в мультипликативной модели есть произведение трендовой и сезонной компонент.

Прогноз прибыли в течение следующего года (17–20-й кварталы) соответственно есть: 90,84; 142,27; 123,38; 83,28.

На рис. 11.3 показаны наблюдаемые и расчетные значения временного ряда.

Рис. 11.3. График зависимости размера дивидендов от времени для примера 11.4▲

Упражнение 11.1. По показателям продаж продовольственных товаров в магазине (y) за последние 11 кварталов (табл. 11.13) разработайте модель продаж и сделайте прогноз объема продаж на следующие два квартала.

Таблица 11.13

Исходные данные для упражнения 11.1, усл. ед.

Квартал	1	2	3	4	5	6	7	8	9	10	11
Объем продаж	4	6	4	5	10	8	7	9	12	14	15

Указание. По графику ряда можно установить наличие приблизительно линейного тренда и сезонных колебаний (период равен 4) одинаковой амплитуды, поэтому используется аддитивная модель.

Ответ: объем продаж в 12 и 13-м кварталах $y(12) = 14{,}48$; $y(13) = 18{,}43$.

Упражнение 11.2. По данным об объеме продаж продовольственных товаров в магазине (y) за последние 11 кварталов разработайте модель продаж и сделайте прогноз объема продаж на следующие два квартала.

Таблица 11.14

Исходные данные для упражнения 11.2, усл. ед.

Квартал	1	2	3	4	5	6	7	8	9	10	11
Объем продаж	63	74	79	120	67	79	88	140	60	82	90

Указание. По графику ряда можно установить наличие приблизительно линейного тренда и сезонных колебаний (период равен 4) возрастающей амплитуды, поэтому используется мультипликативная модель.

Ответ: объем продаж в 12 и 13-м кварталах $y(12) = 139{,}3$; $y(13) = 68{,}0$.

11.4. Адаптивное прогнозирование

Адаптивные модели прогнозирования — это модели дисконтирования данных, способные быстро приспосабливать свою структуру и параметры к изменению условий. Инструментом прогноза в адаптивных моделях выступает математическая модель с единственным фактором, которым является время.

В основе адаптивных методов лежит *модель экспоненциального сглаживания*, которая описывается формулой

$$S_t = \lambda y_t + (1 - \lambda)S_{t-1},$$

где S_t — значение экспоненциальной средней в момент t;

λ — параметр сглаживания, $0 \le \lambda \le 1$.

Когда эта формула применяется рекурсивно, то каждое новое сглаженное значение (которое является также прогнозом) вычисляется как взвешенное среднее текущего наблюдения и сглаженного ряда. Очевидно, результат сглаживания зависит от параметра λ:

- если $\lambda = 1$, то предыдущие наблюдения полностью игнорируются;
- если $\lambda = 0$, то игнорируются текущие наблюдения;
- если $0 < \lambda < 1$, то значения λ дают промежуточные результаты.

Начало развитию методов адаптивного прогнозирования положено работами Ч. Хольта и Р. Г. Брауна. В настоящее время имеется обширная литература, в которой предлагаются различные методы и модели, относящиеся к адаптивному прогнозированию [4, 7, 10].

Общая схема построения адаптивных моделей:

- по нескольким первым наблюдениям ряда *оцениваются значения параметров* модели;
- по имеющейся модели *дается прогноз на один шаг*, причем его отклонение от фактических значений ряда расценивается как ошибка прогнозирования, которая учитывается в соответствии с принятой схемой корректировки параметров модели;
- по модели со скорректированными параметрами *рассчитывается прогнозная оценка на следующий момент времени*, и весь процесс повторяется до исчерпания фактических членов ряда;
- прогнозирование на будущее осуществляется с использованием параметров, определенных на *последнем* шаге по *последним* фактическим наблюдениям ряда.

Рассмотрим две наиболее часто используемые тренд-сезонные модели:

- модель Тейла — Вейджа с аддитивной сезонностью;
- модель Хольта — Уинтерса с мультипликативной сезонностью.

Поведение временного ряда можно объяснить, зная три его характеристики: основание, тренд и сезонность. В табл. 11.15

приведены рекуррентные уравнения параметров $\hat{a}_t, \hat{b}_t, \hat{w}_t$, описывающих поведение соответствующих характеристик ряда, и прогнозные значения $\hat{y}_\tau(t)$.

Таблица 11.15

Тренд-сезонные адаптивные модели

Модель Тейла — Вейджа с аддитивной сезонностью	
Оценка $\hat{a}_t, \hat{b}_t$	$\hat{a}_t = \lambda_1(y_t - \hat{w}_{t-c}) + (1-\lambda_1)(\hat{a}_{t-1} + \hat{b}_{t-1})$, $\hat{b}_t = \lambda_2(\hat{a}_t - \hat{a}_{t-1}) + (1-\lambda_2)\hat{b}_{t-1}$
Оценка сезонности	$\hat{w}_t = \lambda_3(y_t - \hat{a}_t) + (1-\lambda_3)\hat{w}_{t-c}$
Модель прогноза	$\hat{y}_\tau(t) = \hat{a}_t + \hat{b}_t\tau + \hat{w}_{t-c+\tau}$
Модель Хольта — Уинтерса с мультипликативной сезонностью	
Оценка $\hat{a}_t, \hat{b}_t$	$\hat{a}_t = \lambda_1 y_t / w_{t-c} + (1-\lambda_1)(\hat{a}_{t-1} + \hat{b}_{t-1})$, $\hat{b}_t = \lambda_2(\hat{a}_t - \hat{a}_{t-1}) + (1-\lambda_2)\hat{b}_{t-1}$
Оценка сезонности	$\hat{w}_t = \lambda_3 y_t / \hat{a}_t + (1-\lambda_3)\hat{w}_{t-c}$
Модель прогноза	$\hat{y}_\tau(t) = (\hat{a}_t + \hat{b}_t\tau)\hat{w}_{t-c+\tau}$

В приведенных формулах: c — количество периодов в сезонном цикле (например, $c = 12$ месяцев, $c = 4$ квартала); y_t — наблюдаемое в момент времени t значение временного ряда; $\lambda_1, \lambda_2, \lambda_3$ — параметры сглаживания, значения которых определяются из минимизации ошибки прогноза.

В качестве начальных значений $\hat{a}_0$, $\hat{b}_0$ берут оценки МНК коэффициентов линейного тренда, определенные по исходному временному ряду или его части.

Следует иметь в виду, что при расчете прогноза на момент t нужно иметь начальные значения сезонного фактора $\hat{w}_t$. Обычно эти начальные значения не учитываются, что приводит к дополнительной ошибке прогноза.

Начальные значения сезонного фактора для аддитивной модели определяются усреднением отклонений фактических уровней от расчетных для каждой фазы цикла (для одноименных месяцев, кварталов). Для мультипликативной модели — усреднением частного от деления фактических уровней на расчетные для каждой фазы цикла.

Пример 11.5. По данным примера 11.3 об объеме потребления электроэнергии построим модель Тейла — Вейджа с аддитивной сезонностью и сделаем прогноз на четыре квартала вперед.

▼Исходные данные (t, y_t) и расчетные показатели ($\hat{y}_t$, $\hat{a}_t$, $\hat{b}_t$, $\hat{w}_t$), полученные по приведенным выше рекуррентным формулам, представим в *Excel* в виде расчетной таблицы (рис. 11.4).

	A	B	C	D	E	F	G	H
1	t	y_t	$\hat{y}_\tau(t)$	$\hat{a}_t$	$\hat{b}_t$	$\hat{w}_t$	e_t	y_t
2	-3					5,52		
3	-2					-16,76		
4	-1					-10,03		
5	0			53,65	2,28	27,69		
6	1	60	61,45	55,93	2,28	4,58	2,41	55.93
7	2	44	41,45	58,20	2,28	-15,09	5,80	58,20
8	3	50	50,45	60.48	2,28	-10,32	0.89	60.48
9	4	90	90,45	62,76	2,28	27,40	0,50	62,76
10	5	72	69,61	65,03	2,28	6,14	3,32	65,03
11	6	48	52,22	67,31	2,28	-17,84	8,79	67,31
12	7	60	59,26	69,59	2,28	-9,84	1,23	69,59
13	8	100	99,26	71,86	2,28	27,88	0,74	71,86
14	9	80	80,27	74,14	2,28	5,96	0,34	
15	10	56	58,57	76,41	2,28	-19,52	4,59	
16	11	64	68,85	78,69	2,28	-13,01	7,58	
17	12	110	108,85	80,97	2,28	28,63	1,05	
18	13	90	89,20	83,24	2,28	6,48	0,89	
19	14	66	66,00	85,52	2,28	-19,52	0,00	
20	15	70	74,79	87,80	2,28	-16,13	6,84	
21	16	108	118,71	90,07	2,28	21,65	9,91	
22	17		98,83	λ_1	λ_2	λ_3	$\bar{e}$	
23	18		75,11	0,000	0,196	0,652	3,431	
24	19		80,77					
25	20		120,83					

Рис. 11.4. Расчетная таблица для примера 11.5 в *Excel*

Начальные значения $\widehat{a}_0 = 53{,}65$ и $\widehat{b}_0 = 2{,}28$ получены из оценок МНК коэффициентов линейного тренда, определенных по исходным данным:

$$\widehat{y}_t = 53{,}65 + 2{,}28t\ .$$

Начальные значения сезонного фактора определялись по формулам

$$\widehat{w}_{-3} = 0{,}5(y_1 - \widehat{y}_1 + y_5 - \widehat{y}_5); \quad \widehat{w}_{-1} = 0{,}5(y_3 - \widehat{y}_3 + y_7 - \widehat{y}_7);$$
$$\widehat{w}_{-2} = 0{,}5(y_2 - \widehat{y}_2 + y_6 - \widehat{y}_6); \quad \widehat{w}_0 = 0{,}5(y_4 - \widehat{y}_4 + y_8 - \widehat{y}_8).$$

Поясним ввод формул в соответствующие ячейки таблицы (табл. 11.16).

Таблица 11.16

Использование формул *Excel* для решения примера 11.5

Ячейка	Формула
H6:H13	Размещены расчетные значения $\widehat{y}_t = 53.65 + 2.28t$, $t = 1, \ldots, 8$
F2	= (B6-H6+B10-H10)/2
F3:F5	"Протягиваем" формулу из ячейки F2
C6	= (D5+E5)+F2
D6	= $D$23*(B6-F2)+(1-$D$23)*(D5+E5)
E6	= $E$23*(D6-D5)+(1-$E$23)*E5
F6	= $F$23*(B6-D6)+(1-$F$23)*F2
G6	= 100*ABS(B6-C6)/B6
C7:G21	«Протягиваем» формулы из ячеек C6:G6
Прогнозные значения	
C22	= ($D$21+$E$21)+F18
C23	= ($D$21+$E$21*2)+F19
C24	= ($D$21+$E$21*3)+F20
C25	= ($D$21+$E$21*4)+F21
D23:F23 (изменяемые ячейки)	Размещены параметры сглаживания $\lambda_1, \lambda_2, \lambda_3$
G23	= СРЗНАЧ(G6:G21)

Оптимальные значения параметров сглаживания $\lambda_1 = 0$; $\lambda_2 = 0{,}196$; $\lambda_3 = 0{,}652$ определяются из минимума средней абсолютной относительной ошибки $\overline{e}$ с помощью инструмента *Поиск решения*. Ошибка подгонки $\overline{e} = 3{,}43$ %.

Результаты расчета по модели Тейла — Вейджа представлены на рис. 11.5, из которого видно, что значения подгонки хорошо воспроизводят сезонные колебания.

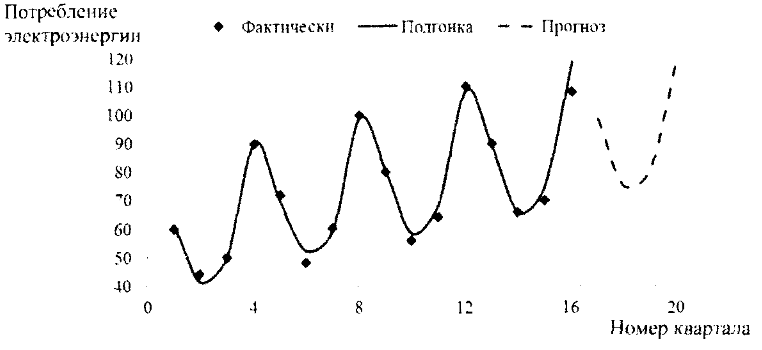


Рис. 11.5. График зависимости поквартального потребления электроэнергии от времени для примера 11.5 ▲

Пример 11.6. По данным об объеме пассажирских авиаперевозок (y_t) одной из авиакомпаний за шесть лет (табл. 11.17) сделаем прогноз на следующий год.

Таблица 11.17

Исходные данные для примера 11.6, тыс. чел.

Месяц	Год					
	1	2	3	4	5	6
Январь	272	238	285	325	367	407
Февраль	224	190	272	300	329	357
Март	284	252	300	316	378	416
Апрель	272	252	319	349	410	452
Май	292	290	363	371	441	485
Июнь	330	347	399	433	520	520
Июль	385	409	487	541	590	590
Август	441	476	521	563	600	617
Сентябрь	352	390	402	465	510	521
Октябрь	289	332	351	402	442	460

Ноябрь	237	308	303	350	390	420
Декабрь	251	300	329	378	400	433

▼В качестве зависимой переменной при анализе временного ряда выступают фактические уровни ряда y_t, а в качестве независимой переменной — время (номер месяца) $t = 1, 2, \ldots, 72$.

По графику ряда можно установить наличие приблизительно линейного тренда и сезонных колебаний (период равен 12) одинаковой амплитуды, поэтому используется аддитивная модель (Тейла — Вейджа).

Исходные данные и расчетные показатели представлены в табл. 11.18.

Таблица 11.18

Расчетная таблица для примера 11.6

t	y_t	$\hat{y}_1(t)$	$\hat{a}_t$	$\hat{b}_t$	$\hat{w}_t$	e_t	$\hat{y}_t$
–11					–38,82		
–10					–89,55		
–9					–31,29		
–8					–40,03		
–7					–13,77		
–6					30,99		
–5					86,76		
–4					145,52		
–3					55,28		
–2					–7,96		
–1					–48,70		
0			274,65	2,74	–48,43		
1	272	238,57	298,63	2,74	–26,63	12,29	277,39
2	224	211,81	309,11	2,74	–85,11	5,44	280,13
3	284	280,56	314,04	2,74	–30,04	1,21	282,87
4	272	276,75	313,76	2,74	–41,76	1,74	285,60
5	292	302,73	309,68	2,74	–17,68	3,67	288,34
6	330	343,41	303,90	2,74	26,10	4,06	291,08
7	385	393,39	301,30	2,74	83,70	2,18	293,82
8	441	449,56	298,60	2,74	142,40	1,94	296,55
9	352	356,62	298,40	2,74	53,60	1,31	299,29
10	289	293,18	298,48	2,74	–9,48	1,45	302,03

t	y_t	$\hat{y}_1(t)$	$\hat{a}_t$	$\hat{b}_t$	$\hat{w}_t$	e_t	$\hat{y}_t$
11	237	252,53	291,36	2,74	–54,36	6,55	304,77
12	251	245,66	297,49	2,74	–46,49	2,13	307,51
13	238	273,59	277,61	2,74	–39,61	14,96	310,24
14	190	195,23	277,02	2,74	–87,02	2,75	312,98
15	252	249,72	281,21	2,74	–29,21	0,90	315,72
16	252	242,18	290,18	2,74	–38,18	3,90	318,46
17	290	275,24	302,30	2,74	–12,30	5,09	321,20
18	347	331,14	315,11	2,74	31,89	4,57	323,93
19	409	401,55	322,59	2,74	86,41	1,82	326,67
20	476	467,72	330,58	2,74	145,42	1,74	329,41
21	390	386,92	335,28	2,74	54,72	0,79	332,15
22	332	328,54	340,22	2,74	–8,22	1,04	334,88
23	308	288,60	355,28	2,74	–47,28	6,30	337,62
24	300	311,54	350,69	2,74	–50,69	3,85	340,36
25	285	313,82	335,11	2,74	–50,11	10,11	
…	…	…	…	…	…	…	…
71	420	420,51	485,36	2,74	–65,36	0,12	
72	433	441,95	482,41	2,74	–49,41	2,07	
73		437,43	λ_1	λ_2	λ_3	$\bar{e}$	
74		398,22	0,635	0	1,000	3,652	
75		450,38					
76		476,44					
77		499,96					
78		545,07					
79		616,77					
80		639,14					
81		545,27					
82		486,84					
83		447,17					
84		465,85					

Начальные значения $\hat{a}_0 = 274{,}65$ и $\hat{b}_0 = 2{,}74$ получены из оценок МНК коэффициентов линейного тренда, определенных по данным наблюдений за два первых года.

Начальные значения сезонного фактора определялись по формулам

$$\hat{w}_{-11} = 0,5(y_1 - \hat{y}_1 + y_{13} - \hat{y}_{13});$$

$$\hat{w}_{-10} = 0,5(y_2 - \hat{y}_2 + y_{14} - \hat{y}_{14});$$

....

$$\hat{w}_0 = 0,5(y_{12} - \hat{y}_{12} + y_{24} - \hat{y}_{24}).$$

Ошибка подгонки $\bar{e}$ = 3,65 %.

Результаты расчета по модели представлены на рис. 11.6, из которого видно, что прогнозные значения хорошо воспроизводят сезонные колебания.

Данную модель можно считать пригодной для прогнозирования.

Рис. 11.6. График зависимости объема пассажирских авиаперевозок от времени для примера 11.6 ▲

Пример 11.7. По данным примера 11.4 о выплате дивидендов построим модель Хольта — Уинтерса с мультипликативной сезонностью и сделаем прогноз на четыре квартала вперед.

▼Исходные данные и расчетные показатели представим в *Excel* в виде расчетной таблицы (рис. 11.7).

Начальные значения $\hat{a}_0$ = 37,50 и $\hat{b}_0$ = 3,75 получены из оценок МНК коэффициентов линейного тренда, определенных по исходным данным:

$$\hat{y}_t = 37,50 + 3,75t.$$

	A	B	C	D	E	F	G	H
1	t	y_t	$y_t(t)$	a_t	b_t	w_t	e_t	y_t
2	-3					0,93		
3	-2					1,33		
4	-1					1,06		
5	0			37,50	3,75	0,66		
6	1	40	38,33	41,25	3,75	0,96	4,17	41,25
7	2	60	60,00	45,00	3,75	1,33	0,00	45,00
8	3	50	51,76	48,75	3,75	1,04	3,53	48,75
9	4	30	34,44	52,50	3,75	0,60	14,81	52,50
10	5	50	53,85	56,25	3,75	0,91	7,70	56,25
11	6	80	80,00	60,00	3,75	1,33	0,00	60,00
12	7	70	66,09	63,75	3,75	1,08	5,58	63,75
13	8	50	40,32	67,50	3,75	0,70	19,36	67,50
14	9	60	64,83	71,25	3,75	0,86	8,04	
15	10	100	100,00	75,00	3,75	1,33	0,00	
16	11	80	84,99	78,75	3,75	1,04	6,24	
17	12	60	57,49	82,50	3,75	0,72	4,19	
18	13	70	74,42	86,25	3,75	0,83	6,32	
19	14	110	120,00	90,00	3,75	1,26	9,09	
20	15	130	97,06	93,75	3,75	1,28	25,34	
21	16	70	70,00	97,50	3,75	0,72	0,00	
22	17		83,76	λ_1	λ_2	λ_3	$\bar{e}$	
23	18		131,91	0	0	0,694	7,148	
24	19		139,10					
25	20		80,77					

Рис. 11.7. Расчетная таблица для примера 11.7 в *Excel*

Начальные значения сезонного фактора определялись по формулам

$$\hat{w}_{-3} = 0,5(y_1 / \hat{y}_1 + y_5 / \hat{y}_5); \quad \hat{w}_{-1} = 0,5(y_3 / \hat{y}_3 + y_7 / \hat{y}_7);$$
$$\hat{w}_{-2} = 0,5(y_2 / \hat{y}_2 + y_6 / \hat{y}_6); \quad \hat{w}_0 = 0,5(y_4 / \hat{y}_4 + y_8 / \hat{y}_8).$$

Поясним ввод формул в соответствующие ячейки таблицы (табл. 11.19).

Таблица 11.19

Использование формул *Excel* для решения примера 11.7

Ячейка	Формула
H6:H13	Размещены расчетные значения $\tilde{y}_t = 37{,}50 + 3{,}75t$, $t = 1, \dots, 8$.
F2	= (B6/H6+B10/H10)/2
F3:F5	"Протягиваем" формулу из ячейки F2
C6	= (D5+E5)*F2
D6	= \$D\$23*B6/F2+(1-\$D\$23)*(D5+E5)
E6	= \$E\$23*(D6-D5)+(1-\$E\$23)*E5
F6	= \$F\$23*B6/D6+(1-\$F\$23)*F2
G6	= 100*ABS(B6-C6)/B6
C7:G21	«Протягиваем» формулы из ячеек C6:G6
Прогнозные значения	
C22	= (\$D\$21+\$E\$21)*F18
C23	= (\$D\$21+\$E\$21*2)*F19
C24	= (\$D\$21+\$E\$21*3)*F20
C25	= (\$D\$21+\$E\$21*4)*F21
D23:F23 (изменяемые ячейки)	Размещены параметры сглаживания $\lambda_1, \lambda_2, \lambda_3$
G23	= СРЗНАЧ(G6:G21)

Ошибка подгонки $\bar{e}$ = 7,15 %.

Результаты расчета по модели Хольта — Уинтерса представлены на рис. 11.8, из которого видно, что значения подгонки хорошо воспроизводят сезонные колебания.

Рис. 11.8. График зависимости размера дивидендов от времени для примера 11.7▲

Экспоненциальное сглаживание — наиболее простой способ построения прогнозов, часто дающий быстрые эффективные результаты. Однако этот метод не позволяет строить доверительные интервалы и, следовательно, рассчитать риски при использовании прогнозов.

11.5. Автокорреляция уровней временного ряда

При наличии во временном ряде тенденции и сезонных колебаний значения каждого последующего уровня ряда зависит от предыдущих.

Корреляционная зависимость между последовательными уровнями временного ряда называется *автокорреляцией* уровней ряда. Для измерения автокорреляции уровней ряда используется коэффициент автокорреляции.

Коэффициент автокорреляции порядка τ определяется как коэффициент корреляции между рядами y_t, $y_{t-\tau}$:

$$r_\tau = \frac{\operatorname{cov}(y_t, y_{t-\tau})}{\sqrt{\operatorname{var}(y_t)\operatorname{var}(y_{t-\tau})}}$$

Число периодов τ, по которым рассчитывается коэффициент автокорреляции, называется *лагом*.

Коэффициенты автокорреляции разных порядков принято обозначать как r_k, где $k = 1, 2, \ldots$ указывает на порядок коэффициента автокорреляции.

Пример 11.8. По данным примера 11.3 об объеме потребления электроэнергии (y) в некотором районе за четыре года определите коэффициенты автокорреляции до 4-го порядка.

▼Создадим таблицу уровней временного ряда (табл. 11.20).

Последовательно вводя данные (y_t, y_{t-1}), (y_t, y_{t-2}), (y_t, y_{t-3}), (y_t, y_{t-4}), с помощью функции *Excel* =КОРРЕЛ получим следующие коэффициенты автокорреляции: $r_1 = 0{,}165$; $r_2 = -0{,}567$; $r_3 = 0{,}114$; $r_4 = 0{,}983$. ▲

Последовательность коэффициентов автокорреляции первого, второго и более высоких порядков называется *автокорреляционной функцией* (АКФ) временного ряда. График зависимости ее значений от величины лага (порядка коэффициента автокорреляции) называется *коррелограммой*.

Таблица 11.20

Таблица уровней временного ряда для примера 11.8

y_t	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-4}
60				
44	60			
50	44	60		
90	50	44	60	
72	90	50	44	60
48	72	90	50	44
60	48	72	90	50
100	60	48	72	90
80	100	60	48	72
56	80	100	60	48
64	56	80	100	60
110	64	56	80	100
90	110	64	56	80
66	90	110	64	56
70	66	90	110	64
108	70	66	90	110

Автокорреляционную функцию обычно используют для выявления во временном ряде наличия или отсутствия трендовой и сезонной компонент:

• если наиболее высоким оказался коэффициент автокорреляции первого порядка, то исследуемый ряд содержит только тенденцию;

• если наиболее высоким оказался коэффициент автокорреляции порядка $\tau > 1$, то ряд содержит сезонные колебания с периодом τ;

• если ни один из коэффициентов автокорреляции не является значимым, то либо ряд не содержит тенденции и сезонных колебаний, либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ.

11.6. Учет тенденции при построении модели регрессии

При изучении развития явления во времени иногда возникает необходимость оценить степень взаимосвязи в изменениях уровней двух и более временных рядов различного содержания, но связанных один с другим.

При установлении связи между уровнями временных рядов может иметь место *ложная корреляция* — установление связи там, где ее на самом деле нет.

Наличие ложной корреляции связано с тенденцией каждого из рядов динамики, с автокорреляцией их уровней. Если ряды динамики характеризуются наличием тренда, то при построении модели регрессии надо исключить тренд, в противном случае корреляция уровней рядов динамики будет преувеличена.

При построении модели регрессии по временным рядам при наличии тенденции используются следующие методы:

• метод отклонений от тренда;

• метод последовательных разностей;

• метод включения в уравнение регрессии фактора времени.

Метод отклонений от тренда. Предположим, имеются два ряда динамики y_t, x_t, содержащие трендовую и случайную компоненты.

Проведение аналитического выравнивания по каждому из этих рядов позволяет найти параметры соответствующих уравнений трендов и определить их расчетные уровни $\hat{y}_t$, $\hat{x}_t$. Эти расчетные значения можно принять за оценку трендовой компоненты каждого ряда. Поэтому влияние тенденции можно устранить путем вычитания этих расчетных значений из фактических уровней ряда.

Дальнейший анализ взаимосвязи рядов проводят с использованием не исходных уровней, а рассчитанных отклонений от тренда

$$dy_t = y_t - \hat{y}_t, \; dx_t = x_t - \hat{x}_t$$

при условии, что последние не содержат тенденцию.

Метод последовательных разностей. Исключить влияние автокорреляции можно путем перехода от исходных уровней ряда y_t, x_t к первым разностям $\Delta y_t = y_t - y_{t-1}$, $\Delta x_t = x_t - x_{t-1}$ в случае линейного тренда и ко вторым разностям в случае параболического тренда. В случае экспоненциального и степенного трендов метод последовательных разностей применяется к логарифмам исходных данных.

Метод включения в уравнение регрессии фактора времени. Модель регрессии по временным рядам может быть построена с включением в нее как отдельной независимой переменной фактора времени t. Считается, что введение фактора времени исключает тенденцию развития всех явлений, представленных исследуемыми рядами динамики.

Доказано, что если временные ряды характеризуются линейной тенденцией, то включение в модель фактора времени t равносильно построению модели регрессии по отклонениям от трендов с последующим переходом от нее к исходным уровням временного ряда зависимой переменной y.

Пример 11.9. Имеются данные о расходах на конечное потребление (y_t) и совокупном доходе (x_t) за 10 лет (усл. ед). Построим модель регрессии и сделаем прогноз на 11-й год, используя методы исключения тенденции и включения фактора времени.

▼Исходные данные и расчетные показатели представим в виде табл. 11.21.

Таблица 11.21

Расчетная таблица для примера 11.9

t	y_t	x_t	$\hat{y}_t$	$\hat{x}_t$	dy_t	dx_t	Δy_t	Δx_t
1	6	24	6,073	23,00	–0,073	1,00		
2	10	29	8,945	27,40	1,055	1,60	4	5
3	12	33	11,818	31,80	0,182	1,20	2	4
4	14	35	14,691	36,20	–0,691	–1,20	2	2
5	17	39	17,564	40,60	–0,564	–1,60	3	4
6	21	45	20,436	45,00	0,564	0,00	4	6
7	23	48	23,309	49,40	–0,309	–1,40	2	3
8	25	50	26,182	53,80	–1,182	–3,80	2	2
9	28	55	29,055	58,20	–1,055	–3,20	3	5
10	34	70	31,927	62,60	2,073	7,40	6	15

Если к исходным данным y_t, x_t применить МНК, то получим уравнение регрессии

$$\hat{y}_t = -8,14 + 0,634x_t,\ R^2 = 0,983.$$

Полученные результаты содержат ложную корреляцию ввиду наличия в каждом из рядов y_t, x_t линейной тенденции.

Применим различные методы устранения тенденции.

Метод отклонения от тренда. По трендам

$$\hat{y}_t = 3,2 + 2,87t,\ \hat{x}_t = 18,6 + 4,4t$$

определим расчетные значения $\hat{y}_t$, $\hat{x}_t$ и отклонения от трендов dy_t, dx_t.

Применяя к рядам dy и dx МНК, получим уравнение линейной регрессии

$$d\hat{y}_t = 0,30 \cdot dx_t,\ R^2 = 0,893.$$

Для прогноза значений y перейдем к уравнению, связывающему между собой уровни временных рядов:

$$y_p - \hat{y}_{t=p} = 0,30(x_p - \hat{x}_{t=p}),$$

где y_p, x_p — прогнозные значения y, x;

$\hat{y}_{t=p}$, $\hat{x}_{t=p}$ — прогнозы y, x по тренду при $t = p$.

Для прогноза на 11-й год имеем:

$\hat{y}_{t=p} = 3,2 + 2,87 \cdot 11 = 34,8$; $\hat{x}_{t=p} = 18,6 + 4,4 \cdot 11 = 67$; $x_p = 71$ (принято представлять как $x_n = 70$ плюс $\Delta n = 1$), тогда

$$y_p = 34,8 + 0,30(71 - 67) = 36.$$

Метод последовательных разностей. Найдем первые разности Δy_t, Δx_t. Используя МНК, получим уравнение регрессии

$$\Delta y_t = 1,48 + 0,32 \cdot \Delta x_t,\ R^2 = 0,856.$$

Для прогноза значений y перейдем к уравнению, связывающему между собой уровни временных рядов.

Прогноз на 11-й год выполним по уравнению

$$y_p - y_n = 1{,}48 + 0{,}32(x_p - x_n),$$

где $y_n = 34$, $x_n = 70$ — конечные уровни рядов.

По уравнению линейного тренда найдем $x_p = 18{,}6 + 4{,}4 \cdot 11 = 67$, тогда

$$y_p = 34 + 1{,}48 + 0{,}32(67 - 70) = 34{,}52.$$

Метод включения фактора времени. Модель регрессии с включением в нее фактора времени имеет вид

$$\hat{y}_t = -2{,}40 + 0{,}30 \cdot x_t + 1{,}55 \cdot t,\ R^2 = 0{,}999.$$

Значение прогноза y_p на 11-й год составляет

$$y_p = -2{,}40 + 0{,}30 \cdot 71 + 1{,}55 \cdot 11 = 36.$$

Результат прогноза совпадает с прогнозом при использовании метода отклонения от тренда. ▲

Полученные прогнозные уравнения хорошо учитывают тенденцию, если отсутствует автокорреляция в остатках $e_t = y_t - \hat{y}_t$, т. е. остатки текущего периода не коррелируют с остатками предыдущего периода. Чтобы судить об отсутствии (наличии) автокорреляции остатков, используют критерий Дарбина — Уотсона. В уравнениях примера 11.9 автокорреляция в остатках отсутствует.

11.7. Учет сезонности при построении модели регрессии

Если в рядах динамики тенденция и сезонная компонента наблюдаются одновременно, то в модель можно включить фактор времени и фиктивные переменные.

Пример 11.10. По данным примера 11.3 об объеме потребления электроэнергии рассмотрим построение модели временного ряда с использованием фактора времени и фиктивных переменных.

▲Возьмем I квартал каждого года в качестве эталонного и будем использовать фиктивные переменные для оценки разности между ним и другими кварталами.

Запишем модель как

$$y = \alpha + \beta t + \delta_1 z_1 + \delta_2 z_2 + \delta_3 z_3 + \varepsilon,$$

где z_1, z_2, z_3 — фиктивные переменные, определяемые следующим образом:

$$z_1 = \begin{cases} 1, & \text{(II квартал)} \\ 0; & \text{(остальные)} \end{cases} \quad z_2 = \begin{cases} 1, & \text{(III квартал)} \\ 0; & \text{(остальные)} \end{cases} \quad z_3 = \begin{cases} 1, & \text{(IV квартал)} \\ 0. & \text{(остальные)} \end{cases}$$

Коэффициенты δ_1, δ_2, δ_3 дают численную величину эффекта, вызываемого сменой года.

Исходные данные с фиктивными переменными представим в табл. 11.22.

Таблица 11.22

Исходные данные и значения фиктивных переменных для примера 11.10

y	t	z_1	z_2	z_3
60	1	0	0	0
44	2	1	0	0
50	3	0	1	0
90	4	0	0	1
72	5	0	0	0
48	6	1	0	0
60	7	0	1	0
100	8	0	0	1
80	9	0	0	0
56	10	1	0	0
64	11	0	1	0
110	12	0	0	1
90	13	0	0	0
66	14	1	0	0
70	15	0	1	0
108	16	0	0	1

Оцененное уравнение регрессии:

$$\hat{y} = 62,375 + 1,875t - 23,875z_1 - 18,25z_2 + 20,875z_3,\ R^2 = 0,985,$$

где все коэффициенты при переменных статистически значимы.

Отдельные уравнения для каждого квартала таковы:

$\hat{y} = 62,375 + 1,875t$, (I квартал)

$\hat{y} = 38,5 + 1,875t$, (II квартал)

$\hat{y} = 44,125 + 1,875t$, (III квартал)

$\hat{y} = 83,25 + 1,875t$. (IV квартал)

Усредняя эти уравнения, получим линейный тренд

$$T = \frac{62,375 + 38,5 + 44,125 + 83,25}{4} + 1,875t = 57,06 + 1,875t.$$

Расстояние между линией регрессии каждого квартала и трендом дает оценку сезонной компоненты в данном квартале:

$S_1 = 62,375 - 57,06 = 5,31$, (I квартал)

$S_2 = 38,50 - 57,06 = -18,56$, (II квартал)

$S_3 = 44,125 - 57,06 = -12,94$, (III квартал)

$S_4 = 83,25 - 57,06 = 26,19$, (IV квартал)

причем сумма сезонных отклонений должна равняться нулю, т. е.

$$S_1 + S_2 + S_3 + S_4 = 0.$$

Прогнозное значение уровня временного ряда в аддитивной модели есть сумма трендовой и сезонной компонент.

Прогнозное потребление электроэнергии в 17–20 кварталах есть:

$y(17) = T(17) + S_1 = 57,06 + 1,875 \cdot 17 + 5,31 = 94,245,$

$y(18) = T(18) + S_2 = 57,06 + 1,875 \cdot 18 - 18,56 = 72,25,$

$y(19) = T(19) + S_3 = 57,06 + 1,875 \cdot 19 - 12,94 = 79,745,$

$y(20) = T(20) + S_4 = 57,06 + 1,875 \cdot 20 + 26,19 = 120,75.$

Средняя абсолютная относительная ошибка $\bar{e} = 2,8$ %.▲

11.8. Прогнозирование с помощью моделей авторегрессии — проинтегрированного скользящего среднего

Часто экономические показатели, представленные временными рядами, имеют настолько сложную структуру, что моделирование таких рядов путем построения моделей тренда, сезонности и применение других традиционных подходов не приводят к удовлетворительным результатам. Временной ряд остатков часто имеет статистические закономерности, которые можно моделировать. Будем рассматривать класс стационарных временных рядов.

Ряд y_t называется *строго стационарным* (или *стационарным в узком смысле*), если совместное распределение m наблюдений $y_{t_1}, y_{t_2}, \ldots, y_{t_m}$ не зависит от сдвига по времени, т. е. совпадает с совместным распределением наблюдений $y_{t_1+\tau}, y_{t_2+\tau}, \ldots, y_{t_m+\tau}$ при любых $m, t_1, t_2, \ldots, t_m, \tau$.

Обычно интересует не все распределение, а средние значения и ковариации. Поэтому часто используется понятие *слабой стационарности* (или *стационарности в широком смысле*), которое состоит в том, что среднее, дисперсия и ковариации y_t не зависят от момента времени t:

$$M(y_t) = \mu,\ D(y_t) = \gamma_0,\ \mathrm{cov}(y_t, y_{t-k}) = \gamma_k.$$

В дальнейшем под стационарностью будем понимать слабую стационарность. Временной ряд, не удовлетворяющий перечисленным выше свойствам, называется нестационарным.

Рассмотрим автокорреляционную функцию (АКФ)

$$\rho(k) = \mathrm{cov}(y_t, y_{t-k}) : D(y_t) = \gamma_k/\gamma_0.$$

Значения АКФ характеризуют тесноту (степень) статистической связи между уровнями временного ряда, разделенными k временными тактами.

Наряду с автокорреляционной функцией при исследовании стационарных временных рядов рассматривается *частная автокорреляционная функция* $\rho_ч(k)$ (ЧАКФ). С помощью ЧАКФ измеряется корреляция между уровнями ряда y_t и y_{t-k}, разделенными k временными тактами, при исключении влияния на эту взаимосвязь всех промежуточных уровней ряда $y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}$.

Примером стационарного по определению временного ряда является ряд с независимыми одинаково распределенными наблюдениями:

$$y_t = \varepsilon_t,\ \varepsilon_t \sim iid(0, \sigma^2),\ t = 1, \ldots, n,$$

т. е. ошибки ε_t независимые, одинаково распределенные с нулевым средним и дисперсией σ^2 (*independent and identically distrib-*

uted). Этот процесс называется *белым шумом*, у него $\mu = 0$, $\gamma_0 = \sigma^2$, $\gamma_k = 0$, $k > 0$.

Большое значение в анализе временных рядов имеют стационарные временные ряды, вероятностные свойства которых не изменяются во времени. Это объясняется тем, что многие временные ряды могут быть приведены к стационарному ряду после выделения и удаления из них тренда, сезонной компоненты или взятия разностей. Как правило, ряд ошибок является стационарным рядом.

Наиболее распространенными моделями стационарных рядов являются модели авторегрессии и модели скользящего среднего.

Модель авторегрессии *AR*(*p*). В авторегрессионной модели каждое значение ряда y_t зависит только от конечного числа p предыдущих своих значений.

Авторегрессионной процесс $AR(p)$ может быть представлен в виде

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \varepsilon_t$$

или в более короткой форме

$$(1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_p L^p) y_t = \varepsilon_t ,$$

где L — оператор сдвига, т. е. преобразование ряда, смещающее его на один временной такт.

Простейшим примером является модель $AR(1)$, или *марковский процесс*:

$$y_t = \alpha y_{t-1} + \varepsilon_t ,$$

где ε_t — белый шум.

Используя оператор сдвига, запишем уравнение в виде

$$(1 - \alpha L) y_t = \varepsilon_t.$$

Разлагая оператор $(1 - \alpha L)^{-1}$ в ряд по αL, получим

$$y_t = (1 - \alpha L)^{-1} \varepsilon_t = (1 + \alpha L + \alpha^2 L^2 + \ldots)\, \varepsilon_t = \varepsilon_t + \alpha\varepsilon_{t-1} + \alpha^2\varepsilon_{t-2} + \ldots,$$

откуда $M(y_t) = 0$, $D(y_t) = \sigma^2/(1 - \alpha^2)$, $\mathrm{cov}(y_t, y_{t\ k}) = \alpha^k\sigma^2/(1 - \alpha^2)$.

Если в уравнение добавить константу:

$$y_t = \delta + \alpha y_{t-1} + \varepsilon_t,$$

то

$$\mu = M(y_t) = \delta/(1 - \alpha).$$

Поскольку среднее, дисперсия и ковариация не зависят от времени t, процесс $AR(1)$ является стационарным. Неравенство $|\alpha| < 1$ является *необходимым условием стационарности* процесса y_t.

Автокорреляционная функция равна $\rho(k) = \alpha^k$, поэтому степень тесноты корреляционной связи между членами последовательности экспоненциально убывает по мере их взаимного удаления друг от друга по времени.

Значения частной автокорреляционной функции $\rho_ч(k)$ равны нулю для всех лагов $k > 1$.

Важным примером является процесс, называемый *случайным блужданием*:

$$y_t = y_{t-1} + \varepsilon_t,$$

где ε_t — белый шум.

Этот процесс по виду похож на $AR(1)$ с $\alpha = 1$, однако существенно отличается по своим свойствам от стационарного процесса $AR(1)$ с $|\alpha| < 1$.

Учитывая, что ошибка ε_t некоррелирована с y_{t-1}, можно получить

$$M(y_t) = M(y_{t-1});\ D(y_t) = D(y_{t-1}) + \sigma^2,$$

т. е. случайное блуждание *нестационарно*, так как $D(y_t) \neq D(y_{t-1})$. При этом дисперсия может неограниченно возрастать со временем.

Тест Дики — Фуллера. Тест Дики — Фуллера (DF-тест) используется для проверки временных рядов на стационарность.

Рассмотрим $AR(1)$ с нулевым средним:

$$y_t = \alpha y_{t-1} + \varepsilon_t,$$

где ε_t — белый шум.

Ряд y_t является стационарным, если $|\alpha| < 1$. Если $\alpha = 1$, то y_t — нестационарный временной ряд случайного блуждания. Как определить по имеющимся наблюдениям, верно ли, что в $AR(1)$-процессе $\alpha = 1$?

Вычтем y_{t-1} из обеих частей уравнения для y_t (взятие первой разности):

$$\Delta y_t = \beta y_{t-1} + \varepsilon_t ,$$

где $\Delta y_t = y_t - y_{t-1}$, $\beta = \alpha - 1$.

Дики и Фуллер рассмотрели три регрессии:

$$\Delta y_t = \beta y_{t-1} + \varepsilon_t , \qquad (11.1)$$

$$\Delta y_t = \alpha_0 + \beta y_{t-1} + \varepsilon_t , \qquad (11.2)$$

$$\Delta y_t = \alpha_0 + \beta y_{t-1} + \alpha_2 t + \varepsilon_t . \qquad (11.3)$$

Регрессия (11.2) содержит константу α_0, и если $|\alpha| < 1$, константа соответствует среднему уровню ряда $\alpha_0 = (1 - \alpha)M(y_t)$.

Регрессия (11.3) содержит, кроме константы, линейный детерминированный тренд $\alpha_2 t$. Нестационарность в этом случае устраняется включением переменной времени в модель. Если $\alpha_2 \neq 0$ и $|\alpha| < 1$, то временной ряд, описываемый моделью (11.3), называется стационарным относительно детерминированного тренда.

Во всех трех регрессиях нас интересует параметр β.

Нулевая гипотеза H_0: $\beta = 0$ (временной ряд нестационарен) против альтернативной H_1: $\beta < 0$ (временной ряд стационарен).

Тест Дики — Фуллера состоит в следующем. Уравнение (11.1) (*или* (11.2), (11.3)) оценивается методом наименьших квадратов. Получают оценку $\hat{y}$, стандартную ошибку и соответствующее значение t-статистики.

Сравнивая значение t-статистики с критическим (табличным), имеющим нестандартное распределение и зависящим от формы регрессии и объема выборки, определяют, принять или отклонить H_0. Нулевая гипотеза будет отвергнута, если наблюдаемое значение t-статистики меньше критического значения $t_{кр}$, взятого из таблиц Дики — Фуллера, и в этом случае ряд может быть отнесен к стационарным. Если нулевая гипотеза не отвергается, то исследуемый ряд является нестационарным.

В приложении 2 приведены односторонние критические значения статистики Дики — Фуллера (DF).

Критические значения не изменятся, если в правые части регрессий (11.1–3) добавить слагаемые вида Δy_{t-1}, Δy_{t-2}, ... Например:

$$\Delta y_t = \beta y_{t-1} + \alpha_1 \Delta y_{t-1} + \varepsilon_t, \qquad (11.4)$$

$$\Delta y_t = \alpha_0 + \beta y_{t-1} + \alpha_1 \Delta y_{t-1} + \varepsilon_t, \qquad (11.5)$$

$$\Delta y_t = \alpha_0 + \beta y_{t-1} + \alpha_1 \Delta y_{t-1} + \alpha_2 t + \varepsilon_t. \qquad (11.6)$$

Тест, соответствующий уравнению с лагированными значениями приращений Δy_{t-1}, Δy_{t-2}, ... в правой части, называется *расширенным тестом Дики — Фуллера* (*ADF*).

Процедура тестирования аналогична простому тесту Дики — Фуллера: оценивается значение t-статистики Стьюдента для параметра β, критические значения для *ADF*-теста те же самые, что и для обычного *DF*-теста.

Тест Дики — Фуллера включен во все современные эконометрические пакеты.

Пример 11.11. По данным примера 11.3 об объеме потребления электроэнергии y_t сделаем на 5%-м уровне прогноз на четыре квартала вперед с использованием авторегрессионной модели.

▼Исходные данные и расчетные показатели представим в табл. 11.23.

Таблица 11.23

Расчетная таблица для примера 11.11

t	y_t	Δy_t	y_{t-1}	Δy_{t-1}
1	60	—	—	—
2	44	–16	60	—
3	50	6	44	–16
4	90	40	50	6
5	72	–18	90	40
6	48	–24	72	–18
7	60	12	48	–24
8	100	40	60	12
9	80	–20	100	40
10	56	–24	80	–20
11	64	8	56	–24
12	110	46	64	8

t	y_t	Δy_t	y_{t-1}	Δy_{t-1}
13	90	–20	110	46
14	66	–24	90	–20
15	70	4	66	–24
16	108	38	70	4

Рассмотрим два уравнения: с константой (11.5) и с константой и трендом (11.6).

Оцененное уравнение (11.5):

$$\Delta y_t = 102,78 - \underset{(-4,37)}{1,381 y_{t-1}} + 0,620 \Delta y_{t-1}, \ R^2 = \ 0,635,$$

где в скобках указана t-статистика коэффициента β.

Критическое значение $t_{кр} = -3,0$; $t_р = -4,37 < t_{кр} = -3,0$.

Оцененное уравнение (11.6):

$$\Delta y_t = 105,63 - \underset{(-10,32)}{1,966 y_{t-1}} + 0,949 \Delta y_{t-1} + 4,073t, \ R^2 = \ 0,915,$$

где в скобках указана t-статистика коэффициента β.

Критическое значение $t_{кр} = -3,60$; $t_р = -10,32 < t_{кр} = -3,60$.

В обоих случаях наблюдаемое значение t-статистики меньше критического $t_{кр}$, взятого из таблиц Дики — Фуллера, и в этом случае ряд может быть отнесен к стационарным.

Из двух построенных моделей более точной является вторая (с константой и трендом). Вычислим по этой модели прогнозные значения объема потребления электроэнергии на четыре следующих квартала:

$\Delta y_{17} = 105,63 - 1,996 \cdot 108 + 0,949 \cdot 38 + 4,073 \cdot 17 = -1,41,$

$y_{17} = y_{16} + \Delta y_{17} = 108 - 1,41 = 106,59;$

$\Delta y_{18} = 105,63 - 1,996 \cdot 106,59 + 0,949 \cdot (-1,41) + 4,073 \cdot 18 = -31,95,$

$y_{18} = y_{17} + \Delta y_{18} = 106,59 - 31,95 = 74,64;$

$\Delta y_{19} = 105,63 - 1,996 \cdot 74,64 + 0,949 \cdot (-31,95) + 4,073 \cdot 19 = 5,97,$

$y_{19} = y_{18} + \Delta y_{19} = 74,64 + 5,97 = 80,61;$

$\Delta y_{20} = 105,63 - 1,996 \cdot 80,61 + 0,949 \cdot 5,97 + 4,073 \cdot 20 = 34,27$,

$y_{20} = y_{19} + \Delta y_{20} = 80,61 + 34,27 = 114,88$. ▲

Модель скользящего среднего *MA*(*q*). При моделировании стационарных временных рядов широкое распространение получили модели скользящего среднего, когда y_t линейно зависит от конечного числа q предыдущих значений ε_t:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q},$$

где ε_t — белый шум.

Модель скользящего среднего порядка q обозначается $MA(q)$. Согласно определению процесс $MA(q)$ стационарен при любом q и любых θ_i.

Имеется «двойственность» между процессом скользящего среднего и процессом авторегрессии. Это означает, что приведенное выше уравнение скользящего среднего можно переписать (обратить) в форму авторегрессии (неограниченного порядка), и наоборот.

Модель авторегрессии и скользящего среднего *ARMA*. Для описания стационарных процессов также может использоваться модель авторегрессии и скользящего среднего порядка (p, q), или модель $ARMA(p, q)$, включающая как члены, описывающие авторегрессионные составляющие, так и члены, моделирующие остаток в виде процесса скользящих средних.

Модель $ARMA(p, q)$ имеет вид

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q},$$

где ε_t — белый шум.

Обычно число параметров p или q не бывает больше 2.

Для процессов $ARMA(p, q)$ сформулированы следующие практические рекомендации по их идентификации:

• $ARMA(1, 0)$: АКФ экспоненциально убывает, ЧАКФ имеет выброс на лаге 1, нет корреляции на других лагах;

• $ARMA(2, 0)$: АКФ имеет форму синусоиды или экспоненциально убывает, ЧАКФ имеет выбросы на лагах 1 и 2, нет корреляции на других лагах;

• $ARMA(0, 1)$: АКФ имеет выброс на лаге 1, нет корреляции на других лагах, ЧАКФ экспоненциально убывает;

• *ARMA*(0, 2): АКФ имеет выбросы на лагах 1 и 2, нет корреляции на других лагах, ЧАКФ имеет форму синусоиды или экспоненциально затухает;

• *ARMA*(1, 1): АКФ экспоненциально убывает с лага 1, ЧАКФ экспоненциально убывает с лага 1.

***ARIMA*-модели.** Некоторые нестационарные временные ряды могут быть сведены к стационарным с помощью операции взятия разности. Такую процедуру называют *интегрированием*.

Обычно необходимо брать разности ряда до тех пор, пока он не станет стационарным (часто также применяют логарифмическое преобразование для стабилизации дисперсии). Число разностей, которые были взяты, чтобы достичь стационарности, определяется параметром d.

Пусть временной ряд y_t после взятия разности d раз стал стационарным, удовлетворяющим *ARMA*(p, q)-модели. В этом случае ряд y_t принято называть интегрированным рядом авторегрессии и скользящего среднего (АРПСС) или *ARIMA*(p, d, q). В специальной литературе она также известна как модель Бокса — Дженкинса.

Методология Бокса — Дженкинса подбора *ARIMA*-модели для описания и прогнозирования временного ряда включает следующие этапы:

• идентификация модели;

• оценивание модели и проверка ее адекватности;

• прогнозирование.

В работе [13] подробно изложены прикладные процедуры обработки данных в пакете *STATISTICA*, в том числе и подбор *ARIMA*-модели.

Пример 11.12. Проведем подбор *ARIMA*-модели по данным о размере золотовалютных резервов (y_t) России с 31.12.05 по 12.10.07 и сделаем прогноз на 5 шагов вперед.

▼Исходные данные и расчетные показатели приведены в табл. 11.24.

1. Идентификация модели. Первый шаг идентификации — получение стационарного ряда. Исходный ряд y_t не является стационарным, поскольку имеет возрастающий тренд (рис. 11.9).

Чтобы ряд стал стационарным, необходимо брать последовательные разности до тех пор, пока он не станет стационарным.

Таблица 11.24

Расчетная таблица для примера 11.12

Дата	t	y_t	$z_t = \Delta y_t$	z_{t-1}	e_t
31.12.05	1	182,2	—	—	—
13.01.06	2	184,6	2,4	—	—
20.01.06	3	185,2	0,6	2,4	–3,9718
03.02.06	4	188,5	3,3	0,6	–0,6879
17.02.06	5	195,4	6,9	3,3	2,0362
03.03.06	5	197,9	2,5	6,9	–3,5317
17.03.06	7	204,1	6,2	2,5	1,5957
31.03.06	8	205,9	1,8	6,2	–4,0046
14.04.06	9	212,0	6,1	1,8	1,7228
21.04.06	10	217,1	5,1	6,1	–0,6722
05.05.06	11	231,1	14	5,1	8,5522
26.05.06	12	243,3	12,2	14	3,8648
09.06.06	13	247,9	4,6	12,2	–3,1512
23.06.06	14	247,2	–0,7	4,6	–5,9856
07.07.06	15	253,2	6	–0,7	2,4339
28.07.06	16	265,6	12,4	6	6,6602
11.08.06	17	277,0	11,4	12,4	3,5839
25.08.06	18	258,5	–18,5	11,4	–25,9917
08.09.06	19	260,7	2,2	–18,5	4,4087
29.09.06	20	266,6	5,9	2,2	1,3931
13.10.06	21	266,5	–0,1	5,9	–5,8073
27.10.06	22	269,1	2,6	–0,1	–1,1608
10.11.06	23	277,0	7,9	2,6	3,2633
24.11.06	24	283,4	6,4	7,9	0,0438
08.12.06	25	293,8	10,4	6,4	4,5305
29.12.06	26	303,0	9,2	10,4	2,0328
12.01.07	27	301,7	–1,3	9,2	–8,0779
26.01.07	28	303,8	2,1	–1,3	–1,2714
09.02.07	29	309,5	5,7	2,1	1,2255
23.02.07	30	311,1	1,6	5,7	–4,0424
09.03.07	31	317,3	6,2	1,6	1,8877
30.03.07	32	338,7	21,4	6,2	15,5954

Дата	t	y_t	$z_t = \Delta y_t$	z_{t-1}	e_t
13.04.07	33	356,6	17,9	21,4	7,1640
27.04.07	34	369,0	12,4	17,9	2,7995
11.05.07	35	386,3	17,3	12,4	9,4839
25.05.07	36	402,2	15,9	17,3	6,4942
08.06.07	37	406,5	4,3	15,9	–4,6516
22.06.07	38	406,6	0,1	4,3	–5,0882
06.07.07	39	408,4	1,8	0,1	–2,0256
20.07.07	40	413,1	4,7	1,8	0,3228
03.08.07	41	416,8	3,7	4,7	–1,6180
17.08.07	42	414,7	–2,1	3,7	–7,0936
31.08.07	43	416,0	1,3	–2,1	–1,8119
14.09.07	44	420,9	4,9	1,3	0,6850
28.09.07	45	425,1	4,2	4,9	–1,1829
05.10.07	46	424,8	–0,3	4,2	–5,4558
12.10.07	47	434,0	9,2	–0,3	5,5041

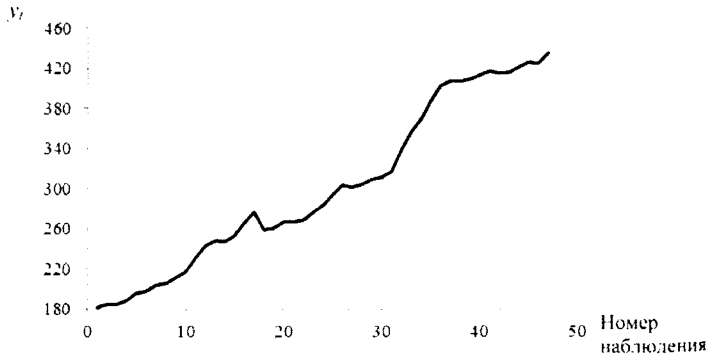


Рис. 11.9. График динамики переменной y_t для примера 11.12

Для определения порядка разности нужно исследовать автокоррелограмму. Если наблюдается медленное убывание выборочных коэффициентов автокорреляции в зависимости от лага, обычно берут разность первого порядка.

На рис. 11.10 показана АКФ переменной y_t, где коэффициенты выборочной АКФ вычисляются по формуле

$$r_k = \frac{\sum_{t=k+1}^{n}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2}, k = 1, 2, ...$$

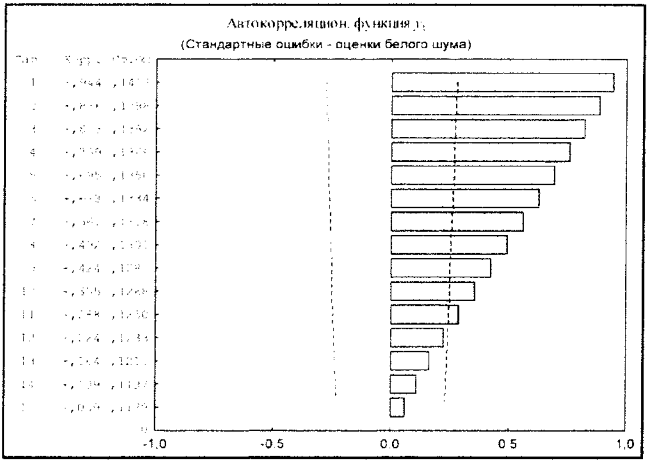


Рис. 11.10. Автокоррелограмма переменной y_t для примера 11.12

Из рис. 11.10 видно, что автокорреляции в зависимости от лага убывают медленно, что говорит о том, что для идентификации модели $ARIMA(p, d, q)$ можно взять разности первого порядка ($d = 1$).

Найдем первую разность $z_t = \Delta y_t$, где $\Delta y_t = y_t - y_{t-1}$ и построим ее график в зависимости от номера наблюдений (рис. 11.11), из которого видно, что ряд стал стационарным, так как тренд отсутствует.

Для стационарного ряда z_t исследуется характер поведения выборочных АКФ и ЧАКФ, которые позволяют сформулировать несколько гипотез о возможных порядках авторегрессии (p) и скользящего среднего (q).

Коэффициенты выборочной АКФ для ряда z_t вычисляются по формуле

$$r_k = \frac{\sum_{t=k+1}^{n}(z_t - \bar{z})(z_{t-k} - \bar{z})}{\sum_{t=1}^{n}(z_t - \bar{z})^2}, k = 1, 2, ...$$

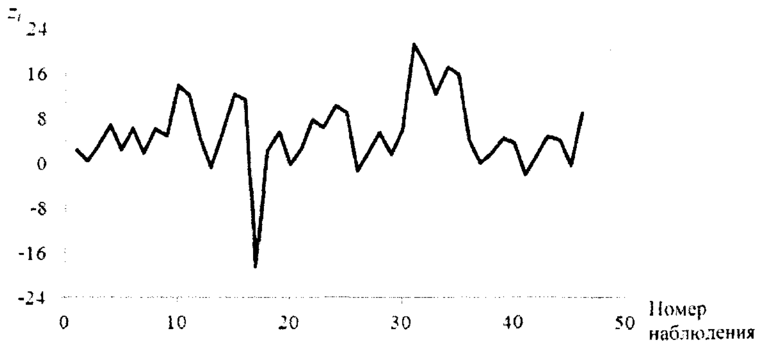


Рис. 11.11. График динамики первой разности z_t для примера 11.12

Для стационарного ряда z_t значение выборочной ЧАКФ вычисляется как МНК-оценка последнего коэффициента β_k в регрессионном уравнении $z_t = \beta_0 + \beta_1 z_{t-1} + \ldots + \beta_k z_{t-k} + \varepsilon_t$.

На рис. 11.12 приведены автокорреляционная и частная автокорреляционная функции переменной z_t .

На рис. 11.12 АКФ имеет небольшой выброс на первом лаге и заметную тенденцию к затуханию, в ЧАКФ значимо отличается от нуля только значение корреляции для первого лага.

В соответствии с указанными ранее практическими рекомендациями по идентификации модели *ARMA* выбираем модель *ARIMA*(1, 1, 0), но можно также использовать и модель *ARIMA*(0, 1, 1).

2. Оценивание ARMA-моделей производится различными методами (линейный и нелинейный МНК, полный или условный метод максимального правдоподобия).

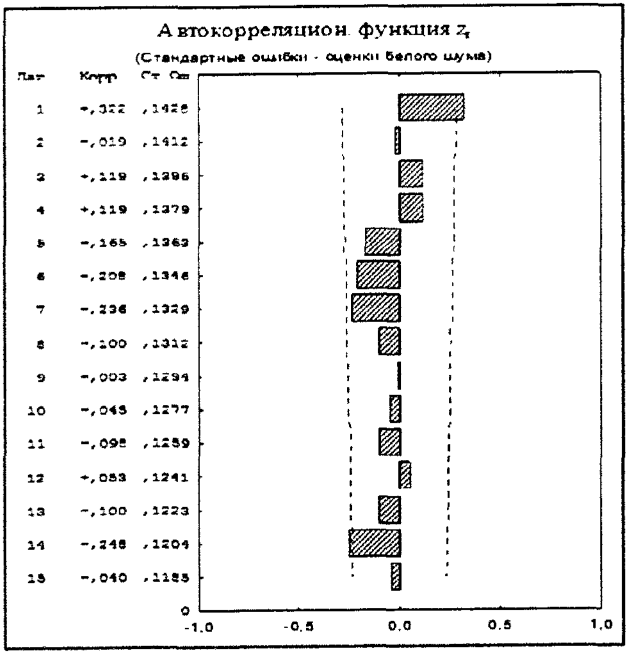
Рассмотрим модель *ARIMA*(1, 1, 0). Оценим модель авторегрессии первого порядка со свободным членом $z_t = \delta + \alpha z_{t-1} + \varepsilon_t$ методом МНК.

В табл. 11.24 приведены расчетные показатели, необходимые для оценивания параметров уравнения в *Excel*.

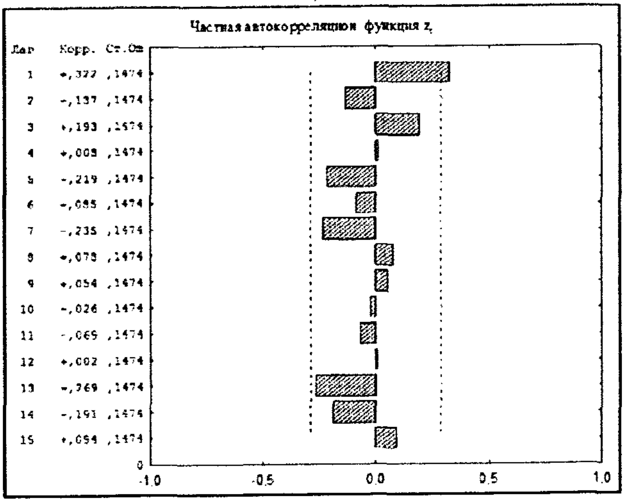
Оцененная статистически значимая модель есть

$$\tilde{z}_t = 3,793 + 0,324 z_{t-1},$$

где $\delta = 3,793$; $\alpha = 0,324$, а остаточная дисперсия (остаток) равна 39,8.



а)



б)

Рис. 11.12. Автокорреляционная (*а*) и частная автокорреляционная (*б*) функции переменной z_t для примера 11.12

Коэффициенты модели статистически значимые.

Запишем преобразованную модель в виде

$$z_t = 5{,}615 + 0{,}324 z_{t-1} + \varepsilon_t,$$

где $5{,}615 = \mu = \delta/(1 - \alpha)$.

Если имеется несколько моделей, успешно прошедших проверку условия адекватности, то выбираем ту модель, у которой дисперсия остатков минимальна.

Для проверки адекватности *ARMA*-моделей существуют различные критерии:

1) оценки коэффициентов модели должны статистически значимо отличаться от нуля;

2) остатки модели ε_t должны быть похожи на белый шум, т. е. иметь нулевую автокорреляцию.

Проверим адекватность модели *ARIMA*(1, 1, 0).

Коэффициенты $\mu = 5{,}615$ и $\alpha = 0{,}324$ статистически значимы (первое условие по проверке адекватности модели выполнено).

При проверке значимости коэффициентов АКФ остатков используются два подхода:

• проверка значимости каждого коэффициента автокорреляции отдельно;

• проверка значимости группы коэффициентов автокорреляции с помощью теста Бокса — Льюнга.

Для проверки выполнения второго условия рассмотрим табл. 11.25, которую можно получить расчетным способом на основании остатков e_t модели *ARIMA*(1, 1, 0) из табл. 11.24.

Таблица 11.25

Таблица результатов автокорреляционной функции остатков модели *ARIMA*(1, 1, 0) для примера 11.12 (стандартные ошибки — ошибки белого шума)

Лаг	Коэффициент автокорреляции	Стандартная ошибка	Статистика Бокса — Льюнга (Q)	Уровень значимости (P)
1	0,039	0,1428	0,07	0,785
2	–0,189	0,1412	1,87	0,392

Лаг	Коэффициент автокорреляции	Стандартная ошибка	Статистика Бокса — Льюнга (Q)	Уровень значимости (P)
3	0,113	0,1396	2,53	0,469
4	0,158	0,1379	3,85	0,426
5	–0,171	0,1363	5,42	0,367
6	–0,125	0,1346	6,29	0,392
7	–0,179	0,1329	8,11	0,323
8	–0,039	0,1312	8,20	0,415
9	0,063	0,1294	8,43	0,491
10	–0,016	0,1277	8,45	0,585
11	–0,134	0,1259	9,58	0,568
12	0,131	0,1241	10,69	0,556
13	–0,042	0,1223	10,81	0,627
14	–0,243	0,1204	14,88	0,386
15	0,045	0,1185	15,02	0,450

Автокорреляция — это корреляция исходного ряда с самим собой, сдвинутым на определенный лаг k. Коэффициенты выборочной автокорреляционной функции остатков определяются по формуле

$$r_k = \frac{\sum_{t=k+1}^{n} e_t e_{t-k}}{\sum_{t=1}^{n} e_t^2}, \; k = 1, 2, \ldots$$

В предположении, что процесс — белый шум (в этом процессе все коэффициенты автокорреляции равны нулю), стандартные ошибки r_k определяются как

$$\text{Стандартная ошибка } (r_k) = \sqrt{(1/n)\cdot(n-k)/(n+2)},$$

где n — число наблюдений ряда.

Из сопоставления полученных значений, представленных в табл. 11.25, следует, что коэффициенты автокорреляции незначимы на всех 15 лагах.

Для проверки равенства нулю k первых значений автокорреляционной функции остатков используется Q-статистика Бокса — Льюнга.

На данном лаге k статистика Бокса — Льюнга Q определяется как

$$Q_k = n(n+2)\sum_{i=1}^{k} r_i^2 \, / \, (n-i) \, .$$

При выполнении нулевой гипотезы отсутствия автокорреляции Q-статистика имеет распределение $\chi^2(k-p-q)$.

Уровни значимости P_k, соответствующие статистике Q_k, можно определить с помощью функции *Excel* = ХИ2РАСП(Q_k, k). Если P_k больше заданного уровня значимости, то k первых значений автокорреляционной функции остатков статистически незначимы.

Из рассмотрения полученных значений последнего столбца табл. 11.25 следует, что все k первых значений автокорреляционной функции остатков статистически незначимы.

В табл. 11.26 показан пример расчета значений Q_k, P_k для лагов k = 1, 2, 3 по приведенным формулам, n = 46.

Таблица 11.26

Расчет значений статистики Бокса — Льюнга и соответствующих уровней значимости

k	r_k	Q_k	P_k
1	0,039	$Q_1 = 46 \cdot 48 \cdot 0{,}039^2 / 45 = 0{,}075$	ХИ2РАСП(0,075;1) = = 0,785
2	–0,189	$Q_2 = Q_1 + 46 \cdot 48 \cdot (-0{,}189)^2 / 44 = 1{,}875$	ХИ2РАСП(1,875;2) = = 0,392
3	0,113	$Q_3 = Q_2 + 46 \cdot 48 \cdot 0{,}113^2 / 43 = 2{,}535$	ХИ2РАСП(2,535;3) = = 0,469

Таким образом, второе условие по проверке адекватности модели выполнено.

3. Прогнозирование в модели ARIMA(1, 1, 0). Рассмотрим нестационарный временной ряд y_t, первые разности которого z_t являются $AR(1)$-процессом:

$$z_t = y_t - y_{t-1},\ z_t - \mu = \alpha(z_{t-1} - \mu) + \varepsilon_t.$$

Многократное применение этих выражений дает следующую рекуррентную формулу прогнозирования:

$$\hat{y}_{t+1} = y_t + \mu + \alpha(y_t - y_{t-1} - \mu).$$

Произведем прогноз на пять шагов. Для двух последних наблюдений имеем $y_{46} = 424{,}8$ и $y_{47} = 434{,}0$.

Прогноз на один шаг:

$\hat{y}_{48} = y_{47} + \mu + \alpha(y_{47} - y_{46} - \mu) = 434{,}0 + 5{,}615 + 0{,}324 \cdot (434{,}0 - 424{,}8 - 5{,}615) = 440{,}8.$

Прогноз на два шага:

$\hat{y}_{49} = y_{48} + \mu + \alpha(y_{48} - y_{47} - \mu) = 440{,}8 + 5{,}615 + 0{,}324 \cdot (440{,}8 - 434{,}0 - 5{,}615) = 446{,}8.$

Прогноз на три шага:

$\hat{y}_{50} = y_{49} + \mu + \alpha(y_{49} - y_{48} - \mu) = 446{,}8 + 5{,}615 + 0{,}324 \cdot (446{,}8 - 440{,}8 - 5{,}615) = 452{,}5.$

Прогноз на четыре шага:

$\hat{y}_{51} = y_{50} + \mu + \alpha(y_{50} - y_{49} - \mu) = 452{,}5 + 5{,}615 + 0{,}324 \cdot (452{,}5 - 446{,}8 - 5{,}615) = 458{,}2.$

Прогноз на пять шагов:

$\hat{y}_{52} = y_{51} + \mu + \alpha(y_{51} - y_{50} - \mu) = 458{,}2 + 5{,}615 + 0{,}324 \cdot (458{,}2 - 452{,}5 - 5{,}615) = 463{,}8.$ ▲

Сезонные модели *ARIMA*. Сезонная модель представляется в виде: $ARIMA(p, d, q)(P, D, Q)_s$, где к параметрам модели p, d, q добавлены сезонные параметры P, D, Q и s — сезонная авторегрессия, сезонная разность, сезонное скользящее среднее и сезонный период соответственно.

Идентификация сезонной модели производится тем же способом, что и идентификация несезонной модели. Поведение автокорреляционной и частной автокорреляционной функций на начальных лагах позволяет идентифицировать стандартным образом несезонную компоненту, а на лагах, кратных сезонному лагу, — сезонную составляющую.

При наличии ярко выраженной сезонной компоненты целесообразно включение в модель сезонного дифференцирования, но при этом желательно, чтобы $d + D \leq 2$.

Существенно облегчить решение задач анализа и прогнозирования финансово-экономических показателей поможет использование современных компьютерных статистических пакетов. В некоторых компьютерных пакетах реализованы процедуры автоматического подбора структуры модели Бокса — Дженкинса (АРПСС).

Процедура построения моделей временного ряда в программе *SPSS* включает в себя инструмент *Эксперт построения моделей*, который автоматически идентифицирует и оценивает наиболее подходящую модель Бокса — Дженкинса или экспоненциального сглаживания, исключая необходимость определения подходящей модели методом проб и ошибок.

Пример 11.13. Используя пакет *SPSS*, проведем подбор *ARIMA*-модели по данным примера 11.6 об объеме пассажирских авиаперевозок за шесть лет и сделаем прогноз на следующий год.

▼Укажем последовательность действий.

• Вводим данные примера в таблицу в один столбец с именем «Авиаперевозки» (рис. 11.13).

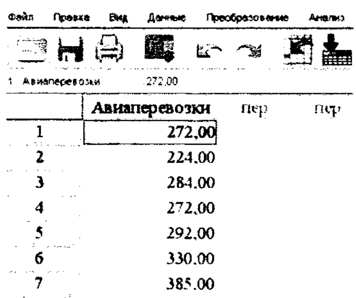


Рис. 11.13. Ввод исходных данных в *SPSS* для примера 11.13

• В верхнем меню выбираем команды *Данные* → *Задать даты*. Откроется диалоговое окно (рис. 11.14).

Задаем дату, связанную с первым наблюдением (например, январь 2010 г.), и интервал времени между последовательными наблюдениями. Это приводит к набору переменных, отмечающих

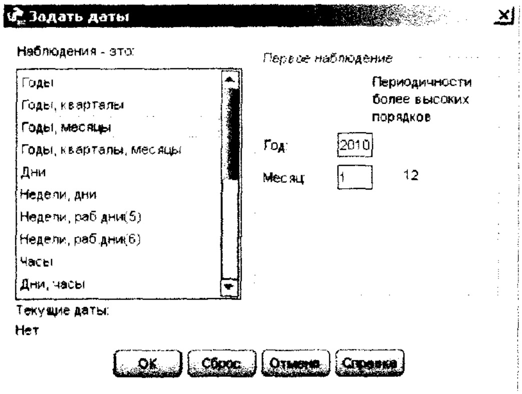
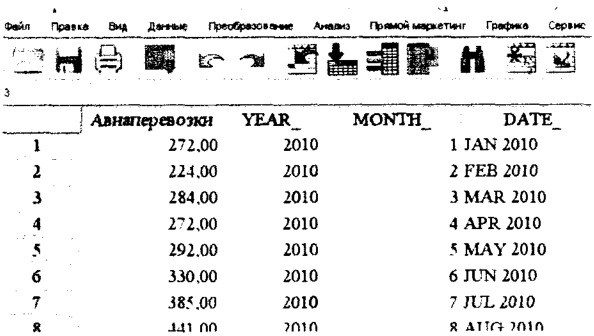


Рис. 11.14. Диалоговое окно *Задать даты* (пример 11.13)

даты, связанные с каждым наблюдением. Этим же задается предполагаемая периодичность данных, например, периодичность 12, если временной интервал между последовательными наблюдениями равен одному месяцу. Эта периодичность необходима, если нужно создать сезонные модели. Если сезонные модели не нужны и метки данных в выводе не требуются, то диалоговое окно *Задать даты* можно пропустить. В этом случае меткой, связанной с каждым наблюдением, будет просто номер наблюдения.

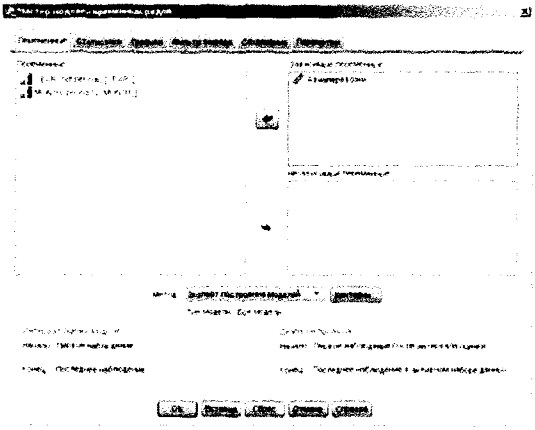
• Щелкнув на кнопку *OK*, перейдем в таблицу данных, в которой добавились новые переменные YEAR, MONTH, DATE (рис. 11.15).



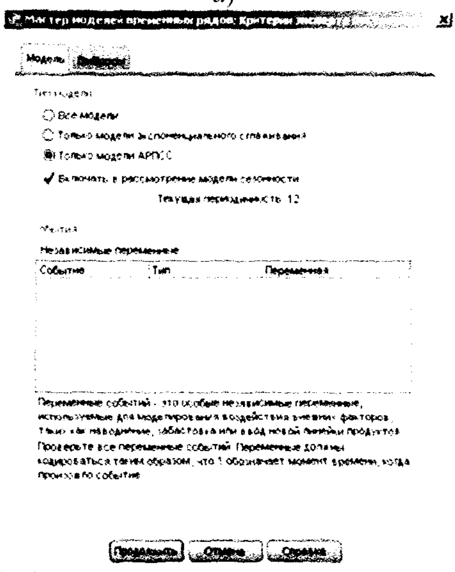
	Авиаперевозки	YEAR_	MONTH_	DATE_
1	272,00	2010		1 JAN 2010
2	224,00	2010		2 FEB 2010
3	284,00	2010		3 MAR 2010
4	272,00	2010		4 APR 2010
5	292,00	2010		5 MAY 2010
6	330,00	2010		6 JUN 2010
7	385,00	2010		7 JUL 2010
8	441,00	2010		8 AUG 2010

Рис. 11.15. Вид таблицы данных после указания даты первого наблюдения (пример 11.13)

• В верхнем меню выбираем команды *Анализ → Прогнозирование → Создать модели*. Откроется диалоговое окно *Мастер моделей временных рядов* (рис. 11.16, *а*).



а)



б)

Рис. 11.16. Вкладка *Переменные* диалогового окна *Мастер моделей временных рядов* (*а*) и задание критериев для метода *Эксперт построения моделей* (*б*)

• Выделим переменную «Авиаперевозки» и при помощи кнопки ➝ перенесем ее в список *Зависимые переменные*. В качестве метода в группе *Метод* устанавливаем *Эксперт построения моделей* и нажмем на кнопку *Критерии*. Откроется диалоговое окно *Мастер моделей временных рядов: Критерии эксперта построения...* (рис. 11.16, *б*).

• Установите флажки, как указанно на рис. 11.16, *б*, и щелкните на кнопке *Продолжить*, чтобы вернуться в диалоговое окно *Мастер моделей временных рядов* (рис. 11.16, *а*).

• Последовательно щелкаем по вкладкам *Статистика*, *Графики*, *Сохранение*, *Параметры* и устанавливаем значения, показанные на рис. 11.17.

• Нажимаем кнопку *ОК* в диалоговом окне *Мастер моделей временных рядов* и получаем результаты.

В табл. 11.27 приведены результаты оценивания параметров модели методом *Эксперт построения моделей*.

Идентификация модели: *ARIMA*(1,1,0)(0,1,1)12 (без свободного параметра). Имело место логарифмическое преобразование исходной переменной, дифференцирование исходного ряда с лагом 1 и сезонное дифференцирование с лагом 12.

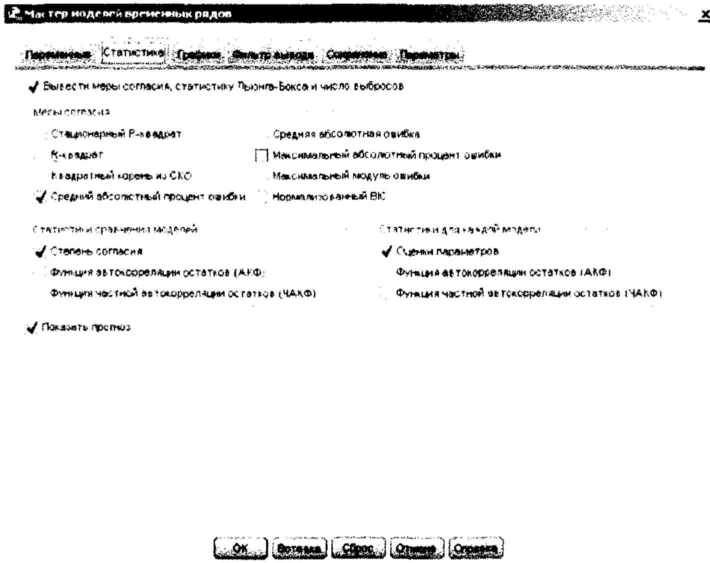
Таблица 11.27

Результаты оценивания параметров модели методом *Эксперт построения моделей* для примера 11.13

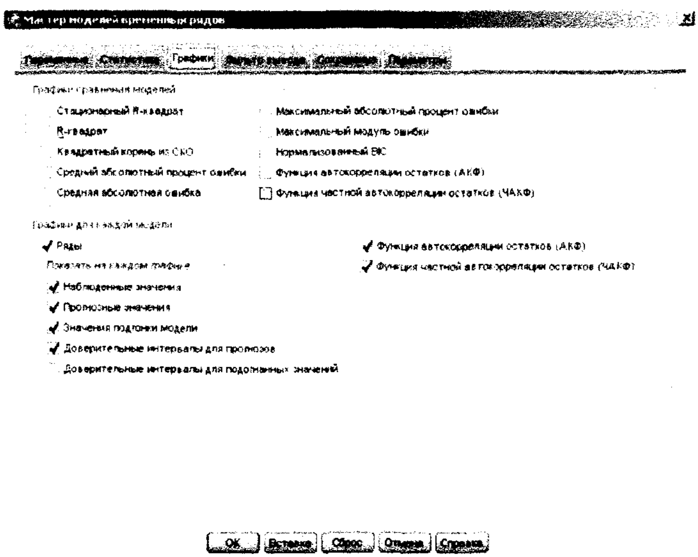
Параметр	Оценка	Стандартная ошибка	*t*	Значение
p(1)	–0,344	0,127	–2,689	0,009
Qs(1)	0,409	0,149	2,738	0,008

Данная модель содержит коэффициент авторегрессии *p*(1) для учета линейного тренда в динамике объема авиаперевозок y_t и коэффициент сезонного скользящего среднего *Qs*(1). Приведенные в таблице параметры модели высокозначимы. Ошибка подгонки $\overline{e}$ = 4,09 %.

В табл. 11.28 приведены результаты прогноза объема авиаперевозок на 12 месяцев вперед и доверительные границы прогнозных значений.

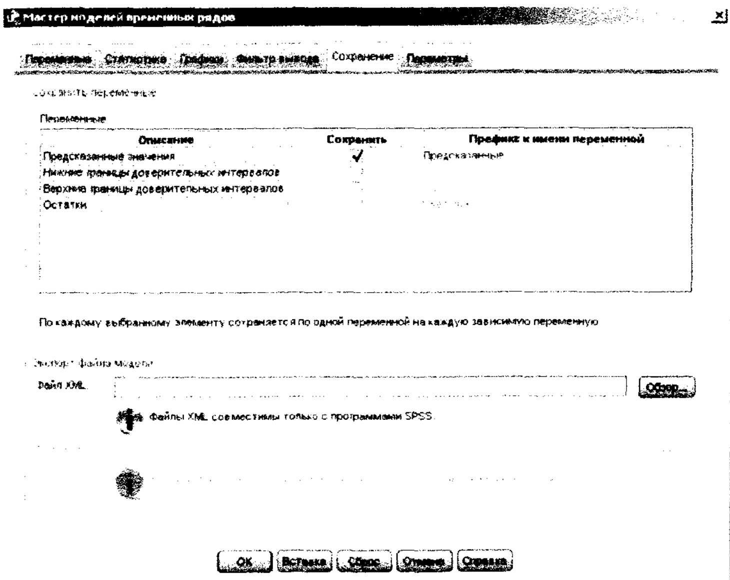


а)

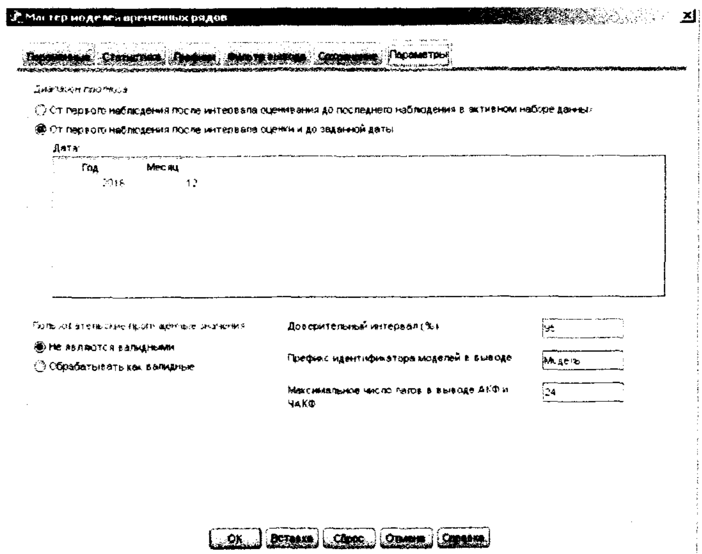


б)

Рис. 11.17. Вкладки *Статистика* (*а*), *Графики* (*б*) диалогового окна *Мастер моделей временных рядов*



в)



г)

Рис. 11.17. Вкладки *Сохранение* (*в*), *Параметры* (*г*) диалогового окна *Мастер моделей временных рядов*

Таблица 11.28

Результаты прогноза на 12 месяцев вперед и доверительные границы прогнозных значений для примера 11.13

Дата	Прогноз	Нижний 95 %	Верхний 95 %
Январь 2016	434,68	390,70	482,29
Февраль 2016	385,40	339,10	436,29
Март 2016	444,06	381,10	514,53
Апрель 2016	481,64	405,37	568,21
Май 2016	518,94	428,86	625,53
Июнь 2016	576,55	468,65	702,09
Июль 2016	663,53	531,08	819,35
Август 2016	691,96	545,86	865,61
Сентябрь 2016	581,63	452,57	736,53
Октябрь 2016	509,65	391,40	652,87
Ноябрь 2016	458,30	347,59	593,57
Декабрь 2016	475,08	356,01	621,79

На рис. 11.18 приведен график динамики переменной y_t (объем авиаперевозок) и прогноз с доверительным интервалом на 12 месяцев вперед.

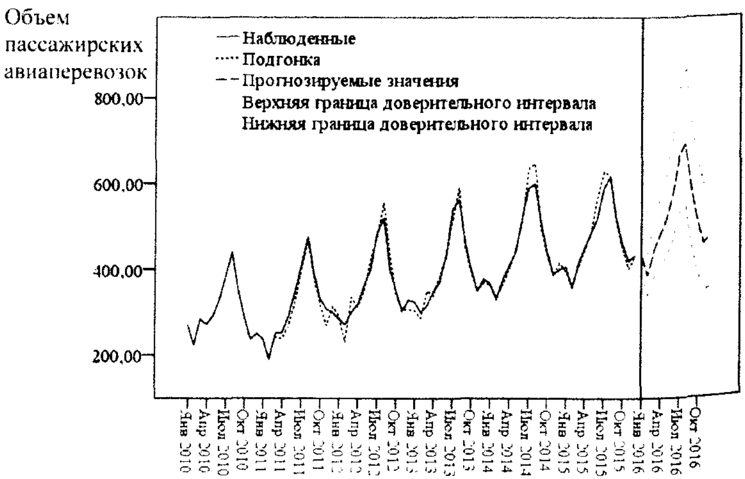


Рис. 11.18. График динамики переменной y_t и прогноз с доверительным интервалом на 12 месяцев вперед для примера 11.13

Статистического различия в значениях прогноза примера 11.6 (модель Тейла — Вейджа) и полученных данным методом нет, но для данного примера предпочтительней является модель Тейла — Вейджа, поскольку для нее ошибка подгонки $\bar{e} = 3{,}65$ % меньше. ▲

Глава 12. СИСТЕМА ЭКОНОМЕТРИЧЕСКИХ УРАВНЕНИЙ

12.1. Общая характеристика системы эконометрических уравнений

Системный подход в экономическом анализе предполагает сложную структуру взаимосвязей между признаками, когда эффективность деятельности экономического объекта характеризуется несколькими показателями. В этом случае одного регрессионного уравнения может быть недостаточно, и для описания явления или процесса может потребоваться система уравнений и тождеств.

Различают следующие виды систем эконометрических уравнений:

- системы независимых уравнений;
- системы рекурсивных уравнений;
- системы взаимозависимых уравнений.

В *системе независимых уравнений* каждая зависимая переменная y рассматривается как функция одного и того же набора факторов x:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1, \\ y_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m + \varepsilon_n. \end{cases}$$

Каждое уравнение такой системы может рассматриваться самостоятельно. Для нахождения его параметров используется метод наименьших квадратов.

В *системе рекурсивных уравнений* зависимая переменная y включает в каждое последующее уравнение в качестве факторов все зависимые переменные предшествующих уравнений и набор факторов x:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2, \\ y_3 = b_{31}y_1 + b_{32}y_2 + a_{31}x_1 + a_{32}x_2 + \ldots + a_{3m}x_m + \varepsilon_3, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = b_{n1}y_1 + b_{n2}y_2 + \ldots + b_{nn-1}y_{n-1} + a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m + \varepsilon_n. \end{cases}$$

В таких моделях уравнения оцениваются последовательно (от первого уравнения к последнему) с использованием МНК.

В *системе взаимозависимых уравнений* одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других — в правую часть системы:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + \ldots + b_{1n}y_n + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1, \\ y_2 = b_{21}y_1 + b_{23}y_3 + \ldots + b_{2n}y_n + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = b_{n1}y_1 + b_{n2}y_2 + \ldots + b_{nn-1}y_{n-1} + a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m + \varepsilon_n. \end{cases}$$

Систему взаимозависимых уравнений также называют системой одновременных уравнений.

12.2. Структурная и приведенная формы уравнений

Система одновременных уравнений в эконометрике получила название структурной формы модели.

Структурной формой модели называется система уравнений, в каждом из которых аргументы помимо объясняющих переменных могут включать в себя также объясняемые переменные из других уравнений системы.

Уравнения, составляющие исходную модель, называются *структурными уравнениями модели.*

Простейшая структурная форма модели имеет вид

$$\begin{cases} y_1 = \alpha_1 + \beta_{12} y_2 + \alpha_{11} x_1 + \alpha_{12} x_2 + \varepsilon_1, \\ y_2 = \alpha_2 + \beta_{21} y_1 + \alpha_{21} x_1 + \alpha_{22} x_2 + \varepsilon_2, \end{cases}$$

где y и x — зависимая и независимая переменные;

ε_1 и ε_2 — случайные члены;

α, β — параметры модели.

Параметры структурной формы модели называются *структурными коэффициентами.*

Структурная форма модели обычно включает в систему не только уравнения, отражающие взаимосвязи между отдельными переменными, но и уравнения, отражающие тенденцию развития явления, а также разного рода уравнения-тождества. Тождества не содержат каких-либо подлежащих оценке параметров, а также не включают случайного члена.

В процессе оценивания параметров одновременных уравнений следует различать эндогенные и экзогенные переменные. Приставки «эндо-» и «экзо-» означают соответственно внутреннее и внешнее.

Эндогенными считаются переменные, значения которых определяются внутри модели. Они являются зависимыми переменными.

Экзогенными считаются переменные, значения которых определяются вне модели. Они являются независимыми переменными.

В качестве экзогенных переменных могут рассматриваться значения эндогенных переменных за предшествующий период времени (лаговые переменные).

Предполагается, что в каждом уравнении экзогенные переменные некоррелированы со случайным членом. В то же время эндогенные переменные, стоящие в правых частях уравнения, как правило, коррелированы со случайным членом, что приводит к смещенности и несостоятельности МНК-оценок структурных коэффициентов.

Для определения структурных коэффициентов структурная форма модели преобразуется в приведенную форму.

Приведенной формой модели называется система уравнений, в каждом из которых эндогенные переменные выражены только через экзогенные переменные и случайные составляющие.

Например, приведенная форма исходной модели имеет вид

$$\begin{cases} y_1 = \alpha_1' + \alpha_{11}' x_1 + \alpha_{12}' x_2 + \nu_1, \\ y_2 = \alpha_2' + \alpha_{21}' x_1 + \alpha_{22}' x_2 + \nu_2, \end{cases}$$

где α' — параметры приведенной формы;

ν_1, ν_2 — случайные члены.

Параметры приведенной формы модели называются *коэффициентами приведенной формы* (приведенными коэффициентами).

Коэффициенты приведенной формы оцениваются обычным МНК, поскольку экзогенные переменные некоррелированы со случайным членом.

Оцененные коэффициенты приведенной формы могут быть использованы для оценивания структурных коэффициентов. Такой способ оценивания структурных коэффициентов называется *косвенным МНК* (КМНК).

Приведенная форма модели аналитически уступает структурной форме модели, так как в ней отсутствуют оценки взаимосвязи между эндогенными переменными.

При переходе от приведенной формы модели к структурной возникает проблема идентификации. **Идентификация** — это единственность соответствия между приведенной и структурной формами модели.

Тот или иной структурный коэффициент может либо однозначно выражаться через приведенные коэффициенты, либо иметь несколько разных оценок, либо совсем не выражаться через них.

Структурный коэффициент называется *идентифицируемым*, если его можно вычислить на основе приведенных коэффициентов, причем *точно идентифицируемым*, если он единственен, и *сверхидентифицируемым*, если имеет несколько разных оценок; в противном случае он называется *неидентифицируемым*.

Какое-либо структурное уравнение называется идентифицируемым, если идентифицируемы все его коэффициенты. Если хотя бы один структурный коэффициент неидентифицируем, то и все уравнение является неидентифицируемым.

Модель считается идентифицируемой, если каждое ее уравнение идентифицируемо. Если хотя бы одно из уравнений системы неидентифицируемо, то вся модель неидентифицируема.

12.3. Методы оценивания структурных уравнений

Рассмотрим различные виды структурных уравнений.

I. Точная идентифицируемость.

Допустим, требуется оценить параметры уравнения функции потребления в простой модели Кейнса формирования доходов:

$$\begin{cases} C_t = \alpha + \beta Y_t + \varepsilon_t, & \text{(функция потребления)} \\ Y_t = C_t + I_t, & \text{(тождество дохода)} \end{cases}$$

где C_t, Y_t, I_t — объем потребления, совокупный доход и инвестиции соответственно;

ε_t — случайный член.

Структурный коэффициент β характеризует предельную склонность к потреблению, т. е. из каждой единицы совокупного дохода расходуется β единиц на конечное потребление.

В данной модели C_t, Y_t — эндогенные переменные, а I_t — экзогенная.

Непосредственное оценивание параметров α, β в структурном уравнении функции потребления дает смещенные и несостоятельные оценки, так как объясняющая переменная Y_t является эндогенной.

Разрешая структурную систему относительно эндогенных переменных, получим приведенную систему

$$\begin{cases} C_t = \dfrac{\alpha}{1-\beta} + \dfrac{\beta}{1-\beta} I_t + \dfrac{\varepsilon_t}{1-\beta}, \\ Y_t = \dfrac{\alpha}{1-\beta} + \dfrac{1}{1-\beta} I_t + \dfrac{\varepsilon_t}{1-\beta}. \end{cases}$$

В приведенной системе коэффициенты при переменной I_t, равные $M_C = \beta/(1-\beta)$ и $M_Y = 1/(1-\beta)$, — *инвестиционные мультипликаторы потребления* и *дохода* соответственно. Это значит, что если объем инвестиций возрастает на единицу, то объем потребления увеличится на $\beta/(1-\beta)$ единиц, а совокупный доход — на $1/(1-\beta)$ единиц.

Косвенный метод наименьших квадратов. Уравнение для C_t в приведенной форме можно также представить в виде

$$C_t = \alpha' + \beta' I_t + \varepsilon'_t,$$

где $\alpha' = \dfrac{\alpha}{1-\beta}$; $\beta' = \dfrac{\beta}{1-\beta}$; $\varepsilon'_t = \dfrac{\varepsilon_t}{1-\beta}$.

В этом уравнении экзогенная переменная I_t некоррелирована со случайным членом ε'_t, поэтому для оценки параметров (α', β') можно использовать обычный МНК.

Оцененное уравнение

$$\hat{C}_t = \alpha' + \beta' I_t,$$

полученное по выборочным данным с помощью МНК, дает несмещенные и состоятельные оценки параметров.

Поскольку оценки (α, β) структурных коэффициентов

$$\alpha = \frac{\alpha'}{1+\beta'}, \quad \beta = \frac{\beta'}{1+\beta'}$$

однозначно выражаются через оценки (α', β') приведенных коэффициентов, то структурное уравнение функции потребления является *точно идентифицируемым.*

Таким образом, для решения точно идентифицируемого уравнения применяется косвенный МНК.

Процедура КМНК производится в несколько этапов:

1) структурная модель преобразуется в приведенную форму;

2) для каждого приведенного уравнения обычным МНК оцениваются приведенные коэффициенты;

3) оценки приведенных коэффициентов преобразуются в оценки параметров структурных уравнений.

Метод инструментальных переменных. Проблема коррелированности объясняющей переменной Y_t со случайным членом ε_t в структурном уравнении для C_t может быть разрешена с помощью метода инструментальных переменных (ИП).

Для применения метода ИП необходимо найти такую инструментальную переменную, которая обладает следующими свойствами:

1) коррелирует с неудачно объясняющей переменной Y_t;

2) не коррелирует со случайным членом ε_t.

В данном случае модель сама предоставляет такую переменную. Величина I_t коррелирует с Y_t, поскольку Y_t зависит от I_t в приведенном уравнении, и I_t не коррелирует с ε_t, поскольку является экзогенной переменной. В общем случае, когда оценка, полученная косвенным методом, единственна, она совпадает с оценкой, полученной методом ИП, т. е. КМНК можно рассматривать как частный случай метода ИП.

Структурное уравнение модели, в которой число экзогенных переменных, которые могут использоваться как инструментальные, равно числу объясняющих эндогенных переменных, является *точно идентифицируемым.*

Пример 12.1. Для некоторой страны имеются данные о совокупном доходе (Y_t), объеме потребления (C_t) и инвестициях (I_t), полученные за 10 лет (табл. 12.1).

Таблица 12.1

Исходные данные для примера 12.1, усл. ед.

C_t	190	198	200	180	200	210	220	210	205	210
I_t	10	20	30	20	10	20	30	20	15	30
Y_t	200	218	230	200	210	230	250	230	220	240

Построим функцию потребления, используя модель Кейнса формирования доходов:

$$\begin{cases} C_t = \alpha + \beta Y_t + \varepsilon_t, & \text{(функция потребления)} \\ Y_t = C_t + I_t. & \text{(тождество дохода)} \end{cases}$$

▼В данной модели C_t, Y_t — эндогенные переменные; I_t — экзогенная.

Первое уравнение является идентифицируемым, поскольку переменная I_t не включена в него и может выступать как инструментальная для Y_t.

Второе уравнение представляет собой тождество, параметры которого известны, поэтому необходимости в его идентификации нет.

Непосредственное оценивание структурного уравнения функции потребления обычным МНК дает

$$\hat{C}_t = 60,9 + 0,635Y_t,$$

где оценки $\alpha = 60,9$; $\beta = 0,635$ смещенные и несостоятельные.

Оцененная система приведенных уравнений есть

$$\begin{cases} \hat{C}_t = 188 + 0,696I_t, \\ \hat{Y}_t = 188 + 1,696I_t. \end{cases}$$

Выразив I_t из второго уравнения системы в виде $I_t = (Y_t - 188)/1,696$ и подставив его в первое, получим

$$\hat{C}_t = 188 + \frac{0,696}{1,696}(Y_t - 188) = 110,9 + 0,41Y_t,$$

где оценки $\alpha = 110,9$; $\beta = 0,41$ несмещенные и состоятельные.

Коэффициент $\beta = 0,41$ определяет склонность к потреблению, а $M_C = 0,696$ и $M_Y = 1,696$ — инвестиционные мультипликаторы потребления и дохода соответственно. Это значит, что если объем инвестиций возрастает на 1 ед., то объем потребления увеличится на 0,696 ед., а совокупный доход — на 1,696 ед.▲

Пример 12.2. По данным 15 торговых предприятий получены сведения о показателях, характеризующих объем продаж, интенсивность рекламы и динамику цен: y_1 — объем продаж, млн руб.; y_2 — число рекламных сообщений; x_1 — индекс цен на продукцию, %; x_2 — индекс цен на рекламу, %.

Пусть исходная модель имеет вид

$$\begin{cases} y_{1t} = \alpha_{10} + b_{11}y_{2t} + \alpha_{11}x_{1t} + \varepsilon_{1t}, \\ y_{2t} = \alpha_{20} + b_{21}y_{1t} + \alpha_{21}x_{2t} + \varepsilon_{2t}. \end{cases}$$

Оценим структурную модель на основании данных табл. 12.2.

▼В исходной модели y_1, y_2 — эндогенные переменные; x_1, x_2 — экзогенные. Первое уравнение является идентифицируемым, поскольку переменная x_2 не включена в него и может выступать как инструментальная для y_2. Аналогично второе уравнение также является идентифицируемым.

Таблица 12.2

Исходные данные для примера 12.2, усл. ед.

t	y_1	y_2	x_1	x_2
1	56,7	270	104,3	97,8
2	64,5	172	94,2	105,7
3	53,3	324	102,8	103,3
4	82,6	428	98,7	95,1
5	62	420	99,8	100,5
6	61,3	473	100,5	101,4
7	25,7	82	112,8	110,1
8	36,1	276	106,7	100,7

t	y_1	y_2	x_1	x_2
9	69,7	409	100,3	102
10	46,4	191	105	101,9
11	53,5	231	105,6	106,5
12	41,2	131	106,2	107,7
13	45,6	115	110,3	109,1
14	48,3	202	105,9	102,9
15	56,1	223	105,8	100,2

Для оценки структурных коэффициентов используем КМНК. Оценивая систему приведенных уравнений, получим

$$\begin{cases} \hat{y}_1 = 369,634 - 2,013x_1 - 1,038x_2, \\ \hat{y}_2 = 2887,483 - 9,259x_1 - 16,138x_2. \end{cases}$$

где коэффициенты уравнений статистически значимы.

Выражая переменную x_2 из второго уравнения и подставляя ее в первое, а переменную x_1 из первого уравнения и подставляя ее во второе уравнение, получим

$$\begin{cases} \hat{y}_1 = 183,935 + 0,064y_2 - 1,418x_1, \\ \hat{y}_2 = 1187,388 + 4,599y_1 - 11,365x_2, \end{cases} \blacktriangle$$

Пример 12.3. Рассматривается модель

$$\begin{cases} y_1 = \beta_{12}y_2 + \alpha_{11}x_1 + \varepsilon_1, \\ y_2 = \beta_{21}y_1 + \alpha_{22}x_2 + \varepsilon_2, \end{cases}$$

где y_1 — годовое потребление свинины на душу населения;

y_2 — оптовая цена свинины;

x_1 — доход на душу населения;

x_2 — расходы по обработке мяса.

Оценим структурную модель на основании данных табл. 12.3.

Таблица 12.3

Исходные данные для примера 12.3, усл. ед.

t	y_1	y_2	x_1	x_2
1	60	5	1300	60
2	62	4	1300	56
3	65	4,2	1500	56
4	62	5	1600	63
5	66	3,8	1800	50
Среднее	63	4,4	1500	57

▼В исходной модели y_1, y_2 — эндогенные переменные, а x_1, x_2 — экзогенные. Первое уравнение является идентифицируемым, поскольку переменная x_2 не включена в него и может выступать как инструментальная для y_2. Аналогично второе уравнение также является идентифицируемым.

Для оценки структурных коэффициентов используем КМНК.

Структурная форма модели преобразуется в приведенную форму:

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + u_1, \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + u_2. \end{cases}$$

При расчете параметров приведенной формы модели преобразуем исходные данные в центрированные (табл. 12.4).

Таблица 12.4

Центрированные данные для примера 12.3, усл. ед.

t	y_1	y_2	x_1	x_2
1	–3	0,6	–200	3
2	–1	–0,4	–200	–1
3	2	–0,2	0	–1
4	–1	0,6	100	6
5	3	–0,6	300	–7
Итого	0	0,0	0	0

Оцененная приведенная форма модели имеет вид

$$\begin{cases} \hat{y}_1 = 0,00609x_1 - 0,26481x_2, \\ \hat{y}_2 = 0,00029x_1 + 0,11207x_2. \end{cases}$$

Выражая переменную x_2 из второго уравнения и подставляя ее в первое, а переменную x_1 из первого уравнения и подставляя ее во второе, получим следующую структурную форму модели:

$$\begin{cases} \hat{y}_1 = -2,363y_2 + 0,0068x_1, \\ \hat{y}_2 = 0,048y_1 + 0,1248x_2. \end{cases}$$

Если в структурную модель включить свободные члены, т. е.

$$\begin{cases} y_1 = \alpha_{10} + \beta_{12}y_2 + \alpha_{11}x_1 + \varepsilon_1, \\ y_2 = \alpha_{20} + \beta_{21}y_1 + \alpha_{22}x_2 + \varepsilon_2, \end{cases}$$

то свободные члены уравнений определим по формулам

$$\alpha_{10} = \overline{y}_1 - \beta_{12}\overline{y}_2 - \alpha_{11}\overline{x}_1 = 63,21;$$
$$\alpha_{20} = \overline{y}_2 - \beta_{21}\overline{y}_1 - \alpha_{22}\overline{x}_2 = -5,76.$$

Тогда структурная модель имеет вид

$$\begin{cases} \hat{y}_1 = 63,21 - 2,363y_2 + 0,0068x_1, \\ \hat{y}_2 = -5,76 + 0,048y_1 + 0,1248x_2. \end{cases} \blacktriangle$$

II. Сверхидентифицируемость.

Рассмотрим следующую простую модель Кейнса формирования доходов:

$$\begin{cases} C_t = \alpha + \beta Y_t + \varepsilon_t, & \text{(функция потребления)} \\ Y_t = C_t + I_t + G_t, & \text{(тождество дохода)} \end{cases}$$

где G_t — объем государственных расходов.

В исходной модели C_t, Y_t — эндогенные переменные, а I_t, G_t — экзогенные. Обе экзогенные переменные I_t, G_t не присутствуют в структурном уравнении функции потребления и могут использоваться как инструментальные для эндогенной переменной Y_t.

Структурное уравнение модели, в которой число экзогенных переменных, которые могут использоваться как инструментальные, больше, чем необходимо, является *сверхидентифицируемым.*

Наилучшее решение в данном случае — применение *двухшагового метода наименьших квадратов* (ДМНК) и построение инструментальной переменной, которая является комбинацией I_t, G_t.

Процедура ДМНК состоит из двух этапов.

1. На основе приведенной формы модели получают для сверхидентифицируемого уравнения теоретические (расчетные) значения эндогенных переменных, содержащихся в правой части уравнения.

2. Подставляя теоретические значения эндогенных переменных вместо их фактических значений в сверхидентифицируемое уравнение и применяя обычный МНК, определяют его структурные коэффициенты.

Двухшаговый МНК можно рассматривать как способ конструирования наилучшей из возможных комбинаций инструментальных переменных в случае, когда в уравнении имеется избыток экзогенных переменных, которые можно использовать как инструментальные для объясняющей эндогенной переменной.

Пример 12.4. Для некоторой страны имеются данные о совокупном доходе (Y_t), объеме потребления (C_t), инвестициях (I_t) и государственных расходах (G_t), полученные за 10 лет. Построим функцию потребления, используя модель Кейнса формирования доходов:

$$\begin{cases} C_t = \alpha + \beta Y_t + \varepsilon_t, & \text{(функция потребления)} \\ Y_t = C_t + I_t + G_t. & \text{(тождество дохода)} \end{cases}$$

▼Исходные данные и расчетные показатели представим в табл. 12.5.

В данной модели C_t, Y_t — эндогенные переменные, а I_t, G_t — экзогенные.

Непосредственное оценивание структурного уравнения функции потребления обычным МНК приводит к следующим результатам:

$$\tilde{C}_t = 109{,}4 + 0{,}4Y_t,$$

где оценки α = 109,4; β = 0,4 — смещенные и несостоятельные.

Таблица 12.5

Исходные данные расчетные значения $\tilde{Y}$ для примера 12.4, усл. ед.

C_t	I_t	G_t	Y_t	$\tilde{Y}_t$
195	10	20	225	237,486
203	20	10	233	238,953
210	30	20	260	263,258
200	20	40	260	273,210
215	10	30	255	248,905
215	20	10	245	238,953
210	30	20	260	263,258
215	20	10	245	238,953
225	15	40	280	266,767
220	30	20	270	263,258

Структурное уравнение для функции потребления является сверхидентифицируемым, поскольку переменные I_t, G_t не включены в него и могут выступать как инструментальные для Y_t. Для оценки параметров сверхидентифицируемого уравнения используем ДМНК.

Оценивая приведенное уравнение для совокупного дохода

$$Y_t = \gamma_0 + \gamma_1 I_t + \gamma_2 G_t + \varepsilon_t$$

обычным МНК, получим

$$\hat{Y}_t = 201{,}8 + 1{,}29 I_t + 1{,}14 G_t.$$

Подставим расчетные значения $\hat{Y}$ вместо фактических значений в структурное уравнение функции потребления, получим уравнение

$$C_t = \alpha + \beta \hat{Y}_t + \varepsilon_t,$$

и оценка полученного уравнения обычным МНК дает

$$\hat{C}_t = 171,3 + 0,156\,Y_t,$$

где оценки $\alpha = 171,3$; $\beta = 0,156$ состоятельны. ▲

Пример 12.5. Имеются поквартальные данные по РФ за десять лет об объемах валового внутреннего продукта (ВВП), расходов на конечное потребление (КП), валового накопления (ВН) и чистого экспорта (Э) в среднегодовых ценах первого года периода. Оценим следующую структурную модель:

$$\begin{cases} \text{КП}_t = \alpha_1 + \beta_{11}\text{ВВП}_t + \varepsilon_{1t}, \\ \text{ВН}_t = \alpha_2 + \beta_{21}\text{ВВП}_{t-4} + \varepsilon_{2t}, \\ \text{ВВП}_t = \text{КП}_t + \text{ВН}_t + \text{Э}_t, \end{cases}$$

где ВВП_{t-4} — объем ВВП за аналогичный квартал предыдущего года.

▼Исходные данные и расчетные показатели представим в виде табл. 12.6.

Таблица 12.6

Исходные данные и расчетные значения $\widehat{\text{ВВП}}$ для примера 12.4, трлн руб.

Квартал	t	ВВП_t	КП_t	ВН_t	Э_t	ВВП_{t-4}	$\widehat{\text{ВВП}}_t$
1	1	329	260	43	26	—	—
2	2	341	248	73	20	—	—
3	3	395	266	124	5	—	—
4	4	360	251	105	4	—	—
1	5	322	257	55	10	329	322
2	6	329	254	64	11	341	335
3	7	373	244	116	13	395	393
4	8	349	241	86	22	360	361
1	9	320	255	52	13	322	316
2	10	327	258	63	6	329	320
3	11	385	267	113	5	373	366
4	12	362	278	78	6	349	341

Окончание табл. 12.6

Квартал	t	$ВВП_t$	$КП_t$	$ВН_t$	$Э_t$	$ВВП_{t-4}$	$\widehat{ВВП}_t$
1	13	316	247	73	–4	320	306
2	14	323	237	87	–1	327	314
3	15	350	270	62	18	385	385
4	16	329	275	–7	61	362	383
1	17	311	244	24	43	316	325
2	18	333	232	51	50	323	336
3	19	390	242	89	59	350	369
4	20	369	245	41	83	329	359
1	21	346	222	40	84	311	340
2	22	367	222	61	84	333	364
3	23	432	253	100	79	390	421
4	24	399	250	85	64	369	392
1	25	346	235	46	65	346	368
2	26	364	242	71	51	367	383
3	27	424	261	116	47	432	450
4	28	400	265	101	34	399	408
1	29	363	272	53	38	346	354
2	30	382	269	69	44	364	376
3	31	449	289	111	49	424	442
4	32	418	287	88	43	400	414
1	33	377	272	50	55	363	381
2	34	399	280	74	45	382	396
3	35	469	296	127	46	449	467
4	36	445	300	100	45	418	434
1	37	405	298	58	49	377	392
2	38	430	296	80	54	399	418
3	39	498	316	122	60	469	495
4	40	478	306	113	59	445	469
1	41	435	314	61	60	405	428
2	42	463	309	84	70	430	459
3	43	534	336	127	71	498	531
4	44	509	318	128	63	478	506

В структурной модели три эндогенные переменные ($КП_t$, $ВН_t$, $ВВП_t$) и две экзогенные переменные ($Э_t$, $ВВП_{t-4}$).

Первое структурное уравнение является сверхидентифицируемым, поскольку переменные $Э_t$, $ВВП_{t-4}$ не включены в него и могут выступать как инструментальные для $ВВП_t$. Для оценки параметров сверхидентифицируемого уравнения используем ДМНК.

Оценивая приведенное уравнение

$$ВВП_t = \gamma_0 + \gamma_1 Э_t + \gamma_2 ВВП_{t-4} + \varepsilon_t$$

обычным МНК, получим

$$\widehat{ВВП}_t = -30{,}66 + 0{,}50 Э_t + 1{,}06 ВВП_{t-4}.$$

Подставим расчетные значения $\widehat{ВВП}_t$ вместо фактических значений в первое структурное уравнение функции $КП_t$, получим уравнение

$$КП_t = \alpha_1 + \beta_{11} \widehat{ВВП}_t + \varepsilon_{1t}.$$

Оценка полученного уравнения обычным МНК дает

$$\widehat{КП}_t = 110{,}66 + 0{,}41 ВВП_t.$$

Поскольку второе структурное уравнение не содержит эндогенных переменных в качестве факторов, то для оценки его параметров используем обычный МНК и в результате получим

$$\widehat{ВН}_t = -103{,}67 + 0{,}48 ВВП_{t-4}.$$

Третье структурное уравнение-тождество не содержит неизвестных параметров.

Таким образом, получены следующие параметры структурной модели

$$\begin{cases} \widehat{КП}_t = 110{,}66 + 0{,}41 ВВП_t, \\ \widehat{ВН}_t = -103{,}67 + 0{,}48 ВВП_{t-4}, \\ ВВП_t = КП_t + ВН_t + Э_t, \end{cases}$$

где все коэффициенты модели статистически значимы. ▲

III. Неидентифицируемость.

Рассмотрим следующую модель спроса и предложения:

$$\begin{cases} Y^D = \alpha + \beta P + u^D, & \text{(спрос)} \\ Y^S = \delta + \varepsilon P + u^S, & \text{(предложение)} \\ Y^D = Y^S = Y, & \text{(равновесие)} \end{cases}$$

где P — цена товара;

u^D, u^S — случайные члены.

Переменные Y, P являются эндогенными, и их значения определяются в процессе установления равновесия.

Структурное уравнение модели, в которой число экзогенных переменных, которые могут использоваться как инструментальные, меньше числа объясняющих эндогенных переменных, является неидентифицируемым.

В рассматриваемой модели нет экзогенных переменных, поэтому ни одно из этих уравнений не является идентифицируемым. Чтобы модель имела статистическое решение, в нее вводятся экзогенные переменные.

Предположим, что продавцы товара облагаются специальным налогом T, который они должны платить с выручки. При этом уравнение спроса останется неизменным, если переменная P означает рыночную цену, а уравнение предложения изменится:

$$\begin{cases} Y^D = \alpha + \beta P + u^D, & \text{(спрос)} \\ Y^S = \delta + \varepsilon P + \sigma T + u^S, & \text{(предложение)} \\ Y^D = Y^S = Y, & \text{(равновесие)} \end{cases}$$

где T — экзогенная переменная.

Уравнение спроса будет идентифицируемым, поскольку переменная T не включена в него и может выступать как инструментальная для P, а уравнение предложения — неидентифицируемым.

Включим в уравнение спроса экзогенную переменную X — доход на душу населения:

$$\begin{cases} Y^D = \alpha + \beta P + \gamma X + u^D, \text{(спрос)} \\ Y^S = \delta + \varepsilon P + \sigma T + u^S, \text{ (предложение)} \\ Y^D = Y^S = Y. \qquad \text{(равновесие)} \end{cases}$$

Экзогенную переменную X можно использовать как инструментальную вместо P для уравнения предложения.

В итоге получили в целом *точно идентифицируемую модель спроса и предложения.*

Пусть структурное уравнение спроса имеет временной тренд (скажем, потому что привычки медленно меняются со временем):

$$\begin{cases} Y^D = \alpha + \beta P + \gamma X + \rho t + u^D, \text{ (спрос)} \\ Y^S = \delta + \varepsilon P + \sigma T + u^S, \qquad \text{(предложение)} \\ Y^D = Y^S = Y, \qquad \text{(равновесие)} \end{cases}$$

где t — переменная времени, а ρ — коэффициент при ней.

В модели спроса имеются две экзогенные переменные X, t, которые можно использовать в качестве инструментальных для P в уравнении предложения.

В итоге получили *сверхидентифицируемое уравнение предложения и точно идентифицируемое уравнение спроса.*

Пример 12.6. Имеется следующая модель спроса и предложения:

$$\begin{cases} Y = \alpha + \beta P + \gamma X + u_1, \text{(спрос)} \\ Y = \delta + \varepsilon P + u_2, \qquad \text{(предложение)} \end{cases}$$

где Y, P — количество товара и его цена соответственно;

X — доход потребителей.

Оценим структурные уравнения на основании данных табл. 12.7.

▼В исходной модели Y, P — эндогенные переменные, а X — экзогенная.

Таблица 12.7

Исходные данные для примера 12.6, усл. ед.

Y	5	4	8	5	9	7	11
P	2	3	4	3	7	5	6
X	1	2	3	4	5	6	7

Уравнение спроса является неидентифицируемым, а уравнение предложения — идентифицируемым, поскольку переменная X не включена в него и может выступать как инструментальная для P.

Оценивая систему приведенных уравнений, получим

$$\begin{cases} \hat{P} = 1,571 + 0,679\,X, \\ \hat{Y} = 3,429 + 0,893\,X. \end{cases}$$

Выразив X из первого уравнения системы в виде $X = \dfrac{P - 1,571}{0,679}$ и подставив его во второе, получим структурное уравнение предложения

$$\hat{Y} = 3,429 + 0,893 \cdot \frac{P - 1,571}{0,679} = 1,361 + 1,316P,$$

где оценки $\delta = 1,361$; $\varepsilon = 1,316$.

Функция спроса неидентифицируема. ▲

Упражнение 12.1. Имеется следующая модель спроса и предложения:

$$\begin{cases} Y = \alpha + \beta P + \gamma X + u_1, & \text{(спрос)} \\ Y = \delta + \varepsilon P + \sigma I + u_2, & \text{(предложение)} \end{cases}$$

где Y, P — эндогенные переменные: количество некоторого товара и его цена соответственно;

X, I — экзогенные переменные: доход на душу населения и инвестиции в производство соответственно.

Оцените структурные уравнения на основании данных табл. 12.8.

Таблица 12.8

Исходные данные для упражнения 12.1, усл. ед.

Y	*P*	*X*	*I*
20	3	34	5
33	3	43	6
28	5	51	6
41	4	49	7
40	5	55	7
36	6	62	6
42	6	70	8
38	7	68	8
51	7	78	12

Ответ: $\begin{cases} \hat{Y} = -2,743 - 12,667P + 1,842X, & \text{(спрос)} \\ \hat{Y} = 7,297 + 1,843P + 2,742I. & \text{(предложение)} \end{cases}$ ▲

Упражнение 12.2. Имеется следующая модель спроса и предложения:

$$\begin{cases} Y = \alpha + \beta P + u_1, & \text{(спрос)} \\ Y = \delta + \varepsilon P + \sigma W + u_2, & \text{(предложение)} \end{cases}$$

где Y, P — количество товара и цена;

W — зарплата потребителей.

Оцените структурные уравнения на основании данных табл. 12.9.

Таблица 12.9

Исходные данные для упражнения 12.2, усл. ед.

Y	10	12	10	8	9	7	6
P	7	6	9	7	8	11	10
W	1	2	3	4	5	6	7

Ответ: структурное уравнение спроса $\hat{Y} = 19,43 - 1,28P$, а функция предложения неидентифицируема. ▲

12.4. Ненулевое ограничение

Добавление экзогенной переменной не единственный способ, который может привести к идентифицируемости уравнения. В некоторых случаях неидентифицируемая модель может быть идентифицируема путем *задания соотношения между структурными коэффициентами.*

Рассмотрим неидентифицируемую модель спроса и предложения

$$\begin{cases} Y^D = \alpha + \beta P + u^D, & \text{(спрос)} \\ Y^S = \delta + \varepsilon P + \sigma T + u^S, & \text{(предложение)} \\ Y^D = Y^S = Y. & \text{(равновесие)} \end{cases}$$

Улучшим спецификацию модели, введя ограничение $\sigma = -\varepsilon$:

$$\begin{cases} Y^D = \alpha + \beta P + u^D, & \text{(спрос)} \\ Y^S = \delta + \varepsilon (P - T) + u^S, & \text{(предложение)} \\ Y^D = Y^S = Y. & \text{(равновесие)} \end{cases} \qquad (12.1)$$

Благодаря введению ограничения на коэффициенты $\sigma = -\varepsilon$ уравнение предложения также стало идентифицируемым.

Действительно, при использовании ИП можно рассмотреть новую версию модели как систему из четырех уравнений:

$$\begin{cases} Y^D = \alpha + \beta P + u^D, \\ Y^S = \delta + \varepsilon P_1 + u^S, \\ P_1 = P - T, \\ Y^D = Y^S, \end{cases} \qquad (12.2)$$

где P_1 — цена товара для продавца (сумма, остающаяся у него после уплаты налога).

Последние два уравнения системы (12.2) являются уравнениями-тождествами и не требуют проверки на идентификацию.

Переменная T не включена в уравнение спроса, поэтому она может использоваться как инструментальная для P. Точно так же эта переменная не включена в уравнение предложения, поэтому она может использоваться как инструментальная для P_1.

В итоге модель в целом является *точно идентифицируемой*.

Вывод. Ограничение на коэффициенты позволяет исключить одну объясняющую переменную из уравнения. Если эта переменная эндогенная, для нее не нужно искать инструментальную переменную, если экзогенная, то она освобождается на роль инструментальной для одной из эндогенных переменных, оставшихся в уравнении.

Пример 12.7. Оценим структурную модель спроса и предложения (12.1) по данным табл. 12.10.

Таблица 12.10

Исходные данные для примера 12.7, усл. ед.

T	0	2	5	8	10	12	14
P	40	42	43	44	45	48	49
Y	70	68	63	61	60	56	52

▼Рассмотрим как исходную (12.1) так и новую (12.2) версии модели.

Было показано, что исходная модель точно идентифицируема, и поэтому для оценки ее структурных коэффициентов используем КМНК.

Оцененные уравнения приведенной системы, полученные по выборочным данным обычным МНК, соответственно для моделей (12.1) и (12.2) есть

$$\begin{cases} \hat{P} = 40 + 0,6T, \\ \hat{Y} = 70,2 - 1,2T \end{cases} \text{ и } \begin{cases} \hat{P}_1 = P - T = 40 - 0,4T, \\ \hat{Y} = 70,2 - 1,2T. \end{cases}$$

Перейти от приведенной формы модели к структурной можно следующим образом:

• выразив T из первого уравнения приведенной формы в виде $T = \dfrac{P-40}{0,6}$ и подставив его во второе, получим

$$\hat{Y} = 70,2 - 1,2 \cdot \frac{P-40}{0,6} = 150,1 - 2P,$$

т. е. $\alpha = 150,1$, $\beta = -2$;

• выразив T из первого уравнения приведенной формы в виде $T = \dfrac{40 - P_1}{0,4}$ и подставив его во второе, получим

$$\hat{Y} = 70,2 - 1,2 \cdot \frac{40 - P_1}{0,4} = -50,6 + 3P_1,$$

т. е. $\delta = -50,6$, $\varepsilon = 3$.

Таким образом, получили следующую оцененную модель спроса и предложения:

$$\begin{cases} \hat{Y} = 150,1 - 2P, & \text{(спрос)} \\ \hat{Y} = -50,6 + 3(P - T). & \text{(предложение)} \end{cases} \quad \blacktriangle$$

Пример 12.8. Оценим структурную модель вида

$$\begin{cases} Y = \alpha_1 + \beta_1(C + D) + \varepsilon_1, \\ C = \alpha_2 + \beta_2 Y + \beta_3 Y_{-1} + \varepsilon_2, \end{cases}$$

где Y — валовой доход;

Y_{-1} — валовой доход предшествующего года;

C — личное потребление;

D — конечный спрос (помимо личного потребления);

ε_1, ε_2 — случайные члены.

Исходные данные представлены в табл. 12.11.

Таблица 12.11

Исходные данные для примера 12.8, усл. ед.

t	Y	C	D	Y_{-1}
1	3	7	5	47
2	23	30	23	3
3	8	1	2	23
4	21	9	12	8
5	18	25	6	21
6	37	8	45	18
7	36	30	24	37
8	47	32	50	36
9	56	40	32	47
10	32	26	20	56

▼В данной модели две эндогенные переменные (Y, C) и две экзогенные переменные (D, Y_{-1}).

Учитывая ограничение на коэффициенты в первом уравнении, а также то, что переменная Y_{-1} не включена в это уравнение, переменные D, Y_{-1} могут выступать как инструментальные для C, следовательно, первое уравнение сверхидентифицируемо. Переменная D не включена во второе уравнение, поэтому она может использоваться как инструментальная для Y, т. е. второе уравнение точно идентифицировано.

Оценивая систему приведенных уравнений, получим

$$\begin{cases} \hat{Y} = 4,146 + 0,8249D + 0,1989Y_{-1}, \\ \hat{C} = 7,328 + 0,3521D + 0,1946Y_{-1}. \end{cases}$$

Для определения параметров сверхидентифицируемого уравнения используем ДМНК.

По сверхидентифицируемому уравнению структурной формы модели заменяем фактические значения C на теоретические $\hat{C}$ и рассчитываем новую переменную $\hat{C} + D$ (табл. 12.12).

Таблица 12.12

Расчетные данные для примера 12.8, усл. ед.

Y	$\hat{C}$	D	$\hat{C}+D$
3	18,236	5	23,236
23	16,010	23	39,010
8	12,509	2	14,509
21	13,110	12	25,110
18	13,528	6	19,528
37	26,676	45	71,676
36	22,980	24	46,980
47	31,940	50	81,940
56	27,743	32	59,743
32	25,269	20	45,269

Оценивая полученное структурное уравнение обычным МНК, имеем

$$\hat{Y} = 1,443 + 0,624(C + D).$$

Для определения параметров точно идентифицированного второго уравнения используем КМНК.

Из первого уравнения приведенной системы выразим D в виде $D = \dfrac{Y - 4,146 - 0,1989Y_{-1}}{0,8249}$ и, подставив его во второе, получим

$$\hat{C} = 5,558 + 0,427Y + 0,110Y_{-1}.$$

Таким образом, получили следующую оцененную структурную модель:

$$\begin{cases} \hat{Y} = 1,443 + 0,624(C + D), \\ \hat{C} = 5,558 + 0,427Y + 0,110Y_{-1}. \end{cases}$$ ▲

Пример 12.9. Оценим следующую идентифицируемую эконометрическую модель с двумя эндогенными (y_1, y_2) и двумя экзогенными (x_1, x_2) переменными

$$\begin{cases} y_1 = \alpha_1 + \beta_{12}y_2 + \alpha_{11}x_1 + \varepsilon_1, \\ y_2 = \alpha_2 + \beta_{21}y_1 + \alpha_{22}x_2 + \varepsilon_2 \end{cases}$$

по выборочным данным, представленным в табл. 12.13.

Таблица 12.13

Исходные данные для примера 12.9, усл. ед.

y_1	y_2	x_1	x_2
2	5	1	3
3	6	2	1
4	7	3	2
5	8	2	5
6	5	4	6

▼1. Для точно идентифицируемой структурной модели применим КМНК.

Приведенная форма модели:

$$\begin{cases} y_1 = \delta_1 + \delta_{11}x_1 + \delta_{12}x_2 + \nu_1, \\ y_2 = \delta_2 + \delta_{21}x_1 + \delta_{22}x_2 + \nu_2. \end{cases}$$

Оцененные уравнения приведенной системы, полученные по выборочным данным с использованием МНК, есть

$$\begin{cases} \hat{y}_1 = 0,685 + 0,8524\,x_1 + 0,3733\,x_2, \\ \hat{y}_2 = 6,393 - 0,0724\,x_1 - 0,0056\,x_2. \end{cases}$$

Перейдем от приведенной формы модели к структурной. Для этой цели из первого уравнения приведенной формы модели надо исключить x_2, выразив его из второго уравнения:

$$x_2 = \frac{6,393 - 0,0724\,x_1 - y_2}{0,0056},$$

и подставить в первое, а из второго уравнения приведенной формы модели следует исключить x_1, выразив его из первого уравнения приведенной формы:

$$x_1 = \frac{y_1 - 0,685 - 0,3733x_2}{0,8524},$$

и подставить во второе. В результате получим следующую структурную форму модели:

$$\begin{cases} \hat{y}_1 = 429 - 67\,y_2 - 4\,x_1 + \varepsilon_1, \\ \hat{y}_2 = 6,45 - 0,085\,y_1 + 0,026\,x_2 + \varepsilon_2. \end{cases}$$

2. Покажем, что для точно идентифицируемых уравнений ДМНК дает тот же результат, что и КМНК.

Из приведенной системы уравнений можно найти расчетные значения эндогенных переменных $\hat{y}_1, \hat{y}_2$. Подставляя их вместо фактических y_1, y_2 в правую часть структурной модели и применяя обычный МНК к каждому уравнению модели, получим тот же результат, что и при КМНК.

Расчетные данные для использования ДМНК приведены в табл. 12.14.

Таблица 12.14

Расчетные данные для примера 12.9, усл. ед.

y_1	$\hat{y}_2$	x_1
2	6,304	1
3	6,242	2
4	6,164	3
5	6,220	2
6	6,070	4

y_2	$\hat{y}_1$	x_2
5	2,657	3
6	2,763	1
7	3,989	2
8	4,256	5
5	6,334	6

3. Усложним теперь задачу. Пусть в исходной идентифицируемой модели наложено *ограничение на ее параметры* $\beta_{12} = \alpha_{11}$, тогда придем к новой модели

$$\begin{cases} y_1 = \alpha_1 + \beta_{12}(y_2 + x_1) + \varepsilon_1, \\ y_2 = \alpha_2 + \beta_{21}y_1 + \alpha_{22}x_2 + \varepsilon_2. \end{cases}$$

В результате второе (точно идентифицируемое) уравнение не изменилось, следовательно, ее структурные коэффициенты не изменятся, а первое уравнение стало сверхидентифицируемым.

Для определения структурных коэффициентов первого, сверхидентифицируемого уравнения новой системы используем ДМНК. Для этого на основе второго уравнения приведенной системы находим расчетные значения $\hat{y}_2$ эндогенной переменной. Подставляя их вместо фактических y_2 в первое уравнение полученной системы и применяя обычный МНК, получим решение поставленной задачи.

Исходные данные при использовании ДМНК приведены в табл. 12.15.

Таблица 12.15

Исходные данные при использовании ДМНК для примера 12.9, усл. ед.

y_1	x_1	$\hat{y}_2$	$\hat{y}_2 + x_1$
2	1	6,304	7,304
3	2	6,242	8,242
4	3	6,164	9,164
5	2	6,220	8,220
6	4	6,070	10,070

Окончательно рассматриваемая структурная система уравнений есть

$$\begin{cases} y_1 = -6,693 + 1,243(y_2 + x_1), \\ y_2 = 6,45 - 0,085 y_1 + 0,026 x_2. \end{cases} \blacktriangle$$

12.5. Условия для идентификации

Приступать к оцениванию того или иного структурного уравнения системы имеет смысл после того, как установлена его идентифицируемость. Для установления идентифицируемости можно использовать инструментальные переменные (ИП).

В полностью определенной модели будет столько уравнений, сколько имеется эндогенных переменных.

Пусть D — число не включенных в уравнение, но присутствующих в системе экзогенных переменных, а H — число включенных в уравнение эндогенных переменных.

Необходимое условие идентификации. Уравнение в структурной модели может быть идентифицировано, если число не включенных в него экзогенных переменных не меньше числа включенных в него объясняющих эндогенных переменных, уменьшенного на единицу, т. е.

$$D \geq H - 1 \text{ (\textit{порядковое условие})}.$$

Данное условие является *необходимым*, но *не достаточным* для идентификации.

В частности:

- если $D = H - 1$, то уравнение *точно идентифицируемо*;
- если $D > H - 1$, то уравнение *сверхидентифицируемо*;
- если $D < H - 1$, то уравнение *неидентифицируемо*.

Достаточное условие идентификации. Уравнение идентифицируемо, если ранг матрицы, составленной из коэффициентов при переменных (эндогенных и экзогенных), отсутствующих в исследуемом уравнении, не меньше $N - 1$, где N — число эндогенных переменных системы.

Пример 12.10. Проверим на идентификацию каждое уравнение модели:

$$\begin{cases} y_1 = \alpha_{01} + \beta_{13}y_3 + \beta_{14}y_4 + \varepsilon_1, \\ y_2 = \alpha_{02} + \beta_{23}y_3 + \alpha_{21}x_1 + \varepsilon_2, \\ y_3 = \alpha_{03} + \beta_{34}y_4 + \alpha_{31}x_1 + \varepsilon_3, \\ y_4 = y_1 + y_2 + x_2, \end{cases}$$

где y_1 — расходы на конечное потребление текущего года;

y_2 — валовые инвестиции в текущем году;

y_3 — расходы на заработную плату в текущем году;

y_4 — валовой доход за текущий год;

x_1 — валовой доход предыдущего года;

x_2 — государственные расходы текущего года;

ε — случайные ошибки.

В данной модели четыре эндогенные переменные (y_1, y_2, y_3, y_4), т. е. $N = 4$, и две экзогенные (x_1, x_2).

Первое уравнение: $H = 3$ (y_1, y_3, y_4 присутствуют), $D = 2$ (x_1, x_2 отсутствуют) и $D = H - 1$, поэтому уравнение точно идентифицируемо (необходимое условие).

Для проверки на достаточное условие идентификации выпишем матрицу A коэффициентов при переменных, не входящих в первое уравнение (табл. 12.16).

Таблица 12.16

Матрица коэффициентов при переменных, не входящих в первое уравнение, для примера 12.10

Уравнение	y_2	x_1	x_2
2	–1	α_{21}	0
3	0	α_{31}	0
4	1	0	1

Определитель матрицы $\det(A) = -\alpha_{31} \neq 0$, следовательно, ранг матрицы равен $3 \geq N - 1$, т. е. достаточное условие идентификации выполняется, и первое уравнение точно идентифицируемо.

Второе уравнение: $H = 2$ (y_2, y_3 присутствуют), $D = 1$ (x_2 отсутствует) и $D = H - 1$, поэтому уравнение точно идентифицируемо (необходимое условие). Выпишем матрицу A коэффициентов при переменных, не входящих во второе уравнение (табл. 12.17).

Таблица 12.17

Матрица коэффициентов при переменных, не входящих во второе уравнение, для примера 12.10

Уравнение	y_1	y_4	x_2
1	–1	β_{14}	0
3	0	β_{34}	0
4	1	–1	1

Выполняется также достаточное условие идентификации: $\det(A) = -\beta_{34} \neq 0$, ранг матрицы равен $3 \geq N - 1$.

Третье уравнение: $H = 2$ (y_3, y_4 присутствуют), $D = 1$ (x_2 отсутствует) и $D = H - 1$, поэтому уравнение точно идентифицируемо (необходимое условие).

Выпишем матрицу A коэффициентов при переменных, не входящих в третье уравнение (табл. 12.18).

Таблица 12.18

Матрица коэффициентов при переменных, не входящих в третье уравнение, для примера 12.10

Уравнение	y_1	y_2	x_2
1	–1	0	0
2	0	–1	0
4	1	1	1

Здесь также выполняется достаточное условие идентификации: $\det(A) = 1$, ранг матрицы равен $3 \geq N - 1$.

Четвертое уравнение представляет собой тождество, параметры которого известны, поэтому необходимости в его идентификации нет.

Таким образом, все уравнения модели точно идентифицируемы. ▲

Пример 12.11. Выполним идентификацию следующей модели:

$$\begin{cases} C_t = \alpha_{01} + \beta_{11} Y_t + \alpha_{12} C_{t-1} + \varepsilon_1, & \text{(функция потребления)} \\ I_t = \alpha_{02} + \beta_{21} r_t + \alpha_{22} I_{t-1} + \varepsilon_2, & \text{(функция инвестиций)} \\ r_t = \alpha_{03} + \beta_{31} Y_t + \alpha_{32} M_t + \varepsilon_3, & \text{(функция денежного рынка)} \\ Y_t = C_t + I_t + G_t, & \text{(тождество дохода)} \end{cases}$$

где C_t — расходы на потребление;

Y_t — совокупный доход;

I_t — инвестиции;

r_t — процентная ставка;

M_t — денежная масса;

G_t — государственные расходы;

t — текущий период;

t–1 — предыдущий период.

▼В данной модели четыре эндогенные переменные (C_t, I_t, Y_t, r_t), т. е. $N = 4$, и четыре экзогенных (M_t, G_t, C_{t-1}, I_{t-1}).

Для *первого уравнения*: $H = 2$ (C_t, Y_t присутствуют), $D = 3$ (M_t, G_t, I_{t-1} отсутствуют) и $D > H - 1$, поэтому уравнение сверхидентифицируемо (необходимое условие).

Для проверки на достаточное условие идентификации выпишем матрицу коэффициентов при переменных, не входящих в первое уравнение (табл. 12.19).

Таблица 12.19

Матрица коэффициентов при переменных, не входящих в первое уравнение, для примера 12.11

Уравнение	I_t	r_t	I_{t-1}	M_t	G_t
2	–1	β_{21}	α_{22}	0	0
3	0	–1	0	α_{32}	0
4	1	0	0	0	1

Минор 3-го порядка данной матрицы $\begin{vmatrix} -1 & \beta_{21} & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \end{vmatrix} \neq 0$, следовательно, ранг матрицы равен $3 \geq N - 1$, т. е. достаточное условие идентификации выполняется.

Для *второго уравнения*: $H = 2$ (I_t, r_t присутствуют), $D = 3$ (M_t, G_t, C_{t-1} отсутствуют) и $D > H - 1$, поэтому уравнение сверхидентифицируемо.

Выпишем матрицу коэффициентов при переменных, не входящих во второе уравнение (табл. 12.20).

Таблица 12.20

Матрица коэффициентов при переменных, не входящих во второе уравнение, для примера 12.11

Уравнение	C_t	Y_t	C_{t-1}	M_t	G_t
1	–1	β_{11}	α_{12}	0	0
3	0	β_{31}	0	α_{32}	0
4	1	–1	0	0	1

Минор 3-го порядка данной матрицы $\begin{vmatrix} -1 & 0 & 0 \\ 0 & \alpha_{32} & 0 \\ 1 & 0 & 1 \end{vmatrix} \neq 0$, следовательно, ранг матрицы равен $3 \geq N - 1$, т. е. достаточное условие идентификации выполняется.

Для *третьего уравнения*: $H = 2$ (Y_t, r_t присутствуют), $D = 3$ (G_t, C_{t-1}, I_{t-1} отсутствуют) и $D > H - 1$, поэтому уравнение сверхидентифицируемо.

Выпишем матрицу коэффициентов при переменных, не входящих в третье уравнение (табл. 12.21).

Таблица 12.21

Матрица коэффициентов при переменных, не входящих в третье уравнение, для примера 12.11

Уравнение	C_t	C_{t-1}	I_t	I_{t-1}	G_t
1	–1	α_{12}	0	0	0
2	0	0	–1	α_{22}	0
4	1	0	1	0	1

Минор 3-го порядка данной матрицы $\begin{vmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 1 & 1 \end{vmatrix} \neq 0$, следовательно, ранг матрицы равен $3 \geq N - 1$, т. е. достаточное условие идентификации выполняется.

Четвертое уравнение представляет собой тождество, параметры которого известны, и необходимости в его идентификации нет.

Таким образом, все уравнения модели сверхидентифицируемы. ▲

Выводы.

1. Для установления идентифицируемости можно использовать ИП или условия для идентификации.

2. Для решения точно идентифицируемого уравнения применяется КМНК, а для решения сверхидентифицируемого уравнения — ДМНК.

3. Для точно идентифицируемой системы применение к ней КМНК или ДМНК дает одинаковые результаты.

4. Сверхидентифицируемая структурная модель может быть двух типов:

• все уравнения системы сверхидентифицируемы. В этом случае для оценки структурных коэффициентов каждого уравнения используется ДМНК;

• система наряду со сверхидентифицируемыми содержит точно идентифицируемые уравнения. В этом случае структурные коэффициенты по ним находятся из системы приведенных уравнений.

Глава 13.
МОДЕЛИ ФИНАНСОВОГО РЫНКА

13.1. Понятие риска финансового актива

Риск и доходность в финансовом менеджменте и анализе рассматриваются как две взаимосвязанные категории. Они используются для характеристики как отдельных видов финансовых активов, так и их комбинаций.

Существуют различные интерпретации понятия «риск». В наиболее общем виде под *риском* понимают вероятность возникновения убытков или недополучения доходов по сравнению с прогнозируемым вариантом развития событий.

В приложении к финансовым активам используют следующую интерпретацию риска и его меры: рисковость актива характеризуется степенью вариабельности дохода, который может быть получен благодаря владению данным активом.

Доходность акции r, % за некоторый период времени рассчитывается как

$$r = \frac{d + (P_1 - P_0)}{P_0} \cdot 100,$$

где d — дивиденды, выплаченные в течение периода, усл. ед.;

P_0 — курс акции на начало периода, усл. ед.;

P_1 — курс акции на конец периода, усл. ед.

Пусть имеется набор акций разных видов. Рассмотрим критерии выбора акций только одного вида из их набора с учетом риска.

Имеются два подхода к определению риска:

- на основе выделения будущих состояний экономики;
- основе статистических наблюдений.

1. Определение меры риска на основе выделения будущих состояний экономики.

В процессе обоснования рисковых инвестиционных решений, связанных с вложениями в ценные бумаги, приходится вводить предположение о том, что мы не знаем точно будущих результатов инвестиционной деятельности, но можем оценить некоторый набор их возможных значений.

Оценка сводится к выделению так называемых *будущих состояний экономики*, порождаемых совокупностью условий и факторов, определяющих каждый из ожидаемых результатов.

В числе этих состояний могут быть выделены подъем, спад, сохранение тенденции развития. В каждом таком состоянии будущие доходы определяются однозначно.

Если известны перечень будущих состояний экономики и доходы от инвестиционных альтернатив в каждом из этих состояний, то говорят о принятии инвестиционных решений в условиях неопределенности.

Если наряду с указанной информацией известны также вероятности наступления будущих состояний экономики, то говорят о принятии инвестиционных решений в условиях риска.

В качестве вероятностей наступления будущих состояний экономики рассматриваются так называемые *субъективные вероятности*, которые отражают степень, или уровень, убежденности каждого отдельного инвестора в том, что то или иное будущее состояние фактически наступит.

Будем считать, что выделен определенный набор будущих состояний экономики и известны субъективные вероятности их наступления.

В данном подходе к измерению риска учитываются основные характеристики распределения вероятностей получения будущих доходов, которые могут быть непосредственно использованы для определения наиболее предпочтительных решений.

Ожидаемая доходность $\bar{r}$ акций одного вида, дисперсия σ^2 и стандартное отклонение σ доходности вычисляются по формулам математического ожидания, дисперсии и стандартного отклонения дискретной случайной величины:

$$\bar{r} = M(r) = \sum_k r_k p_k, \quad \sigma^2 = \sigma^2(r) = \sum_k (r_k - \bar{r})^2 p_k, \quad \sigma = \sqrt{\sigma^2},$$

где p_k — субъективные вероятности наступления k-го состояния.

Стандартное отклонение при анализе финансовых рисков часто называют просто риском. Измерение риска в форме стандартного отклонения предполагает нормальное распределение доходности по акциям.

2. Определение меры риска на основе наблюдений.

Для оценки ожидаемой доходности и риска по акции конкретного вида можно использовать ряды наблюдений за фактической доходностью этой акции в прошедшие периоды.

Анализируя разброс фактических значений доходности от его среднего значения за выбранный период наблюдения, можно оценить колеблемость этих результатов в форме выборочной дисперсии или выборочного стандартного отклонения.

Выборочная доходность $\overline{r}$ *и выборочная дисперсия* σ^2 по каждому виду акций определяются по формулам

$$\overline{r} = \frac{1}{n}\sum_{t=1}^{n} r_t \text{ , } \sigma^2 = \frac{1}{n-1}\sum_{t=1}^{n}(r_t - \overline{r})^2 \text{ , } \sigma = \sqrt{\sigma^2} \text{ ,}$$

где r_t — доходность на дату t;

n — число дат.

Предполагается, что любой инвестор при прочих равных условиях всегда предпочитает вариант вложений с большей ожидаемой доходностью.

Если ожидаемая доходность акций одинакова, то инвестор, склонный к риску, выбирает акции с бóльшим риском, а инвестор, не склонный к нему, предпочитает акции с меньшим риском.

Поскольку риск представляет собой меру изменчивости, или колеблемости, доходности, то ожидаемая доходность и риск по инвестициям в ценные бумаги находятся в определенном соотношении.

Как правило, более высокому значению ожидаемой доходности по данному финансовому инструменту соответствует и больший уровень риска, и наоборот, чем меньше ожидаемая доходность, тем меньше и риск по этому инструменту. Выбирая акции с высокой ожидаемой доходностью, следует учитывать и более высокий риск, что практически означает и более высокие шансы этот доход не получить. Выбирая акции с относительно малым риском, следует рассчитывать и на относительно малую доходность: шансы на получение относительно большой доходности невелики.

Если оказывается, что стандартные отклонения доходности двух и более финансовых активов оказываются одинаковыми, то для измерения риска в данном случае используют коэффициент вариации, который представляет собой отношение стандартного отклонения к ожидаемой доходности:

$$V = \frac{\sigma}{\bar{r}} 100\ \%,$$

причем в качестве актива с наименьшей степенью риска выбирается актив с наименьшим значением коэффициента вариации.

Пример 13.1. Пусть рынок может находиться в одном из четырех состояний. Известны вероятности этих состояний и доходности трех активов (табл. 13.1).

Таблица 13.1

Исходные данные для примера 13.1

Вид акций	Доходность акций в зависимости от состояния экономики, %			
	$p_1 = 0,4$	$p_2 = 0,2$	$p_3 = 0,1$	$p_4 = 0,3$
1	3	10	16	9
2	0	14	–4	6
3	8	4	22	10

Определим рисковую характеристику каждого актива.

▼Ожидаемая доходность акций:

$\bar{r}_1 = M(r_1) = 3 \cdot 0,4 + 10 \cdot 0,2 + 16 \cdot 0,1 + 9 \cdot 0,3 = 7,5\ \%;$

$\bar{r}_2 = M(r_2) = 0 \cdot 0,4 + 14 \cdot 0,2 + (-4) \cdot 0,1 + 6 \cdot 0,3 = 4,2\ \%;$

$\bar{r}_3 = M(r_3) = 8 \cdot 0,4 + 4 \cdot 0,2 + 22 \cdot 0,1 + 10 \cdot 0,3 = 9,2\ \%.$

Дисперсия и среднее квадратическое отклонение доходности акций:

$\sigma_1^2 = (3 - 7,5)^2 \cdot 0,4 + (10 - 7,5)^2 \cdot 0,2 + (16 - 7,5)^2 \cdot 0,1 +$
$+ (9 - 7,5)^2 \cdot 0,3 = 17,25;$

$\sigma_1 = \sqrt{17,25} = 4,153\ \%;$

$\sigma_2^2 = (0 - 4{,}2)^2 \cdot 0{,}4 + (14 - 4{,}2)^2 \cdot 0{,}2 + (-4 - 4{,}2)^2 \cdot 0{,}1 + (6 - 4{,}2)^2 \cdot 0{,}3 = 34{,}44;$

$\sigma_2 = \sqrt{34{,}44} = 5{,}869\ \%;$

$\sigma_3^2 = (8 - 9{,}2)^2 \cdot 0{,}4 + (4 - 9{,}2)^2 \cdot 0{,}2 + (22 - 9{,}2)^2 \cdot 0{,}1 + (10 - 9{,}2)^2 \cdot 0{,}3 = 22{,}56;$

$\sigma_3 = \sqrt{22{,}56} = 4{,}750\ \%;$

Результаты расчета рисковой характеристики каждого актива представим в виде табл. 13.2.

Таблица 13.2

Рисковые характеристики активов для примера 13.1

Вид акций	Ожидаемая доходность ($\bar{r}_i$), %	Риск (σ_i), %
1	7,5	4,153
2	4,2	5,869
3	9,2	4,750

Получили, что стандартное отклонение для акций всех видов примерно одинаково. Из сравнения коэффициентов вариации активов

$$V_1 = 100 \cdot 4{,}153/7{,}5 = 55{,}4\ \%;$$

$$V_2 = 100 \cdot 5{,}869/4{,}2 = 139{,}7\ \%;$$

$$V_3 = 100 \cdot 4{,}750/9{,}2 = 51{,}6\ \%$$

следует, что наиболее предпочтительным является вариант вложения капитала в акции 3-го вида, имеющие наименьшее значение коэффициента вариации. ▲

Пример 13.2. По известным доходностям трех видов акций на 5 моментов времени (табл. 13.3) составим их рисковые характеристики.

▼Рисковые характеристики акций представим в виде табл. 13.4.

При расчете ожидаемой доходности $\bar{r}_i$ и риска σ_i использовались функции *Excel* СРЗНАЧ и СТАНДОТКЛОН.

Таблица 13.3

Исходные данные для примера 13.2

Вид акций	Доходность акций в различные моменты времени, %				
	1	2	3	4	5
1	9,6	10,1	11,4	11,7	12,2
2	14,2	15,9	15,3	14,1	15,5
3	7,9	8,2	6,8	8,7	8,4

Таблица 13.4

Рисковые характеристики активов для примера 13.2

Вид акций	Ожидаемая доходность ($\bar{r}_i$), %	Риск (σ_i), %
1	11	1,102
2	15	0,806
3	8	0,731

▲

13.2. Модель Марковица

Портфелем ценных бумаг инвестора называется совокупность ценных бумаг, принадлежащих данному инвестору.

При формировании портфеля следует различать рисковые и безрисковые активы.

Рисковые активы — это активы, доходность которых в будущем неопределенна.

Безрисковые активы — это активы, будущая доходность которых известна в момент погашения. Как правило, это краткосрочные правительственные облигации.

Портфель ценных бумаг, содержащий самые разнообразные типы ценных бумаг, называется *диверсифицированным портфелем*.

Подход Г. Марковица предполагает, что все инвестиции вложены в рисковые активы.

Структура портфеля акций или других ценных бумаг описывается показателями x_i, характеризующими долю стоимости акций данного вида в общей стоимости приобретаемого или имеющегося портфеля, причем выполняется соотношение

$$\sum_{i}^{n} x_i = 1.$$

Доходность портфеля — это сумма доходностей всех составляющих его бумаг:

$$r_p = \sum_{i=1}^{n} x_i r_i ,$$

где x_i — доля i-го актива в портфеле;

r_i — доходность i-го актива;

n — число активов в портфеле.

Поскольку доходность составляющих портфель ценных бумаг случайна, то и доходность r_p портфеля есть также случайная величина и имеет математическое ожидание и дисперсию.

Ожидаемая доходность портфеля акций в целом m_p при заданной его структуре определяется как математическое ожидание значений его доходности, достигаемых в каждом возможном будущем состоянии экономики:

$$m_p = M(r_p) = \sum_{i} x_i M(r_i) = \sum_{i} x_i \overline{r_i} .$$

Дисперсия доходности портфеля определяется выражением

$$\sigma_p^2 = \sigma^2(r_p) = \sum_{i,j} x_i x_j \sigma_{ij} = \sum_{i} x_i^2 \sigma_i^2 + \sum_{i \neq j} x_i x_j \sigma_{ij}$$

или

$$\sigma_p^2 = \sum_{i} x_i^2 \sigma_i^2 + 2\sum_{i<j} x_i x_j \sigma_{ij} ,$$

где производится суммирование по одинаковым нижним индексам.

Рискованность одного актива измеряется дисперсией или стандартным отклонением доходности по этому активу, а риск портфеля — дисперсией или стандартным отклонением доходности портфеля. Для оценки риска портфеля обычно используется стандартное отклонение $\sigma_p = \sqrt{\sigma_p^2}$.

Чтобы измерить риск портфеля, нужно не только знать вариацию доходов отдельных ценных бумаг, но и степень, с которой доходы ценных бумаг колеблются вместе, т. е. знать ковариацию или корреляцию доходов каждой пары активов в портфеле.

Ковариация и корреляция активов в портфеле

Ковариация — это мера связи двух случайных величин, которую можно задать в виде ковариационной матрицы.

Ковариация между доходностями двух ценных бумаг i и j можно определить выражением

$$\sigma_{ij} = \mathrm{cov}(r_i, r_j) = \sum_k (r_{ik} - \overline{r}_i)(r_{jk} - \overline{r}_j) p_k$$

или

$$\sigma_{ij} = \frac{1}{n-1} \sum_k (r_{ik} - \overline{r}_i)(r_{jk} - \overline{r}_j),$$

причем $\sigma_{ij} = \sigma_{ji}$, $\sigma_{ii} = \sigma_i^2$.

Положительное значение ковариации показывает, что доходности акций изменяются в одном направлении, отрицательное — в разных. Относительно небольшое или нулевое значение ковариации означает, что связь между доходностями акций отсутствует.

Другим показателем, измеряющим степень взаимосвязи доходностей двух акций, является *коэффициент корреляции*, определяемый выражением

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}.$$

Коэффициент корреляции нормирует ковариацию для облегчения сравнения с другими парами случайных переменных.

Коэффициент корреляции является безразмерной величиной и изменяется в пределах $-1 \leq r \leq 1$. Положительное значение коэффициента показывает, что доходности акций изменяются в одном направлении при изменении конъюнктуры, отрицательное — в противоположных. При нулевом значении коэффициента корреляции взаимосвязь между доходностями акций отсутствует.

Таким образом, ковариация и корреляция измеряют степень согласованности изменений значений двух переменных.

Пример 13.3. По данным примера 13.1 определим матрицы ковариации и корреляции трех активов.

▼Матрица ковариации доходности трех активов имеет вид:

$$(\sigma)_{ij} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix},$$

где $\sigma_{11} = \sigma_1^2 = 17{,}25$, $\sigma_{22} = \sigma_2^2 = 34{,}44$, $\sigma_{33} = \sigma_3^2 = 22{,}56$.

$\sigma_{12} = \sigma_{21} = (3 - 7{,}5) \cdot (0 - 4{,}2) \cdot 0{,}4 + (10 - 7{,}5) \cdot (14 - 4{,}2) \cdot 0{,}2 +$
$+ (16 - 7{,}5) \cdot (-4 - 4{,}2) \cdot 0{,}1 + (9 - 7{,}5) \cdot (6 - 4{,}2) \cdot 0{,}3 = 6{,}3$;

$\sigma_{13} = \sigma_{31} = (3 - 7{,}5) \cdot (8 - 9{,}2) \cdot 0{,}4 + (10 - 7{,}5) \cdot (4 - 9{,}2) \cdot 0{,}2 +$
$+ (16 - 7{,}5) \cdot (22 - 9{,}2) \cdot 0{,}1 + (9 - 7{,}5) \cdot (10 - 9{,}2) \cdot 0{,}3 = 10{,}8$;

$\sigma_{23} = \sigma_{32} = (8 - 9{,}2) \cdot (0 - 4{,}2) \cdot 0{,}4 + (4 - 9{,}2) \cdot (14 - 4{,}2) \cdot 0{,}2 +$
$+ (22 - 9{,}2) \cdot (-4 - 4{,}2) \cdot 0{,}1 + (10 - 9{,}2) \cdot (6 - 4{,}2) \cdot 0{,}3 = -18{,}24$.

В результате матрица ковариации доходности есть:

$$(\sigma)_{ij} = \begin{pmatrix} 17{,}25 & 6{,}3 & 10{,}8 \\ 6{,}3 & 34{,}44 & -18{,}24 \\ 10{,}8 & -18{,}24 & 22{,}56 \end{pmatrix}.$$

Определим матрицу корреляции доходности.

$\rho_{11} = \rho_{22} = \rho_{33} = 1$;

$$\rho_{12} = \rho_{21} = \frac{\sigma_{12}}{\sigma_1\sigma_2} = \frac{6{,}3}{4{,}153 \cdot 5{,}869} = 0{,}258;$$

$$\rho_{13} = \rho_{31} = \frac{\sigma_{13}}{\sigma_1\sigma_3} = \frac{10{,}8}{4{,}153 \cdot 4{,}750} = 0{,}547;$$

$$\rho_{23} = \rho_{32} = \frac{\sigma_{23}}{\sigma_2\sigma_3} = \frac{-18,24}{5,869 \cdot 4,750} = -0,654 .$$

В результате матрица корреляции доходности есть:

$$(\rho)_{ij} = \begin{pmatrix} 1,000 & 0,258 & 0,547 \\ 0,258 & 1,000 & -0,654 \\ 0,547 & -0,654 & 1,000 \end{pmatrix}. \quad \blacktriangle$$

Пример 13.4. По данным примера 13.2 определим матрицы ковариации и корреляции трех активов.

▼Для расчета ковариационной матрицы доходности можно воспользоваться инструментом *Ковариация* из пакета *Анализ данных Excel*. Заметим, что при использовании инструмента *Ковариация* получаем элементы ковариационной матрицы $\sigma_{ij} = \frac{1}{n}\sum_k (r_{ik} - \overline{r}_i)(r_{jk} - \overline{r}_j)$ по генеральной совокупности. Умножая каждый элемент этой матрицы на множитель $n/(n-1)$, получим ковариационную матрицу по выборке.

В итоге находим следующую матрицу ковариации:

$$(\sigma)_{ij} = \begin{pmatrix} 1,215 & 0,1 & 0,1125 \\ 0,1 & 0,65 & -0,1325 \\ 0,1125 & -0,1325 & 0,535 \end{pmatrix}.$$

Для расчета корреляционной матрицы доходности можно воспользоваться инструментом *Корреляция* из пакета *Анализ данных Excel*.

В итоге получим следующую матрицу корреляции:

$$(\rho)_{ij} = \begin{pmatrix} 1,000 & 0,113 & 0,140 \\ 0,113 & 1,000 & -0,225 \\ 0,140 & -0,225 & 1,000 \end{pmatrix}. \quad \blacktriangle$$

Эффективное множество портфелей

Для практического использования модели Марковица необходимо определить для каждой акции ожидаемую доходность, стандартное отклонение доходности и ковариацию между доходностями акций. Если имеется эта информация, то, как показал Марковиц, с помощью квадратичного программирования можно определить набор «эффективных портфелей».

Допустимым портфелем называется любой портфель, который может построить инвестор из имеющихся в наличии активов. Набор допустимых портфелей называется допустимым множеством портфелей.

Допустимый портфель называется **эффективным**, если никакой другой портфель не обеспечивает более высокую ожидаемую доходность при том же уровне риска или более низкий риск при том же уровне доходности.

Набор эффективных портфелей называется эффективным множеством портфелей Марковица. Эффективное множество портфелей Марковица называют также эффективной границей Марковица, поскольку эффективные портфели лежат на границе множества допустимых портфелей. Портфели, лежащие вне эффективной границы, недостижимы, а лежащие внутри неэффективны (рис. 13.1).

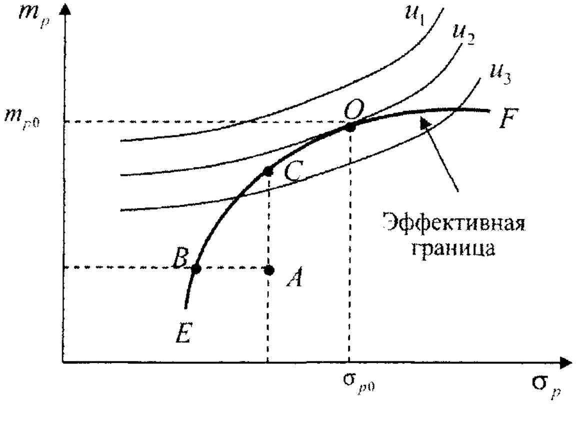


Рис. 13.1. Кривые безразличия и эффективная граница

Согласно трактовке Марковица, если имеется некий портфель *A*, то он является неэффективным, так как портфель *B* мог бы обеспечить тот же самый уровень ожидаемой доходности с меньшей степенью риска, в то время как портфель *C* при той же степени риска мог бы обеспечить более высокую ожидаемую доходность. Таким образом, все эффективные портфели должны лежать на кривой *EF* — эффективной границе. Инвестор всегда выбирает портфель, лежащий на эффективной границе.

Портфели, которые лежат в средней части кривой, обычно содержат много ценных бумаг, в то время как портфели, которые находятся ближе к краям, — всего несколько.

Эффективная граница начинается с портфеля *E*, в котором сочетание нескольких акций обеспечивает наименьшую степень риска портфеля, и заканчивается портфелем *F*, в котором все инвестиции вложены в акции одного вида с максимальной ожидаемой доходностью.

Выбор наиболее приемлемого для инвестора портфеля ценных бумаг зависит от предпочтений инвестора. Предпочтения инвестора описываются функцией полезности, которая графически представляется при помощи набора кривых безразличия. Кривая безразличия определяет комбинации риска и ожидаемой доходности, дающие одинаковый уровень полезности.

Выпуклость кривых безразличия вниз связана с несклонностью инвесторов к риску. Наклон кривых безразличия характеризует степень избегания риска инвестором. Чем выше лежит кривая, тем больше полезность. Инвестор должен выбирать портфель, лежащий на кривой безразличия, расположенной выше и левее всех остальных кривых.

На рис. 13.1 изображены три кривые безразличия u_1, u_2, u_3 и эффективная граница. **Оптимальный портфель** — это такой портфель, для которого кривая безразличия касается эффективной границы.

Определение функции полезности и построение кривых безразличия представляют собой непростую задачу. Экономисты еще не пришли к единому мнению, как измерять полезность.

Из набора *n* ценных бумаг можно сформировать бесконечное число портфелей. Модель Марковица не дает возможности выбрать оптимальный портфель, а определяет набор эффективных портфелей, каждый из которых:

1) обеспечивает максимальную ожидаемую доходность для некоторого уровня риска;

2) обеспечивает минимальный риск для некоторого значения ожидаемой доходности.

В такой постановке задачу оптимизации портфеля можно сформулировать следующим образом.

1. Найти x_i, максимизирующие ожидаемую доходность портфеля

$$m_p = \sum_i x_i \overline{r}_i \to \max$$

при условии, что обеспечивается заданное значение риска портфеля σ_{p0}, т. е.

$$\sigma_p = \sqrt{\sum_{i,j} x_i x_j \sigma_{ij}} = \sigma_{p0},$$

и поскольку x_i — доли, то в сумме они должны составлять единицу:

$$\sum_i x_i = 1.$$

Назовем данную формализацию *портфелем Марковица максимальной эффективности.*

2. Найти x_i, минимизирующие риск портфеля

$$\sigma_p = \sqrt{\sum_{i,j} x_i x_j \sigma_{ij}} \to \min$$

при условии, что обеспечивается заданное значение доходности портфеля m_{p0}, т. е.

$$m_p = \sum_i x_i \overline{r}_i = m_{p0},$$

и поскольку x_i — доли, то в сумме они должны составлять единицу:

$$\sum_i x_i = 1.$$

Назовем данную формализацию *портфелем Марковица минимального риска.*

В модели Марковица требуется еще выполнение условий $x_i \geq 0$.

А в модели Блэка допустимы любые значения переменных x_i, причем $x_i^* \geq 0$ означает рекомендацию вложить долю x_i^* наличного капитала в ценные бумаги i-го вида, а $x_i^* < 0$ — провести операцию «короткая продажа». «Короткая продажа» — продажа акций (или иных активов), взятых взаймы и не являющихся собственностью инвесторов; предполагает возврат акций в конце установленного периода.

Нахождение эффективной границы легко и наглядно решается для случая двух бумаг, а в общем случае — численными методами на компьютере.

Пример 13.5. Определим эффективное множество портфелей, составленных из рисковых активов двух видов: $\overline{r_1} = 10\ \%$, $\sigma_1 = 2\ \%$; $\overline{r_2} = 30\ \%$, $\sigma_2 = 3\ \%$ и $\sigma_{12} = -2$.

▼Ожидаемая доходность такого портфеля и ее дисперсия:

$m_p = 10x_1 + 30x_2$, $\sigma_p^2 = 4x_1^2 + 9x_2^2 - 4x_1x_2$, а риск равен σ_p.

Для определения эффективного множества портфелей, составленных из двух рисковых активов, построим по точкам зависимость «риск — доходность». Все доступные комбинации риска и доходности показаны в табл. 13.5 и на рис. 13.2.

Таблица 13.5

Доступные комбинации риска и доходности для примера 13.5

Задано		**Вычислено**			**Портфель**
x_1	x_2	σ_p^2	σ_p	m_p	
1,00	0,00	4,00	2,00	10	*A*
0,80	0,20	2,28	1,51	14	
0,70	0,30	1,93	1,39	16	
0,60	0,40	1,92	1,39	18	
0,50	0,50	2,25	1,50	20	
0,40	0,60	2,92	1,71	22	

Задано		Вычислено			Портфель
x_1	x_2	σ_p^2	σ_p	m_p	
0,20	0,80	5,28	2,30	26	
0,00	1,00	9,00	3,00	30	*B*
Вычислено					
0,65	0,35	1,88	1,37	17	*D*

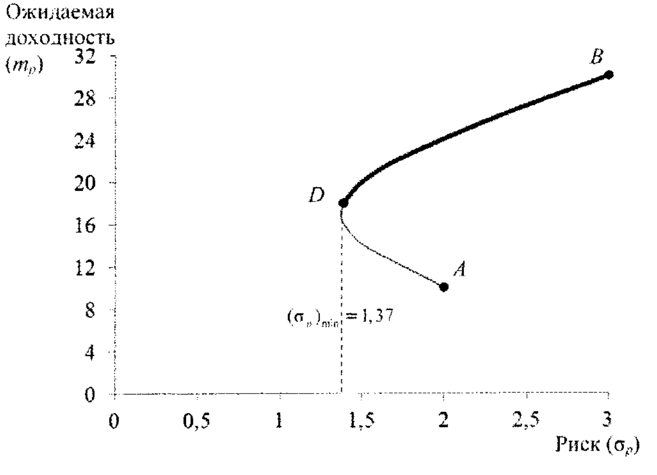


Рис. 13.2. Определение эффективного множества портфелей, составленных из двух рисковых активов

Дадим анализ кривой, приведенной на рис. 13.2. Точка *A* на графике соответствует портфелю, полностью составленному из первого актива, а точка *B* — полностью составленному из второго. При перемещении из точки *A* к точке *B* наблюдается не только повышение средней ставки доходности, но и снижение стандартного отклонения до значения, соответствующего точке *D*, после чего стандартное отклонение повышается.

Данный график показывает наличие структуры, обеспечивающей минимальный риск портфеля $(\sigma_p)_{min}$, составленного из двух рисковых активов (самая левая точка *D* на графике). Найдем координаты точки *D*, соответствующей минимальному значению σ_p.

Для этого воспользуемся необходимым условием экстремума функции от одной переменной x_1

$$\sigma_p^2 = x_1^2\sigma_1^2 + (1-x_1)^2\sigma_2^2 + 2x_1(1-x_1)\sigma_{12}.$$

Приравняв к нулю ее производную по этой переменной, получим

$$x_1 = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}},\ x_2 = 1 - x_1.$$

Координаты точки D есть x_1 = 0,65; x_2 = 0,35; при этом $(\sigma_p)_{\min}$ = 1,37.

Портфели, расположенные от минимально рискованного портфеля и выше (кривой отрезок DB), принадлежат **эффективному множеству** в модели Марковица. ▲

13.3. Модель Тобина

Подход Г. Марковица предполагает, что все инвестиции вложены в рисковые активы. В модели Дж. Тобина инвестору разрешается инвестировать не только в рисковые, но и в безрисковые активы. Под безрисковым активом $r_б$ понимается актив, по которому доход является строго установленным. По определению стандартное отклонение доходности по безрисковому активу равно нулю. Следовательно, ковариация между доходностями безрискового актива и любого рискового актива равна нулю. Инвестирование в безрисковый актив называется также *безрисковым кредитованием*, поскольку оно фактически является предоставлением займа государству.

Кроме того, инвестору разрешается одалживать деньги под процент, равный доходности безрискового актива (*безрисковое заимствование*).

Учет возможности безрискового кредитования

С появлением на рынке безрискового актива инвестор получил возможность вкладывать часть своих денег в этот актив, а остаток — в любой из рискованных портфелей, содержащихся во множестве достижимости Марковица. Появление новых возможностей существенно расширяет множество достижимости и, что важнее, изменяет расположение значительной части эффективного множества Марковица.

Рассмотрим портфель A, составленный каким-то способом из n рисковых ценных бумаг. Зафиксируем портфель A и составим новый портфель, состоящий на $(1 - V)$ из безрисковой ценной бумаги и на V из портфеля A.

Доходность такого портфеля имеет вид

$$r_p = Vr_A + (1 - V)r_б.$$

Ожидаемая доходность портфеля и ее дисперсия определяются как

$$\begin{aligned} &m_p = V\overline{r}_A + (1-V)r_б \text{ или } m_p = r_б + V(\overline{r}_A - r_б), \\ &\sigma_p^2 = V^2\sigma_A^2, \sigma_p = V\sigma_A, \end{aligned} \quad (13.1)$$

где $\overline{r}_A$ и σ_A — ожидаемая доходность и риск портфеля A.

Исключая V из системы уравнений, получим уравнение прямой линии HA (рис. 13.3):

$$m_p = r_б + \frac{\overline{r}_A - r_б}{\sigma_A}\sigma_p.$$

На рис. 13.3 показано эффективное множество EF для портфелей из рисковых ценных бумаг (кривая линия) и возможные сочетания риска и ожидаемой доходности при добавлении в рисковый портфель безрисковой ценной бумаги (прямая HA).

Из всех линий, которые могут быть проведены из точки H, соответствующей безрисковому активу, и соединяют эту точку с рисковым портфелем, ни одна не имеет больший наклон, чем линия, идущая в точку T.

Портфель, располагающийся в точке касания T, называется *касательным*. Именно объединением этого портфеля рисковых активов с безрисковым активом достигается формирование *максимально эффективного портфеля*.

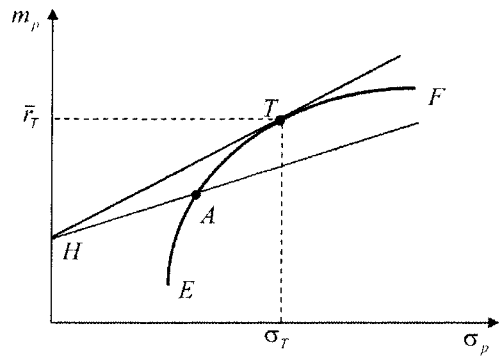


Рис. 13.3. Максимально эффективные портфели в модели Тобина

Уравнение линии «риск — доходность» для комбинированного портфеля, составленного из безрискового актива и касательного портфеля, есть

$$m_p = r_б + \frac{\bar{r}_T - r_б}{\sigma_T}\sigma_p.$$

Часть эффективного множества модели Марковица отсекается этой линией. В частности, портфели, которые принадлежали эффективному множеству в модели Марковица и располагались между минимально рискованным портфелем E и портфелем T, с введением возможности инвестирования в безрисковые активы не являются эффективными. Теперь эффективное множество состоит из прямого и искривленного отрезка. Прямой отрезок идет от безрискового актива в точку T и поэтому представляет портфели, составленные из различных комбинаций безрискового актива и портфеля T. Искривленный отрезок расположен выше и правее точки T и представляет портфели из эффективного множества модели Марковица.

Пример 13.6. Сформируем множество портфелей, составленных из рискового актива ($\bar{r}_1 =$ 16,20 %, σ_1 = 12,08 %) и безрискового актива с фиксированной ставкой $r_б$ = 4 %.

▼Пусть x_1 обозначает долю средств, вложенную в рисковый актив, и $x_б = 1 - x_1$ обозначает долю, инвестированную в безрисковый актив.

Все доступные комбинации риска и доходности показаны в табл. 13.6 и на рис. 13.4.

Таблица 13.6

Доступные комбинации риска и доходности для примера 13.6

Доля рискового актива в портфеле (x_1)	**Доля безрискового актива в портфеле (x_6)**	**Ожидаемая доходность (m_p)**	**Стандартное отклонение (σ_p)**	**Портфель**
0	1	4	0	*H*
0,25	0,75	7,05	3,02	*B*
0,5	0,5	10,1	6,04	*C*
0,75	0,25	13,15	9,06	*D*
1	0	16,2	12,08	*E*

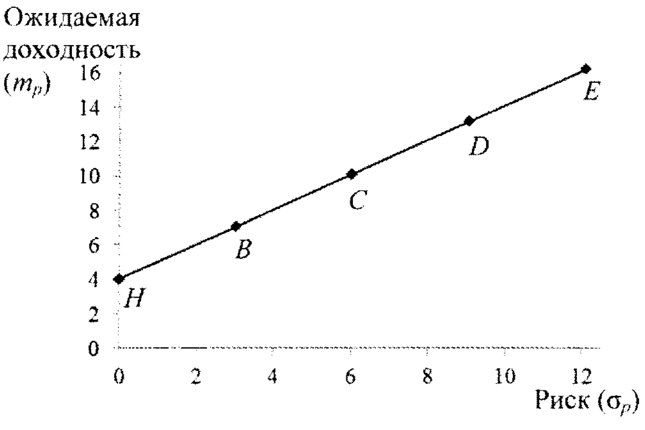


Рис. 13.4. Сочетание безрискового кредитования с инвестированием в рисковый актив (пример 13.6)

Например, для портфеля *B*, используя формулы (13.1), имеем:

$$m_p = 16{,}20 \cdot 0{,}25 + 4 \cdot 0{,}75 = 7{,}05;\ \sigma_p = 0{,}25 \cdot 12{,}08 = 3{,}02.$$

Точка *H* на рис. 13.4 соответствует ситуации, когда все деньги вкладываются в безрисковый актив, а точка *E* — ситуации, когда все деньги инвестируются в рисковый актив. Линия *HE* представляет множество портфелей, составленных из рискового и безрискового актива. Так, портфель C наполовину состоит из рискового актива, наполовину — из безрискового.

Следовательно, если инвестировать 10 000 усл. ед. в портфель C, то 5000 усл. ед. следует инвестировать в безрисковый актив и 5000 усл. ед. — в рисковый актив. ▲

Пример 13.7. Сформируем максимально эффективный портфель, составленный из двух рисковых активов примера 13.5 и безрискового актива с фиксированной ставкой $r_б = 8$ %.

▼На рис. 13.5 показано графическое представление всех возможных комбинаций риск — доходность.

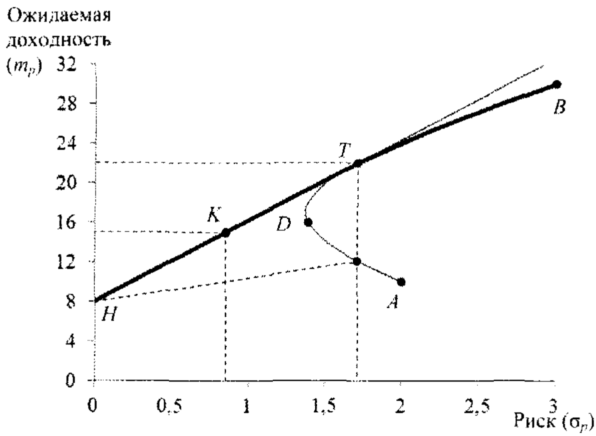


Рис. 13.5. Достижимое и эффективное множества при возможности безрискового кредитования для примера 13.7

Кривая ADB есть кривая, изображенная на рис. 13.2. Прямая линия, соединяющая точку H с любой точкой кривой ADB, представляет собой график, описывающий соотношение риск — доходность для всех комбинаций следующих трех активов: рискованных активов 1 и 2 с безрисковым активом. Наибольшее значение этого соотношения, которое можно достичь, находится на линии, соединяющей точки H и T.

Точка T является общей точкой прямой линии, выходящей из точки H, и кривой, и, кроме того, это точка касания прямой и кривой. Исходя из этих условий, определим состав касательного портфеля T:

$$x_1 = \frac{(\bar{r}_1 - r_б)\sigma_2^2 - (\bar{r}_2 - r_б)\sigma_{12}}{(\bar{r}_1 - r_б)\sigma_2^2 + (\bar{r}_2 - r_б)\sigma_1^2 - (\bar{r}_1 + \bar{r}_2 - 2r_б)\sigma_{12}},\ x_2 = 1 - x_1.$$

Подставляя данные примера, получим

$$x_1 = \frac{(10-8)\cdot 9 - (30-8)\cdot(-2)}{(10-8)\cdot 9 + (30-8)\cdot 4 - (10+30-2\cdot 8)\cdot(-2)} = 0{,}40;\ x_2 = 0{,}60,$$

при этом ожидаемая доходность и риск касательного портфеля есть

$$\bar{r}_T = x_1\bar{r}_1 + x_2\bar{r}_2 = 0{,}40 \cdot 10\ \% + 0{,}60 \cdot 30\ \% = 22\ \%;$$

$$\sigma_T = \sqrt{x_1^2\sigma_1^2 + x_2^2\sigma_2^2 + 2x_1x_2\sigma_{12}} =$$
$$= \sqrt{0{,}4^2 \cdot 4 + 0{,}6^2 \cdot 9 - 2\cdot 0{,}4\cdot 0{,}6\cdot 2} = 1{,}71\ \%.$$

Именно объединением этого портфеля рискованных активов с безрисковым активом достигается формирование максимально эффективного портфеля.

Уравнение линии «риск — доходность» для комбинированного портфеля, составленного из безрискового актива и касательного портфеля, есть

$$m_p = r_б + \frac{\bar{r}_T - r_б}{\sigma_T}\sigma_p = 8 + \frac{22-8}{1{,}71}\sigma_p = 8 + 8{,}187\sigma_p.$$

С учетом только безрискового кредитования эффективное множество на графике приобретает вид прямого отрезка *HT*, исходящего из точки *H*, соответствующей безрисковой ставке, к точке касания *T* с эффективным множеством Марковица, а также искривленного отрезка *TB* части эффективного множества Марковица, лежащей выше и правее точки касания.

Инвестор может выбрать позицию в любой точке на отрезке *HT* рис. 13.5 в зависимости от своего отношения к риску. Пусть на этом отрезке выбрана точка *K* с ожидаемой доходностью $m_K = 15$ %.

Для определения структуры оптимального портфеля вначале решим относительно V следующее уравнение: $mK = m_TV + r_б(1 - V)$, где m_T = 22 % и $r_б$ = 8 %. При этом данное уравнение будет записано так: 15 % = 22 % · V + 8 % · (1 – V) $\Rightarrow$ V = 0,5. Это означает, что оптимальный портфель состоит на 0,5 из касательного портфеля и на 0,5 из безрискового актива.

В терминах инвестиций в рискованные активы 1 и 2 это означает:

$$VX_T = 0{,}5\begin{pmatrix}0{,}4\\0{,}6\end{pmatrix} = \begin{pmatrix}0{,}2\\0{,}3\end{pmatrix}.$$

Таким образом, инвестор должен инвестировать 50 % начального капитала в долях 20 % и 30 % в рискованные активы 1 и 2 соответственно, а 50 % начального капитала должны быть использованы для покупки безрискового актива.

Следовательно, если инвестировать 10 000 усл. ед. в портфель K, то 5000 усл. ед. следует инвестировать в безрисковый актив, 2000 усл. ед. — в рисковый актив 1 и 3000 усл. ед. — в рисковый актив 2. ▲

Понимая взаимосвязь между риском и доходом и влияние ковариации, можно определить задачу оптимизации портфеля. Произвести оптимизацию портфеля можно, используя **модель Тобина максимальной эффективности**.

Постановка задачи оптимизации. Найти $x_i, x_б$, максимизирующие ожидаемую доходность портфеля

$$m_p = \sum_{i=1}^{n} x_i \bar{r}_i + x_б r_б \to \max$$

при условии, что обеспечивается заданное значение риска портфеля σ_{p0}, т. е.

$$\sigma_p = \sqrt{\sum_{i,j} x_i x_j \sigma_{ij}} = \sigma_{p0},$$

и поскольку $x_i, x_б$ — доли, то в сумме они должны составлять единицу:

$$\sum_i x_i + x_б = 1.$$

Назовем данную формализацию портфелем Тобина максимальной эффективности.

Пример 13.8. Добавим к трем рисковым активам ($\bar{r}_1$ = 7 %, σ_1 = 3 %; $\bar{r}_2$ = 11 %, σ_2= 3,5 %; $\bar{r}_3$ = 15 %, σ_3 = 4 % и σ_{12} = 4, σ_{13} = –5, σ_{23} = 8) безрисковый актив с фиксированной ставкой $r_б$ = 6 %. Произведем оптимизацию портфеля, используя модель Тобина максимальной эффективности.

▼Ожидаемая доходность и риск такого портфеля:

$$m_p = 7x_1 + 11x_2 + 15x_3 + 6x_б,$$

$$\sigma_p = \sqrt{9x_1^2 + 12{,}25x_2^2 + 16x_3^2 + 8x_1x_2 - 10x_1x_3 + 16x_2x_3}\ .$$

Для нахождения оптимального портфеля Тобина максимальной эффективности воспользуемся средствами *Excel* и процедурой *Поиск решения*. Вид рабочего листа с исходными данными и необходимыми для расчета показателями представлен на рис. 13.6.

	A	B	C	D
1	**Исходные данные**			
2	$\bar{r}_1$	$\bar{r}_2$	$\bar{r}_3$	$\bar{r}_б$
3	7	11	15	6
4	σ_1	σ_2	σ_3	
5	3	3.5	4	
6	σ_{12}	σ_{13}	σ_{23}	
7	4	-5	8	
8	**Изменяемые ячейки**			
9	x_1	x_2	x_3	$x_б$
10	0.39	0.00	0.55	0.07
11	**Задаем**			
12		σ_{p0} =	2	
13	**Целевая ячейка**			
14	m_p		11.30	$\to$max
15	**Ограничения**			
16	$\sum x = 1$		1.00	1
17	$\sigma_p = \sigma_{p0}$		2.00	2.00

Рис. 13.6. Условия примера 13.8 в *Excel*

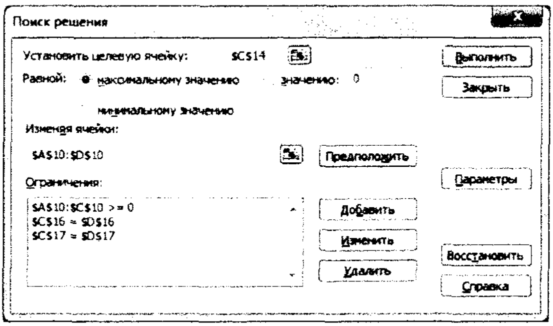
Поясним ввод формул в соответствующие ячейки таблицы (табл. 13.7).

Таблица 13.7

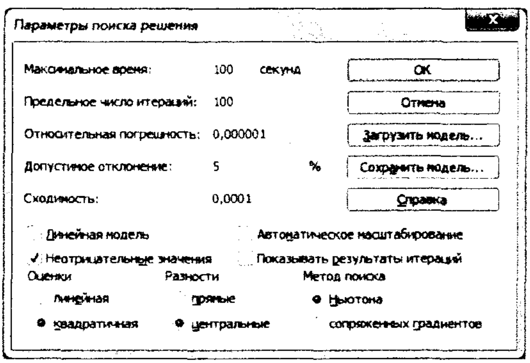
Использование формул *Excel* для решения примера 13.8

Ячейка	Формула
C12	Размещено заданное значение риска портфеля
C14	=СУММПРОИЗВ(A3:D3;A10:D10)
C16	=СУММ(A10:D10)
C17	=КОРЕНЬ(A5^2*A10^2+B5^2*B10^2+C5^2*C10^2+2*A7*A10*B10+2*B7*A10*C10+2*C7*B10*C10)
D17	=C12

На рис. 13.7, *а* показано заполненное диалоговое окно *Поиск решения*, соответствующее исходным (расчетным) данным.



а)



б)

Рис. 13.7. Заполненные диалоговые окна *Поиск решения* (*а*) и *Параметры поиска решения* (*б*) для примера 13.8

С помощью окна *Параметры поиска решения* (рис. 13.7, *б*) устанавливаем дополнительное ограничение, активируя флажок *Неотрицательные значения*, если возможно только безрисковое кредитование (а безрисковое заимствование невозможно).

Задавая последовательно значения σ_{p0} (увеличивая риск), вычисляем отвечающие ему значения максимальной ожидаемой доходности m_p. Результаты расчета приведены в табл. 13.8 и на рис. 13.8.

Таблица 13.8

Результаты вычисления максимальной ожидаемой доходности при заданных значениях риска для примера 13.8

Задано	Вычислено					Портфель
σ_p	m_p	x_1	x_2	x_3	x_6	
0,00	6,00	0,00	0,00	0,00	1,00	*H*
1,50	9,97	0,29	0,00	0,41	0,30	
2,00	11,30	0,39	0,00	0,55	0,07	
2,14	11,67	0,41	0,00	0,58	0,00	*T*
2,5	12,48	0,31	0,00	0,69	0,00	
3	13,40	0,20	0,00	0,80	0,00	
3,5	14,22	0,10	0,00	0,90	0,00	
4	15	0,00	0,00	1,00	0,00	*C*

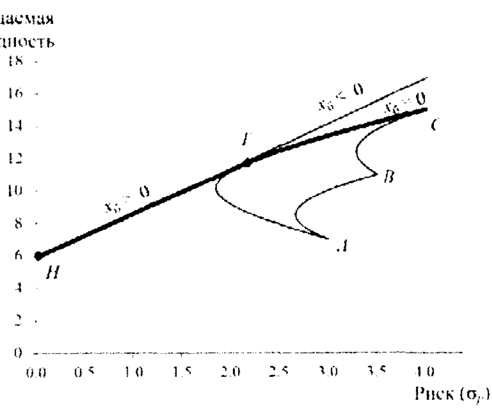


Рис. 13.8. Достижимое и эффективное множества при возможности безрискового кредитования для примера 13.8

С учетом только безрискового кредитования эффективное множество на графике приобретает вид прямого отрезка *НТ*, исходящего из точки, соответствующей безрисковой ставке, к точке касания *Т* с эффективным множеством Марковица, а также искривленного отрезка *ТС* части эффективного множества Марковица, лежащей выше и правее точки касания. ▲

Одновременный учет безрискового заимствования и кредитования

Рассмотрим, как изменяется допустимое множество, если введена возможность как предоставления, так и получения займа по одной и той же безрисковой процентной ставке.

Пример 13.9. Сформируем множество портфелей по данным примера 13.6.

▼Все доступные комбинации риска и доходности показаны в табл. 13.9 и на рис. 13.9.

Таблица 13.9

Доступные комбинации риска и доходности для примера 13.9

Доля рискового актива в портфеле (x_1)	Доля безрискового актива в портфеле ($x_б$)	Ожидаемая доходность (m_p)	Стандартное отклонение (σ_p)	Портфель
0	1	4	0	*H*
0,25	0,75	7,05	3,02	*B*
0,5	0,5	10,1	6,04	*C*
0,75	0,25	13,15	9,06	*D*
1	0	16,2	12,08	*E*
1,25	–0,25	19,25	15,1	*F*
1,5	–0,5	22,3	18,12	*G*
1,75	–0,75	25,35	21,14	*I*
2	–1	28,4	24,16	*J*

Рис. 13.9 показывает, что все четыре портфеля, содержащие безрисковое заимствование (*F*, *G*, *I*, *J*), лежат на той же самой прямой линии, что и четыре портфеля, включающих безрисковое кредитование (*H*, *B*, *C*, *D*). При этом чем больше величина взято-

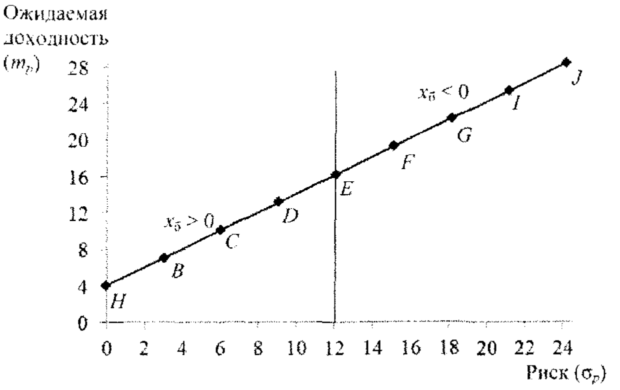


Рис. 13.9. *Сочетание безрискового заимствования и кредитования с инвестированием в рисковый актив для примера 13.9*

го займа, т. е. чем меньше $x_б$, тем дальше на прямой располагается портфель. Можно показать, что любая комбинация заимствования и инвестирования в рисковый актив лежит на этой прямой, ее точное расположение зависит от величины займа. ▲

Произвести оптимизацию портфеля можно, используя также **модель Тобина минимального риска**.

Постановка задачи оптимизации. Найти $x_i, x_б$, минимизирующие риск портфеля

$$\sigma_p = \sqrt{\sum_{i,j} x_i x_j \sigma_{ij}} \rightarrow \min$$

при заданном значении его ожидаемой доходности m_{p0}, т. е.

$$m_p = \sum_{i=1}^{n} x_i \overline{r}_i + x_б r_б = m_{p0},$$

и поскольку $x_i, x_б$ — доли, то в сумме они должны составлять единицу:

$$\sum_i x_i + x_б = 1.$$

Назовем данную формализацию портфелем Тобина минимального риска.

Пример 13.10. Рассмотрим портфель, составленный из рисковых активов *A*, *B*, *C* со следующими ожидаемыми доходностями: $\overline{r_1}$ = 16,2 %, $\overline{r_2}$ = 24,6 %, $\overline{r_3}$ = 22,8 %, ковариационная матрица которых представляет собой

$$(\sigma)_{ij} = \begin{pmatrix} 146 & 187 & 145 \\ 187 & 854 & 104 \\ 145 & 104 & 289 \end{pmatrix},$$

и безрискового актива со ставкой $r_б$ = 4 %.

Произведем оптимизацию портфеля, используя модель Тобина минимального риска.

▼Ожидаемая доходность и риск такого портфеля:

$$m_p = 16{,}2x_1 + 24{,}6x_2 + 22{,}8x_3 + 4x_б,$$

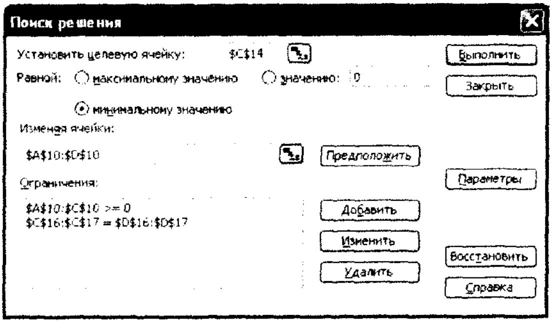
$$\sigma_p = \sqrt{146x_1^2 + 854x_2^2 + 289x_3^2 + 374x_1x_2 + 290x_1x_3 + 208x_2x_3}\ .$$

Для нахождения оптимального портфеля Тобина минимального риска воспользуемся средствами *Excel* и процедурой *Поиск решения*. Вид рабочего листа с исходными данными и необходимыми для расчета показателями представлен на рис. 13.10.

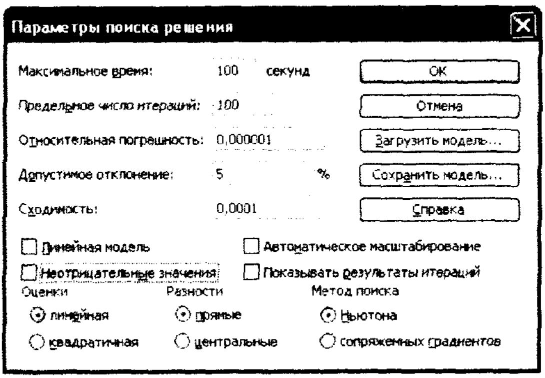
	A	B	C	D
1	**Исходные данные**			
2	r_1	r_2	r_3	$r_б$
3	16.2	24.6	22.8	4
4	σ_1	σ_2	σ_3	
5	12.08	29.22	17	
6	σ_{12}	σ_{13}	σ_{23}	
7	187	145	104	
8	**Изменяемые ячейки**			
9	x_1	x_2	x_3	$x_б$
10	*0.12*	*0.19*	*0.69*	*0.00*
11	**Задаем**			
12		m_{p0} =	22.4	
13	**Целевая ячейка**			
14		σ_p	15.2	→ min
15	**Ограничения**			
16	$\sum x = 1$		1	1
17	$m_p - m_{p0}$		22.4	22.4

Рис. 13.10. Условия примера 13.10 в *Excel*

На рис. 13.11, *а* показано заполненное диалоговое окно *Поиск решения*, соответствующее исходным (расчетным) данным.



а)



б)

Рис. 13.11. Заполненные диалоговые окна *Поиск решения* (*а*) и *Параметры поиска решения* (*б*) для примера 13.10

В окне *Параметры поиска решения* (рис. 13.11, *б*) можно ввести дополнительное ограничение, установив флажок *Неотрицательные значения*, если возможно только безрисковое кредитование. Если допускается безрисковое заимствование, то флажок не устанавливается.

Задавая последовательно значения m_{p0} (увеличивая доходность), вычисляем отвечающие ему значения минимального риска σ_p.

Результаты расчета приведены в табл. 13.10 и на рис. 13.12.

Таблица 13.10

Результаты вычисления минимального риска при заданных значениях доходности для примера 13.10

Задано	Вычислено					Портфель
m_p	σ_p	x_1	x_2	x_3	x_6	
4,0	0,0	0,00	0,00	0,00	1,00	*H*
14,0	8,3	0,06	0,11	0,37	0,46	*K*
20,0	13,2	0,10	0,17	0,60	0,13	
22,4	15,2	0,12	0,19	0,69	0,00	*T*
22,8	15,6	0,06	0,21	0,73	0,00	
23,0	15,7	0,03	0,22	0,75	0,00	
24,0	21,4	0,00	0,67	0,33	0,00	
24,6	29,2	0,00	1,00	0,00	0,00	
23,0	15,7	0,12	0,20	0,71	–0,03	
23,5	16,1	0,13	0,21	0,73	–0,06	
24,0	16,5	0,13	0,21	0,75	–0,09	
27,0	19,0	0,15	0,24	0,86	–0,25	*P*
30,0	21,5	0,17	0,27	0,97	–0,42	

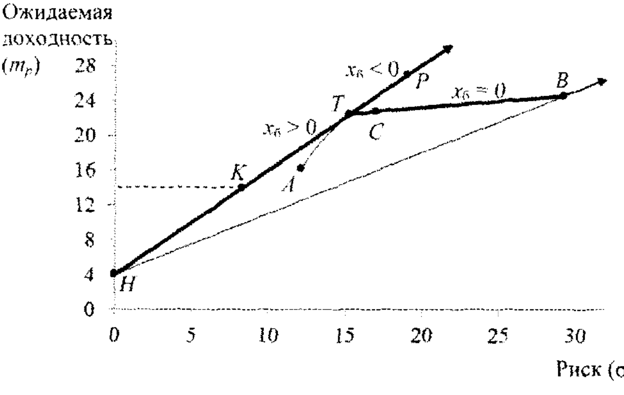


Рис. 13.12. Достижимое и эффективное множества в случае возможности безрискового заимствования и кредитования

Рис. 13.12 изображает, как изменяется допустимое множество, если введена возможность как представления, так и получения займа по одной и той же безрисковой ставке. Множество допустимости представлено областью, расположенной между двумя лучами, выходящими из точки, соответствующей безрисковой ставке, и проходящими через точки, соответствующие акциям *B* и портфелю *T*. Эти два луча уходят в бесконечность при условии, что нет ограничений на величину получаемого займа.

Луч, идущий через портфель *T*, является особенно важным, поскольку он представляет эффективное множество. Это означает, что на нем располагаются портфели, предлагающие наилучшие возможности для инвестора. Касательный портфель *T* состоит из инвестиций в рисковые акции трех видов в пропорции 0,12 : 0,19 : 0,69 соответственно.

Линия, идущая через *T*, является касательной к эффективному множеству модели Марковица. Кроме портфеля *T* ни один из портфелей, которые находились в эффективном множестве модели Марковица, не является эффективным после введения возможности предоставления и получения безрисковых займов. Любой портфель (кроме *T*), принадлежащий эффективному множеству модели Марковица, уступает портфелям, лежащим на верхнем луче и имеющим больший ожидаемый доход при том же самом значении риска (стандартном отклонении доходности).

С учетом только безрискового кредитования (без заимствования) эффективное множество на графике приобретает вид прямого отрезка *HT*, исходящего из точки, соответствующей безрисковой ставке, к точке касания с эффективным множеством Марковица, а также искривленного отрезка *TB* части эффективного множества Марковица, лежащей выше и правее точки касания.

Инвестор может выбрать позицию в любой точке на луче *HT* (рис. 13.12) в зависимости от своего отношения к риску. Пусть на этом луче выбрана точка *K* с ожидаемой доходностью $m_K = 14$ %. Для определения структуры оптимального портфеля вначале решим относительно V следующее уравнение: $m_K = m_T V + r_б(1 - V)$, где $m_T = 22{,}4$ % и $r_б = 4$ %. При этом данное уравнение будет записано так: $14\ \% = 22{,}4\ \% \cdot V + 4\ \% \cdot (1 - V) \Rightarrow V = 0{,}54$. Это означает, что оптимальный портфель состоит на 0,54 из касательного портфеля и на 0,46 из безрискового актива.

В терминах инвестиций в акции A, B и C это означает:

$$VX_T = 0{,}54\begin{pmatrix}0{,}12\\0{,}19\\0{,}69\end{pmatrix} = \begin{pmatrix}0{,}07\\0{,}10\\0{,}37\end{pmatrix}.$$

Таким образом, инвестор должен вложить 54 % начального капитала в долях 7, 10 и 37 % в акции A, B, C соответственно и 46 % начального капитала использовать для покупки безрискового актива.

Аналогично, если на луче HT (рис. 13.12) выбрана точка P с ожидаемой доходностью m_P = 27 %, то при этом решением уравнения 27 % = 22,4 % · V + 4 % · (1 – V) будет V = 1,25. То есть оптимальный портфель состоит из получения займа в размере 25 % начального капитала и из инвестирования занятых денег и начального капитала в портфель T. В терминах инвестиций в акции A, B и C получим:

$$VX_T = 1{,}25\begin{pmatrix}0{,}12\\0{,}19\\0{,}69\end{pmatrix} = \begin{pmatrix}0{,}15\\0{,}24\\0{,}86\end{pmatrix}.$$

Таким образом, инвестор должен инвестировать деньги в долях 15, 24 и 86 % его начального капитала в акции A, B и C соответственно. ▲

Упражнение 13.1. Рискованный портфель имеет 15%-ю ожидаемую доходность. Безрисковая доходность равна 5 %. Определите, какова ожидаемая доходность нового портфеля, если следующая доля средств инвестируется в рискованный портфель, а остаток — в безрисковый актив: а) 120 %; б) 90 %; в) 75%.

Ответ: ожидаемая доходность нового портфеля равна:

а) m_p = 17 %;

б) m_p = 14 %;

в) m_p = 12,5 %.

Упражнение 13.2. Рассмотрим рискованный портфель, имеющий ожидаемую доходность 18 %. Определите, как можно составить портфель, имеющий ожидаемую доходность 24 %, если безрисковая ставка равна 5 %.

Ответ: доля рискованного портфеля — 1,46.

13.4. Рыночная модель Шарпа

В 1964 г. У. Шарпом была предложена однофакторная рыночная модель доходности актива. Он сформулировал и обосновал утверждение о том, что доходность любого капитального финансового актива, обращающегося на фондовом рынке, точно коррелирует с некоторым фактором, присущим данному рынку. В качестве такого фактора может быть выбран уровень доходности рыночного индекса.

Рыночный индекс — это взвешенная сумма курсов наиболее важных для рынка ценных бумаг, а доходность рыночного индекса представляет собой их усредненную доходность.

Одним из наиболее широко известных рыночных индексов в США является *S&P* 500, который представляет собой средневзвешенную величину курсов акций 500 наиболее крупных компаний, а наиболее часто цитируемым рыночным индексом является индекс Доу — Джонса (*DJIA*).

В России наиболее известным является *индекс РТС*, определяемый по акциям наиболее крупных российских компаний.

Рыночный портфель — совокупность всех акций, обращающихся на фондовом рынке. Однако в связи с тем, что точно определить структуру рыночного портфеля не удается, в качестве рыночного портфеля используют рыночный индекс.

В рыночной модели предполагается, что доход по акции связан с доходом по рыночному индексу следующим образом:

$$r_i = \alpha_{iM} + \beta_{iM} r_M + \varepsilon_{iM},$$

где r_i, r_M — фактические доходности акции данного вида и рыночного индекса в определенные моменты времени;

α_{iM} — ордината точки пересечения прямой с вертикальной осью;

β_{iM} — величина наклона прямой;

ε_{iM} — величина случайной ошибки.

Коэффициент наклона рыночной модели часто называют коэффициентом бета (или просто бетой).

Рыночный индекс r_M в определенной степени отражает состояние экономики в целом, а рыночная модель показывает, насколько доходность ценной бумаги соответствует экономической динамике страны.

На основе выборочных наблюдений определяется *характеристическая линия* данной ценной бумаги:

$$m_i = \alpha_i + \beta_i \overline{r}_M ,$$

где $m_i = M(r_i)$, $\overline{r}_M$ — ожидаемая доходность акции данного вида и рыночного портфеля соответственно;

$$\alpha_i = \alpha_{iM},\ \beta_i = \beta_{iM}.$$

Коэффициенты α_i и β_i рассчитываются по формулам

$$\beta_i = \sigma_{iM} / \sigma_M^2 ,\ \alpha_i = \overline{r}_i - \beta_i \overline{r}_M ,$$

где σ_{iM} — ковариация между доходностью i-й ценной бумаги и доходностью рыночного индекса;

$\overline{r}_M$ — среднее значение доходности индекса;

σ_M^2 — дисперсия доходности индекса.

Диверсификация

Исходя из рыночной модели, общий риск ценной бумаги i, измеряемый ее дисперсией σ_i^2, состоит из двух частей: 1) рыночный (или систематический) риск; 2) собственный, нерыночный (или несистематический) риск:

$$\sigma_i^2 = \beta_i^2 \sigma_M^2 + \sigma_{\varepsilon i}^2 ,$$

где σ_M^2 — дисперсия доходности рыночного индекса;

$\beta_i^2 \sigma_M^2$ — рыночный риск ценной бумаги i;

$\sigma_{\varepsilon i}^2$ — собственный риск ценной бумаги i, мерой которого является дисперсия случайной погрешности ε_{iM} .

Общий риск портфеля. Если долю фондов инвестора, вложенную в ценную бумагу i данного портфеля p, обозначить через x_i, то доходность портфеля

$$r_p = \sum_{i=1}^{N} x_i r_i = \sum x_i (\alpha_i + \beta_i r_M + \varepsilon_{iM}) = \alpha_p + \beta_p r_M + \varepsilon_{pM} ,$$

где $\alpha_p = \sum x_i \alpha_i$; $\beta_p = \sum x_i \beta_i$; $\varepsilon_{pM} = \sum x_i \varepsilon_{iM}$.

Таким образом, рыночная модель портфеля является прямым обобщением рыночных моделей отдельных ценных бумаг.

Общий риск портфеля есть

$$\sigma_p^2 = \beta_p^2 \sigma_M^2 + \sigma_{\varepsilon p}^2,$$

где $\beta_p^2 = \left(\sum x_i \beta\right)^2$.

Предполагая, что случайные отклонения доходности ценных бумаг являются некоррелированными, имеем $\sigma_{\varepsilon p}^2 = \sum x_i^2 \sigma_{\varepsilon i}^2$.

Таким образом, общий риск портфеля состоит из двух компонентов, аналогичных двум компонентам общего риска отдельных ценных бумаг. Эти компоненты также носят название рыночного риска ($\beta_p^2 \sigma_M^2$) и собственного риска ($\sigma_{\varepsilon p}^2$).

Рыночный риск портфеля. Чем более диверсифицирован портфель (т. е. чем большее количество ценных бумаг в него входит), тем меньше каждая доля x_i. Так как коэффициент бета портфеля является средним значением коэффициентов бета ценных бумаг, входящих в портфель, то значение β_p при диверсификации не меняется существенным образом. Можно утверждать, что диверсификация приводит к усреднению рыночного риска.

Собственный риск портфеля. Если предположить, что во все ценные бумаги инвестировано одинаковое количество средств, то доля x_i составит $1/N$, а уровень собственного риска будет равен

$$\sigma_{\varepsilon p}^2 = \sum_{i=1}^{N} \left(\frac{1}{N}\right)^2 \sigma_{\varepsilon i}^2 = \frac{1}{N}\left(\frac{\sigma_{\varepsilon 1}^2 + \sigma_{\varepsilon 2}^2 + \ldots + \sigma_{\varepsilon N}^2}{N}\right).$$

Значение, находящееся внутри круглых скобок последнего выражения, является средним собственным риском ценных бумаг, образующих портфель. Но собственный риск портфеля в N раз меньше данного значения. Если портфель становится более диверсифицированным, то количество бумаг в нем (N) становится больше. Это означает, что величина $1/N$ уменьшается, что приводит к уменьшению собственного риска портфеля.

Можно утверждать, что диверсификация существенно уменьшает собственный риск.

Таким образом, увеличение диверсификации может привести к снижению общего риска портфеля. Это происходит вследствие сокращения собственного риска портфеля, в то время как рыночный риск портфеля остается приблизительно таким же. Рис. 13.13 показывает, как диверсификация приводит к снижению собственного риска и усреднению рыночного риска.

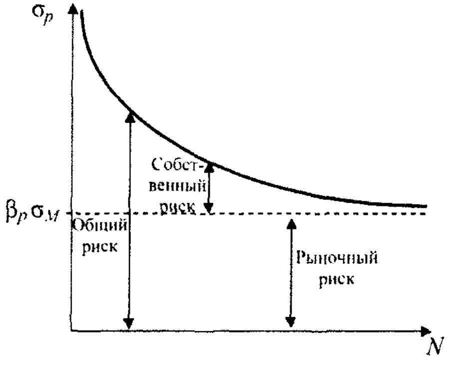


Рис. 13.13. Риск и диверсификация

Упражнение 13.3. Рыночная модель портфеля имеет вид

$$r_p = 1{,}5\ \% + 0{,}90 r_M + \varepsilon_{pM}\ .$$

Какой будет ожидаемая доходность портфеля, если ожидаемая доходность на индекс рынка составляет 12 %?

Ответ: 12,3 %.

Упражнение 13.4. Портфель составлен из трех ценных бумаг, характеристики которых представлены в табл. 13.11

Таблица 13.11

Исходные данные для упражнения 13.4

Ценная бумага	**Бета-коэффициент**	**Стандартное отклонение случайной погрешности, %**	**Доля в портфеле**
A	1,20	5	0,30
B	1,05	8	0,50
C	0,90	2	0,20

Вычислите общий риск портфеля, если стандартное отклонение доходности рыночного индекса равняется 18 %.

Ответ: 19,7 %.▲

Пример 13.11. Рассмотрим два портфеля: один, состоящий из четырех ценных бумаг; а второй — из десяти. Все ценные бумаги имеют бета-коэффициент, равный единице, и собственный риск в 30 %. В обоих портфелях доли всех ценных бумаг одинаковы. Вычислим общий риск обоих портфелей, если стандартное отклонение доходности индекса рынка составляет 20 %.

▼Пусть N — число ценных бумаг в портфеле, тогда доля x_i составит $1/N$.

Уровень собственного риска портфеля будет равен

$$\sigma_{\varepsilon p}^2 = \sum_{i=1}^{N}\left(\frac{1}{N}\right)^2 \sigma_{\varepsilon i}^2 = \frac{\sigma_{\varepsilon i}^2}{N} = \frac{900}{N}.$$

Бета портфеля: $\beta_p = 1$.

Общий риск портфеля:

$$\sigma_p^2 = \beta_p^2 \sigma_M^2 + \sigma_{\varepsilon p}^2 = 400 + \frac{900}{N}.$$

Стандартное отклонение доходности первого портфеля ($N = 4$):

$$\sigma_p = \sqrt{400 + \frac{900}{4}} = \sqrt{625} = 25\ \%.$$

Стандартное отклонение доходности второго портфеля ($N = 10$):

$$\sigma_p = \sqrt{400 + \frac{900}{10}} = \sqrt{490} = 22{,}1\ \%. \ ▲$$

Пример 13.12. Рассмотрим две ценные бумаги A и B с коэффициентами бета, равными $\beta_A = 1{,}2$ и $\beta_B = 0{,}8$; собственными рисками $\sigma_{\varepsilon A}^2 = 37$ и $\sigma_{\varepsilon B}^2 = 23$. Дисперсия доходности рыночного индекса $\sigma_M^2 = 64$.

▼ Значения дисперсии доходности для ценных бумаг A и B есть: $\sigma_A^2 = (1{,}2^2 \cdot 64) + 37 = 129$ и $\sigma_B^2 = (0{,}8^2 \cdot 64) + 23 = 64$.

a) *Портфель, состоящий из двух ценных бумаг.* Рассмотрим комбинацию ценных бумаг A и B в портфеле, образованном вложением равного количества денег инвестора в каждую ценную бумагу, т. е. $x_A = x_B = 0{,}5$.

Бета данного портфеля:

$$\beta_p = (0{,}5 \cdot 1{,}2) + (0{,}5 \cdot 0{,}8) = 1{,}0.$$

Дисперсия случайного отклонения доходности портфеля:

$$\sigma^2_{\varepsilon p} = (0{,}5^2 \cdot 37) + (0{,}5^2 \cdot 23) = 15.$$

Дисперсия доходности портфеля: $\sigma^2_p = (1{,}0^2 \cdot 64) + 15 = 79.$ Данное выражение показывает общий риск портфеля, состоящего из двух ценных бумаг.

b) *Портфель, состоящий из трех ценных бумаг.* Добавим к двум ценным бумагам A и B третью (C) ценную бумагу с $\beta_C = 1$ и $\sigma^2_{\varepsilon C} = 30$. Сформируем портфель, состоящий из трех ценных бумаг, взятых в равных пропорциях ($x_A = x_B = x_C = 0{,}33$).

Дисперсия доходности ценной бумаги C есть

$$\sigma^2_C = (1{,}0^2 \cdot 64) + 30 = 94.$$

Бета портфеля:

$$\beta_p = (0{,}33 \cdot 1{,}2) + (0{,}33 \cdot 0{,}8) + (0{,}33 \cdot 1{,}0) = 1{,}0.$$

Таким образом, увеличение диверсификации не привело к изменению уровня рыночного риска. Вместо этого оно привело к усреднению рыночного риска.

Дисперсия случайного отклонения доходности портфеля:

$$\sigma^2_{\varepsilon p} = (0{,}33^2 \cdot 37) + (0{,}33^2 \cdot 23) + (0{,}33^2 \cdot 30) = 10.$$

Отметим, что дисперсия случайного отклонения доходности портфеля, состоящего из трех ценных бумаг, меньше дисперсии доходности портфеля, состоящего из двух ценных бумаг (т. е. $10 < 15$). Таким образом, в данном примере увеличение диверсификации действительно уменьшило собственный риск.

Доходность портфеля, состоящего из трех ценных бумаг, имеет дисперсию

$$\sigma_p^2 = (1{,}0^2 \cdot 64) + 10 = 74.$$

Это выражение показывает общий риск портфеля, значение которого меньше, чем значение общего риска портфеля, состоящего из двух ценных бумаг (74 < 79). Таким образом, увеличение диверсификации привело к снижению общего риска.

Для определения риска портфеля можно также использовать формулу, предложенную Марковицем:

$$\sigma_p^2 = \sum_i x_i^2\sigma_i^2 + \sum_{i,j} x_i x_j \beta_i \beta_j \sigma_M^2\ ,$$

тогда дисперсия доходности портфеля, состоящего:

а) из двух ценных бумаг: $\sigma_p^2 = x_A^2\sigma_A^2 + x_B^2\sigma_B^2 + 2x_A x_B \beta_A \beta_B \sigma_M^2 =$ $= 0{,}5^2 \cdot 129 + 0{,}5^2 \cdot 64 + 2 \cdot 0{,}5 \cdot 0{,}5 \cdot 1{,}2 \cdot 0{,}8 \cdot 64 = 79;$

b) трех ценных бумаг: $\sigma_p^2 = x_A^2\sigma_A^2 + x_B^2\sigma_B^2 + x_C^2\sigma_C^2 + 2x_A x_B \beta_A \beta_B \sigma_M^2 +$ $+ 2x_A x_C \beta_A \beta_C \sigma_M^2 + 2x_B x_C \beta_B \beta_C \sigma_M^2 = 0{,}33^2 \cdot 129 + 0{,}33^2 \cdot 64 + 0{,}33^2 \cdot 94 +$ $+ 2 \cdot 0{,}33 \cdot 0{,}33 \cdot 1{,}2 \cdot 0{,}8 \cdot 64 + 2 \cdot 0{,}33 \cdot 0{,}33 \cdot 1{,}2 \cdot 1{,}0 \cdot 64 +$ $+ 2 \cdot 0{,}33 \cdot 0{,}33 \cdot 0{,}8 \cdot 1{,}0 \cdot 64 = 74.$ ▲

Бета-коэффициент. Для измерения рыночного риска актива (портфеля) используется величина бета. Она показывает зависимость между доходностью актива (портфеля) и доходностью рынка. Доходность рынка — это доходность рыночного портфеля.

Поскольку величина бета определяется по отношению к рыночному портфелю, то бета самого рыночного портфеля равна единице, так как ковариация доходности рыночного портфеля с самой собой есть ее дисперсия, отсюда

$$\beta_M = \frac{\sigma_M^2}{\sigma_M^2} = 1\ ,$$

где β_M — бета рыночного портфеля.

Бета актива (портфеля) без риска равна нулю, потому что нулю равна ковариация доходности актива (портфеля) без риска с доходностью рыночного портфеля.

Величина β актива (портфеля) говорит о том, насколько его риск больше или меньше риска рыночного портфеля.

Если для акции некоторой компании:

• $|\beta_i| = 1$, то доходность этой акции в среднем совпадает с доходностью рыночного портфеля, или, с другой стороны, акции данной компании имеют среднюю степень риска, сложившуюся на рынке в целом (такие акции называются среднерисковыми);

• $|\beta_i| > 1$, то доходность этой акции растет в среднем быстрее, чем по рыночному портфелю, или, с другой стороны, акции данной компании более рискованные, чем в среднем на рынке (такие акции называются агрессивными). Такие акции следует иметь в своем портфеле, когда ожидается рост доходности рыночного портфеля. Они могут обеспечить инвестору более высокий уровень доходности, чем в среднем по рынку;

• $|\beta_i| < 1$, то доходность этой акции растет в среднем медленнее, чем по рыночному портфелю, или, с другой стороны, акции данной компании менее рискованные, чем в среднем на рынке (такие акции называются оборонительными). Такие акции следует иметь в своем портфеле, когда ожидается падение доходности рыночного портфеля, поскольку сокращение доходности по ним будет меньше, чем по рынку.

Бета может быть как положительной, так и отрицательной величиной.

Если:

• $\beta_i > 0$, то доходность бумаг данного вида колеблется в такт с рынком;

• $\beta_i < 0$, то поведение бумаги прямо противоположно колебаниям доходности рынка в целом.

Активы с отрицательной бетой являются ценными инструментами для диверсификации портфеля, поскольку в этом случае можно построить портфель с «нулевой бетой», который не будет нести риска.

Рассмотрим портфель из акций двух видов. Ожидаемая доходность такого портфеля и ее дисперсия есть:

$$m_p = x_1\overline{r_1} + x_2\overline{r_2},$$

$$\sigma_p^2 = x_1^2\sigma_1^2 + x_2^2\sigma_2^2 + 2x_1x_2\sigma_{12}, \text{ или } \sigma_p^2 = x_1^2\sigma_1^2 + x_2^2\sigma_2^2 + 2x_1x_2\sigma_1\sigma_2\rho_{12}.$$

Из формулы для дисперсии следует, что уменьшение риска портфеля по сравнению с риском вложений в каждый вид ценных бумаг зависит от степени коррелированности ρ_{12} доходности этих ценных бумаг, а также от выбора структуры портфеля. Можно показать, что при коэффициенте $\rho_{12} = -1$ существует структура портфеля с нулевым риском.

Действительно, из условий

$$\begin{cases} \sigma_p^2 = 0, \\ x_1 + x_2 = 1 \end{cases} \quad \text{или} \quad \begin{cases} x_1\sigma_1 - x_2\sigma_2 = 0, \\ x_1 + x_2 = 1 \end{cases}$$

следует, что портфель, содержащий рисковые бумаги с отрицательной корреляцией $\rho_{12} = -1$ в пропорциях

$$x_1 = \frac{\sigma_2}{\sigma_1 + \sigma_2}; \; x_2 = \frac{\sigma_1}{\sigma_1 + \sigma_2},$$

имеет нулевой риск. Доходность такого безрискового портфеля равна

$$m_p = x_1\bar{r}_1 + x_2\bar{r}_2 = \frac{\sigma_2\bar{r}_1 + \sigma_1\bar{r}_2}{\sigma_1 + \sigma_2}.$$

Пример 13.13. Имеется портфель, состоящий из двух ценных бумаг, характеристики которых приведены в табл. 13.12.

Таблица 13.12

Исходные данные для примера 13.13

Ценная бумага	**Ожидаемая доходность, %**	**Стандартное отклонение, %**	**Доля в портфеле**
A	10	20	0,35
B	15	25	0,65

Для различных уровней корреляции этих ценных бумаг определим максимальное и минимальное значения стандартного отклонения доходности портфеля.

▼Из выражения дисперсии доходности портфеля

$$\sigma_p^2 = x_1^2\sigma_1^2 + x_2^2\sigma_2^2 + 2x_1x_2\sigma_1\sigma_2\rho_{12}$$

следует: при $\rho_{12} = 1$ имеем $\max \sigma_p = x_1\sigma_1 + x_2\sigma_2 = 0{,}35 \cdot 20 + 0{,}65 \times$ $\times\ 25 = 23{,}25$ %; при $\rho_{12} = -1$ имеем $\min \sigma_p = |x_1\sigma_1 - x_2\sigma_2| = 0{,}65 \cdot 25 -$ $-\ 0{,}35 \cdot 20 = 9{,}25$ %. ▲

Пример 13.14. Рассматриваются две акции: акция 1, акция 2. Пусть известны индекс РТС и курсы акций на конец месяца (табл. 13.13).

Таблица 13.13

Исходные данные для примера 13.14, долл.

Период наблюдения	Индекс РТС	Курс акции 1	Курс акции 2
Январь	55,12	0,071	0,027
Февраль	70,03	0,099	0,044
Март	80,36	0,129	0,046
Апрель	91,83	0,153	0,05
Май	97,64	0,151	0,057
Июнь	125,65	0,16	0,087
Июль	116,49	0,183	0,083
Август	102,5	0,154	0,069
Сентябрь	83,12	0,128	0,052
Октябрь	97,8	0,167	0,058
Ноябрь	112,38	0,189	0,073
Декабрь	177,74	0,304	0,113

При определении доходности будем учитывать только изменения курса акций (без учета дивидендов).

▼Преобразуя данные табл. 13.13, определим доходности индекса РТС и акций обоих видов в течение указанного периода (табл. 13.14).

Например, доходность в феврале.

Индекс РТС: $100 \cdot (70{,}03 - 55{,}12)/55{,}12 = 27{,}05$ %;

Курс акции 1: $100 \cdot (0{,}099 - 0{,}071)/0{,}071 = 39{,}44$ %;

Курс акции 2: $100 \cdot (0{,}044 - 0{,}027)/0{,}027 = 62{,}96$ %.

Таблица 13.14

Фактическая доходность для примера 13.14, %

Период наблюдения	Индекс РТС	Курс акции 1	Курс акции 2
Февраль	27,05	39,44	62,96
Март	14,75	30,30	4,55
Апрель	14,27	18,60	8,70
Май	6,33	–1,31	14,00
Июнь	28,69	5,96	52,63
Июль	–7,29	14,38	–4,60
Август	–12,01	–15,85	–16,87
Сентябрь	–18,91	–16,88	–24,64
Октябрь	17,66	30,47	11,54
Ноябрь	14,91	13,17	25,86
Декабрь	58,16	60,85	54,79

Преобразование данных табл. 13.13 в данные табл. 13.14 удобно производить с использованием электронных таблиц *Excel*. Используя пакет *Анализ данных Excel* (инструмент *Регрессия*), получим характеристические линии соответствующих акций:

$$m_1 = 4,17 + 0,93\overline{r}_M\,,\ m_2 = 1,60 + 1,19\overline{r}_M\,,$$

где m_i, $\overline{r}_M$ — ожидаемая доходность i-й акции и рыночного портфеля соответственно.

При этом

- акция 1: $\alpha_1 = 4,17$; $\beta_1 = 0,93$; $R_1^2 = 0,72$; $\sigma_{\varepsilon 1} = 12,96$;
- акция 2: $\alpha_2 = 1,60$; $\beta_2 = 1,19$; $R_2^2 = 0,77$; $\sigma_{\varepsilon 2} = 14,65$.

На рис. 13.14 представлены графики характеристических линий обеих акций.

Полученные уравнения можно использовать для прогнозирования ожидаемой доходности акций в зависимости от прогноза ожидаемой доходности по фондовому рынку.

Анализ полученных результатов.

- Расчетные значения коэффициентов альфа показывают, что при нулевой доходности фондового рынка бóльшая доходность достигается по акциям 1:

$$\alpha_1 = 4,17 > 1,60 = \alpha_2\,.$$

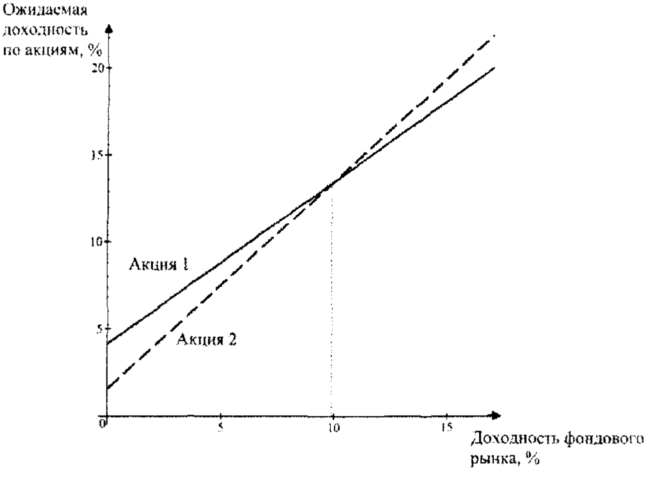


Рис. 13.14. Характеристические линии акций для примера 13.14

• Из сравнения значений коэффициентов бета для акций обоих видов следует, что с ростом доходности фондового рынка доходность по акциям 2 будет возрастать быстрее, чем по в среднем по рынку, а при падении доходности фондового рынка доходность по акциям 1 будет падать медленнее, чем в среднем по рынку:

$$\beta_1 = 0{,}93 < 1, \beta_2 = 1{,}19 > 1.$$

• Коэффициенты $R_1^2 = 0{,}72$ и $R_2^2 = 0{,}77$ показывают долю рыночного риска в общем риске по соответствующим акциям в форме дисперсии, а доля нерыночного риска: $(1 - R_1^2) = 1 - 0{,}72 = 0{,}28$ и $(1 - R_2^2) = 1 - 0{,}77 = 0{,}23$.

• Из рис. 13.14 следует, что с ростом доходности фондового рынка ожидаемая доходность акций обоих видов возрастает. При относительно небольшой доходности фондового рынка бóльшую ожидаемую доходность обеспечивают акции 1. В точке пересечения прямых ожидаемые доходности по акциям обоих видов совпадают. При дальнейшем увеличении ожидаемой доходности фондового рынка бóльшую доходность обеспечивают акции 2, для которых коэффициент бета больше единицы.

Анализ полученных данных показывает, что при ожидаемом уменьшении доходности рыночного портфеля (соответствующего биржевого индекса) целесообразнее иметь в портфеле акции с коэффициентом бета меньше единицы, а при прогнозируемом увеличении доходности рыночного портфеля — акции с коэффициентом бета больше единицы. ▲

Информация о значениях коэффициентов α, β, R^2 и σ_ε для различных ценных бумаг, определенных подобным образом в зависимости от выбранного рыночного портфеля и установленного периода наблюдения, регулярно публикуется в специальных бюллетенях.

Пример 13.15. Акции компании имеют бета-коэффициент, равный 1,20. В течение пяти лет акции этой компании и индекс рынка демонстрировали доходность, представленную в табл. 13.15.

Таблица 13.15

Исходные данные о доходности для примера 13.15, %

Год	Акции	Индекс рынка
1	17,2	14
2	–3,1	–3
3	13,3	10
4	28,5	25
5	9,8	8

Предполагая, что коэффициент смещения рыночной модели равен 0 %, вычислим стандартное отклонение случайной погрешности рыночной модели за данный период.

▼По условию ожидаемые доходности акций компании задаются выражением $m_i = 1{,}2r_M$. В табл. 13.16 приведены значения ожидаемой и фактической доходности за рассматриваемый период.

Таблица 13.16

Расчетная таблица доходности акций для примера 13.15

Год	Фактическая доходность, %	Ожидаемая доходность, %	Отклонение (σ), %	σ^2
1	17,2	16,8	0,4	0,16
2	–3,1	–3,6	0,5	0,25
3	13,3	12	1,3	1,69

Окончание табл. 13.16

Год	Фактическая доходность, %	Ожидаемая доходность, %	Отклонение (σ), %	σ^2
4	28,5	30	−1,5	2,25
5	9,8	9,6	0,2	0,04
			$\Sigma\sigma^2$ =	4,39

Стандартное отклонение случайной погрешности за данный период: $\Sigma\sigma^2/5 = 4{,}39/5 = 0{,}88$. ▲

13.5. Модель оценки финансовых активов (*CAPM*)

Модель оценки финансовых активов (*CAPM*) показывает взаимосвязь между риском и равновесной ожидаемой доходностью рискованных активов.

Модель *CAPM* опирается на следующие предположения.

1. Инвесторы имеют одинаковые представления в отношении прогнозов по ожидаемым ставкам доходности, показателей стандартных отклонений доходности (риска) и корреляции между рискованными ценными бумагами. Следовательно, они вкладывают свои средства в рискованные активы таким образом, что в итоге сосредотачивают их в своих портфелях в одних и тех же пропорциях.

2. Инвесторам присуще оптимальное поведение. Поэтому на находящемся в равновесии рынке курс ценных бумаг устанавливается таким образом, что если инвесторы владеют оптимальными портфелями ценных бумаг, то совокупный спрос на ту или иную ценную бумагу равняется ее совокупному предложению.

3. Существует безрисковая процентная ставка, по которой инвестор может дать взаймы (т. е. инвестировать) или взять в долг денежные средства.

Исходя из этих предположений, приходим к выводу, что фондовый рынок может находиться в состоянии равновесия только в том случае, если оптимальные пропорции владения ценными бумагами соответствуют пропорциям, в которых активы представлены на рынке.

Рыночный портфель — это портфель, состоящий из всех ценных бумаг, в котором доля каждой соответствует ее относительной рыночной стоимости. Относительная рыночная стоимость ценной бумаги равна ее совокупной рыночной стоимости, деленной на сумму совокупных рыночных стоимостей всех ценных бумаг.

Инвесторы выбирают портфели из одного и того же эффективного множества. Исходное положение *CAPM*: при равновесии на рынке ценных бумаг рыночный портфель *M* как совокупность всех обращающихся на рынке рисковых активов совпадает с оптимальным эффективным для инвесторов касательным портфелем *T*.

Если для всех участников рынка имеется единая безрисковая ставка получения и предоставления займов $r_б$, то в состоянии равновесия оптимальным будет являться портфель, представляющий собой комбинацию рыночного портфеля и безрискового актива.

Это означает, что рыночный и касательный портфели имеют одинаковую структуру, в частности: $\bar{r}_M = \bar{r}_T$, $\sigma_M = \sigma_T$. В зависимости от своей меры неприятия риска инвесторы обладают различными наборами безрисковых и рисковых активов, однако процентное соотношение рисковых ценных бумаг в портфелях инвесторов оказывается для всех них одинаковым.

Таким образом, можно определить касательный портфель как рыночный и обозначать его через *M* вместо *T*. На практике рыночный портфель определяют индексом, например *S&P* 500.

Рыночная линия (*CML*)

В модели *CAPM* простым образом определяется связь между риском и доходностью эффективных портфелей (рис. 13.15).

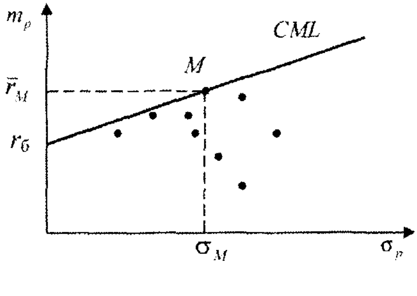


Рис. 13.15. Рыночная линия *CML*

Точка M обозначает рыночный портфель, а $r_б$ представляет собой безрисковую ставку доходности. Эффективные портфели находятся вдоль прямой, пересекающей ось ординат в точке с координатами $(0, r_б)$ и проходящей через точку M, и образуются альтернативными комбинациями риска и доходности, получаемыми в результате сочетания рыночного портфеля с безрисковым заимствованием или кредитованием. Это линейное эффективное множество в *CAPM* известно под названием **рыночная линия** (*Capital Market Line, CML*). Все остальные портфели, не использующие рыночный портфель в комбинации с безрисковым заимствованием или кредитованием, будут лежать ниже рыночной прямой.

Наклон *CML* равен разнице между ожидаемой доходностью рыночного портфеля и безрисковой бумаги $(\overline{r}_M - r_б)$, деленной на разницу их рисков $(\sigma_M - 0)$, или $(\overline{r}_M - r_б)/\sigma_M$. Так как *CML* пересекает вертикальную ось в точке с координатами $(0, r_б)$, то уравнение *CML* имеет вид

$$m_p = r_б + \left(\frac{\overline{r}_M - r_б}{\sigma_M}\right)\sigma_p,$$

где m_p и σ_p обозначают ожидаемую доходность и среднеквадратическое отклонение доходности эффективного портфеля.

Пример 13.16. В примере 13.10 рыночный портфель при безрисковой ставке 4 % состоял из трех акций, взятых в пропорции 0,12 : 0,19 : 0,69. Ожидаемая доходность и среднеквадратическое отклонение доходности для такого портфеля составляли 22,4 и 15,2 % соответственно. Составим уравнение *CML*.

▼Уравнение *CML* есть:

$$m_p = 4 + \left(\frac{22,4-4}{15,2}\right)\sigma_p = 4 + 1,21\sigma_p. \ ▲$$

Состояние равновесия на рынке ценных бумаг может быть охарактеризовано двумя ключевыми величинами. Первая — это ордината точки пересечения *CML* с вертикальной осью (т. е. безрисковая ставка), которую часто называют наградой за ожидание.

Вторая — это наклон *CML*, который называют наградой за единицу принятия риска. По сути, фондовый рынок позволяет осуществлять торговлю временем и риском по ценам, определяемым спросом и предложением. Таким образом, две эти величины можно интерпретировать как цены времени и риска. В примере 13.16 они равны 4 % и 1,21 соответственно.

Упражнение 13.5. В рыночный портфель входят две ценные бумаги, характеристики которых представлены в табл. 13.17.

Таблица 13.17

Исходные данные для упражнения 13.5

Ценная бумага	Ожидаемая доходность, %	Стандартное отклонение, %	Доля в рыночном портфеле
A	10	20	0,4
B	15	28	0,6

При условии, что корреляция этих ценных бумаг составляет 0,3, а безрисковая ставка равна 5 %, определите уравнение рыночной линии.

Ответ: $m_p = 5 + 0{,}39\sigma_p$. ▲

Рыночная линия ценной бумаги (*SML*)

Рыночная линия *CML* представляет собой равновесное соотношение ожидаемой доходности и риска для эффективных портфелей. Отдельные рискованные бумаги всегда будут находиться ниже этой прямой, так как единичная рискованная бумага сама по себе является неэффективным портфелем.

Пусть *P* — портфель, состоящий из i-й ценной бумаги и рыночного портфеля *M* в пропорции t и $1-t$ соответственно. Ожидаемая доходность портфеля *P* и стандартное отклонение будут равны

$$\bar{r}_p = t\bar{r}_i + (1-t)\bar{r}_M\,,\quad \sigma_p = \sqrt{t^2\sigma_i^2 + (1-t)^2\sigma_M^2 + 2t(1-t)\sigma_{iM}}\,,$$

где σ_{iM} — ковариация доходностей i-й ценной бумаги и рыночного портфеля.

На рис. 13.16 изображено достижимое множество модели Марковица, а также безрисковая ставка $r_б$ и соответствующее эффективное множество, представленное рыночной линией (*CML*). Каждой ценной бумаге соответствует точка множества модели Марковица.

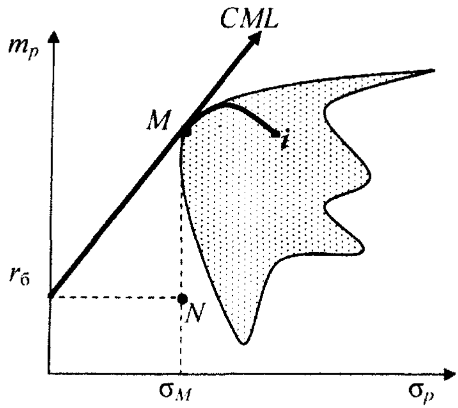


Рис. 13.16. Построение рыночной линии ценной бумаги

Кривая *iM*, отмеченная на рис. 13.16, представляет собой множество портфелей, состоящих из *i*-й ценной бумаги и рыночного портфеля *M*.

Из выражений для $\bar{r}_p$ и σ_p получаем:

$$\frac{d\bar{r}_p}{dt} = \bar{r}_i - \bar{r}_M\,, \quad \frac{d\sigma_p}{dt} = \frac{t\sigma_i^2 - (1-t)\sigma_M^2 + (1-2t)\sigma_{iM}}{\sigma_p}\,.$$

В точке $t = 0$ портфель *P* совпадает с *M*, и выражение для производной $\frac{d\sigma_p}{dt}$ упрощается: $\frac{d\sigma_p}{dt} = \frac{\sigma_{iM} - \sigma_M^2}{\sigma_M}$, откуда $\frac{d\bar{r}_p}{d\sigma_p} = \frac{(\bar{r}_i - \bar{r}_M)\sigma_M}{\sigma_{iM} - \sigma_M^2}$. С другой стороны, из треугольника $r_б MN$ (рис. 13.16) имеем равенство: $\frac{d\bar{r}_p}{d\sigma_p} = \frac{\bar{r}_M - r_б}{\sigma_M}$.

Сравнивая две последние формулы, получаем уравнение

$$m_i = r_б + \left(\frac{\bar{r}_M - r_б}{\sigma_M^2} \right) \sigma_{iM} ,$$

где $m_i = \bar{r}_i$ — ожидаемая доходность ценной бумаги;

σ_{iM} — ковариация доходности этой бумаги с доходностью рыночного портфеля.

Эта зависимость ковариации и ожидаемой доходности известна под названием *рыночная линия ценной бумаги* (*Security Market Line, SML*).

На рис. 13.17, *а* это уравнение описывает прямую, пересекающую вертикальную ось в точке с ординатой $r_б$ и имеющую наклон $(\bar{r}_M - r_б) / \sigma_M^2$. Так как величина наклона положительна, то уравнение указывает на то, что курсы ценных бумаг с большим значением ковариации с рыночным портфелем σ_{iM}, будут обеспечивать большую ожидаемую доходность (m_i).

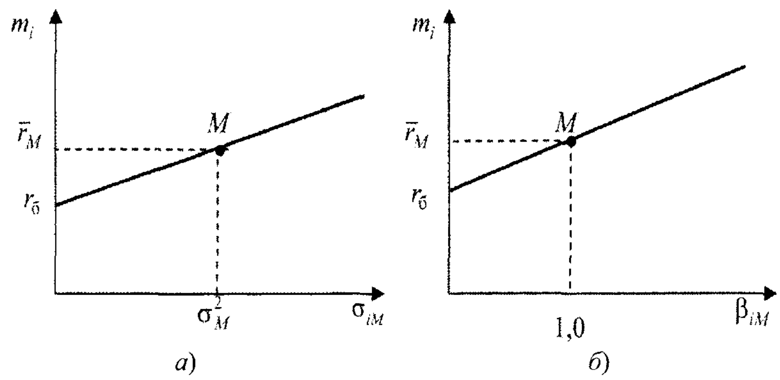


Рис. 13.17. Рыночная линия ценной бумаги:
а — версия ковариации; *б* — версия коэффициента бета

Таким образом, мерой риска ценной бумаги является ковариация ее доходности с доходностью рыночного портфеля.

Рискованная ценная бумага с $\sigma_{iM} = 0$ будет иметь ожидаемую доходность, равную ставке процента безрисковой бумаги $r_б$. Объясняется это тем, что такая рискованная бумага, так же как и безрисковая, не добавляет риска в рыночный портфель.

Рискованная бумага с $\sigma_{iM} = \sigma_M^2$ будет иметь ожидаемую доходность, равную ожидаемой доходности рыночного портфеля $\overline{r}_M$. Это связано с тем, что такая бумага вносит среднюю величину риска в рыночный портфель.

Уравнение *SML* может быть записано также и в следующей форме:

$$m_i = r_б + (\overline{r}_M - r_б)\beta_{iM},$$

где $\beta_{iM} = \sigma_{iM} / \sigma_M^2$.

Величина β_{iM} называется коэффициентом бета (или просто бетой) для бумаги i и является альтернативным способом представления ковариации бумаги.

На рис. 13.17, *а*, *б* приведены графики рыночной линии ценной бумаги в двух формах. Хотя обе прямые пересекают ось ординат в одной и той же точке, они имеют различный наклон. Наклон прямой на рис. 13.17, *а* равен $(\overline{r}_M - r_б) / \sigma_M^2$, а на рис. 13.17, *б* равен $(\overline{r}_M - r_б)$.

Одно из свойств коэффициента бета портфеля заключается в том, что он представляет собой взвешенное среднее коэффициентов бета входящих в него ценных бумаг, где в качестве весов выступают доли инвестиций в эти бумаги.

Выражение для вычисления коэффициента бета портфеля выглядит следующим образом:

$$\beta_{pM} = \sum x_i \beta_{iM}.$$

Ранее было показано, что ожидаемая доходность портфеля представляет собой взвешенную среднюю ожидаемых доходностей входящих в его состав ценных бумаг, где в качестве весов представлены доли инвестирования в эти бумаги. Это означает, что так как каждая бумага лежит на *SML*, то на этой же прямой будет лежать и каждый портфель. Следовательно, эффективные портфели лежат как на *CML*, так и на *SML*, а неэффективные лежат на *SML*, но ниже *CML*.

Следует отметить, что *SML* должна проходить через точку, изображающую рыночный портфель. Значение беты для этой точки равно 1, а ожидаемая доходность равна $\overline{r}_M$.

Величины коэффициентов бета (β_{iM}) в модели *CAPM* и в рыночной модели Шарпа (β_i) сходны по смыслу. Однако в отличие от *CAPM* рыночная модель не является моделью равновесия финансового рынка. Более того, рыночная модель использует рыночный индекс, который в общем случае не охватывает рыночный портфель, используемый в *CAPM*.

Пример 13.17. Пусть ожидаемая доходность рыночного портфеля равна 10 %, безрисковая ставка — 6 %, значения беты для акций *A* и *B* составляют 0,85 и 1,20 соответственно. Составим уравнение *SML*. Определим, каковы равновесные значения ожидаемых доходностей акций *A* и *B*.

▼Уравнение *SML*:

$$m_i = r_б + (\overline{r}_M - r_б)\beta_{iM} = 6 + (10 - 6)\beta_{iM} = 6 + 4\beta_{iM}.$$

Равновесные значения ожидаемых доходностей акций *A* и *B*:

$$m_A = 6 + 4 \cdot 0{,}85 = 9{,}4\ \%;$$
$$m_B = 6 + 4 \cdot 1{,}2 = 10{,}8\ \%. \blacktriangle$$

Пример 13.18. Рыночный портфель состоит из четырех ценных бумаг с характеристиками, представленными в табл. 13.18.

Таблица 13.18

Исходные данные для примера 13.18

Ценная бумага	Ковариация с рынком	Доля в портфеле
A	242	0,2
B	360	0,3
C	155	0,2
D	210	0,3

Исходя из этих данных, подсчитаем стандартное отклонение доходности рыночного портфеля.

▼Среднеквадратическое отклонение доходности для рыночного портфеля равно квадратному корню из взвешенного среднего ковариации доходности всех бумаг с доходностью рыночного портфеля, где в качестве весов выступают доли инвестиций в бумаги, входящие в состав этого портфеля:

$$\sigma_M = \left(\sum_i x_{iM}\sigma_{iM}\right)^{1/2} =$$
$$= \sqrt{0,2\cdot 242 + 0,3\cdot 360 + 0,2\cdot 155 + 0,3\cdot 210} = 15,8\,. \blacktriangle$$

Пример 13.19. Пусть две бумаги *A* и *B* образуют рыночный портфель, причем доля в портфеле и дисперсия доходности соответственно равны 0,39 и 160 для *A* и 0,61 и 340 для *B*. Ковариация доходностей бумаг равна 190. Рассчитаем значение беты для каждой бумаги.

▼Дисперсия доходности рыночного портфеля:

$$\sigma_M^2 = x_1^2\sigma_1^2 + x_2^2\sigma_2^2 + 2x_1x_2\sigma_{12} =$$
$$= 0,39^2 \cdot 160 + 0,61^2 \cdot 340 + 2 \cdot 0,39 \cdot 0,61 \cdot 190 = 241,3.$$

Ковариации каждой бумаги с рыночным портфелем составляют

$$\sigma_{1M} = \sum x_{jM}\sigma_{1j} = (0,39 \cdot 160) + (0,61 \cdot 190) = 178,3;$$
$$\sigma_{2M} = \sum x_{jM}\sigma_{2j} = (0,39 \cdot 190) + (0,61 \cdot 340) = 281,5.$$

Значение беты для каждой бумаги:

$\beta_{1M} = \sigma_{1M}/\sigma_M^2 = 178,3/241,3 = 0,74;$

$\beta_{2M} = \sigma_{2M}/\sigma_M^2 = 281,5/241,3 = 1,17.$ ▲

Пример 13.20. В примере 13.10 заданы ожидаемые доходности трех рисковых акций $\overline{r_1}$ = 16,2 %, $\overline{r_2}$ = 24,6 %, $\overline{r_3}$ = 22,8 % с ковариационной матрицей

$$(\sigma)_{ij} = \begin{pmatrix} 146 & 187 & 145 \\ 187 & 854 & 104 \\ 145 & 104 & 289 \end{pmatrix}.$$

Рыночный портфель при безрисковой ставке 4 % состоял из трех акций, взятых в пропорции 0,12 : 0,19 : 0,69. Как было показано, ожидаемая доходность и ее среднеквадратическое отклонение для такого портфеля составляли 22,4 и 15,2 % соответственно. Составим уравнение *SML*. Вычислим значения рыночного и собственного риска для каждой акции.

▼Уравнение *SML*, в котором мера риска бумаги выражена ковариацией ее доходности с доходностью рыночного портфеля, имеет вид:

$$m_i = r_б + \left(\frac{\overline{r}_M - r_б}{\sigma_M^2}\right)\sigma_{iM} = 4 + \left(\frac{22,4 - 4}{15,2^2}\right)\sigma_{iM} = 4 + 0,08\sigma_{iM}.$$

Ковариации доходностей каждой *i*-й бумаги с доходностью рыночного портфеля вычисляются по формуле $\sigma_{iM} = \sum x_{jM}\sigma_{ij}$ и составляют

$$\sigma_{1M} = \sum x_{jM}\sigma_{1j} = (0,119 \cdot 146) + (0,194 \cdot 187) + (0,689 \cdot 145) = 153,6;$$
$$\sigma_{2M} = \sum x_{jM}\sigma_{2j} = (0,119 \cdot 187) + (0,194 \cdot 854) + (0,689 \cdot 104) = 259,6;$$
$$\sigma_{3M} = \sum x_{jM}\sigma_{3j} = (0,119 \cdot 145) + (0,194 \cdot 104) + (0,689 \cdot 289) = 236,6.$$

С помощью уравнения *SML* можно вычислить ожидаемую доходность каждой акции:

- акция 1: $m_1 = 4 + 0,0795 \cdot 153,6 = 16,2$ %;
- акция 2: $m_2 = 4 + 0,0795 \cdot 259,6 = 24,6$ %;
- акция 3: $m_3 = 4 + 0,0795 \cdot 236,6 = 22,8$ %.

Каждое из полученных значений ожидаемой доходности соответствует определенному заданному значению ожидаемой доходности.

Полученное уравнение *SML* можно использовать для вычисления беты каждой акции: $\beta_{1M} = \sigma_{1M}/\sigma_M^2 = 153,6/15,2^2 = 0,66$; $\beta_{2M} = \sigma_{2M}/\sigma_M^2 = 259,6/15,2^2 = 1,12$; $\beta_{3M} = \sigma_{3M}/\sigma_M^2 = 236,6/15,2^2 =$ $= 1,02$.

Бета рыночного портфеля:

$$\beta_p = \sum x_{iM}\beta_{iM} = 0,12 \cdot 0,66 + 0,19 \cdot 1,12 + 0,69 \cdot 1,02 = 1.$$

Уравнение *SML*, в котором мера риска бумаги выражена ее коэффициентом бета, имеет вид

$$m_i = r_б + (\overline{r}_M - r_б)\beta_{iM} = 4 + (22,4 - 4)\beta_{iM} = 4 + 18,4\beta_{iM}.$$

С помощью этого уравнения также можно вычислить ожидаемые доходности трех акций:

- акция 1: $m_1 = 4 + 18{,}4 \cdot 0{,}66 = 16{,}2$ %;
- акция 2: $m_2 = 4 + 18{,}4 \cdot 1{,}12 = 24{,}6$ %;
- акция 3: $m_3 = 4 + 18{,}4 \cdot 1{,}02 = 22{,}8$ %.

Если в качестве рыночного портфеля выбран другой, отличный от того, в котором акции находились в пропорции 0,12 : 0,19 : 0,69, то равновесного отношения между ожидаемой доходностью и коэффициентами бета (или ковариациями) уже не будет. Например, портфель с равными долями всех трех акций не может быть рыночным.

Значения рыночного риска для каждой акции:

- акция 1: $(\beta_{1M} \cdot \sigma_M)^2 = (0{,}66 \cdot 15{,}2)^2 = 101$;
- акция 2: $(\beta_{2M} \cdot \sigma_M)^2 = (1{,}12 \cdot 15{,}2)^2 = 290$;
- акция 3: $(\beta_{3M} \cdot \sigma_M)^2 = (1{,}02 \cdot 15{,}2)^2 = 241$.

Собственный риск для каждой акции:

- акция 1: $\sigma_{\varepsilon 1}^2 = 146 - 101 = 45$;
- акция 2: $\sigma_{\varepsilon 2}^2 = 854 - 290 = 564$;
- акция 3: $\sigma_{\varepsilon 3}^2 = 289 - 241 = 48$.

Собственный риск в виде стандартного отклонения:

- акция 1: $\sigma_{\varepsilon 1} = \sqrt{45} = 6{,}7$ %;
- акция 2: $\sigma_{\varepsilon 2} = \sqrt{564} = 23{,}7$ %;
- акция 3: $\sigma_{\varepsilon 3} = \sqrt{48} = 6{,}9$ %. ▲

Рыночный риск связан с риском рыночного портфеля и значением коэффициента бета данной ценной бумаги. Для бумаги с бóльшими значениями беты значение рыночного риска больше. В рамках модели *CAPM* у таких бумаг также больше ожидаемые доходности. Отсюда следует, что ценные бумаги с бóльшими значениями рыночного риска должны иметь бóльшие ожидаемые доходности.

Нерыночный риск не связан с бетой. Поэтому увеличение собственного риска не ведет к росту ожидаемой доходности. Итак, согласно *CAPM* инвесторы вознаграждаются за рыночный риск, но их нерыночный риск не компенсируется.

Согласно *CAPM* цена актива будет изменяться до тех пор, пока он не окажется на *SML*. На практике можно обнаружить активы, которые неверно оценены рынком относительно уровня его равновесной ожидаемой доходности.

Показатель, который говорит о величине переоценки или недооценки актива рынком, называется **альфой**. Альфа представляет собой разность между фактической ожидаемой доходностью актива и равновесной ожидаемой доходностью, т. е. доходностью, которую требует рынок для данного уровня риска.

Положительное значение альфы свидетельствует о *недооценке* актива, отрицательное — о *переоценке*. Для равновесной ожидаемой доходности альфа равна нулю (такие активы называются справедливо оцененными).

В частности, одна из задач финансового анализа состоит в нахождении недооцененных рынком бумаг и в рекомендации инвестору приобретать их.

На рис. 13.18 представлены два актива A и B, которые неверно оценены рынком по отношению к уровню их риска.

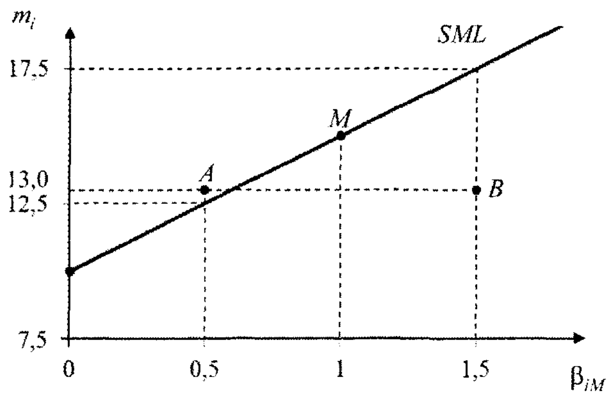


Рис. 13.18. Положение активов, неверно оцененных рынком

Согласно SML доходность A в условиях равновесия должна составлять 12,5 %, фактическая оценка — 13 %, альфа = +0,5 %, т. е. актив недооценен.

Ожидаемая доходность B должна составлять 17,5 %, фактическая оценка — 13 %, альфа = –4,5 %, т. е. актив переоценен.

Пример 13.21. Известны доходности рискового актива r_i и рыночного индекса r_M на фондовой бирже за последние десять месяцев (табл. 13.19), а также безрисковая ставка доходности $r_б = 15$ %.

Таблица 13.19

Исходные данные для примера 13.21, %

Показатель	Месяц										Среднее значение
	1	2	3	4	5	6	7	8	9	10	
r_i	17	18	19	17	16	20	23	22	17	16	18,5
r_M	16	17	17	18	18	19	21	20	18	16	18

Запишем уравнения характеристической линии и рыночной линии ценной бумаги и сделаем выводы.

▼Оценив регрессию между r_i, r_M, получим уравнение характеристической линии данной ценной бумаги:

$$m_i = -4 + 1,25\overline{r}_M,$$

где m_i, $\overline{r}_M$ — ожидаемая доходность акции данного вида и рыночного индекса соответственно.

Поскольку $\beta_i = 1,25 > 1$, то бумага является агрессивной, ее доходность растет быстрее, чем доходность рыночного индекса.

Уравнение рыночной линии ценной бумаги есть

$$m_i = r_б + (\overline{r}_M - r_б)\beta_{iM} = 15 + (18 - 15)\beta_{iM} = 15 + 3\beta_{iM}.$$

Ожидаемое значение доходности данной ценной бумаги $m_i = 15 + 3 \cdot 1,25 = 18,75$. Фактическая средняя доходность актива составляет $\overline{r}_i = 18,5$.

Поскольку альфа = 18,5 – 18,75 = –0,25 % < 0, то эта бумага переоценена рынком. ▲

13.6. Арбитражная теория ценообразования (*APT*)

Модель *CAPM* является равновесной моделью, объясняющей, почему различные ценные бумаги обладают разными ожидаемыми доходностями.

Эта модель образования цен на финансовые активы, в частности, утверждает, что ценные бумаги обладают различными доходностями вследствие различных коэффициентов бета. Однако существует альтернативная модель ценообразования, разработанная Стефаном Россом, известная как теория арбитражного ценообразования (*APT*).

Теория *APT* является также равновесной моделью цен на финансовые активы, но исходит из меньшего числа предположений о поведении инвестора, чем *CAPM*. Главным предположением теории *APT* является то, что каждый инвестор стремится использовать возможность увеличения доходности своего портфеля без увеличения риска. Механизмом, способствующим реализации данной возможности, является арбитражный портфель.

Арбитраж — это получение безрисковой прибыли путем использования разных цен на одинаковую продукцию или ценные бумаги. Арбитраж, являющийся широко распространенной инвестиционной тактикой, обычно состоит из продажи ценной бумаги по относительно высокой цене и одновременной покупки такой же ценной бумаги по относительно низкой цене.

Арбитражная деятельность является важной составляющей современных эффективных рынков ценных бумаг. Поскольку арбитражные доходы являются безрисковыми по определению, то все инвесторы стремятся получить такие доходы при каждой возможности. Возможность безрискового арбитража очень кратковременна.

Арбитражные портфели. В соответствии с *APT* инвестор исследует возможности формирования арбитражного портфеля для увеличения ожидаемой доходности своего текущего портфеля без увеличения риска.

APT исходит из предположения о связи доходности ценных бумаг с некоторым количеством неизвестных факторов.

Доходность отдельных видов ценных бумаг, как и фондового рынка в целом, описывается следующим уравнением множественной регрессии:

$$r_i = a_i + b_{i1}F_1 + b_{i2}F_2 + \ldots + b_{ik}F_k + \varepsilon_i, \qquad (13.2)$$

где r_i — доходность i-й ценной бумаги;

F_k — значение k-го фактора;

b_{ik} — коэффициент чувствительности ценной бумаги i к фактору k;

a_i — свободный член уравнения;

ε_i — случайная переменная с нулевым математическим ожиданием и отличной от нуля дисперсией, не зависящая от рассматриваемых факторов.

При формировании арбитражного портфеля следует соблюсти два условия:

• это портфель, который не нуждается в дополнительных ресурсах инвестора. Если через v_i обозначить изменение в стоимости ценной бумаги i в портфеле инвестора (а значит, и ее вес в арбитражном портфеле), то это требование к арбитражному портфелю может быть записано так:

$$\sum_i v_i = 0, \tag{13.3}$$

причем, $v_i > 0$ означает покупку i-й ценной бумаги, а $v_i < 0$ — ее продажу;

• арбитражный портфель не чувствителен ни к какому фактору. Поскольку чувствительность портфеля к фактору является взвешенной средней чувствительностей ценных бумаг портфеля, то это требование арбитражного портфеля может быть записано так:

$$\sum_i v_i b_{ij} = 0,\ j = 1, 2, \ldots, k. \tag{13.4}$$

Если ожидаемая доходность по данному портфелю

$$\bar{r}_p = \sum_i v_i \bar{r}_i > 0, \tag{13.5}$$

то найден арбитражный портфель, т. е. при нулевых инвестициях имеет место положительная ожидаемая доходность.

Инвесторы будут формировать такие арбитражные портфели, пока не будет достигнуто равновесие. Это означает, что равновесие будет достигнуто, когда любой портфель, удовлетворяющий уравнениям (13.3) и (13.4), будет иметь нулевую ожидаемую доходность: $\sum_i v_i \bar{r}_i = 0$.

При этом связь между доходностями и коэффициентами чувствительности будет линейной:

$$\bar{r}_i = \lambda_0 + \lambda_1 b_{i1} + \ldots + \lambda_k b_{ik}, \tag{13.6}$$

где $\lambda_0, \lambda_1, \ldots, \lambda_k$ являются константами.

Уравнение (13.6) представляет собой основное уравнение арбитражной теории ценообразования, которое должно выполняться для каждой рисковой ценной бумаги, фондового рынка в целом и безрискового актива.

Поскольку безрисковая ставка процента не зависит от рассматриваемых факторов, то из условия (13.6) следует, что $\lambda_0 = r_б$.

Если ввести в рассмотрение портфели, каждый из которых чувствителен только к одному из выделенных факторов k и не чувствителен ко всем остальным, то для каждого такого портфеля уравнение (13.6) превращается в однофакторное, причем соответствующий коэффициент λ_k можно представить как

$$\lambda_k = \delta_k - r_б,$$

где δ_k — ожидаемая доходность портфеля, чувствительного к k-му фактору и не чувствительного ко всем остальным.

Следовательно, λ_i является ожидаемой избыточной доходностью (т. е. ожидаемой доходностью сверх безрисковой ставки) портфеля, имеющего единичную чувствительность к фактору. Поэтому λ_i называется *премией за факторный риск*.

Следовательно, уравнение (13.6) можно представить как

$$\overline{r_i} = r_б + b_{i1}(\delta_1 - r_б) + \ldots + b_{ik}(\delta_k - r_б).$$

Из полученного уравнения следует, что премия за риск по каждой ценной бумаге в форме превышения ожидаемой доходности по бумаге над безрисковой ставкой процента, источником которого является каждый рассматриваемый фактор в отдельности, определяется премией за риск по портфелю, чувствительному только к данному фактору.

Уравнение линии рынка ценных бумаг, используемое в модели *CAPM*:

$$\overline{r_i} = r_б + \beta_{iM}(\overline{r_M} - r_б),$$

где $\overline{r_M}$ — ожидаемая доходность рыночного портфеля, фактически является частным случаем уравнения (13.6) с единственным фактором — доходностью рыночного портфеля — и коэффициентом чувствительности $b_i = \beta_{iM}$.

Однофакторная модель. Доходность ценных бумаг определяется в соответствии со следующей однофакторной моделью:

$$r_i = a_i + b_i F + \varepsilon_i . \qquad (13.7)$$

Покажем, что рыночная модель является примером однофакторной модели, в которой фактором служит доходность по рыночному индексу.

Сравнение рыночной модели $r_i = \alpha_{iM} + \beta_{iM} r_M + \varepsilon_{iM}$ с однофакторной моделью (13.7) показывает их очевидное сходство. Смещение из уравнения рыночной модели соответствует значению нулевого фактора в однофакторной модели, наклон аналогичен чувствительности. Доходность по рыночному индексу играет роль единственного фактора. Однако в однофакторной модели могут быть использованы многие другие факторы (предсказанный ВВП, объем промышленной продукции).

В однофакторной модели дисперсия доходности любой ценной бумаги i равняется

$$\sigma_i^2 = b_i^2 \sigma_F^2 + \sigma_{\varepsilon i}^2 , \qquad (13.8)$$

где σ_F^2 — дисперсия фактора F;

$\sigma_{\varepsilon i}^2$ — дисперсия случайной ошибки.

Первый член в правой части уравнения ($b_i^2 \sigma_F^2$) называется факторным риском, а второй ($\sigma_{\varepsilon i}^2$) называется нефакторным риском.

Ковариация доходностей любых двух ценных бумаг i и j равняется

$$\sigma_{ij} = b_i b_j \sigma_F^2 . \qquad (13.9)$$

Дисперсия доходности портфеля задается выражением

$$\sigma_p^2 = b_p^2 \sigma_F^2 + \sigma_{\varepsilon p}^2 ,$$

где

$$b_p = \sum_i^N x_i b_i , \; \sigma_{\varepsilon p}^2 = \sum_i^N x_i^2 \sigma_{\varepsilon i}^2 .$$

Общий риск σ_p^2 любого портфеля можно представить в виде двух компонент: факторного ($b_p^2\sigma_F^2$) и нефакторного ($\sigma_{\varepsilon p}^2$) риска портфеля.

Диверсификация приводит к усреднению факторного риска и уменьшению нефакторного риска.

Предположение о том, что доходности всех ценных бумаг реагируют на единственный общий фактор, значительно упрощает задачу определения касательного портфеля. Для определения его состава инвестор должен оценить все ожидаемые доходности, их дисперсии и ковариации. В однофакторной модели это можно сделать, оценив a_i, b_i и $\sigma_{\varepsilon i}$ для любой из N рискованных ценных бумаг.

Необходимо также иметь ожидаемое значение фактора $\bar{F}$ и его стандартное отклонение σ_F. Используя эти оценки в уравнениях (13.7), (13.8) и (13.9), можно вычислить ожидаемые доходности ценных бумаг, их дисперсии и ковариации. С помощью этих параметров можно определить кривую эффективного множества Марковица. Наконец, отсюда может быть определен касательный портфель для заданной безрисковой ставки.

Общая чувствительность ценных бумаг к фактору устраняет необходимость непосредственного вычисления ковариации между ценными бумагами. Эти ковариации уже учтены в коэффициентах чувствительности ценных бумаг к фактору и дисперсии его доходности.

APT и *CAPM* могут согласовываться друг с другом. Если доходы по ценной бумаге генерируются по факторной модели и выполняется *CAPM*, то коэффициент бета ценной бумаги зависит от чувствительности ценной бумаги к значению фактора и от ковариации фактора и доходности по рыночному индексу, а также рыночного риска:

$$\beta_{iM} = \frac{\mathrm{cov}(F_1, r_M)}{\sigma_M^2} b_i ,$$

где $\mathrm{cov}(F_1, r_M)$ — ковариация между значением фактора и доходностью рыночного портфеля;

σ_M^2 — дисперсия доходности рыночного портфеля.

В случае двухфакторной модели это выражение обобщается:

$$\beta_{iM} = \frac{\operatorname{cov}(F_1, r_M)}{\sigma_M^2} b_1 + \frac{\operatorname{cov}(F_2, r_M)}{\sigma_M^2} b_2,$$

где $\operatorname{cov}(F_1, r_M)$ и $\operatorname{cov}(F_2, r_M)$ означают ковариацию между первым фактором и доходностью рыночного портфеля и между вторым фактором и доходностью рыночного портфеля.

Пример 13.22. Предположим, что инвестор обладает акциями трех видов и текущая рыночная цена каждого его актива равна 4 000 000 долл., следовательно, стоимость инвестированного капитала S_0 равна 12 000 000 долл.

Предположим, что в результате оценки параметров некоторого однофакторного уравнения в форме $r_i = a_i + b_i F + \varepsilon_i$ были установлены следующие данные по трем видам акций (табл. 13.20).

Таблица 13.20

Исходные данные для примера 13.22

Вид акций (i)	Ожидаемая доходность (m_i), %	Коэффициент чувствительности (b_i)
1	15	0,9
2	21	3,0
3	12	1,8

Покажем возможности формирования арбитражного портфеля для увеличения ожидаемой доходности текущего портфеля при нулевых чистых инвестициях без увеличения риска.

▼Если через v_i обозначить изменение в стоимости ценной бумаги i в портфеле инвестора (а значит, и ее вес в арбитражном портфеле), то требования (13.3) и (13.4) к арбитражному портфелю следует записать так:

$$\begin{cases} v_1 + v_2 + v_3 = 0, \\ 0{,}9v_1 + 3{,}0v_2 + 1{,}8v_3 = 0. \end{cases}$$

В данном случае имеются три неизвестные (v_1, v_2, v_3) и два уравнения, что означает существование бесконечного числа комбинаций значений v_1, v_2, v_3, удовлетворяющих этим двум уравнениям. Для того чтобы найти одну комбинацию, предположим, что $v_1 = 0{,}1$. В результате получим два уравнения с двумя неизвестными:

$$\begin{cases} 0{,}1 + v_2 + v_3 = 0, \\ 0{,}09 + 3{,}0v_2 + 1{,}8v_3 = 0. \end{cases}$$

Решением этих уравнений является $v_2 = 0{,}075$ и $v_3 = -0{,}175$. Следовательно, потенциальным арбитражным портфелем является портфель с полученными показателями (долями).

Чтобы определить, является ли портфель арбитражным, необходимо определить его ожидаемую доходность. Если доходность положительна, то портфель является арбитражным. Для данного портфеля ожидаемая доходность равна (15 % · 0,1) + + (21 % · 0,075) – (12 % · 0,175) = 0,975 %. Так как доходность положительна, то данный портфель является арбитражным.

Найденный арбитражный портфель предполагает покупку акций 1-го вида на сумму $v_1S_0 = 0{,}1 \cdot 12\,000\,000 = 1\,200\,000$ долл., 2-го вида — на сумму $v_2S_0 = 0{,}075 \cdot 12\,000\,000 = 900\,000$ долл. и продажу акций 3-го вида на сумму $v_3S_0 = 0{,}175 \cdot 12\,000\,000 = 2\,100\,000$ долл.

Таким образом, этот арбитражный портфель не требует дополнительных долларовых инвестиций и обладает положительной ожидаемой доходностью.

В определенный момент каждый инвестор должен выбрать между: 1) владением как старым, так и новым арбитражным портфелем; 2) владением только новым портфелем. Для этого он может, например, оценить долю акций 1-го вида. Эта доля в старом портфеле равнялась 0,33, а в арбитражном портфеле — 0,10, что в сумме дает 0,43. Заметим, что долларовая стоимость акций 1-го вида в новом портфеле возрастает до 5 200 000 (4 000 000 + + 1 200 000), т. е. их доля равна 0,43 (5 200 000/12 000 000), что совпадает с суммой долей этих акций в старом и арбитражном портфелях.

Аналогично ожидаемая доходность нового портфеля равна ожидаемой доходности старого и арбитражного портфелей, или 16,975 % (16 + 0,975). Ожидаемая доходность нового портфеля также может быть подсчитана с использованием долей акций в новом портфеле и ожидаемой доходности акций: (0,43 · 15 %) + + (0,41 · 21 %) + (0,16 · 12 %) = 16,975 %.

Чувствительность нового портфеля равна 1,9 (0,43 · 0,9 + + 0,41 · 3,0 + 0,16 · 1,8). Это то же самое, что и сумма чувствительностей старого и арбитражного портфелей (1,9 + 0,0).

Определим рискованность нового портфеля. Предположим, что стандартное отклонение доходности для старого портфеля равно 11 %. Дисперсия доходности арбитражного портфеля будет мала, поскольку единственным источником риска является нефактороный риск. Соответственно, дисперсия доходности нового портфеля будет отличаться от дисперсии доходности старого портфеля только вследствие изменения нефакторного риска. Таким образом, можно заключить, что рискованность нового портфеля приблизительно равна 11 %.

В табл. 13.21 приведены данные, иллюстрирующие приведенные выше рассуждения.

Таблица 13.21

Характеристики нового портфеля для примера 13.22

Параметры портфеля	Старый портфель +	Арбитражный портфель =	Новый портфель
Структура:			
x_1	0,333	0,100	0,433
x_2	0,333	0,075	0,408
x_3	0,333	–0,175	0,158
Свойства:			
$\bar{r}_p$, %	16,000	0,975	16,975
b_p	1,900	0,000	1,900
σ_p, %	11,000	≈ 0	≈ 11

Каковы последствия покупки акций 1-го и 2-го видов и продажи акций 3-го вида? Если каждый инвестор будет поступать таким образом, то это повлияет на курсы акций и, соответственно, на их ожидаемые доходности.

Курсы акций 1-го и 2-го видов поднимутся вследствие увеличения спроса, это повлечет за собой падение ожидаемой доходности по ним. Возросшие объемы продаж акций 3-го вида, наоборот, повлекут за собой падение курса этих акций и повышение ожидаемой доходности.

Подобная деятельность по покупке и продаже будет продолжаться до тех пор, пока все арбитражные возможности не будут существенно сокращены или исчерпаны. В этом случае существует близкая к линейной зависимость между ожидаемыми доходностями и чувствительностями

$$\bar{r}_i = \lambda_0 + \lambda_1 b_i, \tag{13.10}$$

где λ_0 и λ_1 являются константами. Это уравнение является уравнением ценообразования для финансового актива в модели *APT*, когда доходы генерируются одним фактором. Отметим, что это уравнение является линейным, т. е. в состоянии равновесия зависимость между ожидаемыми доходностями и чувствительностя*ми линейна.*

В данном примере одним из возможных равновесных сочетаний является $\lambda_0 = 8$ и $\lambda_1 = 4$. Следовательно, уравнением ценообразования будет:

$$\bar{r}_i = 8 + 4b_i.$$

Таким образом, мы придем к следующим равновесным значениям ожидаемых доходностей для акций всех трех видов:

$$\bar{r}_1 = 8 + 4 \cdot 0,9 = 11,6\ \%;$$
$$\bar{r}_2 = 8 + 4 \cdot 3,0 = 20,0\ \%;$$
$$\bar{r}_3 = 8 + 4 \cdot 1,8 = 15,2\ \%.$$

В результате получаем, что ожидаемая доходность акций 1-го и 2-го видов упадет с 15 и 21 % до 11,6 и 20 % соответственно вследствие увеличения покупательного спроса. При этом увеличение предложения акций 3-го вида приведет к повышению их ожидаемой доходности с 12 до 15,2 %. По сути дела, в ситуации равновесия ожидаемая доходность любой ценной бумаги является линейной функцией от чувствительности ценной бумаги к фактору b_i.

Графически линия оценки финансовых активов в модели *APT* представляет собой прямую линию (рис. 13.19).

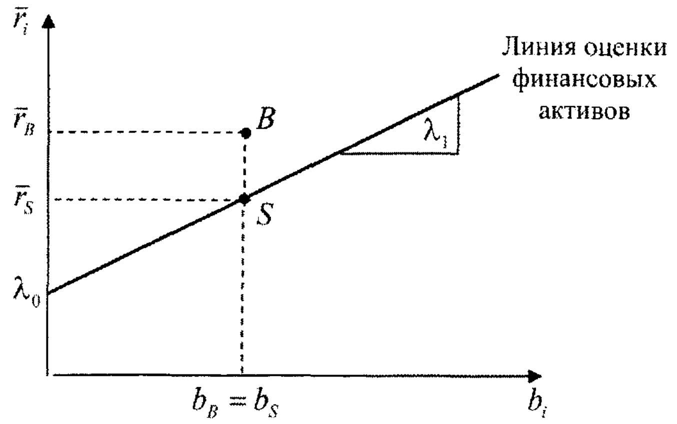


Рис. 13.19. Линия оценки финансовых активов в модели *APT*

Любая ценная бумага, для которой сочетание ожидаемой доходности и чувствительности к фактору лежит вне прямой линии, будет по теории *APT* неправильно оцененной бумагой, что предоставит инвестору возможность сформировать арбитражный портфель. Примером подобной бумаги является ценная бумага *B*.

Если инвестор купит ценную бумагу *B* и продаст ценную бумагу *S* на равные суммы денежных средств, то тем самым он сформирует арбитражный портфель. Во-первых, продавая некоторое количество бумаг *S*, инвестор не прибегает к новым фондам. Во-вторых, поскольку ценные бумаги *B* и *S* обладают одинаковыми чувствительностями к фактору, то продажа бумаг *S* и покупка бумаг *B* приведут к формированию портфеля, нечувствительного к фактору. Таким образом, арбитражный портфель будет обладать положительной ожидаемой доходностью, потому что ожидаемая доходность ценной бумаги *B* больше, чем ожидаемая доходность ценной бумаги *S*. В результате покупок инвесторами бумаги *B* ее цена будет повышаться и, следовательно, ее ожидаемая доходность будет понижаться до тех пор, пока точка, соответствующая характеристикам ценной бумаги *B* не окажется на линии оценки финансовых активов в модели *APT*.

Интерпретация уравнения ценообразования. Если безрисковый актив существует, то ставка доходности такого актива является постоянной величиной. Следовательно, этот актив нечувствителен к фактору. Из уравнения (13.10) следует, что $\overline{r}_i = \lambda_0$ для любого актива, имеющего $b_i = 0$. В случае безрискового актива также известно, что $\overline{r}_i = r_б$ и, следовательно, $\lambda_0 = r_б$.

Подставляя в уравнение (13.10) $r_б$ вместо λ_0, получим

$$\overline{r}_i = r_б + \lambda_1 b_i \qquad (13.11)$$

Рассмотрим чистый факторный портфель p^*, имеющий единичную чувствительность к фактору, т. е. $b_{p^*} = 1$. В соответствии с уравнением (13.10) такой портфель обладает следующей ожидаемой доходностью:

$$\overline{r}_{p^*} = r_б + \lambda_1 \text{, или } \overline{r}_{p^*} - r_б = \lambda_1 .$$

Следовательно, λ_1 является ожидаемой избыточной доходностью (т. е. ожидаемой доходностью сверх безрисковой ставки) портфеля, имеющего единичную чувствительность к фактору. Поэтому λ_1 называется *премией за факторный риск*. Пусть $\delta_1 = \overline{r}_{p^*}$ обозначает ожидаемую доходность портфеля с единичной чувствительностью к фактору, тогда $\lambda_1 = \delta_1 - r_б$.

Подставляя λ_1 в уравнение (13.10), получим вторую версию уравнения ценообразования в модели *APT*:

$$\overline{r}_i = r_б + (\delta_1 - r_б) b_i .$$

Так как в рассматриваемом примере $r_б = \lambda_0 = 8$ % и $\lambda_1 = \delta_1 - r_б = 4$ %, то получаем, что $\delta_1 = 12$ %. Это означает, что ожидаемая доходность портфеля с единичной чувствительностью к первому фактору равна 12 %. ▲

Пример 13.23. Предположим, что в результате оценки параметров некоторого однофакторного уравнения в форме $r_i = a_i + b_i F + \varepsilon_i$ были установлены следующие данные по трем видам акций (табл. 13.22).

Таблица 13.22

Исходные данные для примера 13.23

Вид акций (i)	Ожидаемая доходность (m_i), %	Коэффициент чувствительности (b_i)
1	12	0,6
2	25	2,5
3	8	3,4

Покажем возможности формирования арбитражного портфеля из акций указанных видов, который при нулевых чистых инвестициях обеспечивал бы положительную ожидаемую доходность.

▼Обозначим структуру этого портфеля через $v = (v_1, v_2, v_3)$, где v_1, v_2, v_3 — доли стоимостей акций 1, 2, 3 соответственно. Такой портфель будет арбитражным в том случае, когда он будет иметь положительную ожидаемую доходность. Система уравнений (13.3), (13.4) и (13.5) относительно неизвестных компонент структуры арбитражного портфеля в данном случае имеет вид:

$$\begin{cases} v_1 + v_2 + v_3 = 0, \\ 0,6v_1 + 2,5v_2 + 3,4v_3 = 0, \\ 12v_1 + 25v_2 + 8v_3 = \bar{r}_p. \end{cases}$$

Получили систему из трех линейных уравнений с четырьмя неизвестными v_1, v_2, v_3, $\bar{r}_p$, где $\bar{r}_p$ — ожидаемая доходность портфеля.

Решения системы при различных значениях $\bar{r}_p$ приведены в табл. 13.23.

Таблица 13.23

Возможные структуры арбитражного портфеля для примера 13.23

Ожидаемая доходность арбитражного портфеля ($\bar{r}_p$), %	Доля вида акции в арбитражном портфеле		
	1	**2**	**3**
5	–0,1023	0,3182	–0,2159
10	–0,2045	0,6364	–0,4318
15	–0,3068	0,9545	–0,6477
20	–0,4091	1,2727	–0,8636

В условиях данного примера арбитражная стратегия состоит в том, чтобы продавать акции 1-го и 3-го видов и приобретать на эту же сумму акции 2-го вида. Проверим, что полученный портфель обеспечивает получение дополнительной доходности 5 %. В самом деле

$$\bar{r}_p = 12 \cdot (-0{,}1023) + 25 \cdot 0{,}3182 + 8 \cdot (-0{,}2159) = 5\ \%.$$

Предположим, что инвестор имел портфель из акций указанных видов стоимостью 200 тыс. руб., тогда он должен продать акций 1-го и 3-го вида на сумму $200 \cdot (0{,}1023 + 0{,}2159) = 200 \cdot 0{,}3182 = 63{,}64$ тыс. руб. и купить акции 2-го вида на ту же сумму. Это увеличит ожидаемую доходность его портфеля на 5 %. При этом риск нового портфеля инвестора не изменится, поскольку по условиям формирования арбитражного портфеля его риск относительно мал и близок к нулю.

Если предположить, что исходный (старый) портфель инвестора имел структуру $X = (0{,}4;\ 0{,}2;\ 0{,}4)$ и риск 25 %, то ожидаемая доходность такого портфеля составляла: $\bar{r}_p = 12 \cdot 0{,}4 + 25 \cdot 0{,}2 + 8 \cdot 0{,}4 = 13$ %. Коэффициент чувствительности портфеля к рассматриваемому фактору риска можно определить так: $b_p = 0{,}6 \cdot 0{,}4 + 2{,}5 \cdot 0{,}2 + 3{,}4 \cdot 0{,}4 = 2{,}1$.

Таблица 13.24

Характеристики нового портфеля для примера 13.23

Параметры портфеля	**Старый портфель** +	**Арбитражный портфель** =	**Новый портфель**
Структура:			
x_1	0,4	–0,1023	0,2977
x_2	0,2	0,3182	0,5182
x_3	0,4	–0,2159	0,1841
Коэффициент чувствительности	2,1	0,0	2,1
Ожидаемая доходность, %	13	5	18
Риск, %	25	≈ 0	≈ 25

Используя арбитражный портфель, имеющий ожидаемую доходность, равную 5 %, и нулевую чувствительность, можно сформировать новый портфель (табл. 13.24), основная особенность которого состоит в том, что при том же самом уровне чувствительности и риска достигается более высокая ожидаемая доходность — 18 %.

Стоимость сформированного портфеля равна стоимости исходного, изменилась только структура портфеля, повысилась его ожидаемая доходность. ▲

Упражнение 13.6. В рамках трехфакторной модели рассмотрим портфель, состоящий из трех ценных бумаг, характеристики которых приведены в табл. 13.25.

Определите, каковы коэффициенты чувствительности портфеля к факторам 1, 2 и 3.

Ответ: $b_{p1} = 0{,}28$; $b_{p2} = 4{,}6$; $b_{p3} = 0{,}24$.

Таблица 13.25

Исходные данные для упражнения 13.6

Ценная бумага	Коэффициент чувствительности к факторам			Доля в портфеле
	1	2	3	
A	–0,20	3,60	0,05	0,60
B	0,50	10,00	0,75	0,20
C	1,50	2,20	0,30	0,20

ПРИЛОЖЕНИЯ

Приложение 1

Значения d_1 и d_2 критерия Дарбина — Уотсона при уровне значимости 0,05

(n — число наблюдений,
m — число объясняющих переменных)

n	$m = 1$		$m = 2$		$m = 3$		$m = 4$		$m = 5$	
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
6	0,61	1,40	—	—	—	—	—	—	—	—
7	0,70	1,36	0,47	1,90	—	—	—	—	—	—
8	0,76	1,33	0,56	1,78	0,37	2,29	—	—	—	—
9	0,82	1,32	0,63	1,70	0,46	2,13	—	—	—	—
10	0,88	1,32	0,70	1,64	0,53	2,02	—	—	—	—
11	0,93	1,32	0,66	1,60	0,60	1,93	—	—	—	—
12	0,97	1,33	0,81	1,58	0,66	1,86	—	—	—	—
13	1,01	1,34	0,86	1,56	0,72	1,82	—	—	—	—
14	1,05	1,35	0,91	1,55	0,77	1,78	—	—	—	—
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	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

n	m = 1		m = 2		m = 3		m = 4		m = 5	
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

Приложение 2

Критические значения (односторонние) статистики Дики — Фуллера

Уровень значимости	Размер выборки					
	25	50	100	250	500	∞
AR-модель вида $\Delta y_t = \beta y_{t-1} + \varepsilon_t$						
0,010	–2,66	–2,62	–2,60	–2,58	–2,58	–2,58
0,025	–2,26	–2,25	–2,24	–2,23	–2,23	–2,23
0,050	–1,95	–1,95	–1,95	–1,95	–1,95	–1,95
AR-модель с константой вида $\Delta y_t = \alpha_0 + \beta y_{t-1} + \varepsilon_t$						
0,010	–3,75	–3,58	–3,51	–3,46	–3,44	–3,43
0,025	–3,33	–3,22	–3,17	–3,14	–3,13	–3,12
0,050	–3,00	–2,93	–2,89	–2,88	–2,87	–2,86
AR-модель с константой и трендом вида $\Delta y_t = \alpha_0 + \beta y_{t-1} + \alpha_2 t + \varepsilon_t$						
0,010	–4,38	–4,15	–4,04	–3,99	–3,98	–3,96
0,025	–3,95	–3,80	–3,69	–3,69	–3,68	–3,66
0,050	–3,60	–3,50	–3,45	–3,43	–3,42	–3,41

ЛИТЕРАТУРА

1. *Вентцель Е. С.* Исследование операций. — М.: КноРус, 2010.

2. *Горчаков А. А., Орлова И. В.* Компьютерные экономико-математические модели. — М.: ЮНИТИ, 1995.

3. *Дубров А. М., Мхитарян В. С., Трошин Л. И.* Многомерные статистические методы. — М.: Финансы и статистика, 2011.

4. *Дуброва Т. А.* Статистические методы прогнозирования в экономике. — М.: Издат. центр ЕАОИ, 2008.

5. Исследование операций в экономике: Учеб. пособие для вузов / Под ред. Н. Ш. Кремера. — М.: Юрайт, 2013.

6. *Калихман И. Л., Войтенко М. А.* Динамическое программирование в примерах и задачах. — М.: Высшая школа, 1979.

7. *Кендалл М. Дж., Стьюарт А.* Многомерный статистический анализ и временные ряды: Пер. с англ. — М.: Наука, 1976.

8. *Косоруков О. А., Мищенко А. В.* Исследование операций: Учебник для вузов. — М.: Экзамен, 2003.

9. *Кузнецов Ю. Н., Кузубов В. И., Волощенко А. Б.* Математическое программирование. — М.: Высшая школа, 1980.

10. *Лукашин Ю. П.* Адаптивные методы краткосрочного прогнозирования временных рядов. — М.: Финансы и статистика, 2003.

11. *Магнус Я. Р., Катышев П. К., Пересецкий А. А.* Эконометрика: Начальный курс. — М.: Дело, 2005.

12. *Орлова И. В., Половников В. А.* Экономико-математические методы и модели: компьютерное моделирование: Учеб. пособие. — М.: НИЦ ИНФРА-М, 2014.

13. *Плохотников К. Э.* Основы эконометрики в пакете STATISTICA: Учеб. пособие. — М.: Вузовский учебник, 2010.

14. *Просветов Г. И.* Математические методы и модели в экономике: задачи и решения. — М.: Альфа-Пресс, 2016.

15. *Просветов Г. И.* Управление запасами: задачи и решения. — М.: Альфа-Пресс, 2009.

16. *Таха Х. А.* Введение в исследование операций: Пер. с англ. — 7-е изд. — М.: Вильямс, 2007.

17. *Федосеев В. В., Гармаш А. Н., Орлова И. В.* и др. Экономико-математические методы и прикладные модели: Учеб. пособие / Под ред. В. В. Федосеева. — 4-е изд., перераб. и доп. — М.: Юрайт, 2014.

18. *Шапкин А. С., Мазаева Н. П.* Математические методы и модели исследования операций: Учебник. — 7-е изд. — М.: ИТК «Дашков и К°», 2019.

19. *Шелобаев С. И.* Математические методы и модели в экономике, финансах, бизнесе. — М.: ЮНИТИ, 2001.

20. Экономико-математические методы и модели: Учеб. пособие / Под ред. С. И. Макарова. — М.: КНОРУС, 2016.

Главный редактор — *Т. А. Смирнова*
Редактор, корректор — *А. Д. Бондарь*
Художник — *Т. И. Тактамов*
Верстка — *Н. А. Кирьянова*

Учебное издание

Новиков Анатолий Иванович

Экономико-математические методы и модели

Сертификат соответствия № РОСС RU.AB51.HO5316

Подписано в печать 10.10.2019. Формат 60×90 1/16.
Бумага офсетная № 1. Печ. л. 33,25. Тираж 50 экз. Заказ 145670

Издательско-торговая корпорация «Дашков и К°»
129347, Москва, Ярославское шоссе, д. 142, к. 732
Тел.: 8 (495) 668-12-30, 8 (499) 183-93-23
E-mail: sales@dashkov.ru — отдел продаж:
office@dashkov.ru — офис; http://www.dashkov.ru

Отпечатано: Акционерное общество
«Т8 Издательские Технологии»
109316, Москва, Волгоградский проспект, дом 42, корпус 5
Тел.: 8 (499) 322-38-30

9785394037825